Modeling with Itô Stochastic Differential Equations

MATHEMATICAL MODELLING:
Theory and Applications

VOLUME 22

This series is aimed at publishing work dealing with the definition, development and application of fundamental theory and methodology, computational and algorithmic implementations and comprehensive empirical studies in mathematical modelling. Work on new mathematics inspired by the construction of mathematical models, combining theory and experiment and furthering the understanding of the systems being modelled are particularly welcomed.

Manuscripts to be considered for publication lie within the following, non-exhaustive list of areas: mathematical modelling in engineering, industrial mathematics, control theory, operations research, decision theory, economic modelling, mathematical programming, mathematical system theory, geophysical sciences, climate modelling, environmental processes, mathematical modelling in psychology, political science, sociology and behavioural sciences, mathematical biology, mathematical ecology, image processing, computer vision, artificial intelligence, fuzzy systems, and approximate reasoning, genetic algorithms, neural networks, expert systems, pattern recognition, clustering, chaos and fractals.

Original monographs, comprehensive surveys as well as edited collections will be considered for publication.

The titles published in this series are listed at the end of this volume.

Modeling with Itô Stochastic Differential Equations

By

E. Allen
Texas Tech University, USA

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-13 978-1-4020-5952-0 (HB)
ISBN-13 978-1-4020-5953-7 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

To Linda and Anna

Contents

Preface

The purpose of this book is to provide an introduction to the theory, computation, and application of Itô stochastic differential equations. In particular, a procedure for developing stochastic differential equation models is described and illustrated for applications in population biology, physics, and mathematical finance. The modeling procedure involves first constructing a discrete stochastic process model. The discrete model is developed by studying changes in the process over a small time interval. An Itô stochastic differential equation model is then formulated from the discrete stochastic model. The procedure is straightforward and is useful for many dynamical processes that experience random influences.

The main topics in the theory and application of stochastic differential equations include random variables, stochastic processes, stochastic integration, stochastic differential equations, and models. These topics are introduced and examined in separate chapters. Many examples are described to illustrate the concepts. The emphasis in the explanations is to provide a good understanding of the concepts. Results are not necessarily presented in their most general form. Simplicity of presentation is chosen over generality. For the first four chapters, the theory of random processes and stochastic differential equations is presented in a Hilbert space setting. A Hilbert space setting is chosen to unify and simplify the presentation of the material. The last chapter concentrates on explaining a model development procedure that is useful for constructing stochastic differential equation models for many kinds of dynamical systems experiencing random changes. The procedure produces, in a natural manner, an Itô stochastic differential equation model, in contrast with, for example, a Stratonovich stochastic differential equation model.

There are many excellent books available on the theory, application, and numerical treatment of stochastic differential equations. The bibliography lists many of these books. It is hoped that the present book will complement these previous books in providing an introduction to the development and testing of stochastic differential equation models.

One of the objectives of this book is to provide a basic introduction to the theory, approximation, and application of stochastic differential equations for anyone who is interested in pursuing research in this area. The intent of this book is to provide a background to stochastic differential equations so that the reader will be in a position to understand the mathematical literature in this area, including more advanced texts. To understand the material presented in this book, proficiency in probability theory and differential equations is assumed. In particular, prerequisite courses for thoroughly understanding the concepts in this book include probability theory or statistics, differential equations, and intermediate analysis. In addition, some knowledge of scientific computing and programming would be very helpful. Throughout the book, approximation procedures are described. Problems involving stochastic integration and stochastic differential equations can rarely be solved exactly and numerical procedures must be employed. In each chapter, one or two computer programs are listed in the computer languages MATLAB or Fortran. Each program is useful in solving a representative stochastic problem. The computer programs are listed in the book for convenience and to illustrate that the programs are generally short and straightforward. At the end of each chapter, analytical and computational exercises are provided. Several additional computer programs are listed in these exercise sets. The exercises are designed to complement the material in the text.

I am grateful to Texas Tech University for providing me a one-semester faculty development leave to write much of this book and the opportunity later to use this book in teaching a one-semester graduate course. I thank my wife, Linda Allen, for her encouragement on the writing of this book and for her many helpful comments and suggestions. I thank my friends and colleagues, Robert Paige, Henri Schurz, and Zhimin Zhang for their many positive suggestions and criticisms on the manuscript. I thank Lynn Brandon, Springer Mathematics Publishing Editor, for her efficient assistance in the publication process. I am grateful to the several anonymous reviewers of the manuscript for their positive comments and their recommended revisions and additions. Finally, I am grateful to the colleagues and graduate students who worked with me on research projects.

Lubbock, Texas, December 2006 *Edward Allen*

1

Random Variables

1.1 Introduction

A random variable is a real-valued function defined on the set of outcomes of a random experiment. Random variables are important in stochastic integration and stochastic differential equations. Indeed, a stochastic integral is a random variable and the solution of a stochastic differential equation at any fixed time is a random variable. One of the main purposes of this chapter is to define Hilbert spaces of random variables. Hilbert spaces unify the presentation of the first four chapters of this book and are used to describe the properties of stochastic integration and stochastic differential equations. For example, in Chapter 3 a particular sequence of random variables in a Hilbert space will be shown to converge to a random variable and this random variable will define a stochastic integral. To describe a Hilbert space of random variables, it is necessary to understand the concept of random experiment. In particular, it is useful to understand the possible events of a random experiment and the probabilities of these events. Furthermore, to define a metric on the Hilbert space of random variables, expectations are needed.

In this chapter, after probability spaces, random variables, and expectation are introduced, a Hilbert space of random variables is discussed. Next, the important concept of convergence of random variables is considered. An advantage of using Hilbert spaces of random variables is that any Cauchy sequence of random variables in the Hilbert space converges to a unique random variable in the space. This completeness property of Hilbert spaces simplifies the presentation of stochastic differential equations. Finally, for computational studies involving random variables, computer generation of random numbers is useful. In the final two sections of this chapter, random number generation is described along with the associated Monte Carlo procedures for using random numbers in problem solving.

A basic introduction is given in this chapter to random variables with priority given to results needed in describing stochastic differential equations.

Thorough treatments of random variables and probability theory are presented, for example, in [11, 51, 62, 63, 67, 85, 87, 89, 96, 106].

1.2 Probability Space

A process that has random outcomes is called a random experiment. The set of all possible outcomes of a random experiment is called the sample space and is denoted Ω. A combination of outcomes, a subset of Ω, is called an event. The set of events is denoted $\mathcal{A}$. It is assumed that the set $\mathcal{A}$ is a σ-algebra which means that the following properties hold for $\mathcal{A}$:

(a) $\Omega \in \mathcal{A}$

(b) $A^c \in \mathcal{A}$ if $A \in \mathcal{A}$, where A^c is the complement of set A

(c) $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$ if $A_1, A_2, \cdots \in \mathcal{A}$.

Suppose that the random experiment is repeated N times and N_A is the number of times that event A occurs. Let $f_A = N_A/N$ be the relative frequency of event A. The probability of an event $A \in \mathcal{A}$ is denoted as $P(A)$. One can regard the probability $P(A)$ as the value that f_A approaches as the experiment is repeated an increasing number of times. The probability P is a set function that maps $\mathcal{A}$ into $[0, 1]$ and P is called a probability measure if the following conditions hold:

(d) $P(\Omega) = 1$

(e) $P(A^c) = 1 - P(A)$

(f) $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$, if $A_i \cap A_j = \phi$ for $i \neq j$.

The triplet $(\Omega, \mathcal{A}, P)$ consisting of the sample space Ω, the σ-algebra $\mathcal{A}$ of subsets of Ω, and a probability measure P defined on $\mathcal{A}$ is called a probability space.

The following four examples illustrate the concept of a probability space.

Example 1.1. Flipping a coin twice; generating a σ-algebra set

Consider the random experiment of flipping a coin twice. The possible outcomes are $\omega_1 = HH$, $\omega_2 = HT$, $\omega_3 = TH$, $\omega_4 = TT$. Thus, the sample space is $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. There are, however, many possible sets $\mathcal{A}$ that satisfy the properties of a σ-algebra. The smallest is $\mathcal{A} = \{\phi, \Omega\}$. If it is desired that $\{\omega_1\}, \{\omega_2\} \in \mathcal{A}$, then the smallest σ-algebra containing $\{\omega_1\}$ and $\{\omega_2\}$ is

$$\mathcal{A} = \{\phi, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}, \Omega\}.$$

This set is said to be generated by the outcomes ω_1 and ω_2. The largest σ-algebra is

$$\mathcal{A} = \{\phi, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_4\}, \{\omega_2, \omega_3\},$$

$$\{\omega_2, \omega_4\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_1, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}, \Omega\}.$$

For N different outcomes, the smallest σ-algebra will have two elements, ϕ and Ω, and the largest σ-algebra will have $\sum_{i=0}^{N} C_i^N = 2^N$ elements where $C_i^N = N!/((N-i)!i!)$.

For this example, let $P(\{\omega_1\}) = P(\{\omega_2\}) = P(\{\omega_3\}) = P(\{\omega_4\}) = \frac{1}{4}$ define the probability measure. Then using properties (d)-(f), the probability of any event can be calculated. For example, the probability of the event $\{\omega_1$ or ω_3 or $\omega_4\}$ is $P(\{\omega_1, \omega_3, \omega_4\}) = 1 - P(\{\omega_2\}) = \frac{3}{4}$. The triplet $(\Omega, \mathcal{A}, P)$ is the probability space for this example.

Example 1.2. Flipping a coin until a tail shows; an infinite number of outcomes

Consider again the random experiment of flipping a coin. In this experiment, however, the coin is flipped until a tail turns up. In this case, $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$ where ω_i is the outcome where $i-1$ tosses are heads and the ith toss is a tail. For example, let B be the event where the first tail occurs after an odd number of tosses. Then, $B = \{\omega_1, \omega_3, \omega_5, \dots\}$ which is an infinite union of subsets of Ω. Let $\mathcal{A} = \{\phi, \{\omega_1\}, \{\omega_2\}, \dots, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \dots,\}$ be a σ-algebra of Ω generated so that $\{\omega_i\} \in \mathcal{A}$ for $i = 1, 2, \dots$. Finally, let $P(\{\omega_i\}) = (\frac{1}{2})^i$ define a probability measure on $\mathcal{A}$. Then the triplet $(\Omega, \mathcal{A}, P)$ is a probability space. Notice, for example, that the probability of an odd number of tosses is $P(B) = \sum_{i=1}^{\infty} P(\{\omega_{2i-1}\}) = \sum_{i=1}^{\infty} (\frac{1}{2})^{2i-1} = \frac{2}{3}$.

Example 1.3. A continuous set of outcomes; Lebesgue measure

Consider the experiment of randomly choosing a real number x from the interval $[0, 1]$. Then $\Omega = \{x : 0 \le x \le 1\}$. Let $(a, b]$ be an interval in $[0, 1]$. Define an event A to be $x \in (a, b]$. Let the σ-algebra $\mathcal{A}$ be defined as the set generated by all intervals of the form $(a, b]$. Therefore, all intervals of the form $(a, b]$, unions of such intervals, and complements of all the resulting sets are contained in the σ-algebra $\mathcal{A}$. This σ-algebra $\mathcal{A}$ is called the σ-algebra of Borel sets [65, 73]. Now define the probability measure P so that $P(A) = b - a$ for the event A. Then $P(A)$ is the probability that an element $x \in [0, 1]$ belongs to A. This probability measure is called Lebesgue measure for this σ-algebra $\mathcal{A}$. Notice that because countable unions and complements are allowed by properties (b)–(c), a very large number of types of sets are included in $\mathcal{A}$. For example, $(a, b) \in \mathcal{A}$ as $(a, b) = \cup_{n=1}^{\infty}(a, b - \frac{1}{n}]$. By De Morgan's laws, countable intersections are also in $\mathcal{A}$. In particular, the singleton $\{\omega\} = \{x\}$ is in $\mathcal{A}$ as $\{\omega\} = \cap_{n=1}^{\infty}(x - \frac{1}{n}, x]$. Consider the following three particular examples.

Let $B_1 \in \mathcal{A}$ where $B_1 = \cup_{n=1}^{\infty} \left(2^{1-2n}, 2^{2-2n}\right)$. Then, $P(B_1) = \frac{2}{3}$.

Let $B_2 = \left\{\frac{n-1}{n}, \text{ for } n = 1, 2, 3, \dots\right\}$. Then $B_2 \in \mathcal{A}$ and $P(B_2) = 0$.

Finally, $B_3 = \{\text{rational numbers on } [0, 1]\} \in \mathcal{A}$ and $P(B_3) = 0$.

Example 1.4. Number of observations; Poisson distributed

Consider an experiment where the number of observations of a result in time interval $[0, t]$ is of interest. It is supposed that the number of observations of the result in any time interval Δt has probability $\lambda\Delta t + o(\Delta t)$ and the probability is independent of time. For example, in radioactive decay of a long-lived isotope, the probability of a decay in a small time interval Δt satisfies this assumption. The probability that a car passes a certain intersection may satisfy this assumption. In addition, if a wire passes through an instrument at a constant speed, the probability of a defect in time Δt may satisfy this assumption. Consider the number of results where t is large compared with the time interval Δt. Let ω_n equal the outcome where n results occur in the interval $[0, t]$. Then, it is clear that $\Omega = \{\omega_0, \omega_1, \omega_2, \dots\}$. Let $\mathcal{A} = \{\phi, \{\omega_0\}, \{\omega_1\}, \dots, \Omega\}$ be the σ-algebra of events generated by assuming that $\{\omega_i\} \in \mathcal{A}$ for $i = 0, 1, 2, \dots$. It is of interest to determine a probability measure for Ω. For notational convenience, let $P(\{\omega_n\}) = P_n(t)$ be the probability of n results in the interval $[0, t]$. Using the above assumptions, it is clear that $P_0(0) = 1$ and $P_n(0) = 0$ for $n \geq 1$. In addition,

$$P_0(t + \Delta t) = (1 - \lambda\Delta t)P_0(t) + o(\Delta t)$$

and

$$P_n(t + \Delta t) = (1 - \lambda\Delta t)P_n(t) + \lambda\Delta t P_{n-1}(t) + o(\Delta t)$$

for $n \geq 1$ as $1 - \lambda\Delta t$ is the probability of no result in time interval Δt and λt is the probability of one result in time interval Δt. Letting $\Delta t \to 0$ in the above expressions yields the differential equations

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t), \quad P_0(0) = 1$$

and

$$\frac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad P_n(0) = 1, \quad \text{for} \quad n \geq 1.$$

Solving this system gives $P(\{\omega_n\}) = P_n(t) = \exp(-\lambda t)(\lambda t)^n/n!$ for $n = 0, 1, 2, \dots$ as the probability of n results in time t. With this probability measure, $(\Omega, \mathcal{A}, P)$ is a probability space. To check the probability measure, notice that

$$P(\Omega) = \sum_{n=0}^{\infty} P(\{\omega_n\}) = \sum_{n=0}^{\infty} \exp(-\lambda t)\frac{(\lambda t)^n}{n!} = 1$$

for any time $t \geq 0$. For this random experiment, the number of results is said to be Poisson distributed.

1.3 Random Variable, Probability Distribution

A random variable X is a real-valued function that assigns the value $X(\omega) \in \mathbb{R}$ to each outcome $\omega \in \Omega$. That is, $X : \Omega \to \mathbb{R}$. The function $F_X(x) =$

$P(\{\omega \in \Omega : X(\omega) \leq x\})$ is called the distribution function or the probability distribution of the random variable X. It is assumed that X is measurable, specifically, X has the property that the set $A(x) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$ for each $x \in \mathbb{R}$. This technical condition is required so that the distribution function can be defined. Notice that the distribution function involves $P(A(x))$ as $F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(A(x))$ and cannot be determined unless $A(x) \in \mathcal{A}$.

A random variable can be discrete or continuous. A random variable is called discrete if it takes values in a countable subset $\{x_1, x_2, x_3, \dots\} \subset \mathbb{R}$. That is, $X(\omega) \in \{x_1, x_2, x_3, \dots\}$ for each $\omega \in \Omega$. The probability mass function p of a discrete random variable X is the function $p : \{x_1, x_2, x_3, \dots\} \to [0, 1]$ given by $p(x) = P(X = x)$. Notice that for a discrete random variable $F_X(x) = \sum_{x_i < x} p(x_i)$. A random variable is called continuous if there exists a piecewise continuous nonnegative function $p(x)$ such that $F_X(x) = \int_{-\infty}^{x} p(s)\, ds$. In this case, $p(x)$ is called the probability density function of X. Notice that $P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b p(s)\, ds$.

It is useful to notice that if X is a random variable and $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, then $Y = g(X)$ is also a random variable. Indeed, $Y(\omega) = g(X(\omega)) = g(x)$ if $X(\omega) = x$. In addition, if $p(x)$ is the probability mass function for a discrete random variable X and the inverse g^{-1} exists, then $P(Y = y = g(x)) = P(X = x) = p(x) = p(g^{-1}(y))$. So $q(y) = p(g^{-1}(y))$ is the probability mass function for Y. In addition, if X takes on the discrete values $\{x_1, x_2, x_3, \dots\}$, then Y takes on the discrete values $\{y_1, y_2, y_3, \dots\}$ where $y_i = g(x_i)$.

The following examples illustrate concepts involving random variables. It is interesting that if X is defined and $F_X(x)$ is determined, then the underlying probability space is not explicitly required to understand many of the properties of the random variable.

Example 1.5. Flipping a coin once

Consider the random experiment of flipping a coin one time. Then $\Omega = \{H, T\}$. Let $\mathcal{A} = \{\phi, \{H\}, \{T\}, \Omega\}$ and let $P(\{H\}) = P(\{T\}) = \frac{1}{2}$. Define the discrete random variable X on Ω so that $X(T) = 0$ and $X(H) = 5$. Then $F_X(x)$ has the form

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1/2, & 0 \leq x < 5 \\ 1, & x \geq 5. \end{cases}$$

Example 1.6. Poisson distributed

Consider the random experiment of Example 1.4 where the probability of a result in time interval Δt is $\lambda \Delta t + o(\Delta t)$. Let $\gamma = \lambda t$ and recall that w_n is defined to be the outcome of having n results occurring in $[0, t]$. Let $X(\omega_n) = n$ define the random variable X. Notice that X is a discrete random variable and is integer-valued. For this example, the probability mass function $p(n) = P(\omega_n) = \exp(-\gamma)\gamma^n/n!$ for $n = 0, 1, 2, \dots$ and thus, the probability distribution $F_X(x)$ has the form

$$F_X(x) = \exp(-\gamma) \sum_{k=0}^{n} \frac{\gamma^k}{k!} \quad \text{for} \quad n \le x < n+1$$

for $n = 0, 1, 2, \ldots$. The random variable X is said to be Poisson distributed with intensity λ.

Example 1.7. Uniformly distributed on $[0, 1]$

Consider the random experiment of selecting a number x randomly from the interval $[0, 1]$ described in Example 1.3. Recall that $\mathcal{A}$ includes all intervals and unions of intervals in $[0, 1]$ and the probability measure is defined as $P(A) = d - c$ where the event $A = \{x \in [0, 1] : c < x \le d\}$. For this example, define the random variables X_1 and X_2 as $X_1(x) = x$ and $X_2(x) = \exp(x)$. Notice that X_1 and X_2 are continuous random variables. In this case,

$$F_{X_1}(x) = \int_{-\infty}^{x} p_1(s)\, ds, \quad \text{where} \quad p_1(s) = \begin{cases} 0, & s < 0 \\ 1, & 0 \le s \le 1 \\ 0, & s > 1 \end{cases}$$

and

$$F_{X_2}(x) = \int_{-\infty}^{x} p_2(s)\, ds, \quad \text{where} \quad p_2(s) = \begin{cases} 0, & s < 1 \\ 1/s, & 1 \le s \le e \\ 0, & s > e. \end{cases}$$

Notice that if $a, b \in [0, 1]$, then $P(a \le X_1 \le b) = \int_a^b p_1(s)\, ds = b - a$ and $P(e^a \le X_2 \le e^b) = \int_{e^a}^{e^b} p_2(s)\, ds = b - a$. However, if $c, d \in [1, e]$, then $P(c \le X_2 \le d) = \ln d - \ln c$. For this example, X_1 is said to be distributed uniformly on $[0, 1]$ as the probability that $X_1(x)$ lies within any subinterval of $[0, 1]$ is proportional to the width of the subinterval. It is denoted that $X_1 \sim U[0, 1]$.

Example 1.8. Uniformly distributed on $[u, v]$

The random experiment described in Example 1.7 can be generalized to any interval $[u, v]$. Specifically, numbers are randomly selected on $[u, v]$ rather than on $[0, 1]$. As in the previous example, define the random variable $X(x) = x$. Then

$$F_X(x) = \int_{-\infty}^{x} p(s)\, ds, \quad \text{where} \quad p(s) = \begin{cases} 0, & s < u \quad \text{or} \quad s > v \\ \dfrac{1}{v-u}, & u \le s \le v. \end{cases}$$

Notice that if $a, b \in [u, v]$, then $P(a \le X \le b) = \int_a^b p(s)\, ds = (b-a)/(v-u)$. In this example, X is said to be distributed uniformly on $[u, v]$ and it is denoted that $X \sim U[u, v]$.

Example 1.9. Normally distributed on $(-\infty, \infty)$

Consider $\Omega = \{x : -\infty < x < \infty\}$. Let $\mathcal{A}$ be a $\sigma-$algebra generated by intervals of the form $(a, b]$, that is, $(a, b] \in \mathcal{A}$ for any $a, b \in \mathbb{R}$ along with

countable unions of such intervals and complements of the resulting sets. The set $\mathcal{A}$ is similar to the σ-algebra described in Example 1.3 but for the interval $(-\infty, \infty)$ rather than $[0, 1]$. Define the random variable X as $X(x) = x$. Let $A \subset \mathcal{A}$ and let $\mu \in \mathbb{R}$ and $\sigma > 0$ be constants. Define

$$P(A) = \int_A p(s)\, ds, \quad \text{where} \quad p(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(s-\mu)^2}{2\sigma^2}\right).$$

That is,

$$P(a \le X \le b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(s-\mu)^2}{2\sigma^2}\right) ds.$$

For this example, X is said to be normally distributed with mean μ and variance σ^2 and it is denoted that $X \sim N(\mu, \sigma^2)$. Notice, for example, that

$$\begin{aligned} P(\mu - \sigma \le X \le \mu + \sigma) &= \int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(s-\mu)^2}{2\sigma^2}\right) ds \\ &= \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} \exp\frac{-s^2}{2}\, ds = 0.6826. \end{aligned}$$

If $\mu = 0$ and $\sigma = 1$, then X is said to have a standard normal distribution, i.e., $X \sim N(0, 1)$.

1.4 Expectation

Suppose that X is a discrete random variable, where $X(\omega) \in \{x_1, x_2, \dots\}$ for $\omega \in \Omega$. Let $p(x)$ be the probability mass function of X, that is, $p(x) = P(X = x)$. The expectation of X is defined as

$$\mu = E(X) = \sum_i x_i p(x_i) = \sum_i X(\omega_i) P(\{\omega_i\}).$$

whenever the sum is convergent. Let g be a function $g : \mathbb{R} \to \mathbb{R}$. Recalling that $Y = g(X)$ is a random variable, the expectation of $g(X)$ is

$$E(g(X)) = \sum_i g(x_i) p(x_i).$$

In particular, the kth moment of X is

$$E(X^k) = \sum_i x_i^k p(x_i)$$

and the kth central moment is defined as

$$E((X-\mu)^k) = \sum_i (x_i - \mu)^k p(x_i) \quad \text{for} \quad k = 1, 2, \dots.$$

The variance of X is defined as the second central moment:

$$\mathrm{Var}(X) = E((X-\mu)^2).$$

It is useful to note that for constants a and b, $E(ag(X)+bh(X)) = aE(g(X))+bE(h(X))$. So, for example, $\mathrm{Var}(X) = E((X-\mu)^2) = E(X^2) - \mu^2$. Before considering expectations of functions of a continuous random variable, the following example is given.

Example 1.10. Mean and variance for a Poisson distributed variable

Consider the mean and variance for the Poisson distributed variable of Example 1.6, where $\gamma = \lambda t$. Clearly,

$$E(X) = \sum_{k=0}^{\infty} kp(k) = \sum_{k=0}^{\infty} ke^{-\gamma}\frac{\gamma^k}{k!} = e^{-\gamma}\gamma\sum_{k=1}^{\infty}\frac{\gamma^{k-1}}{(k-1)!} = \gamma$$

and

$$E(X^2) = \sum_{k=0}^{\infty} k^2 p(k) = e^{-\gamma}\gamma^2\sum_{k=1}^{\infty}\left(\frac{(k-1)\gamma^{k-2}}{(k-1)!} + \frac{\gamma^{k-2}}{(k-1)!}\right) = \gamma^2+\gamma.$$

Thus, $E(X) = \gamma$ and $\mathrm{Var}(X) = \gamma$.

Expectations are defined for continuous random variables in an analogous way as for discrete random variables. Suppose that X is a continuous random variable where $X(x) = x$ and with probability density $p(x)$. Notice that $p(x)\Delta x$ is the approximate probability that X takes a value in the interval $(x-\Delta x/2, x+\Delta x/2)$. It follows that the expectation of X can be approximated as $E(X) \approx \sum xp(x)\Delta x$ and as $\Delta x \to 0$,

$$E(X) = \int_{-\infty}^{\infty} xp(x)\,dx.$$

The expectation of a function g of X is defined as

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)p(x)\,dx.$$

In particular, the kth moment of X is defined as

$$E(X^k) = \int_{-\infty}^{\infty} x^k p(x)\,dx.$$

Letting $\mu = E(X)$, the variance of X is defined as the second central moment

$$\mathrm{Var}(X) = E((X-\mu)^2) = E(X^2) - \mu^2.$$

Example 1.11. Mean and variance for a uniformly distributed random variable
Let $X \sim U[a,b]$. Then $P(c \le X \le d) = \int_c^d p(x)\,dx = \int_c^d \frac{dx}{b-a} = \frac{d-c}{b-a}$, assuming that $a \le c \le d \le b$. It is straightforward to see that

$$E(X) = \int_a^b x\,\frac{dx}{b-a} = \frac{b+a}{2}$$

and

$$E(X^2) = \int_a^b x^2\,\frac{dx}{b-a} = \frac{1}{3}(b^2+ab+a^2).$$

Thus, $\mathrm{Var}(X) = \frac{1}{12}(b-a)^2$. Also, for example, $E(e^x) = \int_a^b e^x\,\frac{dx}{b-a} = \frac{e^b-e^a}{b-a}$. Indeed, $E(f(X)) = \int_a^b f(x)\,\frac{dx}{b-a}$.

Example 1.12. Central moments of a normally distributed variable
Let $X \sim N(\mu,\sigma^2)$. Then, one obtains for any positive integer k that

$$E((X-\mu)^{2k+1}) = 0 \quad \text{and} \quad E((X-\mu)^{2k}) = 1\cdot 3\cdot 5\cdots(2k-1)\sigma^{2k}.$$

1.5 Multiple Random Variables

Consider a random experiment with sample space Ω, set of events $\mathcal{A}$, and probability measure P. Let X_1 and X_2 be two random variables defined on this probability space. The bivariate random variable or random vector $\mathbf{X} = [X_1, X_2]^T$ maps Ω into $\mathbb{R}^2$. Notice that if $A_1, A_2 \in \mathcal{A}$ such that $A_1 = \{\omega \in \Omega : X_1(\omega) \le x_1\}$ and $A_2 = \{\omega \in \Omega : X_2(\omega) \le x_2\}$ then $A_1 \cap A_2 \in \mathcal{A}$ and

$$P(X_1 \le x_1, X_2 \le x_2) = P(A_1 \cap A_2).$$

The joint cumulative distribution function of X_1 and X_2 is denoted $F_{X_1X_2}(x_1,x_2)$ and is defined as

$$F_{X_1X_2}(x_1,x_2) = P(X_1 \le x_1, X_2 \le x_2) = P(A_1 \cap A_2).$$

In addition, if A_1 and A_2 are independent events, then

$$F_{X_1X_2}(x_1,x_2) = P(A_1 \cap A_2) = P(A_1)P(A_2) = F_{X_1}(x_1)F_{X_2}(x_2).$$

Suppose that X_1, X_2 are discrete random variables that take on the values $(x_{1,i}, x_{2,j})$ for $1 \le i \le M, 1 \le j \le N$. Let

$$p_{X_1X_2}(x_{1,i}, x_{2,j}) = P(X_1 = x_{1,i}, X_2 = x_{2,j}).$$

Then $p_{X_1X_2}$ is called the joint probability mass function of $\mathbf{X} = [X_1, X_2]^T$ and

$$F_{X_1X_2}(x_1, x_2) = \sum_{x_{1,i} \le x_1} \sum_{x_{2,j} \le x_2} p_{X_1X_2}(x_{1,i}, x_{2,j}).$$

If X_1, X_2 are continuous random variables, $p_{X_1X_2}(x_1, x_2)$ is the joint probability density function of the continuous bivariate random variable $\mathbf{X} = [X_1, X_2]^T$ if

$$F_{X_1X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p_{X_1X_2}(s_1, s_2)\, ds_2\, ds_1.$$

Furthermore, the conditional probability mass density or probability density functions satisfy

$$p_{X_1|X_2}(x_1|x_2)p_{X_2}(x_2) = p_{X_1X_2}(x_1, x_2)$$

and

$$p_{X_2|X_1}(x_2|x_1)p_{X_1}(x_1) = p_{X_1X_2}(x_1, x_2).$$

The variance of X_1 and X_2 satisfies

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1X_2)$$

where the covariance of X_1 and X_2 is defined as

$$\mathrm{Cov}(X_1X_2) = E((X_1 - \mu_1)(X_2 - \mu_2)) = E(X_1X_2) - E(X_1)E(X_2).$$

If $p_{X_1X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$ and it follows that $F_{X_1X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$, then X_1 and X_2 are said to be independent random variables. Notice that if X_1 and X_2 are independent, then

$$\begin{aligned} E(f(X_1)g(X_2)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1)g(x_2)p_{X_1X_2}(x_1, x_2)\, dx_1\, dx_2 \\ &= E(f(X_1))E(g(X_2)) \end{aligned}$$

and

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$$

where $\mathrm{Cov}(X_1X_2) = 0$ if X_1 and X_2 are independent.

If $\mathbf{X} = [X_1, X_2, \ldots, X_n]^T$ is a vector of n random variables each defined on sample space Ω, then $\boldsymbol{\mu} = E(\mathbf{X})$ is the mean vector of length n and $E((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T)$ is the $n \times n$ matrix defined to be the covariance matrix. The probability distribution $F_{\mathbf{X}}$ is related to the joint probability density $p_{\mathbf{X}}$ by

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} p_{\mathbf{X}}(s_1, s_2, \ldots, s_n)\, ds_n \ldots ds_2\, ds_1.$$

Example 1.13. Correlated random variables in coin tossing

Suppose that a coin is flipped twice and the sample space is $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ where $\omega_1 = HH, \omega_2 = HT, \omega_3 = TH$, and $\omega_4 = TT$. Also, $P(\{\omega_i\}) = \frac{1}{4}$ for $i = 1, 2, 3, 4$. Let $X_1(\omega_i) =$ number of heads in outcome ω_i and $X_2(\omega_i) =$ number of tails in outcome ω_i. Then, it is easy to show that the joint probability mass function has the form

$$p_{X_1X_2}(x_1, x_2) = \begin{cases} 1/4, & x_1 = 2, x_2 = 0 \quad \text{or} \quad x_1 = 0, x_2 = 2 \\ 1/2, & x_1 = 1, x_2 = 1 \\ 0, & \text{otherwise.} \end{cases}$$

For this random experiment, $E(\mathbf{X}) = \boldsymbol{\mu} = [1, 1]^T$ and the covariance matrix is

$$\begin{aligned} E((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T)) &= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_1 - \mu_1)(X_2 - \mu_2) & (X_2 - \mu_2)^2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

Notice that X_1 and X_2 are not independent and $\text{Cov}(X_1X_2) = -\frac{1}{2} \neq 0$.

Example 1.14. Correlated normally distributed random variables

Let $G_1, G_2 \sim N(0, 1)$ be independent normally distributed random variables. Then $E(G_1G_2) = E(G_1)E(G_2) = 0$ and $\text{Var}(G_1 + G_2) = 2$. Let $X_1 = h^{\frac{1}{2}}G_1$ and $X_2 = \frac{1}{2}h^{\frac{3}{2}}(G_1 + \frac{1}{\sqrt{3}}G_2)$ for a constant $h > 0$. Then $X_1 \sim N(0, h)$ and $X_2 \sim N(0, h^3/3)$. However, X_1 and X_2 are correlated Gaussian random variables and $E(X_1X_2) = h^2/2 \neq E(X_1)E(X_2) = 0$.

Before describing more examples, it is useful to consider functions of several random variables. Given two random variables X_1 and X_2 and a function $g : \mathbb{R}^2 \to \mathbb{R}$, then $X = g(X_1, X_2)$ defines a new random variable. If $p_{X_1X_2}(x_1, x_2)$ is the joint probability density of $\mathbf{X} = [X_1, X_2]^T$, then the expectation of X is

$$E(g(X_1, X_2)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) p_{X_1X_2}(x_1, x_2)\, dx_1\, dx_2.$$

In addition, letting $D_x = \{(x_1, x_2) \in \mathbb{R}^2 : g(x_1, x_2) \leq x\}$, then $F_X(x) = P(X \leq x) = P(g(x_1, x_2) \leq x) = P(D_x)$ is the probability distribution function of X.

Example 1.15. Mean distance between two uniformly distributed points

Let X_1 and X_2 be independent uniformly distributed random variables on $[0, 1]$, that is, $X_1, X_2 \sim U[0, 1]$. Let $g(X_1, X_2) = |X_1 - X_2|$ be the distance between two points. In this example, it is desired to find the mean

distance between the points, that is, $E(g(X_1, X_2))$. Clearly, as X_1 and X_2 are independent,

$$\begin{aligned} E(g(X_1, X_2)) &= \int_0^1 \int_0^1 |x_1 - x_2| p_{X_1 X_2}(x_1, x_2)\, dx_1\, dx_2 \\ &= \int_0^1 \int_0^1 |x_1 - x_2| p_{X_1}(x_1) p_{X_2}(x_2)\, dx_1\, dx_2 \\ &= \int_0^1 \int_0^1 |x_1 - x_2|\, dx_1\, dx_2 \\ &= \int_0^1 \left(\int_0^{x_2} (x_2 - x_1)\, dx_1 + \int_{x_2}^1 (x_1 - x_2)\, dx_1 \right) dx_2 = \frac{1}{3} \end{aligned}$$

and hence, the mean distance between two randomly selected points on $[0, 1]$ is $1/3$.

Example 1.16. Sum of two uniformly distributed random variables

Let $X_1, X_2 \sim U[0, 1]$ be independent uniformly distributed random variables and notice that $X_1, X_2 : [0, 1] \to [0, 1]$. Let $X = X_1 + X_2$. Then $X : [0, 1] \to [0, 2]$. It is of interest to find the probability density function for X. It follows that

$$F_X(x) = \int_{-\infty}^{x} p_X(s)\, ds,$$

where

$$p_X(s) = \begin{cases} s, & 0 < s < 1 \\ 2 - s, & 1 \le s \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

It is easy, for example, to compute $E((X_1 + X_2)^2) = \int_0^1 \int_0^1 (x_1 + x_2)^2\, dx_1\, dx_2$ and verify that $E((X_1 + X_2)^2) = E(X^2)$, where

$$E(X^2) = \int_{-\infty}^{\infty} s^2 p_X(s)\, ds = \frac{7}{6}.$$

Example 1.17. Sum of two Poisson distributed random variables

Let X_1 and X_2 be two independent Poisson variables with $E(X_1) = \mu_1 = \gamma_1$ and $E(X_2) = \mu_2 = \gamma_2$. Let $X = X_1 + X_2$. Then X is Poisson distributed with mean $\mu = \gamma_1 + \gamma_2$. By direct calculation,

$$\begin{aligned} P(X = x) &= \sum_{x_1=0}^{x} P(X_1 = x_1) P(X_2 = x - x_1) \\ &= \sum_{x_1=0}^{x} e^{-\gamma_1} \frac{\gamma_1^{x_1}}{x_1!} e^{-\gamma_2} \frac{\gamma_2^{x - x_1}}{(x - x_1)!} \\ &= \sum_{x_1=0}^{x} e^{-(\gamma_1 + \gamma_2)} \frac{\gamma_1^{x_1} \gamma_2^{x - x_1}}{x_1!(x - x_1)!} = e^{-(\gamma_1 + \gamma_2)} \frac{(\gamma_1 + \gamma_2)^x}{x!}. \end{aligned}$$

This result implies that X is Poisson distributed with mean $\mu = \gamma_1 + \gamma_2$.

Example 1.18. Sum of two normally distributed random variables

Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ be two independent normally distributed random variables. Then, $X = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Therefore, $X = X_1 + X_2$ is also normally distributed but with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. This important result is shown for the special case of two standard normal distributions, i.e., $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$. In this case,

$$\begin{aligned}
F_X(x) = P(X_1 + X_2 \le x) &= \int_{-\infty}^{\infty} \int_{-\infty}^{x-s_1} p_{X_1X_2}(s_1, s_2)\, ds_2\, ds_1 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x-s_1} \frac{1}{\sqrt{2\pi}} e^{\frac{-s_2^2}{2}} e^{\frac{-s_1^2}{2}}\, ds_2\, ds_1 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-(z-s_1)^2}{2}} e^{\frac{-s_1^2}{2}}\, dz\, ds_1 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-z^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-s_1^2 + s_1 z}\, ds_1\, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-z^2}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(s_1 - z/2)^2}\, ds_1\, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-z^2}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2}\, du\, dz = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-z^2}{4}}\, dz
\end{aligned}$$

which implies that $X \sim N(0, 2)$ and thus X is normally distributed with mean 0 and variance 2.

Illustrated in the previous two examples is an interesting and important property of Poisson and normal random variables. Specifically, the sum of two independently distributed normal variables is also normally distributed and the sum of two independently distributed Poisson variables is also Poisson distributed.

1.6 A Hilbert Space of Random Variables

Some properties of Hilbert spaces are discussed in this section. Hilbert spaces of random variables and stochastic processes unify and simplify the development of stochastic integration and stochastic differential equations presented in Chapters 3 and 4. A vector space with a metric or norm defined on it is called a metric space or a normed linear space. If the metric space is complete, i.e., all Cauchy sequences converge in the space, then the metric space is called a Banach space. A complete inner product space is a special type of Banach space. In a complete inner product space, the metric $\|\cdot\|$ is defined in terms of an inner product $(\cdot, \cdot)$, specifically, $\|f\| = (f, f)^{\frac{1}{2}}$ for any f in the space. Complete inner product spaces are very useful and are called Hilbert spaces.

Let H be a Hilbert space with inner product $(f,g) \in \mathbb{R}$ for $f,g \in H$ and norm $\|f\| = (f,f)^{\frac{1}{2}}$. Two important inequalities for Hilbert spaces are the triangle inequality and the Cauchy-Schwarz inequality:

$$\|f+g\| \leq \|f\| + \|g\|$$

$$|(f,g)| \leq \|f\|\|g\|.$$

Suppose that S is an inner product space but S is not complete. A useful result [65, 73] is that S can be completed to form a Hilbert space H by adding elements to S. Furthermore, this can be accomplished so that $S \subset H$ is dense in H. Hence, if S is completed to form a Hilbert space H, then given any $f \in H$ and given any $\epsilon > 0$, there is a $g \in S$ so that $\|f-g\| < \epsilon$.

Now let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $A \in \mathcal{A}$ and let I_A be the indicator function for A. That is, I_A is the random variable defined so that

$$I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \text{otherwise.} \end{cases}$$

Then, for example, $E(I_A) = P(A)$. Finite linear combinations of indicator functions are called simple random variables. If X is a simple random variable, then X has the form

$$X(\omega) = \sum_{i=1}^{n} c_i I_{A_i}(\omega) \quad \text{and} \quad E(X) = \sum_{i=1}^{n} c_i P(A_i).$$

Now let S_{RV} be the set of simple random variables defined on the probability space. Specifically, $S_{RV} = \{X : X \text{ is a simple random variable defined on probability space } (\Omega, \mathcal{A}, P)\}$. The set S_{RV} is a vector space of random variables. Now let $X, Y \in S_{RV}$. The inner product (X,Y) is defined on S_{RV} as

$$(X,Y) = E(XY) \quad \text{for} \quad X,Y \in S_{RV}.$$

Notice that for $X, Y \in S_{RV}$, then

$$(X,Y) = E(XY) = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} c_i I_{A_i} d_j I_{B_j}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i d_j P(A_i \cap B_j)$$

and the norm has the form

$$\|X\|_{RV} = (X,X)^{\frac{1}{2}} = (E|X|^2)^{\frac{1}{2}}.$$

In general, this inner product space of simple random variables is not complete. However, it can be completed to form a Hilbert space H_{RV} where S_{RV} is dense in H_{RV}. In the act of completing S_{RV}, many random variables may be added to S_{RV} to form H_{RV}. Therefore, suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence

of random variables in H_{RV} such that given $\epsilon > 0$ there is an integer N so that $\|X_n - X_m\|_{RV} < \epsilon$ when $m, n > N$. Then, as H_{RV} is complete, there is a random variable $X \in H_{RV}$ such that $\|X_n - X\|_{RV} \to 0$ as $n \to \infty$. Furthermore, as $S_{RV} \subset H_{RV}$ is dense in H_{RV}, given $\epsilon > 0$ there is a simple random variable $Y \in S_{RV}$ such that $\|X - Y\| < \epsilon$.

The Hilbert space H_{RV}, which we have just constructed, will be very useful in the next three chapters. Notice, in particular, that the inner product on this space is $(X, Y) = E(XY)$, the norm for this Hilbert space is $\|X\|_{RV} = (E(|X|^2))^{1/2}$, the set of simple functions S_{RV} is dense in H_{RV}, and H_{RV} depends on the probability distribution through the expectation E. Some examples are given below to illustrate the usefulness of this Hilbert space. First, for emphasis:

H_{RV} is a Hilbert space of Random Variables with norm $\|X\|_{RV} = (E(|X|^2))^{1/2}$.

Example 1.19. Hilbert space $L_2[0, 1]$

Consider Example 1.3 where the sample space is the set of points on $[0, 1]$, i.e., $\Omega = \{x : 0 \le x \le 1\}$. The event space $\mathcal{A}$ for this example is the σ-algebra of sets generated by all intervals of the form $(a, b] \subset [0, 1]$. The probability measure P is Lebesgue measure where $P(A) = b - a$ if $A = (a, b] \in \mathcal{A}$. Let S_{RV} be all simple random variables defined on $\mathcal{A}$. If $X \in S_{RV}$, then the random variable X has the form

$$X(x) = \sum_{i=1}^{n} c_i I_{A_i}(x), \quad \text{where} \quad A_i \in \mathcal{A} \quad \text{for each} \quad i$$

and

$$I_{A_i}(x) = \begin{cases} 1, & x \in A_i \\ 0, & \text{otherwise.} \end{cases}$$

Let H_{RV} be the completion of S_{RV}. Hilbert space H_{RV} includes, for example, all random variables that are continuous on $[0, 1]$. To see that continuous random variables are included in H_{RV}, let $f : [0, 1] \to \mathbb{R}$ be a continuous function. Let $x_i = (i - 1)/n$ for $i = 1, 2, \ldots, n$ and define

$$f_n(x) = \sum_{i=1}^{n} f(x_i) I_{n,i}(x), \quad \text{where} \quad I_{n,i}(x) = \begin{cases} 1, & (i-1)/n \le x < i/n \\ 0, & \text{otherwise.} \end{cases}$$

Then it can be shown that this sequence of simple random variables $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence in H_{RV}. In addition, $\|f - f_n\|_{RV} \to 0$ as $n \to \infty$ so $f_n \to f$ in H_{RV}. Thus, f is the limit of a sequence of simple random variables in H_{RV} and $f \in H_{RV}$. Also notice that if $X(x) = x$, then X is uniformly distributed on $[0, 1]$, i.e., $X \sim U[0, 1]$.

The Hilbert space H_{RV} for this example is the well-known space $L_2[0, 1]$, that is, $H_{RV} = L_2[0, 1] = \{$Lebesgue measurable functions f on [0,1] such

that $\int_0^1 |f(x)|^2\,dx < \infty$ }. In particular, many of the functions needed in the completion of H_{RV} may be highly discontinuous. Therefore, integration in the Riemann sense is not possible for many of these functions and integration is defined in the Lebesgue sense. However, if the Riemann integral exists, then the Lebesgue integral exists and the two integrals are equal. Furthermore, as piecewise continuous functions which are square integrable are dense in $L_2[0,1]$, this Hilbert space can be roughly considered as the set of piecewise continuous functions which are square integrable on $[0,1]$. Notice that for $X, Y \in H_{RV}$ then

$$(X,Y) = \int_0^1 X(x)Y(x)\,dx \quad \text{and} \quad \|X\|_{RV}^2 = \int_0^1 |X(x)|^2\,dx.$$

Example 1.20. Convergence example in Hilbert space $H_{RV} = L_2[0,1]$
Let H_{RV} be defined as in the previous example. Let $Y \sim U[0,1]$ and define the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ by

$$X_n(x) = \begin{cases} \frac{1}{2}Y(x), & 1/n \le Y(x) \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then, it is straightforward to see that $\|X_n - X_m\|_{RV} \to 0$ as $m, n \to \infty$. Thus, $\{X_n\} \subset H_{RV}$ is a Cauchy sequence in H_{RV}. Indeed, X_n converges in H_{RV} to $X = \frac{1}{2}Y$ as $n \to \infty$.

Example 1.21. A nonconvergent example
Let H_{RV} be defined as in Example 1.20. Let $Y_n \sim U[0,1]$ be independent uniformly distributed random variables for $n = 1, 2, \ldots$. Then

$$\|Y_n\|_{RV} = \left(\int_0^1 x^2\,dx\right)^{\frac{1}{2}} = \frac{1}{\sqrt{3}} \quad \text{for each } n.$$

Let $X = 1$ and define the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ by

$$X_n(x) = \begin{cases} 1, & 1/\sqrt{n} \le Y_n(x) \le 1 \\ 1 + nY_n(x), & \text{otherwise.} \end{cases}$$

Then

$$\|X_n - X\|_{RV} = \left(\int_0^{1/\sqrt{n}} n^2x^2\,dx\right)^{\frac{1}{2}} = \left(\frac{\sqrt{n}}{3}\right)^{\frac{1}{2}} \to \infty \quad \text{as } n \to \infty.$$

Indeed, the sequence $\{X_n\}_{n=1}^{\infty}$ is not a Cauchy sequence in Hilbert space H_{RV}.

Example 1.22. A weighted normed Hilbert space
Consider the probability space described in Example 1.9. In particular, the sample space is $\Omega = \{x : -\infty < x < \infty\}$ and the event space $\mathcal{A}$ is the

σ-algebra generated by intervals of the form $(a, b]$ on $\mathbb{R}$. Let $\mu \in \mathbb{R}$ and $\sigma > 0$ be given constants. Let $P(A) = \int_A p(s)\, ds$ where

$$p(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(s-\mu)^2}{2\sigma^2}\right).$$

Recall that if the random variable X is defined as $X(x) = x$ then X is said to be normally distributed with mean μ and variance σ^2, i.e., $X \sim N(\mu, \sigma^2)$. Let S_{RV} be the inner product space of simple functions on this probability space where the inner product is defined as

$$(f, g) = E(fg) = \int_{-\infty}^{\infty} f(s)g(s)p(s)\, ds \quad \text{for} \quad f, g \in S_{RV}$$

and, as in Example 1.19, the above integral is a Lebesgue integral. Let H_{RV} be the completion of S_{RV}. Then H_{RV} is a Hilbert space of random variables defined on $\mathbb{R}$ with norm

$$\|f\|_{RV}^2 = \int_{-\infty}^{\infty} |f(s)|^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(s-\mu)^2}{2\sigma^2}\right) ds \quad \text{for} \quad f \in H_{RV}.$$

In addition, it is useful to note that the set of piecewise continuous random variables f such that $\int_{-\infty}^{\infty} |f(s)|^2 p(s)\, ds < \infty$ is dense in H_{RV}.

1.7 Convergence of Sequences of Random Variables

The convergence of sequences of random variables is important in the study of stochastic differential equations. Consider a sequence of random variables $\{X_i\}_{i=1}^{\infty}$ defined on a probability space $(\Omega, \mathcal{A}, P)$. Of interest is the existence of a random variable X to which the sequence approaches as $n \to \infty$. It is important to clearly characterize the manner in which X_n approaches X as $n \to \infty$.

There are several different types of convergence criteria which are useful to understand for random variables. An important kind of convergence is mean square convergence which is the kind that is used throughout this book. In mean square convergence, one wishes to show that $\lim_{n\to\infty} E(X_n - X)^2 = 0$. However, for $\{X_i\}_{i=1}^{\infty} \subset H_{RV}$, this convergence is equivalent to $\|X_n - X\|_{RV} \to 0$ as $n \to \infty$. Furthermore, as H_{RV} is a Hilbert space, the existence of the random variable $X \in H_{RV}$ is guaranteed if $\{X_i\}_{i=1}^{\infty}$ is a Cauchy sequence in H_{RV}, i.e., given $\epsilon > 0$ there exists a positive integer N such that $\|X_n - X_m\|_{RV} < \epsilon$ whenever $m, n > N$. Other types of convergence criteria are briefly discussed in this section.

Consider first strong convergence. The sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is said to converge strongly to X if $\lim_{n\to\infty} E(|X_n - X|) = 0$. It is interesting that mean square convergence implies strong convergence. This result

follows from the Cauchy-Schwarz inequality for $X \in H_{RV}$ or from the Lyapunov inequality [51, 106]:

Lyapunov Inequality: $\left(E(|X|^p)\right)^{1/p} \leq \left(E(|X|^r)\right)^{1/r}$ for $0 < p < r$.

In particular, for $p = 1$ and $r = 2$, Lyapunov's inequality reduces to the inequality $E(|X|) \leq \left(E(|X|^2)\right)^{1/2}$.

A third form of convergence is convergence in probability. The sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is said to converge in probability to X if given any $\epsilon > 0$ then $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$. Mean square convergence also implies convergence in probability. This result follows from the Chebyshev-Markov inequality [51, 106] stated below.

Chebyshev-Markov Inequality: $P(\{\omega : |X(\omega)| \geq \epsilon\}) \leq \frac{1}{\epsilon^p} E(|X|^p)$ for any $p, \epsilon > 0$.

The Chebyshev-Markov inequality, for $p = 2$, reduces to $P(|X| \geq \epsilon) \leq \frac{1}{\epsilon^2} E(|X|^2)$. Although mean square convergence implies convergence in probability as well as strong convergence, convergence in probability or strong convergence does not necessarily imply mean square convergence.

Another type of convergence is convergence with probability one (w.p.1) or almost sure convergence. The sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is said to converge w.p.1 to X if $P\left(\left\{\omega \in \Omega : \lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0\right\}\right) = 1$. The following result [51] is sometimes useful in determining almost sure convergence.

Lemma on Almost Sure Convergence: If $\sum_{n=1}^{\infty} P(|X_n - X| \geq \epsilon) < \infty$ for all $\epsilon > 0$, then X_n converges almost surely to X.

Almost sure convergence implies convergence in probability. However, mean square convergence does not imply almost sure convergence and almost sure convergence does not imply mean square convergence [51].

There are two other weaker forms of convergence which are useful. Either of these forms of convergence is implied by almost sure convergence or by mean square convergence. The first is convergence in distribution. The sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is said to converge in distribution to X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points of continuity of the distribution function F_X. In the second form of convergence, referred to as weak convergence, the sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is said to converge weakly to X if

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f(y)\,dF_{X_n}(y) = \int_{-\infty}^{\infty} f(y)\,dF_X(y)$$

for all smooth functions f. A sequence converges weakly if and only if it converges in distribution [20].

Several examples are given to illustrate these forms of convergence.

Example 1.23. Almost sure and mean square convergence

Let X be a random variable uniformly distributed on $[0,1]$, i.e., $X \sim U[0,1]$, and define the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ by

$$X_n(\omega) = \begin{cases} 0, & 0 \le X(\omega) \le 1/n^2 \\ X(\omega), & 1/n^2 < X(\omega) \le 1 \end{cases}$$

for $n = 1, 2, \ldots,$. Then

$$\sum_{n=1}^{\infty} P(|X_n - X| \ge \epsilon) \le \sum_{n=1}^{\infty} 1/n^2 < \infty$$

for any $\epsilon > 0$. Therefore, $X_n \to X$ almost surely. Also notice that

$$E(|X_n - X|^2) = \int_0^{1/n^2} x^2\,dx = \frac{1}{3n^6} \to 0 \quad \text{as} \quad n \to \infty.$$

Hence, X_n converges to X in the mean square sense also.

Example 1.24. Weak but not mean square convergence

As in Example 1.19 let the sample space be $\Omega = \{x : 0 \le x \le 1\}$ with the event space $\mathcal{A}$ the σ- algebra of Borel sets generated by intervals of the form $(a, b]$ in $[0, 1]$. Let

$$F_{X_n}(x) = \int_0^x p_n(s)\,ds,$$

where

$$p_n(s) = \begin{cases} \dfrac{n}{n-2}, & s \in [1/n, 1-1/n] \\ 0, & \text{otherwise} \end{cases} \quad \text{for} \quad n \ge 3.$$

Let

$$F_X(x) = \int_0^x p(s)\,ds, \quad \text{where} \quad p(s) = \begin{cases} 1, & s \in [0,1] \\ 0, & \text{otherwise.} \end{cases}$$

That is, $X \sim U[0,1]$ and $X_n \sim U[1/n, 1-1/n]$, where X and X_n are independent. Then for any $f \in C[0,1]$,

$$\lim_{n\to\infty}\int_{1/n}^{1-1/n} f(x)p_n(x)\,dx = \int_0^1 f(x)p(x)\,dx$$

so X_n converges weakly to X. Notice, as X_n and X are independent for each n, that

$$E(|X - X_n|^2) = E(X^2 - 2XX_n + X_n^2) \to 1/6 \quad \text{as} \quad n \to \infty$$

so X_n does not converge in the mean square sense to X.

