STOCHASTIC DIFFERENTIAL EQUATIONS FROM A MODELING POINT OF VIEW WITH SPECIAL EMPHASIS ON BIOLOGICAL APPLICATIONS

Marjolein van der Vaart Smith

Inst. of Statistics Mimeo Series #1636

ABSTRACT

SMITH, MARJOLEIN VAN DER VAART. Stochastic Differential Equations from a Modeling Point of View with Special Emphasis on Biological Applications. (Under the direction of HARVEY GOLD.)

The stochastization of nonlinear models was studied. The models were assumed to be systems of n first order differential equations. The stochastic processes used as input to these equations were not restricted to white noise but were assumed to be generated by a system of d-n first order differential equations, nonlinear in their solutions but linear in their white noise input. The augmented system was studied as a necessarily degenerate d-dimensional Ito equation.

Some known theorems showing existence and uniqueness of solution processes to Ito systems of stochastic differential equations were extended to accommodate models not restricted by the traditional linear growth conditions yet allowing the above described degeneracy.

Boundary behavior of one-dimensional equations was studied and an easily applied sufficient condition for repelling boundaries was developed.

Numerical techniques were investigated for use with these augmented systems. These included a Monte Carlo simulation technique for use directly with the stochastic differential equations. The Kolmogorov-Smirnov statistic was used to obtain a confidence region for these Monte Carlo generated approximations.

Since the solution processes of Ito equations can also be described by means of partial differential equations (the forward and backward Kolmogorov equations), finite-difference techniques for use with these partial differential equations were investigated for comparison with the Monte Carlo routines.

All these developments were applied to various stochastizations of the logistic equation. The solution processes obtained from these new stochastizations of the logistic equation were shown to be more realistic than those satisfying earlier stochastizations of the equation.

BIOGRAPHY

Marjolein van der Vaart Smith was born in Leiden, The Netherlands, on April 19, 1952. She moved to Raleigh, North Carolina, in 1961. In 1970 she graduated from Needham Broughton High School.

She attended North Carolina State University, where she received a Bachelor of Science degree in Applied Mathematics in May 1974.

She then went to the University of Kentucky at Lexington, where she received a Master of Science degree in Statistics in December 1975.

In 1975 the author returned to North Carolina State University where she received a Master of Science degree in Applied Mathematics with an Ecology minor.

ACKNOWLEDGMENTS

The author wishes to express her deepest appreciation to Professor Harvey Gold, Chairman of her Advisory Committee. She would also like to thank the members of her committee, Professors Frank Benford, John Bishir, Robert Martin, Ronald Stinner, and Robert White.

The author would also like to thank her husband and father for a great deal of moral support.

Finally the author wishes to thank Ann Etheridge and Nancy Evans for their help, and Mrs. Anne Richardson for typing the manuscript.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The modeling process in biology can be thought of as a sequence of models, each hopefully a better tool than the last one for the specific purpose needed. It is not uncommon for such a sequence to start with deterministic differential equations involving parameters that upon closer examination may seem to exhibit stochastic behavior through time. The inclination in such cases is to replace the parameter by a continuous time stochastic process, thereby modeling the phenomenon of interest as a set of stochastic differential equations.

For the most part such stochastic differential equations are dealt with by a variety of ad hoc methods, each with very specific assumptions about the modeling equation(s) as well as the "noise" or stochastic process(es) which replace(s) the parameter(s) [e.g., A. R. Kiester and R. Barakat (1974)].

There is, however, one class of problems for which a body of results and examples exists. A model falls into this class when (1) the stochastic process(es) replacing the parameter(s) is(are) uncorrelated "white noise(s)" to be defined later and when (2) the resulting set of stochastic differential equations can be written as a first order system whose right-hand side is linear in independent white noise(s). That is, if one ends up with the so-called Ito equation:

$$dX = f(X,t)dt + G(X,t)dW \tag{1.1}$$

where $X(.)$, $f(.,.)$, $W(.)$ may be vectors and $G(.,.)$ a matrix. $W(.)$ is the Wiener process which is also called Brownian motion. Note the solution $X(t)$ is now itself a stochastic process.

The Wiener process is a stochastic process whose sample paths with probability one are continuous functions of time but of unbounded variation, i.e., nowhere differentiable. If we let $\xi(t)$ denote the process whose sample paths are the derivatives of the sample paths of the Wiener process, then this $\xi(t)$ cannot exist mathematically nor can it describe any physical process. However, the Fourier transform of its autocorrelation function $E(\xi(t+\tau)\xi(t))$ does formally exist and is a constant function. This constant function can be considered to be the frequency spectrum of a source emitting waves that hit all frequencies equally often. Again this is impossible, but white light has a frequency spectrum that is constant for a while before falling off to zero. For this reason, engineers who first used this process called it white noise.

The equation (1.1) then is written in terms of differentials and is interpreted as the integral equation:

$$X(t) = X_o \int_o^t f(X(\tau),\tau)d\tau + \int_o^t G(X(\tau),\tau)dW(\tau) \qquad (1.2)$$

requiring some meaning to the last integral. This integral does not, in general, exist in the Riemann sense because both $G(X,t)$ and $W(t)$ may be of unbounded variation. This means that if we try to write the integral as the limit of a Riemann sum:

$$\lim_{n\to\infty} \sum_{i=1}^{n} G(X(\tau_i),\tau_i)(W(t_i) - W(t_{i-1})), \text{ where } \tau_i \varepsilon [t_{i-1},t_i] \quad (1.3)$$

the limit will not exist except possibly for particular cases. If we write $G(X(\tau_i),\tau_i)$ as $G(W(\tau_i),\tau_i)$ [Stratonovich (1966)], then the $\lim_{n\to\infty} \sum^{n} G(\lambda W(t_i) + (1-\lambda)W(t_{i-1}), \lambda t_i + (1-\lambda)t_{i-1})[W(t_i) - W(t_{i-1})]$ may exist for each value of λ but have a different value. Each different choice of λ will lead to a different and self-consistent calculus.

The advantage of working with white noise is that the transition probability density has certain properties. The function $p(t,x,t+h,y)$ is defined to be the transition probability density of the process $X(t)$, if $\int_B p(t,x,t+h,y)dy$ is the probability that $X(t+h)$ will have values in set B, assuming that $X(t)$ has the value x. Given restrictions to be specified later on the functions $f(.,.)$ and $G(.,.)$ from equation (1.1), the process $X(t)$ will be Markovian and the transition density will satisfy the following limit equations.

$$\lim_{h\to o} \frac{1}{h} \int_{|x-y| > \varepsilon} p(t,x,t+h,y)dy = 0 \quad (1.4)$$

$$\lim_{h\to o} \frac{1}{h} \int_{|x-y| < \varepsilon} (x_i-y_i)p(t,x,t+h,y)dy = f_i(x,t), \text{ for all } i \quad (1.5)$$

$$\lim_{h\to o} \frac{1}{h} \int_{|x-y| < \varepsilon} (x_i-y_i)(x_j-y_j)p(t,x,t+h,y)dy$$

$$= \frac{1}{2} \sum_{k\leq d} G_{ik}(x,t)G_{kj}(x,t), \text{ for all } i,j \quad (1.6)$$

Equation (1.4) amounts to saying that as h gets small, the probability of being far from where you were at t is small.

Equations (1.5) and (1.6) give values to the first two truncated infinitesimal moments. The first is seen to equal the coefficient of dt in (1.1) and is also called the drift coefficient. The second one is equal to one-half the square of the coefficient of dW in (1.1) and this is called the diffusion coefficient. A continuous Markov process whose transition density satisfies these equations is often called a diffusion process, because its density will then also satisfy two generalized diffusion equations that are derived using the above infinitesimal moment equations, and whose coefficients are, in fact, the right-hand sides of the last two equations. These two partial differential equations are the forward (also called Fokker-Planck) and backward Kolmogorov equations (let $B_{ij}(x,t) = \frac{1}{2}\sum_{k=1}^{d} G_{ik}(x,t)G_{kj}(x,t)$):

$$\frac{\partial p}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left(f_i(x,t)p(s,x_0,t,x)\right) + \lambda \sum_{k=1}^{d}\sum_{j=1}^{d} G_{kj}\frac{\partial G_{ij}}{\partial x_k}(x,t)p(s,x_0,t,x)$$

$$- \sum_{k,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}\left(B_{ij}(x,t)p(s,x_0,t,x)\right) = 0 \quad \text{(forward equation)} \tag{1.7}$$

$$\frac{\partial p}{\partial t} + \sum_{i=1}^{d} f_i(x_0,s)\frac{\partial}{\partial x_{oi}}p(s,x_0,t,x) + \lambda \sum_{i,k,j=1}^{d} G_{kj}\frac{\partial G_{ij}}{\partial x_k}(x,t)\frac{\partial p}{\partial x_{oi}}(s,x_0,t,x)$$

$$+ \sum_{i,j=1}^{d} B_{ij}(x_0,s)\frac{\partial^2}{\partial x_{oi}\partial x_{oj}}p(s,x_0,t,x) = 0 \quad \text{(backward equation)} \tag{1.8}$$

Notice that the density is considered as a function of four variables, two of which may be vectors. That is, $p(s,x_0,t,x)$ is a function of the variables x and t when considered as a solution to

the forward equation, but it is a function of the variables x_0 and s when considered as a solution of the backward equation. We will be mainly interested in the forward equation.

For the modeler starting with a well-defined deterministic system of equations, a major disadvantage of using Ito equations in this way is that, upon stochasticizing the model, he no longer has a well-defined problem. He must somehow pick a value of the λ used in the Riemann sum for the stochastic integral; that is, pick a stochastic calculus with which to interpret the second integral of (1.2). Two common choices are $\lambda = 0$ or $\lambda = 1/2$.

Ito's choice is $\lambda = 0$. It produces a calculus whose rules differ from those of ordinary calculus but has the advantage that the integral under this interpretation will be a martingale (that is, $E(X_t | X_r \text{ for } r \leq s) = X_s$) so that many of the results of martingale theory may be used. $\lambda = 0$ is the only choice that has this property [p. 67, Friedman (1975)].

$\lambda = 1/2$ leads to the Stratonovich calculus. This is the only stochastic calculus whose rules are the same as those of deterministic calculus.

Turelli (1977), Feldman and Roughgarden (1975), Tuckwell (1974) and many others have tried to lay down rules and justifications for choosing one calculus over the others based on the physical system modeled. The results have been mostly the Ito or Stratonovich calculus.

Each of these "rules" leads to problems, because the origin of the ambiguity lies in the nonphysical nature of the noise.

Turelli (1977) admits this but finds that the above mentioned advantages outweigh the problems (p. 142, first paragraph).

Besides the ambiguity of calculus, this approach causes other problems. As mentioned before, the white noise must be entered linearly. This means that the carrying capacity in the logistic, for example, cannot be stochasticized directly but only through 1/k; i.e.,

$$dX = RX(1-X)(Z + (1/k))dt \text{ instead of } dX = RX(1-X/(Z+k))dt,$$

where Z(t) is an appropriately specified random process.

Another problem comes from the fact that, for any fixed t, the Wiener process (and therefore white noise) is normally distributed. This means that with positive probability extremely large values both positive and negative may be attained. Not only are these large values physically unrealizable, most deterministic models are not equipped to handle them.

For example, in the above logistic equations, large negative values for Z would lead to large negative values for the carrying capacity which have the mathematical effect of increasing the growth rate of X!

In this thesis, it is proposed that some of these questions might be resolved by reframing the problem. That is, instead of adding white noise directly to the modeling equation, one could add it to a second equation (or system of equations) whose solutions can then be used as stochastic input (not necessarily linearly) to the modeling equations. We may then end up with a system that looks like this:

$$dX = f(X,Y)dt$$

$$dY = g(Y)dt + h(Y)dW \tag{1.9}$$

Overall this has the form of an Ito equation where what we had called the vector f is now $\begin{bmatrix} f(X,Y) \\ g(Y) \end{bmatrix}$ and the matrix G becomes $\begin{bmatrix} 0 & 0 \\ 0 & h(Y) \end{bmatrix}$. We hope that the nice properties still hold. Since white noise is not directly added to the modeling equation, there will be no ambiguity in its interpretation. There may be trouble of this kind in the second equation, but this equation is not meant to have a physical interpretation at all; we will choose whatever calculus is most convenient for describing the process that we want to use as input. For consistency, we will use the Ito calculus throughout this thesis.

The Y-noise produced in this fashion will also be autocorrelated as well as possibly having a finite range of values depending on the g and h functions.

This thesis will be mainly concerned with the problems associated with this approach.

An important problem concerns those "certain conditions" referred to earlier which ensure that our solution process exists, is unique and is, in fact, a continuous Markov process whose transition probability density satisfies those partial differential equations so easily written down formally.

The most elegant and traditional proofs require that f and G of (1.1) both be bounded by functions linear in the solution process, or that G be nonsingular. Notice that the nonsingularity of G is

equivalent to the non-degeneracy of the Kolmogorov equations; that is, the nonsingularity of the matrix B in (1.7) and (1.8). Also notice that our G, and therefore $B = \frac{1}{2} G^T G$, will always be singular and our f is likely to have at least quadratic terms.

Nevertheless, both of these restrictions have good reasons behind them. Without having the coefficient functions bounded by linear functions, the solution process may with positive probability (or even probability one) become infinite in finite time. This is called an explosion.

It can be dealt with by looking at properties of the process only up to explosion time. Of course, the actual time of explosion will also be a random variable, since different sample paths will hit infinity at different times.

The nonsingularity of G is nice because it ensures that the solution process does not degenerate into a Dirac delta type function, i.e., N-R-dimensional probability density functions on N space. It turns out that the hypoellipticity property of partial differential operators can substitute for the lack of nonsingularity. These problems will be examined in detail in Chapter 2.

In Chapter 3, we will examine exactly what kind of noises we can substitute for the white noise. This includes finding out all we will need to know about these noises, including the exact shape of their densities, their autocorrelation functions, their ranges, etc.

We will examine only one-dimensional input noises so that we can use some of the existing literature for one-dimensional processes.

Finally assuming that the above problems, including that of finding the autocorrelation function, can be satisfactorily dealt with, what can we actually find out about the process we are really interested in, namely the X-component of the solution vector? Our approach makes this more difficult since we will increase the dimensionality of the problem as a whole. Although for very simple cases the probability density of X can still be found analytically, for the most part this problem will require numerical techniques. Here again we run into some problems when we try to apply standard numerical techniques. A rather simple Monte Carlo simulation is used to get rough estimates of the density function. The Kolmogorov-Smirnov statistic is used for a rough sort of error analysis. The numerical problems will be dealt with in Chapter 4.

Since most of the literature dealing with stochastic models in biology use the logistic equation, we will also use this very well-known model to test our techniques. In particular, we will compare the effects of applying noises of different range and auto-correlation to this model. These results and comparisons will be presented in Chapter 5. Finally, we will try to summarize our results and put them into perspective with future projects in Chapter 6.

CHAPTER 2

EXISTENCE AND UNIQUENESS OF SOLUTION PROCESSES

In this chapter, we look more closely at questions of existence and uniqueness of solutions to systems of n first order differential equations,

$$dX = f(X,Y)dt \tag{2.1}$$

where Y is a stochastic process. The right-hand side is not assumed to be linear in stochastic input or bounded by linear growth. We do assume that the right-hand side is continuously differentiable in all n state variables, X, as well as all d-n noise variables, Y. Adjoined to these n equations are d-n first order stochastic differential equations describing the stochastic input, Y. They are assumed to be linear in white noise input (i.e., the first differential of the Brownian motion) and to have continuously differentiable drift and diffusion coefficients.