Two important results involving sequences of random variables are the Law of Large Numbers and the Central Limit Theorem [51, 63, 67, 87]. These are stated below for convenience.

Law of Large Numbers (Strong): Let $X_1, X_2, \ldots$ be independent and identically distributed random variables. Let $\mu = E(X_n)$ and $\sigma^2 = \text{Var}(X_n)$. Define $S_n = \sum_{i=1}^{n} X_i$. Then $S_n/n \to \mu$ almost surely and in the mean square sense. That is,

$$\lim_{n\to\infty} E\left(\left|\frac{S_n}{n} - \mu\right|^2\right) = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{S_n}{n} = \mu \quad \text{w.p.1.}$$

Central Limit Theorem: Define S_n as above and let $Z_n = (S_n - n\mu)/(\sigma\sqrt{n})$. Then Z_n converges in distribution to $Z \sim N(0,1)$. That is,

$$\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x),$$

where F_{Z_n} is the distribution function of Z_n and F_Z is the standard normal distribution function.

Example 1.25. Sum of Poisson distributed variables

Let X_i for $i = 1, 2, \ldots$ be independent Poisson distributed variables with parameter γ. That is, $E(X_i) = \text{Var}(X_i) = \gamma$ or $\mu = \sigma^2 = \gamma$. Let $S_n = \sum_{i=1}^{n} X_i$. Then, by Example 1.17, it is known that S_n is Poisson distributed with mean $n\gamma$ and variance $n\gamma$. However, by the Central Limit Theorem, $\dfrac{S_n - n\gamma}{\sqrt{n\gamma}}$ converges in distribution to a normally distributed random variable with mean 0 and variance 1. Now let $\lambda = n\gamma$ and let $R_\lambda = S_n$. Then R_λ is Poisson distributed with parameter λ. However, by the above argument, as λ increases then $\dfrac{R_\lambda - \lambda}{\sqrt{\lambda}}$ converges in distribution to a normally distributed random variable with mean 0 and variance 1. This implies that a Poisson distributed variable with parameter λ converges in distribution to a normally distributed variable with mean λ and variance λ as the value of λ increases.

1.8 Computer Generation of Random Numbers

To approximate solutions to stochastic problems using a computer, large numbers of random numbers are required [11, 17, 35, 37, 84]. Simple deterministic algorithms to generate sequences of random variables are called pseudo-random number generators. There are many kinds of pseudo-random number generators for producing uniformly distributed random numbers on [0,1]. A popular type is the linear congruential generator which has the form:

$$X_{n+1} = (aX_n + c) \bmod(m), \quad \text{for} \quad n = 0, 1, 2, \ldots,$$

where a, c, and m are nonnegative integers with m typically large and X_0 is a starting number. As $d \bmod(m)$ is the remainder when dividing d by m, then $0 \leq d \bmod(m) \leq m - 1$. The sequence U_n is calculated where

$$U_n = \frac{X_n}{m} \quad \text{for} \quad n = 0, 1, 2, \ldots$$

and $0 \leq U_n \leq 1$ for each n. For certain values of the parameters a, c, and m, this sequence of numbers may possess statistical properties for numbers that are randomly distributed uniformly on $[0, 1]$.

Linear congruential generators eventually repeat. If $X_{i+p} = X_i$, then the smallest such value of p is called the cycle length or period of the generator. For a linear congruential generator, $p \leq m$. The following result [84] is useful for determining the cycle length of certain linear congruential generators when $c \neq 0$.

Lemma on Period Length: The period of a linear congruential generator is m if and only if

(i) c and m are relatively prime,
(ii) every prime factor of m divides $a - 1$, and
(iii) if 4 divides m, then 4 divides $a - 1$.

When $c = 0$ and m is a prime number, the longest possible period is $m-1$ when a satisfies the property that $a^k - 1$ is not divisible by m for $k = 1, 2, \ldots, m-2$. A popular linear congruential generator that satisfies this property is [17]:

$$X_{n+1} = 16807X_n \bmod(2^{31} - 1), \quad \text{for} \quad n = 0, 1, 2, \ldots.$$

Here $a = 7^5, c = 0$, and $m = 2^{31} - 1 = 2,147,483,647$, and the cycle length of this generator is $m - 1$. Notice that $m = 2^{31} - 1$ is a Mersenne prime for this generator.

Assume now that we have generated a sequence $\{U_n\}$ of uniformly distributed numbers on $[0, 1]$. In addition, suppose that we need a sequence $\{Y_n\}$ which are distributed according to a distribution F_Y which may not be a uniform distribution but is strictly monotone increasing. One way to compute the sequence $\{Y_n\}$ from the sequence $\{U_n\}$ is to set $Y = g(U)$ for a function g. Now notice that $g^{-1}(Y) = U$. To find g^{-1}, consider

$$F_Y(y) = P(\{g(U) \leq y\}) = P(\{U \leq g^{-1}(y)\}) = \int_0^{g^{-1}(y)} ds$$
$$= g^{-1}(y) \quad \text{for} \quad 0 \leq g^{-1}(y) \leq 1.$$

If $F_Y(y) = \int_{-\infty}^{y} p_y(s)\, ds$, then

$$\int_{-\infty}^{Y_n} p_y(s)\, ds = U_n \quad \text{for} \quad n = 1, 2, 3, \ldots$$

and often this formula can be used to compute Y_n given that U_n for $n = 1, 2, 3, \ldots$ are $U[0, 1]$ uniformly distributed pseudo-random numbers.

Example 1.26. Generation of exponentially distributed random numbers

Suppose that we need the sequence $\{Y_n\}$ to be exponentially distributed on $[0, \infty)$. It is assumed that we have the sequence $\{U_n\}$ of uniformly distributed numbers on $[0, 1]$. In this case, the probability density has the form $p_y(s) = \exp(-s)$ for $s \geq 0$. By the previous argument, we need to find Y_n such that

$$\int_0^{Y_n} \exp(-s)\, ds = 1 - \exp(-Y_n) = U_n \quad \text{for} \quad n = 1, 2, 3, \ldots.$$

Thus, Y_n is given by $Y_n = -\ln(1 - U_n)$ for $n = 1, 2, \ldots$.

Example 1.27. Generation of Poisson distributed numbers

Suppose that we need the sequence $\{Y_n\}$ to be Poisson distributed with parameter γ, i.e., $E(Y_n) = \gamma$. It is assumed that we have the sequence $\{U_n\}$ of uniformly distributed numbers on $[0, 1]$. By Example 1.6, we know that

$$F_Y(y) = e^{-\gamma} \sum_{k=0}^{m} \frac{\gamma^k}{k!} \quad \text{for} \quad m \leq y < m + 1.$$

Therefore, to find Y_n given U_n, the sum is computed until m is found such that

$$e^{-\gamma} \sum_{k=0}^{m-1} \frac{\gamma^k}{k!} < U_n \leq e^{-\gamma} \sum_{k=0}^{m} \frac{\gamma^k}{k!}.$$

Then Y_n is set equal to m. In particular, notice that if $\gamma = \lambda \Delta t$ and $\lambda \Delta t$ is much less than unity, then to $o(\Delta t)$,

$$Y_n = \begin{cases} 0 & \text{if} \quad 0 < U_n \leq 1 - \lambda \Delta t \\ 1 & \text{if} \quad 1 - \lambda \Delta t < U_n \leq 1. \end{cases}$$

For the common situation where normally distributed numbers are required, the above procedure is computationally awkward. This is because if U_n is a $U[0, 1]$ uniformly distributed number, then Y_n is an $N(0, 1)$ normally distributed number provided that Y_n satisfies the integral

$$\int_{-\infty}^{Y_n} \frac{1}{\sqrt{2\pi}} \exp(-s^2/2)\, ds = U_n.$$

This integral cannot be solved analytically and must be evaluated numerically. For generation of normally distributed numbers, the Box-Muller method [26] is effective. In this method, two independent uniformly distributed numbers on $[0, 1]$ are transformed into two independent standard normally distributed numbers. The Box-Muller method uses the transformation between random variables

$$X_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2)$$
$$X_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2).$$

If $U_1, U_2 \sim U[0,1]$, then $X_1, X_2 \sim N(0,1)$. To see why this transformation is valid, let $R = \sqrt{-2\ln(U_1)}$ and $\Theta = 2\pi U_2$. Then, R and Θ are independent and have the joint probability density

$$p(r,\theta) = \frac{1}{2\pi} r\exp(-r^2/2) = p_1(r)p_2(\theta) \quad \text{for} \quad r \in [0,\infty) \quad \text{and} \quad \theta \in [0, 2\pi].$$

Thus,

$$P(R \le r^*, \Theta \le \theta^*) = \int_0^{r^*}\int_0^{\theta^*} \frac{1}{2\pi} r\exp(-r^2/2)\, dr\, d\theta.$$

Converting to polar coordinates, with $R^2 = X_1^2 + X_2^2$, $\sin\Theta = X_2/R$, $\cos\Theta = X_1/R$, leads to

$$P(X_1 \le x_1^*, X_2 \le x_2^*) = \int_{-\infty}^{x_1^*}\int_{-\infty}^{x_2^*} \frac{1}{2\pi}\exp(-(x_1^2 + x_2^2)/2)\, dx_1\, dx_2.$$

But this is the joint probability distribution of independent random variables $X_1, X_2 \sim N(0,1)$. In addition,

$$X_1 = R\cos\Theta = \sqrt{-2\ln(U_1)}\cos(2\pi U_2)$$

and

$$X_2 = R\sin\Theta = \sqrt{-2\ln(U_1)}\sin(2\pi U_2).$$

Finally, notice that if $X \sim N(0,1)$ then $\mu + \sigma X \sim N(\mu, \sigma^2)$. Thus, standard normally distributed random numbers can be easily converted to normally distributed random numbers having mean μ and variance σ^2.

It is useful to point out that the Polar-Marsaglia method [70] is similar to the Box-Muller method but avoids trigonometric function evaluations. In this method, let $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, and $Z = \sqrt{V_1^2 + V_2^2}$. If $Z > 1$, then U_1, U_2, V_1, V_2, and Z are recomputed. Otherwise, the independent $N(0,1)$ normally distributed random numbers X_1, X_2 are computed as follows:

$$X_1 = \frac{V_1}{Z}\sqrt{-4\ln(Z)}$$
$$X_2 = \frac{V_2}{Z}\sqrt{-4\ln(Z)}.$$

1.9 Monte Carlo

Monte Carlo methods [17, 35, 37, 84] are numerical methods where random numbers are used to conduct computational experiments. Numerical solution

of stochastic differential equations can be viewed as a type of Monte Carlo calculation. It is useful to understand some basic properties of Monte Carlo computation. Monte Carlo methods will be illustrated in the following examples.

Example 1.28. Monte Carlo Estimation of an Integral

Consider the classic Monte Carlo problem of estimating the value of an integral [35]. A one-dimensional problem is described in this example but the method easily generalizes to multiple integrals. Consider estimating the integral

$$I(f) = \int_0^1 f(x)\,dx, \quad \text{where} \quad f \in C[0,1].$$

Note that

$$I(f) = E(f) = \int_0^1 f(x)p(x)\,dx, \quad \text{where} \quad p(x) = \begin{cases} 1 & \text{if} \quad 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $I(f)$ is the expected value of random variable f in the Hilbert space discussed in Example 1.19 where x is uniformly distributed on $[0,1]$. Let

$$I_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(x_i), \quad \text{where} \quad x_i \quad \text{are uniformly distributed on} \quad [0,1].$$

Then it follows that $I_n(f) \to I(f)$ as $n \to \infty$. The error can be estimated using the Central Limit Theorem. First, let

$$E(f) = \int_0^1 f(x)\,dx \quad \text{and} \quad E(f^2) = \int_0^1 |f(x)|^2\,dx$$

so

$$\mu = E(f) \quad \text{and} \quad \sigma^2 = E(f^2) - (E(f))^2.$$

The Central Limit Theorem says that the sample mean $I_n(f)$ is approximately normally distributed with mean $I(f)$ and variance σ^2/n. This implies that

$$P\left(I_n(f) - \frac{\lambda\sigma}{\sqrt{n}} \le I(f) \le I_n(f) + \frac{\lambda\sigma}{\sqrt{n}}\right) \approx \frac{1}{\sqrt{2\pi}}\int_{-\lambda}^{\lambda} \exp(-x^2/2)\,dx.$$

Hence, for example, the probability that $|I(f) - I_n(f)| \le \dfrac{1.96\sigma}{\sqrt{n}}$ is equal to 0.95. This example illustrates an important property of Monte Carlo methods. That is, the error in a Monte Carlo calculation is statistical in nature and the error is proportional to $1/\sqrt{n}$, where n is the sample size. This property also holds for numerical solution of stochastic differential equations.

Example 1.29. Monte Carlo Estimation of Mean Distance Between Two Points

In Example 1.15, it was determined that the mean distance between two points randomly picked on the interval $[0,1]$ is exactly $1/3$. In this example,

the mean distance between two random points on the interval $[0,1]$ will be estimated as well as the mean distance between two points picked on the square $[0,1] \times [0,1]$. Each of these problems can be cast in the form of an integral. In the first case,

$$I = \int_0^1 \int_0^1 |x_1 - x_2| \, dx_1 \, dx_2$$

is the mean distance between two randomly-picked points on $[0,1]$ whereas in the second case,

$$J = \int_0^1 \int_0^1 \int_0^1 \int_0^1 \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \, dx_1 \, dx_2 \, dy_1 \, dy_2$$

is the mean distance between two randomly-picked points on $[0,1] \times [0,1]$. Monte Carlo estimates of these values are calculated using the sums:

$$I_n = \frac{1}{n} \sum_{i=1}^{n} |x_{1,i} - x_{2,i}|,$$

where $x_{1,i}, x_{2,i}$ are uniformly distributed on $[0,1]$ and

$$J_n = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(x_{1,i} - x_{2,i})^2 + (y_{1,i} - y_{2,i})^2},$$

where $x_{1,i}, x_{2,i}, y_{1,i}, y_{2,i}$ are uniformly distributed on $[0,1]$. It is useful to notice that multiple integrals are as easy to set up for Monte Carlo calculations as are single integrals. A multiple integral still only involves a single sum. In addition, it is straightforward to show that the same error estimates hold as given in the previous example. In particular, the statistical error is proportional to $1/\sqrt{n}$ for either case. A Fortran program for this example is given at the end of this chapter. The subroutine in the program applies the linear congruential random number generator described in the previous section. Using one million samples, i.e., $n = 10^6$, I_n and J_n were calculated as $I_n = 0.33335$ and $J_n = 0.52163$. For comparison, the exact values of I and J to five significant digits are 0.33333 and 0.52141, respectively.

Example 1.30. Monte Carlo Estimation of Mean Set Length

This example illustrates the flexibility of Monte Carlo techniques. Let $S = \{000, 001, \ldots, 999\}$ be the set of three-digit numbers. In this experiment, sets are randomly selected from S having certain properties. In particular, each set $B \subset S$ is randomly selected so that all the elements of B differ from each other in at least two positions. That is, if $x, y \in B$, then the three-digit numbers x and y differ from each other in at least two positions. Furthermore, set B is constructed as large as possible so that for any $x \in B^c$ there is a $y \in B$ so that x and y are identical in two positions. Let

$\Omega = \{B_1, B_2, B_3, \ldots, B_N\}$ with $N = 10^{16}$ be a large collection of randomly selected sets having the described properties. (The total number of possible sets with the described properties exceeds 10^{35}.) Define the random variable X on Ω so that $X(B)$ equals the number of elements in set B. It can be shown [6] that $X : \Omega \to [50, 100]$. Define the probability measure P by $P(B_i) = \frac{1}{N}$ for $i = 1, 2, \ldots, N$. It is of interest in this example to estimate the average number of elements in the sets $B \in \Omega$, i.e., $E(X) = \frac{1}{N}\sum_{i=1}^{N} X(B_i)$. As N is very large, all the sets of Ω cannot be computed. Therefore, $E(X)$ must be estimated using $E(X) \approx \frac{1}{M}\sum_{i=1}^{M} X(B_i)$ for $M < N$. The sets $\{B_i\}_{i=1}^{M}$ are constructed using a random number generator. Computing M=10,000 sets, $E(X)$ and $\mathrm{Var}(X)$ are estimated as $E(X) \approx 86.13$ and $\mathrm{Var}(X) \approx 3.35$. Based on these calculational results, the average length of sets in Ω is between 86.09 and 86.17 with probability 0.95, i.e., $P(86.09 \le E(X) \le 86.17) = 0.95$.

Example 1.31. Method Error and Statistical Error in Monte Carlo Estimation

This example illustrates the two kinds of error involved in estimating stochastic integrals or in approximating solutions to stochastic differential equations. Let

$$I_N = \sum_{i=1}^{N} \frac{i-1}{N^{3/2}}\,\eta_i, \quad \text{where} \quad \eta_i \sim N(0,1) \quad \text{for each} \;\; i.$$

The expression I_N will be defined in Chapter 3 and is an approximation to the stochastic integral $\int_0^1 t\,dW(t)$ using N subintervals of $[0, 1]$, where $W(t)$ is a Wiener process. Indeed, $E(I_N^2) \to E(I^2) = \dfrac{1}{3}$ as $N \to \infty$. However, $E(I_N^2)$ can be calculated exactly as

$$E(I_N^2) = \sum_{i=1}^{N} \frac{(i-1)^2}{N^3} = \frac{N(N-1)(2N-1)}{6N^3} = \frac{1}{3} - \frac{1}{2N} + \frac{1}{6N^2}.$$

The method error in this approximation is therefore

$$E(I^2) - E(I_N^2) = \frac{1}{2N} - \frac{1}{6N^2}$$

which goes to zero as $N \to \infty$. The method error is due to approximating the stochastic integral using N subintervals in the numerical procedure. However, there is also a statistical error in approximating $E(I^2)$ by $E(I_N^2)$ which is due to using, for example, M samples to estimate $E(I_N^2)$. Suppose, therefore, that $I_{N,m}$ for $m = 1, 2, \ldots, M$ are M samples of I_N using the random numbers $\eta_{i,m}$ for $1 \le i \le N, 1 \le m \le M$. That is,

$$I_{N,m} = \sum_{i=1}^{N} \frac{i-1}{N^{3/2}}\,\eta_{i,m},$$

where $\eta_{i,m}$ are $N(0, 1)$ normally distributed numbers. Then,

$$E(I_N^2) \approx \frac{1}{M}\sum_{m=1}^{M}(I_{N,m})^2 = \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{i=1}^{N}\frac{i-1}{N^{3/2}}\,\eta_{i,m}\right)^2$$

and there is a statistical error proportional to $1/\sqrt{M}$ in the estimate of $E(I_N^2)$. Therefore, when estimating $E(I^2)$ using $\sum_{m=1}^{M}(I_{N,m})^2/M$, there exist two errors. There is a statistical error proportional to $1/\sqrt{M}$, where M is the number of samples. However, even as $M \to \infty$ the approximation may not be satisfactory unless the error due to the method is also small, i.e., the value of N is sufficiently large. Consider the calculational results for this problem which are given in Table 1.1. Recall that $E(I^2) = 1/3$ exactly. Observe that as M increases in any column, the statistical error decreases and the total error approaches the error resulting from the method error. For large N, the error is due primarily to statistical error which can be high for small sample sizes, i.e., small values of M. The first column (N=10) in the table gives values that have large method errors. The first row (M=10) in the table gives values that have large statistical errors. This example illustrates that to obtain accurate results, the method error as well as the statistical error must be kept small.

Table 1.1. Calculated estimates of $E(I^2)$ for Example 1.31

Value of M	$N = 10^1$	$N = 10^2$	$N = 10^3$
10^1	0.25880	0.34465	0.25033
10^2	0.30833	0.28372	0.35280
10^3	0.29179	0.32875	0.32994
10^4	0.28495	0.32399	0.32916
10^5	0.28545	0.32591	0.33201
10^6	0.28430	0.32853	0.33296

Exercises

1.1. Consider the random experiment of rolling one die.
(a) Find the sample space Ω.
(b) Carefully determine the σ-algebra, $\mathcal{A}$, of sets generated by $A_1 = \{1, 2\}$ and $A_2 = \{2, 3\}$.
(c) Define a probability measure P on the sample space Ω.

1.2. Consider rolling one die until an odd number turns up. Find the probability that the first odd number occurs with an even number of rolls.

1.3. Let $C_i^N = \dfrac{N!}{(N-i)!\,i!}$. Prove that $\sum_{i=0}^{N} C_i^N = 2^N$.

1.4. Let $\Omega = \{x : 0 \le x < \infty\}$. Let $\mathcal{A}$ be the σ-algebra generated by intervals of the form $(a, b] \subset [0, \infty)$, that is, $(a, b] \in \mathcal{A}$ for any $0 \le a \le b < \infty$. Define the probability measure by $P(A) = \int_A p(s)\, ds$ where $p(s) = se^{-s}$. Define the random variable, X, by $X(x) = x$. Find $\mu = E(X)$ and $\sigma^2 = E((X - \mu)^2)$. (Note that $\int_0^\infty s^n e^{-s}\, ds = n!$ for $n = 0, 1, 2, \ldots$. Also, for subsequent exercises, note that $\int se^{-s}\, ds = e^{-s}(-s - 1)$ and $\int s^2 e^{-s}\, ds = e^{-s}(-s^2 - 2s - 2)$.)

1.5. Let X_1, X_2, X_3 be independent and identically distributed with the probability measure defined in Exercise 1.4. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2 - X_3$. Show that $E(Y_1 Y_2) \ne E(Y_1)E(Y_2)$ and thus infer that Y_1 and Y_2 are not independent.

1.6. Let X_1 and X_2 be independent and identically distributed with the probability measure defined in Exercise 1.4. Let $X = X_1 + X_2$. Find $F_X(x)$ where $F_X(x) = P(X_1 + X_2 \le x)$ and find $p_X(x) = \dfrac{dF_X(x)}{dx}$. Is X distributed with the same probability measure as X_1 and X_2? (Note that $P(X_1 + X_2 \le x) = \displaystyle\int_0^x \int_0^{x - s_1} s_2 e^{-s_2} s_1 e^{-s_1}\, ds_2\, ds_1$.)

1.7. Let $C[-1, 1]$ be the vector space of real continuous functions on $[-1, 1]$. Defining $(f, g) = \int_{-1}^1 f(x)g(x)\, dx$ and $\|f\| = (f, f)^{1/2}$ for $f, g \in C[-1, 1]$, then $C[-1, 1]$ becomes a normed inner product space. However, this space is not complete. To show this, find and analyze a sequence $\{f_n\}_{n=1}^\infty$ such that the sequence is Cauchy in this normed space, but the sequence does not converge to a function in $C[-1, 1]$.

1.8. Suppose that a sequence of numbers $\{Y_n\}$ is distributed on $[0, \infty)$ with density $p_Y(s) = se^{-s}$. (Refer to Exercise 1.4.) Let $\{U_n\}$ be uniformly distributed numbers on $[0, 1]$.
(a) Find an algebraic expression for U_n in terms of Y_n.
(b) Compute Y_1 given that $U_1 = 0.65$.

1.9. Let $X_n(x) = X(x) + X^2(x)(1 - \frac{1}{n})$, for $n = 0, 1, 2, \ldots$, where $X \sim N(0, 4)$. Prove that $X_n \to X + X^2$ in the mean square sense. That is, prove that $\|X_n - (X + X^2)\|_{RV} \to 0$ as $n \to \infty$.

1.10. Consider the random number generator $X_{n+1} = (8X_n + 9) \bmod (7)$ for $n = 0, 1, 2, \ldots$, with $U_n = X_n/7$. Let $X_0 = 2$ and calculate $X_1, X_2, \ldots, X_{10}$. Determine the period of the generator.

1.11. Assume that the function f is integrable and maps $[0, 1]$ into $[0, 1]$. Consider estimating $\int_0^1 f(x)\, dx$ using two different Monte Carlo approaches. The standard approximation is applied in the first approach, that is, $\int_0^1 f(x)\, dx = E(f) \approx \sum_{i=1}^n f(S_i)/n$ where S_i are uniformly distributed numbers on $[0, 1]$

for $i = 1, 2, \ldots, n$. In this approach, $\sigma_1^2 = E(f^2) - (E(f))^2$. In the second approach, $\int_0^1 f(x)\,dx = p$ where p is the probability that a random point chosen in $[0,1] \times [0,1]$ lies in the region defined by $\{(x,y) \in [0,1] \times [0,1] : y \leq f(x)\}$. In this approach, $\sigma_2^2 = p(1-p)$.
(a) Explain why $\sigma_2^2 = p(1-p)$.
(b) Prove that the first approach always has a variance less than or equal to that of the second approach, i.e., prove that $\sigma_1^2 \leq \sigma_2^2$.

1.12. Modify the MATLAB program given below and approximate $\int_0^1 \sin(\pi x)\,dx$ using the two Monte Carlo procedures described in Exercise 1.11 with $n = 10000$ points.

```
% Two Monte Carlo procedures are used for approximating
% the integral of 0 < f(x) < 1 on [0,1].
% xx is used for the random number generator.
% Function randmonte is needed.
clear
xx=4830887.0;
for j=1:5
 n=10000;
 s1=0.0;
 s2=0.0;
 s3=0.0;
 for i=1:n
    rand=randmonte(xx);
    xx=rand(2);
    x=rand(1);
    f=exp(-.5*x);
    s1=s1+f/n;
    s2=s2+f*f/n;
 end;
 v1=s2-s1*s1;
 disp((sprintf(' %10.4e %10.4e', s1, v1)))
 for i=1:n
    rand=randmonte(xx);
    xx=rand(2);
    x=rand(1);
    rand=randmonte(xx);
    xx=rand(2);
    y=rand(1);
    f=exp(-.5*x);
    hlp=1.0;
    if (y > f)
       hlp=0.0;
       end;
    s3=s3+hlp/n;
```

```
 end;
 v3=s3-s3*s3;
 disp((sprintf(' %10.4e %10.4e', s3, v3)))
 end;
%---------------------------------------------------------------
% random number generator (uniform on [0,1])
% congruential generator xx=16807*xx mod(2^31-1)
% xx=rand(2) is input to the generator
function rand = randmonte(xx)
       a=16807;
       b=2147483647;
       d=fix(a*xx/b);
       xx=a*xx-d*b;
       rand(1)=xx/b;
       rand(2)=xx;
```

Computer Programs

Program 1.1. A Monte Carlo program for computing the average distance between two points

Listed is a Monte Carlo code in Fortran for computing the average distance between two randomly selected points on the interval [0,1] and also for computing the average distance between two randomly selected points on the square $[0,1] \times [0,1]$. In the program, the value given for xx is the initial starting value for the random number generator and nrun is the number of samples. The subroutine uses a linear congruential generator to generate uniformly distributed numbers on $[0,1]$. The output of the program is listed following the code.

```
       real*8 xx
       xx=5320859.
c A Monte Carlo code for distance between random points
       nrun=1000000
       arun=nrun
       s1=0.0
       s2=0.0
       do 400 nr=1,nrun
       call random(xx,x1)
       call random(xx,x2)
       f= abs(x1-x2)
       s1=s1+f/arun
       s2=s2+f*f/arun
 400   continue
```

```
      sd=sqrt(s2-s1*s1)
      write(6,420) nrun,s1,s2,sd
 420  format(2x,i8,3(2x,f9.5))
      s1=0.0
      s2=0.0
      do 450 nr=1,nrun
      call random(xx,x1)
      call random(xx,x2)
      call random(xx,y1)
      call random(xx,y2)
      f= sqrt((x1-x2)*(x1-x2)+(y1-y2)*(y1-y2))
      s1=s1+f/arun
      s2=s2+f*f/arun
 450  continue
      sd=sqrt(s2-s1*s1)
      write(6,420) nrun,s1,s2,sd
      stop
      end
      subroutine random(xx,r)
      real*8 xx,a,b,d,c
c A linear congruential generator
      a=16807.
      ib=2147483647
      b=ib
      c=0.0
      id=(a*xx+c)/b
      d=id
      xx=(a*xx+c)-d*b
      r=xx/b
      return
      end
  1000000    0.33335    0.16659    0.23552
  1000000    0.52163    0.33364    0.24808
```

2

Stochastic Processes

2.1 Introduction

A stochastic process is a family of random variables $\{X(t), t \in \tau\}$ defined on a probability space $(\Omega, \mathcal{A}, P)$ and indexed by a parameter t where t varies over a set τ. If the set τ is discrete, the stochastic process is called discrete. If the set τ is continuous, the stochastic process is called continuous. The parameter t usually plays the role of time and the random variables can be discrete-valued or continuous-valued at each value of t. For example, a continuous stochastic process can be discrete-valued. For modeling purposes, it is useful to understand both continuous and discrete stochastic processes and how they are related.

Stochastic processes occur throughout the remainder of this book. Indeed, solutions of stochastic differential equations are stochastic processes. Properties of stochastic differential equations are studied in Chapter 4 with regard to a Hilbert space of stochastic processes. A Hilbert space setting for random variables and stochastic processes unifies and simplifies the presentation of stochastic integration and stochastic differential equations. Furthermore, the procedure described in Chapter 5 for developing a stochastic differential equation model involves first constructing a discrete stochastic process model. As time is made continuous, the probability distribution of the discrete stochastic model approaches that of the continuous stochastic model.

In this chapter, discrete and continuous stochastic models are discussed. Then, a Hilbert space of stochastic processes is described. The important Wiener process is introduced and computer generation of stochastic processes is considered. Finally, several well-known stochastic processes are described. A basic introduction is given in this chapter to stochastic processes with priority given to results needed in describing stochastic differential equations and modeling. Excellent and thorough treatments of stochastic processes are presented, for example, in [11, 41, 51, 63, 67, 85, 87, 96, 106].

2.2 Discrete Stochastic Processes

In this section, discrete stochastic processes are considered. For these processes, let $\tau = \{t_0, t_1, t_2, \dots\}$ be a set of discrete times. Let the sequence of random variables $X(t_0), X(t_1), X(t_2), \dots$ each be defined on the sample space Ω. This sequence may describe, for example, the evolution of a physical, biological, or financial system over the discrete times $t_0, t_1, t_2, \dots$. If only the present value of the random variable $X(t_n) = X_n$ is needed to determine the future value of X_{n+1}, the sequence $\{X_n\}$ is said to be a Markov process. Throughout this section, the discrete stochastic processes that are discussed are Markov processes. Such processes are common and are useful in developing the stochastic differential equation models described in Chapter 5. A discrete-valued Markov process is called a Markov chain. Let $P(X_{n+1} = x_{n+1}|X_n = x_n)$ define the one-step transition probabilities for a Markov chain. That is, $P(X_{n+1} = x_{n+1} \text{ and } X_n = x_n) = P(X_{n+1} = x_{n+1}|X_n = x_n)P(X_n = x_n)$. If the transition probabilities are independent of time t_n, then the Markov chain is said to have stationary transition probabilities and the Markov chain is referred to as a homogeneous Markov chain. Consider the following two examples.

Example 2.1. A continuous-valued Markov process

Let $t_i = i\Delta t$ and $\eta_i \sim N(0,1)$ for $i = 0, 1, \dots, N$, where $\Delta t = 1/N$. Let $X_i = X(t_i)$ be defined by

$$X_{i+1} = X_i + \eta_i\sqrt{\Delta t} \quad \text{for} \quad i = 0, 1, \dots, N-1,$$

where $X_0 = 0$. Then, $\{X_i\}_{i=0}^N$ is a Markov process with continuous values of X_i and discrete values of time t_i. Note that $X_N = \sum_{i=0}^{N-1} \eta_i\sqrt{1/N}$ so $X_N \sim N(0,1)$.

Example 2.2. A homogeneous Markov chain

Let $X_i = X(t_i)$ with $t_i = i\Delta t$ for $i = 0, 1, \dots, N$ where $\Delta t = 1/N$ and $X_0 = 0$. Define the probability distribution of the discrete random variable δ so that δ takes on the values $-\alpha, 0, \alpha$ with probabilities

$$p(\delta = -\alpha) = \gamma\Delta t, \quad p(\delta = \alpha) = \gamma\Delta t, \quad \text{and} \quad p(\delta = 0) = 1 - 2\gamma\Delta t$$

assuming that $1 - 2\gamma\Delta t > 0$. Let

$$X_{i+1} = X_i + \delta_i \quad \text{for} \quad i = 0, 1, \dots, N-1,$$

where δ_i are independent identically distributed values with the same distribution as δ. Then, $E(\delta_i) = 0$ and $\text{Var}(\delta_i) = 2\alpha^2\gamma\Delta t = 2\alpha^2\gamma/N$. Notice that the stochastic process X_i takes on discrete values and time is also discrete-valued. For this example, the transition probabilities have the form

$$P(X_{n+1} = l\alpha|X_n = k\alpha) = \begin{cases} 1 - 2\gamma\Delta t, & l = k \\ \gamma\Delta t, & l = k-1 \quad \text{or} \quad l = k+1 \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, note that

$$X_N = \sum_{i=0}^{N-1} \hat{\delta}_i/\sqrt{N}, \quad \text{where} \quad \hat{\delta}_i = \delta_i\sqrt{N}, \quad E(\hat{\delta}_i) = 0, \quad \text{and} \quad \text{Var}(\hat{\delta}_i) = 2\alpha^2\gamma.$$

Then, by the Central Limit Theorem,

$$X_N \sim N(0, 2\alpha^2\gamma) \quad \text{for large} \quad N.$$

In particular, if $2\alpha^2\gamma = 1$, then $X_N \sim N(0,1)$ for large N. Thus, as N increases, the distribution of X_N approaches the same distribution as the random variable in the previous example.

Before considering nonhomogeneous Markov chains, it is interesting to briefly discuss some properties of homogeneous Markov chains. Let $\{X_n, n \geq 0\}$ be a homogeneous Markov chain defined at discrete times $\tau = \{t_0, t_1, t_2, \dots\}$, where $t_n = n\Delta t$ so that $t_{n+k} = t_n + t_k$. Let X_n be nonnegative and integer-valued for each $t_n, n = 0, 1, \dots$. That is, $X_n \in \{0, 1, 2, \dots\}$. Let

$$p_{i,j} = P\{X_{n+1} = j | X_n = i\}, \ i \geq 0, \ j \geq 0$$

define the transition probabilities. The transition probability matrix is defined as $P = [p_{i,j}]$ and $\sum_{j=0}^{\infty} p_{i,j} = 1$ for $i = 0, 1, 2, \dots$.

The probability distribution of X_n for $n \geq 1$ can be computed using the transition probability matrix P. Define the kth power of P as $P^k = [p_{i,j}^{(k)}]$. As $P^{l+n} = P^l P^n$, then by matrix multiplication

$$p_{i,j}^{(l+n)} = \sum_{m=0}^{\infty} p_{i,m}^{(l)} p_{m,j}^{(n)} \quad \text{for} \quad l, n \geq 0,$$

where P^0 is defined as $P^0 = I$. This relation is known as the Chapman-Kolmogorov formula for a homogeneous Markov chain. Let $p_i(t_k) = P(X(t_k) = i)$ for $i = 0, 1, 2, \dots$ be the probability distribution of X_k . Let $\mathbf{p}(t_k) = [p_0(t_k), p_1(t_k), \dots, p_r(t_k), \dots]^T$, where $(\mathbf{p}(t_0))_i = P(X(t_0) = i)$ is the initial probability distribution of $X(t_0)$. It is straightforward to see that

$$\begin{aligned}
(\mathbf{p}(t_1))^T &= (\mathbf{p}(t_0))^T P \\
(\mathbf{p}(t_2))^T &= (\mathbf{p}(t_1))^T P = (\mathbf{p}(t_0))^T P^2 \\
&\vdots \\
(\mathbf{p}(t_n))^T &= (\mathbf{p}(t_{n-1}))^T P = (\mathbf{p}(t_0))^T P^n.
\end{aligned}$$

Thus, $p_i(t_n) = \sum_{m=0}^{\infty} p_m(t_{n-1}) p_{m,i} = \sum_{m=0}^{\infty} p_m(t_0) p_{m,i}^{(n)}$.

Example 2.3. Approximation to a Poisson process

Consider the discrete homogeneous stochastic process defined by the transition probabilities

$$p_{i,k} = \begin{cases} 1 - \lambda \Delta t, & \text{for} \quad k = i \\ \lambda \Delta t, & \text{for} \quad k = i + 1 \\ 0, & \text{otherwise.} \end{cases}$$

Assume that $\mathbf{p}(0) = [1, 0, 0, \dots]^T$. In this example, the transition probability matrix P is bidiagonal and the equation $(\mathbf{p}(t + \Delta t))^T = (\mathbf{p}(t))^T P$ has the componentwise form:

$$p_0(t + \Delta t) = p_0(t)(1 - \lambda \Delta t)$$

and

$$p_i(t + \Delta t) = p_i(t)(1 - \lambda \Delta t) + p_{i-1}(t)\lambda \Delta t \quad \text{for} \quad i \geq 1.$$

Rearranging these expressions yields

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = -\lambda p_0(t) \quad \text{and}$$

$$\frac{p_i(t + \Delta t) - p_i(t)}{\Delta t} = -\lambda p_i(t) + \lambda p_{i-1}(t) \quad \text{for} \quad i \geq 1,$$

where $p_0(0) = 1$ and $p_i(0) = 0$ for $i \geq 1$. As $\Delta t \to 0$, the above Markov chain probabilities approach those satisfied by the Poisson process. That is,

$$p_i(t) \approx \frac{\exp(-\lambda t)(\lambda t)^i}{i!} \quad \text{for} \quad \Delta t \quad \text{small.}$$

Now consider Markov chains which are not necessarily homogeneous. Let $\tau = \{t_0, t_1, \dots\}$ where $t_n = n\Delta t$. Let $\{X_n, n \geq 0\}$ be the Markov chain satisfying $X_n \in M = \{z_{-m}, z_{-m+1}, \dots, z_0, z_1, \dots z_m\}$ where m may be arbitrarily large and where $z_i = i\Delta x$ for each i for a positive number Δx. Let

$$p_{i,j}^{(n)} = P_n\{X_{n+1} = z_j | X_n = z_i\}, \ -m \leq i, j \leq m$$

define the transition probabilities which now may depend on time t_n. The transition probability matrix is defined as the $(2m + 1) \times (2m + 1)$ matrix $P_n = [p_{i,j}^{(n)}]$ where $\sum_{j=-m}^{m} p_{i,j}^{(n)} = 1$ for each i and n.

Similar to the homogeneous Markov chain, the probability distribution for X_n for $n \geq 1$ can be computed using the probability transition matrices P_n for $n \geq 0$. Let $p_i(t_n) = P(X_n = z_i)$ for $i = -m, -m + 1, \dots, m$ define the probability distribution at time t_n. Let $\mathbf{p}(t_n) = [p_{-m}(t_n), p_{-m+1}(t_n), \dots, p_m(t_n)]^T$ where $\mathbf{p}(t_0) = [p_{-m}(t_0), p_{-m+1}(t_0), \dots, p_m(t_0)]^T$ is the initial probability distribution. Noticing that $p_i(t_1) = \sum_{l=-m}^{m} p_l(t_0) p_{l,i}^{(0)}$ for $i = -m, -m+1, \dots, m$, it is straightforward to see that

$$
\begin{aligned}
(\mathbf{p}(t_1))^T &= (\mathbf{p}(t_0))^T P_0 \\
(\mathbf{p}(t_2))^T &= (\mathbf{p}(t_1))^T P_1 = (\mathbf{p}(t_0))^T P_0 P_1 \\
&\vdots \\
(\mathbf{p}(t_n))^T &= (\mathbf{p}(t_{n-1}))^T P_{n-1} = (\mathbf{p}(t_0))^T P_0 P_1 \cdots P_{n-1}.
\end{aligned}
$$

If the transition matrix P_n is independent of time t_n, the above expressions reduce to those derived for the homogeneous Markov chain, for example, $(\mathbf{p}(t_n))^T = (\mathbf{p}(t_0))^T P^n$.

Consider the following important nonhomogeneous discrete stochastic process which introduces the forward Kolmogorov equations. This process will be of interest when developing models using stochastic differential equations.

Example 2.4. Forward Kolmogorov equations

Let $t_i = i\Delta t$ for $i = 0, 1, \ldots, N$ and let $x_j = j\delta$, for $j = \ldots, -2, -1, 0, 1, 2, \ldots$. Let X_0 be given. Define the transition probabilities of a discrete stochastic process by the following:

$$
p_{i,k}(t) = \begin{cases} r(t, x_i)\Delta t/\delta^2, & \text{for} \quad k = i+1 \\ 1 - r(t, x_i)\Delta t/\delta^2 - s(t, x_i)\Delta t/\delta^2, & \text{for} \quad k = i, \\ s(t, x_i)\Delta t/\delta^2, & \text{for} \quad k = i-1 \end{cases}
$$

where r and s are smooth nonnegative functions. Notice that with the above transition probabilities, if ΔX is the change in the stochastic process at time t fixing $X(t) = x_i$, then the mean change $E(\Delta X)$ and variance in the change $\mathrm{Var}(\Delta X)$ can be determined. Specifically, $E(\Delta X) = (r(t, X) - s(t, X))\Delta t/\delta$ and $\mathrm{Var}(\Delta X) = (r(t, X) + s(t, X))\Delta t$. It is assumed that $\Delta t/\delta^2$ is small so that $1 - r(t, x_k)\Delta t/\delta^2 - s(t, x_k)\Delta t/\delta^2$ is positive. Let $p_k(t) = P(X(t) = x_k)$ be the probability distribution at time t. Then, $p_k(t + \Delta t)$ satisfies

$$
\begin{aligned}
p_k(t + \Delta t) \; &= p_k(t) + \big[p_{k+1}(t)s(t, x_{k+1}) \\
&\quad - p_k(t)(r(t, x_k) + s(t, x_k)) + p_{k-1}(t)r(t, x_{k-1})\big]\Delta t/\delta^2.
\end{aligned} \tag{2.1}
$$

Rearranging this expression yields:

$$
\begin{aligned}
\frac{p_k(t + \Delta t) - p_k(t)}{\Delta t} &= -\left(\frac{p_{k+1}(t)a(t, x_{k+1}) - p_{k-1}(t)a(t, x_{k-1})}{2\delta^2}\right) \\
&+ \left(\frac{p_{k+1}(t)b(t, x_{k+1}) - 2p_k(t)b(t, x_k) + p_{k-1}(t)b(t, x_{k-1})}{2\delta^2}\right),
\end{aligned}
$$

where, for notational simplicity, $a(t, x) = (r(t, x) - s(t, x))/\delta$ and $b(t, x) = r(t, x) + s(t, x)$. As $\Delta t \to 0$, the discrete stochastic process approaches a continuous-time process. Then as $\Delta t \to 0$, $p_k(t)$ satisfies the initial-value problem:

$$
\begin{aligned}
\frac{dp_k(t)}{dt} &= -\left(\frac{p_{k+1}(t)a(t, x_{k+1}) - p_{k-1}(t)a(t, x_{k-1})}{2\delta}\right) \\
&+ \left(\frac{p_{k+1}(t)b(t, x_{k+1}) - 2p_k(t)b(t, x_k) + p_{k-1}(t)b(t, x_{k-1})}{2\delta^2}\right)
\end{aligned} \tag{2.2}
$$

for $k = \ldots, -2, -1, 0, 1, 2, \ldots$ where $\{p_k(0)\}_{k=-m}^{m}$ are known. Equations (2.2) with the given initial conditions are the forward Kolmogorov equations for the continuous-time stochastic process.

Now assume that δ is small so that the stochastic process approaches a continuous-valued process. As

$$\frac{F(x+\delta) - F(x-\delta)}{2\delta} = F'(x) + \frac{\delta^2}{6}F'''(\xi_1) \quad \text{and}$$

$$\frac{F(x+\delta) - 2F(x) + F(x-\delta)}{\delta^2} = F''(x) + \frac{\delta^2}{12}F''''(\xi_2)$$

for some values ξ_1, ξ_2 such that $x - \delta \leq \xi_1, \xi_2 \leq x + \delta$, then the above system of differential equations approximates the partial differential equation:

$$\frac{\partial p(t,x)}{\partial t} = -\frac{\partial (a(t,x)p(t,x))}{\partial x} + \frac{1}{2}\frac{\partial^2 (b(t,x)p(t,x))}{\partial x^2}. \tag{2.3}$$

Equation (2.2) is a central-difference approximation to (2.3). This approximation is accurate for small δ and when comparing the solutions of (2.1) and (2.3), it can be shown [91] that

$$p(t, x_k) = p_k(t) + O(\Delta t) + O(\delta^2).$$

In Chapter 4, it is shown that (2.3) is the forward Kolmogorov equation corresponding to a diffusion process having the stochastic differential equation

$$dX(t) = a(t, X)dt + \sqrt{b(t, X)}dW(t). \tag{2.4}$$

The probability density of solutions to the stochastic differential equation (2.4) satisfies the partial differential equation (2.3). Therefore, there exists a close relationship between the discrete stochastic process defined by (2.1) and the continuous process defined by (2.4). In particular, for small Δt and δ, the probability distribution of the solutions to (2.4) will be approximately the same as the probability distribution of solutions to the discrete stochastic process. It is shown in Chapter 5 that this result is useful for modeling purposes. It may be straightforward to construct a realistic discrete stochastic process model for a dynamical system under investigation. Then, an appropriate stochastic differential equation model is inferred from the above argument.

Finally, it is useful to note in the above argument that the coefficients at time t of the stochastic differential equation (2.4) are related to the discrete stochastic model (2.1) through the mean and variance in the change in the process ΔX over a short time interval Δt fixing $X(t) = x$. Specifically,

$$E(\Delta X) = a(t, X)\Delta t \quad \text{and} \quad \text{Var}(\Delta X) = b(t, X)\Delta t.$$

Example 2.5. Specific example of forward Kolmogorov equations

As a specific instance of Example 2.4, consider a birth-death process, where $\delta = 1$, $x_j = j$ for $j = 0, 1, 2, \ldots$, b is the per capita birth rate, and d is the

per capita death rate. It is assumed that b and d are constants. The transition probabilities for this example have the form

$$p_{i,k}(t) = \begin{cases} bx_i\Delta t, & \text{for} \quad k = i+1 \\ 1 - bx_i\Delta t - dx_i\Delta t, & \text{for} \quad k = i \\ dx_i\Delta t, & \text{for} \quad k = i-1. \end{cases}$$

It follows that the probability distribution in continuous time (letting $\Delta t \to 0$) satisfies the forward Kolmogorov equations

$$\begin{aligned} \frac{dp_k(t)}{dt} &= -(b-d)\left(\frac{p_{k+1}(t)x_{k+1} - p_{k-1}(t)x_{k-1}}{2}\right) \\ &\quad + \frac{b+d}{2}\big(p_{k+1}(t)x_{k+1} - 2p_k(t)x_k + p_{k-1}(t)x_{k-1}\big) \end{aligned} \tag{2.5}$$

with $p_M(0) = P(X(0) = x_M) = 1$ and $p_k(0) = P(X(0) = x_k) = 0$ for $k \neq M$ assuming an initial population of size M. Note that, fixing $X(t) = x_i$ at time t, $E(\Delta X) = (b-d)X\Delta t$ and $\mathrm{Var}(\Delta X) = (b+d)X\Delta t$ to order $(\Delta t)^2$. For large M, the above equations approximately satisfy the Fokker-Planck equation

$$\frac{\partial p(t,x)}{\partial t} = -\frac{\partial\big((b-d)xp(t,x)\big)}{\partial x} + \frac{1}{2}\frac{\partial^2\big((b+d)xp(t,x)\big)}{\partial x^2} \tag{2.6}$$

with $p(0,x) = \delta(x-M)$.

As will be seen in Chapter 4, the probability distribution $p(t,x)$ is the probability distribution of solutions to the Itô stochastic differential equation

$$dX(t) = (b-d)X(t)dt + \sqrt{(b+d)X(t)}\, dW(t) \tag{2.7}$$

with $X(0) = M$. Thus, the above argument implies that solutions to the stochastic differential equation (2.7) have approximately the same probability distribution as the discrete birth-death stochastic process and a reasonable model for the simple birth-death process is the stochastic differential equation (2.7).

2.3 Continuous Stochastic Processes

Now consider a continuous stochastic process $\{X(t), t \in \tau\}$ defined on the probability space $(\Omega, \mathcal{A}, P)$ where $\tau = [0,T]$ is an interval in time and the process is defined at all instants of time in the interval. A continuous-time stochastic process is a function $X : \tau \times \Omega \to \mathbb{R}$ of two variables t and ω and X may be discrete-valued or continuous-valued. In particular, $X(t) = X(t,\cdot)$ is a random variable for each value of $t \in \tau$ and $X(\cdot,\omega)$ maps the interval τ into $\mathbb{R}$ and is called a sample path, a realization, or a trajectory of the stochastic process for each $\omega \in \Omega$. It is interesting that specific knowledge of ω is generally unnecessary but ω is significant as each $\omega \in \Omega$ results in a

different trajectory. As a result, the normal convention is that the variable ω is often suppressed, that is, $X(t)$ represents a random variable for each value of t and $X(\cdot)$ represents a trajectory over the interval $\tau = [0, T]$. In this book, ω is often included for emphasis or to provide clarity that X is a function of two variables.

The stochastic process X is a Markov process if the state of the process at any time $t_n \in \tau$ determines the future state of the process. Specifically, $P\big(X(t_{n+1}) \le x_{n+1} | X(t_n) = x_n\big) = P\big(X(t_{n+1}) \le x_{n+1} | X(t_1) = x_1, \ldots, X(t_n) = x_n\big)$ whenever $t_1 < t_2 < \cdots < t_n < t_{n+1}$. Markov processes are common stochastic processes and most of the stochastic processes discussed in this book are Markov stochastic processes. The following two examples illustrate continuous Markov processes.

Example 2.6. Poisson process with intensity λ

Let $X(t)$ equal the number of observations in time t. Assume that the probability of one observation in time interval Δt is equal to $\lambda \Delta t + o(\Delta t)$. Referring to Example 1.4, it is clear that this is a continuous stochastic process and the probability of n observations in time t is

$$P(X(t) = n) = P_n(t) = \exp(-\lambda t)(\lambda t)^n / n!.$$

The process $X(t)$ is a continuous-time stochastic process which is discrete-valued. Specifically, $X(t)$ is a Poisson process with intensity $\lambda > 0$. Note that $X(0) = 0$ and the number of observations at any time t is Poisson-distributed with mean λt. That is, for any $s \ge 0$,

$$P\big(X(t+s) - X(s) = n\big) = \exp(-\lambda t)(\lambda t)^n / n!.$$

Indeed, the process is a Markov process and

$$P\big(X(t + \Delta t) \le m + \Delta m | X(t) = m\big) = \sum_{l=0}^{\Delta m} \exp(-\lambda \Delta t)(\lambda \Delta t)^l / l!$$

and the probability distribution at time $t + \Delta t$ only depends on the state of the system at time t and not on the history of the system. Also, as seen in Example 1.4,

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t),$$

$$\frac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t) \quad \text{for} \quad n \ge 1,$$

$E(X(t)) = \lambda t$, and $\mathrm{Var}(X(t)) = \lambda t$. It is readily seen that the relations satisfied by the probabilities of the discrete stochastic process for Example 2.3 are finite-difference approximations to the above differential equations and approach these differential equations as $\Delta t \to 0$. In addition, if $Y(t) = X(t+s) - X(s)$ for any $s \ge 0$, then $Y(t)$ is also Poisson-distributed

with intensity λ and $Y(0) = 0$. Figure 2.1 illustrates the random behavior of the discrete jumps in a Poisson process. The average curve given in the figure is for 200 sample paths. Notice that the average closely follows the line λt. The computer program that calculates and plots these two curves is listed at the end of this chapter.

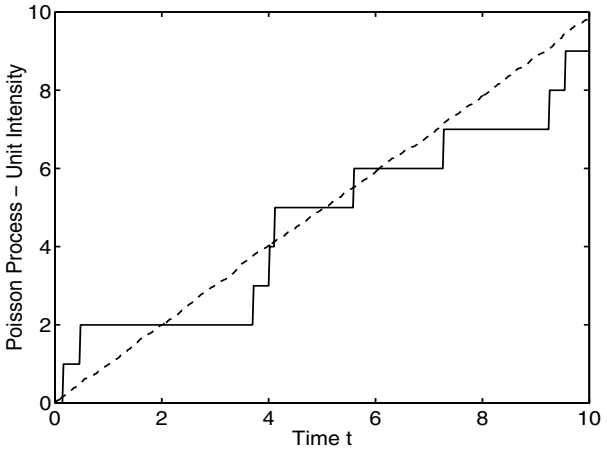

Fig. 2.1. One sample path of the Poisson process (solid curve) and the average of 200 Poisson processes (dashed curve) for $\lambda = 1$

It is interesting to consider the transition probability density function for transition from x at time s to y at time t for a continuous Markov process. Analogous to discrete Markov processes, the transition probability density function satisfies the Chapman-Kolmogorov equation:

$$p(y, t, x, s) = \int_{\mathbb{R}} p(y, t, z, u)p(z, u, x, s)\, dz.$$

A Markov process $X(t)$ is said to be homogeneous if its transition probability satisfies $p(y, t + u, x, s + u) = p(y, t, x, s)$. That is, the transition probability only depends on lapsed time. In this case, it can be written as $p(y, x, t - s)$.

Example 2.7. An approximate Wiener process

Let $X_i(t)$ for $i = 1, 2, \ldots, N$ be N independent Poisson processes with intensity λ as described in Example 2.6. Let $Y_N(t)$ be another stochastic process defined by

$$Y_N(t) = \sum_{i=1}^{N} \frac{X_i(t) - \lambda t}{\sqrt{\lambda N}}.$$

By the Central Limit Theorem, as N increases, $Y_N(t)$ approaches a random variable distributed normally with mean 0 and variance t. Indeed, by

considering Example 2.6, $Y_N(t+s)-Y_N(t)$ approaches a normally distributed variable with mean 0 and variance s for every $s, t \in \tau$.