Although it will not often be directly referred to, we will assume there to be in the background an arbitrary probability space (Ω,F,p) on which our Brownian motion W(t) will be defined. W(0) will be considered to be zero with probability one and W(t) may refer to a vector valued process. As is usual, we will define $\{F_t,t>0\}$ to be an increasing family of sigma-subalgebras of F such that W(t) is F_t-measurable and W(t+h) - W(t) is independent of F_t for all $t \geq 0$ and $h > 0$. Viewed altogether then, we have a d-dimensional system

of first order differential equations whose dxd diffusion matrix is degenerate:

$$\begin{bmatrix} dX \\ dY \end{bmatrix} = \begin{bmatrix} f(X,Y)\,dt \\ 1(Y)\quad dt \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & h(Y) \end{bmatrix} \begin{bmatrix} dW \\ dW \end{bmatrix} \tag{2.2}$$

In the case that the noises themselves are uncoupled, h(Y) will be a d-n x d-n diagonal matrix. The vector (f(X,Y),1(Y)) we will also call b(Z), and the vector (X,Y) we will call Z. The singular matrix $\begin{bmatrix} 0 & 0 \\ 0 & h(Y) \end{bmatrix}$ we will refer to as G(Z).

$$dZ = b(Z)dt + G(Z)dW \tag{2.3}$$

We will show existence and uniqueness of the solution process. We also document the fact that the transition probability density of the solution process is the minimal fundamental solution to the Kolmogorov equations.

We start by quoting a theorem found in the first volume of Friedman (1975), p. 98, theorem 5.1.1:

> Suppose b(z) and G(z) are measurable in z on R^d and
>
> $$|b(z) - b(\bar{z})| \leq K_1|z-\bar{z}|, \quad |G(z) - G(\bar{z})| \leq K_1|z-\bar{z}|$$
>
> $$|b(z)| \leq K_2(1+|z|), \quad |G(z)| \leq K_2(1+|z|), \text{ for all } z, \bar{z} \tag{2.4}$$
>
> where the K_i are constants. Let Z(0) be any d-dimensional random vector independent of the white noise input, such that

$E|Z(0)|^2 < \infty$. Then there exists a unique solution of (2.3) that exists in mean square on [0,T].

Furthermore, under these same conditions, Friedman also shows that the solution process is unique in the sense of probability law, that the solution process is a strong Markov process, and that the solution process is a diffusion process.

To summarize, this theorem applies to our problem in the sense that it holds for degenerate systems like ours. On the other hand, it would not apply to even simple models like the logistic because of the stringent conditions on the growth of the coefficients.

When we look for a result that allows our coefficient functions to grow more steeply, we end up with the following theorem from section 4.3 of McKean's book (1969):

> If G is a positive definite matrix on a manifold M, and the coefficient functions are continuously differentiable, then the local solutions of
>
> $$X(t) = X_0 + \int_0^t G(X)dW + \int_0^t b(X)ds$$
>
> on the patches U of M can be pieced together into a diffusion Z(t) such that the following hold:
>
> (a) the path $Z(t): t \to M$ is defined up to an explosion time, $0 < \ell \leq \infty$,
>
> (b) $\ell = \infty$ if M is compact, while $\lim_{t \to \ell} z(t) = \infty$ if $\ell < \infty$, and M is noncompact. (nota bene, ∞ is the compactifying point of M in the noncompact case),

(c) z begins afresh at its stopping time $\bar{t}$ conditional on the event that the stopping time $\bar{t}$ occurs before the explosion time. A stopping time $\bar{t}$ is a random variable such that the events $\{\bar{t} < s\}$ are measurable with respect to the σ-algebra F_s. That is, if we start a new process at $\bar{t}$ using the same stochastic differential equation and call it $\bar{Z}(t)$, then $\bar{Z}(t) = Z(t+\bar{t})$. (This is the definition of the strong Markov property.),

(d) on the event that $\bar{t} < \ell$ and $\bar{t}$ is a stopping time of Z, and if Z(t) belongs to a patch U with patch map j, then

$$X(t) \equiv j(\bar{Z}) = Z(0) + \int_0^t G(X(s))dW(s) + \int_0^t b(X(s))ds \qquad (2.5)$$

up to the exit time of $\bar{Z}$ from U, for a suitable Brownian motion W depending upon the patch map j. This "matches up" the pieced together diffusion with the local diffusions on the original patches,

(e) the density p(t,z) of the distribution of Z(t) relative to the volume element $(\det[G'G]^{-1})^{1/2}\,dx$ is the smallest elementary solution of the forward Kolmogorov equation with a pole at $Z(0) = z_0$ in M; i.e., it is the smallest function $p \geq 0$ belonging to the class of infinitely differentiable functions on $(0,\infty)xM$ that satisfy the forward Kolmogorov equation and

$$\lim_{t\to o}\int_U p(t,z)\ \sqrt{\det(G'G)^{-1}}\ dz = 1 \qquad (2.6)$$

for each patch U containing z_o.

Here we no longer have the stringent growth conditions on the coefficient functions, but unfortunately we do have a nonsingularity condition on G we cannot hope to meet.

In his introduction to volume 851 of the Springer lecture notes in Mathematics, D. Williams (1981) finds that the (e) part of the result found in McKean can be kept for degenerate Kolmogorov equations provided we have hypoellipticity. To be exact, he states: If the operator:

$$-\frac{\partial p}{\partial t} + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial z_i \partial z_j}[G'G)_{ij}p] - \sum_{i=1}^{d}\frac{\partial}{\partial z_i}(b_i p) \qquad (2.7)$$

is hypoelliptic then the density of the solution to (2.3) is a C^∞ function on $(0,\infty)xR^d$ satisfying the forward Kolmogorov equation and, moreover, is a fundamental solution of this equation with a pole at Z_o. This implies all of (e), since a fundamental solution is one that is smooth, nonnegative, satisfies the equation $\lim_{\varepsilon\to o}\int_R p(\varepsilon,z)dz = 1$, and is the smallest solution having these properties.

Under what conditions would the operator (2.7) be hypoelliptic? Williams answers this question in part, by quoting the following theorem by Hörmander:

Suppose that on some open set in R, an operator can be written as

$$\sum_{j=1}^{r} X_j^2 + X_0 + c$$

where c is an infinitely differentiable function on this set, and $X_0, \ldots, X_r$ are first order homogeneous differential operators again on this set and with C^∞ coefficients. Then if the Lie algebra generated by the X_j contains d linearly independent operators at every point in the open set, the operator is hypoelliptic there.

Note the X_j do not have to be linearly independent at every point, so that they might be chosen in various ways. The trick is to pick the X_j cleverly, and Williams does this as follows. A fuller explanation of what comprises this Lie algebra can be found in Appendix I. To reduce notation, we will assume our G matrix is diagonal. Recall that this corresponds to input noises with uncoupled diffusion coefficients. The same choices for the X_j would hold in the coupled case. For q = 1 to d, let $X_q = \sum_{i=1}^{d} G_{iq}\partial_i = G_{qq}\partial_q$. Note that in our case X_q will actually be zero for q = 1 to n.

$$X_0 = \sum_{i=1}^{d} b_i\partial_i + \frac{3}{2}\sum_{i=1}^{d} G_{ii}\partial_i(G_{ii})\partial_i - \partial_t \equiv \bar{X}_0 - \partial_t \text{ where}$$

$$\partial_i = \frac{\partial}{\partial Z_i} \quad \text{and} \quad \partial_t = \frac{\partial}{\partial t}$$

and

$$c = \frac{1}{2} \sum_{i=1}^{d} \partial_i^2(G_{ii}^2) - \sum_{i=1}^{d} \partial_i(b_i)$$

Our operator then will be hypoelliptic when Lie $(X_1, \ldots, X_d, \bar{X}_0 - \partial_t)$ is full at each point in $(0,\infty)xR^d$. This in turn is true if, and only if, the Lie $(X_1, \ldots, X_d, [X_1, \bar{X}_0], \ldots, [X_d, \bar{X}_0])$ is full at each point in R^d. This second condition turns out to be easily checked in our cases. The details for particular cases are worked out in Appendix I. In particular, for some of our two-dimensional cases, the operator is hypoelliptic when $\frac{\partial}{\partial y}(f(x,y)) \neq 0$. For the logistic equation, this condition fails to take place at $x = 0$, where the population goes extinct. This means the corresponding forward Kolmogorov equation is hypoelliptic away from the line $x = 0$.

We would like to see if we can use the result found in Friedman to extend the proof found in McKean to give us the first four results for our degenerate case.

The range of the (X,Y)-process will be called M in R^d. McKean assumes his process is on an abstract manifold only locally isomorphic to Euclidean space. He then requires the non-degeneracy to patch his local solutions together. I do not require this flexibility of domain space, but do need the degeneracy of the diffusion matrix. Instead of his patches then, I want to consider B_i, concentric spheres about the origin in d-dimensional space having radius i, a positive integer. The range of (X,Y) will be contained in the countable union over all positive i of the intersections between B_i and M.

Suppose the process starts inside B_i, then let ℓ_i be the first exit time from this set, i.e.,

$$\ell_i = \min_s \{ ||(X(s),Y(s))|| = i \},$$

where Euclidean distance is used; ℓ_i will be a stopping time. We next define smooth functions $f_i(x,y)$ such that $f_i(x,y) = f(x,y)$ inside B_i, but is bounded outside of B_i. We define $h_i(y)$ and $l_i(y)$ in the same way. We can clearly do this for each B_i, and can so write a sequence of systems as:

$$dX_i = f_i(X_i,Y_i)dt$$

$$dY_i = l_i(Y_i)\ dt + h_i(Y_i)dW$$

For all $t < \ell_i$, the solutions to each new system will be the same as the solutions to the original one, but since we now have bounded coefficient functions that are continuously differentiable, each of our new systems satisfies the conditions of Friedman's results mentioned earlier. This means that we have unique solutions having the strong Markov property for all $t < \ell_i$, i=1,2 ... . Note that clearly $(X_i,Y_i) = (X_{i-1},Y_{i-1})$ for all $t < \ell_{i-1}$. If with probability one the norm of the solution vector (X,Y) does not go to infinity in finite time, then $p\{\ell=\lim_{n\to\infty} \ell_n = \text{infinity}\} = 1$. In that case, for any finite t, (X,Y) will be contained in $B_m \cap M$ for some m, and for all $s < t$. The system will then have a unique solution process having the strong Markov property and not having explosions. Clearly if M is compact this will always be the case.

If the system "explodes", the norm of (X,Y) will go to infinity in finite time with positive probability. In that event, the time of explosion, $\ell = \lim_{m\to\infty} \ell_m$ will be less than infinity. We need to show that, again on this event, $\lim_{m\to\infty} ||(X(\ell_m),Y(\ell_m))|| = \infty$ with probability one. That is, we need to show that conditional upon the event that the solution process explodes, there are no finite accumulation points.

The only way in which there might be finite accumulation points is if the solution vector bounced back to some inner sphere infinitely often before the explosion time.

McKean argues that this cannot happen by contradiction and using the following a priori bound:

$$p\{\max_{s\leq t} ||(X(s),Y(s)|| \leq R) \leq \exp(-R^2/2d\gamma t)$$

which will hold for all sufficiently small t. This part of the argument, including the proof of the above bound, will hold for our system as well as his, so I will give it only in outline form. Let $\bar{t}_i$ be the ith time the process hits the inner sphere after hitting some outer sphere R units away. Using the above bound and letting $t = 1/n$, we can derive

$$P(\bar{t}_n - \bar{t}_{n-1} \leq \frac{1}{n}) \leq e^{-(R^2 n/2\gamma d)}$$

for all sufficiently large n. This inequality means that

$$\sum_{n=1}^{\infty} P(\bar{t}_n - \bar{t}_{n-1} \leq \frac{1}{n}) < \infty$$

But this means that the event $\{\bar{t}_n - \bar{t}_{n-1} \leq \frac{1}{n}\}$ cannot happen infinitely often by the Borel-Cantelli lemma. Then since there are infinitely many $\bar{t}_n$, it must be that with probability one that the event $\{\bar{t}_n - \bar{t}_{n-1} > \frac{1}{n}\}$ happens for infinitely many n, and we end up with the following inconsistent string of inequalities:

$$\infty > \ell > \lim_{n\to\infty} \bar{t}_n \geq \lim_{n\to\infty} \sum_{m=N}^{n} \bar{t}_m - \bar{t}_{m-1} \geq \text{the tail end of}$$

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

Our degenerate systems then can be considered as producing diffusion processes with all their nice properties as long as the region of positive probability is contained in and surrounded by a region on which the forward Kolmogorov operator is hypoelliptic and the solution to the forward Kolmogorov equation is C^{∞}.

CHAPTER 3

PROPERTIES OF ONE-DIMENSIONAL NOISES

In this chapter, we examine characteristics of noises produced by a one-dimensional Ito equation and some properties we would like these input noises to have. That is, we look at solution processes of

$$dY = g(Y)dt + h(Y)dW \tag{3.1}$$

We then develop some particular noises to use in the later applications.

We want input noises that can be conveniently studied and compared and that conform to some physical interpretation as well. We restrict consideration to stationary noises; that is, stochastic processes whose transition density does not change with time. This will make the mathematics easier and also give results that are more easily interpreted.

Since X does not appear in the equation used to produce the Y-noise, the transition density of the Y-noise (that is, the Y-marginal of the density of the X,Y-process) must satisfy its own lower dimensional Fokker-Planck equation. Furthermore, provided that a stationary solution exists, its density must satisfy the steady state version of the Fokker-Planck:

$$0 = \frac{\partial(g(y)p)}{\partial y} + \frac{1}{2}\frac{\partial^2(h^2(y)p)}{\partial y^2} \tag{3.2}$$

This equation has two independent solutions that can be easily found. To distinguish between them and find out which solution

corresponds to the density of our steady state process, we must consider boundary behavior of the solution process.

There are two different ways in which we can generate behavior of the solution process of a stochastic differential equation at the boundary.

One is to take a stochastic equation that produces a process of known range, then impose barriers on it within that range, and prescribe the behavior of the process when it hits such a barrier. For instance, one can insist that the boundary absorb the "particle" or trajectory, so that the process terminates when the boundary is hit. Since such a boundary is always within the "natural" range of the process, it is always hit eventually.

Imposed boundaries of this type have been discussed a good deal in the literature, starting with the early Feller papers (1952, 1954). More recent work has been done by Anderson and Orey (1976). Often used boundaries of this type are known as absorbing, reflective, and elastic. We will not use imposed boundaries for the following reasons.

First of all, the existence and uniqueness proofs discussed in Chapter 1 do not apply. The papers mentioned above do address precisely this problem, but naturally each of those has its own restrictions on the coefficient functions.

Second, these boundary conditions imposed upon the stochastic differential equations are translated easily into conditions upon only the solution of the backward Kolmogorov equation (Feller, 1952).

Since we are interested in the forward Kolmogorov equation, this is of limited usefulness.

Third, it would be difficult in biological modeling to give a physical reason for preferring one set of imposed boundary conditions to another.

And finally a condition for the existence of a stationary solution mentioned by Prohorov and Rozanov (1969) to be discussed later could not be used.

A second method of generating boundary behavior of a system is to use the "natural range" of the solution process as determined by the stochastic differential equation itself. Existence and pathwise uniqueness would already be established in that case. A result by Yamada and Watanabe (1971) shows that pathwise uniqueness implies uniqueness in law, so that we know there is only one solution of the Fokker-Planck equation that corresponds to the solution process uniquely determined by the stochastic differential equation. This means that all needed boundary conditions to the partial differential equation must come from the coefficients of the stochastic differential equation.

Boundary behavior of this sort is usually described in probabilistic terms. Suppose the range of our process is $[r_1, r_2]$ where either endpoint might be infinite. And suppose that the time of first exit through r_i from this set is τ, that is,

$$\inf_s \{s | y(s) = r_i\} = \tau$$

Then the standard terminology is as follows. The boundary r_i is said to be repelling, if the

$$\text{Prob}(\lim_{s\to\tau} y(s)) = r_i = 0.$$

The boundary r_i is attracting if this probability is positive. Note that the time of first exit may still be infinite here. An attracting boundary with a finite time of first exit is called attainable.

Prohorov and Rozanov (1969) have developed analytical conditions determining which of these boundary types obtains, using just the coefficients of the stochastic equation. These conditions relate to existence of certain integrals built up from the coefficients of the stochastic differential equation.

W. J. Anderson (1976) also has developed what he calls "sufficient conditions for reflecting boundaries." He finds these using just the stochastic equation, under the conditions that g(y) be Lipschitz and that h(y) satisfy the Hölder condition for $\alpha > 1/2$. That is,

$$|h(y) - h(\tilde{y})| \leq k|y - \tilde{y}|^{\alpha}, \text{ for all } y \text{ and } \tilde{y}.$$

He shows that if $h(r) = 0$ and $g(r) > 0$, then if the process starts above r, it will never fall below it. Likewise, r is a upper bound, if $h(r) = 0$ and $g(r) < 0$. Basically this follows from a result he shows earlier, that if m is a Markov time, and $h(y(m)) = 0$ then

$$\lim_{h\to o}[y(m+h) - y(m)]/h = g(y(m))$$

for time step $h > 0$. This result can be used only when looking at what could happen to a process at times after a Markov time. Thus, Anderson can use this result to show that the process cannot pass the boundary r, by assuming that it does and letting the Markov time m be the first time the process passes r. He then gets a contradiction using the above result. He calls this behavior reflective at the boundary.

This is an interesting result but imprecisely described since the word reflective has the following very precise meaning. If a particle or trajectory actually hits a boundary and exhibits behavior that is the reflection of what it would be without the boundary there, that boundary is said to be reflective. As the definition indicates, it is usually reserved for imposed barriers within the range of the process as determined by the stochastic differential equation.

This leads to the natural question: do these assumptions described by Anderson determine a boundary condition that falls into one of the categories of repelling, attainable, etc.? This cannot be determined from Anderson's proof, because to characterize the boundary in this way we must look at what happens before the boundary is hit.

For diffusion processes, this question can be answered by simply applying the following test (Prohorov and Rozanov, 1969):

$$\text{Let} \qquad R(x) = e^{-\int^{x} \frac{2g(\xi)d\xi}{h^2(\xi)}}$$

where $g(.)$ and $h(.)$ are the coefficients of equation (3.1). Then r is a repelling boundary of the solution process to (3.1) if and only

if R(x) is not integrable in an arbitrarily small neighborhood of r.

We need only show then that $\lim_{B\to r} \int_{\bar{y}}^{B} R(x)dx$ does not exist for some $\bar{y} < r$.

We will check this for the case that r is an upper bound. The procedure is entirely similar for the case that r is a lower bound.

By assumption $g(r) < 0$. Since $g(y)$ is continuous, this means there is an $\varepsilon > 0$ and a $y_1 < r$ but sufficiently close to r, so that $g(y) < -\varepsilon < 0$ for $y \in (y_1, r)$.

Also, by the assumptions on $h(y)$, there is a $y_2 < r$ so that $h(y) \leq k(r-y)^{\alpha}$ for $\alpha > 1/2$, and $y \in (y_2, r)$.

If we let $\bar{y} = \max(y_1, y_2)$, then for $y \in (\bar{y}, r)$,

$$\frac{-2g(\xi)}{h(\xi)^2} > \frac{2\varepsilon}{k^2(r-\xi)^{2\alpha}}$$

so that

$$R(x) = e^{-\int^{x} \frac{2g(\xi)}{h^2(\xi)}} > e^{\frac{-2\varepsilon}{k^2} \int^{x} (r-\xi)^{-2\alpha}(-d\xi)}$$

$$= e^{\frac{2\varepsilon}{k^2} \cdot \frac{1}{(2\alpha-1)} \cdot \frac{1}{(r-x)^{2\alpha-1}}} \qquad (2\alpha-1 > 0)$$

Then,

$$\lim_{B\to r^-} \int_{\bar{y}}^{B} R(x)dx = \lim_{B\to r^-} \int_{\bar{y}}^{B} e^{\frac{2\varepsilon}{k^2} \cdot \frac{1}{(2\alpha-1)} \cdot \frac{1}{(r-x)^{2\alpha-1}}} dx$$

Since $2\alpha-1 > 0$, $\exists n$, a positive integer such that $n > \frac{1}{2\alpha-1}$,

and so

$$e^{\frac{2\varepsilon}{k^2} \cdot \frac{1}{2\alpha-1} \cdot \frac{1}{(r-x)^{2\alpha-1}}} > \frac{\left(\frac{2\varepsilon}{k^2}\right)^n (r-x)^{-(2\alpha-1)n}}{(2\alpha-1)^n\, n!}$$

so that

$$\lim_{B\to r^-} \int_{\bar{y}}^{B} R(x)dx > \lim_{B\to r^-} \left[\frac{\left(\frac{2\varepsilon}{k^2}\right)^n (r-x)^{1-(2\alpha-1)n}}{(2\alpha-1)^n\, n!\, [1-(2\alpha-1)n]} \right]_{\bar{y}}^{B}$$

$$= \lim_{B\to r^-} \frac{+(2\varepsilon)^n \left[\frac{1}{(r-B)^{(2\alpha-1)n-1}} - \frac{1}{(r-\bar{y})^{(2\alpha-1)n-1}} \right]}{(k^{2n})(2\alpha-1)^n\, n!\, [(2\alpha-1)n-1]}$$

This limit does not exist since $(2\alpha-1)n-1 > 0$, and in fact grows out of bound to positive infinity.

So, if we have a diffusion process satisfying a stochastic differential equation whose coefficients satisfy Anderson's conditions at a boundary, that boundary will be a repelling one.

Note that if we just have the stochastic differential equation with coefficients satisfying Anderson's conditions, and we do not know a priori that the solution process is a diffusion process, then the above test does not apply. To ensure that a given stochastic differential equation does produce a diffusion process, we must add the conditions from the result of Chapter 2. Specifically, we must replace the Hölder condition of h(y) by the much stronger one of being

locally Lipschitz. Since in this dissertation we are interested only in diffusion processes, the stronger Lipschitz condition will always be met.

These properties of one-dimensional equations will be used mainly to produce easily equations that will generate suitable noise processes to be used as described in Chapter 1.

A diffusion process (as mentioned in Chapter 1) has a transition probability density that satisfies the Fokker-Planck equation. Since we are often interested in producing stationary noise processes, we would like to know if the Fokker-Planck equation has a steady state solution; that is, we consider the equation

$$0 = \frac{\partial}{\partial y}\,(g(y)p(y)) + \frac{\partial^2}{\partial y^2}\,(b(y)p(y)) \qquad (3.3)$$

Recall $b(y) = \frac{1}{2}\,h^2(y)$. Unfortunately, this equation has two independent solutions, both ≥ 0 and, for some $g(y)$ and $h(y)$, both integrable over the range. The two solutions are

$$\frac{1}{b(y)}\,e^{\int^{y} g(x)/b(x)dx} \quad \text{and} \quad \frac{1}{b(y)}\,e^{\int^{y} g(x)/b(x)dx}\int_{y_0}^{y} e^{-\int^{x} g(z)/b(z)dz}\,dx$$

Now suppose we have a process with a non-attainable lower bound r. We will show that the second solution will, in this case, not be integrable over the range of y, so that the density of the stationary solution must be equal to the first solution.

Note first of all that since the second solution must be non-negative over the entire range of y, y_0 must be r. If r is non-attainable, it may be either repelling or attracting but not attainable. If it is repelling, then Prohorov and Rozanov show $e^{-\int^{y} g/b}$ is not integrable in any neighborhood of r. That is,

$$\int_r^y \exp\left(-\int^x g(z)/b(z)^{dz}\right) dx \text{ is not finite for } y > r.$$

By the assumption of the existence of a stationary solution, $\frac{1}{b} e^{\int^{y} g/b}$ is finite for $y > r$ and will be non-zero. Therefore,

$$\frac{e^{\int^{y} g/b}}{b} \int_r^y e^{-\int^{x} g/b} dx$$

does not exist for $y > r$.

Now suppose, r is attractive but non-attainable. Then Prohorov and Rozanov show that

$$e^{-\int^{x} g/b} \int_r^x \frac{e^{\int^{y} g/b}}{b} dy$$

cannot be integrable in a neighborhood of r. Suppose the second solution is integrable; then

$$\int_r^{\bar{y}} \frac{e^{\int^y g/b}}{b} \left(\int_r^y e^{\int^x g/b} dx \right) dy$$

must exist, and we can integrate by parts:

$$\int_r^y e^{\int^x g/b} dx \int_r^y e^{\int^x g/b} dx \Bigg]_r^{\bar{y}} - \int_r^{\bar{y}} \left(e^{-\int^y g/b} \int_r^y \frac{e^{\int^x g/b}}{b} dx \right) dy$$

Of the first term, the first factor exists because we have a stationary distribution, and the second factor exists because r is attractive; however, the second term does not exist by Prohorov and Rozanov's condition for non-attainability. The overall effect of all this is that, if we want a noise with a lower (or upper) bound, it is sufficient to have $h(y) = 0$ at that point, and $g(y)$ have a sign pointing into the domain. Furthermore, in that case, if $\frac{1}{b(y)} e^{\int^y g/b}$ is integrable over the domain, we have a stationary distribution that is actually equal to

$$\frac{1}{b(y)} \exp\left(\int^{y} \frac{g(x)}{b(x)}\, dx\right)$$

times a normalizing constant.

Besides the density function, another characteristic of our noise that we would like to know about is the autocorrelation function, $E(Y_{\tau+t}\, Y_{\tau}) = E(Y_t Y_0)$ in case of stationary processes. To find this, we need to first realize that $p(y_t|y_0)$ satisfies the Fokker-Planck with initial condition $p(y\,|y_0) = \delta(y-y_0)$.

If we multiply both sides of the Fokker-Planck by y_t and integrate over all y_t, we may hope to have an equation for

$$\int_{D_y} yp(y|y_0)dy = E(Y|Y_0)$$

In that case, let $E(Y_t|Y_0) = M(t)$. Then $\int_{D_y} y \frac{\partial p}{\partial t}\, dy =$

$$\frac{\partial}{\partial t}\int_{D_y} yp\, dy = \dot{M}(t) = -\int_{D_y} y \frac{\partial}{\partial y}([g_1 y+g_0]p)\, dy + \int_{D_y} y \frac{\partial^2 (bp)}{\partial y^2}\, dy,$$

where $g(y)$ is now $g_1 y + g_0$, and D_y is the domain of y. Each of these terms we want to integrate by parts, so that we will end up with:

$$- y\,(g_1 y+g_0)p\Big]_{\partial D_y} + \int_{D_y} (g_1 y+g_0)p\,dy + y\,\frac{\partial(bp)}{\partial y}\Big]_{\partial D_y} - \int_{D_y} \frac{\partial(bp)}{\partial y}\,dy$$

$$= +y\,\frac{\partial(bp)}{\partial y} - g(y)p\Big]_{\partial D_y} + g_1\int_{D_y} yp\,dy + g_0\int_{D_y} p\,dy - bp\Big]_{\partial D_y}$$

For noises of finite range, the first term will be zero as can be seen by integrating the Fokker-Planck with respect to y over D_y.

$$\int_{D_y} p\,dy = 1, \text{ which implies } - g(y)p + \frac{\partial(bp)}{\partial y}\Big]_{\partial D_y} = 0$$

The only thing remaining is to see if bp = 0 on the boundary. This will be true for noises with finite range provided that $gh' \neq 0$ at the boundary. In this case, if the coefficients are C-infinity on all D_y (or at least on a set properly containing the set of positive probability), the equation will be hypoelliptic (and the solution smooth) on a set containing the set of positive probability. This means the solution (and all of its derivatives) must be zero on the boundary. Since, if we want to use our previous result, b will be zero on the boundary also, bp will be zero there. In this case, we get an equation for $M(t) = E(Y_t | Y_0)$:

$$M(t) = g_1 M(t) + g_0 ,$$

so that we will have, $M(t) = E(Y_t | Y_0) = y_0[e^{g_1 t}] - \frac{g_0}{g_1}(1 - e^{g_1 t})$.

We can then find the autocorrelation function, since it is $E(Y_t Y_0) = E[Y_0 E(Y_t | Y_0)]$. We have then

$$E(Y_t Y_0) = \frac{-g_0}{g_1} (1 - e^{g_1 t}) E Y_0 + e^{g_1 t} E Y_0^2$$

Since we will be starting with the stationary solution which we know, we will be able to figure out the first two moments.

What about noises of infinite range? We will use the simple Langevin equation

$$dY = (g_1 Y + g_0)dt + h_0 dW$$

to generate our noise of infinite range. This equation can be solved directly and for a given y_0 shown to have a normally distributed solution process with $N(-g_0/g_1, -h_0^2/2g_1)$ as the stationary distribution. In particular, if $g_0 = 0$, $g_1 = -1$, $h_0^2 = 2$, we have the standard normal distribution. This process has been studied in its own right and is often referred to as colored noise or the Ornstein-Uhlenbeck process. For this case again, we are able to calculate the autocorrelation function, since for each t, $p(y_t | y_0)$ will go to zero as the negative exponential, as y_t goes to infinity. The autocorrelation function will be the same as the above provided that the first two moments of the stationary distribution are the same, as well as the parameters g_0 and g_1.

Noises with a semi-infinite range must be treated somewhat differently again. Their finite bound can be treated in general as

those of the noises of finite range. Their infinite bound is more difficult, since we will need to show in general that $\lim_{y\to\infty} y^k p \to 0$, without actually solving for $p(y_t|y_0)$. Fortunately, Friedman (1976) has an estimate at infinity for the solutions of parabolic equations that degenerate on a closed set. To be exact:

If all coefficient functions and their partial derivatives are Hölder continuous in compact subsets of R, and the diffusion coefficient h(y) is positive definite except on a compact set, then if also

$$h^2(y)y^2 \leq C(1+|y|^4), \text{ and}$$

$$\frac{1}{2}h''(y) - g'(y) \leq (\log(2+|y|))^2 \eta(|y|), \text{ where } \eta \to 0 \text{ as } y \to \infty,$$

then for large enough y, the $p(y|y_0) \leq C \exp\left(\frac{-\gamma}{t}(\log y)^2\right)$.

The result applies since we can pick

$$\eta(|y|) = \frac{1}{\log(2+|y|)},$$

which will allow the theorem to be applied to h(y) that are bounded by linear growth. We will again end up with the same autocorrelation function, provided our stationary distribution has the same first two moments, and the same values of the parameters g_0 and g_1.

We now want to use these results to assemble a collection of convenient noises. For comparison purposes, we would like to make their autocorrelation functions the same, if possible. This would involve having the same g_0 and g_1, as well as the same first two moments of the stationary distribution. This may not be possible, since the first two moments will be functions of g_0 and g_1.

CHAPTER 4

NUMERICAL TECHNIQUES

In this chapter, we justify some of the numerical techniques applied and compared in the next chapter. Our goal is to approximate the joint transition density or its antiderivative, the distribution function of the variables in our system. Although we would like to keep in mind our specific examples of finite, semi-infinite, and infinite noise ranges, the collection of results in this chapter is meant to be more generally applicable.

Recall that the system we are interested in can be written as

$$dX = f(X,Y)dt$$

$$dY = g(Y)dt + h(Y)dW \tag{4.1}$$

where X, Y, f, g, and dW may in general all be vectors and h, a matrix. The corresponding Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} (f(x,y)p) - \sum_{j=1}^{d-n} \frac{\partial}{\partial y_j} (g(y)p) + \sum_{i,j \leq d-n} \frac{\partial^2}{\partial y_i \partial y_j} (b(y)p) \tag{4.2}$$

where $b(y) = .5h^T(y)h(y)$. We use the letter G for the dxd matrix,

$$\begin{bmatrix} 0 & 0 \\ 0 & h(Y) \end{bmatrix}$$

The techniques we will look at can be divided into two kinds: numerical methods for partial differential equations (to be applied to the Fokker-Planck equation), and methods like those used for ordinary differential equations (to be used on the stochastic differential equations).

We will start with techniques for partial differential equations. A major advantage of this approach is that it is an application to a deterministic equation, so that we can apply standard nonstochastic techniques such as finite-differences directly. Furthermore, the solution to the partial differential equation is exactly what we are looking for, namely the joint density of all the variables involved. The degeneracy of the partial differential equation will be a problem, even if we can limit ourselves to hypoelliptic operators, which will ensure very smooth solutions. We will be able to circumvent these problems in some cases.