In the previous example, $Y_N(t)$ approaches a Wiener process or Brownian motion $W(t)$ as N increases. A Wiener process $\{W(t), t \geq 0\}$ is a continuous stochastic process with stationary independent increments such that

$$W(0) = 0 \quad \text{and} \quad W(t) - W(s) \sim N(0, t-s) \quad \text{for all} \quad 0 \leq s \leq t.$$

So $E(W(t)) = 0, \text{Var}(W(t) - W(s)) = t - s \quad$ for $0 \leq s \leq t$ and, in particular,

$$W(t_2) - W(t_1) \sim N(0, t_2 - t_1) \quad \text{and} \quad W(t_4) - W(t_3) \sim N(0, t_4 - t_3)$$

are independent Gaussian random variables for $0 \leq t_1 < t_2 \leq t_3 < t_4$. Notice that a Wiener process is a homogeneous Markov process. Furthermore, by convention, $W(t) = W(t, \omega)$ where ω is generally suppressed. Therefore, $W(t)$ represents a random variable at each value of t.

It is of practical interest as well as of conceptual value to see how easily a sample path of a Wiener process $W(t)$ can be generated at a finite number of points. Suppose that a Wiener process trajectory is desired on the interval $[t_0, t_N]$ at the points $\{t_i\}_{i=0}^N$ where $t_0 = 0$. Then, $W(t_0) = 0$ and a recurrence relation that gives the values of a Wiener process trajectory at the points $t_0, t_1, \ldots, t_N$ is given by

$$W(t_i) = W(t_{i-1}) + \eta_{i-1}\sqrt{t_i - t_{i-1}}, \quad \text{for} \quad i = 1, 2, \ldots, N, \tag{2.8}$$

where η_{i-1} are $N(0,1)$ independent normally distributed numbers for $i = 1, 2, \ldots, N$. The values $W(t_i)$, $i = 0, 1, \ldots, N$ determine a Wiener sample path at the points $\{t_i\}_{i=0}^N$. As discussed in later examples, using these $N+1$ values, the Wiener process sample path can now be approximated everywhere on the interval $[t_0, t_N]$.

Another interesting way to generate a Wiener process, which uses a countable number of normally distributed random variables is the Karhunen-Loève expansion [69, 70]. The Karhunen-Loève expansion is derived from a Fourier series expansion of the Wiener process and has the form

$$W(t) = \sum_{n=0}^{\infty} \frac{2\sqrt{2T}}{(2n+1)\pi} \eta_n \sin\left(\frac{(2n+1)\pi t}{2T}\right) \tag{2.9}$$

for $t \in [0, T]$, where η_n are independent identically distributed standard normal random variables, i.e., $\eta_n \sim N(0,1)$ for $n = 0, 1, \ldots$. Indeed, η_n in (2.9) is given explicitly by

$$\eta_n = \frac{(2n+1)\pi}{2^{1/2}T^{3/2}} \int_0^T W(t) \sin\left(\frac{(2n+1)\pi t}{2T}\right) dt \quad \text{for} \quad n = 0, 1, 2, \ldots.$$

This integral will be considered in Example 3.4 in the next chapter. To see that the series (2.9) has the required properties of the Wiener process, let

$$S_N(t) = \sum_{n=0}^{N} \frac{2\sqrt{2T}}{(2n+1)\pi} \eta_n \sin\left(\frac{(2n+1)\pi t}{2T}\right)$$

be the Nth partial sum of this series. It is easy to show that $S_N(t) \in H_{RV}$ for each $t \in [0,T]$ and that $\{S_N(t)\}$ is Cauchy in Hilbert space H_{RV}. (Recall the definition of Hilbert space H_{RV} in Section 1.6.) Therefore, $S_N(t) \to S(t)$ as $N \to \infty$ in H_{RV} for each $t \in [0,T]$. Indeed, as $\eta_n \sim N(0,1)$ for each n, then $S_N(t) \sim N(0, \sigma_N^2(t))$ where

$$\sigma_N^2(t) = t - \sum_{n=N+1}^{\infty} \frac{8T}{(2n+1)^2\pi^2} \sin^2\left(\frac{(2n+1)\pi t}{2T}\right)$$

noting that

$$\begin{aligned} t &= \sum_{n=0}^{\infty} \frac{8T}{(2n+1)^2\pi^2} \sin^2\left(\frac{(2n+1)\pi t}{2T}\right) \\ &= \sum_{n=0}^{\infty} \frac{4T}{(2n+1)^2\pi^2} \left(1 - \cos\left(\frac{(2n+1)2\pi t}{2T}\right)\right) \end{aligned}$$

for $0 < t < T$. In addition, it can be shown using the trigonometric identity

$$\sin(at)\sin(as) - \sin^2(at) = \frac{1}{2}\cos(at - as) - \frac{1}{2}\cos(at + as) - \frac{1}{2} + \frac{1}{2}\cos(2at)$$

that

$$E\big((S(s) - S(t))S(t)\big) = 0 \quad \text{for} \quad s \geq t.$$

Notice that at each $t \in [0,T]$, $W(t) \in H_{RV}$. In addition, W is continuous in the mean square sense but does not possess a derivative. To see continuity of W, consider

$$\|W(t + \Delta t) - W(t)\|_{RV}^2 = E\big(W(t + \Delta t) - W(t)\big)^2 = \Delta t.$$

Thus, $\|W(t + \Delta t) - W(t)\|_{RV} = \sqrt{\Delta t}$ so given $\epsilon > 0$ there is a $\delta > 0$ such that $\|W(t + \Delta t) - W(t)\|_{RV} < \epsilon$ when $\Delta t < \delta$. However, as

$$\left\| \frac{W(t + \Delta t) - W(t)}{\Delta t} \right\|_{RV}^2 = \frac{1}{\Delta t}$$

there is no $F(t) \in H_{RV}$ such that

$$\left\| \frac{W(t + \Delta t) - W(t)}{\Delta t} - F(t) \right\|_{RV}^2 \to 0 \quad \text{as} \quad \Delta t \to 0.$$

It is useful to discuss expectations of functions of $W(t)$ for $0 \leq t \leq T$. First, recall that

$$p(t,x,y) = \frac{1}{(2\pi|t|)^{1/2}} \exp\left(\frac{-(x-y)^2}{2|t|}\right) \quad \text{for} \quad x,y \in \mathbb{R}$$

is the probability density of normally distributed random variables with mean y and variance $|t|$. Let $W(t)$ be a Wiener process on $[0,T]$. Clearly, for $t_1 \in [0,T]$ and $G : \mathbb{R} \to \mathbb{R}$,

$$E\big(G(W(t_1))\big) = \int_{-\infty}^{\infty} G(x_1) p(t_1, x_1, 0)\, dx_1.$$

In addition,

$$P\big(W(t_1) \le z_1\big) = \int_{-\infty}^{z_1} p(t_1, x_1, 0)\, dx_1.$$

Now consider a partition of $[0,T]$, $0 = t_0 \le t_1 \le t_2 \le \cdots \le t_k \le T$. For $G : \mathbb{R}^2 \to \mathbb{R}$,

$$E\big(G(W(t_1), W(t_2))\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x_1,x_2) p(t_1,x_1,0) p(t_2 - t_1, x_2, x_1)\, dx_1\, dx_2.$$

Furthermore, for $G : \mathbb{R}^k \to \mathbb{R}$,

$$E(G(W(t_1), W(t_2), \ldots, W(t_k))) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} G(x_1, \ldots, x_k) p(t_1, x_1, 0) \ldots p(t_k - t_{k-1}, x_k, x_{k-1})\, dx_1 \ldots dx_k.$$

The densities $p(t_m - t_{m-1}, x_m, x_{m-1})$ for $m = 1, 2, 3, \ldots, k$ define a set of finite-dimensional probability measures on $\mathbb{R}^k$. The probability distribution on this partition satisfies

$$\begin{aligned} F_{t_1 t_2 \ldots t_k}(z_1, z_2, \ldots, z_k) &= P(W(t_1) \le z_1, \ldots, W(t_k) \le z_k) \\ &= \int_{-\infty}^{z_k} \int_{-\infty}^{z_{k-1}} \cdots \int_{-\infty}^{z_1} p(t_1, x_1, 0) \ldots p(t_k - t_{k-1}, x_k, x_{k-1})\, dx_1 \ldots dx_k. \end{aligned} \tag{2.10}$$

It is interesting that this probability measure can be extended through finer and finer partitions to all $[0,T]$ where the measure is identical to the finite-dimensional measure for any partition $0 = t_0 \le t_1 \le t_2 \le \cdots \le t_k \le T$ of $[0,T]$. As these finite-dimensional probability measures satisfy certain symmetry and compatibility conditions [41, 92], Kolmogorov's extension theorem can be applied which says that there exists a probability space and a stochastic process such that the finite-dimensional probability distributions are identical to those defined above. The stochastic process is the Wiener process or Brownian motion $W(t)$ and over any partition of $[0,T]$, the finite-dimensional distributions of $W(t)$ reduce to expression (2.10).

Finally, consider the transition probability density $p(y,t,x,s)$ for the Wiener process from x at time s to y at time t. In this case,

$$p(y,t,x,s) = \frac{1}{(2\pi|t-s|)^{1/2}} \exp\left(\frac{-(x-y)^2}{2|t-s|}\right)$$

and clearly $p(y,t,x,s) = p(y,x,|t-s|)$ so the Wiener process is a continuous homogeneous Markov process. In addition, one can directly verify the Chapman-Kolmogorov equation for this transition probability, that is, for $s < u < t$,

$$p(y,t,x,s) = \int_{-\infty}^{\infty} p(z,u,x,s)p(y,t,z,u)\,dz.$$

2.4 A Hilbert Space of Stochastic Processes

It is useful to define a Hilbert space of stochastic processes. This space is needed in Chapters 3 and 4 for discussing stochastic integrals and stochastic differential equations. For example, demonstrating convergence of a sequence of stochastic processes can be facilitated if the sequence is in a Hilbert space. In particular, Cauchy sequences in the Hilbert space will converge in the space. In this section, a metric space consisting of elementary stochastic processes will be described first. This space will then be completed to a Hilbert space and the set of elementary stochastic processes will be dense in the Hilbert space.

Consider continuous stochastic processes defined on the interval $[0,T]$ and probability space $(\Omega, \mathcal{A}, P)$. Let $f(t) = f(t,\omega)$ be an elementary stochastic process which is a random step function defined on $[0,T] \times \Omega$. That is, f has the form

$$f(t,\omega) = \sum_{i=0}^{N-1} f(t_i,\omega)I_i(t),$$

where $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$ is a partition of $[0,T]$ and $I_i(t)$ is the characteristic function

$$I_i(t) = \begin{cases} 1 & \text{for} \quad t_i \le t < t_{i+1} \\ 0, & \text{otherwise} \end{cases} \qquad \text{for} \quad i = 0,1,2,\ldots,N-1.$$

Recall from Section 1.6 that H_{RV} is a Hilbert space of random variables. It is assumed that the random variable $f(t_i,\cdot) \in H_{RV}$ for each t_i, in particular, $E(f^2(t_i)) < \infty$ for each i. Now, the metric space S_{SP} is defined as

$$S_{SP} = \{\text{random step functions} \quad f(t,\omega) \quad \text{defined on} \quad [0,T] \times \Omega \quad \text{such that} \quad \int_0^T E(f(t))^2\,dt = \textstyle\sum_{i=0}^{N-1} E(f^2(t_i))(t_{i+1} - t_i) < \infty\}.$$

On S_{SP}, the inner product $(\cdot,\cdot)_{SP}$ is defined as

$$(f,g)_{SP} = \int_0^T E(f(t)g(t))\,dt$$

and the norm is defined as

$$\|f\|_{SP} = (f,f)_{SP}^{1/2} = \left(\int_0^T E|f(t)|^2 \, dt \right)^{1/2}.$$

The space S_{SP} is a metric space with the metric $\|\cdot\|_{SP}$. However, S_{SP} is not complete and it is straightforward to show that not all Cauchy sequences converge in S_{SP}. This space can be completed by adding to it additional stochastic processes. The completed space is defined as H_{SP} and S_{SP} is dense in H_{SP}. That is, given $f \in H_{SP}$ and given $\epsilon > 0$ there is a $g \in S_{SP}$ such that $\|f - g\|_{SP} < \epsilon$. In the act of completing S_{SP}, many stochastic processes may be added to S_{SP} to form H_{SP}.

Suppose, for example, that a stochastic process $f(t,\omega)$ satisfies, for some positive constants k_1 and k_2, the inequalities $\|f(0)\|_{RV}^2 \le k_1$ and $\|f(t_2) - f(t_1)\|_{RV}^2 \le k_2|t_2 - t_1|$ for all $t_1, t_2 \in [0,T]$. Then, $f \in H_{SP}$ and $f_N(t,\omega) = \sum_{i=0}^{N-1} f(t_i,\omega) I_i(t)$ forms a Cauchy sequence in $S_{SP} \subset H_{SP}$ that converges to f. Indeed,

$$\|f\|_{SP}^2 \le 2\int_0^T E|f(t) - f(0)|^2 \, dt + 2\int_0^T E|f(0)|^2 \, dt \le k_2 T^2 + 2k_1 T.$$

In addition, Fubini's theorem [41, 65, 73] states that

$$\int_0^T E|f(t)| \, dt = E\int_0^T |f(t)| \, dt \quad \text{and} \quad \int_0^T E|f(t)|^2 \, dt = E\int_0^T |f(t)|^2 \, dt.$$

For $f \in H_{SP}$, the Cauchy-Schwarz inequality, $|(f,g)_{SP}| \le \|f\|_{SP}\|g\|_{SP}$, is very useful and written explicitly has the form

$$\left| \int_0^T E(f(t)g(t)) \, dt \right| \le \left(\int_0^T E|f(t)|^2 \, dt \right)^{1/2} \left(\int_0^T E|g(t)|^2 \, dt \right)^{1/2}.$$

Thus, for example, applying the Cauchy-Schwarz inequality and Fubini's theorem,

$$E\int_0^T |f(t)| \, dt = \int_0^T E|f(t)| \, dt \le T^{1/2} \left(\int_0^T E|f(t)|^2 \, dt \right)^{1/2}.$$

Furthermore, the triangle inequality, $\|f+g\|_{SP} \le \|f\|_{SP} + \|g\|_{SP}$, for $f, g \in H_{SP}$ is explicitly

$$\left(\int_0^T E|f(t)+g(t)|^2 \, dt \right)^{1/2} \le \left(\int_0^T E|f(t)|^2 \, dt \right)^{1/2} + \left(\int_0^T E|g(t)|^2 \, dt \right)^{1/2}.$$

Sometimes it is useful to apply a set of stochastic processes which are even more elementary than those in S_{SP}. Let $\hat{S}_{SP} \subset S_{SP}$ be the set of simple random step functions. That is, $f \in \hat{S}_{SP}$ has the form

$$f(t,\omega) = \sum_{i=0}^{N-1} \sum_{j=1}^{M} f_{ij}(\omega) I_{A_j} I_i(t), \quad \text{where} \quad A_j \in \mathcal{A}.$$

As simple functions are dense in H_{RV}, $\hat{S}_{SP}$ is also dense in H_{SP}.

Several examples help to clarify Hilbert space H_{SP}. First, for emphasis:

H_{SP} is a Hilbert space of Stochastic Processes with norm $\|f\|_{SP} = \left(\int_0^T E|f(t)|^2\, dt\right)^{1/2}$.

Example 2.8. A converging sequence of stochastic processes

Define the stochastic process $f_N(t)$ as

$$f_N(t) = \sum_{i=0}^{N-1} I_i(t) W(t_i), \quad \text{where} \quad h = \frac{T}{N} \quad \text{and} \quad t_i = ih \quad \text{for} \quad i = 0, 1, \ldots, N.$$

Clearly, $f_N \in H_{SP}$ for each N. Also,

$$\begin{aligned} \|f_N - W\|_{SP}^2 &= \int_0^T E\left(\sum_{i=0}^{N-1} (W(t) - W(t_i)) I_i(t)\right)^2 dt \\ &= \int_0^T \sum_{i=0}^{N-1} (t - t_i) I_i(t)\, dt = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (t - t_i)\, dt \\ &= \sum_{i=0}^{N-1} \frac{h^2}{2} = \frac{T^2}{2N}. \end{aligned}$$

Since $\|f_N - W\|_{SP} \to 0$ as $N \to \infty$, the sequence of stochastic processes $\{f_N\}_{N=1}^{\infty}$ converges to W in H_{SP}.

Example 2.9. Another converging sequence of stochastic processes

Let $W \in H_{SP}$ be a Wiener process on $[0, T]$. Define the stochastic process $X_n(t)$ in the following way:

$$X_n(t) = \frac{n-1}{n} W(t) \quad \text{for} \quad n = 1, 2, \ldots.$$

Then

$$\|X_n - W\|_{SP}^2 = \int_0^T E(X_n(t) - W(t))^2\, dt = \int_0^T E(W(t))^2 \frac{1}{n^2}\, dt = \frac{T^2}{2n^2}.$$

Thus, X_n converges to W in H_{SP} as $n \to \infty$.

Example 2.10. Integration of a function of a Poisson process

Consider $J(e^{-X}) = \int_0^T \exp(-X(t))\,dt$ where X is a Poisson process with intensity λ on the interval $[0,T]$. Suppose that $X(t)$ experiences unit increases at the times $t_1 < t_2 < t_3 < \cdots < t_{N-1}$ on $[0,T]$ and let $t_0 = 0$ and $t_N = T$. Then $X(t)$ can be written in the form

$$X(t) = i \quad \text{for} \quad t_i \le t < t_{i+1} \quad \text{for} \quad i = 0, 1, \ldots, N-1.$$

Then, it is easily seen that

$$J(e^{-X}) = \int_0^T \exp(-X(t))\,dt = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \exp(-i)\,dt = \sum_{i=0}^{N-1} \exp(-i)(t_{i+1} - t_i).$$

Furthermore, it is interesting that

$$E\left(J(e^{-X})\right) = E\int_0^T \exp(-X(t))\,dt = \int_0^T \sum_{i=0}^{\infty} \exp(-i)\frac{(\lambda t)^i \exp(-\lambda t)}{i!}\,dt$$

$$= \sum_{i=0}^{\infty} \frac{\exp(-i)}{i!} \int_0^T (\lambda t)^i \exp(-\lambda t)\,dt = \sum_{i=0}^{\infty} \frac{\exp(-i)}{i!\lambda} \int_0^{\lambda T} (x)^i \exp(-x)\,dx$$

$$= \sum_{i=0}^{\infty} \frac{\exp(-i) B_i(\lambda T)}{i!\lambda} \quad \text{where} \quad B_i(\lambda T) = \int_0^{\lambda T} x^i \exp(-x)\,dx.$$

Also, as $T \to \infty$, then $B_i(\lambda T) \to i!$ and $E\big(\int_0^{\infty} \exp(-X(t))\,dt\big) = \dfrac{e}{\lambda(e-1)}$.

Example 2.11. A commonly used approximation to the Wiener process

Consider the interval $0 \le t \le T$ and let $t_i = ih$ for $i = 0, 1, 2, \ldots, N$ where $h = T/N$. Let $W(t) \sim N(0,t)$ be a Wiener process. Define the continuous piecewise linear stochastic process $X_N(t)$ on this partition of $[0,T]$ by

$$X_N(t) = W(t_i)\frac{t_{i+1} - t}{h} + W(t_{i+1})\frac{t - t_i}{h}$$

for $t_i \le t \le t_{i+1}$ and $i = 0, 1, \ldots, N-1$. Notice that $X_N(t_i) = W(t_i)$ for $i = 0, 1, \ldots, N$ and $X_N(t)$ is continuous on $[0,T]$. Also,

$$\|X_N - W\|_{SP}^2 = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\left(W(t_i)\frac{t_{i+1} - t}{h} + W(t_{i+1})\frac{t - t_i}{h} - W(t)\right)^2 dt$$

$$= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\left(\big(W(t_i) - W(t)\big)\frac{t_{i+1} - t}{h} + \big(W(t_{i+1}) - W(t)\big)\frac{t - t_i}{h}\right)^2 dt$$

$$= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \frac{2(t - t_i)(t_{i+1} - t)}{h}\,dt$$

$$= \sum_{i=0}^{N-1} \frac{h^2}{3} = \frac{T^2}{3N}.$$

Thus, $\|X_N - W\|_{SP}^2 \to 0$ as $N \to \infty$, i.e., $X_N \to W$ in H_{SP} as $N \to \infty$. For a large value of N, the graph of a sample path of $X_N(t)$ is indistinguishable from the graph of the corresponding sample path of $W(t)$. The graph of a Wiener process trajectory is often represented by plotting $X_N(t)$ for a large value of N. In Fig. 2.2 two Wiener process sample paths are plotted on the interval $[0, 10]$. (In this figure, two sample paths $X_{1000}(t)$ are actually graphed where the recurrence relation (2.8) is used to generate the 1000 values $W(t_i)$ for $i = 1, 2, \ldots, 1000$ for each Wiener process trajectory.)

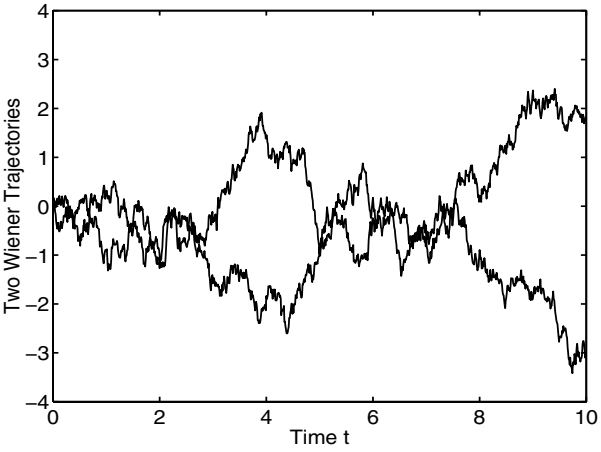

Fig. 2.2. Two Wiener sample paths on [0,10]

A few additional properties of stochastic processes in Hilbert space H_{SP} are useful. First, if $X \in H_{SP}$, then $\int_0^T E|X(t)|^2\, dt < \infty$ which implies that $X(t) \in H_{RV}$ for almost every $t \in [0, T]$. In addition, if $\{X_n\}_{n=1}^{\infty} \subset H_{SP}$ converges to a stochastic process $X \in H_{SP}$, then $\|X_n - X\|_{SP} \to 0$ as $n \to \infty$. Hence, $E|X_n(t) - X(t)|^2 \to 0$ for almost every $t \in [0, T]$. That is, convergence in Hilbert space H_{SP} implies convergence in probability on $[0, T]$. Specifically, $\|X_n - X\|_{SP} \to 0$ as $n \to \infty$ implies that for any $\epsilon > 0$, $P(\{\omega : |X_n(t, \omega) - X(t, \omega)| > \epsilon\}) \to 0 \quad \text{as} \quad n \to \infty$ for almost every $t \in [0, T]$.

Now, suppose that $X(0) \in H_{RV}$ and $X \in H_{SP}$ satisfies

$$\|X(t_2) - X(t_1)\|_{RV}^2 = E|X(t_2) - X(t_1)|^2 \leq K|t_2 - t_1|$$

for any $t_1, t_2 \in [0, T]$ for a constant $K > 0$. Then a bound on $\|X\|_{SP}$ can be found and X is continuous on $[0, T]$ with respect to the H_{RV} norm. First, if $X(0) \in H_{RV}$, then $X(0) \in H_{SP}$ and

$$\|X\|_{SP} \leq \|X - X(0)\|_{SP} + \|X(0)\|_{SP} \leq \left(\frac{KT^2}{2}\right)^{1/2} + T^{1/2}\|X(0)\|_{RV}.$$

Thus, $\|X\|_{SP}$ is bounded by $\left(\frac{KT^2}{2}\right)^{1/2} + T^{1/2}\|X(0)\|_{RV}$. Also, it is easy to see that X is continuous on $[0,T]$ with respect to the H_{RV} norm. Given $\epsilon > 0$, then

$$\|X(t_2) - X(t_1)\|_{RV} \leq K^{1/2}|t_1 - t_2|^{1/2} \leq \epsilon$$

whenever $|t_1 - t_2| \leq \epsilon^2/K$.

Finally, it is useful to present some terminology used in the literature regarding the Hilbert space H_{SP} that is applied in Chapters 3 and 4 for Itô stochastic integrals and differential equations. Let $W(t,\omega)$ be a Wiener process defined on a probability space $(\Omega, \mathcal{A}, P)$. Let $\{\mathcal{A}(t) : 0 \leq t \leq T\}$ be a family of sub-σ-algebras of $\mathcal{A}$ satisfying $\mathcal{A}(t_1) \subset \mathcal{A}(t_2)$ if $t_1 < t_2$, $W(t)$ is $\mathcal{A}(t)$-measurable, and $W(t+s) - W(t)$ is independent of $\mathcal{A}(t)$. $\mathcal{A}(t)$ is the $\sigma-$algebra of events generated by the values of the Wiener process until time t [70]. A stochastic process $f(t,\omega)$ is said to be adapted [40, 92] to $\mathcal{A}(t)$ if $f(t,\cdot)$ is $\mathcal{A}(t)$-measurable for almost all $t \in [0,T]$. If f is measurable on $[0,T] \times \Omega$, then f is said to be nonanticipative [40, 41]. A nonanticipative function $f(t)$ is independent of a Wiener increment $W(t+s)-W(t)$ for $s > 0$. Furthermore, the Hilbert space H_{SP} is the set of nonanticipative stochastic processes f such that f satisfies $\int_0^T E|f(t)|^2\,dt < \infty$.

2.5 Computer Generation of Stochastic Processes

It is useful to consider how stochastic processes can be computationally simulated using pseudo-random numbers. First, consider simulation of a discrete stochastic process, in particular, a Markov chain $\{X_n\}$ on $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$ where $X_0 = z_0$ and X_n is a discrete random variable for each time t_n, $n = 0, 1, \ldots, N$. Specifically, $X_n \in M = \{z_{-m}, z_{-m+1}, \ldots, z_m\}$. Suppose that the transition probability matrix

$$P_n = [p_{i,j}^{(n)}] \quad \text{where} \quad p_{i,j}^{(n)} = P\{X_{n+1} = z_j | X_n = z_i\}$$

depends on time t_n. Consider generation of one trajectory or sample path $\{X_n,\, 0 \leq n \leq N\}$. At time t_0, $X_0 = z_0$. To find X_1, $p_{0,j}^{(0)}$ are first computed for $j = -m, -m+1, \ldots, m$. Next, a pseudo-random number η_0 uniformly distributed on $[0,1]$ is generated. Then, r_0 is calculated so that

$$\sum_{j=-m}^{r_0-1} p_{0,j}^{(0)} < \eta_0 \leq \sum_{j=-m}^{r_0} p_{0,j}^{(0)}.$$

Finally, X_1 is set equal to z_{r_0}. To find X_2, $p_{r_0,j}^{(1)}$ are computed for $j = -m, -m+1, \ldots, m$. Then, η_1 uniformly distributed on [0,1] is generated and r_1 is calculated so that

$$\sum_{j=-m}^{r_1-1} p_{r_0,j}^{(1)} < \eta_1 \leq \sum_{j=-m}^{r_1} p_{r_0,j}^{(1)}.$$

Then X_2 is set equal to z_{r_1}. These steps are repeated N times to give one realization $\{X_k\}_{k=0}^N$ of the discrete stochastic process.

Now consider generating a trajectory for a continuous Markov process $\{X(t), t \in [0,T]\}$. Generally, as illustrated in the examples described in this section, trajectories of continuous processes are determined at a discrete set of times. Specifically, a trajectory $X(t)$ is calculated at the times $t_0, t_1, \ldots, t_N$ where $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$. Then, $X(t)$ may be approximated between these points using, for example, piecewise linear interpolation.

Example 2.12. Simulation of a Poisson process

Consider a Poisson process $X(t)$ with intensity λ. Recall that the process $X(t)$ equals the number of observations in time t where the probability of one observation in time Δt is equal to $\lambda \Delta t + o((\Delta t)^2)$. From Example 1.4, we saw that

$$P(X(t) = n) = \exp(-\lambda t)\frac{(\lambda t)^n}{n!}.$$

Consider now simulating this continuous stochastic process at the discrete times $t_k = kh$ for $k = 0, 1, 2, \ldots, N$ where $h = T/N$. Let

$$X(t_{k+1}) = X(t_k) + \hat{\eta}_k \quad \text{for} \quad k = 0, 1, \ldots, N-1, \quad \text{where} \quad X_{t_0} = 0$$

and the random numbers $\hat{\eta}_k$ are chosen so that

$$P(\hat{\eta}_k = n) = \exp(-\lambda h)\frac{(\lambda h)^n}{n!} \quad \text{for} \quad n = 0, 1, 2, \ldots.$$

Then, $X(t_k)$ are Poisson distributed with intensity λ at the discrete times $t_0, t_1, \ldots, t_N$. Notice that to find $\hat{\eta}_k$ given η_k uniformly distributed on $[0,1]$, one uses the relation

$$\sum_{j=0}^{\hat{\eta}_k - 1} \exp(-\lambda h)\frac{(\lambda h)^j}{j!} < \eta_k \leq \sum_{j=0}^{\hat{\eta}_k} \exp(-\lambda h)\frac{(\lambda h)^j}{j!}.$$

Example 2.13. Simulation of a Wiener process sample path

Consider the Wiener process $W(t)$ on $[0,T]$. Consider simulating this continuous stochastic process at the discrete times $t_k = kh$ for $k = 0, 1, 2, \ldots, N$ where $h = T/N$. Let

$$X(t_{k+1}) = X(t_k) + \eta_k \quad \text{for} \quad k = 0, 1, \ldots, N-1,$$

where $X_{t_0} = 0$ and η_k are normally distributed numbers with mean 0 and variance h. As in the previous example, each sample path of the continuous stochastic process is computed at the discrete times $t_0, t_1, \ldots, t_N$. Thus,

$W(t_k) = X(t_k)$ for $k = 0, 1, 2, \ldots, N$. To estimate $W(t)$, at a time $t \neq t_k$ for any k, a continuous linear interpolant can be used as was shown in Example 2.11. In particular,

$$W(t) \approx X(t_k)\frac{t_{k+1} - t}{h} + X(t_{k+1})\frac{t - t_k}{h} \quad \text{for} \quad t_k \leq t \leq t_{k+1}.$$

Example 2.14. Simulation of a Wiener process by a discrete process

Let $t_k = kh$ for $k = 0, 1, 2, \ldots, N$ where $h = T/N$. Define the discrete stochastic process $\{X_n\}_{n=0}^N$ on the partition $0 = t_0 < t_1 < \cdots < t_N = T$ in the following way. Let $X_0 = 0$ and let the transition probabilities $p_{i,k} = P\{X_{n+1} = k\delta | X_n = i\delta\}$ be

$$p_{i,k}(t) = \begin{cases} \lambda \Delta t/2\delta^2, & \text{for} \quad k = i - 1 \\ 1 - \lambda \Delta t/\delta^2, & \text{for} \quad k = i \\ \lambda \Delta t/2\delta^2, & \text{for} \quad k = i + 1 \end{cases}$$

assuming that $\lambda \Delta t/\delta^2 < 1$. Then, as explained in Section 2.2, the probability distribution for X_k as $\Delta t \to 0$ satisfies the forward Kolmogorov equations

$$\frac{dp_k(t)}{dt} = \frac{\lambda}{2}\left(\frac{p_{k+1}(t) - 2p_k(t) + p_{k-1}(t)}{\delta^2}\right),$$

where $p_0(0) = 1$ and $p_k(0) = 0$ for $k \neq 0$. For δ small, $p_k(t_n) = P(X(t_n) = k\delta)$ approximately equals $p(t_n, k\delta)\delta$ where $p(t, x)$ satisfies the partial differential equation

$$\frac{\partial p(t,x)}{\partial t} = \frac{\lambda}{2}\frac{\partial^2 p(t,x)}{\partial x^2} \quad \text{with} \quad p(0, x) = \delta(x - 0).$$

Solving this partial differential equation gives

$$p(t, x) = \frac{1}{(2\pi\lambda t)^{1/2}} \exp\left(-\frac{x^2}{2\lambda t}\right).$$

In particular, for $\lambda = 1$, $\Delta t/\delta^2 < 1$, and Δt small, X_n is approximately normally distributed with mean 0 and variance $n\Delta t$. Furthermore, $X_n - X_j$ is approximately normally distributed with mean 0 and variance $(n - j)\Delta t$ and is independent of $X_l - X_m$ for $j \geq l$. Indeed, $\{X_n\}_{n=0}^N$ approximates a Wiener process on the partition $0 = t_0 < t_1 < \cdots < t_N = T$.

2.6 Examples of Stochastic Processes

Stochastic processes are common in physics, biology, meteorology, and finance. Indeed, stochastic processes occur whenever dynamical systems experience random influences. A classic physical stochastic process is radioactive decay where atoms of unstable isotopes transform spontaneously to other isotopes. Suppose that there are initially present n_0 atoms of a radioactive isotope. Let

λ be the decay constant of the isotope. This means that the probability that an atom transforms in small time interval Δt is equal to $\lambda \Delta t + O((\Delta t)^2)$. Let $N(t)$ be the number of atoms at time t. Consider finding the expected number of atoms at time t, i.e., $E(N(t))$. Let $p_n(t)$ be the probability that there are n atoms at time t. Then, considering the possible transitions in time interval Δt, one obtains that

$$p_n(t + \Delta t) = p_{n+1}(t)\lambda(n + 1)\Delta t + p_n(t)(1 - \lambda n \Delta t) + O((\Delta t)^2).$$

Thus, letting $\Delta t \to 0$,

$$\begin{cases} \dfrac{dp_{n_0}(t)}{dt} = -\lambda n_0 p_{n_0}(t) \quad \text{with} \quad p_{n_0}(0) = 1 \quad \text{and} \\ \dfrac{dp_n(t)}{dt} = -\lambda n p_n(t) + \lambda(n + 1)p_{n+1}(t) \quad \text{with} \quad p_n(0) = 0 \quad \text{for} \quad 0 \le n < n_0. \end{cases}$$

The expected number of atoms can now be computed as

$$E(N(t)) = \sum_{n=0}^{n_0} n p_n(t) \quad \text{and so} \quad \frac{dE(N(t))}{dt} = \sum_{n=0}^{n_0} n \frac{dp_n(t)}{dt}.$$

This leads to

$$\begin{aligned} \frac{dE(N(t))}{dt} &= \sum_{n=0}^{n_0} n \frac{dp_n(t)}{dt} = \sum_{n=0}^{n_0} -\lambda n^2 p_n(t) + \sum_{n=0}^{n_0-1} \lambda n(n + 1)p_{n+1}(t) \\ &= \sum_{n=0}^{n_0} -\lambda n^2 p_n(t) + \sum_{n=1}^{n_0} \lambda(n - 1)n p_n(t) \\ &= \sum_{n=0}^{n_0} -\lambda n p_n(t) = -\lambda E(N(t)). \end{aligned}$$

Hence,

$$\frac{dE(N(t))}{dt} = -\lambda E(N(t)) \quad \text{with} \quad E(N(0)) = n_0.$$

So, $E(N(t)) = n_0 \exp(-\lambda t)$ is the expected number of atoms at time t.

Population biology is rich in stochastic processes. Mathematical models are useful for understanding these random biological processes. The birth-death process, in itself, is a random process. Also, variability in the environment introduces additional random influences which are time and spatially varying. As a result, growth of a population exhibits random behavior. This random behavior is exhibited in the Aransas-Wood Buffalo population of whooping cranes [27]. These cranes nest in Wood Buffalo National Park in Canada and winter in Aransas National Wildlife Refuge in Texas. The population size is graphed in Fig. 2.3 over the years 1939–1985. This population of whooping cranes will be studied in more detail in Chapter 4.

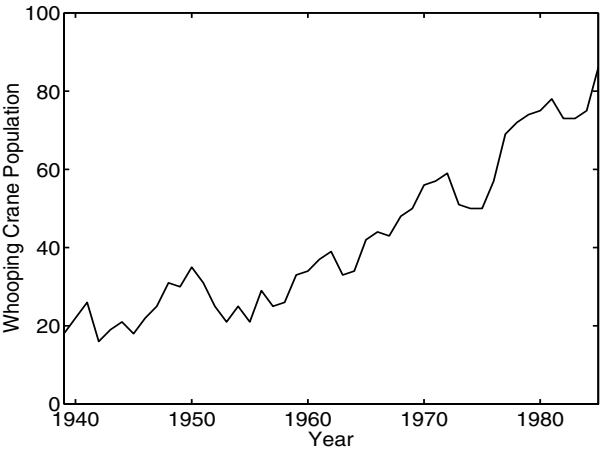

Fig. 2.3. Aransas-Wood Buffalo population of whooping cranes from 1939 to 1985

Because of the actions of many different influences, weather is highly variable and climatic quantities can be considered stochastic processes. Consider, for example, the annual precipitation in Lubbock, Texas which is plotted in Fig. 2.4 from 1911 to 2004. This data exhibits a Wiener-like behavior. The rainfall data will be modeled in Chapter 5.

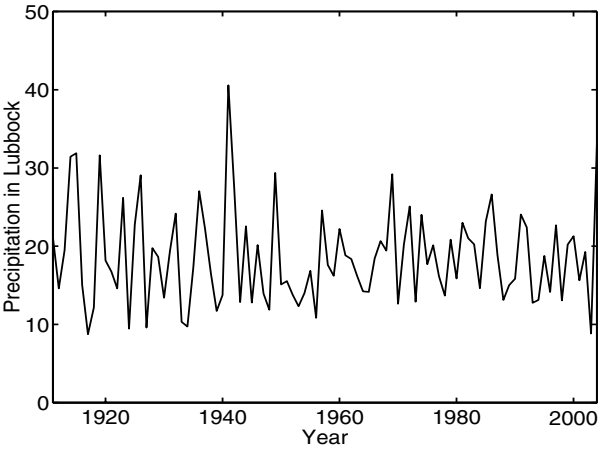

Fig. 2.4. Annual rainfall (in inches) for Lubbock from 1911 to 2004

In addition to physical and biological phenomena, human activities such as stock trading which are quantitative and involve many uncertainties can exhibit a random behavior. Plotted in Fig. 2.5 are stock prices of Best Buy Company and Mattel Inc. The stock prices are graphed as a function of trading

day from January 13, 2004 to January 12, 2005. A model for stock-price dynamics will be presented in Chapter 5.

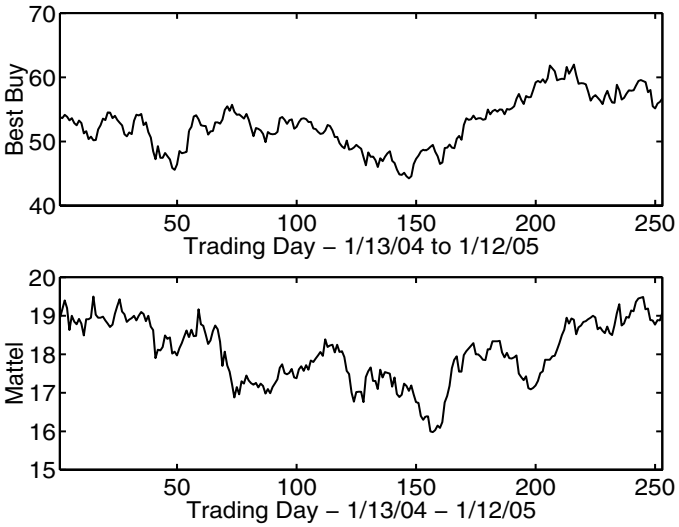

Fig. 2.5. Best Buy Co. and Mattel Inc. stock prices from 1/13/2004 to 1/12/2005

Exercises

2.1. Consider rolling a die once each minute for $t = 1, 2, \ldots$. Define $\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$ where $\omega_k = k$ and $P(\omega_k) = 1/6$ for $k = 1, 2, \ldots, 6$. Define the stochastic process, X, by $X(t, \omega_k) = (k+1)/t^2$. Notice that $X(t)$ is a random variable for each $t \in \tau = \{1, 2, \ldots\}$. Calculate $E(X(t))$ and $\mathrm{Var}(X(t))$.

2.2. Let $\tau = \{0, 1, 2, \ldots\}$ and define the stochastic process, X, on τ by $X(t+1) = X(t) + \gamma_t$ for $t = 0, 1, 2, \ldots$ where $\gamma_t \sim U[0, 1]$ for each t and $X(0) = 1$. Find $E(X(t))$ and $\mathrm{Var}(X(t))$. Notice that X is a continuous-valued discrete stochastic process.

2.3. Let $\tau = \{0, \Delta t, 2\Delta t, 3\Delta t, \ldots\}$ and let $t_n = n\Delta t$ for $n = 0, 1, \ldots$. Let $X_n = X(t_n)$ be a homogeneous Markov chain defined on τ. Assume that the process X has the transition probabilities

$$p_{i,k} = \begin{cases} 1 - \Delta t/2, & \text{for} \quad k = i \\ \Delta t/4, & \text{for} \quad k = i - 1 \\ \Delta t/4, & \text{for} \quad k = i + 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Let $p_k(t) = P(X(t) = k)$ be the probability distribution for X at time t. Assume that $p_0(0) = 1$ and $p_k(0) = 0$ for $k \neq 0$. Show that the probability distribution satisfies

$$p_k(t + \Delta t) = \left(1 - \frac{\Delta t}{2}\right) p_k(t) + p_{k+1}(t)\frac{\Delta t}{4} + p_{k-1}(t)\frac{\Delta t}{4}$$

for $k = \ldots, -2, -1, 0, 1, 2, \ldots$.
(b) Show that as $\Delta t \to 0$, then $p_k(t)$ approximately satisfies

$$\frac{dp_k(t)}{dt} = \frac{1}{4}\left(\frac{p_{k+1}(t) - 2p_k(t) + p_{k-1}(t)}{1^2}\right) \quad \text{for} \quad k = \ldots, -2, -1, 0, 1, 2, \ldots.$$

(c) Show that $p(t, x) = \dfrac{1}{\sqrt{\pi t}} \exp(\frac{-x^2}{t})$ solves

$$\frac{\partial p(t, x)}{\partial t} = \frac{1}{4}\frac{\partial^2 p(t, x)}{\partial x^2}$$

assuming initial condition $p(0, x) = \delta(x - 0)$.
(d) For $\frac{x^2}{t}$ small, $p(t, x) = \dfrac{1}{\sqrt{\pi t}} \exp(\frac{-x^2}{t}) \approx \dfrac{1}{\sqrt{\pi t}}\left(1 - \frac{x^2}{t}\right)$ and thus, $p(t, k) \approx \dfrac{1}{\sqrt{\pi t}}\left(1 - \frac{k^2}{t}\right)$ for $\frac{k^2}{t}$ small. Show that $p_k(t) = \dfrac{1}{\sqrt{\pi t}}\left(1 - \frac{k^2}{t}\right)$ approximately satisfies the differential equation in part (b) for $\frac{k^2}{t}$ small.

2.4. Suppose that $X(t)$ is the number of observations in time t of a randomly occurring phenomenon and the probability for zero observations in time interval Δt is exactly equal to $\exp(-\lambda \Delta t)$ where λ is a positive constant. Derive the probability of one occurrence in time t. That is, derive $P_1(t) = P(X(t) = 1) = \lambda t \exp(-\lambda t)$. Note that $P_0(t + \Delta t) = P_0(t) \exp(-\lambda \Delta t)$ and $P_1(t + \Delta t) = P_1(t)P_0(\Delta t) + P_0(t)P_1(\Delta t)$. (Let $\Delta t \to 0$ and find and solve the differential equations for $P_0(t)$ and $P_1(t)$. Also note that $P_1(\Delta t) = P_1(0) + P_1'(0)\Delta t + O((\Delta t)^2)$.)

2.5. Let $X(t) = \gamma_{i+1}(t - t_i)/h + \gamma_i(t_{i+1} - t)/h$ for $t_i \leq t \leq t_{i+1}$ for $i = 0, 1, 2, \ldots, N - 1$ where $h = T/N$, $t_i = ih$, and $\gamma_i \sim U[-1/2, 1/2]$ are independent and identically distributed for $i = 0, 1, 2, \ldots, N$. Notice that X is a continuous stochastic process. For Δt small compared with h, prove that $E((X(t + \Delta t) - X(t))^2) \leq (\Delta t)^2/h^2$ for any $0 \leq t \leq T$. (Consider two cases: $t_i \leq t < t + \Delta t \leq t_{i+1}$ and $t_i \leq t \leq t_{i+1} < t + \Delta t < t_{i+2}$ for any i.)

2.6. The Wiener process has transition probability density

$$p(y, t, x, s) = \frac{1}{(2\pi(t - s))^{1/2}} \exp\left(\frac{-(x - y)^2}{2(t - s)}\right).$$

Prove, by direct integration, that the Chapman-Kolmogorov formula

$$p(y,t,x,s) = \int_{-\infty}^{\infty} p(z,u,x,s)p(y,t,z,u)\,dz$$

holds for this transition probability density.

2.7. Let

$$W(t) - S_N(t) = E_N(t) = \sum_{n=N+1}^{\infty} \frac{2\sqrt{2T}}{(2n+1)\pi}\eta_n \sin\left(\frac{(2n+1)\pi t}{2T}\right)$$

where $S_N(t)$ is the Nth partial sum of the Karhunen-Loève expansion of $W(t)$ on $[0,T]$. Prove that $\|E_N(t)\|_{RV}^2 \le 2T/(\pi^2 N)$ for $0 \le t \le T$.

2.8. Define the stochastic process X on $[0,T]$ by the formula

$$X(t) = \frac{(t_{i+1}-t)U_i}{h} + \frac{(t-t_i)U_{i+1}}{h} \quad \text{for} \quad t_i \le t \le t_{i+1}$$

for $i = 0,1,\ldots,N-1$ where $h = T/N$, $t_i = ih$, and $U_i \sim U[0,1]$ are independent random variables for $i = 0,1,\ldots,N$. Find $\|X\|_{SP}$.

2.9. Let $\tau = [0,T]$, $X \in H_{SP}$, and $X_k \in H_{SP}$ for $k = 0,1,2,\ldots$. Suppose that $\|X(t) - X_N(t)\|_{RV} \le t/N$ for $0 \le t \le T$.
(a) Prove that $\|X - X_N\|_{SP} \le T^{3/2}/(N\sqrt{3})$.
(b) Prove that given $\epsilon > 0$, $P(|X(t) - X_N(t)| > \epsilon) < \dfrac{1}{\epsilon^2}\dfrac{T^2}{N^2}$ for all $0 \le t \le T$. Thus, for example, given $\epsilon = 10^{-6}$ there is a value of N sufficiently large such that $P(|X(t) - X_N(t)| > 10^{-6}) < 10^{-12}$ for any $t \in [0,T]$.

2.10. Modify the MATLAB program given below and graph three individual Wiener process trajectories, the average of 200 trajectories, and the variance of 200 trajectories from $t = 0$ to $t = 5$. Use 500 points equally spaced on $t = 0$ to $t = 5$. Put all plots on one figure and hand in the figure.

```
% This program can be used to solve a system of SDEs.
% If a1=0, a2=0, b11=1, b12=0, b21=0, and b22=1, then
% Wiener processes are plotted.
clear
clf
xx=56430.;
n=500;
nrun=200;
tf=5.0;
h=tf/n;
hs=sqrt(h);
```

```
x=linspace(0,tf,n);
y3=zeros(1,n);
y4=zeros(1,n);
ya=zeros(1,n);
yv=zeros(1,n);
sa=0.0;
sv=0.0;
for ii=1:nrun
for i=1:n-1
   rr=random(xx);
   xx=rr(3);
   a1=0;
   a2=0;
   b11=1;
   b12=0;
   b21=0;
   b22=1;
   y3(i+1)=y3(i)+a1*h+hs*b11*rr(1)+hs*b12*rr(2);
   y4(i+1)=y4(i)+a2*h+hs*b21*rr(1)+hs*b22*rr(2);
   ya(i+1)=ya(i+1)+y4(i+1)/nrun;
   yv(i+1)=yv(i+1)+y4(i+1)*y4(i+1)/nrun;
  end
  sa=sa+y4(n)/nrun;
  sv=sv+y4(n)*y4(n)/nrun;
  end
  set(gca,'fontsize',15,'linewidth',1.5);
  plot(x,y3,'k-','linewidth',1.5)
  hold on
  plot(x,y4,'k-','linewidth',1.5)
  hold on
  plot(x,ya,'k--','linewidth',1.5)
  hold on
  plot(x,yv,'k:','linewidth',1.5)
  axis([0,tf,-6.0,6.0])
  xlabel('Time t')
  ylabel('Wiener Processes, Average, and Variance')
  hold off
%-------------------------------------------------
% Function random generates random numbers using
%  the congruential generator xx=16807*xx mod(2^31-1).
% The Box-Muller method converts to normal random numbers.
% xx=rand(3) is input to the function rand=random(xx).
       a=16807;
       b=2147483647;
   for i=1:2
```

```
        d=fix(a*xx/b);
        xx=a*xx-d*b;
        rng(i)=xx/b;
    end
        p=3.141592654;
        u1=rng(1);
        u2=rng(2);
        hlp=sqrt(-2.0*log(u1));
        rand(1)=hlp*cos(p*2.0*u2);
        rand(2)=hlp*sin(p*2.0*u2);
        rand(3)=xx;
```

2.11. Suppose for interval $[0,T]$ that $f \in H_{SP}$ and $f_n \in H_{SP}$ for $n = 1, 2, \ldots$. Also, suppose that $\|f - f_n\|_{SP}^2 < \delta_n$ where $\delta_n \to 0$ as $n \to \infty$. Let $\mu_n = \{t \in [0,T] : E(f(t) - f_n(t))^2 > \sqrt{\delta_n}\}$ and define $\text{meas}(\mu_n) = \int_{t\in\mu_n} dt$.
(a) Show that $\sqrt{\delta_n}\,\text{meas}(\mu_n) \le \int_{t\in\mu_n} E(f(t) - f_n(t))^2\, dt \le \delta_n$ and thus show that $\text{meas}(\mu_n) \le \sqrt{\delta_n} \to 0$ as $n \to \infty$.
(b) Show, using Chebyshev's inequality, that $P(\{\omega : |f(t,\omega) - f_n(t,\omega)| > \epsilon\}) \le \dfrac{1}{\epsilon^2}\sqrt{\delta_n}$ for $t \notin \mu_n$. Notice that parts (a) and (b) imply, given $\epsilon > 0$ and $t \in [0,T]$, that the likelihood decreases as n increases that the values on the sample paths, $f(t,\omega)$ and $f_n(t,\omega)$, differ by more than ϵ.

2.12. Let X be a stochastic process on $[0,T]$ and define the elementary stochastic process X_N by $X_N(t) = \sum_{i=0}^{N-1} X(t_i) I_i(t)$, where

$$I_i(t) = \begin{cases} 1, & t_i \le t < t_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

$h = T/N$, and $t_i = ih$ for $i = 0, 1, 2 \ldots, N$. Suppose that $X_N \in S_{SP}$ for $N = 1, 2, \ldots$, and suppose given $\epsilon > 0$ there is an integer M such that $\|X - X_N\|_{SP} < \epsilon$ for $N \ge M$. Let $Y = G(X)$ define another stochastic process where G satisfies the Lipschitz condition $|G(r) - G(s)| \le L|r - s|$ for all $r, s \in \mathbb{R}$. Let $Y_N(t) = \sum_{i=0}^{N-1} G(X(t_i)) I_i(t)$. Show that $\|Y - Y_N\|_{SP} < L\epsilon$ for $N \ge M$. Hence, $\{Y_N\}$ are elementary stochastic processes that converge to Y in H_{SP}.

Computer Programs

Program 2.1. A computer program to generate Poisson processes

Listed is a MATLAB program that makes Fig. 2.1. A sample path is plotted along with the average of nrun sample paths for a Poisson process with intensity λ.

```
% MATLAB program to plot a Poisson process.
% Function file unirand.m is needed.
% Function unr=unirand(xx) returns two numbers:
%  unr(1) is uniformly distributed on [0,1], and
%  xx is set equal to unr(2) for the next call to unirand.m.
% The average of nrun paths and one sample path are plotted.
% The program can readily be modified to plot two paths.
 clear
 clf
 xx=4034218;
% xx starts the random number sequence
 nt=500;
 nrun=200;
 time=10;
 lambda=1;
% nt is number of time intervals, time is total time
% nrun is the number of different paths
 h=time/nt;
 tt=linspace(0,time,nt+1);
 sm=zeros(nt+1,1);
 s2=zeros(nt+1,1);
 patha=zeros(nt+1,1);
 pathb=zeros(nt+1,1);
 for jj=1:nrun
  y=0;
  % y=0 is initial value for all paths
   sm(1)=y;
   s2(1)=y*y;
   patha(1)=y;
   pathb(1)=y;
   for i=1:nt
   unr=unirand(xx);
   xx=unr(2);
% need xx=unr(2) for next call to unirand.m
   if unr(1) < lambda *h
      y=y+1;
   end
   sm(i+1)=sm(i+1)+y/nrun;
   s2(i+1)=s2(i+1)+y*y/nrun;
   pathb(i+1)=patha(i+1);
   patha(i+1)=y;
   end
   end
   set(gca,'fontsize',18,'linewidth',1.5);
   plot(tt,patha,'-',tt,sm,'--','linewidth',1.5);
```

```
   axis([0,10,0,10]);
   xlabel('Time t')
   ylabel('Poisson Process - Unit Intensity')
   set(gca,'linewidth', 1.5);
%-------------------------------------------
% Function file unirand.m generates random numbers using
%  the congruential generator xx=16807*xx mod(2^31-1).
% The random number unr(1) is uniform on [0,1].
% The value xx=unr(2) is input to the generator.
% Call using two statements: unr=unirand(xx); xx=unr(2);
  function [unr] = unirand(xx)
  a=16807;
  b=2147483647;
  d=fix(a*xx/b);
  xx=a*xx-d*b;
  unr(1)=xx/b;
  unr(2)=xx;
```

3

Stochastic Integration

3.1 Introduction

Integrals of the form $\int_a^t f(s,\omega)\,ds$ and $\int_a^t g(s,\omega)\,dW(s,\omega)$ for $a \le t \le b$, where f and g are stochastic processes on $(\Omega, \mathcal{A}, P)$, are studied in this chapter. If f and g satisfy certain conditions and are stochastic processes in Hilbert space H_{SP}, then the integrals will also be stochastic processes in this Hilbert space. Properties of these integrals will be useful in the next chapter where stochastic differential equations are discussed.