And there are other problems. Remember that although we have existence and pathwise uniqueness (i.e., uniqueness of solution to the stochastic differential equation), we do not have a general way of showing uniqueness of the solution to the partial differential equations or even a standard way of generating boundary conditions from the stochastic differential equations. The remarks made about hypoellipticity in the previous chapter allow us to make the observation that, if the region of positive probability were properly contained in and shared no boundaries with the region of hypoellipticity, then the density (i.e., the solution to the partial differential equation) and all its derivatives would have to be zero on the

boundary. Clearly this is possible only when the range of both the noise and the output of the model are bounded. However, if these conditions are met, we can apply the following weak maximum principle of Friedman (1974) to establish uniqueness:

Suppose we have the following partial differential equation:

$$\sum_{i,j \leq d} G_{ij}(t,z) \frac{\partial^2 p}{\partial z_i \partial z_j} + \sum_{i \leq d} b_i(t,z) \frac{\partial p}{\partial z_i} + c(t,z)p - \frac{\partial p}{\partial t} = 0$$

Let G be a non-negative definite matrix for all (t,z) in a bounded domain Q in $[0,T]xR^d$, and assume that $c(t,z) \leq 0$. If the solution p(t,z) is continuous on the closure of Q, and $\frac{\partial p}{\partial t}$, $\frac{\partial p}{\partial z_i}$, $\frac{\partial^2 p}{\partial z_i \partial z_j}$ are continuous on the interior of Q, and if the right-hand side is 0 in the interior of Q, and the maximum of the solution is positive, then the least upper bound of p over the interior of Q is less than or equal to the maximum of p over the boundary of Q.

Another problem with all methods dealing directly with the partial differential equation is that, although the initial information about the noise will usually be given only as a probability distribution, we often want to investigate cases where the initial value of some of the state variables is taken to be known exactly. This means the joint initial density of all the variables would be non-zero only on a hyperplane. Since the integral over the whole space still has to be one, this means we must have a Dirac-delta

type initial function. This would need to be approximated by a somewhat smoother legitimate function, before we could apply the partial differential equation techniques. Furthermore, the accuracy of the result would depend on how smooth we made the approximation.

In approximating the partial differential equation by a difference equation, we can consider both explicit and implicit schemes. In both cases, error analysis is incomplete because of non-constant coefficient functions. In both cases, the unknown solution is approximated by a vector of values each of which approximates the solution at a particular grid point. Since our partial differential equations will all be linear in the solution as well as its derivatives, the differencing schemes can be thought of as linear algebraic systems.

Explicit difference methods are those whose entire right-hand side is evaluated at the earlier time step. The linear system has the form

$$p(t+\Delta t) = Ap(t)$$

where A is a matrix derived from the coefficient functions of the original differential equation. The solution vector at the previous time step, p(t), is assumed known, so that this is an easy system to solve. Also, explicit differencing does not require boundary conditions.

Unfortunately explicit methods are unstable, even in the case of constant coefficients, unless the time interval is sufficiently small. This means that although for any one time step the solution to the difference equation may provide a reasonably good approximation

to the solution of the partial differential equation, the error that does exist will grow during repeated applications of the difference scheme. This might happen, for example, if the matrix A has an eigenvalue larger than one. Convenient bounds on the size of the time interval to prevent this can be found only in very simple cases of partial differential equations with constant coefficients or coefficients of slow growth. These bounds are usually proportional to the sum of the squares of the step lengths of the state variables. Working by trial and error in a relatively simple non-degenerate example of the kind of system we are looking at, the explicit scheme turns out to be unstable even at very small time steps.

Furthermore, unless we do use boundary conditions, the region in the domain covered by the approximation shrinks at each time step, so that we need to start with a region much larger than we are ultimately interested in. This becomes a serious drawback if we want to estimate the solution of a system with several variables or the solution at the end of a long time interval. Both are likely events since we are adding noise variables to the system and often like to see numerically if a steady state solution might be approached.

Implicit difference methods have the solution evaluated at the later time step present on both sides of the equation. This means that after solving for the unknown solution vector at the later time step, we have a linear system that looks like

$$Ap(t+\Delta t) = Bp(t)$$

This we can then solve by iterative methods, for example. Implicit

schemes are usually stable for simple equations with constant coefficients, and do not have a shrinking region of approximation. The implicit methods do, however, require boundary conditions. For noises that have either an infinite or a semi-infinite range, we need to approximate the infinite domain by a finite one large enough that we can approximate the values of the solution on its boundary by those at infinity.

This is a problem not only because the boundary values are approximated, but also because it increases greatly the number of grid points at which we want to approximate the solution. This again is a problem especially with multiple dimensions. We will show results of using these methods in the next chapter.

The second approach is to apply numerical techniques to the stochastic differential equations. Although the methods are very similar to those used for deterministic differential equations, convergences must be shown using different norms. For example, Rümelin (1982) shows that, given bounded partial derivatives of the coefficient functions, the general explicit Runge-Kutta approximations converge to the solution of the stochastic differential equation in quadratic mean and uniformly in t. That is, if z(t) is the solution to the stochastic differential equation and $z_h(t)$ is the corresponding approximation,

$$\lim_{h \to o} E(Z(t) - Z_h(t))^2 = 0$$

Specifically we use his next result concerning the equation:

$$dX = a(X,t)dt + \sigma(X,t)dW$$

Given the one-dimensional stochastic equation with the coefficient functions and their partial derivatives up to fourth order bounded and continuous, let the value of the Wiener process be known at discrete times. Then, the best integration scheme has a one-step error $E(X(t) - X_h(t))^2$ of order $C(h^3)$ if

$$\sigma \cdot \frac{\partial a}{\partial X} - \frac{\partial \sigma}{\partial t} - a \frac{\partial \sigma}{\partial X} - \frac{1}{2} \sigma^2 \frac{\partial^2 \sigma}{\partial X^2} \neq 0$$

In this case, either the Heun method or the Milstein method are sufficient.

The boundedness conditions will in effect be determined by the largest number the computer can calculate with. The Heun method is the very simplest predictor-corrector method with an explicit Euler predictor and a Crank-Nicolson corrector. The Milstein method is easier to calculate while having the same properties as the Heun method. It can be written out as follows:

$$X_{i+1} = X_i + (a(X_i,t_i) - \frac{1}{2} \frac{\partial \sigma}{\partial x} (X_i,t_i)\sigma(X_i,t_i))h$$

$$+ \sigma(X_i,t_i)\Delta W_i + \frac{1}{2} \frac{\partial \sigma}{\partial X} (X_i,t_i)\sigma(X_i,t_i)(\Delta W_i)^2 \qquad (4.3)$$

For multidimensional cases, the same results hold provided

$$(DG_i)G_j = (DG_j)G_i \quad \text{for all } i,j < d$$

where the G_i correspond to the columns of our matrix G, mentioned at the beginning of this chapter.

We can easily see that this condition does hold for our case as long as the only coupling between the noises occurs in the drift coefficients, g(Y), and not in h(Y). In that case, the G matrix will be diagonal and both sides of the above equality will be zero. The corresponding error over a finite interval of the above method is $O(h^2)$, i.e., $E(Z_i(t) - Z_{hi}(t))^2 = O(h^2)$, for all components Z_i and Z_{hi}.

Now we would like to use this result to derive a bound on the difference between the distribution function of the solution and that of its approximation. Since we know that the solution to the difference equation converges to that of the differential equation in quadratic mean, we know immediately from the Markov inequality that we also have convergence in probability. That is,

$$P\left(\sqrt{\sum_{i=1}^{d} (Z_{ih} - Z_i)^2} \geq \varepsilon\right) \leq \frac{\sum_{i=1}^{d} E(Z_{ih} - Z_i)^2}{\varepsilon^2} = \frac{O(h^2)}{\varepsilon^2} \qquad (4.4)$$

In fact, it is well known that the distribution of z_h, say F_h, will converge to F, the distribution function of z (Billingsley, 1968). We can use the proof of this last result to find the order of error between the two distribution functions as a function of the time step, h.

From Billingsley, we get the following inequality:

$$P(Z_h \varepsilon A) \leq P\left(\sqrt{\sum_{i=1}^{d} (Z_{ih} - Z_i)^2} \geq \varepsilon\right) + P(Z \varepsilon A_\varepsilon) \qquad (4.5)$$

Here A is any set, and $A_\varepsilon = \{x | d(z,A) < \varepsilon\}$, where we will continue to use Euclidean distance. We will pick $A = \{y | y_i \leq z_i,\ i = 1,d\}$. The above inequality can then be rewritten as

$$F_h(z) \leq \frac{O(h^2)}{\varepsilon^2} + F(z')$$

where $z' = z + \sqrt{d}\,\varepsilon$ since $P(Z \varepsilon A_\varepsilon) < F(z')$. By subtracting F(z) from both sides and applying the mean value theorem (we must here assume that F(z) is differentiable, as it certainly will be if we are dealing with a hypoelliptic operator), we get the following:

$$F_h(z) - F(z) \leq \frac{O(h^2)}{\varepsilon^2} + \sqrt{d} \left(\sum_{i \leq d} \frac{\partial}{\partial z_i} F(\xi) \right) \varepsilon \qquad (4.6)$$

To be able to add absolute value signs, we must also be able to get the inequality

$$F(z) - F_h(z) \leq \frac{O(h^2)}{\varepsilon^2} + \sqrt{d} \left(\sum_{i \leq d} \frac{\partial}{\partial z_i} F(\tilde{\xi}) \right) \varepsilon \qquad (4.7)$$

Naturally the proportionality constants will not be the same in both cases, but we can always pick the larger right-hand side. We can get this inequality by defining the set

$$A_{-\varepsilon} = \{z | d(z,A^c) \geq \varepsilon\}$$

The following inequality is then true:

$$P(Z \varepsilon A_{-\varepsilon}) \leq P\left(\sqrt{\sum_{i=1}^{d} (Z_{ih} - Z_i)^2} \geq \varepsilon \right) + P(Z_h \varepsilon A) \qquad (4.8)$$

Now suppose z'' is defined by $z'' = z - \sqrt{d}\,\varepsilon$, then

$$P(Z \varepsilon A_{-\varepsilon}) \geq P(Z_i \leq z_i'' \text{ for } i = 1, d) = F(z'')$$

The inequality (4.8) can then be rewritten as

$$F(z'') \leq \frac{O(h^2)}{\varepsilon^2} + F_h(z)$$

Rearranging this inequality and adding $F(z)$ to both sides gives us

$$F(z) - F_h(z) \leq \frac{O(h^2)}{\varepsilon^2} + F(z) - F(z'')$$

Application of the mean value theorem then gives the desired result (4.7). We next remark that since this inequality will hold for any $\varepsilon > 0$, we can pick the value most advantageous to ourselves, which will be $\varepsilon = h^{2/3}$, so that the error between F and F_h is of order $h^{2/3}$. This is true at any z uniformly provided $F'(z)$ is bounded.

We must now approximate the density of the solution to the difference scheme using its trajectories. We do this by using the computer to pick a finite sample of these trajectories then compute the frequency function at any particular time we are interested in. That is, we divide the range space into a grid, compute a trajectory up to the time we are interested in, and see which square the coordinates of the trajectory at that time fall into. After doing this for all the trajectories in our sample, we count up how many fell into each square. After normalizing these numbers, we can use them to construct a multi-dimensional "bar graph" that will approximate the density

$p(t,x,y)$. This is the same technique used in constructing a frequency function or histogram. We can clearly use the numbers found in this way to compute the sample distribution function of this sample.

The question now becomes, how many trajectories to pick in our sample to get a "good" enough approximation. To answer this question for the marginal distributions of one-dimensional variables, we use the Kolmogorov-Smirnov statistic. We let F_{hn} be the sample distribution of any one-dimensional marginal distribution calculated in the above manner using a sample of n trajectories; then the quantity $\sup_z |F_h - F_{hn}|$ is the Kolmogorov-Smirnov statistic, whose critical values have been tabulated. In fact, since our sample will likely be larger than 100, we will need only the following asymptotic values (Lindgren, 1968):

Significance levels	.20	.15	.10	.05	.01
Critical values	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

For example, this means that $P(\sup|F_{hn} - F_h| \leq 1.36/\sqrt{n}) = 95\%$.

So we know that if we make the h small enough and the n large enough, we can find a good enough approximation to the distribution function that we want. The principal use of this method, however, will be to get rough approximations relatively quickly. We will see precise comparisons between the two methods in the next chapter.

CHAPTER 5

COMPUTER RUNS

In this chapter, we will examine numerical results from the following three particular noise equations:

$$dY = g_1 \cdot Y dt + b_0(1-Y^2)dW \tag{5.1}$$

$$dY = g_1 \cdot Y dt + b_0(1+Y)dW \tag{5.2}$$

$$dY = g_1 \cdot Y dt + b_0\, dW \tag{5.3}$$

We pick these three equations because, as was shown in Chapter 3, they all have the same autocorrelation function and zero mean. We can also pick values for the parameters b_0 and g_1, in such a way that the means and variances of the respective stationary distributions are the same (see Appendix II). In this way, we can examine the effect on whatever model we are investigating of the noises having different value sets.

Before we go to the higher dimensional system of the model equation together with the noise equation, we describe briefly the numerical difficulties in generating and maintaining the stationary densities of the noise equations by themselves.

The first equation can easily be seen to have a value set contained in [-1,1]. In fact, using some results of Chapter 3, we can see immediately that the boundary points are repelling and that a stationary solution exists as well as what it must look like (see Appendix II for the actual expression for any stationary solution).

The repelling boundaries ensure [see Prohorov and Rozanov (1969)] that the solution process is recurrent or ergodic. This, in turn, means that the stationary density satisfying the steady state Fokker-Planck equation is, in fact, asymptotically stable.

We would hope then that, if we use the computer to find the numerical solution to an approximating difference scheme to a partial differential equation with our steady state solution as an initial condition, we would be able to reproduce some time independent solution to the difference equation (hopefully close to the steady state solution of our differential equation) over as many time steps as we want. The actual equation we want to find a difference scheme for is

$$\frac{\partial p}{\partial t} = (6b_0^2y^2 - 2b_0^2 - g_1)p + (4b_0^2y^3 - 4b_0^2y - g_1y)\frac{\partial p}{\partial y} + \frac{b_0^2}{2}(1 - y^2)^2\frac{\partial^2 p}{\partial y^2} \tag{5.4}$$

This is a parabolic differential equation with singularities at $y = -1$ and $y = 1$. We investigated this equation using finite-difference techniques with different values of a Crank-Nicolson parameter λ.

To find the boundary behavior of the solution to the partial differential equation, we use the following result from the third chapter. If any solution to the stochastic differential equation starts between -1 and +1, it has to stay in that interval with probability one. For our partial differential equation, this means that if we start with an initial function whose support is within [-1,1], then the support of the time dependent solution must also stay inside

the interval [-1,1]. Since the partial differential operator is also hypoelliptic (see Appendix I), we know the solution and all its derivatives with respect to y must be zero on the boundary points, -1 and 1. This gives us more information than we can use for the difference scheme.

To examine more closely what selection of boundary conditions would work best, we ran several numerical experiments for the particular values $b_0 = .5$ and $g_1 = -1$. To begin with, we set the function and its first derivative equal to zero on the left-hand boundary (that is, at y = -1). Clearly this meant that we could impose no restrictions on the rightmost end point. Naturally, the solution vector after one time step was zero at the leftmost grid point (since that is part of the boundary conditions). The effect of having the first derivative be zero there as well is that the solution vector becomes zero at the first two grid points at the time step after that, and so on until the solution is zero at all time steps. Writing out the matrix equations makes it clear that this would happen for any initial condition.

We next set the derivative of the solution function with respect to y equal to zero at both endpoints but left the function value there undetermined. This time the solution grew out of control on the left part of the interval. We then set the function value equal to zero on both endpoints, and left the derivative undetermined. Again the solution grew out of control on the left grid points. This indicated to us that perhaps there was an eigenvalue of the matrix

with modulus greater than one, whose eigenvector had its largest components corresponding to the left grid points of the interval.