As stochastic differential equations are written in the form of stochastic integrals, stochastic integrals are important in the study of stochastic differential equations and properties of stochastic integrals determine properties of stochastic differential equations. It is assumed in this chapter that the integrands are elements of Hilbert space H_{SP}. This assumption unifies and simplifies the presentation of stochastic integrals and stochastic differential equations. Many texts are available which present excellent introductions to stochastic integrals such as references [20, 29, 40, 41, 42, 69, 70, 90, 92, 96].

In this chapter, Itô stochastic integrals are defined and some important properties are examined. A method to approximate Itô stochastic integrals is described. Stochastic differentials are described and the important Itô's formula is derived. Application of Itô's formula in Chapter 4 enables certain stochastic differential equations to be solved exactly. Furthermore, Itô's formula can be used to derive additional important results for stochastic differential equations. Finally, another commonly used stochastic integral, a Stratonovich stochastic integral, is briefly discussed.

3.2 Integrals of the Form $\int_a^t f(s,\omega)\,ds$

In this section, integrals of the forms $J(f) = J(f)(\omega) \equiv \int_a^b f(s,\omega)\,ds$ and $J(f)(t) = J(f)(t,\omega) \equiv \int_a^t f(s,\omega)\,ds$ for $a \le t \le b$ are defined and studied.

These integrals are mappings from the Hilbert spaces H_{SP} to H_{RV} or H_{SP} to H_{SP}, respectively. That is, for $f \in H_{SP}$, the first integral is a random variable in H_{RV} while the second integral is a stochastic process in H_{SP}.

In this section and the next section, three conditions are assumed on the stochastic process f, in addition to $f \in H_{SP}$. These conditions are the following:

Condition (c1): $f(a) \in H_{RV}$. Hence, $\|f(a)\|_{RV}^2 = E|f(a)|^2 \leq k_1$ for a positive constant k_1.

Condition (c2): $\|f(t_2) - f(t_1)\|_{RV}^2 = E|f(t_2) - f(t_1)|^2 \leq k_2|t_2 - t_1|$ for any $t_1, t_2 \in [a, b]$ for a positive constant k_2.

Condition (c3): f is nonanticipating on [a,b].

Notice that if $f \in H_{SP}$ satisfies (c1) and (c2), then $\|f(t)\|_{RV} \leq k_2^{1/2}(b - a)^{1/2} + \|f(a)\|_{RV}$ for any $t \in [a, b]$. This boundedness property follows from the triangle inequality $\|f(t)\|_{RV} \leq \|f(t) - f(a)\|_{RV} + \|f(a)\|_{RV}$. The third condition, that f is nonanticipating on $[a, b]$, means essentially that $f(t, \omega)$ does not depend on time t' for $t' > t$. Hence, $E\big(f(t)(W(t') - W(t))\big) = E(f(t))E(W(t') - W(t)) = 0$ for all $a \leq t \leq t' \leq b$. For example, $f_1(t) = 3\cos(W^2(t)) + 4W(t) - 5t$ is nonanticipating while $f_2(t) = W((t + b)/2)$ is anticipating for $a < t < b$. Indeed, $E\big(f_1(t)(W(t') - W(t))\big) = 0$ for $t' > t$ while

$$E\big(f_2(t)(W(t') - W(t))\big) = \begin{cases} t' - t, & \text{for} \quad t < t' \leq (t + b)/2 \\ (b - t)/2, & \text{for} \quad (t + b)/2 \leq t' \leq b. \end{cases}$$

To motivate the definition of $J(f) = J(f)(\omega) \equiv \int_a^b f(s, \omega)\, ds$ for $f \in H_{SP}$, the integral for a step function approximation to f is considered. Recall that the set of step functions $S_{SP} \subset H_{SP}$ is dense in H_{SP}. Let $a = t_0 < t_1 < \cdots < t_{m-1} < t_m = b$ be of a family of partitions of $[a, b]$ where $\max\limits_{1 \leq i \leq m} |t_i - t_{i-1}| \to 0$ as $m \to \infty$ and let

$$I_i(t) = \begin{cases} 1 & \text{for} \quad t_i \leq t < t_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

for $i = 0, 1, 2, \ldots, N - 1$. Furthermore, let

$$f_m(t, \omega) = \sum_{i=0}^{m-1} f_i^{(m)}(\omega) I_i(t), \quad \text{where,} \quad f_i^{(m)} \in H_{RV} \quad \text{for each} \quad i \quad \text{and} \quad m$$

be a sequence of step functions in S_{SP} that converges to f. That is, given $\epsilon > 0$ there is an $M > 0$ such that

$$\|f - f_m\|_{SP} < \epsilon \quad \text{when} \quad m \geq M.$$

Before continuing, the integral of step function $f_m(t, \omega)$ is defined in a standard way [73].

Definition 3.1. *Let* $f_m(t,\omega) = \sum_{i=0}^{m-1} f_i^{(m)}(\omega) I_i(t)$ *be an element of* S_{SP}. *Then the integral* $\int_a^b f_m(s)\,ds$ *is denoted as* $J(f_m)$ *and is defined as*

$$J(f_m) = \int_a^b f_m(s)\,ds = \sum_{i=0}^{m-1} f_i^{(m)}\,(t_{i+1} - t_i).$$

Notice that $J(f_m) \in H_{RV}$ since by the Cauchy-Schwarz inequality,

$$\|J(f_m)\|_{RV}^2 = E\left|\int_a^b f_m(s)\,ds\right|^2 \le \sum_{i=0}^{m-1}(t_{i+1}-t_i) E \sum_{i=0}^{m-1} |f_i^{(m)}|^2\,(t_{i+1}-t_i)$$
$$= (b-a)E\int_a^b |f_m(s)|^2\,ds = (b-a)\|f_m\|_{SP}^2.$$

Also, notice that $\{J(f_m)\}_{m=1}^{\infty}$ is a Cauchy sequence in H_{RV} and therefore has a limit in H_{RV}. To see that $\{J(f_m)\}_{m=1}^{\infty}$ is a Cauchy sequence in H_{RV} consider

$$\|J(f_m) - J(f_n)\|_{RV}^2 = E\left|\int_a^b (f_m(t) - f_n(t))\,dt\right|^2 \le (b-a)\|f_m - f_n\|_{SP}^2.$$

The right-hand side can be made arbitrarily small as $\{f_m\}_{m=1}^{\infty}$ is a Cauchy sequence in H_{SP}. Thus, $J(f_m)$ converges in H_{RV} as $m \to \infty$. This limit is defined to be $J(f)$. This discussion is summarized in the following definition.

Definition 3.2. *Given* $f \in H_{SP}$ *and a sequence* $\{f_m\}_{m=1}^{\infty}$ *in* S_{SP} *such that* $\|f - f_m\|_{SP} \to 0$ *as* $m \to \infty$, *then the integral* $\int_a^b f(s)\,ds$ *is denoted as* $J(f)$ *and is defined as*

$$J(f) = \int_a^b f(s)ds = \lim_{m\to\infty}\int_a^b f_m(s)\,ds = \lim_{m\to\infty}\sum_{i=0}^{m-1} f_i^{(m)}\,(t_{i+1}-t_i).$$

A useful inequality follows from this definition. By definition, if $\|f - f_m\|_{SP} \to 0$, then $\int_a^b f(s)\,ds = \lim_{m\to\infty}\int_a^b f_m(s)\,ds$. However, by the Cauchy-Schwarz inequality, $E\left(\int_a^b f_m(s)\,ds\right)^2 \le T\int_a^b E|f_m(s)|^2\,ds$. Hence, for $f \in H_{SP}$,

$$E\left(\int_a^b f(s)\,ds\right)^2 \le T\int_a^b E|f(s)|^2\,ds.$$

Also, notice for $f \in H_{SP}$ that satisfies inequality (c2), i.e., $\|f(t_2) - f(t_1)\|_{RV}^2 = E|f(t_2) - f(t_1)|^2 \le k_2|t_2 - t_1|$ for any $t_1, t_2 \in [a,b]$ for a positive constant k_2, then

$$\int_a^b f(s)\,ds = \lim_{m\to\infty}\sum_{i=0}^{m-1} f(t_i)\Delta t,$$

where $\Delta t = (b-a)/m, t_i = i\Delta t + a$, and the limit is in H_{RV}. This result follows by letting $f_m(t) = \sum_{i=0}^{m-1} f(t_i) I_i(t)$ and seeing that $\|f - f_m\|_{SP} \to 0$ as $m \to \infty$.

Now consider integrals of the form $J(f) = J(f)(t,\omega) \equiv \int_a^t f(s,\omega)\, ds$ for $a \le t \le b$ where $f \in H_{SP}$. In this case, the integral is defined as in Definition 3.2 for each value of t for $a \le t \le b$. This integral, however, maps a stochastic process in H_{SP} to another stochastic process in H_{SP}. Indeed, $\|J(f)\|_{SP} \le (b-a)\|f\|_{SP}/\sqrt{2}$. To see this, the Cauchy-Schwarz inequality and Fubini's theorem are applied in the following argument.

$$\begin{aligned}\|J(f)\|_{SP}^2 &= \int_a^b E\left|\int_a^t f(s)\, ds\right|^2 dt \le \int_a^b E\left(\int_a^t |f(s)|^2\, ds\right)(t-a)\, dt \\ &\le \int_a^b (t-a)\left(\int_a^b E|f(s)|^2\, ds\right) dt = \frac{(b-a)^2}{2}\|f\|_{SP}^2.\end{aligned}$$

In addition, it is useful to notice that $J(f)(t) = \int_a^t f(s)\, ds$ is continuous in H_{RV} with respect to t on $[a,b]$. To see this, note that

$$\|J(f)(t_1) - J(f)(t_2)\|_{RV}^2 = E\left(\int_{t_1}^{t_2} f(s)\, ds\right)^2 \le |t_2 - t_1|\|f\|_{SP}^2.$$

Consider now the following two examples.

Example 3.3. Integration of a function of a Wiener process

Consider $J(e^{-W}) = \int_0^T \exp(-W(t))\, dt$ where $W(t)$ is a Wiener process on the interval $[0,T]$. For this example, it is interesting that

$$\begin{aligned}E\left(J(e^{-W})\right) &= E\int_0^T \exp(-W(t))\, dt \;=\; E\int_0^T \sum_{k=0}^{\infty} \frac{(-W(t))^k}{k!}\, dt \\ &= \int_0^T E\sum_{k=0}^{\infty} \frac{(-W(t))^k}{k!}\, dt \;=\; \int_0^T \sum_{j=0}^{\infty} \frac{(t/2)^j}{j!}\, dt \\ &= \int_0^T e^{t/2}\, dt \;=\; 2(e^{T/2} - 1).\end{aligned}$$

Example 3.4. An integral related to the Karhunen-Loève expansion

Consider the integral

$$\eta_n = \frac{(2n+1)\pi}{2^{1/2}T^{3/2}} \int_0^T W(t) \sin\left(\frac{(2n+1)\pi t}{2T}\right) dt,$$

where n is a nonnegative integer. This stochastic integral appears in the sine series expansion (Karhunen-Loève expansion) of the Wiener process discussed in Section 2.3. In this example, it will be verified that $E(\eta_n \eta_m) = 0$ when

$n \neq m$ and $E(\eta_n \eta_m) = 1$ if $n = m$, i.e., $E(\eta_n \eta_m) = \delta_{mn}$ where δ_{mn} is the Kronecker delta. As $E(W(t)W(s)) = \min\{t, s\}$, then

$$\begin{aligned} E(\eta_n \eta_m) &= \left(\frac{2a_n a_m}{T}\right) \int_0^T \sin(a_n s) \left[\int_0^s t \sin(a_m t)\, dt + \int_s^T s \sin(a_m t)\, dt \right] ds \\ &= \left(\frac{2a_n a_m}{T}\right) \int_0^T \frac{\sin(a_n s) \sin(a_m s)}{a_m^2}\, ds \\ &= \frac{2(2n+1)}{T(2m+1)} \int_0^T \sin(a_n s) \sin(a_m s)\, ds \\ &= \delta_{mn}, \end{aligned}$$

where $a_n = \dfrac{(2n+1)\pi}{2T}$.

3.3 Itô Stochastic Integrals

In this section, integrals of the form $I(f) = I(f)(\omega) = \int_a^b f(s, \omega)\, dW(s, \omega)$ and $I(f)(t) = I(f)(t, \omega) = \int_a^t f(s, \omega)\, dW(s, \omega)$ for $a \leq t \leq b$ are defined and studied. These integrals, like the integrals in the previous section are mappings from H_{SP} to H_{RV} or from H_{SP} to H_{SP}, respectively. As in the previous section, the stochastic process $f \in H_{SP}$ is assumed to satisfy the three conditions (c1), (c2), and (c3).

Consider the stochastic integral $I(f) = I(f)(\omega) = \int_a^b f(s, \omega)\, dW(s, \omega)$. As was done for the stochastic integrals $J(f)$ described in the previous section, $I(f)$ is first defined for elementary functions $f_m \in S_{SP} \subset H_{SP}$. In particular, the integral $\int_c^d dW(t)$ is defined as $W(d) - W(c)$ for $a \leq c \leq d \leq b$. The integral $\int_a^b f_m(t)\, dW(t)$ is defined in Definition 3.5 Motivation for this definition is given in the following argument. Notice that

$$\begin{aligned} I(f_m) &= I(f_m)(\omega) \\ &= \int_a^b f_m(t, \omega)\, dW(t, \omega) = \int_a^b \sum_{i=0}^{m-1} f_i^{(m)}(\omega) I_i(t)\, dW(t, \omega) \\ &= \sum_{i=0}^{m-1} f_i^{(m)}(\omega) \big(W(t_{i+1}, \omega) - W(t_i, \omega) \big) = \sum_{i=0}^{m-1} f_i^{(m)}(\omega) \Delta W_i, \end{aligned}$$

where $a = t_0 < t_1 < \cdots < t_{m-1} < t_m = b$ is a partition of $[a, b]$, $\Delta W_i = \Delta W_i(\omega) = W(t_{i+1}, \omega) - W(t_i, \omega)$, and

$$I_i(t) = \begin{cases} 1, & \text{for} \quad t_i \leq t \leq t_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

for $i = 0, 1, 2, \ldots, N-1$.

Definition 3.5. Stochastic integral for step functions *For nonanticipating $f_m \in S_{SP}$ where $f_m(t,\omega) = \sum_{i=0}^{m-1} f_i^{(m)}(\omega) I_i(t)$ and $f_i^{(m)} \in H_{RV}$ for each i and m, then the integral $\int_a^b f_m(s)\,dW(s)$ is defined to be*

$$I(f_m) = \int_a^b f_m(s)\,dW(s) = \sum_{i=0}^{m-1} f_i^{(m)}\,\Delta W_i,$$

where $\Delta W_i = W(t_{i+1}) - W(t_i)$.

Notice that $I(f_m) \in H_{RV}$ and

$$\begin{aligned}\|I(f_m)\|_{RV}^2 &= E\left|\sum_{i=0}^{m-1} f_i^{(m)}\,\Delta W_i\right|^2 = \sum_{i=0}^{m-1} \|f_i^{(m)}\|_{RV}^2 \Delta t_i \\ &= \int_a^b E|f_m(t)|^2\,dt = \|f_m\|_{SP}^2,\end{aligned}$$

where $\Delta t_i = t_{i+1} - t_i$ for $i = 0, 1, \ldots, m-1$. So, for $f_m \in S_{SP}$, $\|I(f_m)\|_{RV} = \|f_m\|_{SP}$.

Now consider $f \in H_{SP}$ where f satisfies conditions (c1)–(c3). Define a family of partitions of $[a,b]$ by $a = t_0^{(m)} < t_1^{(m)} < \cdots < t_m^{(m)} = b$ where $t_i^{(m)} = i\Delta t + a$ for $i = 0, 1, \ldots, m$ and $\Delta t = (b-a)/m$. Define a sequence $\{f_m\}_{m=0}^{\infty} \subset S_{SP}$ by

$$f_m(t,\omega) = \sum_{i=0}^{m-1} f(t_i^m, \omega) I_i^{(m)}(t) \quad \text{where} \quad I_i^{(m)}(t) = \begin{cases} 1, & \text{for} \quad t_i^{(m)} \le t < t_{i+1}^{(m)} \\ 0, & \text{otherwise.} \end{cases}$$

Notice that by condition (c2),

$$\begin{aligned}\|f - f_m\|_{SP}^2 &= \int_a^b E|f(t) - f_m(t)|^2\,dt \\ &= \sum_{i=0}^{m-1} \int_{t_i^{(m)}}^{t_{i+1}^{(m)}} E|f(t) - f_m(t)|^2\,dt \le k_2 \sum_{i=0}^{m-1} \int_{t_i^{(m)}}^{t_{i+1}^{(m)}} (t - t_i^{(m)})\,dt \\ &= \frac{k_2}{2} \sum_{i=0}^{m-1} \left(\frac{b-a}{m}\right)^2 = \frac{k_2(b-a)^2}{2m}.\end{aligned}$$

Thus, $f_m \to f$ in H_{SP} as $m \to \infty$. The sequence $\{f_m\}_{m=0}^{\infty}$ is a Cauchy sequence in S_{SP} that converges to $f \in H_{SP}$. Consider $I(f_m) \in H_{RV}$ for $m = 0, 1, 2, \ldots$. In particular,

$$\begin{aligned}\|I(f_m) - I(f_n)\|_{RV}^2 &= E\left(\int_a^b (f_m(t) - f_n(t))\,dW(t)\right)^2 \\ &= \|f_m - f_n\|_{SP}^2 \to 0 \quad \text{as} \quad m, n \to \infty.\end{aligned}$$

The sequence $\{I(f_m)\}_{m=1}^{\infty}$ is a Cauchy sequence in H_{RV} and therefore has a limit in H_{RV}. The above discussion leads to the following definition of $\int_a^b f(t)\,dW(t)$ for $f \in H_{SP}$.

Definition 3.6. Itô stochastic integral $\int_a^b f(s)\,dW(s)$ *Let $f \in H_{SP}$ satisfy conditions (c1)–(c3). The integral $\int_a^b f(t)\,dW(t)$ is defined to be*

$$I(f) = \lim_{m\to\infty} \int_a^b f_m(t)\,dW(t) = \lim_{m\to\infty} \sum_{i=0}^{m-1} f(t_i^{(m)}) \left(W(t_{i+1}^{(m)}) - W(t_i^{(m)})\right),$$

where $t_i^{(m)} = a + i\left(\dfrac{b-a}{m}\right)$ and the convergence is in H_{RV}.

Now consider stochastic integrals of the form

$$I(f) = I(f)(t,\omega) = \int_a^t f(s,\omega)\,dW(s,\omega) \quad \text{for} \quad a \le t \le b, \quad \text{where} \quad f \in H_{SP}$$

and f satisfies conditions (c1)–(c3). As in the situation for the integral $J(f)$, this integral can be defined as in Definition 3.6 for each value of t in $[a,b]$. In this case, I is a mapping from Hilbert space H_{SP} to H_{SP}, i.e., $I : H_{SP} \to H_{SP}$. Indeed, using property (d) of Table 3.1,

$$\|I(f)\|_{SP}^2 = \int_a^b E\left|\int_a^t f(s)\,dW(s)\right|^2 dt = \int_a^b \int_a^t E|f(s)|^2\,ds\,dt \le (b-a)\|f\|_{SP}^2.$$

In addition, $I(f)(t) \in H_{RV}$ for each $t \in [a,b]$ is continuous with respect to variable t on $[a,b]$ in H_{RV}. To see this, notice that by condition (c2)

$$\begin{aligned}
\|I(f)(t_2) - I(f)(t_1)\|_{RV}^2 &= E\left|\int_{t_1}^{t_2} f(s)\,dW(s)\right|^2 = \int_{t_1}^{t_2} E|f(s)|^2\,ds \\
&\le 2\int_{t_1}^{t_2} E|f(s) - f(t_1)|^2\,ds + 2\int_{t_1}^{t_2} E|f(t_1)|^2\,ds \\
&\le k_2(t_2 - t_1)^2 + 2|t_2 - t_1|\|f(t_1)\|_{RV}^2 \\
&\le |t_2 - t_1|\left(5k_2(b-a) + 4\|f(a)\|_{RV}^2\right)
\end{aligned}$$

which follows from $\|f(t)\|_{RV} \le k_2^{1/2}(b-a)^{1/2} + \|f(a)\|_{RV}$. In summary, $\int_a^t f(s)\,dW(s)$ is defined in the following way.

Definition 3.7. Itô stochastic integral $\int_a^t f(s)\,dW(s)$ *Let $f \in H_{SP}$ satisfy conditions (c1)–(c3). The integral $\int_a^t f(t)\,dW(t)$ is defined to be*

$$I(f)(t) = I(f)(t) = \lim_{m\to\infty} \sum_{i=0}^{m-1} f(t_i^{(m)})(W(t_{i+1}^{(m)}) - W(t_i^{(m)})),$$

where $t_i^{(m)} = a + i\left(\dfrac{t-a}{m}\right)$ and the convergence is in H_{RV}.

Some important and useful properties of the integrals $J(f) = \int_a^b f(t,\omega)\,dt$ and $I(f) = \int_a^b f(t,\omega)\,dW(t,\omega)$ for $f \in H_{SP}$ are listed in Table 3.1. The integral $J(f)$ satisfies properties (a) and (b) but not properties (c) and (d) while $I(f)$ satisfies properties (a)–(d). In properties (a)–(d), it is assumed that $f, g \in H_{SP}$ satisfy conditions (c1)–(c3) and that c is a constant.

Table 3.1. Some useful properties of Itô stochastic integrals

	Property
(a)	$I(f+g) = I(f) + I(g)$
(b)	$I(cf) = cI(f)$
(c)	$E(I(f)) = 0$
(d)	$E\|I(f)\|^2 = \int_a^b E\|f(t)\|^2\,dt$

Consider properties (c) and (d). To prove property (c), let

$$I(f) = \lim_{m\to\infty} I(f_m), \quad \text{where} \quad f_m(t) = \sum_{i=0}^{m-1} f(t_i^{(m)}) I_i^{(m)}(t)$$

and

$$I_i^{(m)}(t) = \begin{cases} 1, & \text{for} \quad t_i^{(m)} \le t < t_{i+1}^{(m)} \\ 0, & \text{otherwise.} \end{cases}$$

Then, given $\epsilon > 0$ there is an M such that $\|I(f) - I(f_m)\|_{RV} < \epsilon$ when $m > M$. By the Lyapunov inequality,

$$|E(I(f) - I(f_m))| \le (E|I(f) - I(f_m)|^2)^{1/2} < \epsilon \quad \text{when} \quad m > M.$$

In addition, it is clear that

$$E(I(f_m)) = E\left(\sum_{i=0}^{m-1} f(t_i^{(m)})\,\Delta W_i\right) = 0 \quad \text{for every value of} \quad m.$$

Summarizing the above,

$$|E(I(f))| \le |E(I(f) - I(f_m))| + |E(I(f_m))| < \epsilon \quad \text{when} \quad m > M.$$

As ϵ is arbitrary, $E(I(f)) = 0$.

To prove property (d), the above argument is continued. By the triangle inequality,

$$\|I(f_m)\|_{RV} - \|I(f) - I(f_m)\|_{RV} \le \|I(f)\|_{RV} \le \|I(f_m)\|_{RV} + \|I(f) - I(f_m)\|_{RV}.$$

Thus,

$$\|I(f_m)\|_{RV} - \epsilon \;\le\; \|I(f)\|_{RV} \;\le\; \|I(f_m)\|_{RV} + \epsilon \quad \text{when} \quad m > M.$$

However, it is straightforward to see that

$$\begin{aligned}\|I(f_m)\|_{RV} &= \left(\sum_{i=0}^{m-1} E|f(t_i^{(m)})|^2 \frac{b-a}{m}\right)^{1/2} \\ &\to \left(\int_a^b E|f(t)|^2\,dt\right)^{1/2} \qquad \text{as} \quad m \to \infty.\end{aligned}$$

Since ϵ is arbitrary,

$$\|I(f)\|_{RV}^2 = E\left|\int_a^b f(t)\,dW(t)\right|^2 = \int_a^b E|f(t)|^2\,dt$$

which completes the proof of property (d).

Example 3.8. Mean and mean square of a stochastic integral

Let $I(f) = \int_0^1 t\,dW(t)$. Then, by the properties (c) and (d) of Itô integrals,

$$E(I(f)) = 0 \quad \text{and} \quad E(|I(f)|^2) = \int_0^1 t^2\,dt = \frac{1}{3}.$$

Example 3.9. Mean and mean square of a stochastic integral

Let $I(f) = \int_0^1 W(t)\,dW(t)$. Then, by the properties (c) and (d) of Itô integrals,

$$E(I(f)) = 0 \quad \text{and} \quad E(|I(f)|^2) = \int_0^1 E|W(t)|^2\,dt = \int_0^1 t\,dt = \frac{1}{2}.$$

Example 3.10. An important double stochastic integral

A double stochastic integral that is commonly seen in numerical methods for solving stochastic differential equations [69, 70, 90] is the integral

$$I(1,2) = \int_t^{t+\Delta t} \int_t^s dW_1(r)\,dW_2(s),$$

where W_1 and W_2 are two independent Wiener processes. This double integral cannot be evaluated exactly so approximations must be used to estimate the integral. One approximation is [90]:

$$\begin{aligned}\tilde{I}(1,2) &= \sum_{i=0}^{M-1} \int_{t_i}^{t_{i+1}} \int_{t_0}^{t_i} dW_1(r)\,dW_2(s) \\ &= \sum_{i=0}^{M-1} (W_1(t_i) - W_1(t_0))(W_2(t_{i+1}) - W_2(t_i)),\end{aligned}$$

where $t_i = t + i\Delta t/M$ for $i = 0, 1, 2, \ldots, M$ so $t_0 = t$ and $t_M = t + \Delta t$. The purpose of this example is to show that

$$\|I(1,2) - \tilde{I}(1,2)\|_{RV}^2 = E|I(1,2) - \tilde{I}(1,2)|^2 = \frac{(\Delta t)^2}{2M}$$

indicating that the error in the H_{RV} norm is proportional to $\Delta t/\sqrt{M}$. To see this result, consider

$$\begin{aligned}
&E|I(1,2) - \tilde{I}(1,2)|^2 \\
&\quad = E\left|\sum_{i=0}^{M-1} \int_{t_i}^{t_{i+1}} \left[(W_1(s) - W_1(t_0)) - (W_1(t_i) - W_1(t_0))\right] dW_2(s)\right|^2 \\
&\quad = \sum_{i=0}^{M-1} E\left|\int_{t_i}^{t_{i+1}} \big(W_1(s) - W_1(t_i)\big)\, dW_2(s)\right|^2 \\
&\quad = \sum_{i=0}^{M-1} \int_{t_i}^{t_{i+1}} E|W_1(s) - W_1(t_i)|^2\, ds \\
&\quad = \sum_{i=0}^{M-1} \int_{t_i}^{t_{i+1}} (s - t_i)\, ds \;=\; \frac{1}{2}\sum_{i=0}^{M-1} \left(\frac{\Delta t}{M}\right)^2 \;=\; \frac{(\Delta t)^2}{2M}.
\end{aligned}$$

Finally, recall that if $f \in H_{SP}$ satisfies condition (c2) on $[a, b]$, then $f(t)$ is continuous in H_{RV} for $a \le t \le b$. Specifically, given $\epsilon > 0$, there is a $\delta > 0$ such that $\|f(t_1) - f(t_2)\|_{RV} < \epsilon$ when $|t_2 - t_1| < \delta$. In addition, $I(f)(t) = \int_a^t f(s)\, dW(s)$ is also continuous in H_{RV} for $a \le t \le b$ as

$$\begin{aligned}
\|I(f)(t_2) - I(f)(t_1)\|_{RV}^2 &= E\left(\int_{t_1}^{t_2} f(s)\, dW(s)\right)^2 = \int_{t_1}^{t_2} E|f(t)|^2\, dt \\
&\le 2|t_2 - t_1|\|f(t_1)\|_{RV}^2 + k_2(t_2 - t_1)^2.
\end{aligned}$$

3.4 Approximation of Stochastic Integrals

It is useful for conceptual as well as practical purposes to understand how stochastic integrals can be approximated. Consider first an approximation of $J(f) = \int_a^b f(t)\, dt$, where $f \in H_{SP}$ satisfies conditions (c1)–(c3). Let $a = t_0 < t_1 < \cdots < t_N = b$ be a partition of $[a, b]$, where $t_i = a + i\Delta t$ for $i = 0, 1, \ldots, N$ and $\Delta t = (b - a)/N$. Let $J_N(f)$ be an approximation to $J(f)$, where

$$J(f) \approx J_N(f) = \sum_{i=0}^{N-1} f(t_i)\Delta t.$$

Then

$$\begin{aligned}
\|J(f) - J_N(f)\|_{RV}^2 &= E\left|\int_a^b f(t)\,dt - \sum_{i=0}^{N-1} f(t_i)\Delta t\right|^2 \\
&= E\left|\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \big(f(t) - f(t_i)\big)\,dt\right|^2 \\
&\le (b-a)\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|f(t) - f(t_i)|^2\,dt \\
&\le k_2(b-a)\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} (t - t_i)\,dt = \frac{(b-a)^3 k_2}{2N}
\end{aligned}$$

implying that $\|J_N(f) - J(f)\|_{RV} = O(1/\sqrt{N})$.

Now consider approximation of $I(f) = \int_a^b f(t)\,dW(t)$ for $f \in H_{SP}$ where f satisfies conditions (c1)-(c3). Again let $a = t_0 < t_1 < \cdots < t_N = b$ be a partition of $[a, b]$, where $t_i = a + i\Delta t$ for $i = 0, 1, \ldots, N$. Let $I_N(f)$ be an approximation to $I(f)$, where

$$I(f) \approx I_N(f) = \sum_{i=0}^{N-1} f(t_i)\,\eta_i \quad \text{and} \quad \eta_i = W(t_{i+1}) - W(t_i).$$

In calculating the error in this approximation,

$$\begin{aligned}
\|I(f) - I_N(f)\|_{RV}^2 &= E\left|\int_a^b f(t)\,dW(t) - \sum_{i=0}^{N-1} f(t_i)\,\eta_i\right|^2 \\
&= E\left|\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \big(f(t) - f(t_i)\big)\,dW(t)\right|^2 \\
&= \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} E|f(t) - f(t_i)|^2\,dt \\
&\le k_2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} (t - t_i)\,dt \;=\; \frac{(b-a)^2 k_2}{2N}.
\end{aligned}$$

The approximation for this integral is also $O(1/\sqrt{N})$ in the H_{RV} norm.

Notice that in either approximation, $J_N(f)$ or $I_N(f)$, the respective integral is estimated for only one sample path. To approximate expectations involving stochastic integrals, many sample paths are needed. (Recall Example 1.31.) For example, assuming that the expectation of a function g of $I(f)$ is to be estimated, then

$$E\big(g(I(f))\big) \approx \frac{1}{M}\sum_{j=1}^{M} g(I_N^{(j)}(f)),$$

where $I_N^{(j)}(f)$ is the jth sample-path approximation for $j = 1, 2, \ldots, M$. Consider the following example.

Example 3.11. Approximation of an Itô integral

In this example, the stochastic integral $\int_0^1 t\,W(t)\,dW(t)$ is considered. The expected value of the integral and the expected value of the square of the integral are estimated using $M = 1,000,000$ sample paths. The number of intervals used on $[0,1]$ is N where $N = 4, 8, 16, \ldots, 256$. The preceding approximate method is used, which for this problem, has the explicit form

$$E\big((I(f))^\alpha\big) = E\left(\int_0^1 t\,W(t)\,dW(t)\right)^\alpha$$
$$\approx \frac{1}{M}\sum_{j=1}^{M}\left(\sum_{i=0}^{N-1} t_i^{(N)} W^{(j)}(t_i^{(N)})\big(W^{(j)}(t_{i+1}^{(N)}) - W^{(j)}(t_i^{(N)})\big)\right)^\alpha$$

for $\alpha = 1$ or 2 and $t_i^{(N)} = i/N$ for $i = 0, 1, 2, \ldots, N$. Notice that $(W^{(j)}(t_{i+1}^{(N)}) - W^{(j)}(t_i^{(N)})) = \eta_i^{(j)}/\sqrt{N}$ where $\eta_i^{(j)} \sim N(0,1)$ and also that $W^{(j)}(t_i^{(N)}) = W^{(j)}(t_{i-1}^{(N)}) + \eta_{i-1}^{(j)}/\sqrt{N}$. A computer program that performs this calculation is listed at the end of this chapter. The calculational results are given in Table 3.2 for $M = 1,000,000$ sample paths. Notice that the results improve as N increases. The exact values are $E(I(f)) = 0$ and $E\big((I(f))^2\big) = 0.25$.

Table 3.2. Calculated estimates of $|E(I(f))|$ and $E((I(f))^2)$ for Example 3.11

Value of N	$\lvert E(I(f))\rvert$	$E((I(f))^2)$
2^2	0.00107	0.14071
2^3	0.00012	0.19151
2^4	0.00069	0.21906
2^5	0.00038	0.23508
2^6	0.00007	0.24209
2^7	0.00001	0.24486
2^8	0.00002	0.24821

3.5 Stochastic Differentials and Itô's Formula

In this section, the following stochastic process is considered:

$$X(t) = X(a) + \int_a^t f(s)\,ds + \int_a^t g(s)\,dW(s) \quad \text{for} \quad a \le t \le b, \tag{3.1}$$

where $f, g \in H_{SP}$, $X(a) \in H_{RV}$, and f and g satisfy conditions (c1)–(c3). First, note that as $\int_a^t f(s)\,ds \in H_{SP}$ and $\int_a^t g(s)\,dW(s) \in H_{SP}$, then $X \in H_{SP}$. Second, notice that if f (or g) satisfies conditions (c1)–(c3), then

$$\|f(t)\|_{RV} \le \|f(t) - f(a)\|_{RV} + \|f(a)\|_{RV} \le k_2^{1/2}(b-a)^{1/2} + \|f(a)\|_{RV}$$

for $a \le t \le b$. Therefore,

$$E|f^2(t)| \le 2k_2(b-a) + 2\|f(a)\|_{RV}^2 \quad \text{for any} \quad t \in [a,b].$$

Finally, it is useful to notice that $X(t)$ is a continuous stochastic process in H_{RV} for $a \le t \le b$. That is, given $\epsilon > 0$, there is a $\delta > 0$ such that $\|X(t_1) - X(t_2)\|_{RV} < \epsilon$ when $|t_1 - t_2| < \delta$ for $t_1, t_2 \in [a,b]$. To see this, consider

$$\begin{aligned}
&\|X(t_1) - X(t_2)\|_{RV}^2 = E\left|\int_{t_1}^{t_2} f(s)\,ds + \int_{t_1}^{t_2} g(s)\,dW(s)\right|^2 \\
&\le 2E\left|\int_{t_1}^{t_2} f(s)\,ds\right|^2 + 2E\left|\int_{t_1}^{t_2} (g(s) - g(t_1) + g(t_1))\,dW(s)\right|^2 \\
&\le 2(t_2 - t_1)\int_{t_1}^{t_2} E|f(s)|^2\,ds + 4\int_{t_1}^{t_2} E|g(s) - g(t_1)|^2\,ds + 4\int_{t_1}^{t_2} E|g(t_1)|^2\,ds \\
&\le 2(t_2 - t_1)\|f\|_{SP}^2 + 2k_2(t_2 - t_1)^2 + 4(t_2 - t_1)E|g(t_1)|^2 \\
&\le 2(t_2 - t_1)\Big(\|f\|_{SP}^2 + 5k_2(b-a) + 4\|g(a)\|_{RV}^2\Big).
\end{aligned}$$

Furthermore, notice that the preceding inequality implies that $P(|X(t_1) - X(t_2)| \ge \epsilon) \to 0$ as $|t_2 - t_1| \to 0$. That is, X has continuous sample paths on $[a,b]$ in probability.

If X satisfies (3.1), it is said that X satisfies the stochastic differential

$$dX(t) = f(t)\,dt + g(t)\,dW(t) \quad \text{for} \quad a < t < b. \tag{3.2}$$

An important result is Itô's formula which says that a smooth function, $F(t, X(t))$, of the stochastic process $X(t)$ also satisfies a stochastic differential. The following two conditions on a function $G : [a,b] \times \mathbb{R} \to \mathbb{R}$ are useful in stating and proving this theorem.

Condition (c4): For function $G : [a,b] \times \mathbb{R} \to \mathbb{R}$, there exists a nonnegative constant k_3 such that for any $t_1, t_2 \in [a,b]$ and any $X \in H_{SP}$ then $E|G(t_2, X(t_2)) - G(t_1, X(t_1))|^2 \le k_3\Big(|t_2 - t_1| + E|X(t_2) - X(t_1)|^2\Big)$.

Condition (c5): For function $G : [a,b] \times \mathbb{R} \to \mathbb{R}$, then $G(a, X(a)) \in H_{RV}$ if $X(a) \in H_{RV}$.

Theorem 3.12. (Itô's Formula)
Let $X \in H_{SP}$ satisfy (3.1) for $t \in [a,b]$, where f and g satisfy conditions (c1)–(c3) and

$$\|f^2(t)\|_{RV},\ \|g^2(t)\|_{RV} \le k_4 \quad \textit{for} \quad t \in [a,b].$$

Let F be a function of t and x. Assume that $F(t,x)$ has the continuous derivatives

$$\frac{\partial F(t,x)}{\partial t}, \frac{\partial F(t,x)}{\partial x}, \frac{\partial^2 F(t,x)}{\partial x^2}, \frac{\partial^2 F(t,x)}{\partial t^2}, \frac{\partial^2 F(t,x)}{\partial x \partial t} \quad \textit{for} \quad t \in [a,b] \quad \textit{and} \quad x \in \mathbb{R}$$

and that F and these derivatives satisfy conditions (c4) and (c5). Suppose also that the functions

$$f(t)\frac{\partial F(t,x)}{\partial x}, \quad \frac{1}{2}g^2(t)\frac{\partial^2 F(t,x)}{\partial x^2}, \quad g(t)\frac{\partial F(t,x)}{\partial x}$$

satisfy conditions (c4) and (c5). Let

$$\tilde{f}(t,x) = \frac{\partial F(t,x)}{\partial t} + f(t)\frac{\partial F(t,x)}{\partial x} + \frac{1}{2}g^2(t)\frac{\partial^2 F(t,x)}{\partial x^2}$$

and

$$\tilde{g}(t,x) = g(t)\frac{\partial F(t,x)}{\partial x}.$$

Then, F satisfies the stochastic differential

$$dF(t,X(t)) = \tilde{f}(t,X(t))dt + \tilde{g}(t,X(t))dW(t). \tag{3.3}$$

Proof. Let $\hat{f}(t) \equiv \tilde{f}(t,X(t))$ and $\hat{g}(t) \equiv \tilde{g}(t,X(t))$ where $X(t)$ satisfies the stochastic differential (3.1). It is readily seen that $\hat{f}$ and $\hat{g}$ satisfy conditions (c1)–(c3). Therefore, let $Y(t)$ be the stochastic process in H_{SP} that satisfies the stochastic differential

$$dY(t) = \hat{f}(t)\,dt + \hat{g}(t)\,dW(t) \quad \text{with} \quad Y(a) = F(a,X(a)).$$

Hence,

$$Y(t) = F(a,X(a)) + \int_a^t \hat{f}(s)\,ds + \int_a^t \hat{g}(s)\,dW(s) \quad \text{for} \quad a \le t \le b.$$

Now notice that if a function G satisfies condition (c4) and X satisfies (3.1), then there is a constant $k_5 > 0$ such that

$$E|G(t_2,X(t_2)) - G(t_1,X(t_1))|^2 \le k_5|t_2 - t_1| \quad \text{for any} \quad t_1, t_2 \in [a,b].$$

Also, useful is the result that if G satisfies conditions (c4)–(c5) and X satisfies (3.1), then there is a constant $k_6 > 0$ such that

$$\|G(t, X(t))\|_{RV} \le k_6 \quad \text{for any} \quad t \in [a, b].$$

Fix $t \in [a, b]$ and let $t_k = a + k\Delta t$, where $\Delta t = (t - a)/m$, and $\Delta W_k = W(t_{k+1}) - W(t_k)$ for $k = 0, 1, \ldots, m$. By Taylor's Theorem, there is a number $\theta_k, 0 \le \theta_k \le 1$, such that

$$\begin{aligned}
F(t_{k+1}, X(t_{k+1})) &= F(t_k, X(t_k)) + \frac{\partial F(t_k, X(t_k))}{\partial t}\Delta t + \frac{\partial F(t_k, X(t_k))}{\partial x}\Delta X_k \\
&+ \frac{1}{2}g^2(t_k)\frac{\partial^2 F(t_k, X(t_k))}{\partial x^2}\Delta t + \frac{1}{2}\frac{\partial^2 F(t_k + \theta_k\Delta t, X(t_k) + \theta_k\Delta X_k)}{\partial t^2}(\Delta t)^2 \\
&+ \frac{\partial^2 F(t_k + \theta_k\Delta t, X(t_k) + \theta_k\Delta X_k)}{\partial x \partial t}\Delta t\Delta X_k \\
&+ \frac{1}{2}(\Delta X_k)^2\left(\frac{\partial^2 F(t_k + \theta_k\Delta t, X(t_k) + \theta_k\Delta X_k)}{\partial x^2} - \frac{\partial^2 F(t_k, X(t_k))}{\partial x^2}\right) \\
&+ \frac{1}{2}\frac{\partial^2 F(t_k, X(t_k))}{\partial x^2}\left((\Delta X_k)^2 - g^2(t_k)\Delta t\right),
\end{aligned}$$

where

$$\begin{aligned}
\Delta X_k &= X(t_{k+1}) - X(t_k) = f(t_k)\Delta t + g(t_k)\Delta W_k + \epsilon_k, \\
\epsilon_k &= \int_{t_k}^{t_{k+1}} (f(t) - f(t_k))\, dt + \int_{t_k}^{t_{k+1}} (g(t) - g(t_k))\, dW(t),
\end{aligned}$$

and by condition (c2),

$$\|\epsilon_k\|_{RV}^2 \le k_2\big((\Delta t)^3 + (\Delta t)^2\big).$$

Summing the Taylor expression from $k = 0, 1, \ldots, m - 1$ yields

$$\begin{aligned}
F(t, X(t)) - F(a, X(a)) - \int_a^t \hat{f}(s)\, ds - \int_a^t \hat{g}(s)\, dW(s) \\
= E_1^{(m)}(t) + E_2^{(m)}(t) + \cdots + E_{10}^{(m)}(t),
\end{aligned}$$

where

$$\begin{aligned}
E_1^{(m)}(t) &= \sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} \left(\frac{\partial F(t_k, X(t_k))}{\partial t} - \frac{\partial F(t, X(t))}{\partial t}\right) dt \\
E_2^{(m)}(t) &= \sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} \left(\frac{\partial F(t_k, X(t_k))}{\partial x} f(t_k) - \frac{\partial F(t, X(t))}{\partial x} f(t)\right) dt \\
E_3^{(m)}(t) &= \frac{1}{2}\sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} \left(\frac{\partial^2 F(t_k, X(t_k))}{\partial x^2} g^2(t_k)\left(\frac{(\Delta W_k)^2}{\Delta t} - 1\right)\right) dt \\
E_4^{(m)}(t) &= \frac{1}{2}\sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} \left(\frac{\partial^2 F(t_k, X(t_k))}{\partial x^2} g^2(t_k) - \frac{\partial^2 F(t, X(t))}{\partial x^2} g^2(t)\right) dt
\end{aligned}$$

$$E_5^{(m)}(t) = \sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} \left(\frac{\partial F(t_k, X(t_k))}{\partial x} g(t_k) - \frac{\partial F(t, X(t))}{\partial x} g(t) \right) dW(t)$$

$$E_6^{(m)}(t) = \frac{1}{2} \sum_{k=0}^{m-1} \frac{\partial^2 F(t_k, X(t_k))}{\partial x^2} \left((\Delta X_k)^2 - g^2(t_k)(\Delta W_k)^2 \right)$$

$$E_7^{(m)}(t) = \frac{1}{2} \sum_{k=0}^{m-1} \frac{\partial^2 F(t_k + \theta_k \Delta t, X(t_k) + \theta_k \Delta X_k)}{\partial t^2} (\Delta t)^2$$

$$E_8^{(m)}(t) = \sum_{k=0}^{m-1} \frac{\partial^2 F(t_k + \theta_k \Delta t, X(t_k) + \theta_k \Delta X_k)}{\partial t \partial x} \Delta t \Delta X_k$$

$$E_9^{(m)}(t) = \frac{1}{2} \sum_{k=0}^{m-1} \frac{\partial^2 \big(F(t_k + \theta_k \Delta t, X(t_k) + \theta_k \Delta X_k) - F(t_k, X(t_k)) \big)}{\partial x^2} (\Delta X_k)^2$$

$$E_{10}^{(m)}(t) = \sum_{k=0}^{m-1} \frac{\partial F(t_k, X(t_k))}{\partial x} \epsilon_k.$$

Subtracting $Y(t)$ from the expression for $F(t, X(t))$ and using the triangle inequality, it follows that

$$\|Y(t) - F(t, X(t))\|_{RV} \ \le \ \sum_{i=1}^{10} \|E_i^{(m)}(t)\|_{RV} \quad \text{for any} \quad t \in [a, b].$$

But $\|E_i^{(m)}(t)\|_{RV} \to 0$ as $m \to \infty$ for each i, $1 \le i \le 10$ and for each $t \in [a, b]$. Thus, F satisfies the differential (3.3).

To see that $\|E_i^{(m)}(t)\|_{RV} \to 0$ as $m \to \infty$ for each $1 \le i \le 10$, the terms $E_1^{(m)}$ and $E_3^{(m)}$ are considered here in more detail. The other terms can be be analyzed in a similar manner. For $E_1^{(m)}$, applying the Cauchy-Schwarz inequality and condition (c3),

$$\|E_1^{(m)}(t)\|_{RV}^2 = E \left| \sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} \left(\frac{\partial F(t_k, X(t_k))}{\partial t} - \frac{\partial F(t, X(t))}{\partial t} \right) dt \right|^2$$

$$\le E \left| \sum_{k=0}^{m-1} (\Delta t)^{1/2} \left(\int_{t_k}^{t_{k+1}} \left(\frac{\partial F(t_k, X(t_k))}{\partial t} - \frac{\partial F(t, X(t))}{\partial t} \right)^2 dt \right)^{1/2} \right|^2$$

$$\le (b - a) \sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} E \left| \frac{\partial F(t_k, X(t_k))}{\partial t} - \frac{\partial F(t, X(t))}{\partial t} \right|^2 dt$$

$$\le k_5 (b - a) \sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} (t - t_k)\, dt \ \le \ k_5 (b - a)^2 / (2m)$$

and so $\|E_1^{(m)}(t)\|_{RV} \to 0$ as $m \to \infty$.

Now consider the term $E_3^{(m)}(t)$. Applying the Cauchy-Schwarz inequality, using the independence of $\dfrac{\partial^2 F(t_k, X(t_k))}{\partial x^2} g^2(t_k)$ and $\left(\dfrac{(\Delta W_k)^2}{\Delta t} - 1\right)$, and the fact that $E\left(\dfrac{(\Delta W_k)^2}{\Delta t} - 1\right) = 0$, it follows that

$$\begin{aligned}
\|E_3^{(m)}(t)\|_{RV}^2 &= \frac{1}{4} E \left| \sum_{k=0}^{m-1} \Delta t \left(\frac{\partial^2 F(t_k, X(t_k))}{\partial x^2} g^2(t_k) \right) \left(\frac{(\Delta W_k)^2}{\Delta t} - 1 \right) \right|^2 \\
&= \frac{1}{4} \sum_{k=0}^{m-1} (\Delta t)^2 E \left| \frac{\partial^2 F(t_k, X(t_k))}{\partial x^2} g^2(t_k) \right|^2 E \left| \frac{(\Delta W_k)^2}{\Delta t} - 1 \right|^2 \\
&\leq \frac{1}{2} k_6^2 \sum_{k=0}^{m-1} (\Delta t)^2 = \frac{k_6^2 (b-a)^2}{2m},
\end{aligned}$$

where boundedness of processes satisfying conditions (c3) and (c4) was also used. Thus, $\|E_3^{(m)}(t)\|_{RV} \to 0$ as $m \to \infty$.

Several examples are given here that illustrate the usefulness of Itô's formula. Additional applications of Itô's formula are discussed in the next chapter.

Example 3.13. Simplifying an Itô integral

Consider the stochastic integral $\int_0^t s\, dW(s)$. Let $dX(t) = dW(t)$ so $g = 1$ and $f = 0$. Let $F(t, x) = tx$. Using Itô's formula,

$$\int_0^t d(sW(s)) = \int_0^t W(s)\, ds + \int_0^t s\, dW(s).$$

Thus,

$$\int_0^t s\, dW(s) = -\int_0^t W(s)\, ds + tW(t).$$

Example 3.14. Evaluating an Itô integral

Consider the stochastic integral $\int_0^t W(s)\, dW(s)$. Let $dX(t) = dW(t)$ so $g = 1$ and $f = 0$. Let $F(t, x) = \frac{1}{2}x^2$. By Itô's formula,

$$\int_0^t d\left(\frac{1}{2}W^2(s)\right) = \int_0^t \frac{1}{2}\, ds + \int_0^t W(s)\, dW(s).$$

Thus,

$$\int_0^t W(s)\, dW(s) = \frac{1}{2}(W^2(t) - W^2(0)) - \frac{t}{2}.$$

Example 3.15. Simplifying a double stochastic integral

Let $dX(t) = W(t)dt$ and $X(0) = 0$. Then $f = W$ and $g = 0$. Let $F(t, x) = tx$. Then by Itô's formula,

$$d(tX(t)) = (X(t) + tW(t))\,dt.$$

Therefore,

$$tX(t) = \int_0^t X(s)\,ds + \int_0^t sW(s)\,ds.$$

But $X(t) = \int_0^t W(s)\,ds$. Therefore, using Example 3.13,

$$\begin{aligned}\int_0^t \int_0^s W(z)\,dz\,ds &= t\int_0^t W(s)\,ds - \int_0^t sW(s)\,ds \\ &= (t+1)\int_0^t W(s)\,ds - tW(t).\end{aligned}$$

3.6 Stratonovich Stochastic Integrals

Recall that the Itô stochastic integral is defined as

$$\int_a^b f(t,\omega)\,dW(t,\omega) = \lim_{m\to\infty} \sum_{i=0}^{m-1} f(t_i^{(m)},\omega)\big(W(t_{i+1}^{(m)},\omega) - W(t_i^{(m)},\omega)\big),$$

where $t_i^{(m)} = a + (\frac{b-a}{m})j$ for $j = 0, 1, 2, \ldots, m$ and the sum converges in H_{RV}. Also, as $f(t_i^{(m)})$ and $\Delta W_i^{(m)} = (W(t_{i+1}^{(m)}) - W(t_i^{(m)}))$ are independent, some useful properties are obtained such as

$$E\left|\int_a^b f(t)\,dW(t)\right|^2 = \int_a^b E|f(t)|^2\,dt.$$

However, it is important to be aware that there are alternate ways of defining stochastic integrals. Besides the Itô stochastic integral, the most common stochastic integral is the Stratonovich stochastic integral. The Stratonovich integral is defined as

$$\int_a^b f(t) \circ dW(t) = \lim_{m\to\infty} \sum_{i=0}^{m-1} \frac{1}{2}\big(f(t_i^{(m)}) + f(t_{i+1}^{(m)})\big)\big(W(t_{i+1}^{(m)}) - W(t_i^{(m)})\big),$$

where the symbol "$\circ$" is used to denote this integral as the Stratonovich integral. In the summation, $\frac{1}{2}\big(f(t_i^{(m)}) + f(t_{i+1}^{(m)})\big)$ and $\big(W(t_{i+1}^{(m)}) - W(t_i^{(m)})\big)$ are not likely to be independent. Consequently, Itô and Stratonovich integrals generally have different values.

Consider, for example, the integrals

$$\int_0^t W(s)\,dW(s) \quad \text{and} \quad \int_0^t W(s) \circ dW(s).$$

By Example 3.14, we know that the value of the Itô integral is

$$\int_0^t W(s)\,dW(s) = \frac{1}{2}(W^2(t) - W^2(0)) - \frac{t}{2}.$$

Now consider the Stratonovich integral. By the above definition of the Stratonovich integral,

$$\begin{aligned}
\int_0^t W(s) \circ dW(s) &= \lim_{m\to\infty} \sum_{i=0}^{m-1} \frac{1}{2}\left(W(t_{i+1}^{(m)}) + W(t_i^{(m)})\right)\left(W(t_{i+1}^{(m)}) - W(t_i^{(m)})\right) \\
&= \lim_{m\to\infty} \sum_{i=0}^{m-1} \frac{1}{2}\left(W^2(t_{i+1}^{(m)}) - W^2(t_i^{(m)})\right) \\
&= \lim_{m\to\infty} \frac{1}{2}\left(W^2(t) - W^2(0)\right) \\
&= \frac{1}{2}\left(W^2(t) - W^2(0)\right).
\end{aligned}$$

Hence, for this stochastic integral,

$$\int_0^t W(s)\,dW(s) - \int_0^t W(s) \circ dW(s) = -\frac{t}{2}.$$

Throughout Chapters 4 and 5, Itô stochastic differential equations are considered. Itô stochastic differential equations involve Itô integrals. The modeling procedure described in Chapter 5 produces, in a natural manner, an Itô stochastic differential equation model. However, Stratonovich stochastic differential equations can be defined by using Stratonovich integrals rather than Itô integrals. It is useful to be aware that Itô and Stratonovich stochastic differential equations are closely related. In particular, if $X(t)$ satisfies the Stratonovich stochastic differential equation

$$X(t) = X(0) + \int_0^t f(s, X(s)))\,ds + \int_0^t g(s, X(s)) \circ dW(s),$$

then $X(t)$ satisfies the Itô stochastic differential equation

$$\begin{aligned}
X(t) = X(0) &+ \int_0^t \left(f(s, X(s)) + \frac{1}{2}\frac{\partial g(s, X(s))}{\partial x} g(s, X(s))\right) ds \\
&+ \int_0^t g(s, X(s))\,dW(s).
\end{aligned}$$

In addition, this relationship can be extended [20, 41, 69] to systems of Stratonovich and Itô stochastic differential equations.

3.7 Multidimensional Itô's Formula

Considered in this section is the multidimensional stochastic differential

$$d\mathbf{X}(t,\omega) = \mathbf{f}(t,\omega)dt + G(t,\omega)d\mathbf{W}(t,\omega) \quad \text{for} \quad a < t < b, \tag{3.4}$$

where

$$\mathbf{X}(t,\omega) = [X_1(t,\omega), X_2(t,\omega), \ldots, X_n(t,\omega)]^T,$$

$$\mathbf{f}(t,\omega) = [f_1(t,\omega), f_2(t,\omega), \ldots, f_n(t,\omega)]^T,$$

$$\mathbf{W}(t,\omega) = [W_1(t,\omega), W_2(t,\omega), \ldots, W_m(t,\omega)]^T, \quad \text{and}$$

$$(G(t,\omega))_{ij} = g_{ij}(t,\omega) \quad \text{where} \quad G(t,\omega) \quad \text{is an} \quad n \times m \quad \text{matrix.}$$

In particular, notice that $\mathbf{W}(t)$ is an m-dimensional Wiener process where the elements $W_i(t)$ and $W_j(t)$ are independent if $i \neq j$. Equation (3.4) can be put in componentwise form as

$$dX_i(t,\omega) = f_i(t,\omega)dt + \sum_{j=1}^{m} g_{ij}(t,\omega)dW_j(t,\omega) \tag{3.5}$$

for $i = 1, 2, \ldots, n$, and, hence,

$$X_i(t,\omega) = X_i(a,\omega) + \int_a^t f_i(s,\omega)\, ds + \int_a^t \sum_{j=1}^{m} g_{ij}(s,\omega)\, dW_j(s,\omega) \tag{3.6}$$

for $i = 1, 2, \ldots, n$.

Now let $F(t, \mathbf{X})$ be a smooth function of t and $\mathbf{X}$. That is, $F : [a,b] \times H_{SP}^n \to \mathbb{R}$. Then, Itô's formula can be generalized [41, 69, 70] in this multidimensional setting to yield the stochastic differential for F of the form

$$\begin{aligned} dF(t,\mathbf{X}) = &\left(\frac{\partial F}{\partial t} + \sum_{i=1}^{n} \frac{\partial F}{\partial x_i} f_i + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} \frac{1}{2} \frac{\partial^2 F}{\partial x_i \partial x_j} g_{ik} g_{jk} \right) dt \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\partial F}{\partial x_i} g_{ij}\, dW_j(t). \end{aligned} \tag{3.7}$$

Consider the following example where $n = 1$ and $m = 2$.

Example 3.16. Itô's formula applied to a multidimensional problem

Let $dX_1(t) = a\, dW_1(t) + b\, dW_2(t)$ with $X_1(0) = 0$ and where a and b are nonzero constants. Let $F(t, X_1) = X_1^2$. Then, using Itô's formula with $n = 1, m = 2, f_1 = 0, g_{11} = a$, and $g_{12} = b$,

$$d(X_1^2(t)) = (a^2 + b^2)\, dt + 2aX_1(t)\, dW_1(t) + 2bX_1(t)\, dW_2(t)$$

with $X_1^2(0) = 0$. Thus, a stochastic differential for X_1^2 is readily derived by applying Itô's formula. By continuing this example, using some of the results obtained earlier, an interesting formula can be obtained. Hence, noting that $X_1(t) = a(W_1(t) - W_1(0)) + b(W_2(t) - W_2(0))$ and substituting this identity into the integral $\int_0^t dX_1^2(s) = X_1^2(t) - X_1^2(0)$ leads to

$$\begin{aligned}\big(a(W_1(t) - W_1(0)) + b(W_2(t) - W_2(0))\big)^2 &= (a^2 + b^2)t \\ &+ 2a \int_0^t [a(W_1(s) - W_1(0)) + b(W_2(s) - W_2(0))]\, dW_1(s) \\ &+ 2b \int_0^t [a(W_1(s) - W_1(0)) + b(W_2(s) - W_2(0))]\, dW_2(s).\end{aligned}$$

Now using the relation $\int_0^t W(s)\, dW(s) = \frac{1}{2}(W^2(t) - W^2(0)) - \dfrac{t}{2}$, which was derived in Example 3.14, the following result is obtained that

$$2\big(W_1(t)W_2(t) - W_1(0)W_2(0)\big) = \int_0^t W_1(s)\, dW_2(s) + \int_0^t W_2(s)\, dW_1(s).$$

Exercises

3.1. Consider the stochastic integral $F_N = \int_0^T f_N(t)\, dt$ where f_N is the stochastic process defined on $[0, T]$ by the formula

$$f_N(t) = \frac{(t_{i+1} - t)U_i}{h} + \frac{(t - t_i)U_{i+1}}{h} \quad \text{for} \quad t_i \le t \le t_{i+1}$$

for $i = 0, 1, \ldots, N-1$ where $h = T/N$, $t_i = ih$, and $U_i \sim U[-\frac{1}{2}, \frac{1}{2}]$ are independent and identically distributed random variables for $i = 0, 1, \ldots, N$.
(a) Show that $F_N = \dfrac{T}{2N} \sum_{i=0}^{N-1} (U_i + U_{i+1})$.
(b) Calculate $E(F_N^2)$.
(c) Find an $F \in H_{RV}$ such that $\|F_N - F\|_{RV} \to 0$ as $N \to \infty$.

3.2. Prove that $E(e^{-W(t)}) = e^{t/2}$. (Apply the Taylor expansion $e^{-W(t)} = \sum_{j=0}^{\infty} (-W(t))^j / j!$ and use the formulas $E(W(t))^{2k} = (1 \cdot 3 \cdot 5 \cdots (2k-1))t^k$ and $E(W(t))^{2k+1} = 0$.)

3.3. Let $I(f) = \int_0^2 \exp(-W(t)/2)\, dW(t)$. Find $E(I(f))$ and $E((I(f))^2)$. (Use the relation $E(\exp(-W(t)) = \exp(t/2)$.)

3.4. Approximate $E((I(f))^2)$ in Exercise 3.3 using the relation $I(f) \approx I_N(f) = \sum_{i=0}^{N-1} f(t_i) \Delta W_i$ where $t_i = 2i/N$. That is, use

$$E((I(f))^2) \approx \frac{1}{M}\sum_{j=1}^{M}(I_{N,j}(f))^2 = \frac{1}{M}\sum_{j=1}^{M}\left(\sum_{i=0}^{N-1} f(t_i,\omega_j)\Delta W_{i,j}\right)^2,$$

where $\Delta W_{i,j} = W(t_{i+1},\omega_j) - W(t_i,\omega_j)$. Use $N = 200$ and $M = 20000$. Modify the MATLAB program listed below to perform the calculations. The program approximates expectations of $\int_0^2 f(W(t))\,dt$ where the function f is specified in the program. Hand in your calculational results along with a listing of your program.

```
% Expectations of a stochastic integral are approximated.
% Matrix array operations are used for efficiency.
% nvec paths are simultaneously computed.
   clear
   randn('state',3)
   for icase=1:5
   nvec=10000;
   nint=200;
   time=2;
   h=time/nint;
   hs=sqrt(h);
   w=zeros(nvec,1);
   y=zeros(nvec,1);
   for i=1:nint
      r=randn(nvec,1);
      tt=(i-1)*h;
      f=w.*cos(w);
      y=y + f.*r*hs;
      w=w+hs*r;
   end
   mean=sum(y)/nvec;
   meansqr=sum(y.*y)/nvec;
   check=sum(exp(-w))/nvec;
   disp((sprintf(' \%12.0f',icase)))
   disp((sprintf(' \%12.5f \%12.5f',mean,meansqr)))
   end
```

3.5. Consider the Stratonovich stochastic integral

$$\int_a^b f(t)\circ dW(t) = \lim_{m\to\infty}\sum_{i=0}^{m-1}\frac{1}{2}(f(t_i^{(m)}) + f(t_{i+1}^{(m)}))(W(t_{i+1}^{(m)}) - W(t_i^{(m)})),$$

where $t_i^{(m)} = a + i(b-a)/m$ for $i = 0, 1, 2, \ldots, m$. Find $E\left(\int_a^b f(t)\circ dW(t)\right)$ and $E\left(\int_a^b f(t)\circ dW(t)\right)^2$ when $f(t) = 3tW(t)$. Compare these values with the values obtained for the corresponding Itô stochastic integrals.

3.6. Suppose that X satisfies

$$X(t) = 2 + \int_0^t (3s + \exp(s^2))\, ds + \int_0^t \cos(s)\, dW(s).$$

Let $Y(t) = F(t, X(t)) = (2t+3)X(t) + 4t^2$. Find $Y(0), \hat{f}(s)$ and $\hat{g}(s)$ such that

$$Y(t) = Y(0) + \int_0^t \hat{f}(s)\, ds + \int_0^t \hat{g}(s)\, dW(s).$$

3.7. (a) Use Itô's formula to show that

$$e^{W(t)} - 1 = \int_0^t \frac{1}{2}\, e^{W(s)}\, ds + \int_0^t e^{W(s)}\, dW(s).$$

(b) Let $r(t) = E(e^{W(t)})$ and use part (a) to show that $r(t)$ satisfies the initial-value problem $\dfrac{dr(t)}{dt} = \frac{1}{2} r(t)$ with $r(0) = 1$. Thus, $r(t) = E(e^{W(t)}) = e^{t/2}$.

3.8. Suppose that $f \in C^\infty\big([a,b] \times \mathbb{R}\big)$. Use a Taylor expansion in two variables and the definition of the Stratonovich integral to prove that

$$E\left[\int_a^b f(t, W(t)) \circ dW(t)\right] = \frac{1}{2} \int_a^b E\left(\frac{\partial f(t,x)}{\partial x}\Big|_{x=W(t)}\right) dt.$$

3.9. Let $Y_0(t) = 1$ for $t \in [0,1]$ and define $\{Y_n\}_{n=1}^\infty$ by the recurrence formula

$$Y_n(t) = 1 + \frac{1}{2} \int_0^t Y_{n-1}(s)\, dW(s) \quad \text{for} \quad 0 \le t \le 1 \quad \text{and} \quad n = 1, 2, \ldots.$$

(a) Prove that $\{Y_n\}_{n=1}^\infty$ is a Cauchy sequence in H_{SP}. Thus, there exists a $Y \in H_{SP}$ such that $Y_n \to Y$ as $n \to \infty$.
(b) Use the equation $Y(t) = 1 + \frac{1}{2} \int_0^t Y(s)\, dW(s)$ to show that $E(Y(t)) = 1$ and $E(Y^2(t)) = e^{t/4}$.

3.10. Consider the stochastic differential $dX(t) = -\frac{1}{8}\, dt + \frac{1}{2}\, dW(t)$, with $X(0) = 0$. Thus, $X(t) = -\frac{1}{8}\, t + \frac{1}{2}\, W(t)$. Let $Z(t) = e^{X(t)}$. Find the stochastic differential for $Z(t)$ using Itô's formula. Compare your result with the differential satisfied by $Y(t)$ in Exercise 3.9. Explain what this means about $Y(t)$ in Exercise 3.9.

3.11. Approximate $E((I(f))^2) = E\left(\left(\int_0^2 e^{W(t)/2}\, dW(t)\right)^2\right)$ using $I(f) \approx I_N(f) = \sum_{i=0}^{N-1} f(t_i) \Delta W_i$ where $t_i = 2i/N$. That is, use

$$E((I(f))^2) \approx \frac{1}{M}\sum_{j=1}^{M}(I_{N,j}(f))^2 = \frac{1}{M}\sum_{j=1}^{M}\left(\sum_{i=0}^{N-1} f(t_i,\omega_j)\Delta W_{i,j}\right)^2$$

where $\Delta W_{i,j} = W(t_{i+1},\omega_j) - W(t_i,\omega_j)$. Use $N = 4, 32, 256$ and $M = 500, 5000, 50000$ for a total of nine calculations. Modify the MATLAB program listed below to perform the calculations. In the calculational results, there are statistical errors and method errors. Explain your calculational results in terms of these two kinds of errors. Hand in your explanation, calculational results, and a listing of your program.

```
\% A stochastic integral is computed.
\% Matrix array operations are used for efficiency.
\% nvec paths are simultaneously computed.
clear
 randn('state',6)
 nvec=50;
 for i1=1:3
 nvec=nvec*10;
 nint=1/2;
 for i2=1:3
 nint=8*nint;
 time=2;
 h=time/nint;
 hs=sqrt(h);
 w=zeros(nvec,1);
 y=zeros(nvec,1);
   for i=1:nint
      r=randn(nvec,1);
      tt=(i-1)*h;
      f=w.*w;
      y=y + f.*r*hs;
      w=w+hs*r;
   end
   mean=sum(y)/nvec;
   meansqr=sum(y.*y)/nvec;
   disp((sprintf(' %12.0f %12.0f',nvec,nint)))
   disp((sprintf(' %12.5f %12.5f',mean,meansqr)))
   end
   end
```

Computer Programs

Program 3.1. Solution of a stochastic integral $\int_a^b f(t, W(t))\, dW(t)$

This Fortran program estimates the quantities $E\left(\int_a^b f(t, W(t))\, dW(t)\right)$ and $E\left(\int_a^b f(t, W(t))\, dW(t)\right)^2$. For the code listed, the particular integral calculated is $\int_0^1 tW(t)\, dW(t)$. In the program, the value of j is the number of sample paths. The value of n is the number of intervals. Output of the program is listed following the program listing.

```
      real*8 xx
c This program estimates a stochastic integral on [a,b].
      xx=9055301.
      a=0.0
      b=1.0
      j=1000000
      aj=j
      n=2
      do 500 nn=1,7
      n=2*n
      Anna=n
      h=(b-a)/Anna
      hs=sqrt(h)
      s1=0.0
      s2=0.0
      do 400  k=1,j
      s=0.0
      t=-h
      w=0.0
      do 300 i=1,n
      t=t+h
      call random(xx,rand1,rand2)
      call fcalc(t,w,f)
      s=s+f*hs*rand1
      w=w+rand1*hs
  300 continue
      s1=s1+s/aj
      s2=s2+s*s/aj
  400 continue
      write(6,200) n,j,s1,s2
  200 format(3x,i8,3x,i8,3x,f10.5,3x,f10.5)
  500 continue
      stop
      end
```

```
      subroutine fcalc(t,w,f)
      f=t*w
      return
      end
      subroutine random(xx,rand1,rand2)
      real*8 xx,a,b,d,rng(2)
      a=16807.
      ib=2147483647
      b=ib
      do 55 i=1,2
      id=a*xx/b
      d=id
      xx=a*xx-d*b
55    rng(i)=xx/b
      pi=3.141592654
      u1=rng(1)
      u2=rng(2)
      hlp=sqrt(-2.0*alog(u1))
      rand1=hlp*cos(pi*2.0*u2)
      rand2=hlp*sin(pi*2.0*u2)
      return
      end
```

4	1000000	-0.00107	0.14071
8	1000000	-0.00012	0.19151
16	1000000	-0.00069	0.21906
32	1000000	0.00038	0.23508
64	1000000	-0.00007	0.24209
128	1000000	0.00001	0.24486
256	1000000	-0.00002	0.24821

4

Stochastic Differential Equations

4.1 Introduction

In this chapter, Itô stochastic differential equations are studied. It is proved under certain conditions that a unique solution to the stochastic differential equation exists in Hilbert space H_{SP}. Hilbert spaces H_{RV} and H_{SP} are used throughout this chapter to simplify the presentation of stochastic differential equations and to unify the material in this chapter with that of previous chapters. Several properties of stochastic differential equations are derived. Then, applying these properties and Itô's formula to certain stochastic differential equations, exact solutions and moments of the solution are found. Several numerical methods for approximating solutions of stochastic differential equations are described and computational examples are presented. The forward Kolmogorov partial differential equation is derived. The probability density of solutions to a stochastic differential equation satisfies the forward Kolmogorov equation. In Chapter 5, a modeling procedure is justified by studying the development of the forward Kolmogorov equations as a discrete stochastic process approaches a continuous stochastic process. Finally, a procedure is described for estimating parameters in a stochastic differential equation, assuming that data are available at a discrete set of times. There are many excellent texts on the theory, application, and numerical methods for stochastic differential equations such as references [11, 20, 29, 40, 41, 42, 60, 69, 70, 90, 92, 96, 100, 110].

An Itô stochastic differential equation on the interval $[0, T]$ has the form

$$X(t,\omega) = X(0,\omega) + \int_0^t f(s, X(s,\omega))\, ds + \int_0^t g(s, X(s,\omega))\, dW(s,\omega) \quad (4.1)$$

for $0 \le t \le T$ where $X(0,\cdot) \in H_{RV}$ or in differential form

$$dX(t,\omega) = f(t, X(t,\omega))\, dt + g(t, X(t,\omega))\, dW(t,\omega) \quad (4.2)$$

for $0 \leq t \leq T$ with $X(0,\cdot) \in H_{RV}$. The function f is often called the drift coefficient of the stochastic differential equation while g is referred to as the diffusion coefficient. It is assumed that the functions f and g are nonanticipating and satisfy the following conditions (c6) and (c7) for some constant $k \geq 0$.

Condition (c6): $|f(t,x) - f(s,y)|^2 \leq k(|t-s| + |x-y|^2)$ for $0 \leq s,t \leq T$ and $x, y \in \mathbb{R}$.

Condition (c7): $|f(t,x)|^2 \leq k(1+|x|^2)$ for $0 \leq t \leq T$ and $x \in \mathbb{R}$.

Before considering existence and uniqueness of solutions to (4.1), it is useful to explore some of the properties of functions f and g that satisfy conditions (c6) and (c7). First, let $u(t) = f(t, X(t))$ and notice that condition (c7) implies that $u \in H_{SP}$ when $X \in H_{SP}$ and, of course, the same holds for function g. Indeed, for $X \in H_{SP}$,

$$\begin{aligned} \|u\|_{SP}^2 &= \int_0^T E|f(t,X(t)|^2\,dt \\ &\leq \int_0^T k(1+E|X(t)|^2)\,dt \quad \leq \quad kT + k\|X\|_{SP}^2. \end{aligned}$$

Now consider condition (c6). Condition (c6) implies that given $t_1 \in [0,T]$ and $\epsilon > 0$, then $\|f(t_1,X) - f(t_1,Y)\|_{SP} < \epsilon$ when $\|X - Y\|_{SP} < \delta = \epsilon/k^{1/2}$. In addition, given $X \in H_{SP}$ and $\epsilon > 0$, then $\|f(t_1,X) - f(t_2,X)\|_{SP} < \epsilon$ when $|t_2 - t_1| < \delta = \epsilon^2/kT$. To see the first inequality, consider

$$\begin{aligned} \|f(t_1,X) - f(t_1,Y)\|_{SP}^2 &= \int_0^T E|f(t_1,X(t)) - f(t_1,Y(t))|^2\,dt \\ &\leq \int_0^T kE|X(t) - Y(t)|^2\,dt \quad = \quad k\|X - Y\|_{SP}^2. \end{aligned}$$

Thus, if $\|X - Y\|_{SP} < \epsilon/k^{1/2}$, then $\|f(t_1,X) - f(t_1,Y)\|_{SP} < \epsilon$. To see the second inequality, consider

$$\begin{aligned} \|f(t_1,X) - f(t_2,X)\|_{SP}^2 &= \int_0^T E|f(t_1,X(t)) - f(t_2,X(t))|^2\,dt \\ &\leq \int_0^T kE|t_2 - t_1|\,dt \quad = \quad kT|t_2 - t_1|. \end{aligned}$$

Thus, if $|t_2 - t_1| < \epsilon^2/kT$ then $\|f(t_1,X) - f(t_2,X)\|_{SP} < \epsilon$. Furthermore, conditions (c6) and (c7) are useful in applying Itô's formula to (4.2) when $X \in H_{SP}$.

4.2 Existence of a Unique Solution

In this section, we prove existence and uniqueness for $X \in H_{SP}$ that satisfies (4.1) when f and g satisfy conditions (c6) and (c7). The solution is referred to as a strong solution and has a unique sample path. In contrast, solutions are unique in the weak sense if they have the same probability distribution even though their sample paths may differ [70]. To show existence and uniqueness, a Cauchy sequence of functions in H_{SP} will be constructed. The limit of this sequence will be the solution of (4.1). The procedure employs a Picard iterative sequence [41, 79].

Let the first element of this sequence be $X_0(t) = X(0)$, where $X(0) \in H_{RV}$ is the given initial condition. Notice that $X_0 \in H_{SP}$. Define $X_1(t)$ by the equation

$$X_1(t) = X_0(t) + \int_0^t f(s, X_0(s))\, ds + \int_0^t g(s, X_0(s))\, dW(s).$$

As f and g satisfy conditions (c6) and (c7) and $X_0 \in H_{SP}$, then $X_1 \in H_{SP}$ and

$$\|X_1 - X_0\|_{SP} \leq \left(k(2T^2 + 2T^3)\right)^{1/2} \left(1 + \|X_0\|_{RV}^2\right)^{1/2}.$$

This inequality follows from the following argument.

$$\begin{aligned}
\|X_1 - X_0\|_{SP}^2 &= \int_0^T E \left| \int_0^t f(s, X_0(s))\, ds + \int_0^t g(s, X_0(s))\, dW(s) \right|^2 dt \\
&\leq 2 \int_0^T \left(E \left| \int_0^t f(s, X_0(s))\, ds \right|^2 + E \left| \int_0^t g(s, X_0(s))\, dW(s) \right|^2 \right) dt \\
&\leq 2 \int_0^T \left(t \int_0^t E|f(s, X_0(s))|^2\, ds + \int_0^t E|g(s, X_0(s)|^2\, ds \right) dt \\
&\leq 2 \int_0^T (T + T^2) k (1 + \|X_0\|_{RV}^2)\, dt \\
&= k(2T^2 + 2T^3)(1 + \|X_0\|_{RV}^2).
\end{aligned}$$

Continuing this procedure, the sequence $\{X_n\}_{n=1}^{\infty} \subset H_{SP}$ is defined recursively for $n \geq 1$ as

$$X_n(t) = X_0(t) + \int_0^t f(s, X_{n-1}(s))\, ds + \int_0^t g(s, X_{n-1}(s))\, dW(s).$$

To see that this sequence is Cauchy in H_{SP} notice that

$$\begin{aligned}
X_{n+1}(t) - X_n(t) &= \int_0^t \left(f(s, X_n(s)) - f(s, X_{n-1}(s)) \right) ds \\
&\quad + \int_0^t \left(g(s, X_n(s)) - g(s, X_{n-1}(s)) \right) dW(s).
\end{aligned}$$

Therefore,

$$\begin{aligned}
E|X_{n+1}(t) - X_n(t)|^2 &\le 2t \int_0^t E|f(s, X_n(s)) - f(s, X_{n-1}(s))|^2 \, ds \\
&\quad + 2 \int_0^t E|g(s, X_n(s)) - g(s, X_{n-1}(s))|^2 \, ds \\
&\le 2tk \int_0^t E|X_n(s) - X_{n-1}(s)|^2 \, ds + 2k \int_0^t E|X_n(s) - X_{n-1}(s)|^2 \, ds \\
&\le (2Tk + 2k) \int_0^t E|X_n(s) - X_{n-1}(s)|^2 \, ds \\
&\le L \int_0^t E|X_n(s) - X_{n-1}(s)|^2 \, ds,
\end{aligned}$$

where $L = 2Tk + 2k$. Let $a_n(t) = E|X_{n+1}(t) - X_n(t)|^2$. Then, by the above inequality,

$$\begin{aligned}
a_n(t) &\le L \int_0^t a_{n-1}(s_1) \, ds_1 \le L^2 \int_0^t \int_0^{s_1} a_{n-2}(s_2) \, ds_2 \, ds_1 \\
&\le \cdots \le L^n \int_0^t \int_0^{s_1} \cdots \int_0^{s_{n-1}} a_0(s_n) \, ds_n \ldots ds_2 \, ds_1.
\end{aligned}$$

So,

$$a_n(t) \le L^n \int_0^t \int_0^{s_1} \cdots \int_0^{s_{n-1}} a_0(s_n) \, ds_n \ldots ds_2 \, ds_1.$$

This latter inequality implies that

$$a_n(t) \le \frac{L^n t^{n-1}}{(n-1)!} \|X_1 - X_0\|_{SP}^2$$

and it follows that

$$\|X_{n+1} - X_n\|_{SP}^2 \le \frac{L^n T^n}{n!} \|X_1 - X_0\|_{SP}^2.$$

Consider $m > n$. By repeated application of the triangle inequality,

$$\begin{aligned}
&\|X_m - X_n\|_{SP} \le \|X_m - X_{m-1}\|_{SP} + \|X_{m-1} - X_n\|_{SP} \\
&\le \|X_m - X_{m-1}\|_{SP} + \|X_{m-1} - X_{m-2}\|_{SP} + \cdots + \|X_{n+1} - X_n\|_{SP} \\
&\le \left[\left(\frac{L^n T^n}{n!} \right)^{1/2} + \left(\frac{L^{n+1} T^{n+1}}{(n+1)!} \right)^{1/2} + \cdots + \left(\frac{L^m T^m}{m!} \right)^{1/2} \right] \|X_1 - X_0\|_{SP} \\
&\le \left(\frac{L^n T^n}{n!} \right)^{1/2} \left[1 + \cdots + \left(\frac{L^{m-n} T^{m-n}}{(n+1)(n+2) \ldots m} \right)^{1/2} \right] \|X_1 - X_0\|_{SP} \\
&\le 2 \left(\frac{L^n T^n}{n!} \right)^{1/2} \|X_1 - X_0\|_{SP} \quad \text{assuming that} \quad n, m \ge 4LT.
\end{aligned}$$

As $\|X_1 - X_0\|_{SP}$ is bounded, it follows from the preceding inequality that given any $\epsilon > 0$ there exists an $N > 0$ such that $\|X_n - X_m\|_{SP} < \epsilon$ when $n, m > N$. Hence, the sequence $\{X_n\}_{n=1}^{\infty}$ is Cauchy in H_{SP} and X_n converges to a unique $X \in H_{SP}$ as $n \to \infty$. Now let $Y \in H_{SP}$ where Y satisfies the relation

$$Y(t) = -X(t) + X(0) + \int_0^t f(s, X(s))\, ds + \int_0^t g(s, X(s))\, dW(s).$$

Because

$$0 = -X_n(t) + X(0) + \int_0^t f(s, X_{n-1}(s))\, ds + \int_0^t g(s, X_{n-1}(s))\, dW(s)$$

and $\|X - X_n\|_{SP} \to 0$ as $n \to \infty$, it is clear that $\|Y\|_{SP} = 0$. So $X(t)$ is the unique solution of (4.1) in H_{SP}.

4.3 Properties of Solutions to Stochastic Differential Equations

Some properties of solutions of stochastic differential equations are discussed in this section. The first theorem implies that the solution of (4.1) satisfies $\|X\|_{SP}^2 \le 3T\big(\|X(0)\|_{RV}^2 + kT^2 + kT\big) \exp\big(3k(T + T^2)\big)$. The second theorem states that the solution of (4.1) is continuous on $[0, T]$ in the $\|\cdot\|_{RV}$ norm. Specifically, given $\epsilon > 0$, there is a $\delta > 0$ such that $\|X(t) - X(r)\|_{RV} < \epsilon$ when $|t - r| < \delta$.

Theorem 4.1. *(Boundedness of solutions)*
Assume that f and g satisfy conditions (c6) and (c7) and $X \in H_{SP}$ is the solution of (4.1). Then

$$E|X(t)|^2 \le 3\big(E|X(0)|^2 + kT^2 + kT\big) \exp\big(3k(T + T^2)\big) \quad \textit{for} \quad 0 \le t \le T.$$

Proof. Notice that

$$\begin{aligned}
E|X(t)|^2 &\le E\left|X(0) + \int_0^t f(s, X(s))\, ds + \int_0^t g(s, X(s))\, dW(s)\right|^2 \\
&\le 3E\,|X(0)|^2 + 3E\left|\int_0^t f(s, X(s))\, ds\right|^2 + 3E\left|\int_0^t g(s, X(s))\, dW(s)\right|^2 \\
&\le 3E\,|X(0)|^2 + 3t\int_0^t E|f^2(s, X(s))|\, ds + 3\int_0^t E|g^2(s, X(s))|\, ds \\
&\le 3E\,|X(0)|^2 + (3t + 3)k\int_0^t E|1 + X^2(s)|\, ds \\
&\le 3E\,|X(0)|^2 + (3t^2 + 3t)k + (3T + 3)k\int_0^t E|X^2(s)|\, ds.
\end{aligned}$$

Letting $a(t) = E|X^2(t)|$ and $b(t) = 3E|X^2(0)| + (3t^2 + 3t)k$, the above inequality can be written as

$$a(t) \le b(t) + (3T + 3)k \int_0^t a(s)\, ds.$$

Using the Bellman-Gronwall inequality, it follows that

$$a(t) \le b(t) + (3T + 3)k \int_0^t \exp\big(k(3T + 3)(t - s)\big) b(s)\, ds.$$

As $b(t)$ is increasing on $[0, T]$,

$$a(t) \le b(t) + b(t)(3T + 3)k \int_0^t \exp\big(k(3T + 3)(t - s)\big)\, ds.$$

Thus,

$$E|X(t)|^2 \le 3\big(E|X(0)|^2 + kT^2 + kT\big) \exp\big(3k(T + T^2)\big) \quad \text{for} \quad 0 \le t \le T.$$

Theorem 4.2. *(Continuity of solutions on $[0, T]$)*
Assume that f and g satisfy conditions (c6) and (c7) and $X \in H_{SP}$ is the solution of (4.1). Then, there is a constant $c \ge 0$ such that

$$E|X(t) - X(r)|^2 \le c|t - r| \quad \textit{for} \quad 0 \le r, t \le T.$$

In particular, given $\epsilon > 0$ there is a $\delta > 0$ such that $\|X(t) - X(r)\|_{RV} < \epsilon$ when $|t - r| < \delta$.

Proof. Clearly,

$$X(t) - X(r) = \int_r^t f(s, X(s))\, ds + \int_r^t g(s, X(s))\, dW(s).$$

However, the previous theorem implies that there is an $M > 0$ such that $E|X(s)|^2 \le M$ for $0 \le s \le T$. Using this inequality and condition (c7),

$$\begin{aligned} E|X(t) - X(r)|^2 &\le 2|t - r| \int_r^t E|f(s, X(s))|^2\, ds + 2 \int_r^t E|g(s, X(s))|^2\, ds \\ &\le 2|t - r|k \int_r^t (1 + E|X(s)|^2)\, ds + 2k \int_r^t (1 + E|X(s)|^2)\, ds \\ &\le |t - r|\big(2k|t - r|(1 + M) + 2k(1 + M)\big) \\ &\le c|t - r| \quad \text{where} \quad c = 2k(T + 1)(1 + M). \end{aligned}$$

Remark 4.3. The Bellman-Gronwall inequality

As seen in this section, a useful inequality is the Bellman-Gronwall inequality which states:

$$\text{If} \quad a(t) \le b(t) + c\int_0^t a(s)\,ds, \quad \text{then} \quad a(t) \le b(t) + c\int_0^t \exp(c(t-s))b(s)\,ds.$$

To see this result, suppose that

$$a(t) \le b(t) + c\int_0^t a(s)\,ds. \tag{4.3}$$

Then,

$$a(s) - c\int_0^s a(z)\,dz \le b(s).$$

Therefore,

$$\frac{d}{ds}\left(\exp(-cs)\int_0^s a(z)\,dz\right) \le b(s)\exp(-cs),$$

and so

$$\exp(-ct)\int_0^t a(z)\,dz \le \int_0^t b(s)\exp(-cs)\,ds.$$

Hence,

$$\int_0^t a(s)\,ds \le \int_0^t b(s)\exp(c(t-s))\,ds.$$

Substituting the last inequality into (4.3) gives:

$$a(t) \le b(t) + c\int_0^t \exp(c(t-s))b(s)\,ds. \tag{4.4}$$

4.4 Itô's Formula and Exact Solutions

In this section, Itô's formula is stated for the stochastic differential equation (4.2). Then, several examples are given to illustrate Itô's formula. Itô's formula is helpful in finding exact solutions to certain stochastic differential equations. Also, it is shown how Itô's formula can be used to determine exact moments of the solution for certain problems even though the exact solution may be unknown.

Consider, therefore, the Itô stochastic differential equation in differential form

$$dX(t) = f(t, X(t))\,dt + g(t, X(t))\,dW(t)$$

for $0 \le t \le T$ with $X(0) \in H_{RV}$. Let F be a smooth function and assume that the conditions of Theorem 3.12 are satisfied. Then, Itô's formula can be applied to $F(t, X)$ where X satisfies the stochastic differential (4.2). This yields the stochastic differential for F of the form

$$\begin{aligned} dF(t, X(t)) = &\left(\frac{\partial F(t,X)}{\partial t} + f(t,X)\frac{\partial F(t,X)}{\partial x} + \frac{1}{2}g^2(t,X)\frac{\partial^2 F(t,X)}{\partial x^2}\right)dt \\ &+ g(t,X)\frac{\partial F(t,X)}{\partial x}dW(t), \end{aligned} \tag{4.5}$$

where the notation $\dfrac{\partial F(t,X)}{\partial x} = \dfrac{\partial F(t,x)}{\partial x}\Big|_{x=X}$.

Itô's formula allows us, for example, to determine moments of the solution for certain stochastic differential equations. To find these moments, the following relation, derived in the previous chapter, is useful. That is,

$$E\left(\int_0^t G(t,X(t))\,dW(t)\right) = 0.$$

Before considering two interesting examples, it is important to notice that care must be taken in manipulating stochastic differential equations. For example, applying Itô's formula to $F(t,X(t)) = X^2(t)$ where X satisfies (4.2) yields

$$d(X^2(t)) = [2X(t)f(t,X(t)) + g^2(t,X(t))]\,dt + [2X(t)g(t,X(t))]\,dW(t).$$

Notice, in particular, that

$$d(X^2(t)) \neq 2X(t)dX(t) = 2X(t)[f(t,X(t))\,dt + g(t,X(t))\,dW(t)].$$

Now consider the following two examples where Itô's formula is used to determine exact moments of the solutions for two different stochastic differential equations.

Example 4.4. Finding exact moments for an SDE with linear coefficients

Consider the stochastic differential equation

$$dX(t) = dt + X(t)\,dW(t), \quad X(0) = 0.$$

Then

$$X(t) = t + \int_0^t X(s)\,dW(s).$$

It follows that

$$E(X(t)) = t.$$

Applying Itô's formula to $F(t,X) = X^2$ yields

$$d(X^2(t)) = [2X(t) + X^2(t)]\,dt + 2X^2(t)\,dW(t)$$

so that

$$E(X^2(t)) = E\int_0^t \big(2X(s) + X^2(s)\big)\,ds.$$

This leads to a differential equation for $E(X^2(t))$ of the form

$$\frac{dE(X^2(t))}{dt} = 2E(X(t)) + E(X^2(t)) = 2t + E(X^2(t))$$

with $E(X^2(0)) = 0$. Solving this ordinary differential equation gives the exact second moment for $X(t)$,

$$E(X^2(t)) = -2t - 2 + 2e^t.$$

In addition, it is clear that $\mathrm{Var}(X(t)) = E(X^2(t)) - (E(X(t))^2 = 2e^t - 2 - 2t - t^2$. This procedure can be continued to find higher order moments for this problem. If Itô's formula is applied to $F(t, X) = X^3$, then

$$d(X^3(t)) = [3X^2(t) + 3X^3(t)]\,dt + 3X^3(t)\,dW(t).$$

This leads to the differential equation for $E(X^3(t))$ of the form

$$\frac{dE(X^3(t))}{dt} = 3E(X^2(t)) + 3E(X^3(t)) = -6t - 6 + 6e^t + 3E(X^3(t))$$

with $E(X^3(0)) = 0$. Solving this gives the exact third moment for $X(t)$,

$$E(X^3(t)) = 2t + \frac{8}{3} - 3e^t + \frac{1}{3}e^{3t}.$$

Example 4.5. Finding exact moments for an SDE with nonlinear coefficients
Consider the stochastic differential equation

$$dX(t) = -\frac{1}{4}X^3(t)\,dt + \frac{1}{2}X^2(t)\,dW(t) \quad \text{with} \quad X(0) = \frac{1}{2}.$$

In this example, $E(X(t))$ and $E(X^3(t))$ are to be determined exactly. First,

$$dE(X(t)) = -\frac{1}{4}E(X^3(t))\,dt \quad \text{with} \quad E(X(0)) = \frac{1}{2}$$

so $E(X^3(t))$ is needed in order to find $E(X(t))$. However, applying Itô's formula to the stochastic differential equation gives

$$dX^3(t) = \left[-\frac{3}{4}X^5(t) + \frac{3}{4}X^5(t)\right]\,dt + \frac{3}{2}X^4(t)\,dW(t) \;= \frac{3}{2}X^4(t)\,dW(t)$$

with $E(X^3(0)) = \frac{1}{8}$. Thus, $E(X^3(t)) = \frac{1}{8}$ and it follows that $E(X(t)) = \frac{1}{2} - \frac{1}{32}t$.

Example 4.6. Finding exact moments for an SDE with nonlinear coefficients
Consider the stochastic differential equation

$$dX(t) = \left[\frac{1}{3}X^{1/3}(t) + 6X^{2/3}(t)\right]\,dt + X^{2/3}(t)\,dW(t) \quad \text{with} \quad X(0) = 1.$$

In this example, we wish to determine $E(X(t))$ and $E(X^2(t))$ exactly. First notice that

$$dE(X(t)) \neq \left[\frac{1}{3}\big(E(X(t))\big)^{1/3} + 6\big(E(X(t))\big)^{2/3}\right]dt$$

so an appropriate change of variables is required to find the moments. Let

$$Y_n(t) = (X(t))^{n/3} \quad \text{for} \quad n = 0, 1, 2, \ldots, 6.$$

Next, applying Itô's formula, the stochastic differentials are obtained

$$dY_n(t) = \left[\frac{1}{18}(n^2 - n)X^{\frac{n-2}{3}}(t) + 2nX^{\frac{n-1}{3}}(t)\right] dt + \left[\frac{n}{3}X^{\frac{n-1}{3}}(t)\right] dW(t)$$

with $Y_n(0) = 1$ for $n = 0, 1, 2, \ldots, 6$. Substituting $X^{\frac{n-2}{3}}(t) = Y_{n-2}(t)$ and $X^{\frac{n-1}{3}}(t) = Y_{n-1}(t)$ gives

$$dY_n(t) = \left[\frac{1}{18}(n^2 - n)Y_{n-2}(t) + 2nY_{n-1}(t)\right] dt + \left[\frac{n}{3}Y_{n-1}(t)\right] dW(t)$$

for $n = 0, 1, 2, \ldots, 6$. Finally, letting $Z_n(t) = E(Y_n(t)) = E((X(t))^{n/3})$, one obtains the initial-value system

$$\frac{dZ_n(t)}{dt} = \frac{1}{18}(n^2 - n)Z_{n-2}(t) + 2nZ_{n-1}(t) \quad \text{for} \quad n = 1, 2, \ldots, 6$$

with $Z_n(0) = 1$ for $n = 1, 2, \ldots, 6$ and $Z_0(t) = 1$. Solving this system for $n = 1, 2, 3$ and $n = 6$ produces the results

$$Z_1(t) = E((X(t))^{1/3}) = 2t + 1$$
$$Z_2(t) = E((X(t))^{2/3}) = 4t^2 + \frac{37}{9}t + 1$$
$$Z_3(t) = E((X(t)) = 8t^3 + \frac{38}{3}t^2 + \frac{19}{3}t + 1$$
$$Z_6(t) = E((X(t))^2) = 64t^6 + \frac{656}{3}t^5 + \frac{2660}{9}t^4 + \frac{49145}{243}t^3 + \frac{665}{9}t^2 + \frac{41}{3}t + 1.$$

In particular, $E(X(1)) = 28.0$ and $E(X^2(1)) = 869.0206$.

Finding exact solutions of Itô stochastic differential equations is difficult but is facilitated by appropriately applying Itô's formula. In the following two examples, exact solutions are obtained by applying Itô's formula. Additional examples are given in the exercises at the end of this chapter.

Example 4.7. Exact solution of a stochastic differential equation

Consider the stochastic differential equation

$$dX(t) = -\alpha X(t)\, dt + \sigma\, dW(t), \quad X(0) = X_0,$$

where α, σ, and X_0 are constants. Let $F(t, X) = e^{\alpha t}X(t)$. By Itô's formula,

$$d\big(e^{\alpha t}X(t)\big) = \big[\alpha e^{\alpha t}X(t) - \alpha e^{\alpha t}X(t)\big]\, dt + \sigma e^{\alpha t}\, dW(t).$$

Thus,

$$e^{\alpha t}X(t) - X(0) = \int_0^t e^{\alpha s}\sigma\, dW(s).$$

So the exact solution is

$$X(t) = X(0)e^{-\alpha t} + e^{-\alpha t}\int_0^t e^{\alpha s}\sigma\, dW(s).$$

Example 4.8. Exact solution of a stochastic differential equation
Consider the stochastic differential equation

$$dX(t) = f(t)X(t)\,dt + g(t)X(t)\,dW(t), \quad X(0) = X_0,$$

where X_0 is a constant. For this problem, the exact solution has the form

$$X(t) = X_0 \exp\left(\int_0^t \left(f(s) - \frac{1}{2}g^2(s)\right) ds + \int_0^t g(s)\,dW(s)\right).$$

To see this, let $F(t, X) = \ln(X(t))$. Applying Itô's formula,

$$d(\ln(X(t))) = \left[f(t) - \frac{1}{2}g^2(t)\right] dt + g(t)\,dW(t).$$

Thus,

$$\ln(X(t)) - \ln(X_0) = \int_0^t \left(f(s) - \frac{1}{2}g^2(s)\right) ds + \int_0^t g(s)\,dW(s)$$

which yields the solution.

4.5 Approximating Stochastic Differential Equations

As the exact solution to a stochastic differential equation is generally difficult to obtain, it is useful to be able to approximate the solution. Euler's (or the Euler-Maruyama) method [41, 60, 69, 70, 72, 90, 100, 101, 110] is a simple numerical method. When applied to (4.2), Euler's method has the form

$$X_{i+1}(\omega) = X_i(\omega) + f(t_i, X_i(\omega))\Delta t + g(t_i, X_i(\omega))\Delta W_i(\omega), \quad X_0(\omega) = X(0, \omega),$$

for $i = 0, 1, 2, \ldots, N-1$ where $X_i(\omega) \approx X(t_i, \omega)$, $t_i = i\Delta t$, $\Delta t = T/N$, $\Delta W_i(\omega) = (W(t_{i+1}, \omega) - W(t_i, \omega)) \sim N(0, \Delta t)$, and where ω indicates a sample path. For simplicity, only equidistant time discretizations in the numerical methods are considered in this section although variable step size selection is an issue of current interest [69, 100]. Before describing some computational results for Euler's method, the error in Euler's method is studied. To study the error in this method, it is useful to approximate the solution for all $t \in [0, T]$ and not just at the nodal points t_i for $i = 0, 1, 2, \ldots, N$. To accomplish this, $\hat{X}(t) \approx X(t)$ is defined as

$$\hat{X}(t) = X_i + \int_{t_i}^{t} f(t_i, X_i)\, ds + \int_{t_i}^{t} g(t_i, X_i)\, dW(s)$$

for $t_i \le t \le t_{i+1}$ and $i = 0, 1, \ldots, N-1$ where, for notational simplicity, the dependence on sample path ω is dropped. Notice, in particular, that $\hat{X}$ is identical to Euler's method approximation at the nodal points, that is, $\hat{X}(t_i) = X_i$ for $i = 0, 1, \ldots, N$. Also notice that on the ith subinterval $\hat{X}(t)$ is the solution of the stochastic differential equation

$$\begin{cases} d\hat{X}(t) = f(t_i, X_i)\, dt + g(t_i, X_i)\, dW(t), & t_i \le t \le t_{i+1} \\ \hat{X}(t_i) = X_i \end{cases}$$

for $i = 0, 1, 2, \ldots, N-1$. Recall also that the solution $X(t)$ satisfies the stochastic differential equation

$$dX(t) = f(t, X(t))\, dt + g(t, X(t))\, dW(t), \quad t_i \le t \le t_{i+1}$$

for $i = 0, 1, 2, \ldots, N-1$. Define the error as $\epsilon(t) = X(t) - \hat{X}(t)$. Then the error ϵ satisfies the stochastic differential equation

$$\begin{cases} d\epsilon(t) = (f(t, X(t)) - f(t_i, X_i))\, dt + (g(t, X(t)) - g(t_i, X_i))\, dW(t), \\ \epsilon(t_i) = X(t_i) - \hat{X}(t_i) \end{cases}$$

for $t_i \le t \le t_{i+1}$ and $i = 0, 1, 2, \ldots, N-1$. Using Itô's formula, the stochastic differential for $\epsilon^2(t)$ is obtained as

$$\begin{aligned} d(\epsilon^2(t)) &= 2\big(X(t) - \hat{X}(t)\big)\big(f(t, X(t)) - f(t_i, X_i)\big)\, dt \\ &+ \big(g(t, X(t)) - g(t_i, X_i)\big)^2\, dt + 2\big(X(t) - \hat{X}(t)\big)\big(g(t, X(t)) - g(t_i, X_i)\big)\, dW(t) \end{aligned}$$

for $t_i \le t \le t_{i+1}$. Hence, $E(\epsilon^2(t_{i+1}))$ satisfies

$$\begin{aligned} E(\epsilon^2(t_{i+1})) = E(\epsilon^2(t_i)) &+ E\int_{t_i}^{t_{i+1}} \big(g(t, X(t)) - g(t_i, X_i)\big)^2\, dt \\ &+ E\int_{t_i}^{t_{i+1}} 2\big(X(t) - \hat{X}(t)\big)\big(f(t, X(t)) - f(t_i, X_i)\big)\, dt \\ &+ E\int_{t_i}^{t_{i+1}} 2\big(X(t) - \hat{X}(t)\big)\big(g(t, X(t)) - g(t_i, X_i)\big)\, dW(t). \end{aligned}$$

Using the inequality $|2ab| \le a^2 + b^2$ and properties of stochastic integrals,

$$\begin{aligned} E(\epsilon^2(t_{i+1})) \;\le\; E(\epsilon^2(t_i)) &+ \int_{t_i}^{t_{i+1}} E(X(t) - \hat{X}(t))^2\, dt \\ &+ \int_{t_i}^{t_{i+1}} E(f(t, X(t)) - f(t_i, X_i))^2\, dt + \int_{t_i}^{t_{i+1}} E(g(t, X(t)) - g(t_i, X_i))^2\, dt. \end{aligned}$$

But

$$\begin{aligned}|f(t,X(t))-f(t_i,X_i)|^2 &\le 2|f(t,X(t))-f(t_i,X(t_i))|^2\\ &\quad +2|f(t_i,X(t_i))-f(t_i,X_i)|^2\\ &\le 2k|t-t_i|+2k|X(t)-X(t_i)|^2+2k|X(t_i)-X_i|^2\end{aligned}$$

and similarly for g using property (c6). Hence,

$$\begin{aligned}E(\epsilon^2(t_{i+1})) &\le E(\epsilon^2(t_i))+\int_{t_i}^{t_{i+1}} E(X(t)-\hat{X}(t))^2\,dt\\ &\quad +4k(1+c)\int_{t_i}^{t_{i+1}}(t-t_i)\,dt+4k\int_{t_i}^{t_{i+1}}E(\epsilon^2(t_i))\,dt\end{aligned}$$

using Theorem 4.2 that $E|X(t)-X(t_i)|^2 \le c|t-t_i|$. Therefore,

$$E(\epsilon^2(t_{i+1})) \le E(\epsilon^2(t_i))(1+4k\Delta t)+2k(1+c)(\Delta t)^2+\int_{t_i}^{t_{i+1}} E(\epsilon^2(s))\,ds.$$

By the Bellman-Gronwall inequality with $b(t)=E(\epsilon^2(t_i))(1+4k\Delta t)+2k(1+c)(\Delta t)^2$,

$$\begin{aligned}E(\epsilon^2(t_{i+1})) &\le E(\epsilon^2(t_i))(1+4k\Delta t)+2k(1+c)(\Delta t)^2\\ &\quad +\int_{t_i}^{t_{i+1}} e^{(t_{i+1}-t)}\left[E(\epsilon^2(t_i))(1+4k\Delta t)+2k(1+c)(\Delta t)^2\right]\,dt\\ &= e^{\Delta t}\left[E(\epsilon^2(t_i))(1+4k\Delta t)+2k(1+c)(\Delta t)^2\right].\end{aligned}$$

Letting $a_i=E(\epsilon^2(t_i))$, $R=e^{\Delta t}(1+4k\Delta t)$, and $S=e^{\Delta t}2k(1+c)(\Delta t)^2$, then

$$a_{i+1} \le Ra_i+S \quad \text{for} \quad i=0,1,2,\ldots,N-1.$$

These inequalities yield

$$a_N \le S\frac{R^N-1}{R-1} \quad \text{with} \quad a_0=E(\epsilon^2(0))=0.$$

Hence,

$$E(\epsilon^2(t_N)) \le \frac{e^{\Delta t}2k(1+c)(\Delta t)^2e^{N\Delta t}e^{4kN\Delta t}}{e^{\Delta t}-1+e^{\Delta t}4k\Delta t} \le \Delta t\frac{(1+c)e^{(1+4k)T}}{2}.$$

This result holds for any nodal point and the mean square error in Euler's method satisfies

$$E|X(t_i)-X_i|^2 \le \hat{c}\Delta t$$

for $i=0,1,2,\ldots,N$ where $\hat{c}=\frac{1}{2}(1+c)e^{(1+4k)T}$.

Consider briefly a continuous piecewise linear approximation to the solution $X(t)$ which is identical to Euler's method approximation at the nodal points, t_i for $i=0,1,2,\ldots,N$. This approximate solution is commonly plotted. Let

$$\tilde{X}(t) = X_i(t_{i+1} - t)/\Delta t + X_{i+1}(t - t_i)/\Delta t$$

for $t_i \le t \le t_{i+1}$ and $i = 0, 1, 2, \ldots, N-1$ where $\{X_i\}_{i=0}^N$ is the Euler approximation to (4.2) at the $N+1$ nodal points $\{t_i\}_{i=0}^N$. The function $\tilde{X}(t)$ is a continuous linear approximation to the solution $X(t)$ and it is straightforward to show that $\|X - \tilde{X}\|_{SP} \le \tilde{c}(\Delta t)^{1/2}$ for a constant $\tilde{c} \ge 0$ and hence $\|X - \tilde{X}\|_{SP} \to 0$ as $\Delta t \to 0$. To see this, consider

$$\begin{aligned}
\|X - \tilde{X}\|_{SP}^2 &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E|X(t) - \tilde{X}(t)|^2 \, dt \\
&\le 4 \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(E\big|(X(t) - X(t_i))\big|^2 + E\big|X(t_i) - X(t_{i+1})\big|^2 \frac{(t - t_i)^2}{(\Delta t)^2} \right) dt \\
&\quad + 4 \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\big|X(t_i) - X_i)\big|^2 \frac{(t_{i+1} - t)^2}{(\Delta t)^2} \, dt \\
&\quad + 4 \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} E\big|X(t_{i+1}) - X_{i+1}\big|^2 \frac{(t - t_i)^2}{(\Delta t)^2} \, dt \\
&\le 4 \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \left(c(t - t_i) + \frac{c(t - t_i)^2}{\Delta t} + \frac{\hat{c}(t_{i+1} - t)^2}{\Delta t} + \frac{\hat{c}(t - t_i)^2}{\Delta t} \right) dt \\
&\le 4 \sum_{i=0}^{N-1} (5c + 4\hat{c})(\Delta t)^2/6 \\
&\le \tilde{c}^2 \Delta t \quad \text{where} \quad \tilde{c}^2 = (10c + 8\hat{c})T/3.
\end{aligned}$$

Higher order numerical methods can be developed for stochastic differential equations which are similar in some respects to higher order methods for ordinary differential equations. For example, there are explicit or implicit multistep methods and Runge-Kutta methods [1, 3, 60, 69, 70, 72, 76, 90, 99, 100, 101, 110, 111]. A popular second-order method is Milstein's method. Milstein's method has mean square error proportional to $(\Delta t)^2$ rather than Δt. For one dimension, Milstein's method has the form

$$\begin{aligned}
X_{i+1}(\omega) = X_i(\omega) &+ f(t_i, X_i(\omega))\Delta t + g(t_i, X_i(\omega))\Delta W_i(\omega) \\
&+ \frac{1}{2} g(t_i, X_i(\omega)) \frac{\partial g(t_i, X_i(\omega))}{\partial x} [(\Delta W_i(\omega))^2 - \Delta t]
\end{aligned}$$

for $i = 0, 1, 2, \ldots, N-1$ with $X_0(\omega) = X(0, \omega)$, where $X_i(\omega) \approx X(t_i, \omega)$, $\Delta W_i(\omega) = (W(t_{i+1}, \omega) - W(t_i, \omega)) \sim N(0, \Delta t)$, $t_i = i\Delta t$, $\Delta t = T/N$, and where ω indicates a sample path. Notice that Milstein's method has an additional term at each step in comparison with Euler's method. Listed at the end of this chapter are computer programs for solving a stochastic differential equation using Euler's method or using both Euler's method and Milstein's method.