To test this, we used as initial solution a vector whose largest components were those corresponding to those grid points and calculated the solution over the next 500 time steps. If there were such an eigenvalue with corresponding eigenvector of this type, we would expect to see a sequence of solution vectors approaching this particular eigenvector. Even if the eigenvalue had several independent eigenvectors, we would expect that the program would pick some combination of those eigenvectors (one that in some sense was the first for the computer to run across) and then settle down on it. Instead we got oscillations repeating roughly every 70 time steps. The oscillations took place in only the leftmost 10 or 20 grid points, however, so that our hypothesis might still be correct but with the eigenvalue of largest modulus possibly being complex.

To investigate this further, we wrote down the difference equations corresponding to (5.4) in matrix notation and investigated as best we could the behavior of the eigenvalues of our linear system using the Gerschgorin disc theorem. To reduce notation, let $A_1(y)$ be the quadratic polynomial multiplying the solution function itself in the partial differential equation (5.4); $A_2(y)$, the cubic polynomial multiplying the first derivative; and $A_3(y)$ the quartic multiplying the second derivative. That is, the equation (5.4) now looks like this:

$$\frac{\partial p}{\partial t} = A_1(y)p + A_2(y)\frac{\partial p}{\partial y} + A_3(y)\frac{\partial^2 p}{\partial y^2} \qquad (5.5)$$

We can now write down the difference scheme, using λ as the Crank-Nicolson parameter.

$$\frac{p(t+\Delta t,y) - p(t,y)}{\Delta t} = \lambda A_1(y)p(t+\Delta t,y) + (1-\lambda)A_1(y)p(t,y)$$

$$+ \lambda A_2(y)\left(\frac{p(t+\Delta t,y+\Delta y) - p(t+\Delta t,y)}{\Delta y}\right)$$

$$+ (1-\lambda)A_2(y)\left(\frac{p(t,y+\Delta y) - p(t,y)}{\Delta y}\right)$$

$$+ \lambda A_3(y)\left(\frac{p(t+\Delta t,y+\Delta y) - 2p(t+\Delta t,y) + p(t+\Delta t,y-\Delta y)}{\Delta y^2}\right)$$

$$+ (1-\lambda)A_3(y)\left(\frac{p(t,y+\Delta y) - 2p(t,y) + p(t,y-\Delta y)}{\Delta y^2}\right) \qquad (5.6)$$

We can split this difference equation up in the following way:

$$\left(1-\lambda[A_1\Delta t - A_2\frac{\Delta t}{\Delta y} - 2A_3\frac{\Delta t}{\Delta y^2}]\right)\ p(t+\Delta t,y) - \lambda[A_2\frac{\Delta t}{\Delta y} + A_3\frac{\Delta t}{\Delta y^2}]p(t+\Delta t,y+\Delta y)$$

$$- \lambda A_3\frac{\Delta t}{\Delta y^2}\ p(t+\Delta t,y-\Delta y) = p(t,y) + (1-\lambda)[A_1\Delta t - A_2\frac{\Delta t}{\Delta y} - 2A_3\frac{\Delta t}{\Delta y^2}]p(t,y)$$

$$+ (1-\lambda)[A_2\frac{\Delta t}{\Delta y} + A_3\frac{\Delta t}{\Delta y^2}]p(t,y+\Delta y) + (1-\lambda)A_3\frac{\Delta t}{\Delta y^2}\ p(t,y-\Delta y) \qquad (5.7)$$

We can rewrite this in matrix notation as follows:

$$(I-\lambda A)\bar{p}(t+\Delta t) = (I+[1-\lambda]A)\bar{p}(t) \qquad (5.8)$$

where $\bar{p}(.)$ is used for the vector approximating the solution function at a given time. The matrix A has $A_1(y)\Delta t - A_2(y)\frac{\Delta t}{\Delta y} - 2A_3\frac{\Delta t}{\Delta y^2}$ down the diagonal, $A_2(y)\frac{\Delta t}{\Delta y} + A_3(y)\frac{\Delta t}{\Delta y^2}$ along the superdiagonal, and $A_3(y)\frac{\Delta t}{\Delta y^2}$ along the subdiagonal. Naturally since the $A_1(y)$ are not

constant, these elements will be different in each row. Remember also that we are using for our boundary conditions zero values for the solution vector on the boundaries, $y = +1$ and $y = -1$.

To have a stable solution then, the matrix $(I-\lambda A)^{-1}(I+(1-\lambda)A)$ must not have any eigenvalues of modulus larger than 1. Since we clearly do not wish to calculate $(I-\lambda A)^{-1}$, we will translate this condition into one on $I-\lambda A$. Using algebra we can show that this means we want the eigenvalues of $I-\lambda A$ to be larger than one when $\lambda = .5$. For the particular case we looked at, that is parameter values $b_0 = .5$ and $g_1 = -1$, we calculated the Gerschgorin discs for the $I-\lambda A$ matrix.

Recall that the Gerschgorin theorem states that if the diagonal element of each row is considered to be the center of a circle and the sum of the absolute values of the off-diagonal elements its radius, then the union of all these circles (or discs) must contain all the eigenvalues of the matrix.

The diagonal elements of our $I-\lambda A$ matrix start to the left of 1 with the first row; then with each successive row, the centers of the circles move to the right, towards 1 and past it. This suggests that it is possible that $I-\lambda A$ may have eigenvalues less than one that could be eliminated by throwing out the first several rows.

Recall that this linear system came from the finite-difference scheme by approximating the solution to the differential equations by a vector whose components approximate the true solution at each grid point on the interval [-1,1]. "Taking away" the first five rows of the the matrix $I-\lambda A$ would amount to approximating the solution function by the zero function on the grid points closest to the boundary points

-1 and +1. Geometrically this is plausible since having all derivatives zero on the boundary would force the solution to be very flat near the boundary.

When we tried this, we did get the desired result. That is, the wild behavior at the left-hand boundary was eliminated and a solution that seemed to be numerically stable (it was reproduced for more than 100 steps) was found that approximated the true steady state solution of the differential equation.

When we tried to produce or reproduce the stationary solution to the noise equation using Monte Carlo techniques, we ran into the following problem. We want to do the same thing here as we did with the finite-difference technique; i.e., starting with the known steady state solution as an initial condition, we want to use the stochastic differential equation machinery to produce a stationary noise process with the desired autocorrelations.

The problem we run into with the noise produced by the first equation (5.1) is that the steady state solution to the Fokker-Planck is not a known distribution. This means that there are no random number generators that would pick y_0 with that distribution. This means that we must start Y(t) with some other distribution (for example, a degenerate point distribution at zero) and then generate trajectories over a length of time sufficient to approximate the steady state solution. For the parameter values in question, that length of time was one time unit. This length of time could not be shortened by shortening the time step. For 10,000 trajectories, this involved more computer time than did the finite-difference method.

When we applied the same finite-difference technique to the second noise equation, we ran into the same kind of trouble that we did with the first noise equation, namely wild behavior on the left boundary. We again examined the Gerschgorin discs of the matrix $I-\lambda A$ for $\lambda = .5$, and we found that the discs were centered to the left of one (i.e., at $y = .6, .7, .8$) for the first several rows but moving towards and past one for successive rows, although more slowly than was the case with the finite noise generated by the first equation. And, in fact, although setting the values of the solution equal to zero on the first several grid points did eliminate this uncontrolled behavior, we had to use more grid points in this way than before. The approximation to the steady state solution after 400 time steps was also worse than for the solution to the first equation.

Fortunately, we can now use IMSL subroutines to help us in picking the initial values for the Monte Carlo routine in the semi-finite case. From Appendix II, we know that the steady state noise distribution that we want looks like

$$p(y) = \frac{2}{Nb_o^2} \cdot \frac{e^{2g_1/[b_o^2(1+y)]}}{(1+y)^{2-2g_1/b_0^2}} \tag{5.9}$$

If we use the transformation $u = 1/(1+y)$, we end up with the distribution of u being a gamma distribution. This means that we can take gamma distributed random numbers from an IMSL routine, subtract one from their reciprocal, and have properly distributed random numbers with which to start our Monte Carlo simulations. In this way, in fact, we produced very satisfactory results very quickly.

The third equation causes very little trouble for either the finite-difference method or the Monte Carlo procedure. For the finite-difference method, the corresponding Fokker-Planck equation has a constant coefficient of the second order derivative. This means we have a non-singular parabolic equation. The only drawback to this method is that the infinite range of the y-variable must be approximated by a finite interval, sufficiently large to make zero boundary values for the solution function a reasonably good approximation. This means we can end up trying to estimate the solution function over a large number of grid points.

The Monte Carlo method also has few problems; since the steady state solution is a normal density, initial values are easily gotten using existing random generators. The above results are summarized in Table AIII.1 of Appendix III.

We are now ready to apply these noises to the logistic model. That is, we look at the larger system

$$dX = f(X,Y)dt$$

$$dY = g(Y)dt + h(Y)dW \tag{5.10}$$

This system then has the larger Fokker-Planck equation:

$$\frac{\partial p}{\partial t} = -\frac{\partial(f(x,y)p)}{\partial x} - \frac{\partial(g(y)p)}{\partial y} + \frac{1}{2}\frac{\partial^2(h^2(y)p)}{\partial y^2} \tag{5.11}$$

To begin with, we mention the following analytical results. The differential equation

$$dx = Rx(1-x/k) \tag{5.12}$$

has a known analytical solution, even considering both R and k to be functions of time. That solution can be written as

$$X(t) = \frac{X_0}{e^{\int^{t_R}} + X_0\left(\int_0^t \frac{R(\tau)}{k(\tau)} e^{\int^{\tau_R} d\tau}\right) e^{-\int^{t_R}}} \qquad (5.13)$$

Either the growth rate R of the carrying capacity k may be stochasticized and both have been. Tuckwell (1974) stochasticized the differential equation by adding a zero mean white noise process to a constant R. Most investigators of the logistic model have, however, wanted to stochasticize the carrying capacity k by adding a zero mean white noise term to a constant average value of the parameter 1/k. (They stochasticized 1/k rather than the k directly because they had to enter the input process linearly.) As was mentioned in Chapter 1, this produced the unsatisfactory result of the population growing without bound with probability one. Certainly a major reason for this is the fact that the stochasticized parameter 1/k + dw can take on very nonphysical values such as negative numbers with positive probability.

Can we say anything about the solution process X, if we stochasticize the differential equation (5.12) by adding a continuous noise process that is bounded? Yes, by simply looking at the solution (5.13) and considering R to be a constant and $K(t) = 1/k + AY(t)$, where Y(t) is such a noise process with range [-1,1], say. We see that this solution must then be bounded in the following way:

$$\frac{X_0}{e^{-Rt} + X_0(\frac{1}{k} + A)(1-e^{-Rt})} \leq \text{Solution Process } X(t) \leq \frac{X_0}{e^{-Rt} + X_0(\frac{1}{k} - A)(1-e^{-Rt})}$$

Clearly as time goes to infinity, the solution is caught between

$$\frac{1}{\frac{1}{k} + A} \quad \text{and} \quad \frac{1}{\frac{1}{k} - A}$$

Notice that the solution stays bounded if, instead of a noise of finite range, we picked a noise of semi-infinite range, as long as $1/k + Ay$ stays bounded away from zero. This is, in fact, what we would expect physically since $1/k(t)$ could be interpreted as the amount of land required to support a single beast of whatever species we are studying. If we allow $1/k(t)$ to get very large, that would have a negative influence on the population. Mathematically, it means that the population must stay between zero and $\frac{X_0}{e^{-Rt} + X_0(\frac{1}{k} - A)(1 - e^{-Rt})}$, the previous upper bound.

Furthermore, given the framework of a two-dimensional Ito system as outlined before, we need no longer confine ourselves to stochasticizing the $1/k(t)$, but can stochasticize the $k(t)$ directly, since our noise process no longer needs to enter the model equation linearly. So, we can write the model equation as $dX = RX(1 - \frac{X}{k + AY})$ and combine it with any of the three noise equations we want. For this stochastization, the only noise equation that induces a bounded

solution process x(t) is the first one. Using it and the solution formula, we see that the solution process this time must stay between the following bounds:

$$\frac{X_0}{e^{-Rt} + \frac{(1 - e^{-Rt})X_0}{k - A}} \leq \text{Solution Process } X(t) \leq \frac{X_0}{e^{-Rt} + \frac{(1 - e^{-Rt})X_0}{k + A}}$$

so that as time goes to infinity these bounds go to k - A and k + A.

When we try to analyze the solutions of these two-dimensional systems, we want to know about the boundary behavior. In the above three cases, it is clear that the support of the joint density of x and y is bounded in the x-direction (and in the y-direction in the cases of finite noise). That is, for these two-dimensional systems, we again have the region of positive probability included and surrounded by the region of hypoellipticity of the partial differential operator (at least, in the case that we stay bounded away from x = 0). That means that for these two-dimensional systems also, the solution and all its derivatives must be zero on these boundaries.

For the numerical experiments, we picked two as a value for k , and twelve as the value for R. The large R value was picked so that the solution would grow appreciably during relatively short time periods. Although the large R value did mean the time step had to be made smaller, for both methods the trade-off seemed to be favorable as far as computer time went.

The Monte Carlo program proved to be the one more easily extended from the one-dimensional to the two-dimensional case. The

Milstein method applied to the two-dimensional differential equation system reduced to the same difference scheme as used before for the y-equation, together with a simple forward Euler scheme for the x-equation. While we could use $\Delta t = .1$ to generate the approximation to the steady state process, however, we had to decrease this to $\Delta t = .01$, when we added the x-equation. The choice of the smaller time step hinged on the following known result for using the forward Euler method on deterministic logistic equations; namely, that to have a good approximation to the true solution, the quantity $R \cdot \Delta t$ must be significantly less than 1. We chose $\Delta t = .01$ after repeated numerical experiments with the deterministic logistic using $y(t) = 0$.

The finite-difference program produced more problems. The value of the parameter A was chosen to be 1, so that, with y automatically restricted to [-1,1], the X(t)-process would be restricted to [1,3]. The smallest convenient mesh size was 1/40 for both Δx and Δy. These values mean that in going from the noise by itself to the augmented system with the X(t)-process included, the finite-difference method jumps from an 81×81 linear system to a $81^2 \times 81^2$ linear system. Also the large R value in the model equation meant that we needed to decrease Δt to .002 time unit to retain diagonal dominance of the linear system.

Furthermore, since our parabolic equation has no second derivative taken with respect to the x variable, the value of the solution to the difference equation, $p(t+\Delta t,x,y)$ can depend only on values of $p(t,.,y)$ whose second argument is larger than or equal to x. This is because there are no terms containing $p(t,x-\Delta x,y)$ in the difference

scheme. This means that, if the initial density function $p(0,x,y)$ is zero to the right of some x, say, $p(t,x,y)$ must be zero there also for all t. Since the x variable represents a quantity growing with time, this causes problems. One idea is to use a centered difference for the first derivative with respect to x. Unfortunately, this made diagonal dominance even harder to maintain; in fact, the time step Δt had to be made smaller than .0005 unit. This made this solution impractical. For us a more feasible approach was to use an initial distribution that has positive values over the entire domain (although, of course, very small near the boundaries). We used a truncated normal distribution. This allowed us to continue using the .002 step size. This program then seemed to run quite well for seven or eight time steps. We were able to compare its solution to that of the same problem using the Monte Carlo routine at the time .01. The agreement here was quite good. However, the finite-difference solution seemed to be unstable, since about the ninth or tenth time step, obviously wrong values were calculated for some of the grid points. The behavior looked very much like that of a parasitic solution, i.e., an eigenvector of the linear system whose weight in the linear combination of eigenvectors approximating the solution grows disproportionately because of the relatively large size of its eigenvalue. Getting rid of such a parasitic solution for any particular differencing scheme could be very difficult for a system as large as ours and is not attempted here. The same differencing scheme is applied to the other parameter values we need to look at in this chapter to see if this problem is more prominent for some systems than for others.

The comparisons between the finite-difference solutions and the corresponding Monte Carlo solutions are presented in Table AIII.2 of Appendix III.