Example 4.9. Approximation of an SDE using Euler's and Milstein's methods
Consider the stochastic differential equation

$$dX(t) = \left[\frac{1}{3}X^{1/3}(t) + 6X^{2/3}(t)\right] dt + X^{2/3}(t)\, dW(t), \quad X(0) = 1.$$

In Example 4.6, it was proved that $E(X(1)) = 28.0$ and $E(X^2(1)) = 869.0206$. For this problem, Euler's method has the form

$$X_{i+1} = X_i + \left[\frac{1}{3}X_i^{1/3} + 6X_i^{2/3}\right] \Delta t + X_i^{2/3}\sqrt{\Delta t}\, \eta_i \quad \text{where} \quad \eta_i \sim N(0,1)$$

for $i = 0, 1, 2, \ldots, N-1$ with $X_0 = 1$, $t_i = i\Delta t$, and $\Delta t = 1/N$. Milstein's method has the form

$$X_{i+1} = X_i + \left[\frac{1}{3}X_i^{1/3} + 6X_i^{2/3}\right] \Delta t + X_i^{2/3}\sqrt{\Delta t}\, \eta_i + \frac{1}{3}X_i^{1/3}(\eta_i^2 - 1)\Delta t,$$

where $\eta_i \sim N(0,1)$. The calculational results for the mean square error $E|X(1) - X_N|^2$ are given in Table 4.1.

Table 4.1. Calculated errors $E|X(1) - X_N|^2$ for Euler's and Milstein's methods

Value of N	Euler Error	Milstein Error
2^9	2.80×10^{-2}	1.61×10^{-2}
2^{10}	1.04×10^{-2}	4.03×10^{-3}
2^{11}	4.20×10^{-3}	1.01×10^{-3}
2^{12}	1.89×10^{-3}	2.53×10^{-4}
2^{13}	8.76×10^{-4}	6.24×10^{-5}
2^{14}	4.12×10^{-4}	1.60×10^{-5}

In Table 4.1, the number of intervals N used in Euler's method and Milstein's method were selected as $2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}$, and 2^{14} and 10,000 sample paths were calculated for each value of N. For each sample path, the Wiener process was calculated at $2^{15} = 32768$ equally spaced points on the interval $[0, 1]$ and these points supplied the required values of the Wiener process for the computations using 2^9 to 2^{14} intervals. As the value of $X(1)$ depends on the sample path and is unknown, the value of $X(1)$ used in the error estimate was calculated for each sample path based on the Milstein approximation with $N = 2^{15}$ intervals. In Table 4.1, the calculated values of the errors $\frac{1}{10{,}000}\sum_{j=1}^{10{,}000} |X^{(j)}(1) - X_N^{(j)}|^2 \approx E|X(1) - X_N|^2$ are given for each numerical method and for each value of N where $X_N^{(j)}$ is the estimate of $X^{(j)}(1)$ for the jth sample path using N intervals. Notice that the mean square errors are approximately proportional to $\Delta t = 1/N$ for Euler's method and to $(\Delta t)^2 = 1/N^2$ for Milstein's method.

Table 4.2. Calculated estimates of $E(X(1))$

Value of N	Euler Estimate	Milstein Estimate
2^6	27.07 (0.93)	27.08 (0.92)
2^7	27.56 (0.44)	27.56 (0.44)
2^8	27.79 (0.21)	27.79 (0.21)

Table 4.3. Calculated estimates of $E(X^2(1))$

Value of N	Euler Estimate	Milstein Estimate
2^6	810.15 (58.87)	810.18 (58.84)
2^7	840.89 (28.13)	840.93 (28.09)
2^8	855.33 (13.69)	855.31 (13.71)

Next, the calculations for this example were repeated for 100,000 sample paths for $N = 64, 128$, and 256 to estimate $E(X(1)) \approx \sum_{j=1}^{100,000} X_N^{(j)}/100,000$ and $E(X(1))^2 \approx \sum_{j=1}^{100,000} (X_N^{(j)})^2/100,000$ where $X_N^{(j)}$ is the estimate of $X(1)$ for the jth sample path using N intervals. The calculational results are given in Tables 4.2 and 4.3. Adjacent to each value, the error is given in parentheses. Recall that $E(X(1)) = 28.0$ and $E(X^2(1)) = 869.0206$ are the exact values. In addition, for this example, the mean and one sample path are plotted in Fig. 4.1.

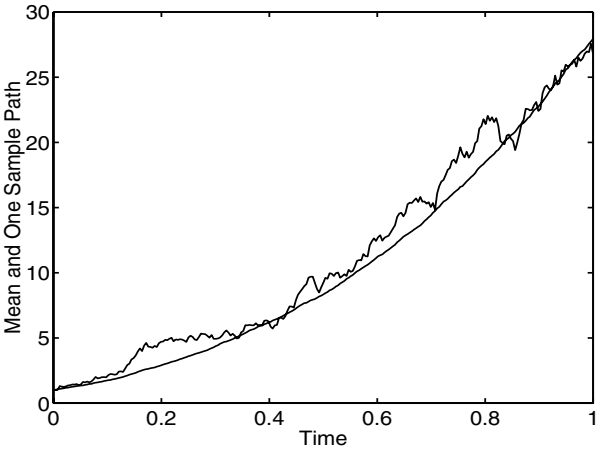

Fig. 4.1. Mean solution and one sample path for Example 4.9

In Tables 4.2 and 4.3, notice that the errors in the mean values are proportional to Δt for either numerical method. In particular, the errors in Euler's

method when estimating mean values are proportional to Δt rather than $(\Delta t)^{1/2}$. Indeed, there are two kinds of approximation commonly discussed in computational solution of stochastic differential equations [69, 70, 100]. A numerical method is said to be a strong approximation of order γ if

$$\|X(T) - X_N\|_{RV} \leq c(\Delta t)^{\gamma}$$

for some constant $c > 0$ where $X(T)$ is the exact solution at time T and X_N is the approximate solution using step length $\Delta t = T/N$. Euler's method and Milstein's methods have strong orders $\frac{1}{2}$ and 1, respectively. However, if expectations of functions of a solution to a stochastic differential equation are desired and not necessarily the pathwise approximation provided by a strong approximation, then a weak numerical method may be sufficient. An approximation X_N is said to converge weakly with order β if there exists a constant $c > 0$ such that

$$|E(F(X(T))) - E(F(X_N))| \leq c(\Delta t)^{\beta}$$

for all smooth functions F, where $\Delta t = T/N$ is the step size of the numerical method. It can be shown [69, 111] that both Euler's method and Milstein's method have weak order 1.

In addition, both of these numerical methods possess another important property. That is, their error expansions in the weak sense have the correct form for applying Richardson extrapolation [69, 76, 110, 111]. Specifically, the weak error for Euler's or Milstein's method has been shown to have the form

$$E(F(X(T))) - E(F(X_N)) = c_1\Delta t + c_2(\Delta t)^2 + c_3(\Delta t)^3 + \ldots,$$

where $\Delta t = T/N$ and $c_1, c_2, c_3, \ldots$ are constants independent of Δt. This result implies that several approximations using different values of N can be applied to obtain a higher order approximation. Suppose that $E(F(X_N))$, $E(F(X_{2N}))$, and $E(F(X_{4N}))$ are three approximations to $E(F(X(T)))$ using step lengths of T/N, $T/2N$, and $T/4N$ in Euler's method or in Milstein's method. To obtain an approximation to $E(F(X(T)))$ of order $(\Delta t)^2$, let

$$E(F(X(T))) - \big[2E(F(X_{2N})) - E(F(X_N))\big] = \hat{c}_2(\Delta t)^2 + \hat{c}_3(\Delta t)^3 + \ldots.$$

To obtain an approximation to $E(F(X(T)))$ of order $(\Delta t)^3$, let

$$E(F(X(T))) - \big[8E(F(X_{4N})) - 6E(F(X_{2N})) + E(F(X_N))\big]/3 = \tilde{c}_3(\Delta t)^3 + \ldots.$$

Example 4.10. Richardson extrapolation of Euler's method

Referring to the values in Table 4.3 for Example 4.9, the following approximations to $E((X(1))^2)$ are obtained using Euler's method:

$$E((X_{64})^2) = 810.15, \quad E((X_{128})^2) = 840.89, \quad \text{and} \quad E((X_{256})^2) = 855.33.$$

To obtain $O((\Delta t)^2)$ and $O((\Delta t)^3)$ approximations, respectively, to $E((X(1))^2)$ we calculate

$$2E((X_{128})^2) - E((X_{64})^2) = 871.63$$

and

$$[8E((X_{256})^2) - 6E((X_{128})^2) + E((X_{64})^2)]/3 = 869.15.$$

As $E((X(1))^2) = 869.02$ exactly, the original Euler approximations are much improved through extrapolation.

It is useful to note that any strong approximation is also a weak approximation. This can be seen using the Lyapunov inequality as

$$\begin{aligned} |E(F(X(T))) - E(F(X_N))| &\le (E|(F(X(T))) - (F(X_N))|^2)^{1/2} \\ &\le L(E|X(T) - X_N|^2)^{1/2} \;=\; L\|X(t) - X_N\|_{RV} \end{aligned}$$

assuming that F satisfies a Lipschitz condition. However, there are weak methods which are not strong approximations. Consider the stochastic differential equation (4.2) and consider the discrete process described in Example 2.4. For a particular trajectory, suppose that at time t_i, $X_i = m\delta$ for some integer m, where $\delta > 0$ is small. Define the three possibilities at time $t_{i+1} = t_i + \Delta t$ as

$$\begin{cases} X_{i+1} = X_i + \delta & \text{with probability} \quad r(t_i, X_i)\Delta t/\delta^2, \\ X_{i+1} = X_i & \text{with probability} \quad 1 - r(t_i, X_i)\Delta t/\delta^2 - s(t_i, X_i)\Delta t/\delta^2, \\ X_{i+1} = X_i - \delta & \text{with probability} \quad s(t_i, X_i)\Delta t/\delta^2, \end{cases}$$

where

$$\begin{cases} r(t_i, X_i) = \big(f(t_i, X_i)\delta + g^2(t_i, X_i)\big)/2 \\ s(t_i, X_i) = \big(-f(t_i, X_i)\delta + g^2(t_i, X_i)\big)/2. \end{cases}$$

It was seen in Example 2.4 the probability distribution of X_N approaches that of $X(T)$ as $\Delta t, \delta \to 0$ implying that $E(F(X_N)) \approx E(F(X(T)))$ for small values of Δt and δ.

Example 4.11. Weak but not strong approximation

Consider the weak method of Example 2.4, where a discrete stochastic process is used to approximate expectations of the solution to a stochastic differential equation. Apply this weak method to the problem that was described in Example 4.6. Let $\Delta t = 1/N$, where $N = 4096, 8192, 16384$, and 32768 and let $\delta = 0.1$. In Table 4.4, the calculational results are presented using 100,000 sample paths for estimating $E(X(1))$ and $E(X(1))^2$. Recalling that $E(X(1)) = 28.00$ and $E(X(1))^2 = 869.02$, the calculational results are reasonable.

Finally, it is worthwhile to point out that rounding errors can be significant in computational solution of stochastic differential equations just as they are in computational solution of ordinary differential equations [18, 70]. Specifically, as Δt is taken smaller and smaller to reduce the method error, rounding errors,

Table 4.4. Calculated estimates of $E(X(1))$ and $E(X(1))^2$

Value of N	$E(X(1))$ Estimate	$E(X(1))^2$ Estimate
2^{12}	17.04	292.42
2^{13}	24.64	630.81
2^{14}	27.88	854.12
2^{15}	27.97	868.12

due to the finite number of digits used in the computer arithmetic, begin to increase. For an important calculation or in a calculation where Δt is very small, the magnitude of the rounding errors can be checked (and reduced) by performing the identical calculation on the computer using higher precision arithmetic.

4.6 Systems of Stochastic Differential Equations

Systems of stochastic differential equations are common in applications. Itô's formula and numerical methods can be extended to systems. Let

$$\mathbf{X}(t,\omega) = [X_1(t,\omega), X_2(t,\omega), \ldots, X_d(t,\omega)]^T$$

$$\mathbf{W}(t,\omega) = [W_1(t,\omega), W_2(t,\omega), \ldots, W_m(t,\omega)]^T$$

$$\mathbf{f} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$$

and

$$g : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m},$$

where $W_i(t,\omega), 1 \le i \le m$ are independent Wiener processes. Then a system of stochastic differential equations has the form

$$d\mathbf{X}(t,\omega) = \mathbf{f}(t,\mathbf{X}(t,\omega))\,dt + g(t,\mathbf{X}(t,\omega))\,d\mathbf{W}(t,\omega). \tag{4.6}$$

In component form, the system is

$$X_i(t) = X_i(0) + \int_0^t f_i(s,\mathbf{X}(s))\,ds + \sum_{j=1}^{m}\int_0^t g_{i,j}(s,\mathbf{X}(s))\,dW_j(s)$$

for $i = 1, 2, \ldots, d$.

Itô's formula can be generalized to systems. Let

$$\mathbf{F} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^k \quad \text{and let} \quad \mathbf{Y}(t,\omega) = \mathbf{F}(t,\mathbf{X}(t,\omega)).$$

Then the pth component of $\mathbf{Y}(t,\omega)$ satisfies [70]:

$$dY_p(t) = \left[\frac{\partial F_p}{\partial t} + \sum_{i=1}^{d} f_i \frac{\partial F_p}{\partial x_i} + \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{l=1}^{m} \frac{1}{2} g_{i,l} g_{j,l} \frac{\partial^2 F_p}{\partial x_i \partial x_j} \right] dt + \sum_{l=1}^{m} \sum_{i=1}^{d} g_{i,l} \frac{\partial F_p}{\partial x_i} dW_l(t) \tag{4.7}$$

for $p = 1, 2, \ldots, k$.

Example 4.12. Itô's formula for a stochastic problem with $d = 1$ and $m = 2$

Consider the stochastic differential equation

$$\begin{cases} dX(t) = t^2 X(t)\, dt + t\, dW_1(t) + X(t)\, dW_2(t), \quad 0 \le t \le T \\ X(0) = 1, \end{cases}$$

where $d = 1$ and $m = 2$. For this problem, $f_1 = t^2 X, g_{1,1} = t$, and $g_{1,2} = X$. Consider using Itô's formula to find the stochastic differential equation for $F = X^2$. Applying Itô's formula, we obtain

$$\begin{cases} d(X^2(t)) = \left[2t^2 X^2(t) + t^2 + X^2(t)\right] dt + 2tX(t)\, dW_1(t) + 2X^2(t)\, dW_2(t) \\ X^2(0) = 1. \end{cases}$$

Euler's method and Milstein's method can also be generalized to a system of stochastic differential equations. Euler's method for systems is straightforward and has the form

$$\mathbf{X}_{n+1}(\omega) = \mathbf{X}_n(\omega) + \mathbf{f}(t_n, \mathbf{X}_n(\omega))\Delta t + g(t_n, \mathbf{X}_n(\omega))\Delta \mathbf{W}_n(\omega) \tag{4.8}$$

for $n = 0, 1, 2, \ldots, N$, where $\mathbf{X}_n(\omega) \approx \mathbf{X}(t_n, \omega)$, $\Delta t = T/N$, $\Delta \mathbf{W}_n = \mathbf{W}(t_{n+1}) - \mathbf{W}(t_n)$. In component form, Euler's method is

$$X_{i,n+1}(\omega) = X_{i,n}(\omega) + f_i(t_n, \mathbf{X}_n(\omega))\Delta t + \sum_{j=1}^{m} g_{i,j}(t_n, \mathbf{X}_n(\omega))\Delta W_{j,n}(\omega) \tag{4.9}$$

for $i = 1, 2, \ldots, d$, where $\Delta W_{j,n} \sim N(0, \Delta t)$ for each j and n.

Milstein's method for multidimensional stochastic differential equations involves the double stochastic integral

$$I_n(j_1, j_2) = \int_{t_n}^{t_n + \Delta t} \int_{t_n}^{s} dW_{j_1}(r)\, dW_{j_2}(s).$$

Milstein's method has the componentwise form

$$X_{i,n+1}(\omega) = X_{i,n}(\omega) + f_i(t_n, \mathbf{X}_n(\omega))\Delta t + \sum_{j=1}^{m} g_{i,j}(t_n, \mathbf{X}_n(\omega))\Delta W_{j,n}(\omega) + \sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \sum_{l=1}^{d} g_{l,j_1} \frac{\partial g_{i,j_2}}{\partial x_l} I_n(j_1, j_2) \tag{4.10}$$

for $i = 1, 2, \ldots, d$.

Example 4.13. Approximation of a stochastic problem with $d = 1$ and $m = 2$

Consider the stochastic differential equation

$$\begin{cases} dX(t) = t^2 X(t)dt + tdW_1(t) + X(t)dW_2(t), \quad 0 \le t \le T \\ X(0) = 1, \end{cases}$$

where $d = 1$ and $m = 2$. For this problem, Euler's method has the form

$$\begin{cases} X_{n+1} = X_n + t_n^2 X_n \Delta t + t_n \Delta W_{1,n} + X_n \Delta W_{2,n} \\ X_0 = 1, \end{cases}$$

for $n = 0, 1, 2, \ldots$, where $\Delta W_{1,n}, \Delta W_{2,n} \sim N(0, \Delta t)$ and $t_n = n\Delta t$. In addition, Milstein's method has the form

$$\begin{cases} X_{n+1} = X_n + t_n^2 X_n \Delta t + t_n \Delta W_{1,n} + X_n \Delta W_{2,n} + t_n I_n(1,2) + X_n I_n(2,2) \\ X_0 = 1, \end{cases}$$

for $n = 0, 1, 2, \ldots$.

It is useful to note that

$$I_n(j_1, j_1) = \int_{t_n}^{t_n + \Delta t} \int_{t_n}^{s} dW_{j_1}(r)\, dW_{j_1}(s) = \frac{1}{2}\big((\Delta W_{j_1,n})^2 - \Delta t\big)$$

but $I_n(j_1, j_2)$ for $j_1 \neq j_2$ does not have an analytical form and must be approximated. This multiple integral can be approximated by a Fourier series expansion [70]. Also, note by Example 3.10, if $[t_n, t_{n+1}]$ is divided into M equal intervals with $t_{j,n} = t_n + j\Delta t/M$ for $j = 0, 1, \ldots, M$, then [90]:

$$I_n(j_1, j_2) \approx \tilde{I}_n(j_1, j_2) = \sum_{j=0}^{M-1} [W_{j_1}(t_{j,n}) - W_{j_1}(t_{0,n})][W_{j_2}(t_{j+1,n}) - W_{j_2}(t_{j,n})].$$

It is shown in Example 3.10 that $E|I_n - \tilde{I}_n|^2 = (\Delta t)^2/(2M)$.

4.7 Forward Kolmogorov (Fokker-Planck) Equation

In Chapter 2, it was shown that the probability distribution of solutions to a discrete-valued continuous stochastic process satisfies a system of differential equations called the forward Kolmogorov equations. An analogous result holds for the probability distribution of solutions to a stochastic differential equation. A derivation of the forward Kolmogorov equation is sketched in this section. Rigorous proofs of the forward Kolmogorov equation are presented, for example, in [40, 42].

Consider the stochastic differential equation

$$dX(t) = f(t, X(t))\, dt + g(t, X(t))\, dW(t)$$

and let $F \in \mathbb{C}_0^\infty(\mathbb{R})$. Applying Itô's formula to $F(X)$, it follows that

$$dF(X) = \left(f(t,X)\frac{\partial F(t,X)}{\partial x} + \frac{1}{2}g^2(t,X)\frac{\partial^2 F(t,X)}{\partial x^2} \right) dt + g(t,X)\frac{\partial F(t,X)}{\partial x}\, dW(t).$$

Because

$$E\int_0^t g(s,X(s))\frac{\partial F(X(s))}{\partial x}\, dW(s) = 0,$$

then

$$\frac{dE(F)}{dt} = E\left[\frac{\partial F}{\partial x}f + \frac{1}{2}g^2\frac{\partial^2 F}{\partial x^2}\right].$$

If $p(t,x)$ is the probability density for solutions to the stochastic differential equation, the above result implies that

$$\frac{d}{dt}\int_{-\infty}^{\infty} p(t,x)F(x)\, dx = \int_{-\infty}^{\infty} p(t,x)\left[\frac{\partial F}{\partial x}f + \frac{1}{2}g^2\frac{\partial^2 F}{\partial^2 x}\right] dx.$$

Integrating by parts the right-hand side of the preceding equation and assuming that the integral and the derivative can be interchanged on the left-hand side yields the relation

$$\int_{-\infty}^{\infty} F(x)\left[\frac{\partial p(t,x)}{\partial t} + \frac{\partial(p(t,x)f(t,x))}{\partial x} - \frac{1}{2}\frac{\partial^2(p(t,x)g^2(t,x))}{\partial^2 x}\right] dx = 0$$

for every $F \in \mathbb{C}_0^\infty(\mathbb{R})$. As the above integral holds for every function $F \in \mathbb{C}_0^\infty(\mathbb{R})$, this implies that

$$\frac{\partial p(t,x)}{\partial t} = -\frac{\partial(p(t,x)f(t,x))}{\partial x} + \frac{1}{2}\frac{\partial^2(p(t,x)g^2(x,t))}{\partial^2 x} \tag{4.11}$$

with $p(0,x) = p_0(x)$. Equation (4.11) is the forward Kolmogorov equation or Fokker-Planck equation for the probability distribution of solutions to stochastic differential equation (4.1). This equation will be useful in the next chapter. Furthermore, the forward Kolmogorov equation for the system of stochastic differential equations (4.4) has the form

$$\frac{\partial p(t,\mathbf{x})}{\partial t} = -\sum_{i=1}^{d}\frac{\partial\big[p(t,\mathbf{x})f_i(t,\mathbf{x})\big]}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{l=1}^{m}\frac{\partial^2}{\partial x_i\partial x_j}\big[g_{i,l}(t,\mathbf{x})g_{j,l}(t,\mathbf{x})p(t,\mathbf{x})\big]. \tag{4.12}$$

Example 4.14. Solution of a Fokker-Planck equation

Consider the stochastic differential equation

$$\begin{cases} dX(t) = a\,dt + b\,dW(t) \\ X(0) = x_0. \end{cases}$$

The probability density of the solutions satisfies the forward Kolmogorov equation

$$\begin{cases} \dfrac{\partial p(t,x)}{\partial t} = -\dfrac{\partial (ap(t,x))}{\partial x} + \dfrac{b^2}{2}\dfrac{\partial^2 (p(t,x))}{\partial^2 x} \\ p(0,x) = \delta(x - x_0). \end{cases}$$

The solution to this partial differential equation is

$$p(t,x) = \frac{1}{\sqrt{2\pi b^2 t}} \exp\left(\frac{-(x - at - x_0)^2}{2b^2 t}\right).$$

4.8 Stability

In this section, stability of stochastic differential equations is introduced. There are several kinds of stability questions and several ways to define stability for stochastic differential equations. To introduce this topic, it is useful to first review stability concepts for ordinary differential equations. Consider the initial-value problem:

$$\begin{cases} \dfrac{d\mathbf{y}(t)}{dt} = \mathbf{f}(\mathbf{y}(t)), \quad \text{for} \quad t > 0 \\ \mathbf{y}(0) = \mathbf{a} \end{cases} \tag{4.13}$$

where $\mathbf{y} : \mathbb{R} \to \mathbb{R}^n$ and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$. Suppose that $\mathbf{z}(t)$ satisfies the same differential equation as $\mathbf{y}(t)$ but with a different initial condition, i.e.,

$$\begin{cases} \dfrac{d\mathbf{z}(t)}{dt} = \mathbf{f}(\mathbf{z}), \quad \text{for} \quad t > 0 \\ \mathbf{z}(0) \neq \mathbf{a}. \end{cases}$$

Suppose that $\mathbf{a} = \boldsymbol{\gamma}$ is a critical point of the differential equation, i.e., $\mathbf{f}(\boldsymbol{\gamma}) = \mathbf{0}$. Then the solution of (4.13) satisfies $\mathbf{y}(t) = \boldsymbol{\gamma}$ for $t \geq 0$. The initial-value problem is said to be stable [23] at $\boldsymbol{\gamma}$ if given $\epsilon > 0$ there is a $\delta > 0$ such that

$$\|\mathbf{z}(t) - \boldsymbol{\gamma}\| < \epsilon \quad \text{for} \quad t \geq 0 \quad \text{whenever} \quad \|\mathbf{z}(0) - \boldsymbol{\gamma}\| < \delta.$$

That is, small changes in the initial condition do not produce large changes in the solution for $t \geq 0$.

Furthermore, in computational solution of initial-value problems for ordinary differential equations, there are two common numerical stability concepts. Suppose that a single-step method for solving (4.13) has the form:

$$\begin{cases} \mathbf{y}_{k+1} = \mathbf{y}_k + h\,\boldsymbol{\phi}(h, \mathbf{y}_k), \quad \text{for} \quad k = 0, 1, 2, \ldots, N-1 \\ \mathbf{y}_0 = \mathbf{a}. \end{cases} \tag{4.14}$$

where $h = T/N$ is the step length, $t_k = kh$, and $\mathbf{y}_k \approx \mathbf{y}(t_k)$ for each $0 \leq k \leq N$. The method (4.14) is numerically stable if small changes in the initial condition do not produce large changes in the computational solution. Specifically, if $\mathbf{z}_k$ for $k = 0, 1, \ldots, N$ satisfies (4.14) but with a different initial condition $\mathbf{z}_0 \neq \mathbf{y}_0$, then the numerical scheme is numerically stable provided that there is a constant $c > 0$ such that

$$\|\mathbf{y}_k - \mathbf{z}_k\| \leq c\epsilon \quad \text{for} \quad 0 \leq k \leq N \quad \text{when} \quad \|\mathbf{y}_0 - \mathbf{z}_0\| < \epsilon.$$

If $\boldsymbol{\phi}$ satisfies an appropriate Lipschitz condition, then the numerical scheme can be shown to be stable [28]. However, the constant c can be extremely large, especially for stiff systems, which motivates another concept of numerical stability. To study stability of stiff systems, the following scalar test problem is studied:

$$\begin{cases} \dfrac{dy(t)}{dt} = \lambda y, & \text{for} \quad t > 0 \\ y(0) = a \end{cases} \tag{4.15}$$

where λ is a constant. Clearly, if $a \neq 0$, then $y(t) \to 0$ as $t \to \infty$ if and only if $Re(\lambda) < 0$. Now consider, for example, applying Euler's method to this test problem. Then

$$\begin{cases} y_{k+1} = (1 + h\lambda)y_k, & \text{for} \quad k = 0, 1, 2, \ldots, \\ y_0 = a. \end{cases}$$

and $y_k \to 0$ as $k \to \infty$, if and only if $-2 < Re(\lambda h) < 0$. The region of absolute stability of Euler's method is $-2 < Re(\lambda h) < 0$. The region of absolute stability gives a condition on the step length. If the method satisfies this condition, then the numerical solution does not "blow up" but decreases to zero behaving like the solution to the initial-value problem. For Euler's method to behave similarly to the solution of (4.15) for a large negative value of λ, the step length h must be selected to be very small. However, for the backward Euler method, which for the test problem has the form:

$$\begin{cases} y_{k+1} = y_k + h\lambda y_{k+1}, & \text{for} \quad k = 0, 1, 2, \ldots, \\ y_0 = a, \end{cases}$$

the region of absolute stability includes the entire left-half of the complex plane, i.e. $-\infty < Re(\lambda h) < 0$, and the backward Euler method is said to be A-stable. For the backward Euler method, the step length h need not be chosen very small for the numerical solution to perform similarly to the actual solution even for an initial-value problem that involves a large negative value of λ. The concept of absolute stability is particularly useful when considering numerical solution of systems. For the test initial-value system

$$\begin{cases} \dfrac{d\mathbf{y}(t)}{dt} = A\mathbf{y}, & \text{for} \quad 0 \leq t \leq T \\ \mathbf{y}(0) = \mathbf{a}. \end{cases}$$

where A is an $n \times n$ matrix, then $\mathbf{y}(t) \to \mathbf{0}$ as $t \to \infty$ provided that $Re(\lambda_i) < 0$ for each eigenvalue λ_i for $1 \le i \le n$. The system is said to be stiff if $\min_{1 \le i \le n} \{Re(\lambda_i)\}$ is much less than $\max_{1 \le i \le n} \{Re(\lambda_i)\}$. For this problem, Euler's method has the form

$$\begin{cases} \mathbf{y}_{k+1} = (I + Ah)\mathbf{y}_k, & \text{for} \quad k = 0, 1, 2, \ldots, \\ \mathbf{y}_0 = \mathbf{a}. \end{cases}$$

The eigenvalues of $I + Ah$ are $1 + \lambda_i h$ for $i = 1, 2, \ldots, n$ and $\mathbf{y}_k \to \mathbf{0}$ as $k \to \infty$ provided that $-2 < Re(\lambda_i)h < 0$ for each eigenvalue λ_i for $1 \le i \le n$. Hence, although the behavior of the actual solution for large time t is determined by the eigenvalues with small negative real parts, the step length h of the numerical method is forced for all time to satisfy a condition determined by the eigenvalues with large negative real parts. Implicit methods with large regions of absolute stability, such as the backward Euler method, are often recommended for stiff systems as such methods give stable calculational results with moderate values of step length.

Now consider stability for stochastic differential equations. First, stability of a steady solution to a stochastic differential equation is studied then numerical stability of an approximation is studied. Consider stability of a steady solution for the stochastic differential equation

$$\begin{cases} dX(t) = f(X(t))\, dt + g(X(t))\, dW(t), & \text{for} \quad 0 \le t \le T \\ X(0) = a. \end{cases} \tag{4.16}$$

It is supposed that $f(0) = g(0) = 0$ so that $X(t) \equiv 0$ is a steady solution of (4.16).

There are many ways to define stochastic stability for a steady solution of a stochastic differential equation [59, 60, 69, 79, 97, 99, 102]. Two ways will be considered in this section, asymptotic stochastic stability (in the large) and mean-square stability [59, 60]. It is assumed that $X(0) \neq 0$. If $\lim_{t \to \infty} |X(t)| = 0$ with probability 1, then $X(t) \equiv 0$ is said to be asymptotically stochastically stable. If $\lim_{t \to \infty} E(|X(t)|^2) = 0$, then $X(t) \equiv 0$ is said to be mean-square stable. It is interesting that some stochastic differential equations may be both asymptotically stochastically stable and mean-square stable while others may be asymptotically stochastically stable but not mean-square stable. To illustrate this behavior [59], stability is analyzed for a stochastic differential equation with linear drift and diffusion coefficients, namely, $f(X) = \lambda X$ and $g(X) = \mu X$. In this case, (4.16) becomes:

$$\begin{cases} dX(t) = \lambda X(t)\, dt + \mu X(t)\, dW(t), & \text{for} \quad 0 \le t \le T \\ X(0) = a \end{cases} \tag{4.17}$$

and $E(X(t)) = X(0) \exp(\lambda t)$. Using Itô's formula, $X^2(t)$ satisfies the stochastic differential equation

$$\begin{cases} d(X^2(t)) = \left(2\lambda X^2(t) + \mu^2 X^2(t)\right) dt + 2\mu X^2(t)\, dW(t), \quad \text{for} \quad t > 0 \\ X^2(0) = a^2. \end{cases}$$

It follows that $E(X^2(t))$ satisfies the differential equation

$$\begin{cases} d(E(X^2(t))) = \left(2\lambda E(X^2(t)) + \mu^2 E(X^2(t))\right) dt, \quad \text{for} \quad t > 0 \\ E(X^2(0)) = a^2. \end{cases}$$

and the solution $E(X^2(t))$ is found to be

$$E(X^2(t)) = X^2(0) \exp((2\lambda + \mu^2)t).$$

This solution implies that the steady solution $X(t) = 0$ is mean-square stable if and only if $\lambda + \mu^2/2 < 0$. Now consider Itô's formula applied to $\ln(X(t))$. Then,

$$\begin{cases} d(\ln(X(t)) = (\lambda - \mu^2/2)\, dt + \mu\, dW(t), \quad \text{for} \quad t > 0 \\ \ln(X(0)) = \ln(a). \end{cases}$$

Let Δt be a given interval width and let $t_i = i\Delta t$ for $t = 0, 1, 2, \ldots$. This stochastic differential equation can be exactly integrated from t_i to t_{i+1} to yield:

$$\ln(X(t_{i+1})) - \ln(X(t_i)) = (\lambda - \mu^2/2)\ (t_{i+1} - t_i) + \mu\eta_i\ \sqrt{(t_{i+1} - t_i)}$$

where $\eta_i \sim N(0, 1)$ for $i = 0, 1, 2, \ldots$. Thus,

$$\ln\left(\frac{X(t_{i+1})}{X(t_i)}\right) = (\lambda - \mu^2/2)\ \Delta t + \mu\eta_i\ \sqrt{\Delta t}$$

and

$$E\left(\ln\left(\frac{X(t_{i+1})}{X(t_i)}\right)\right) = (\lambda - \mu^2/2)\ \Delta t.$$

Let $S_n = \sum_{i=0}^{n-1} \ln\left(\frac{X(t_{i+1})}{X(t_i)}\right)$. By the Law of Large Numbers see Section 1.7),

$$\frac{S_n}{n} = \frac{1}{n} \sum_{i=0}^{n-1} \ln\left(\frac{X(t_{i+1})}{X(t_i)}\right) \ \rightarrow\ (\lambda - \mu^2/2)\ \Delta t \quad \text{w.p.1 as} \quad n \rightarrow \infty.$$

But, letting $t = t_n$,

$$\begin{aligned} \frac{1}{n\Delta t} \sum_{i=0}^{n-1} \ln\left(\frac{X(t_{i+1})}{X(t_i)}\right) &= \frac{1}{n\Delta t} \ln\left(\frac{X(t_n)}{X(0)}\right) \\ &= \frac{1}{t} \ln\left(\frac{X(t)}{X(0)}\right) \\ &\rightarrow\ (\lambda - \mu^2/2) \quad \text{w.p.1 as} \quad t \rightarrow \infty. \end{aligned}$$

Therefore,

$$X(t) \;\to\; X(0)\,\exp((\lambda - \mu^2/2)t) \quad \text{w.p.1 as} \quad t \to \infty.$$

This result implies that the steady solution $X(t) = 0$ is asymptotically stochastically stable if and only if $\lambda - \mu^2/2 < 0$. Hence, for example, if $\lambda = \mu^2/4$ in stochastic differential equation (4.17), then $X(t) \to 0$ with probability 1 as $t \to \infty$ while $E(X(t)) \to \infty$ and $E(X^2(t)) \to \infty$ under the same condition.

Now, numerical stability of stochastic differential equations is considered, in particular, with respect to stiff stochastic problems with additive noise and then, more briefly, with respect to multiplicative noise. (See, for example, [59, 60, 69, 97, 98, 99, 102] for thorough discussions of numerical stability for stochastic differential equations.) The test problem for additive noise has the form

$$\begin{cases} dX(t) = \lambda X(t)\,dt + \mu\,dW(t), & \text{for} \quad t > 0 \\ X(0) = a. \end{cases} \tag{4.18}$$

Two kinds of numerical stochastic stability are numerical asymptotic stochastic stability and numerical mean-square stability. Let X_k and $\tilde{X}_k$ be two approximations of (4.18) with the same numerical method but with different initial values. If $\lim\limits_{k\to\infty} |X_k - \tilde{X}_k| = 0$ with probability 1, then the approximation is said to be asymptotically stochastically stable [59, 60]. If $\lim\limits_{k\to\infty} E(|X_k - \tilde{X}_k|^2) = 0$, then the approximation is said to be mean-square stable [59, 60].

Consider first Euler's method for solution of test problem (4.16):

$$\begin{cases} X_{k+1} = X_k + \lambda X_k\,h + \mu\,\eta_k\sqrt{h}, & \text{for} \quad k = 0, 1, \ldots \\ X_0 = a \end{cases} \tag{4.19}$$

where $X_k \approx X(kh)$, $\eta_k \sim N(0,1)$ for each k, and h is the step length. Furthermore, let $\tilde{X}_k$ be another numerical approximation computed using (4.19) but with a different initial approximation $\tilde{X}_0 = \tilde{a}$. Let $Z_k = X_k - \tilde{X}_k$. Then, Z_k satisfies

$$\begin{cases} Z_{k+1} = Z_k + \lambda h\,Z_k, & \text{for} \quad k = 0, 1, \ldots \\ Z_0 = a - \tilde{a} \end{cases}$$

and therefore,

$$|X_k - \tilde{X}_k| = |Z_k| = |1 + \lambda h|^k\,|Z_0| \quad \text{for} \quad k = 0, 1, \ldots.$$

Thus, Euler's method is asymptotically and mean square stable for the test problem provided that $-2 < \lambda h < 0$. An analogous result holds for stability of Euler's method for stiff systems with additive noise. Specifically, Euler's method is numerically stable for a stochastic system with additive noise

$$\begin{cases} d\mathbf{X}(t) = A\mathbf{X}(t)\,dt + \mu\,d\mathbf{W}(t), & \text{for} \quad t > 0 \\ \mathbf{X}(0) = \mathbf{a}. \end{cases}$$

provided that $-2 < Re(\lambda_i)h < 0$ for each eigenvalue λ_i of A.

Now consider a stochastic trapezoidal method [59, 69] that when applied to (4.16) has the form

$$\begin{cases} X_{k+1} = X_k + \frac{1}{2} f(X_k)\, h + \frac{1}{2} f(X_{k+1})\, h + g(X_k)\eta_k \sqrt{h}, \quad \text{for} \quad k = 0, 1, \ldots \\ X_0 = a. \end{cases}$$

When applied to test problem (4.18), this implicit method reduces to the form

$$\begin{cases} X_{k+1} = X_k + \frac{1}{2}\lambda h X_k + \frac{1}{2}\lambda h X_{k+1} + \mu\eta_k \sqrt{h}, \quad \text{for} \quad k = 0, 1, \ldots \\ X_0 = a. \end{cases}$$

Let $Z_k = X_k - \tilde{X}_k$ where X_k and $\tilde{X}_k$ be two approximations with the trapezoidal method applied to the test problem (4.18). Then, Z_k satisfies

$$Z_{k+1} = \frac{1 + \frac{1}{2}\lambda h}{1 - \frac{1}{2}\lambda h} Z_k \quad \text{for} \quad k = 0, 1, \ldots$$

and hence,

$$|X_k - \tilde{X}_k| = |Z_k| = \left| \frac{1 + \frac{1}{2}\lambda h}{1 - \frac{1}{2}\lambda h} \right|^k |Z_0| \quad \text{for} \quad k = 0, 1, \ldots.$$

It follows that the trapezoidal method is numerically mean square stable for the test problem for any $\lambda h < 0$.

Now consider stability of numerical methods applied to stochastic differential equations with multiplicative noise. Numerical mean-square stochastic stability for multiplicative noise is considered here. (A good explanation of numerical asymptotic stochastic stability is given in [59] for a family of numerical methods.) The test problem for multiplicative noise has the form

$$\begin{cases} dX(t) = \lambda X(t)\, dt + \mu X(t)\, dW(t), \quad \text{for} \quad t > 0 \\ X(0) = a. \end{cases} \tag{4.20}$$

Using Itô's formula, it is straightforward to show that $E(|X(t)|^2) \to 0$ as $t \to \infty$ provided that $\mu^2 + 2\lambda < 0$ and a required condition for the steady solution $X(t) = 0$ to be mean square stable is that $\mu^2 + 2\lambda < 0$. This condition will be assumed on the values of the parameters μ and λ in studying the step size h required for a numerical approximation to likewise approach 0. Consider first Euler's method for solution of test problem (4.20):

$$\begin{cases} X_{k+1} = X_k + \lambda X_k\, h + \mu X_k\, \eta_k \sqrt{h}, \quad \text{for} \quad k = 0, 1, \ldots \\ X_0 = a, \end{cases} \tag{4.21}$$

where $X_k \approx X(kh)$, $\eta_k \sim N(0, 1)$ for each k, and h is the step length. Furthermore, let $\tilde{X}_k$ be another numerical approximation computed using

(4.21) but with a different initial approximation $\tilde{X}_0 = \tilde{a}$. Let $Z_k = X_k - \tilde{X}_k$. Then, Z_k satisfies

$$\begin{cases} Z_{k+1} = Z_k + \lambda h\, Z_k + \mu Z_k\, \eta_k \sqrt{h}, \quad \text{for} \quad k = 0, 1, \ldots \\ Z_0 = a - \tilde{a} \end{cases}$$

and therefore,

$$E(|X_k - \tilde{X}_k|^2) = E(|Z_k|^2) = (1 + 2\lambda h + \lambda^2 h^2 + \mu^2 h)^k\, E(|Z_0|^2) \text{ for } k = 0, 1, \ldots.$$

Thus, Euler's method is mean square stable for the test problem provided that $\mu^2 + 2\lambda < -\lambda^2 h$. Assuming fixed values for μ and λ, such that $\mu^2 + 2\lambda < 0$ for the steady solution $X(t) = 0$ to be mean square stable, the step size h in Euler's method can therefore be selected sufficiently small for the numerical method to be mean square stable.

Now consider a stochastic trapezoidal method [59, 69] that when applied to test problem (4.20) has the form

$$\begin{cases} X_{k+1} = X_k + \frac{1}{2}\lambda h X_k + \frac{1}{2}\lambda h X_{k+1} + \mu X_k \eta_k \sqrt{h}, \quad \text{for} \quad k = 0, 1, \ldots \\ X_0 = a. \end{cases}$$

Let $Z_k = X_k - \tilde{X}_k$ where X_k and $\tilde{X}_k$ be two approximations with the trapezoidal method applied to the test problem (4.20). Then, $E(|Z_k|^2)$ satisfies

$$E(|X_k - \tilde{X}_k|^2) = E(|Z_k|^2) = \left| \frac{(1 + \frac{1}{2}\lambda h)^2 + \mu^2 h}{(1 - \frac{1}{2}\lambda h)^2} \right|^k E(|Z_0|^2) \text{ for } k = 0, 1, \ldots.$$

It follows that the trapezoidal method is numerically mean square stable for the test problem whenever $2\lambda h + \mu^2 h < 0$. That is, for any positive step size, the stochastic trapezoidal method and the stochastic differential equation share the same stability condition on λ and μ, namely $2\lambda + \mu^2 < 0$. Therefore, assuming that μ and λ satisfy $\mu^2 + 2\lambda < 0$ for the steady solution $X(t) = 0$ to be mean square stable, any positive step size h can be selected in the stochastic trapezoidal method so that the numerical solution behaves similarly to the exact solution.

Lyapunov functions [41, 69, 79] are also commonly used to study stability of differential equations (4.13) and (4.15) and are briefly considered here. The idea is to construct a function that behaves like energy in a mechanical system. If the system continually loses "energy," then the system eventually reaches an equilibrium state. Provided certain conditions hold, stability of the system can be inferred using a Lyapunov function. Consider, for example, the deterministic problem $dy(t) = -\frac{1}{2}(y(t) + y^3(t))dt$, $y(0) = a$. Choose the Lyapunov function, V, as $V(t) = y^2(t)$. It follows that V satisfies $dV(t) = (-V(t) - V^2(t))dt$, $V(0) = a^2$. As $dV(t) \le -V(t)dt$, then $V(t) \le a^2 e^{-t}$. Hence, $y^2(t) = V(t) \to 0$ as $t \to \infty$ implying that 0 is a stable fixed point for this problem. Now, consider the stochastic differential equation $dX(t) = -\frac{1}{2}(X(t) + X^3(t))\, dt +$

$\mu X(t)\,dW(t), \quad X(0) = a$. Choosing $V(t) = X^2(t)$, then by Itô's formula, V satisfies $dV(t) = ((-1+\mu^2)V(t) - V^2(t))\,dt + 2\mu V(t)\,dW(t), \;\; V(0) = a^2$. It follows that $d(E(V(t))) \leq (-1+\mu^2)E(V(t))dt$ and thus, $E(V(t)) \leq a^2 e^{(-1+\mu^2)t}$. This result implies that the steady solution $X(t) = 0$ is mean square stable provided that $-1 < \mu < 1$.

4.9 Parameter Estimation for Stochastic Differential Equations

In this section, a stochastic differential equation of the form

$$dX(t) = f(t, X(t); \boldsymbol{\theta})\,dt + g(t, X(t); \boldsymbol{\theta})\,dW(t) \tag{4.22}$$

is considered where $\boldsymbol{\theta} \in \mathbb{R}^m$ is a vector of parameters that are unknown. It is assumed that

$$x_0, x_1, x_2, \ldots, x_N$$

are observed values of $X(t)$ at the respective uniformly distributed times $t_i = i\Delta t$ for $i = 0, 1, \ldots, N$ where $\Delta t = T/N$. The problem is to find an estimate of the vector $\boldsymbol{\theta}$ given these $N + 1$ data points. Two estimation methods are considered in this section: a maximum likelihood estimation method and a nonparametric estimation method.

4.9.1 A maximum likelihood estimation method

Let $p(t_k, x_k | t_{k-1}, x_{k-1}; \boldsymbol{\theta})$ be the transition probability density of (t_k, x_k) starting from (t_{k-1}, x_{k-1}) given the vector $\boldsymbol{\theta}$. Suppose that the density of the initial state is $p_0(x_0|\boldsymbol{\theta})$.

In maximum likelihood estimation of $\boldsymbol{\theta}$ [36, 62, 64], the joint density

$$D(\boldsymbol{\theta}) = p_0(x_0|\boldsymbol{\theta}) \prod_{k=1}^{N} p(t_k, x_k | t_{k-1}, x_{k-1}; \boldsymbol{\theta})$$

is maximized over $\boldsymbol{\theta} \in \mathbb{R}^m$. The value of $\boldsymbol{\theta}$ that maximizes $D(\boldsymbol{\theta})$ will be denoted in this section as $\boldsymbol{\theta}^*$. However, to avoid small numbers on a computer, it is more convenient to minimize the function $L(\boldsymbol{\theta}) = -\ln(D(\boldsymbol{\theta}))$ which has the form

$$L(\boldsymbol{\theta}) = -\ln(p_0(x_0|\boldsymbol{\theta})) - \sum_{k=1}^{N} \ln(p(t_k, x_k | t_{k-1}, x_{k-1}; \theta)).$$

One difficulty in finding the optimal value $\boldsymbol{\theta}^*$ is that the transition densities are not generally known. However, by considering the Euler approximation to (4.22) and letting $X(t_{k-1}) = x_{k-1}$ at $t = t_{k-1}$, then

$$X(t_k) \approx x_{k-1} + f(t_{k-1}, x_{k-1}; \boldsymbol{\theta})\Delta t + g(t_{k-1}, x_{k-1}; \boldsymbol{\theta})\sqrt{\Delta t}\ \eta_k$$

where $\eta_k \sim N(0,1)$. This implies that

$$p(t_k, x_k | t_{k-1}, x_{k-1}; \boldsymbol{\theta}) \approx \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-(x_k - \mu_k)^2}{2\sigma_k^2}\right)$$

where $\mu_k = x_{k-1} + f(t_{k-1}, x_{k-1}; \boldsymbol{\theta})\Delta t$ and $\sigma_k = g(t_{k-1}, x_{k-1}; \boldsymbol{\theta})\sqrt{\Delta t}$. This transition density can be substituted into the expression for $L(\boldsymbol{\theta})$ which can subsequently be minimized over $\mathbb{R}^m$. The second difficulty is now computing the optimal vector $\boldsymbol{\theta}^*$ by minimizing $L(\boldsymbol{\theta})$. This is a nontrivial computation. A numerical optimization algorithm such as the Nelder-Mead method [82] can be useful for computing the minimum of $L(\boldsymbol{\theta})$.

In the above procedure, the transition densities were approximated using the Euler formula. Instead of using Euler's formula to approximate the transition density $p(t_k, x_k | t_{k-1}, x_{k-1}; \boldsymbol{\theta})$, one can approximate the density through simulation as discussed in [64]. First, for a given value of $\boldsymbol{\theta}$, one numerically solves (4.22) starting from x_{k-1} at t_{k-1}. A standard method can be used such as Euler's method or Milstein's method with one or more steps. This calculation is repeated M times to obtain M estimated values $y_1, y_2, \ldots, y_M$ for $X(t)$ at $t = t_k$. (Notice that the Wiener increments, used to find $y_1, y_2, \ldots, y_M$, are saved. These same Wiener increments are applied in finding $y_1, y_2, \ldots, y_M$ at t_k for each value of $\boldsymbol{\theta}$ tested. The problem of finding the optimal value $\boldsymbol{\theta}^*$ is then deterministic.) Then the transition density $p(t_k, x_k | t_{k-1}, x_{k-1}; \boldsymbol{\theta})$ is estimated using

$$p^{(M)}(t_k, x_k | t_{k-1}, x_{k-1}; \boldsymbol{\theta}) = \frac{1}{Mh} \sum_{j=1}^{M} K\left(\frac{x_k - y_j}{h}\right),$$

where K is a nonnegative kernel function and h is a bandwidth. A reasonable kernel K and bandwidth h are [64]:

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \quad \text{and} \quad h = 0.9sM^{-1/5},$$

where

$$s^2 = \frac{1}{M-1}\left[\sum_{j=1}^{M} y_j^2 - \frac{1}{M}\left(\sum_{j=1}^{M} y_j\right)^2\right]$$

gives a formula for the sample standard deviation. Next, one minimizes

$$L^{(M)}(\boldsymbol{\theta}) = -\ln\left(p_0(x_0|\boldsymbol{\theta})\right) - \sum_{k=1}^{N} \ln\left(p^{(M)}(t_k, x_k | t_{k-1}, x_{k-1}; \theta)\right)$$

to find an approximation to the optimal value $\boldsymbol{\theta}^*$.

Example 4.15. Fitting population data to an SDE

Consider the Aransas-Wood Buffalo population of whooping cranes [27, 36]. The population data for the whooping cranes are given in Table 4.5. These whooping cranes nest in Wood Buffalo National Park in Canada and winter in Aransas National Wildlife Refuge in Texas [27, 36]. The population size is graphed in Fig. 2.3 over the years 1939–1985. In this section, this data is fit to the stochastic differential equation

$$dX(t) = \theta_1 X(t)\,dt + \sqrt{\theta_2 X(t)}\,dW(t), \quad X(0) = 18,$$

where $X(t)$ is population size and $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ is to be determined. It is shown in the next chapter that this is a reasonable model for the population dynamics. Time t here means time in years from year 1939, for example, 1941 corresponds to $t = 2$.

Table 4.5. Aransas-Wood Buffalo whooping crane population

Year	Population	Year	Population	Year	Population	Year	Population
1939	18	1951	31	1963	32	1975	49
1940	22	1952	25	1964	33	1976	57
1941	26	1953	21	1965	42	1977	69
1942	16	1954	24	1966	44	1978	72
1943	19	1955	21	1967	43	1998	75
1944	21	1956	28	1968	48	1980	76
1945	18	1957	24	1969	50	1981	78
1946	22	1958	26	1970	56	1982	73
1947	25	1959	32	1971	57	1983	73
1948	31	1960	33	1972	59	1984	75
1949	30	1961	36	1973	51	1985	86
1950	34	1962	39	1974	49		

Both calculational methods described in this section gave optimal values of approximately $\theta_1^* = 0.0361$ and $\theta_2^* = 0.609$. The computer programs that computed these optimal values, using the procedures described in this section, are listed at the end of this chapter. Hence, a reasonable stochastic differential equation based on the data for the whooping crane population size is

$$dX(t) = 0.0361X(t)\,dt + \sqrt{0.609X(t)}\,dW(t), \quad X(0) = 18.$$

The mean population size and two different trajectories for this stochastic differential equation are plotted in Fig. 4.2. In addition, the actual whooping crane population in plotted in the figure. The graphs indicate that the stochastic differential equation model provides a reasonable fit to the data.

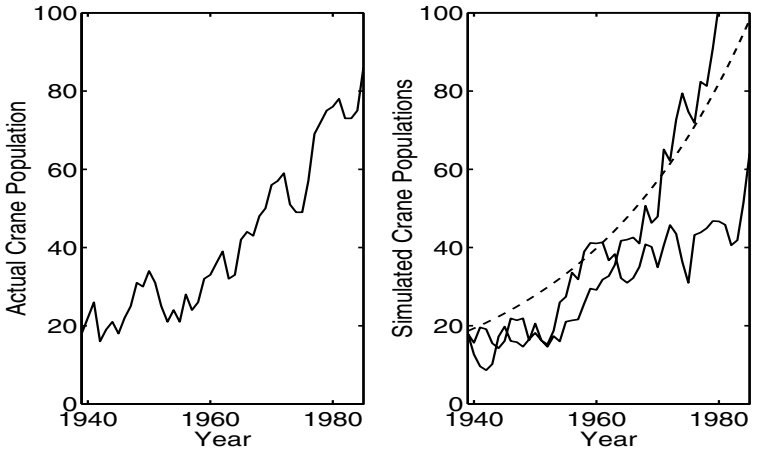

Fig. 4.2. Actual whooping crane population, mean population (dashed line), and two simulated crane populations using the SDE model

4.9.2 A nonparametric estimation method

Described in this section is a simple nonparametric estimation procedure considered in [109] and related to nonparametric methods described by several previous investigators (see, e.g., [30] or [53]). The estimates are only approximately correct but the approach is simple to implement and the approximation errors are likely to be small provided that reasonably frequent data are available [109]. Furthermore, the values estimated by this procedure may also be used as starting values in other computational methods [30, 53, 64] for estimating the parameters.

As in the previous section, it is assumed that

$$x_0, x_1, x_2, \ldots, x_N$$

are observed values of $X(t)$ at the respective uniformly distributed times $t_i = i\Delta t$ for $i = 0, 1, \ldots, N$ where $\Delta t = T/N$. The problem is to find an estimate of the parameter vector $\boldsymbol{\theta}$ given these $N+1$ data points. It is assumed in this estimation procedure that the process can be approximately modeled by the discrete-time process

$$X_{i+1} - X_i = f(t_i, X_i, \boldsymbol{\theta})\Delta t + g(t_i, X_i, \boldsymbol{\theta})\eta_i\sqrt{\Delta t} \tag{4.23}$$

for $i = 0, 1, \ldots, N-1$ where $\eta_i \sim N(0,1)$ for each i. (Notice that (4.23) is an Euler approximation to (4.22).) The method is based on the approximate expectations, conditional on the observed value of X_i, for the discrete-time process given by [109]:

$$E\left[(X_{i+1} - X_i)/\Delta t - f(t_i, X_i, \boldsymbol{\theta})\right] = O(\Delta t) \tag{4.24}$$

$$E\left[(X_{i+1} - X_i)^2/\Delta t - g^2(t_i, X_i, \boldsymbol{\theta})\right] = O(\Delta t). \tag{4.25}$$

If $\boldsymbol{\theta} \in \mathbb{R}^2$, then (4.24)–(4.25) can be used to estimate the parameter vector $\boldsymbol{\theta}$. If $\boldsymbol{\theta} \in \mathbb{R}^m$ for $m > 2$, then (4.24)–(4.25) may be supplemented with additional equations such as [30]:

$$E\left[(X_{i+1} - X_i)X_i/\Delta t - f(t_i, X_i, \boldsymbol{\theta})X_i\right] = O(\Delta t)$$
$$E\left[(X_{i+1} - X_i)^2 X_i/\Delta t - g^2(t_i, X_i, \boldsymbol{\theta})X_i\right] = O(\Delta t).$$

The parameter vector $\boldsymbol{\theta}$ is estimated using the sample counterparts of the previous equations. For example, for $\boldsymbol{\theta} \in \mathbb{R}^2$, $\boldsymbol{\theta}$ may be estimated using:

$$\sum_{i=0}^{N-1} f(t_i, x_i, \boldsymbol{\theta}) = \frac{1}{\Delta t} \sum_{i=0}^{N-1} (x_{i+1} - x_i) \tag{4.26}$$

$$\sum_{i=0}^{N-1} g^2(t_i, x_i, \boldsymbol{\theta}) = \frac{1}{\Delta t} \sum_{i=0}^{N-1} (x_{i+1} - x_i)^2. \tag{4.27}$$

Furthermore, this nonparametric estimation method, like the maximum likelihood procedure described in the previous section, can be extended to estimation of parameters for models that involve systems of stochastic differential equations. For additional information about parameter estimation and using numerical methods to study statistical properties of estimators see, for example, references [69, 70, 71].

Example 4.16. Fitting population data to an SDE

As an example of this nonparametric estimation method, consider the whooping crane population data described in the previous section where θ_1 and θ_2 are to be estimated for the stochastic differential equation model

$$dX(t) = \theta_1 X(t)\, dt + \sqrt{\theta_2 X(t)}\, dW(t), \quad X(0) = 18,$$

where $X(t)$ is population size. For this problem, population sizes are given in Table 4.5 at 47 successive years. In particular, $N = 46$ and $\Delta t = 1$. Applying equations (4.26)-(4.27) to this model gives the estimates:

$$\theta_1 = \frac{\sum_{i=0}^{45}(x_{i+1} - x_i)}{\sum_{i=0}^{45} x_i} = 0.0361 \quad \text{and} \quad \theta_2 = \frac{\sum_{i=0}^{45}(x_{i+1} - x_i)^2}{\sum_{i=0}^{45} x_i} = 0.579$$

which are close to the values obtained for θ_1 and θ_2 in the previous section using a maximum likelihood procedure. The model obtained in this example for the whooping crane population size is therefore

$$dX(t) = 0.0361X(t)\, dt + \sqrt{0.579X(t)}\, dW(t), \quad X(0) = 18.$$

Exercises

4.1. Assume that $X(t)$ satisfies the stochastic differential equation

$$dX(t) = aX(t)\,dt + \sqrt{bX(t)}\,dW(t), \quad X(0) = 1,$$

for $0 \le t \le T$ where a and b are positive constants. Use Itô's formula to find $E(X(t))$, $E(X^2(t))$, and $E(X^3(t))$ exactly.

4.2. Assume that $X(t)$ satisfies the stochastic differential equation

$$dX(t) = (a + bX(t))\,dt + c\,dW(t), \quad X(0) = X_0,$$

for $t \ge 0$ where a, b, and c are constants. Use Itô's formula with an appropriate value of α to show that

$$d(X(t)e^{\alpha t}) = ae^{\alpha t}\,dt + ce^{\alpha t}\,dW(t).$$

Then show that the exact solution is

$$X(t) = X_0 e^{-\alpha t} + \frac{a}{\alpha}(1 - e^{-\alpha t}) + ce^{-\alpha t}\int_0^t e^{\alpha s}\,dW(s).$$

4.3. Assume that $X(t)$ satisfies the stochastic differential equation

$$dX(t) = \left(\frac{2}{5}X^{3/5}(t) + 5X^{4/5}(t)\right)dt + X^{4/5}(t)\,dW(t), \quad X(0) = 1.$$

Use Itô's formula with $F(t, X(t)) = (X(t))^{1/5}$ to prove that the exact solution is

$$X(t) = \left(t + 1 + \frac{1}{5}W(t)\right)^5.$$

4.4. Let $X(t)$ be the solution of

$$X(t) = X(0) + \int_0^t f(X(s))\,ds + \int_0^t g\,dW(s)$$

for $0 \le t \le \Delta t$ where g is a constant and f satisfies conditions (c6) and (c7). In particular, $(f(X(s)) - f(X(r))^2 \le k(X(s) - X(r))^2$ and $E(X(s) - X(r))^2 \le c|s - r|$. Assume that $\hat{X}(\Delta t)$ satisfies

$$\hat{X}(\Delta t) = \hat{X}(0) + \int_0^{\Delta t} f(\hat{X}(\Delta t))\,ds + \int_0^{\Delta t} g\,dW(s).$$

(a) Using the triangle inequality, prove that

$$\|X(\Delta t) - \hat{X}(\Delta t)\|_{RV} \le \left\| \int_0^{\Delta t} \big(f(X(s)) - f(X(\Delta t))\big)\, ds \right\|_{RV}$$
$$+\|X(0) - \hat{X}(0)\|_{RV} + \left\| \int_0^{\Delta t} \big(f(X(\Delta t)) - f(\hat{X}(\Delta t))\big)\, ds \right\|_{RV}.$$

(b) Next, prove that

$$\|X(\Delta t) - \hat{X}(\Delta t)\|_{RV} \le e^{2k^{1/2}\Delta t}\Big(\|X(0) - \hat{X}(0)\|_{RV} + (kc/2)^{1/2}(\Delta t)^{3/2}\Big)$$

assuming that $2k^{1/2}\Delta t < 1$. (Note that $1/(1-x) \le e^{2x}$ for $0 \le x \le 1/2$.)

4.5. The solution of the stochastic differential equation

$$dX(t) = (\alpha + \beta X(t))\, dt + \gamma X(t)\, dW(t), \quad X(0) = X_0,$$

with α, β, and γ constants is [41]:

$$X(t) = R(t)\left(X_0 + \alpha \int_0^t (R(s))^{-1}\, ds\right),$$

where $R(t) = \exp\big((\beta - \gamma^2/2)t + \gamma W(t)\big)$. Use this result to show that the solution of

$$dX(t) = (aX(t) + bX^2(t))\, dt + cX(t)\, dW(t), \quad X(0) = X_0,$$

with a, b and c constants is

$$X(t) = U(t) \Big/ \left((X_0)^{-1} - b \int_0^t U(s)\, ds\right),$$

where $U(t) = \exp\big((a - c^2/2)t + cW(t)\big)$. To show this, let Y(t) = 1/X(t) and use Itô's formula with $F(t, X(t)) = 1/X(t)$.

4.6. Consider the stochastic differential equation

$$dX(t) = (aX(t) + bX^2(t))\, dt + cX(t)\, dW(t), \quad X(0) = 1/2,$$

for $0 \le t \le 1$ where $a = 1/4$, $b = 1/32$, and $c = 1/4$. Use Euler's method with $\Delta t = 1/200$ and 50,000 sample paths to estimate $E(X(1))$ and $E(X^2(1))$. Hand in a listing of your computer program along with your calculational results. In constructing your program, consider modifying a program listed in Chapter 4. Also, if the estimated values of $E(X(1))$ and $E(X^2(1))$ do not agree reasonably well with the results of Exercise 4.7, your program needs to be carefully checked.

4.7. Consider the stochastic differential equation

$$dX(t) = (aX(t) + bX^2(t))\,dt + cX(t)\,dW(t), \quad X(0) = 1/2,$$

for $0 \le t \le 1$ where $a = 1/4$, $b = 1/32$, and $c = 1/4$.
(a) Let $y_k(t) = E(X^k(t))$ for $k = 1, 2, 3$. Use Itô's formula to show that $y_1(t)$ and $y_2(t)$ satisfy the deterministic initial-value system:

$$\begin{cases} y_1'(t) = ay_1(t) + by_2(t), & y_1(0) = 1/2 \\ y_2'(t) = (2a + c^2)y_2(t) + 2by_3(t), & y_2(0) = 1/4. \end{cases}$$

(b) Assuming that $X(t)$ is approximately normally distributed with mean $\mu(t) = y_1(t)$ and variance $\sigma^2(t) = y_2(t) - y_1^2(t)$, show that $y_3(t) \approx 3y_1(t)y_2(t) - 2y_1^3(t)$. Then, modify the system in part (a) to

$$\begin{cases} y_1'(t) \approx ay_1(t) + by_2(t), & y_1(0) = 1/2 \\ y_2'(t) \approx (2a + c^2)y_2(t) + 6by_1(t)y_2(t) - 4by_1^3(t), & y_2(0) = 1/4. \end{cases}$$

(c) Use deterministic Euler's method with $\Delta t = 1/200$ to numerically solve the system in part (b). Calculate $y_1(1)$ and $y_2(1)$. Compare your calculated results with those of Exercise 4.6. If the estimated values of $E(X(1))$ and $E(X^2(1))$ do not agree reasonably well with the results of Exercise 4.6, your program needs to be carefully checked.

4.8. Let $X(t)$ be the solution of

$$X(t) = X_0 + \int_0^t f(X(s))\,ds + \int_0^t g\,dW(s)$$

for $0 \le t \le T$ where g is a constant and f satisfies conditions (c6) and (c7). Let $\Delta t = T/N$ where $2k^{1/2}\Delta t < 1$. Assume that X_i for $i = 0, 1, 2 \ldots, N$ satisfies the implicit numerical method:

$$X_{i+1} = X_i + f(X_{i+1})\,\Delta t + g\,\Delta W_i \quad \text{for} \quad i = 0, 1, 2 \ldots, N-1,$$

where $\Delta W_i = W(t_{i+1}) - W(t_i)$. Using the results of Exercise 4.4, prove that

$$\|X(T) - X_N\|_{RV} \le (\Delta t)^{1/2}(c/8)^{1/2}e^{4k^{1/2}T}.$$

Thus, the method has strong order 1/2 for this problem.

4.9. Consider the stochastic system

$$\begin{cases} dX_1(t) = X_2(t)\,dt + \sqrt{X_1(t)}\,dW_1(t) - \sqrt{X_1(t)}\,dW_2(t), & X_1(0) = 10 \\ dX_2(t) = -X_2(t)\,dt + \sqrt{X_2(t)}\,dW_1(t) + \sqrt{X_1(t)}\,dW_2(t), & X_2(0) = 50 \end{cases}$$

for $0 \le t \le 1/2$. Using Itô's formula, prove that $E(X_2^2(t)) = 30 + 2470e^{-2t}$.

4.10. Use Euler's method on the system in Exercise 4.9 to estimate $E(X_2^2(1/2))$. Let R_N be the estimate of $E(X_2^2(1/2))$ using N intervals on $[0, 1/2]$ and 20,000 sample paths. Compute R_4 and R_8 and extrapolate these values, i.e., $2R_8 - R_4$, to obtain a better approximation to $E(X_2^2(1/2))$. Compare the calculated values with the exact value found in Exercise 4.9. Hand in a listing of your program along with your results.

4.11. Consider the stochastic differential equation

$$dX(t) = -\lambda X(t)\,dt + \sqrt{\mu X(t)}\,dW(t), \quad X(0) = X_0,$$

for $t \geq 0$ where $X_0 > 0$ and λ and μ are positive constants. Prove that the steady solution $X(t) \equiv 0$ is mean square stable. That is, show that $E(X^2(t)) \to 0$ as $t \to \infty$.

4.12. Consider the stochastic system

$$\begin{cases} dX_1(t) = a\,dW_1(t) + b\,dW_2(t), & X_1(0) = 0 \\ dX_2(t) = c\,dW_1(t), & X_2(0) = 0. \end{cases}$$

(a) Show that the forward Kolmogorov equation for this system is:

$$\frac{\partial p(t,x_1,x_2)}{\partial t} = \frac{1}{2}(a^2+b^2)\frac{\partial^2 p(t,x_1,x_2)}{\partial x_1^2} + ac\frac{\partial^2 p(t,x_1,x_2)}{\partial x_1 \partial x_2} + \frac{1}{2}c^2\frac{\partial^2 p(t,x_1,x_2)}{\partial x_2^2}.$$

(b) The probability density for this problem has the form:

$$p(t,\mathbf{x}) = \frac{1}{2\pi(\det \Sigma)^{1/2}} e^{-\frac{1}{2}\mathbf{x}^T \Sigma^{-1}\mathbf{x}} = \frac{1}{2\pi bct} e^{(-c^2x_1^2 + 2acx_1x_2 - (a^2+b^2)x_2^2)/(2b^2c^2t)},$$

where Σ is the 2×2 matrix:

$$\Sigma = \begin{pmatrix} (a^2+b^2)t & act \\ act & c^2t \end{pmatrix}.$$

Verify that $p(t,\mathbf{x})$ satisfies the forward Kolmogorov equation found in part (a).

4.13. Consider the implicit Euler method

$$X_{k+1} = X_k + \lambda h X_{k+1} + \mu X_k \sqrt{h}\,\eta_k, \quad \text{for} \quad k = 0, 1, 2, \ldots$$

for numerically solving the stochastic differential equation

$$dX(t) = \lambda X(t)\,dt + \mu X(t)\,dW(t), \quad X(0) = X_0,$$

for $t \geq 0$ where λ and μ are constants and $h > 0$ is the step size. Recall that the steady solution $X(t) = 0$ of this stochastic differential equation is mean

square stable if and only if $\mu^2 + 2\lambda < 0$.
(a) Show that if $\mu^2 + 2\lambda < 0$, then the implicit Euler method is numerically mean square stable for any step size $h > 0$.
(b) Now suppose that $\mu^2 + 2\lambda > 0$ so that the the stochastic differential equation is not mean square stable at $X(t) = 0$. Show, in this case, however, that if the step size, h, is sufficiently large, specifically $h > c(\mu^2 + 2\lambda)/\lambda^2$ for some constant $c > 1$, then the implicit Euler method is mean square stable.

4.14. Consider the stochastic differential equation

$$dX(t) = X(t)\,dt + \sqrt{\frac{X(t)}{4}}\,dW(t), \quad X(0) = 1,$$

for $0 \le t \le 1$. Notice, using Itô's formula, that $E(X(t)) = e^t$ and $E(X^2(t)) = \frac{5}{4}e^{2t} - \frac{1}{4}e^t$. Calculate 200 sample paths using Euler's method with 500 intervals on $[0, 1]$, i.e. $\Delta t = 1/500$. Plot two sample paths and the average of 200 sample paths. In constructing your computer program, consider modifying the MATLAB program listed in Exercise 2.10. If the sample path average does not appear similar to the graph of the function e^t, then your program needs to be carefully checked.

Computer Programs

Program 4.1. SDE Solution using Euler and Milstein methods

This Fortran program solves a scalar stochastic differential equation using Euler's method and Milstein's method simultaneously with nt equally spaced steps in time. The number of sample paths calculated is specified as nrun. The mean and mean square values of the solution at the final time are output. Then, the calculations are continued with the number of time steps equal to 2*nt and 4*nt each with nrun sample paths. The program is set up to solve the stochastic differential equation in described in Example 4.6. Output of the program is given following the program listing.

```
      real*8 xx
c This program uses Euler and Milstein methods for an SDE.
      nt=32
c nt is the number of intervals
      do 500 nnt=1,3
      nt=nt*2
      ant=nt
      xx=710781.
      nrun=100000
c nrun is the number of sample paths
      arun=nrun
```

```
      sm1=0.0
      sq1=0.0
      sm2=0.0
      sq2=0.0
      ant=nt
      arun=nrun
      do 75 jj=1,nrun
c y1 is the Euler approximation
      y1=1.0
      y2=1.0
c y2 is the Milstein approximation
      time=1.0
      h=time/ant
      hs=sqrt(h)
      t=0.0
      do 600 i=1,nt
      call random(xx,rand1,rand2)
      call fts(t,y1,y2,f1,f2,g1,g21,g22)
      t=t+h
      y1=y1+h*f1+hs*rand1*g1
      y2=y2+h*f2+hs*rand1*g21+(rand1*rand1-1.0)*g22*h
600   continue
      sm1=sm1+y1/arun
      sq1=sq1+y1*y1/arun
      sm2=sm2+y2/arun
      sq2=sq2+y2*y2/arun
75    continue
      write(6,160)nt,nrun
160   format(5x,i6,5x,i6)
      write(6,162) sm1,sq1,sm2,sq2
162   format(8x,4(f12.4,3x))
500   continue
      stop
      end
      subroutine random(xx,rand1,rand2)
      real*8 xx,a,b,d,rng(2)
      a=16807.
      ib=2147483647
      b=ib
      do 55 i=1,2
      id=a*xx/b
      d=id
      xx=a*xx-d*b
55    rng(i)=xx/b
      pi=3.141592654
```

```
u1=rng(1)
u2=rng(2)
hlp=sqrt(-2.0*alog(u1))
rand1=hlp*cos(pi*2.0*u2)
rand2=hlp*sin(pi*2.0*u2)
return
end
subroutine fts(t,y1,y2,f1,f2,g1,g21,g22)
hlp=1.0/3.0
hlp1=(abs(y1))**hlp
if(y1.lt.0.0) hlp1=-hlp1
hlp2=(abs(y2))**hlp
if(y2.lt.0.0) hlp2=-hlp2
f1=hlp*hlp1+6.0*hlp1*hlp1
g1=hlp1*hlp1
f2=hlp*hlp2+6.0*hlp2*hlp2
g21=hlp2*hlp2
g22=hlp*hlp2
return
end
 64   100000
      27.0954   811.0140    27.0993   811.0618
128   100000
      27.5339   839.3231    27.5368   839.3948
256   100000
      27.7886   855.1988    27.7901   855.2540
```

Program 4.2. Solution of a scalar SDE using Euler's method

This MATLAB program solves a scalar stochastic differential equation using Euler's method with nt equally spaced intervals in time. The number of sample paths calculated is specified as nvec. The mean and mean square values of the solution at the final time are the output values. Matrix array operations are applied for efficient computation. The program is set up to solve the stochastic differential equation in described in Example 4.6. Following the program listing, the output of the program is given.

```
% Euler method for stochastic differential equation
% Matrix array operations are used for efficient computation
% nvec sample paths are simultaneously computed
% exact solution is E(y(1))= 28.00 and E(y(1))^2 = 869.02
clear
%t=clock;
 randn('state',2)
 nvec=1000;
```

```
nt=128;
time=1;
h=time/nt;
hs=sqrt(h);
  y=ones(nvec,1);
  for i=1:nt
     r=randn(nvec,1);
     tt=(i-1)*h;
     hlp=y.^(1/3);
     f=hlp/3+6*hlp.^2;
     g=hlp.^2;
     y= y+ f*h + hs*r.*g;
  end
  mean=sum(y)/nvec
  meansqr=sum(y.*y)/nvec
   27.6437
  851.4002
```

Program 4.3. Parameter estimation using an MLE procedure

This Fortran program estimates the parameters of stochastic differential equation using the maximum likelihood procedure described in Section 4.9.1 where the Euler approximation is applied to Eq. (4.22). The program is applied to the Aransas-Wood Buffalo whooping crane population data. The minimum is found on a 1000 by 1000 mesh of [a1, a2] X [b1, b2] where the values of a1, a2, b1, and b2 are specified. Output of the program is given following the program listing.

```
      real*4 x(50)
      np=47
      data (x(i), i=1,47)/ 18, 22, 26, 16, 19, 21, 18, 22,
     * 25, 31, 30, 34, 31, 25, 21, 24, 21, 28, 24, 26, 32,
     * 33, 36, 39, 32, 33, 42, 44, 43, 48, 50, 56, 57, 59,
     * 51, 49, 49, 57, 69, 72, 75, 76, 78, 73, 73, 75, 86/
c This code estimates parameters of an SDE.
      fmin=1000000.0
c The minimum is found on a 1000 by 1000 mesh.
      a1=0.0
      a2=0.1
      da=(a2-a1)/1000
      b1=0.0
      b2=1.0
      db=(b2-b1)/1000
      th1=a1
      do 500 ia=1,1000
```

```
      th1=th1+da
      th2=b1
      do 500 ib=1,1000
      th2=th2+db
      flog=0.0
      do 600 j=1,46
      call functs(x(j),th1,th2,f,g)
      amu=x(j)+f
      sig=g
      hlp1=alog(sqrt(2.0*3.141592654*sig*sig))
      hlp2= -(x(j+1)-amu)**2/(2.0*sig*sig) - hlp1
      flog=flog - hlp2
600   continue
700   format(5x,3(3x,f14.5))
      if(flog.gt.fmin) goto 500
      fmin=flog
      th1min=th1
      th2min=th2
500   continue
      write(6,700) th1min,th2min,fmin
      stop
      end
      subroutine functs(x,th1,th2,f,g)
      f=th1*x
      g=sqrt(th2*x)
      return
      end
      0.03610          0.60900        136.78157
```

Program 4.4. Parameter estimation using an MLE procedure

This Fortran program estimates the parameters of stochastic differential equation using the maximum likelihood procedure described in Section 4.9.1. For this program, the transition densities are estimated through simulation using Euler's method with nt steps. The program is applied to the Aransas-Wood Buffalo whooping crane population data. The minimum is found on a 20 by 20 mesh of [a1, a2] X [b1, b2] where the values of a1, a2, b1, and b2 are specified. Output of the program is given following the program listing.

```
      real*4 x(50),r(50,20000),y(20000)
      real*8 xx
      np=47
      data (x(i), i=1,47)/ 18, 22, 26, 16, 19, 21, 18, 22,
     c 25, 31, 30, 34, 31, 25, 21, 24, 21, 28, 24, 26, 32,
     c 33, 36, 39, 32, 33, 42, 44, 43, 48, 50, 56, 57, 59,
```

```
     c 51, 49, 49, 57, 69, 72, 75, 76, 78, 73, 73, 75, 86/
c This code estimates parameters of an SDE.
      xx=114211.0
c An MLE procedure of Hurn, Lindsay, and Martin is used.
      m=5000
c m is the number of simulations.
      am=m
      nt=4
c nt is the number of Euler steps used in each simulation.
      ant=nt
c Next, random numbers for all simulations are generated.
      do 50 i=1,np
      do 50 j=1,m*nt/2
      call random(xx,rand1,rand2)
      r(i,2*j-1)=rand1
      r(i,2*j)=rand2
 50   continue
c The minimum is found on a 20 by 20 mesh of [a1,a2]X[b1,b2].
      a1=.030
      a2=.040
      da=(a2-a1)/20.
      b1=.500
      b2=.700
      db=(b2-b1)/20.
      fmin=10.**10
      alp=a1
      do 500 ia=1,20
      alp=alp+da
      bet=b1
      do 500 ib=1,20
      bet=bet+db
      flog=0.0
      do 600 j=1,46
      time=1.0
      h=time/ant
      hs=sqrt(h)
      iii=0
      do 75 n=1,m
      zz=x(j)
      do 76 jj=1,nt
      iii=iii+1
      call functs(zz,alp,bet,f,g)
      rrr=r(j,iii)
      zz=zz+h*f+hs*rrr*g
 76   continue
```

```
 75   y(n)=zz
      s1=0.0
      s2=0.0
      do 80 i=1,m
      s1=s1+y(i)
      s2=s2+y(i)*y(i)
 80   continue
      ss=(s2-s1*s1/am)/(am-1.0)
      ss=sqrt(ss)
      hh=.9*ss/(am**.2)
      sum=0.0
      hlp=sqrt(2.0*3.141592654)*hh
      do 85 i=1,m
      sum=sum + exp(-(x(j+1)-y(i))**2/(2.0*hh*hh))/hlp
 85   continue
      sum=sum/am
      flog=flog -log(sum)
600   continue
      if(flog.gt.fmin) goto 500
      fmin=flog
      alpmin=alp
      betmin=bet
 500  continue
      write(6,700) alpmin,betmin,fmin
 700  format(3(3x,f12.5))
      stop
      end
      subroutine random(xx,rand1,rand2)
      real*8 xx,a,b,d,rng(2)
      a=16807.
      ib=2147483647
      b=ib
      do 55 i=1,2
      id=a*xx/b
      d=id
      xx=a*xx-d*b
 55   rng(i)=xx/b
      pi=3.141592654
      u1=rng(1)
      u2=rng(2)
      hlp=sqrt(-2.0*log(u1))
      rand1=hlp*cos(pi*2.0*u2)
      rand2=hlp*sin(pi*2.0*u2)
      return
      end
```

```
subroutine functs(x,alp,bet,f,g)
f=alp*x
g=sqrt(bet*x)
return
end
0.03600        0.61000      137.85956
```

5

Modeling

5.1 Introduction

Most of the stochastic differential equation models described in this chapter are developed using a procedure that is analogous to the procedure used in the development of many ordinary differential equation models. The process under investigation is studied for a small time interval Δt. The resulting information on the changes in the process then leads to the differential equation model. For example, consider the problem of developing a deterministic model for the temperature of an object immersed in a liquid held at a fixed temperature T_L. Let $T(t)$ be the temperature of the object at time t. Suppose that for a small time interval Δt, the change in the temperature of the object is proportional to the difference between the object's temperature $T(t)$ and the liquid's temperature T_L. In addition, suppose that the temperature change is proportional to the interval length Δt. Based on this information, it follows that

$$\Delta T = \alpha(T_L - T(t))\Delta t,$$

where α is a constant of proportionality. Setting $\Delta T = T(t + \Delta t) - T(t)$ and letting $\Delta t \to 0$, one obtains Newton's Law of Cooling. That is,

$$\frac{dT}{dt} = \alpha(T_L - T).$$

Therefore, the procedure used in this chapter is just a natural extension of the procedure used for many years in modeling deterministic dynamical processes in physics, engineering, and biology. However, in the case considered here when the process is stochastic rather than deterministic, a finite Δt produces a discrete stochastic model. The discrete stochastic model then leads to a stochastic differential equation model as $\Delta t \to 0$. This model development procedure has been applied to several interesting biological and physical phenomena [4, 7, 8, 11, 12, 13, 39, 55, 68, 86, 105]. Furthermore, as $\Delta t \to 0$,

a continuous-time stochastic model is obtained from the discrete stochastic model and it appears that similarities between continuous-time Markov chain models and stochastic differential equation models were first noted by Kurtz in 1971 [78].

Based on the preceding discussion, most of the stochastic differential equation models described in this chapter are developed in the following way. First, a discrete stochastic model is developed for the dynamical system under study which is experiencing random influences. Specifically, for a small time interval Δt, the possible changes with their corresponding transition probabilities are determined. Second, the expected change and the covariance matrix for the change are determined for this discrete stochastic process. Third, this information leads to the stochastic differential equation model for the dynamical system. The stochastic differential equation model is inferred by similarities in the forward Kolmogorov equations between the discrete and continuous stochastic processes.

There are ways to develop stochastic differential equation models other than through using the procedure discussed in this chapter. For example, one may hypothesize for a given stochastic dynamical system that the drift and diffusion coefficients in the stochastic differential equation model are linear functions of the solution. Then, assuming data are available, a statistical estimation method may yield values for the unknown parameters. This alternate way of developing a stochastic differential equation model is useful for many stochastic systems. One advantage of the modeling procedure discussed in this section is that, as the parameters in the model are derived from basic assumptions, a better understanding of the parameters in the model is achieved. There are many interesting applications and models of stochastic differential equations. See, for example, the discussions in [4, 11, 29, 32, 41, 56, 69, 70, 74, 89, 90, 92].

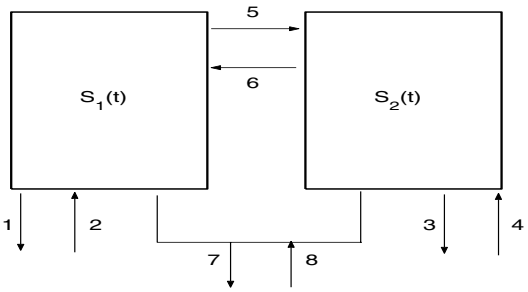

Fig. 5.1. A diagram of a two-state dynamical system

It is useful to begin by considering a representative two-state dynamical process which is illustrated in Fig. 5.1. Let $S_1(t)$ and $S_2(t)$ represent the values of two states in the system at time t. It is assumed that in a small time interval Δt, state S_1 can change by $-\lambda_1, 0$, or $+\lambda_1$ and state S_2 can change by $-\lambda_2, 0$, or $+\lambda_2$ where $\lambda_1, \lambda_2 \geq 0$. Let $\Delta \mathbf{S} = [S_1, S_2]^T$ be the change in a small time interval Δt. As illustrated in Fig. 5.1, there are eight possible changes for the two states in the time interval Δt not including the case where there is no change in the time interval. The possible changes and the probabilities of these changes are given in Table 5.1. It is assumed that the probabilities are given to $O((\Delta t)^2)$. For example, change 1 represents a loss of λ_1 in $S_1(t)$ with probability $d_1 \Delta t$, change 5 represents a transfer of λ_1 out of state S_1 with a corresponding transfer of λ_2 into state S_2 with probability $m_{12}\Delta t$, and change 7 represents a simultaneous reduction in both states S_1 and S_2. The changes λ_i are assumed to be nonnegative. As indicated in the table, all probabilities may depend on $S_1(t), S_2(t)$, and time t. Also notice that it is assumed that the probabilities for the changes are proportional to the time interval Δt.

Table 5.1. Possible changes in the representative two-state system with the corresponding probabilities

Change	Probability
$\Delta \mathbf{S}^{(1)} = [-\lambda_1, 0]^T$	$p_1 = d_1(t, S_1, S_2)\Delta t$
$\Delta \mathbf{S}^{(2)} = [\lambda_1, 0]^T$	$p_2 = b_1(t, S_1, S_2)\Delta t$
$\Delta \mathbf{S}^{(3)} = [0, -\lambda_2]^T$	$p_3 = d_2(t, S_1, S_2)\Delta t$
$\Delta \mathbf{S}^{(4)} = [0, \lambda_2]^T$	$p_4 = b_2(t, S_1, S_2)\Delta t$
$\Delta \mathbf{S}^{(5)} = [-\lambda_1, \lambda_2]^T$	$p_5 = m_{12}(t, S_1, S_2)\Delta t$
$\Delta \mathbf{S}^{(6)} = [\lambda_1, -\lambda_2]^T$	$p_6 = m_{21}(t, S_1, S_2)\Delta t$
$\Delta \mathbf{S}^{(7)} = [-\lambda_1, -\lambda_2]^T$	$p_7 = m_{11}(t, S_1, S_2)\Delta t$
$\Delta \mathbf{S}^{(8)} = [\lambda_1, \lambda_2]^T$	$p_8 = m_{22}(t, S_1, S_2)\Delta t$
$\Delta \mathbf{S}^{(9)} = [0, 0]^T$	$p_9 = 1 - \sum_{i=1}^{8} p_i$

It is useful to calculate the expected change and the covariance matrix for the change $\Delta \mathbf{S} = [\Delta S_1, \Delta S_2]^T$ fixing $\mathbf{S}(t)$ at time t. Using the above table,

$$E(\Delta \mathbf{S}) = \sum_{j=1}^{9} p_j \Delta \mathbf{S}^{(j)} = \begin{bmatrix} (-d_1 + b_1 - m_{12} + m_{21} + m_{22} - m_{11})\lambda_1 \\ (-d_2 + b_2 + m_{12} - m_{21} + m_{22} - m_{11})\lambda_2 \end{bmatrix} \Delta t$$

and

$$E(\Delta \mathbf{S}(\Delta \mathbf{S})^T) = \sum_{j=1}^{9} p_j (\Delta \mathbf{S}^{(j)})(\Delta \mathbf{S}^{(j)})^T$$

$$= \begin{bmatrix} (d_1 + b_1 + m_a)\lambda_1^2 & (-m_{12} - m_{21} + m_{22} + m_{11})\lambda_1\lambda_2 \\ (-m_{12} - m_{21} + m_{22} + m_{11})\lambda_1\lambda_2 & (d_2 + b_2 + m_a)\lambda_2^2 \end{bmatrix} \Delta t,$$

where $m_a = m_{12} + m_{21} + m_{22} + m_{11}$. We now define the expectation vector, $\boldsymbol{\mu}$, and the 2×2 symmetric positive definite covariance matrix, V, as

$$\boldsymbol{\mu}(t, S_1, S_2) = E(\Delta \mathbf{S})/\Delta t \quad \text{and} \quad V(t, S_1, S_2) = E(\Delta \mathbf{S}(\Delta \mathbf{S})^T)/\Delta t.$$

Notice that as Δt is small and $E(\Delta \mathbf{S})(E(\Delta \mathbf{S}))^T = O((\Delta t)^2)$, the covariance matrix V is set equal to $E(\Delta \mathbf{S}(\Delta \mathbf{S})^T)/\Delta t$. Finally, it is useful to define the square root of the covariance matrix V as B, specifically,

$$B(t, S_1, S_2) = (V(t, S_1, S_2))^{1/2} \quad \text{and thus,} \quad B^2(t, S_1, S_2) = V(t, S_1, S_2).$$

Now, as described in Section 2.2 for a one-dimensional problem, a forward Kolmogorov equation can be determined for the probability distribution at time $t + \Delta t$ in terms of the distribution at time t. Letting $p(t, x_1, x_2)$ be the probability that $S_1 = x_1$, and $S_2 = x_2$ at time t, then referring to Table 5.1,

$$p(t + \Delta t, x_1, x_2) = p(t, x_1, x_2) + \Delta t \sum_{i=1}^{10} T_i, \tag{5.1}$$

where

$$\begin{aligned}
T_1 &= p(t, x_1, x_2)(-d_1(t, x_1, x_2) - b_1(t, x_1, x_2) - d_2(t, x_1, x_2) - b_2(t, x_1, x_2)) \\
T_2 &= p(t, x_1, x_2)(-m_a(t, x_1, x_2)) \\
T_3 &= p(t, x_1 + \lambda_1, x_2) d_1(t, x_1 + \lambda_1, x_2) \\
T_4 &= p(t, x_1 - \lambda_1, x_2) b_1(t, x_1 - \lambda_1, x_2) \\
T_5 &= p(t, x_1, x_2 - \lambda_2) b_2(t, x_1, x_2 - \lambda_2) \\
T_6 &= p(t, x_1, x_2 + \lambda_2) d_2(t, x_1, x_2 + \lambda_2) \\
T_7 &= p(t, x_1 + \lambda_1, x_2 - \lambda_2) m_{12}(t, x_1 + \lambda_1, x_2 - \lambda_2) \\
T_8 &= p(t, x_1 - \lambda_1, x_2 + \lambda_2) m_{21}(t, x_1 - \lambda_1, x_2 + \lambda_2) \\
T_9 &= p(t, x_1 + \lambda_1, x_2 + \lambda_2) m_{11}(t, x_1 + \lambda_1, x_2 + \lambda_2) \\
T_{10} &= p(t, x_1 - \lambda_1, x_2 - \lambda_2) m_{22}(t, x_1 - \lambda_1, x_2 - \lambda_2).
\end{aligned}$$

Expanding out the terms T_3 through T_{10} in Taylor series about the point (t, x_1, x_2), it follows that

$$\begin{aligned}
T_3 &\approx p d_1 + \frac{\partial (p d_1)}{\partial x_1} \lambda_1 + \frac{1}{2} \frac{\partial^2 (p d_1)}{\partial x_1^2} \lambda_1^2 \\
T_4 &\approx p b_1 - \frac{\partial (p b_1)}{\partial x_1} \lambda_1 + \frac{1}{2} \frac{\partial^2 (p b_1)}{\partial x_1^2} \lambda_1^2 \\
T_5 &\approx p b_2 - \frac{\partial (p b_2)}{\partial x_2} \lambda_2 + \frac{1}{2} \frac{\partial^2 (p b_2)}{\partial x_2^2} \lambda_2^2
\end{aligned}$$

$$T_6 \approx pd_2 + \frac{\partial(pd_2)}{\partial x_2}\lambda_2 + \frac{1}{2}\frac{\partial^2(pd_2)}{\partial x_2^2}\lambda_2^2$$

$$T_7 \approx pm_{12} + \frac{\partial(pm_{12})}{\partial x_1}\lambda_1 - \frac{\partial(pm_{12})}{\partial x_2}\lambda_2 + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}(-1)^{i+j}\frac{\partial^2(pm_{12})}{\partial x_i \partial x_j}\lambda_i\lambda_j$$

$$T_8 \approx pm_{21} - \frac{\partial(pm_{21})}{\partial x_1}\lambda_1 + \frac{\partial(pm_{21})}{\partial x_2}\lambda_2 + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}(-1)^{i+j}\frac{\partial^2(pm_{21})}{\partial x_i \partial x_j}\lambda_i\lambda_j$$

$$T_9 \approx pm_{11} + \frac{\partial(pm_{11})}{\partial x_1}\lambda_1 + \frac{\partial(pm_{11})}{\partial x_2}\lambda_2 + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial^2(pm_{11})}{\partial x_i \partial x_j}\lambda_i\lambda_j$$

$$T_{10} \approx pm_{22} - \frac{\partial(pm_{22})}{\partial x_1}\lambda_1 - \frac{\partial(pm_{22})}{\partial x_2}\lambda_2 + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial^2(pm_{22})}{\partial x_i \partial x_j}\lambda_i\lambda_j.$$

Substituting these expressions into (5.1) and assuming that Δt, λ_1, and λ_2 are small, then it is seen that $p(t, x_1, x_2)$ approximately solves the Fokker-Planck equation

$$\begin{aligned}\frac{\partial p(t,x_1,x_2)}{\partial t} &= -\sum_{i=1}^{2}\frac{\partial}{\partial x_i}\left[\mu_i(t,x_1,x_2)p(t,x_1,x_2)\right] \\ &\quad + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial^2}{\partial x_i \partial x_j}\left[\sum_{k=1}^{2} b_{i,k}(t,x_1,x_2)b_{j,k}(t,x_1,x_2)p(t,x_1,x_2)\right],\end{aligned} \tag{5.2}$$

where μ_i is the ith component of $\boldsymbol{\mu}$ and $b_{i,j} = (B)_{i,j}$ for $1 \le i, j \le 2$. (Alternatively, as in Example 2.4, it can be shown that (5.1) is a form of finite-difference approximation to (5.2) for small Δt, λ_1, and λ_2.) However, as discussed in the previous chapter, the probability distribution $p(t, x_1, x_2)$ that exactly satisfies (5.2) is identical to the distribution of solutions to the stochastic differential equation system

$$\begin{cases} d\mathbf{S}(t) = \boldsymbol{\mu}(t, S_1, S_2)\, dt + B(t, S_1, S_2)\, d\mathbf{W}(t) \\ \mathbf{S}(0) = \mathbf{S}_0, \end{cases} \tag{5.3}$$

where $\mathbf{W}(t) = [W_1(t), W_2(t)]^T$. Therefore, as also illustrated for the one-dimensional examples in Chapter 2 (Examples 2.4 and 2.5), the discrete stochastic model is closely related to a stochastic differential equation model. In particular, the drift and diffusion terms, $\boldsymbol{\mu}$ and B, respectively, of the stochastic differential equation model are equal to the expected change divided by Δt and the square root of the covariance matrix divided by Δt obtained from the discrete stochastic model.

In summary, the procedure described here in deriving a stochastic differential model for a dynamical process requires three steps. First, a discrete stochastic model for the process is developed by carefully listing the possible changes along with the corresponding probabilities for a short time step

Δt. Second, the expected change and covariance matrix for the change is calculated for the discrete stochastic process. Third, the stochastic differential equation system is obtained by letting the expected change divided by Δt be the drift coefficient and the square root of the covariance matrix divided by Δt be the diffusion coefficient. The procedure described in this section provides, in a natural manner, an Itô stochastic differential equation model, rather than, for example, a Stratonovich stochastic differential equation model.

It is interesting to notice that

$$E\left[\left|E\left(\frac{\Delta \mathbf{S}}{\Delta t}\right) - \boldsymbol{\mu}(t, S_1, S_2)\right|^2\right] \to 0 \quad \text{as} \quad \Delta t \to 0$$

and

$$E\left[\left|E\left(\frac{\Delta \mathbf{S}(\Delta \mathbf{S})^T}{\Delta t}\right) - B(t, S_1, S_2)B^T(t, S_1, S_2)\right|^2\right] \to 0 \quad \text{as} \quad \Delta t \to 0$$

for fixed values $\mathbf{S}(t) = [S_1(t), S_2(t)]^T$ at time t. These relations indicate that the discrete stochastic process is weakly consistent [69] with the stochastic differential equation (5.3). In addition, under suitable assumptions, weak consistency implies the weak convergence result

$$E(g(\mathbf{S}(T))) - E(g(\mathbf{S}_\Delta(T))) \to 0 \quad \text{as} \quad \Delta t \to 0$$

for smooth functions $g : \mathbb{R}^2 \to \mathbb{R}$, where $\mathbf{S}(T)$ is the solution at time T to the stochastic differential equation system (5.3) and $\mathbf{S}_\Delta(T)$ is the solution to the discrete stochastic system [69]. Therefore, another way to regard this modeling procedure is through weak convergence of the discrete stochastic system to a continuous stochastic system. This convergence is also suggested by considering the Kolmogorov equations for the two stochastic systems. In either way of regarding the procedure, a discrete stochastic process is first constructed and the expected change and the covariance in the change are calculated for a small time step Δt which leads to the stochastic differential equation model. Several examples of this approach are illustrated in the remainder of this chapter for problems in biology, physics, engineering, and finance. Of course, it is generally recommended, after a mathematical model is formulated for a given process, that the model be thoroughly tested with Monte Carlo calculations or with experimental data to verify its accuracy.

Example 5.1. A simple example with exact probability distributions

Consider development of one-state discrete and continuous stochastic models with probability distributions that can be exactly determined and compared. Let $t_i = i\Delta t$ for $i = 0, 1, \ldots, N$ and let a discrete stochastic process take a value from the set $\{x_j = j\delta \quad \text{for} \quad j = 0, 1, \ldots\}$ at each discrete time t_i where δ is a small positive constant. Define the transition probabilities of the discrete stochastic process by the following:

$$p_{i,k}(t) = \begin{cases} r\Delta t/\delta^2, & \text{for} \quad k = i+1 \\ 1 - r\Delta t/\delta^2, & \text{for} \quad k = i, \end{cases}$$

where r is a positive constant and the change in the process is either δ or 0 at each time step. The transition probabilities for this example are similar to those in Example 2.4. Notice that with the above transition probabilities, if ΔX is the change in the stochastic process for time interval Δt, then $E(\Delta X) = r\Delta t/\delta$ and $\text{Var}(\Delta X) = r\Delta t + O((\Delta t)^2)$. Let $p_k(t) = P(X(t) = x_k = k\delta)$ be the probability distribution at time t. Then, $p_k(t + \Delta t)$ satisfies

$$p_k(t + \Delta t) = p_k(t) - p_k(t) r\Delta t/\delta^2 + p_{k-1}(t) r\Delta t/\delta^2.$$

As $\Delta t \to 0$, the discrete stochastic process approaches a continuous-time process and $p_k(t)$ satisfies the differential equation:

$$\frac{dp_k(t)}{dt} = -\frac{r}{\delta^2}\, p_k(t) + \frac{r}{\delta^2}\, p_{k-1}(t)$$

for $k = 0, 1, 2, \ldots$ where $p_{-1}(t) = 0$. Assuming that $p_0(0) = 1$ and $p_k(0) = 0$ for $k \geq 1$, these equations can be solved exactly to yield

$$p_n(t) = \exp(-\lambda t)(\lambda t)^n/n! \quad \text{for} \quad n = 0, 1, 2, \ldots,$$

where $\lambda = \dfrac{r}{\delta^2}$. That is, the discrete stochastic model is Poisson distributed as $\Delta t \to 0$. Notice that

$$E(X(t)) = \sum_{n=0}^{\infty} n\delta p_n(t) = \frac{rt}{\delta} \quad \text{and} \quad \text{Var}(X(t)) = \sum_{n=0}^{\infty} n^2\delta^2 p_n(t) - \frac{r^2t^2}{\delta^2} = rt.$$

Now, using the procedure described in this section, the forward Kolmogorov equation satisfied by the corresponding continuous-valued model has the form

$$\frac{\partial p(t,x)}{\partial t} = -\frac{r}{\delta}\frac{\partial p(t,x)}{\partial x} + \frac{r}{2}\frac{\partial^2 p(t,x)}{\partial x^2}$$

with $p(0, x) = \delta(x - 0)$. The Itô stochastic differential equation model of the process is therefore

$$dX(t) = \frac{r}{\delta}\, dt + \sqrt{r}\, dW(t) \quad \text{with} \quad X(0) = 1.$$

For the continuous model, $p(t,x) = \dfrac{1}{\sqrt{2\pi rt}} \exp(-(x - rt/\delta)^2/2rt)$. That is, the continuous stochastic model is normally distributed with mean rt/δ and variance rt. However, referring to Example 1.25, the Poisson distributed variable converges in distribution to a normally distributed variable as λt increases (as δ decreases for fixed $t > 0$). Therefore, the discrete and continuous models are closely related for small Δt and small δ. This example illustrates that the probability distributions of the discrete and continuous stochastic models are very similar under certain conditions.

Remark 5.2. Appropriateness of an Itô or a Stratonovich SDE model

There have been discussions regarding whether an Itô or a Stratonovich stochastic differential equation provides the most appropriate model for a given random dynamical system. (See, e.g., [41], [92], and [112].) Whether the stochastic differential equation is regarded as Itô or Stratonovich is important. For example, if

$$dX(t) = \lambda X(t)\,dt + \mu X(t)\,dW(t)$$

is regarded as Itô, then $X(t) \to X(0)\exp((\lambda - \mu^2/2)t)$ w.p.1 as $t \to \infty$ and, thus, $X(t) \to 0$ with probability 1 if $\lambda < \mu^2/2$. (To see this, refer to the discussion following equation (4.16) in Section 4.8.) On the other hand, if this stochastic differential equation is regarded as Stratonovich, then the equivalent Itô stochastic differential equation is

$$dX(t) = (\lambda + \mu^2/2)X(t)\,dt + \mu X(t)\,dW(t).$$

In this case, $X(t) \to X(0)\exp(\lambda t)$ w.p.1 as $t \to \infty$ and, thus, $X(t) \to 0$ with probability 1 if $\lambda < 0$. It is clear from this example that specification of the stochastic differential equation model as Itô or Stratonovich is important. In the modeling procedure described in this chapter, a discrete stochastic model is first developed by studying the random dynamical system for a small time interval. An Itô stochastic differential equation model is then inferred by the similarities between the forward Kolmogorov equations of the discrete-time and the continuous-time stochastic models. Assuming that the discrete stochastic model is accurate, a specific Itô stochastic differential equation model is obtained using this derivation procedure. Of course, a Stratonovich stochastic differential equation model is simultaneously produced as the Itô model can be readily transformed into a Stratonovich model. As previously noted [41], whether a stochastic differential equation is interpreted as an Itô equation or as a Stratonovich equation may often be resolved under careful derivation. The derivation procedure described in this chapter leads in a natural manner to a particular Itô stochastic differential equation model which, if desired, can be immediately transformed into a Stratonovich stochastic differential equation model.

Remark 5.3. Calculating square roots of matrices

Before studying several model derivations in the following sections, notice that the procedure described in this section produces a term in the stochastic differential equation system that involves the square root of a symmetric positive definite matrix, that is, $B = V^{1/2}$. Solution of the stochastic system entails computation of square roots of matrices. It is well-known (see, e.g., [61, 80]) that a symmetric positive definite matrix has a unique symmetric positive definite square root. For a 2×2 matrix, the square root can be readily calculated. Indeed,

$$V^{1/2} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}^{1/2} = \frac{1}{d}\begin{bmatrix} a+w & b \\ b & c+w \end{bmatrix},$$

where $w = \sqrt{ac - b^2}$ and $d = \sqrt{a + c + 2w}$. However, for a general $n \times n$ symmetric positive definite matrix V with $n \geq 3$, there is no explicit formula for $V^{1/2}$ and so it must be calculated numerically. Clearly, when V is put in the canonical form $V = P^T DP$, where $P^T P = I$ and $d_{ii} > 0$ for $i = 1, 2, \ldots, n$, then $V^{1/2} = P^T D^{1/2} P$. However, for a large matrix, it is computationally intensive to accurately compute all of the eigenvalues and eigenvectors of V which are needed to determine P and D. Fortunately, there are available many numerical procedures for computing $V^{1/2}$ directly. (See, for example, [61, 80].) A fast robust third-order method for symmetric positive definite matrices uses the algorithm in [80]:

$$\begin{aligned} R_{m+1} &= R_m \left(\tfrac{3}{8} I + \tfrac{3}{4} S_m (I - \tfrac{1}{6} S_m)\right) \\ S_{m+1} &= S_m \left(\tfrac{3}{8} I + \tfrac{3}{4} S_m (I - \tfrac{1}{6} S_m)\right)^{-2}, \end{aligned}$$

where $R_0 = I$ and $S_0 = V/||V||$. In this method, $\sqrt{||V||} R_m \to V^{1/2}$ as $m \to \infty$.

Furthermore, given a vector $\mathbf{c}$, a numerical method for directly calculating $V^{1/2}\mathbf{c}$ without explicitly finding $V^{1/2}$ can be formulated. (Generally this is all that is required in solving a stochastic differential equation; computing $V^{1/2}$ is not required.) This numerical method is based on solving an initial-value system. Specifically, consider the initial-value problem

$$\frac{d\mathbf{x}(t)}{dt} = C(t)\mathbf{x}(t) \quad \text{with} \quad \mathbf{x}(0) = \mathbf{c},$$

where

$$C(t) = -\frac{1}{2}(\hat{V}t + (1-t)I)^{-1}(I - \hat{V}) \quad \text{with} \quad \hat{V} = \frac{V}{2||V||_\infty}.$$

The solution to this initial-value problem at $t = 1$ is

$$\mathbf{x}(1) = \hat{V}^{1/2}\mathbf{c} = \frac{V^{1/2}\mathbf{c}}{\sqrt{2||V||_\infty}}.$$

Thus, if this initial-value problem is solved numerically from $t = 0$ to $t = 1$ and $\mathbf{x}(1)$ is computed, it follows that

$$V^{1/2}\mathbf{c} = \sqrt{2||V||_\infty}\, \mathbf{x}(1).$$

This method is robust and efficient for computing the product of the square root of a symmetric positive definite matrix with a vector [5].

Remark 5.4. Including additional Wiener processes as an alternative to the matrix square root

Often, the calculation of the square root of the covariance matrix $B = V^{1/2}$ can be avoided by including additional Wiener processes in the stochastic

system (5.3) [9, 14]. Consider a stochastic modeling problem that involves N states $S_1, S_2, \ldots, S_N$ with a total of $M \geq N$ possible random changes to these states at each time step Δt. Suppose that the probabilities of the changes are $p_j \Delta t = p_j(t, \mathbf{S})\Delta t$ for $j = 1, 2, \ldots, M$ where the jth change alters the ith state by the amount λ_{ji} for $i = 1, 2, \ldots, N$. The ith element of vector $\boldsymbol{\mu}$ for this problem is then

$$\mu_i = \sum_{j=1}^{M} p_j \lambda_{ji} \quad \text{for} \quad i = 1, 2, \ldots, N.$$

The covariance matrix, V, can also be computed and the i, l entry of V has the form

$$v_{il} = \sum_{j=1}^{M} p_j \lambda_{ji} \lambda_{jl} \quad \text{for} \quad 1 \leq i, l \leq N.$$

As explained in Remark 5.3, it is generally difficult to compute the $N \times N$ matrix $V^{1/2}$ in the stochastic differential equation system

$$d\mathbf{S} = \boldsymbol{\mu}\, dt + V^{1/2}\, d\mathbf{W}(t),$$

where $\mathbf{W}(t)$ is a vector consisting of N independent Wiener processes. However, it is interesting that an $N \times M$ matrix C can be readily found such that $V = CC^T$ and the stochastic differential equation system can be modified to the system

$$d\mathbf{S} = \boldsymbol{\mu}\, dt + C\, d\mathbf{W}^*(t),$$

where $\mathbf{W}^*(t)$ is a vector consisting M independent Wiener processes. Indeed, Itô's formula and the forward Kolmogorov equation are identical for both stochastic differential equation systems. The entries of matrix C have the form:

$$c_{ij} = \lambda_{ji} p_j^{1/2} \quad \text{for} \quad 1 \leq i \leq N, 1 \leq j \leq M.$$

To verify this formula, notice that

$$(CC^T)_{il} = \sum_{j=1}^{M} c_{ij} c_{lj} = \sum_{j=1}^{M} \lambda_{ji} p_j^{1/2} \lambda_{jl} p_j^{1/2} = v_{il} \quad \text{for} \quad 1 \leq i, l \leq N.$$

For chemically reacting systems, such as the example in Section 5.3.6, the stochastic differential equation system with $Cd\mathbf{W}^*(t)$ replacing $V^{1/2}d\mathbf{W}(t)$ is referred to as the chemical Langevin equation [45]. Finally, notice that if the number of changes, M, is much greater than the number of states, N, then this approach loses much of its effectiveness. More information about alternate stochastic differential equation models is given in Section 5.6.

5.2 Population Biology Examples

5.2.1 General model of two interacting populations

Consider first two interacting populations although it is straightforward to extend the derivations to more than two populations. The approach will be applicable to populations of the same species or to populations of different species. Populations of the same species may differ, for example, by geographic location or by status in an epidemic, e.g., infective or susceptible. In such cases, the populations may interact, respectively, by migration or by transmitting and recovering from a disease. Populations of different species may interact, for example, as competitors or as predator and prey. More details about modeling the stochastic dynamics of populations using the procedure described in this section can be found, for example, in references [4, 7, 8, 11, 12, 39, 68, 86].