We next describe and compare results of logistic equations with different parameter values. To begin with, we want to compare the results of the following two ways of stochasticizing the logistic equation. As was mentioned earlier, the second way listed has been preferred in the literature, although only with white noise added directly. The first equation is the one we have worked with above.

$$dX = 12\ X\ (1-X/(2+Y)) \tag{5.14}$$

$$dX = 12\ X\ (1-X(1/2+\tilde{Y})) \tag{5.15}$$

Clearly the two input processes are not the same. To make the results comparable, we would like to make the two processes $1/(2+Y)$ and $1/2 + A\tilde{Y}$ to be as much alike as possible. We do this by comparing their ranges and their first two moments.

Recall that the Y noise we used in (5.14) had a finite range and a zero mean. Nevertheless, the process $1/(2+Y)$ does not have mean 1/2 but approximately .5136. The second process $1/2 + A\tilde{Y}$ we can then write as $.5136 + A\hat{Y}$ where now $\hat{Y}$ has a zero mean.

Since our y process had range [-1,1], we know the process $1/(2+Y)$ must have range [1/3,1]. If we require the same to be true of the process $.5136 + A\hat{Y}$, we see that $A\hat{Y}$ must have range [-.1803, .4864]. Now let A be the constant .1803 and we get that $\hat{Y}$ must have range [-1,2.68].

The formulas for a skewed noise of finite range are difficult to work with, but we can reasonably approximate this range by a noise of semi-infinite range, i.e., use the second noise equation rather than the first.

Now, the second noise equation also has two parameters, g_1 and b_0^2. The g_1, we pick the same as before, so that the autocorrelation functions will be the same if the variances are. This leaves us the second parameter b_0 to adjust to make the second moments equal. That is, we want

$$E(.5136 + .1803\hat{Y})^2 = E(1/(2+Y))^2 \tilde{=} .27124$$

Since $E(\hat{Y})^2 = - b_0^2/(2g_1 + b_0^2)$ (see Appendix II), this means that $- b_0^2/(-2 + b_0^2) = .229$, or $b_0^2 = .373$. And these are the parameters we used for this comparison.

We also want to look at the effect of decreasing the parameter g_1 from -1 to -10. This would make the "carry-over" effect in the Y noise much less severe. When changing the g_1 by some factor, the b_0^2 parameter must be changed by the same factor. If that is done, other characteristics of the noise process (such as the steady state solution and the first two moments) stay the same. This program we then carried out, as the one above, for both finite-differences and the Monte Carlo routine. The finite-difference method, however, required too much computer time to examine change in the transition density with time.

We now present three sequences of graphs showing marginal density of the x-variable at times .01, .05, .10, .15, and .20. These

were all constructed using the Monte Carlo routine. The top graph in each figure represents the density of the solution to equation (5.14) using a Y-process of finite range with parameter $g_1 = -1.0$. The second graph represents the density of the same equation, but with parameter value $g_1 = -10.0$. The third graph represents the density of the solution to equation (5.15) using a Y-process of semi-infinite range with parameters picked exactly as described above.

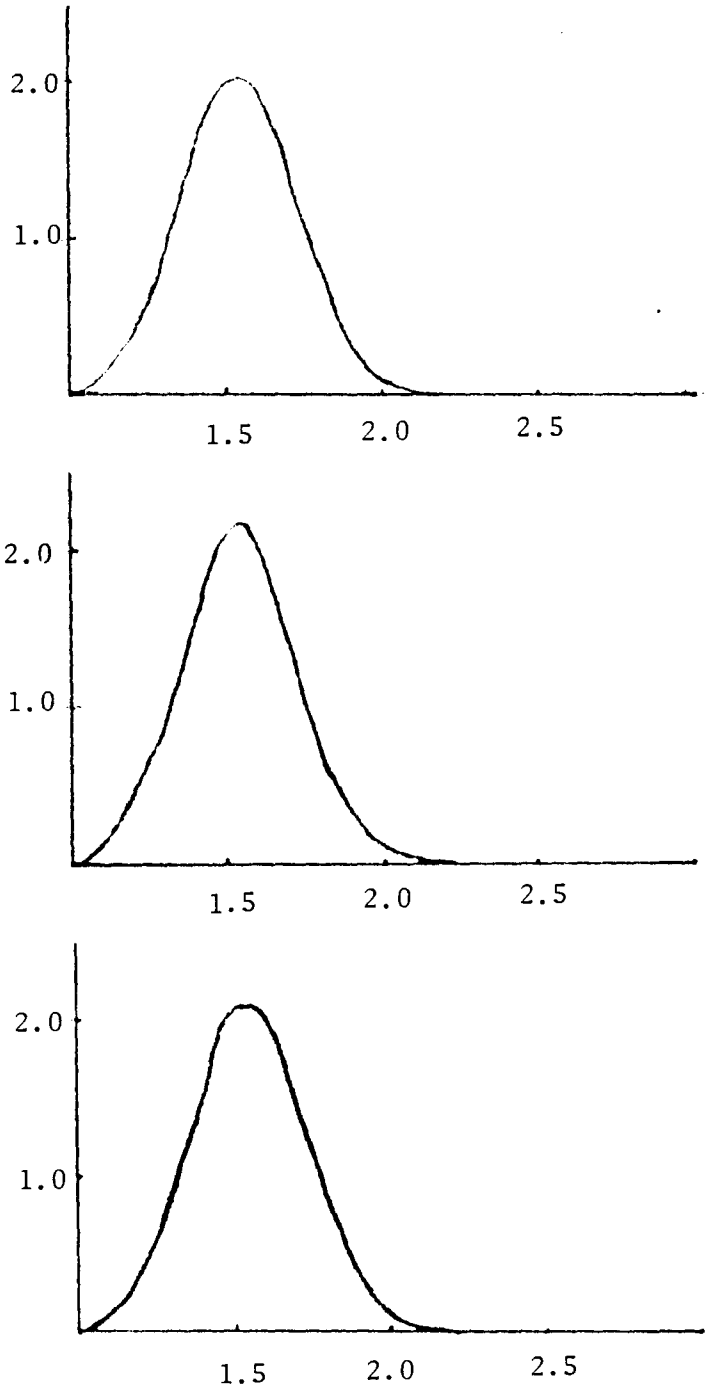

Figure 5.1 Comparison of marginals of the X-process at t = .01.

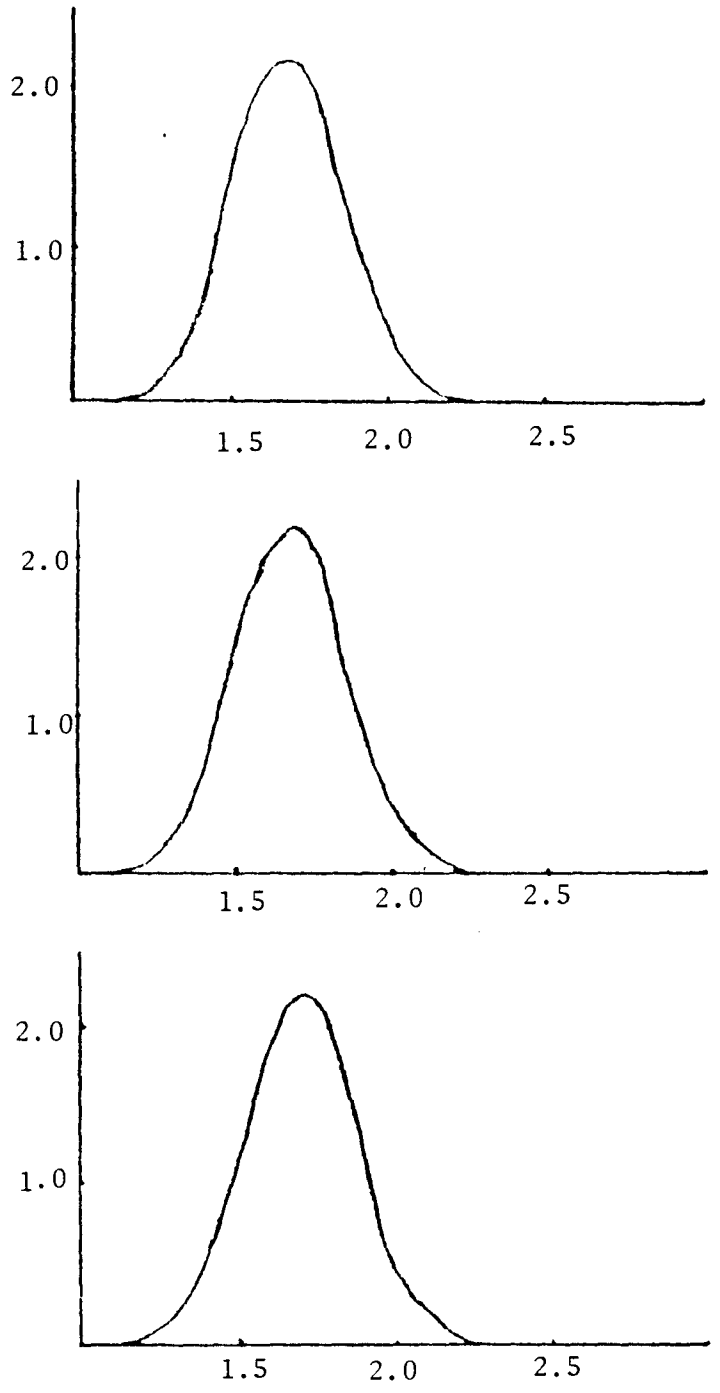

Figure 5.2 Comparison of marginals of the X-process at t = .05.

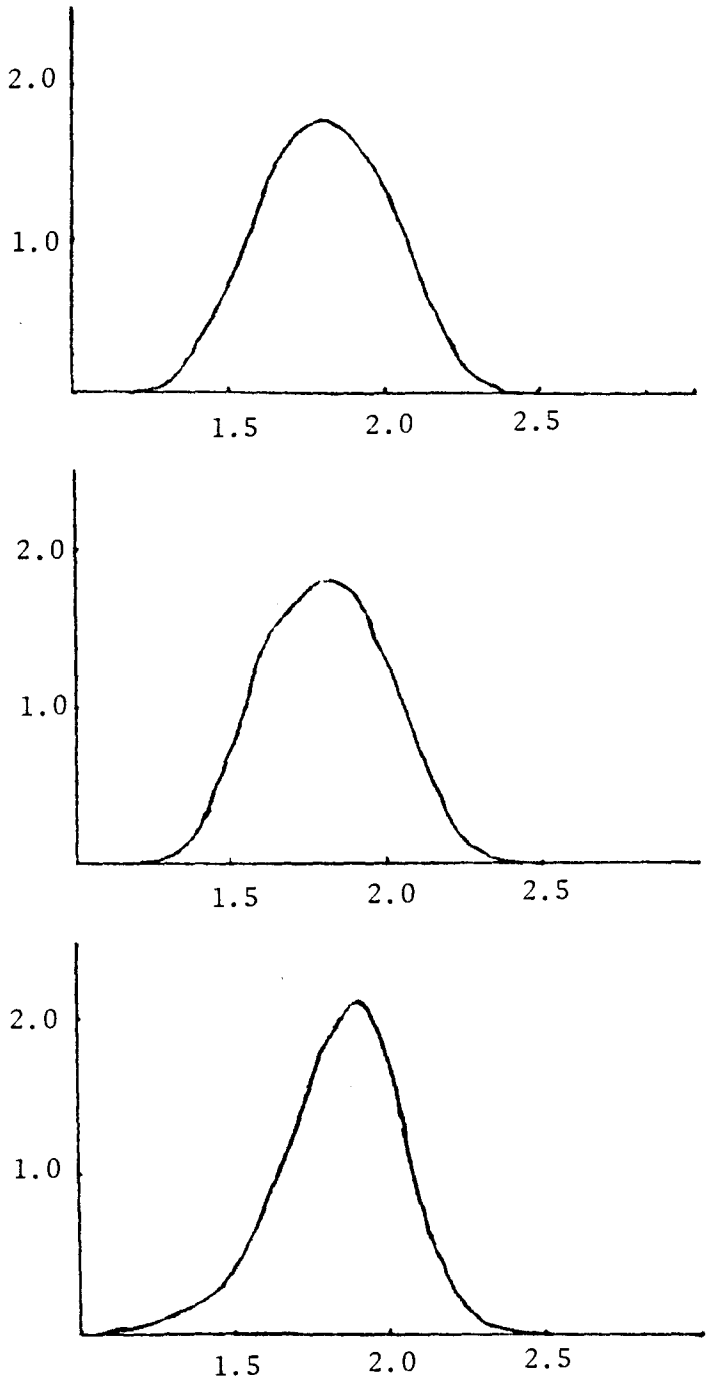

Figure 5.3 Comparison of marginals of the X-process at t = .10.

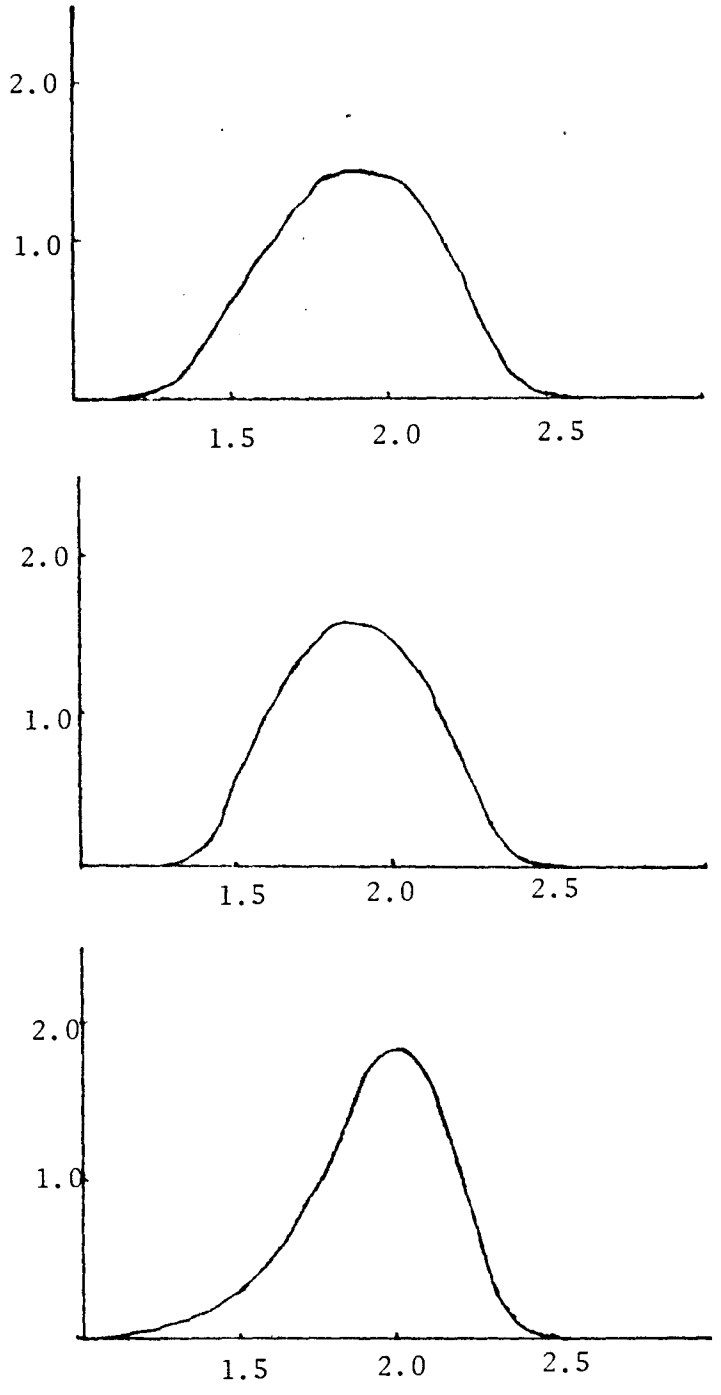

Figure 5.4 Comparison of marginals of the X-process at t = .15.

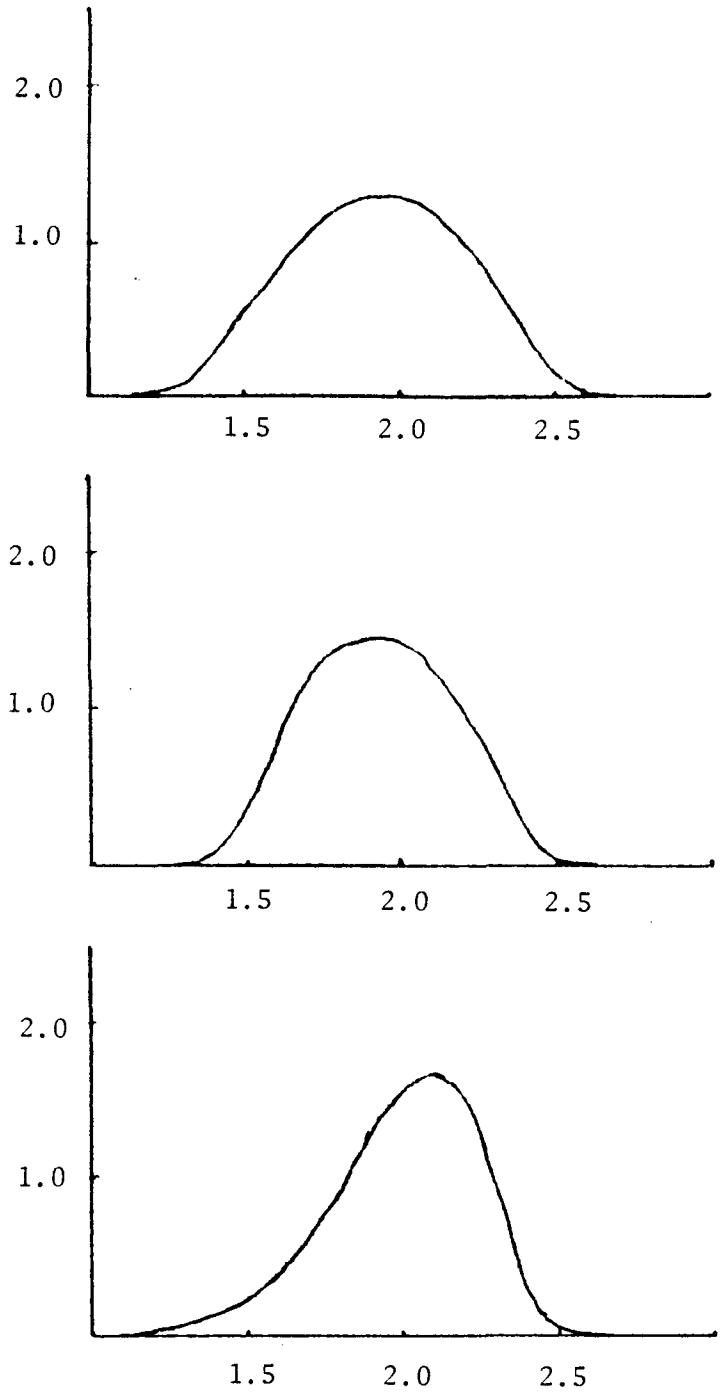

Figure 5.5 Comparison of marginals of the X-process at t = .20.

CHAPTER 6

SUMMARY

The Ito stochastic differential equation (possibly with vector-valued solutions) provides a very tempting framework for stochasticizing first order deterministic differential equations, because it gives the investigator two different approaches to the same problem. That is, one has the stochastic differential equation itself

$$dX = a(X,t)dt + b(X,t)dW$$

to be satisfied by the stochastic solution process $x(t,w)$, and the partial differential equation

$$\frac{\partial p}{\partial t} = - \sum_{i=1}^{n} \frac{\partial (a_i p)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial \left[\sum_{k=1}^{n} b_{ik} b_{kj} p \right]}{\partial x_i \partial x_j}$$

to be satisfied by its transition probability. Furthermore, under known conditions on the coefficient functions, the solution is a Markov process.

The drawbacks to this framework stem mainly from the modeling point of view. Some of these are:

1. The deterministic system must be stochasticized in such a way that the resulting stochastic system is linear in the formal derivatives or differentials of the Wiener process. This constrains the decision of which parameters to stochasticize.

2. Unless the coefficient b(X,t) is independent of the solution process, the resulting stochastic system is not well defined in that the stochastic calculus needs to be specified. Previous efforts have dealt with picking the calculus on the basis of the physical system to be modeled. This is an unsatisfactory approach, because the ambiguity arises from an essentially non-physical cause, namely, the unbounded variation of the trajectories of the Wiener process.

3. There are no positive autocorrelations in the noise process dW.

4. At any fixed time, the noise process dw is normally distributed. This means that any stochasticized parameter must take on values in any interval on the real line with positive probability.

These problems are addressed by increasing the dimensionality of the system by the number of parameters to be stochasticized. That is, instead of using the differentials of the Wiener process directly in the "modeling" equations, they are used in a second set of "noise" equations, whose solution processes are then used in the "modeling" equations. The augmented system looks like this

$$dX = f(X,Y)dt$$

$$dY = g(Y)dt + h(Y)dW$$

Note that there is now more freedom in how the noise (that is, the Y-process) enters in the modeling equation. Although there is still some ambiguity left with respect to the calculus rules in the second

set of equations, the choice here is a matter of convenience since the equation is now used to generate a known process. Some differentials of Wiener processes may, of course, still be added directly to the modeling equation but, in general, the system will exhibit the following kind of degeneracy. The coefficient matrix of the vector of differentials of the Wiener process, $\begin{bmatrix} 0 & 0 \\ 0 & h(Y) \end{bmatrix}$, in the above system may be degenerate at all points of the state space.

Several important questions must be dealt with in using the system described above. First, if the coefficient functions are not bounded by linear functions, a degenerate system of this type cannot be dealt with by the standard existence and uniqueness theorems. Second, work reported in the literature on boundary conditions for the corresponding partial differential equation has dealt principally with the one-dimensional case. Finally, it is harder numerically as well as analytically to solve a higher dimensional system.

In this thesis, these questions have been addressed in the following ways:

(a) The multi-dimensional existence and uniqueness theorem of McKean (1969) has been extended to apply to degenerate systems over a Euclidean domain;

(b) Using noise equations whose coefficients are infinitely differentiable, hypoellipticity was shown for the corresponding partial differential operators, and used to deduce boundary conditions for important special cases;

(c) Confidence regions for approximations to the marginal transition densities generated by Monte Carlo simulations were

developed. These approximations were used as a basis for comparison with the finite-difference approximations of the solutions of the corresponding partial differential equations.

Of all the modeling improvements possible with this approach, the most important is the ability to generate noises whose ranges of values are finite or bounded in at least one direction. Using noises of this type in the stochasticization of the carrying capacity of a logistic equation results in a solution that stays bounded as time goes to infinity. Previous stochastizations of the carrying capacity using differentials of the Wiener process have produced only unbounded solutions [see, for example, Turelli (1977)].

The Monte Carlo simulation technique proved on the whole to be easy to work with, and (for systems of more than one dimension) cheaper than finite-difference methods.

There are several areas for further research indicated by the work in this thesis. Some of them are as follows:

1. Numerical techniques for application either to the stochastic differential equations or the partial differential equations could be improved.

2. The graphs of Chapter 5 indicate that a stationary density for the x-process may exist. Available stability techniques might be investigated for possible application to problems of this type.

3. Autocorrelations in the X-process have been studied very little in the literature. For systems of multiple dimensions, the analytical approach might be very difficult, but there are time

series techniques that could be applied to generated sequences of values of the X-process.

4. Noises that are correlated with each other could be studied for use in larger systems. The numerical techniques described in this thesis could still be applied provided the coupling between the noises takes place in the drift coefficients only. Again this has been studied very little in the literature.

5. Estimation techniques for parameters in stochastic differential equations are needed to apply this machinery to real data. There is a lot of work available in estimation theory, but applications of biological data to mathematical structures of this type are still relatively rare.

BIBLIOGRAPHY

Anderson, R. F. and S. Orey (1976) "Small Random Perturbations of Dynamical Systems with Reflecting Boundary," Nagoya Math Journal 60:189-216.

Anderson, W. J. (1972) "Local Behavior of Solutions of Stochastic Integral Equations," TAMS 164:309-321.

Anderson, W. J. (1976) "A Note on Reflecting Boundaries for Solutions of Stochastic Differential Equations," Canadian Journal of Statistics 4:127-131.

Billingsley, P. (1968) Convergence of Probability Measures, Wiley, New York.

Feldman, M. W. and J. Roughgarden (1975) "A Population's Stationary Distribution and Chance of Extinction in a Stochastic Environment with Remarks on the Theory of Species Packing," Theoretical Population Biology 7:197-207.

Feller, W. (1952) "The Parabolic Differential Equation and the Associated Semigroups of Transformations," Ann. Math. 55:468-519.

Feller, W. (1954) "Diffusion Processes in One Dimension," TAMS 97: 1-31.

Friedman, A. (1975) Stochastic Differential Equations and Applications, Vol. I, Academic Press, New York.

Friedman, A. (1976) Stochastic Differential Equations and Applications, Vol. II, Academic Press, New York.

Hörmander, L. (1967) "Hypoelliptic Second-order Differential Equations," Acta Math. 119:147-171.

Kiester, A. R. and R. Barakat (1974) "Exact Solutions to Certain Stochastic Differential Equation Models of Population Growth," Theoretical Population Biology 6:199-216.

Lindgren, B. W. (1968) Statistical Theory, Macmillan Company, Collier-Macmillan Limited, London.

McKean, H. P. (1969) Stochastic Integrals, Academic Press, New York.

Prohorov, Y. V. and Y. A. Rozanov (1969) Probability Theory, Springer-Verlag, New York, Heidelberg and Berlin.

Rümelin, W. (1982) "Numerical Treatment of Stochastic Differential Equations," SIAM J. Numer. Anal. 19:604-613.

Stratonovich, R. L. (1966) "A New Representation for Stochastic Integrals and Equations," SIAM J. Control 4:362-371.

Tuckwell, H. C. (1974) "A Study of Some Diffusion Models of Population Growth," Theoretical Population Biology 5:345-357.

Turelli, M. (1977) "Random Environments and Stochastic Calculus," Theoretical Population Biology 12:140-178.

Williams, D. (1981) "To Begin at the Beginning: ... ," Stochastic Integrals Lecture Notes in Mathematics 851, Ed. D. Williams, Springer-Verlag, New York, Heidelberg and Berlin, pp. 1-55.

Yamada, T. and S. Watanabe (1971) "On the Uniqueness of Solutions to Stochastic Differential Equations: I, II," J. Math. Kyoto Univ. 11:155-167.

APPENDIX I

PROOFS OF HYPOELLIPTICITY OF SPECIFIC OPERATORS USED IN THE THESIS

In this appendix, we will work out the details of showing hypoellipticity for the forward Kolmogorov equations corresponding to particular systems of the form:

$$dX = f(X,Y)dt$$

$$dY = g(Y)dt + h(Y)dW \qquad \text{(AI.1)}$$

In particular, the second equations might be any of the three noise equations we are most interested in:

$$dY = g_1 Ydt + b_0(1-Y^2)dW$$

$$dY = g_1 Ydt + b_0(1+Y)dW$$

$$dY = g_1 Ydt + b_0 dW$$

We will use the approach developed by D. Williams (1981) based on the theorem by Hörmander (1967). This is explained in greater detail in Chapter 2. Since the g and h functions are functions of y only, we will use the notation g' and h' to refer to derivatives of these functions with respect to y. Since f(x,y)is a function of both x and y, we will use f_x and f_y to refer to partial derivatives with respect to x and y.

As mentioned in Chapter 2, the elements X_j of the Lie algebras we are considering are first order differential operators on R with C^∞ coefficients. The set of all first order operators at a point y forms a vector space $T_y(R^n)$ there over R with dimension n. A set of first order differential operators is said to be full at some point

y if the vectors of that set (that is, the differential operators with their coefficients evaluated at that point) span the vector space $T_y(R^n)$.

The Lie algebra formed by a generating set of first order differential operators will be a vector space over the field of real numbers but also a module over the ring of continuous functions. This means that the Lie algebra will contain, besides the vector space of the generating operators over the real numbers, all linear combinations of these operators using continuous functions as scalars. In addition to this, the Lie algebra will also contain all first order operators found by the Lie bracket operation [X,Y], defined as XY-YX for any two operators X and Y in the Lie algebra.

We will see precisely how this works for the one-dimensional noise equations $dy = g(y)dt + h(y)dw$. Using the Williams' choice of the X_j mentioned in Chapter 2 of the text, we compute the generating set to be $\{U,[U,V]\}$, where $U = h(y)\partial_y$ and $V = g(y)\partial_y - 3/2\ hh'\partial_y$. We are again using ∂_y to mean d/dy. Note that we cannot be sure V itself will be in the Lie algebra as it appears in the generating set only as part of a Lie bracket. We next compute [U,V].

$$[U,V] = h\partial_y(g\partial_y - 3/2\ hh'\partial_y) - g\partial_y(h\partial_y) + 3/2\ hh'\partial_y(h\partial_y)$$

As is generally the case, all the second order terms will vanish, so that we will be left with

$$[U,V] = hg'\partial_y - 3/2\ h(h')^2\partial_y - 3/2\ h^2h''\partial_y - gh'\ _y + 3/2\,h\,(h')\partial_y$$

Some of these are terms containing $h\partial_y$ but multiplied by another continuous function. Since $h\partial_y$ is in the set by itself as U, so are all of these products. This means we can subtract out all of these terms. This leaves us with $gh'\partial_y$ by itself having to be in the Lie algebra.

Now consider the first noise equation where $g(y) = g_1 y$ and $h(y) = b_0(1-y^2)$. Then we know our Lie algebra contains $U = b_0(1-y^2)\partial_y$ and $[U,V] = -2b_0 g_1 y^2 \partial_y$. By multiplying [U,V] by the appropriate constants and adding it to U, we see that ∂_y by itself must be in the Lie algebra. Since this coefficient of ∂y (being constant) never vanishes, it is clear that this Lie algebra will be full for all y. The corresponding operator

$$\frac{\partial p}{\partial t} + \frac{\partial(g,yp)}{\partial y} - \frac{b_0^2}{2}\frac{\partial^2([1-y^2]^2 p)}{\partial y^2}$$

will therefore be hypoelliptic over all of R. The other two noise equations work out in the same way.

For the two-dimensional system, the U will stay the same, $h\partial_y$, but V will be $g\partial_y - 3/2\ hh'\partial_y + f(x,y)\partial_x$. The generating set will again be {U,[U,V]}, but this time [U,V] implies the inclusion of the operator

$$-gh'\partial_y + hf_y\partial_x \qquad \text{(AI.2)}$$

(after eliminating multiples of $h\partial_y$). Note at this point that if h(y) is a constant not equal to zero (such as is the case for the third noise equation producing a noise of infinite range), the

U-operator will produce ∂_y by itself, and the Lie algebra will be full wherever $f_y \neq 0$.

If this is not the case, we must calculate the Lie bracket $[U,[U,V]]$. After we again subtract out all the terms containing $h\partial_y$, we are left with the vector

$$h'hf_y\partial_x + h^2f_{yy}\partial_x + (h')^2 g\,\partial_y \qquad \text{(AI.3)}$$

Now, if $f(x,y)$ is linear in y, f_{yy} will be zero everywhere, so that we will have

$$h'hf_y\partial_x + (h')^2 g\,\partial_y \qquad \text{(AI.4)}$$

We can clearly multiply (AI.2) by h' and subtract it from (AI.4) so that we are left with $2(h')^2 g\,\partial_y$. At this point, we turn to our specific noises. The most complicated one to work with is again the first one, with $g(y) = g_1 y$ and $h(y) = b_0(1-y^2)$. In this case, $2(h')^2 g\,\partial_y = 4b_0 g_1 y^3\partial_y$.

To see that we can again isolate the ∂_y, multiply $U = h\partial_y = b_0(1-y^2)\partial_y$ by $-4g_1$ and add $-4b_0g_1y^3\partial_y$. This will give us $-4g_1b_0y\partial_y$. This we can multiply by $-y/(4g_1)$ and add to $U = h\partial_y$. The semi-infinite noise again works out the same way. Since $[\partial_y, hf_y\partial_x] = h'f_y\partial_x$, we can use the same sort of tricks to see that $f_y\partial_x$ must be in the Lie algebra. Again the Lie algebra will be full (and the corresponding partial differential operator hypoelliptic) wherever $f_y \neq 0$.

Finally, the case that $f(x,y) = Rx - Rx^2(k+Ay)^{-1}$ is clearly not covered by the assumption that $f(x,y)$ be linear in y and must

therefore be covered separately. Note that since the case for our noise of infinite range (with h(y) equal to a constant) has already been covered, we will assume that y can be bounded in such a way that k + Ay can be kept bounded away from zero. That is, the support for the joint probability stays bounded away from the line y = -k/A. Then $(k + Ay)^{-1}$ can be assumed to be a continuous function on this support and its closure.

The generating set for the Lie algebra will be {U,[U,V]} exactly as before, but this time we cannot eliminate the f_{yy} term from [U,[U,V]] which remains as it was in (AI.3). However, now we know that

$$f_y = Rx^2A(k+Ay)^{-2} = -(k+Ay)f_{yy}/2A \quad .$$

since $f_{yy} = -Rx^2A^2(k+Ay)^{-2}2$.

Note first that we can multiply (AI.2) by h and use $U = h\partial_y$ to get $h^2f_y\partial_x$ by itself in the Lie algebra. Next we multiply (AI.3) by (k+Ay)/2A:

$$(k+Ay)hh'f_y\partial_x/2A + h^2(k+Ay)f_{yy}\partial_x/2A + \frac{(k+Ay)g(h')^2\partial_y}{2A} \qquad (AI.5)$$

The second term will be exactly $-h^2f_y\partial_x$, and can therefore be subtracted out. We then multiply (AI.2) by -(k+Ay)h'/2A to get

$$-(k+Ay)h'hf_y\partial_x/2A + (k+Ay)g(h')^2\partial_y/2A \qquad (AI.6)$$

This can be added to what is left of (AI.5) so that we have

$(k+Ay)g(h')^2\partial_y/A$ and therefore $g(h')^2\partial_y$ must again be in the Lie algebra. The fact that ∂_y is then in the Lie algebra and the subsequent hypoellipticity of the corresponding partial differential operator away from the locus $f_y = 0$, follows exactly as before.

APPENDIX II

SPECIFICS OF STATIONARY NOISE PROCESSES

In this appendix, we present the most frequently mentioned characteristics of the three particular noise equations examined in Chapter 5. All the results presented were calculated by evaluating the pertinent integrals directly. Recall the processes actually used in the model equations are the stationary processes produced by using an initial condition, Y_0, distributed according to the steady state solution of the corresponding Fokker-Planck equation.

The y-process produced by the equation

$$dY = g_1 Y\, dt + b_0(1-Y^2)dW$$

has its range of values within the interval [-1,1]. Its stationary solution has density

$$\frac{2}{Nb_0^2}\,\frac{e^{\frac{g_1}{b_0^2(1-y^2)}}}{(1-y^2)^2}$$

where N is the normalizing constant and equal to

$$\frac{e^{g_1/2b_0^2}}{b_0^2}\left[K_1\left(\frac{-g_1}{2b_0^2}\right) + K_0\left(\frac{-g_1}{2b_0^2}\right)\right]$$

K_0 and K_1 are modified Bessel functions. The variance of the stationary process is equal to

$$\frac{K_1\left(\frac{-g_1}{2b_0^2}\right) - K_0\left(\frac{-g_1}{2b_0^2}\right)}{K_1\left(\frac{-g_1}{2b_0^2}\right) + K_0\left(\frac{-g_1}{2b_0^2}\right)}$$

The Y process produced by the equation

$$dY = g_1 Y\, dt + b_0(1+Y)dW$$

has its range of values within the interval $[-1,\infty)$. Its stationary solution has density

$$\frac{2}{Nb_0^2}\ \frac{e^{\frac{2g_1}{b_0^2(1+y)}}}{(1+y)^{2-2g_1/b_0^2}}$$

where N is the normalizing constant and equal to

$$\left(\frac{2}{b_0^2}\right)^{\frac{-2g_1}{b_0^2}}\ \prod^{\frac{2g_1}{b_0^2}-1}\left(1-\frac{2g_1}{b_0^2}\right)$$

The variance of the stationary process is equal to

$$\frac{-b_0^2}{2g_1 + b_0^2} \quad .$$

The y-process produced by the equation

$$dy = g\ y\ dt + b\ dw$$

has its range of values equal to the entire real number line. Its stationary solution has density

$$\frac{2}{Nb_0^2}\ e^{\frac{g_1}{b_0^2} y^2}$$

where N is the normalizing constant and equal to

$$\frac{2}{b_0^2}\ \sqrt{\pi}\ \sqrt{\frac{-b_0^2}{g_1}} \quad .$$

The variance of the stationary process is equal to

$$\frac{-b_0^2}{2g_1} \quad .$$

The means of all three noises are zero. The autocorrelation function can in all three cases be written as

$$\text{variance} \cdot e^{g_1 t} \quad ,$$

as was shown in Chapter 4.

The noise produced by the first equation (having a range of values of [-1,1]) has one more characteristic. The stationary solution has a unimodal density function for low variances and a bimodal one for high variances. That is, if the parameters are picked so that $\frac{g_1}{2b_0^2} \leq -1$, we have a unimodal distribution function with its maximum at $y = 0$. For parameters obeying this inequality, the variance will be less than or equal to .177. For $\frac{g_1}{2b_0^2} > -1$, we have a bimodal distribution with the two values at

$$y = \pm \sqrt{\frac{g_1}{2b_0^2} + 1} \quad .$$

For comparison, recall that the variance of the uniform distribution over [-1,1] is .333.

APPENDIX III

TABLES FOR COMPARISON BETWEEN FINITE-DIFFERENCE METHOD AND MONTE CARLO TECHNIQUE

Table AIII.1 Steady State Densities of y-Processes

	Grid Points	-1	-.8	-.6	-.4	-.2	0	.2	.4	.6	.8	1
Finite noise (unimodal)	Correct	0	.01	.27	.71	.98	1.07	.98	.71	.27	.01	0
	FD (t=1)	0	0	.25	.69	.97	1.07	.98	.72	.29	.01	0
	FD (t=4)	0	0	.25	.68	.96	1.07	.99	.73	.30	.01	0
	MC (t=1)	0	0	.25	.73	.99	1.09	.94	.70	.24	0	0
Finite noise (bimodal)	Correct	0	.71	.65	.53	.46	.44	.46	.53	.65	.71	0
	FD (t=.16)	0	.69	.65	.53	.46	.44	.46	.52	.66	.74	0
	MC (t=2)	0	.15	.47	.70	.79	.85	.80	.66	.48	.15	0
	MC (t=4)	0	.545	.623	.543	.54	.511	.565	.547	.651	.598	0
Infinite noise	Correct	.01	.06	.22	.57	1.03	1.25	1.03	.57	.22	.06	.01
	FD (t=1)	.01	.04	.19	.55	1.02	1.25	1.02	.55	.19	.04	.01
	FD (t=3.92)	.01	.04	.18	.53	1.00	1.25	1.05	.61	.24	.07	.01
	MC (t=1)	.00	.06	.22	.55	1.08	1.23	1.01	.59	.22	.04	.01
Semi-infinite noise	Grid pts.	-.5	-.4	-.2	0	.2	.4	.6	.8	1.0	1.2	1.4
	Correct	.19	.67	1.52	1.30	.76	.38	.18	.09	.04	.02	.01
	FD (t=1)	.13	.61	1.48	1.30	.78	.40	.20	.09	.05	.02	.01
	FD (t=4)	.12	.57	1.43	1.28	.79	.42	.21	.10	.05	.02	.01
	MC (t=1)	.22	.67	1.50	1.24	.74	.42	.18	.09	.04	.01	.01

Table AIII.2 Comparison of Finite-Difference Method with Monte Carlo for x-Marginals

	Grid Points	1.0	1.2	1.4	1.6	1.8	2.0	2.2
Bimodal noise								
-1.0 autocorr.	FD (10 steps)	0.00	.260	1.218	2.146	.925	.142	.007
at time = .02 unit	MC (2 steps)	0.00	.318	1.531	2.015	.965	.174	.006
Unimodal								
-10 autocorr.	FD (10 steps)	0.00	.212	1.316	2.270	1.029	.123	.004
at time = .02 unit	MC (2 steps)	.001	.288	1.322	2.255	.995	.135	.004
Unimodal								
-1.0 autocorr.	FD (5 steps)	.000	.407	1.602	2.045	.813	.100	.000
at time = .01 unit	MC (1 step)	.002	.444	1.621	1.975	.836	.101	.003

Appendix IV

```
//MCLOG WATFIV NCS.ES.G7382,VSMITH,P=15,T=(5,30)
C
C THIS PROGRAM APPROXIMATES THE JOINT DENSITY OF THE
C Y-NOISE TOGETHER WITH THE X-MODEL OUTPUT BY
C PLOTTING THE FREQUENCY FUNCTION OF N TRAJECTORIES
C AFTER T TIMESTEPS OF LENGTH H.
C
      REAL P(20,41,41),BOSQ,GI
      REAL H,Y(50)/50*0.0/,W,YO,U,UU,X(21)
      INTEGER N,I,J,L,K,T,IR,IC
      INTEGER ROUND
      INTEGER M,NM,KM,B,BB
      COMMON B,BB,M,NM,KM
C
C N   IS THE NUMBER OF TRAJECTORIES CALCULATED.
C GI*Y IS THE DRIFT COEFFICIENT OF THE NOISE EQUATION.
C BOSQ*(1-Y**2) IS THE DIFFUSION COEFFICIENT OF THE NOISE EQUATION.
C THE MODEL EQUATION IS DX=12*X(1-X/(2+Y)).
C THE COMMON STATEMENT IS USED ONLY WITH THE SUBROUTINE RAND.
C
      N=10000
      M=314159269
      GI=-1.0
      BOSQ=1.143
      KM=2**31
      T=40
      B=273148
      BB=B
C
      DO 10 I=1,20
         DO 11 J=1,41
            DO 12 K=1,41
               P(I,K,J)=0.0
12          CONTINUE
11       CONTINUE
10    CONTINUE
C
      X(1)=1.5
C
C
C
      DO 25 NJ=1,N
 14      CONTINUE
         CALL RAND
         U=ABS(B/(KM*1.0))
         UU=ABS(BB/(KM*1.0))
         W=SQRT(-ALOG(U)*2.0)
         V=.2*W*COS(6.28*UU)+1.5
         W=.2*W*SIN(6.28*UU)+1.5
         IF(V.LT.1.OR.V.GT.3)  GO TO 18
         X(1)=V
         GO TO 16
 18      CONTINUE
         IF(W.LT.1.OR.W.GT.3) GO TO 14
         X(1)=W
 16      CONTINUE
         H=.05
         CALL RAND
         YO=0.0
41       CALL RAND
         U=ABS(B/(KM*1.0))
         UU=ABS(BB/(KM*1.0))
         W=SQRT(-ALOG(U)*2.0)
         V=W*COS(6.28*UU)
         W=W*SIN(6.28*UU)
```

Appendix IV (continued)

```
      Y(1)=YO+((BOSQ+GI)*YO-BOSQ*YO**3)*H
     1           +SQRT(BOSQ*H)*(1-YO**2)*V
     1           -BOSQ*YO*(1-YO**2)*H*V**2
      Y(2)=Y(1)+((BOSQ+GI)*Y(1)-BOSQ*Y(1)**3)*H
     1           +SQRT(BOSQ*H)*(1-Y(1)**2)*W
     1           -BOSQ*Y(1)*(1-Y(1)**2)*H*W**2
      DO 20 IR=3,T,2
42       CALL RAND
         U=ABS(B/(KM*1.0))
         UU=ABS(BB/(KM*1.0))
         W=SQRT(-ALOG(U)*2.0)
         V=W*COS(6.28*UU)
         W=W*SIN(6.28*UU)
         L=IR-1
         Y(IR)=Y(L)+((BOSQ+GI)*Y(L)-BOSQ*Y(L)**3)*H
     1             +SQRT(BOSQ*H)*(1-Y(L)**2)*W
     1             -BOSQ*Y(L)*(1-Y(L)**2)*H*W**2
         L=IR+1
         Y(L)=Y(IR)+((BOSQ+GI)*Y(IR)-BOSQ*Y(IR)**3)*H
     1            +SQRT(BOSQ*H)*(1-Y(IR)**2)*V
     1            -BOSQ*Y(IR)*(1-Y(IR)**2)*H*V**2
20    CONTINUE
      DO 30 I=1,9,2
         H=.01
43       CALL RAND
         U=ABS(B/(KM*1.0))
         UU=ABS(BB/(KM*1.0))
         W=SQRT(-ALOG(U)*2.0)
         V=W*COS(6.28*UU)
         W=W*SIN(6.28*UU)
         L=I-1
         Y(10+I)=Y(10+L)+((BOSQ+GI)*Y(10+L)-BOSQ*Y(10+L)**3)*H
     1                  +SQRT(BOSQ*H)*(1-Y(10+L)**2)*W
     1                  -BOSQ*Y(10+L)*(1-Y(10+L)**2)*H*W**2
         X(1+I)=X(L+1)+H*(12*X(L+1)-12*X(1+L)**2/(2+Y(10+I)))
         K=21-ROUND(Y(10+I)*10)
         L=1+ROUND(X(1+I)*10)
         IF(K.LT.1) K=1
         IF(K.GT.41) K=41
         IF(L.LT.1) L=1
         IF(L.GT.41) L=41
         P(I,K,L)=P(I,K,L)+1
         L=I+1
         Y(10+L)=Y(10+I)+((BOSQ+GI)*Y(10+I)-BOSQ*Y(10+I)**3)*H
     1                  +SQRT(BOSQ*H)*(1-Y(10+I)**2)*V
     1                  -BOSQ*Y(10+I)*(1-Y(10+I)**2)*H*V**2
         X(1+L)=X(I+1)+H*(12*X(I+1)-12*X(1+I)**2/(2+Y(10+L)))
         K=21-ROUND(Y(10+L)*10)
         L=1+ROUND(X(1+L)*10)
         IF(K.LT.1) K=1
         IF(K.GT.41) K=41
         IF(L.LT.1) L=1
         IF(L.GT.41) K=41
         P(I+1,K,L)=P(I+1,K,L)+1
30    CONTINUE
25  CONTINUE
C
C
C
      PRINT,N
      DO 60 I=1,10
         PRINT,I
         DO 82 K=11,31
            DO 84 L=11,31
               P(I,K,L)=P(I,K,L)*100/N
```

Appendix IV (continued)

```
84          CONTINUE
82        CONTINUE
65        FORMAT(5X,21F5.2,5X)
          DO 67 K=11,31
             WRITE(3,65)(P(I,K,L),L=11,31)
67        CONTINUE
60     CONTINUE
C
C
C
55     RETURN
       END
C
C THIS FUNCTION TAKES ANY REAL NUMBER AND ROUNDS IT OFF TO
C NEAREST INTEGER. THE RESULT IS A DECLARED INTEGER.
C
       INTEGER FUNCTION ROUND(R)
       REAL R
       IF(ABS(R-IFIX(R)).GE..50) GO TO 15
       ROUND=IFIX(R)
       GO TO 25
15     IF(R.LT.0) ROUND=IFIX(R)-1
       IF(R.GT.0) ROUND=IFIX(R)+1
25     RETURN
       END
C
C THIS SUBROUTINE USES THE CONGRUENCE METHOD TO PRODUCE 2
C INDEPENDENT RANDOM VARIABLES DISTRIBUTED UNIF(0,1).
C THE NTH AND 2NTH MEMBERS OF THE PSEUDO RANDOM SEQUENCE ARE
C CALCULATED SO THAT A CYCLE CHECK CAN BE MADE BY CHECKING
C IF X=Y.
C
       SUBROUTINE RAND
       INTEGER X,Y,M,N,K
       COMMON X,Y,M,N,K
       X=M*X+1
       Y=M*Y+1
       Y=M*Y+1
       RETURN
       END
$DATA
/*
```