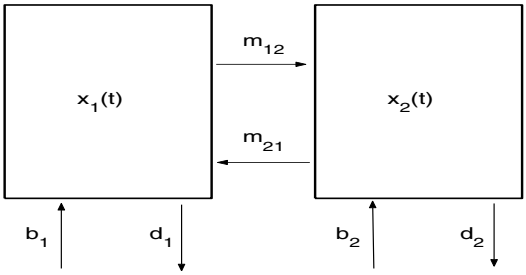

Fig. 5.2. A diagram of two interacting populations

Let the sizes of the two populations at time t be denoted by $x_1(t)$ and $x_2(t)$. Important parameters for the two populations are denoted by $b_1, d_1, b_2, d_2, m_{12}$, and m_{21}. The parameters b_i and d_i are per capita birth and death rates, respectively, for population i and m_{ij} is the rate population i is transformed to population j. For geographically isolated populations, m_{ij} may represent the migration rate of population i to j. For a population undergoing an epidemic, m_{12} may represent the rate a susceptible becomes infected and m_{21} may represent the rate an infected recovers. Of course, each parameter may depend on population sizes x_1 and x_2 and time t, i.e., $b_i = b_i(t, x_1, x_2)$, $d_i = d_i(t, x_1, x_2)$, and $m_{ij} = m_{ij}(t, x_1, x_2)$ where it assumed that each parameter is a smooth function of x_1, x_2, and t. For notational simplicity, the dependence of the parameters on x_1, x_2, and t is often not explicitly written.

Figure 5.2 illustrates the interactions considered here between two populations. In a small time interval Δt, there are seven possibilities for a population change $\Delta \mathbf{x}$ neglecting multiple births, deaths, or transformations in time Δt

which have probabilities of order $(\Delta t)^2$. These possibilities are listed in Table 5.2 along with their corresponding probabilities. For example, $\Delta\mathbf{x}_2 = [-1, 1]^T$ represents the movement of one individual from population x_1 to population x_2 during time interval Δt and the probability of this event is proportional to the size of population x_1 and the time interval Δt, that is, $p_2 = m_{12}x_1\Delta t$. As a second example, $\Delta\mathbf{x}_4 = [0, 1]^T$ represents a birth in population x_2 with probability $p_4 = b_2x_2\Delta t$. It is assumed that $\Delta t > 0$ is sufficiently small so that $p_7 > 0$. Notice that $\sum_{i=1}^{7} p_i = 1$.

Table 5.2. Possible changes in the two-population system with the corresponding probabilities

Change	Probability
$\Delta\mathbf{x}_1 = [-1, 0]^T$	$p_1 = d_1x_1\Delta t$
$\Delta\mathbf{x}_2 = [-1, 1]^T$	$p_2 = m_{12}x_1\Delta t$
$\Delta\mathbf{x}_3 = [0, -1]^T$	$p_3 = d_2x_2\Delta t$
$\Delta\mathbf{x}_4 = [0, 1]^T$	$p_4 = b_2x_2\Delta t$
$\Delta\mathbf{x}_5 = [1, -1]^T$	$p_5 = m_{21}x_2\Delta t$
$\Delta\mathbf{x}_6 = [1, 0]^T$	$p_6 = b_1x_1\Delta t$
$\Delta\mathbf{x}_7 = [0, 0]^T$	$p_7 = 1 - \sum_{i=1}^{6} p_i$

It is now of interest to find the mean change $E(\Delta\mathbf{x})$ and the covariance matrix $E(\Delta\mathbf{x}(\Delta\mathbf{x})^T)$ for the time interval Δt. Neglecting terms of order $(\Delta t)^2$,

$$E(\Delta\mathbf{x}) = \sum_{j=1}^{7} p_j\Delta\mathbf{x}_j = \begin{bmatrix} b_1x_1 - d_1x_1 - m_{12}x_1 + m_{21}x_2 \\ \\ b_2x_2 - d_2x_2 - m_{21}x_2 + m_{12}x_1 \end{bmatrix} \Delta t$$

and

$$E(\Delta\mathbf{x}(\Delta\mathbf{x})^T) = \sum_{j=1}^{7} p_j\Delta\mathbf{x}_j(\Delta\mathbf{x}_j)^T$$

$$= \begin{bmatrix} b_1x_1 + d_1x_1 + m_{12}x_1 + m_{21}x_2 & -m_{12}x_1 - m_{21}x_2 \\ \\ -m_{12}x_1 - m_{21}x_2 & b_2x_2 + d_2x_2 + m_{12}x_1 + m_{21}x_2 \end{bmatrix} \Delta t.$$

As the product $E(\Delta\mathbf{x})(E(\Delta\mathbf{x}))^T$ is of order $(\Delta t)^2$, the covariance matrix V is set equal to $E(\Delta\mathbf{x}(\Delta\mathbf{x})^T)/\Delta t$. It is straightforward to show that V is positive definite and hence has a positive definite square root $B = V^{1/2}$. Hence, the vector $\boldsymbol{\mu}$ and the matrix V are defined as

$$\boldsymbol{\mu} = E(\Delta\mathbf{x})/\Delta t = \begin{bmatrix} b_1x_1 - d_1x_1 - m_{12}x_1 + m_{21}x_2 \\ \\ b_2x_2 - d_2x_2 - m_{21}x_2 + m_{12}x_1 \end{bmatrix} \tag{5.4}$$

and

$$V = \begin{bmatrix} b_1x_1 + d_1x_1 + m_{12}x_1 + m_{21}x_2 & -m_{12}x_1 - m_{21}x_2 \\ -m_{12}x_1 - m_{21}x_2 & b_2x_2 + d_2x_2 + m_{12}x_1 + m_{21}x_2 \end{bmatrix}. \quad (5.5)$$

For this two-dimensional system, $B = V^{1/2}$ can be found exactly and is given by

$$B = V^{1/2} = \frac{1}{d}\begin{bmatrix} a + w & b \\ b & c + w \end{bmatrix},$$

where $w = \sqrt{ac - b^2}$ and $d = \sqrt{a + c + 2w}$ with

$$\begin{aligned} a &= d_1x_1 + m_{12}x_1 + m_{21}x_2 + b_1x_1, \\ b &= -m_{12}x_1 - m_{21}x_2, \\ c &= m_{12}x_1 + d_2x_2 + b_2x_2 + m_{21}x_2. \end{aligned}$$

Based on the discussion in the introduction to this chapter, the stochastic differential equation model for the dynamics of two interacting populations has the form

$$d\mathbf{x} = \boldsymbol{\mu}(t, x_1, x_2)\, dt + B(t, x_1, x_2)\, d\mathbf{W}(t) \quad (5.6)$$

with $\mathbf{x}(0) = \mathbf{x}_0$ and where $\mathbf{W}(t)$ is the two-dimensional Wiener process, i.e., $\mathbf{W}(t) = [W_1(t), W_2(t)]^T$. Equation (5.6) is a stochastic differential equation system that describes the population dynamics. Notice that if matrix B is set equal to zero, then (5.6) reduces to a standard deterministic model for the population dynamics. Of course, $\boldsymbol{\mu} = \boldsymbol{\mu}(x_1, x_2, t)$ and $B = B(x_1, x_2, t)$ as the parameters b_i, d_i, m_{ij} may all depend on x_1, x_2, and t. For a single population, (5.6) reduces to

$$dx_1 = (b_1x_1 - d_1x_1)\, dt + \sqrt{b_1x_1 + d_1x_1}\, dW_1(t) \quad (5.7)$$

which is commonly seen in mathematical models of population dynamics [4, 8, 11, 12, 21].

5.2.2 Epidemic model and predator-prey model

In this section two examples of two interacting populations are considered. The first model is for an epidemic consisting of susceptible and infected subpopulations. The second model is for a predator-prey system.

Consider an SIS epidemic model for a single species. In this model, susceptibles become infected, recover, and become susceptible again. There is no immunity to the disease. A deterministic form of the SIS model is [10, 13, 15]:

$$\frac{dS}{dt} = \gamma I - \alpha IS/N \quad (5.8)$$

$$\frac{dI}{dt} = -\gamma I + \alpha IS/N, \quad (5.9)$$

where $S(0)+I(0)=N$ and therefore $S(t)+I(t)=N$ for $t \geq 0$. In this model, $S(t)$ is the susceptible population size, $I(t)$ is the infected population size, $\alpha > 0$ is the contact rate for passing the infection, i.e., the average number of individuals with whom an infected individual makes sufficient contact to pass an infection, and $\gamma > 0$ is the probability that an infected individual is removed from the infection process during a unit time interval (relative removal rate). In terms of the parameters defined in the first subsection,

$$x_1(t) = S(t), \;\; x_2(t) = I(t), \;\; d_1 = d_2 = b_1 = b_2 = 0,$$

$$m_{12} = \alpha I/(I+S) = \alpha x_2/(x_1+x_2), \;\; \text{and} \;\; m_{21} = \gamma.$$

Referring to (5.6), the stochastic SIS model has the form

$$dx_1 = (-m_{12}x_1 + m_{21}x_2)dt + \sqrt{\frac{m_{12}x_1 + m_{21}x_2}{2}}\big(dW_1 - dW_2\big) \quad (5.10)$$

$$dx_2 = (m_{12}x_1 - m_{21}x_2)dt + \sqrt{\frac{m_{12}x_1 + m_{21}x_2}{2}}\big(-dW_1 + dW_2\big) \quad (5.11)$$

and thus, for this problem,

$$B = \begin{bmatrix} m_{12}x_1 + m_{21}x_2 & -m_{12}x_1 - m_{21}x_2 \\ -m_{12}x_1 - m_{21}x_2 & m_{12}x_1 + m_{21}x_2 \end{bmatrix} \Big/ \sqrt{2(m_{12}x_1 + m_{21}x_2)}.$$

Note that the sum $x_1(t) + x_2(t)$ is constant for $t \geq 0$ in the stochastic model as well as in the deterministic model. Also, to simplify model (5.10)–(5.11), a Wiener process $W(t)$ can be substituted for the process $(W_1(t) - W_2(t))/\sqrt{2}$.

Table 5.3. Expected population sizes and the standard deviations in population sizes at time $t = 100$ for the stochastic SIS model

Estimate	SDE Model Results	Monte Carlo Results
$E(S(100))$	561.7	562.2
$E(I(100))$	438.7	437.8
$\sigma((S(100)))$	41.5	41.0
$\sigma((I(100)))$	41.5	41.0

For a computational example, let $\alpha = 0.04$, $\gamma = 0.01$, with $S(0) = 950$, and $I(0) = 50$. Let the final time be $t = 100$. In Table 5.3, the calculational results in solving (5.10)–(5.11) numerically are given for $E(S(100))$, $E(I(100))$, $\sigma((S(100))) = \sqrt{\text{Var}(S(100))}$, and $\sigma((I(100))) = \sqrt{\text{Var}(I(100))}$. In these computations, 10,000 sample paths were used to obtain the estimated means and standard deviations. Also listed in the table are the expected values and the standard deviations using a Monte Carlo approach. In the Monte

Carlo calculations, each individual in the population is checked after each time step of $\Delta t = 1/5$ day to determine whether a susceptible individual contracts the disease or if an infected individual recovers. These calculations continue until time $t = 100$ for 10,000 sample paths. As can be seen in the table, close agreement is obtained between the two different calculational approaches. Computer programs are listed at the end of this chapter for the two different calculational procedures. Finally, for this example, one sample path and the average of 100 sample paths for the SIS stochastic differential equation model (5.10)–(5.11) are displayed in Fig. 5.3. In addition, the computer program that produced this figure is listed at the end of this chapter.

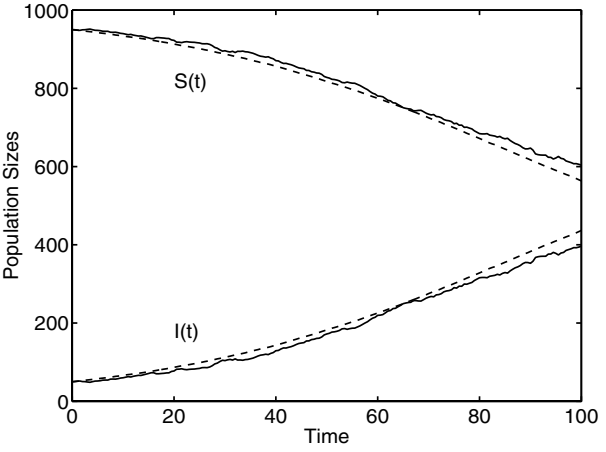

Fig. 5.3. Expected population sizes (dashed lines) and one sample path for susceptible and infected

As a second example of two interacting populations, consider the two species predator-prey system. A deterministic model for the predator and prey populations takes the form

$$dx_1(t) = \big(b_1(x_1, x_2)x_1(t) - d_1(x_1, x_2)x_1(t)\big)dt \tag{5.12}$$

$$dx_2(t) = \big(b_2(x_1, x_2)x_2(t) - d_2(x_1, x_2)x_2(t)\big)dt, \tag{5.13}$$

where $x_1(t)$ is the population size of the prey and $x_2(t)$ is the population size of the predator. For example, in the standard Lotka-Volterra model [15, 38],

$$b_1(x_1, x_2) = b_1, \quad d_1(x_1, x_2) = c_1 x_2,$$

$$b_2(x_1, x_2) = c_2 x_1, \quad \text{and} \quad d_2(x_1, x_2) = d_2,$$

where $b_1, d_2, c_1, c_2 > 0$. In this case, equations (5.12)–(5.13) have the equilibrium solution $x_1(t) = d_2/c_2$ and $x_2(t) = b_1/c_1$ (other than $x_1(t) = x_2(t) = 0$).

The solutions form closed curves in the x_1x_2-plane about the equilibrium point $(d_2/c_2,\ b_1/c_1)$. A variation of this model allows logistic prey growth [11, 12, 15, 38, 48] with

$$b_1(x_1, x_2) = \max(a_1(k_1 - x_1)/k_1, 0). \tag{5.14}$$

In this case, the steady-state solution is $x_1(t) = d_2/c_2$ and $x_2(t) = (a_1c_2k_1 - a_1d_2)/c_1c_2k_1$. The solutions spiral in the x_1x_2-plane to the steady state $\big(d_2/c_2,\ (a_1c_2k_1 - a_1d_2)/c_1c_2k_1\big)$.

Now, consider the stochastic differential equation model of the predator-prey system (5.12) and (5.13). For this problem, as $m_{12} = 0 = m_{21}$, the covariance matrix V is diagonal and the square root matrix B is then also diagonal. Based on equations (5.6), the stochastic predator-prey equations have the form:

$$dx_1 = (b_1(x_1, x_2) - d_1(x_1, x_2))x_1dt + \sqrt{(b_1(x_1, x_2) + d_1(x_1, x_2))x_1}dW_1 \tag{5.15}$$

$$dx_2 = (b_2(x_1, x_2) - d_2(x_1, x_2))x_2dt + \sqrt{(b_2(x_1, x_2) + d_2(x_1, x_2))x_2}dW_2 \tag{5.16}$$

for either the Lotka-Volterra model or the Lotka-Volterra model with logistic growth by substituting in the appropriate birth and death rates.

5.2.3 Persistence-time estimation

For a system of two interacting populations, the mean persistence time of the system is defined in this section to be the expected time it takes for the size of either population to reach zero. The mean persistence time of the system can be estimated through numerical solution of (5.6) or by solving the backward Kolmogorov differential equation to find the mean persistence time directly.

Individual populations can be simulated using (5.6) and followed until a population fails to persist. Averaging many such calculated persistence times yields an estimate for the mean persistence time. In computationally solving (5.6) the iterations are continued until either $x_1 \leq 0$ or $x_2 \leq 0$. When this exit criterion is satisfied, then the persistence time of that population is saved and another population trajectory is simulated. After many individual population trajectories are computed, the mean persistence time is estimated by averaging the individual persistence times.

Besides numerically solving stochastic system (5.6), the mean persistence time for system (5.6) can also be obtained by solving the backward Kolmogorov equation. Suppose that the size of population 1 cannot exceed K_1 and the size of population 2 cannot exceed K_2. The reliability function $R(t, y_1, y_2)$ is the probability that the persistence (exit) time is greater than t assuming initial populations $x_1(0) = y_1$ and $x_2(0) = y_2$. The reliability function is the integral of the transition probability density function over the safe region [94]. The reliability function $R(t, y_1, y_2)$ satisfies the backward Kolmogorov equation [12, 41, 81, 94, 104]:

$$\frac{\partial R}{\partial t} = \sum_{k=1}^{2} \mu_k(t, y_1, y_2) \frac{\partial R}{\partial y_k} + \frac{1}{2} \sum_{k=1}^{2} \sum_{m=1}^{2} v_{km}(t, y_1, y_2) \frac{\partial^2 R}{\partial y_k \partial y_m} \tag{5.17}$$

with

$$R(0, y_1, y_2) = 1 \quad \text{for} \quad (y_1, y_2) \in (0, K_1) \times (0, K_2)$$

$$R(t, 0, y_2) = 0 \quad \text{for} \quad y_2 \in (0, K_2)$$

$$R(t, y_1, 0) = 0 \quad \text{for} \quad y_1 \in (0, K_1)$$

$$\frac{\partial R(t, K_1, y_2)}{\partial y_1} = 0 \quad \text{for} \quad y_2 \in (0, K_2)$$

$$\frac{\partial R(t, y_1, K_2)}{\partial y_2} = 0 \quad \text{for} \quad y_1 \in (0, K_1)$$

and where $v_{km}(t, y_1, y_2) = \sum_{j=1}^{2} b_{kj}(t, y_1, y_2)\ b_{mj}(t, y_1, y_2)$. The probability density of persistence times is given by $-\dfrac{\partial R(t, y_1, y_2)}{\partial t}$ and $T(y_1, y_2) = \displaystyle\int_0^\infty R(t, y_1, y_2)\, dt$ is the mean persistence time with initial populations $x_1(0) = y_1$ and $x_2(0) = y_2$.

As a simple example, consider a single population with per capita birth and death rates $b(y)$ and $d(y)$, respectively, for $0 \le y \le K$. The corresponding backward Kolmogorov equation is

$$\frac{\partial R}{\partial t} = (yb(y) - yd(y)) \frac{\partial R}{\partial y} + \frac{1}{2}(yb(y) + yd(y)) \frac{\partial^2 R}{\partial y^2} \tag{5.18}$$

with $R(0, y) = 1$ for $y \in (0, K)$, $R(t, 0) = 0$, and $\dfrac{\partial R(t, K)}{\partial y} = 0$. Integrating (5.18) over time t from 0 to ∞ yields

$$-1 = (yb(y) - yd(y)) \frac{dT(y)}{dy} + \frac{1}{2}(yb(y) + yd(y)) \frac{d^2 T(y)}{dy^2} \tag{5.19}$$

with $T(0) = 0$ and $T'(K) = 0$, where $T(y)$ is the mean persistence time for a population of initial size y. Now suppose, for a computational comparison, that $K = 20$ and the birth and death rates satisfy

$$yb(y) - yd(y) = \begin{cases} 0, & \text{for } 0 \le y \le 10 \\ -1, & \text{for } 10 < y \le 20 \end{cases}$$

and

$$yb(y) + yd(y) = 1 \quad \text{for} \quad 0 \le y \le 20.$$

Then (5.19) can be solved exactly for the persistence time for an initial population size y to obtain that

$$T(y) = \begin{cases} -y^2 + (21 - e^{-20})y, & \text{for} \quad 0 \le y \le 10 \\ y + 100 - (19e^{-20} + e^{-40}e^{2y})/2, & \text{for} \ \ 10 < y \le 20. \end{cases}$$

In particular, $T(5) = 80.0$, $T(10) = 110.0$, and $T(15) = 115.0$. For comparison purposes, the stochastic differential equation

$$\begin{cases} dx = \big(b(x)x - d(x)x\big)\, dt + \sqrt{(b(x)x + d(x)x)}\, dW(t) \\ x(0) = y \end{cases}$$

was computationally solved using the same birth and death rates. Calculating 10,000 sample paths, where each path was terminated when $x(t) \le 0$, resulted in the estimates $T(5) \approx 80.07$, $T(10) \approx 111.05$, and $T(15) \approx 114.01$. Thus, persistence times calculated using the stochastic differential equation showed good agreement with those obtained by the backward Kolmogorov equation.

5.2.4 A population model with a time delay

Delay equations are common in mathematical biology. For example, Cushing [33] analyzed a predator-prey model with a delay in the response of the predator to the changing density of the prey and Gopalsamy [49] studied a system with a developmental delay such as the delay required for insect larvae to become adults. Models of epidemics with continuously distributed time delays are also of interest [57]. Consider a simple SIS model where there are no births or deaths in the system but there is a delay in the recovery rate of infected individuals. A deterministic SIS epidemic model with a delay in recovery from infected to susceptible has the form:

$$\begin{aligned} \frac{dS}{dt} &= -\lambda SI + \gamma \int_{t-T}^{t} I(u)H(t-u)\, du \\ \frac{dI}{dt} &= \lambda SI - \gamma \int_{t-T}^{t} I(u)H(t-u)\ \, du, \end{aligned}$$

where λ is the rate of contracting the disease from an infected individual and the time T and function H are employed to account for the delay in the recovery of infected individuals. For example, $H(z)$ may have the form

$$H(z) = \begin{cases} \dfrac{1}{T}, & \text{for} \ \ 0 \le z \le T \\ 0, & \text{otherwise} \end{cases} \quad \text{or} \quad H(z) = \begin{cases} \dfrac{2z}{T^2}, & \text{for} \ \ 0 \le z \le T \\ 0, & \text{otherwise,} \end{cases}$$

where H is often assumed to satisfy the condition $\int_0^\infty H(z)\,dz = 1$.

By considering the changes in a small time interval Δt as described in the introduction to this chapter, a stochastic system for the SIS model with delay in recovery is obtained of the form

$$\frac{dS}{dt} = -\lambda SI + \gamma \int_{t-T}^{t} I(u)H(t-u)\,du + b_{11}\frac{dW_1}{dt} + b_{12}\frac{dW_2}{dt}$$
$$\frac{dI}{dt} = \lambda SI - \gamma \int_{t-T}^{t} I(u)H(t-u)\,du + b_{21}\frac{dW_1}{dt} + b_{22}\frac{dW_2}{dt},$$

where

$$b_{11} = b_{22} = -b_{21} = -b_{12} = \frac{1}{\sqrt{2}}\sqrt{\lambda SI + \gamma \int_{t-T}^{t} I(u)H(t-u)\,du}.$$

It is interesting to note that

$$\frac{dS(t)}{dt} + \frac{dI(t)}{dt} = 0 \quad \text{so that} \quad S(t) + I(t) = S(0) + I(0) \quad \text{for} \quad t \geq 0.$$

This is the same relation that the deterministic and the stochastic SIS models satisfy without delay. Finally, it is useful to note that numerical methods have been developed for solving stochastic integro-differential problems. (See, for example, [46, 47].)

5.2.5 A model including environmental variability

In the previous models of population dynamics, the randomness in the births, deaths, and interactions accounted for the nonzero terms in the covariance matrix. However, the environment also varies which is likely to affect the populations in another random manner. One possible way to model the environmental effects would be to include additional variables in the mathematical model such as rainfall, predator populations, competitor populations, and food supply. Including additional variables, however, would quickly complicate the model and destroy the simple nature of the model. Consider the deterministic model for the growth of a single population of size $y(t)$:

$$\frac{dy}{dt} = b(t)y - d(t)y.$$

In a varying environment, the per capita birth and death rates, $b(t)$ and $d(t)$, would be functions of these additional environmental variables and would have the forms $b(t, v_1, v_2, \ldots, v_n)$ and $d(t, v_1, v_2, \ldots, v_n)$, respectively, where $v_1, v_2, \ldots, v_n$ represent n different environmental variables. Thus, as $v_1, v_2, \ldots, v_n$ vary, the per capita birth and death rates also vary. This suggests that an approximate way to include environmental variability, without

modeling additional environmental factors, would be to vary the per capita birth and death rates in a random manner. Thus, in this section, it is hypothesized that changes in the environment produce random changes in a population's per capita birth and death rates that are independent from the changes due to demographic variability. This hypothesis may provide only a rough approximation to the actual biological situation. However, accepting this hypothesis leads to manageable mathematical models that can give some insight into the effects of environmental variability on the dynamics of a population [8].

Table 5.4. Possible changes in the population size and per capita birth and death rates with the corresponding probabilities

Change	Probability
$\Delta y_1 = -1$	$p_1 = dy\Delta t$
$\Delta y_2 = 1$	$p_2 = by\Delta t$
$\Delta y_3 = 0$	$p_3 = 1 - (by + dy)\Delta t$
$\Delta b_1 = -\alpha_b$	$p_4 = (q_b - \beta_b(b_e - b))\Delta t$
$\Delta b_2 = \alpha_b$	$p_5 = (q_b + \beta_b(b_e - b))\Delta t$
$\Delta b_3 = 0$	$p_6 = 1 - 2q_b\Delta t$
$\Delta d_1 = -\alpha_d$	$p_7 = (q_d - \beta_d(d_e - d))\Delta t$
$\Delta d_2 = \alpha_d$	$p_8 = (q_d + \beta_d(d_e - d))\Delta t$
$\Delta d_3 = 0$	$p_9 = 1 - 2q_d\Delta t$

First, as in the previous models, a discrete stochastic process is described for the phenomenon. Let $y(t), b(t)$, and $d(t)$ be the population size, per capita birth rate, and per capita death rate for the population at time t, respectively. Let Δt be a small time interval. The changes in these three variables in time Δt can be considered to be independent under our hypothesis. The possible changes $\Delta y, \Delta b$ and Δd are listed in Table 5.4. A special form is assumed for the probabilities of the per capita birth rates and death rates. Consider here the per capita birth rate as the per capita death rate can be regarded in the same manner. It is assumed that $q_b\Delta t$ represents the probability associated with random diffusion of the per capita birth rate. The term $\pm\beta_b(b_e - b)$ represents the probability associated with drift toward the mean value of b_e. When $b(t) \neq b_e$, where b_e is the average per capita birth rate in the environment, then the probability of moving closer to b_e is greater than the probability of moving further away from b_e. In this way, unrealistic values for the per capita birth and death rates are avoided.

The next step is to find the expected change and the covariance matrix for the change. To order $(\Delta t)^2$, the expected values satisfy

$$
\begin{aligned}
E(\Delta y) &= (b(t) - d(t))y(t)\Delta t \\
E((\Delta y)^2) &= (b(t) + d(t))y(t)\Delta t \\
E(\Delta b) &= 2\alpha_b\beta_b(b_e - b(t))\Delta t \\
E((\Delta b)^2) &= 2\alpha_b^2 q_b \Delta t \\
E(\Delta d) &= 2\alpha_d\beta_d(d_e - d(t))\Delta t \\
E((\Delta d)^2) &= 2\alpha_d^2 q_d \Delta t.
\end{aligned}
$$

Now, for convenience, define $\beta_1, \beta_2, \alpha_1$, and α_2 as

$$
\beta_1 = 2\alpha_b\beta_b, \quad \beta_2 = 2\alpha_d\beta_d, \quad \alpha_1^2 = 2\alpha_b^2 q_b, \quad \text{and} \quad \alpha_2^2 = 2\alpha_d^2 q_d.
$$

As the covariance matrix is diagonal for this model, the following stochastic differential equation system is obtained:

$$
\begin{cases}
dy(t) = \big(b(t)y(t) - d(t)y(t)\big)\,dt + \sqrt{\big(b(t)y(t) + d(t)y(t)\big)}\,dW_1(t) \\
db(t) = \beta_1(b_e - b(t))\,dt + \alpha_1\,dW_2(t) \\
dd(t) = \beta_2(d_e - d(t))\,dt + \alpha_2\,dW_3(t)
\end{cases}
\tag{5.20}
$$

for $\big(y(t), b(t), d(t)\big) \in [0, \infty) \times \mathbb{R}^1 \times \mathbb{R}^1$, where $W_i(t), i = 1, 2, 3$ are independent standard Wiener processes. System (5.20) represents a stochastic model of a single population experiencing variability in the environment.

It is interesting that the stochastic differential equation for $b(t)$ (or $d(t)$) can be solved exactly to yield

$$
b(t) = b_e + \exp(-\beta_1 t)\Big(- b_e + b(0) + \int_0^t \alpha_1 \exp(\beta_1 s)\,dW_2(s)\Big).
$$

This equation implies that, for large time t, the per capita birth rate $b(t)$ is approximately normally distributed with mean b_e and variance $\alpha_1^2/(2\beta_1)$. Thus, it is inherently assumed in this stochastic model that random variations in the environment cause the per capita birth rate to vary normally about a mean value b_e. An analogous result is assumed, of course, for the per capita death rate.

As a computational example, assume $\beta_1 = 1 = \beta_2$, $\alpha_1 = 0.5 = \alpha_2$, $b_e = 1$, $d_e = 1.4$. Suppose that $y(0) = 30$, $b(0) = b_e$, and $d(0) = d_e$. In Fig. 5.4, the probability distribution of the population size $y(t)$ that satisfies stochastic system (5.20) is plotted at time $t = 1.0$. Two different cases are considered, that is, when there is environmental variability and when there is no environmental variability. No environmental variability is obtained by setting $\beta_1 = \beta_2 = \alpha_1 = \alpha_2 = 0$. It is clear for this example that the variability in the environment tends to spread out the population size distribution.

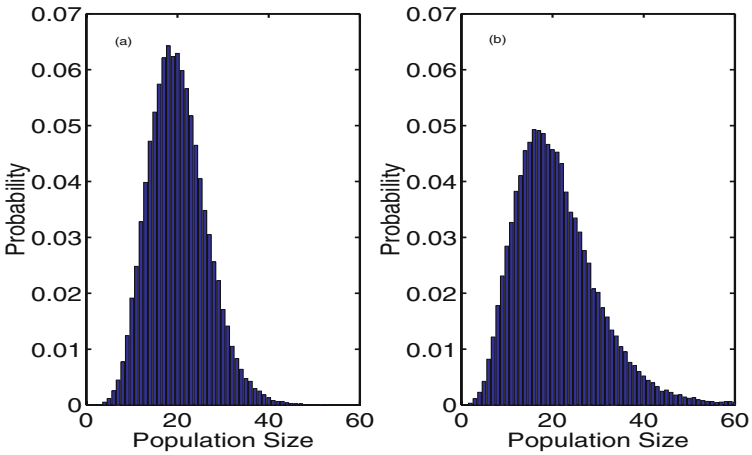

Fig. 5.4. Population distribution at $t = 1$ for an initial population of size 30 obtained through numerical solution of the SDE model (5.20) with no environmental variability (a) and with environmental variability (b)

5.3 Physical Systems

Stochastic differential equation models for seven physical systems that experience random influences in their dynamical behavior will be examined. The dynamical systems discussed include mechanical vibration, seed dispersal, nuclear reactor dynamics, ion transport, precipitation, chemical molecular reactions, and cotton fiber breakage.

5.3.1 Mechanical vibration

The first dynamical physical system considered is mechanical vibration, specifically, a single-degree-of-freedom vibrating system. A single-degree-of-freedom system only requires one coordinate to describe the state of the system at any time. In mechanical vibration, the general governing equation of motion for a single-degree-of-freedom dynamic system has the form [81, 94]:

$$x''(t) + r(x) + c(x, x') = f(t), \tag{5.21}$$

where $x(t)$ is the displacement of the mass from the rest point, $r(x)$ represents restoring forces, $c(x, x')$ models damping forces, and $f(t)$ is a stochastic excitation process. Engineering applications of stochastic differential equation (5.21) arise, for example, in reliability analyses of structures subject to wind, current, or earthquake loads. We will develop a model of the form (5.21) for the spring-mass system, illustrated in Fig. 5.5, where $x(t)$ is the displacement of the mass from equilibrium, $v(t) = x'(t)$ is the velocity, m is the mass, and $M(t) = mv(t)$ is the momentum. Let $kx(t)$ be the spring displacement force

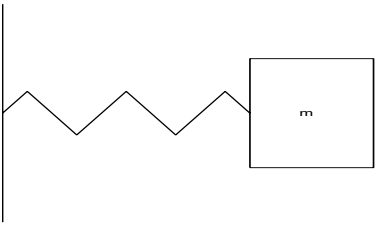

Fig. 5.5. A single-degree-of-freedom spring-mass system

and let $bv(t)$ be the damping force or the force due to friction. The deterministic system of ordinary differential equations that describe the dynamics of the spring-mass system has the form

$$\begin{cases} \dfrac{dx(t)}{dt} = v(t) \\ \dfrac{dM(t)}{dt} = -kx(t) - bv(t) \end{cases} \tag{5.22}$$

with $x(0) = x_0$ and $v(0) = v_0$. Notice that for a small time interval

$$\Delta M = M(t + \Delta t) - M(t) \approx -\big(kx(t) + bv(t)\big)\Delta t.$$

Next assume that impulsive forces are applied randomly to the mass. Specifically, in any small time interval Δt, assume that the probability equals $\lambda \Delta t$ for a positive or negative change in the momentum of magnitude γ. Let ΔM be the change in the momentum in time Δt. There are three possible changes in time Δt and these are listed in Table 5.5 along with their respective probabilities. Notice that for any change in momentum, the term $(-kx - bv)\Delta t$ is included.

Table 5.5. Possible changes in the momentum of the spring-mass system with the corresponding probabilities

Change in Momentum	Probability
$\Delta M_1 = \gamma + (-kx - bv)\Delta t$	$p_1 = \lambda \Delta t + o(\Delta t)$
$\Delta M_2 = -\gamma + (-kx - bv)\Delta t$	$p_2 = \lambda \Delta t + o(\Delta t)$
$\Delta M_3 = (-kx - bv)\Delta t$	$p_3 = 1 - p_1 - p_2$

The next step involves calculating the expected change and the variance in the change for the time interval Δt. To order $(\Delta t)^2$ it follows that

$$E(\Delta M) = (-kx - bv)\Delta t$$

and

$$E((\Delta M)^2) = 2\gamma^2\lambda\Delta t.$$

These values for $E(\Delta M)$ and $E((\Delta M)^2)$ imply that the stochastic differential equation model for the vibrating mass under the conditions assumed in this section for random changes in the momentum has the form:

$$\begin{cases} dx(t) = v(t)dt \\ \\ dM(t) = \big(-kx(t) - bv(t)\big)\,dt + \sqrt{2\gamma^2\lambda}\,dW(t) \end{cases} \tag{5.23}$$

with $x(0) = x_0$ and $v(0) = v_0$. For mass m constant, system (5.23) can be written as the second-order stochastic system

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = \sqrt{2\gamma^2\lambda}\,\frac{dW}{dt} \tag{5.24}$$

with $x(0) = x_0$ and $x'(0) = v_0$. Equations of the form (5.24) are well-known in the study of random vibration for mechanical systems [81, 94]. Also, equation (5.24) is a form of stochastic Liénard equation [103].

5.3.2 Seed dispersal

Consider estimating the horizontal distance that certain seeds traverse when falling from a given height under the influence of a randomly varying wind. It is assumed that the seeds experience a frictional force proportional to the square of the speed of the air on the seeds.

Consider first a deterministic model where the wind speed is constant and not randomly varying. Let $\mathbf{v}_w(t) = v_w(t)\mathbf{i}$ be the wind velocity, $\mathbf{v}(t) = v_x(t)\mathbf{i} + v_y(t)\mathbf{j}$ be the seed velocity, and $\mathbf{v}_a(t) = \mathbf{v}_w(t) - \mathbf{v}(t)$ be the air velocity on the seed at time t. Let $\mathbf{F}_f(t) = k|\mathbf{v}_a(t)|^2\dfrac{\mathbf{v}_a(t)}{|\mathbf{v}_a(t)|} = k|\mathbf{v}_a(t)|\mathbf{v}_a(t)$ be the frictional force on the seed at time t, where k is a constant of proportionality. Let $\mathbf{F}_g(t) = -mg\mathbf{j}$ be the force of gravity on the seed of mass m. Finally, let $\mathbf{s}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$ be the position of the seed at time t. It is assumed that $\mathbf{v}(0) = \mathbf{0} = 0\mathbf{i} + 0\mathbf{j}$ and $\mathbf{s}(0) = h\mathbf{j}$ where h is the initial height. It is straightforward to check that the velocity and the position of the seed at any time t satisfy the initial-value system

$$\begin{cases} \dfrac{d\mathbf{s}(t)}{dt} = \mathbf{v}(t) \\ \\ \dfrac{d\mathbf{v}(t)}{dt} = \dfrac{1}{m}\big(\mathbf{F}_f(t) + \mathbf{F}_g(t)\big) = \dfrac{k}{m}|\mathbf{v}_a(t)|\mathbf{v}_a(t) - g\mathbf{j} \\ \\ \mathbf{v}_a(t) = \mathbf{v}_w(t) - \mathbf{v}(t) \ \text{ with } \ \mathbf{v}_w(t) = v_w\mathbf{i} \\ \\ \mathbf{s}(0) = h\mathbf{j}, \ \ \mathbf{v}(0) = \mathbf{0}. \end{cases} \tag{5.25}$$

Now suppose that the wind speed randomly varies about a mean speed v_e. In particular, the wind speed can experience a change of $\pm\alpha$ during a small time interval Δt with the probabilities listed in Table 5.6, where it is assumed that Δt is sufficiently small so that $p_1, p_2 > 0$. (Notice that p_1 and p_2 can be defined as $p_l = \max\{0, \big(\lambda + (-1)^l \beta(v_e - v_w(t))\big)\Delta t\}$ for $l = 1, 2$ to guarantee nonnegativity of the probabilities.) The value $\lambda \Delta t$ represents the probability associated with random diffusion of the wind speed and does not depend on $v_w(t)$. The term $\pm\beta\big(v_e - v_w(t)\big)$ represents the probability associated with drift toward the mean wind speed of v_e. When $v_w(t) \neq v_e$, the probability that the wind speed moves closer to v_e is greater than the probability that the wind speed moves further from v_e. Thus, v_e can be regarded as a mean wind speed.

Table 5.6. Possible changes in the wind speed with the corresponding probabilities

Change Δv_w	Probability
$(\Delta v_w)_1 = -\alpha$	$p_1 = \big(\lambda - \beta(v_e - v_w(t))\big)\Delta t$
$(\Delta v_w)_2 = \alpha$	$p_2 = \big(\lambda + \beta(v_e - v_w(t))\big)\Delta t$
$(\Delta v_w)_3 = 0$	$p_3 = 1 - p_1 - p_2 = 1 - 2\lambda\Delta t$

The next step is to find the mean wind speed change and the variance in the change. It is straightforward to show that the required expectations to order $(\Delta t)^2$ are

$$E(\Delta v_w) = 2\alpha\beta\big(v_e - v_w(t)\big)\Delta t$$

and

$$E((\Delta v_w)^2) = 2\alpha^2\lambda\Delta t.$$

Based on the arguments in the introduction to this chapter, a reasonable stochastic differential equation model for the randomly varying wind speed is

$$dv_w(t) = 2\alpha\beta\big(v_e - v_w(t)\big)\,dt + \sqrt{2\alpha^2\lambda}\,dW(t). \tag{5.26}$$

Indeed, this stochastic differential equation can be solved exactly to yield

$$v_w(t) = v_e + \exp(-2\alpha\beta t)\Big(-v_e + v_w(0) + \int_0^t \sqrt{2\alpha^2\lambda}\exp(2\alpha\beta s)\,dW(s)\Big).$$

This solution implies, for large time t, that the wind speed $v_w(t)$ is approximately normally distributed with mean v_e and variance $2\alpha^2\lambda/(4\alpha\beta) = \alpha\lambda/(2\beta)$.

The complete model for the seed dispersal dynamics in a randomly varying wind is now formulated as

$$\begin{cases} \dfrac{d\mathbf{s}(t)}{dt} = \mathbf{v}(t) \\ \dfrac{d\mathbf{v}(t)}{dt} = \dfrac{1}{m}\big(\mathbf{F}_f(t) + \mathbf{F}_g(t)\big) = \dfrac{k}{m}|\mathbf{v}_a(t)|\mathbf{v}_a(t) - g\mathbf{j} \\ \dfrac{dv_w(t)}{dt} = 2\alpha\beta\big(v_e - v_w(t)\big) + \sqrt{2\alpha^2\lambda}\,\dfrac{dW(t)}{dt} \\ \mathbf{v}_a(t) = \mathbf{v}_w(t) - \mathbf{v}(t) \\ \mathbf{s}(0) = h\mathbf{j}, \;\; \mathbf{v}(0) = \mathbf{0}, \;\; \mathbf{v}_w(0) = v_w(0)\mathbf{i}. \end{cases} \tag{5.27}$$

This problem illustrates the ease with which a deterministic model can be transformed into a stochastic differential equation model for certain physical problems after agreeing upon a discrete stochastic model. Notice that there are other interesting possibilities for modeling the wind speed besides the model described here. For example, in constructing the probabilities for the wind speed changes, consider letting $p_l = \big(\lambda + (-1)^l\beta(v_e - v_w(t))^{k_1/k_2}\big)\Delta t$ for $l = 1, 2$ for some positive odd integers k_1 and k_2. Selecting the parameter values and the discrete stochastic process that best represents a given physical situation may involve many computational comparisons with physical data.

5.3.3 Ion transport

In this section, another kind of procedure for developing a stochastic differential equation model is discussed. The procedure in this section is different from the procedure described in the introduction to this chapter. In this section, a stochastic differential equation model is formulated for a problem in ion transport. When the system of stochastic differential equations is solved numerically, the resulting numerical method is referred to as the random particle method [2, 54, 66, 114]. A similar procedure is applied in fluid dynamics and is called the random vortex method. (See, for example, [31].)

Considered in this section is the motion of ions in a spatially uniform plasma with magnetic field forces negligible. The distribution of ions in the plasma is governed by the Vlasov-Poisson-Fokker-Planck equations:

$$\begin{aligned} \frac{\partial f_n(t,\mathbf{v})}{\partial t} &= \left(\frac{\partial f_n}{\partial t}\right)_c = \\ &= -\sum_{i=1}^{3}\frac{\partial}{\partial v_i}\big(A_i^n(t,\mathbf{v})f_n(t,\mathbf{v})\big) + \frac{1}{2}\sum_{i,j=1}^{3}\frac{\partial^2}{\partial v_i\partial v_j}\big(B_{ij}^n(t,\mathbf{v})f_n(t,\mathbf{v})\big) \end{aligned} \tag{5.28}$$

for $n = 1, 2, \ldots, N$, where $f_n(0, \mathbf{v})$ is given for $n = 1, 2, \ldots, N$. The function $f_n(t, \mathbf{v})$ is the distribution of ions of kind n with respect to time $t > 0$ and velocity $\mathbf{v} \in \mathbb{R}^3$. Therefore,

$$\int_{\mathbb{R}^3} f_n(t, \mathbf{v})\, d\mathbf{v} = 1 \quad \text{for} \quad t \geq 0$$

and $\left(\dfrac{\partial f_n}{\partial t}\right)_c$ is the rate of change due to collisions. For an inverse square force due to Coulombic interactions between ions, the coefficients have the forms

$$A_i^n(t, \mathbf{v}) = -\frac{1}{4\pi} \sum_{k=1}^{N} \frac{m_n + m_k}{m_k} L_{n,k} \int_{\mathbb{R}^3} \frac{(\mathbf{v} - \mathbf{v}')_i}{|\mathbf{v} - \mathbf{v}'|^3} f_k(t, \mathbf{v}')\, d\mathbf{v}'$$

and

$$B_{ij}^n(t, \mathbf{v}) = \frac{1}{4\pi} \sum_{k=1}^{N} L_{n,k} \int_{\mathbb{R}^3} \frac{f_k(t, \mathbf{v}')}{|\mathbf{v} - \mathbf{v}'|} \left[\delta_{ij} - \frac{(\mathbf{v} - \mathbf{v}')_i (\mathbf{v} - \mathbf{v}')_j}{|\mathbf{v} - \mathbf{v}'|^2}\right] d\mathbf{v}',$$

where $L_{n,k}$ are physical constants [66, 114]. and m_k for $k = 1, 2, \ldots, N$ are ion masses.

From the previous chapter and the introduction to this chapter, these Vlasov-Poisson-Fokker-Planck equations match the forward Kolmogorov equations whose solution gives the probability distribution for a certain system of stochastic differential equations. In particular, the solutions to the following stochastic differential equation system have a probability distribution satisfied by $f_n(t, \mathbf{v})$ of (5.28),

$$d\mathbf{v}^n(t) = \mathbf{A}^n(t, \mathbf{v}^n)\, dt + D^n(t, \mathbf{v}^n)\, d\mathbf{W}^n(t) \tag{5.29}$$

for $n = 1, 2, \ldots, N$ with $\mathbf{v}^n(0)$ distributed according to $f_n(0, \mathbf{v})$. Also, $D^n = (B^n)^{1/2}$ is a 3×3 positive definite matrix and $\mathbf{W}^n(t) = [W_1^n(t), W_2^n(t), W_3^n(t)]^T$ is a three-dimensional Wiener process. As the probability distribution of $\mathbf{v}^n(t)$ is $f_n(t, \mathbf{v})$, this suggests a numerical procedure for approximating $f_n(t, \mathbf{v})$. In the random particle method, (5.29) is solved numerically to estimate $f_n(t_i, \mathbf{v})$ for $n = 1, 2, \ldots, N$ at a set of discrete times $t_1, t_2, \ldots, t_M$. More information on the random particle method and the related random vortex method is available, for example, in references [2, 31, 54, 66, 114].

5.3.4 Nuclear reactor kinetics

In this section, the time-dependent behavior of a nuclear reactor is modeled. Several simplifying assumptions are made. First, it is assumed that the nuclear reactor is large and homogeneous so that spatial effects can be ignored. Second, it is assumed that the neutrons in the reactor have the same energy, that is, only one neutron energy group is treated. Third, it is assumed that only one kind of fission-product isotope decays by neutron emission.

The time-dependent behavior of a nuclear reactor is governed by the neutron population $n(t)$ [25, 58]. With the above assumptions, Fig. 5.6 illustrates

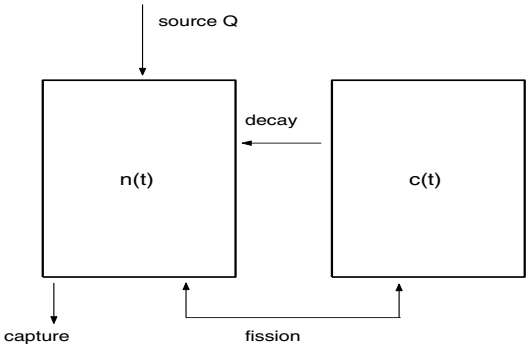

Fig. 5.6. Schematic diagram of a nuclear reactor dynamical system

the important dynamic processes affecting the neutron population. In the figure, an extraneous neutron source provides an average rate of Q neutrons per second. The variable $c(t)$ is the number of atoms at time t of a radioactive isotope which spontaneously decay by neutron emission. The isotope $c(t)$ is called a fission product and is formed in fission events. To describe the dynamics of a nuclear reactor, the following physical parameters are useful.

σ_f = probability per unit distance for a neutron to cause a fission

σ_c = probability per unit distance of a neutron loss by capture in an atom

ν = total number of neutrons per fission ($(1-\beta)\nu$ prompt and $\beta\nu$ delayed)

λ = decay constant of fission product $c(t)$

$\beta\nu$ = number of atoms of fission product $c(t)$ produced per fission

v = neutron speed

It is important to consider the changes $[\Delta n, \Delta c]^T$ in the neutron and fission product levels which can occur in time interval Δt along with their probabilities. These changes and their probabilities are listed in Table 5.7. Notice, for example, that $[-1, 0]^T$ is the loss of a neutron due to capture in an atom when no fission occurs, and $[(1-\beta)\nu - 1, \beta\nu]^T$ is a fission event where $\beta\nu$ atoms of isotope c are produced and $(1-\beta)\nu - 1$ neutrons are immediately produced. Notice that $(1-\beta)\nu$ of these neutrons are new neutrons but one neutron is lost in the fission event so $(1-\beta)\nu - 1$ neutrons result from a fission. In addition, the extraneous neutron source is assumed to provide neutrons at the average rate Q but the number in any interval Δt is Poisson distributed so the probability of one source neutron in time interval Δt is $Q\Delta t + O((\Delta t)^2)$.

Table 5.7. Possible changes in the neutron and fission product levels with the corresponding probabilities

Change $[\Delta n, \Delta c]^T$	Probability
$[\Delta n, \Delta c]_1^T = [1, 0]^T$	$p_1 = Q\Delta t$
$[\Delta n, \Delta c]_2^T = [-1, 0]^T$	$p_2 = \sigma_c n(t) v \Delta t$
$[\Delta n, \Delta c]_3^T = [1, -1]^T$	$p_3 = \lambda c(t) \Delta t$
$[\Delta n, \Delta c]_4^T = [(1-\beta)\nu - 1, \beta\nu]^T$	$p_4 = \sigma_f n(t) v \Delta t$
$[\Delta n, \Delta c]_5^T = [0, 0]^T$	$p_5 = 1 - \sum_{i=1}^4 p_i$

Applying the values in Table 5.7, we can find the expected change and the covariance matrix for the change. Indeed,

$$E\begin{bmatrix}\Delta n\\ \Delta c\end{bmatrix} = \sum_{i=1}^{4} p_i \begin{bmatrix}\Delta n\\ \Delta c\end{bmatrix}_i = \begin{bmatrix}\lambda c + ((1-\beta)\nu - 1)n\sigma_f v - n\sigma_c v + Q\\ \beta\nu n\sigma_f v - \lambda c\end{bmatrix}\Delta t$$

and

$$E\left(\begin{bmatrix}\Delta n\\ \Delta c\end{bmatrix}\begin{bmatrix}\Delta n, \Delta c\end{bmatrix}\right) = \sum_{i=1}^{4} p_i \left(\begin{bmatrix}\Delta n\\ \Delta c\end{bmatrix}_i \begin{bmatrix}\Delta n, \Delta c\end{bmatrix}_i\right) = V(n,c)\Delta t,$$

where

$$V = V(n,c) =$$

$$\begin{bmatrix}\lambda c + ((1-\beta)\nu - 1)^2 n\sigma_f v + n\sigma_c v + Q & -\lambda c + \beta\nu((1-\beta)\nu - 1)n\sigma_f v\\ -\lambda c + \beta\nu((1-\beta)\nu - 1)n\sigma_f v & \lambda c + \beta^2\nu^2 n\sigma_f v\end{bmatrix}.$$

These expectations lead directly to the stochastic differential equation system

$$d\begin{bmatrix}n\\ c\end{bmatrix} = \begin{bmatrix}\lambda c + ((1-\beta)\nu - 1)n\sigma_f v - n\sigma_c v + Q\\ \beta\nu n\sigma_f v - \lambda c\end{bmatrix} dt + V^{1/2}\, d\mathbf{W}, \qquad (5.30)$$

where $(V(n,c))^{1/2}$ is the square root of matrix $V(n,c)$ and $\mathbf{W}(t) = [W_1(t), W_2(t)]^T$ is the two-dimensional Wiener process.

As a numerical example, consider the values $\lambda = 0.077$/sec, $\beta = 0.0079$, $\nu = 2.432$, $\sigma_f v = 4111.84$/sec, $\sigma_c v = 5858.16$/sec, and $Q = 10,000$/sec which are reasonable values for a nuclear reactor involving thermal fission of a uranium-235 fuel [58]. (These parameter values give a neutron lifetime of 10^{-4} seconds and a reactivity of 0.003.) For this example, the reaction is self-sustaining and the neutron level, $n(t)$, and the fission product level, $c(t)$, are increasing with time. A plot of the average neutron population from $t = 0$ to $t = 0.1$ seconds is shown in Fig. 5.7 for 10,000 sample paths. In the same figure, the neutron level for one sample path is plotted. The figure illustrates

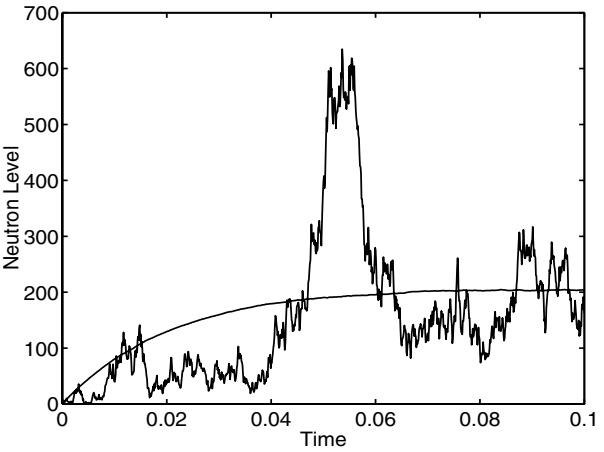

Fig. 5.7. Mean neutron level and one sample path for a nuclear reactor system with time in seconds

how much the neutron level can vary, for example, at reactor start-up. Furthermore, in Table 5.8, the calculational results are presented for the mean neutron level and mean fission product level with their standard deviations at time $t = 0.1$ seconds. In the table, the calculational results are listed using the stochastic differential equation model (5.30) along with those obtained using a Monte Carlo approach. In either method, 10,000 sample paths were calculated. In the Monte Carlo procedure, at each time step of length 10^{-7} seconds, the system is checked to see if a fission, decay, or capture occurs or if a source neutron is produced. As can be seen from the values in Table 5.8, there is good agreement between the two different calculational procedures.

Table 5.8. A comparison of mean neutron level and mean fission product level with their standard deviations calculated using SDE model (5.30) and a Monte Carlo method at $t = 0.1$ seconds

Value Estimated	SDE Model (5.30)	Monte Carlo
$E(n(0.1))$	204.52	199.15
$\sqrt{\text{Var}(n(0.1))}$	174.03	152.63
$E(c(0.1))$	1294.0	1254.5
$\sqrt{\text{Var}(c(0.1))}$	620.68	613.94

Finally, it is worthwhile to note that the procedure in this section was applied to model the Godiva reactor [55] to determine the time it takes for the neutron level to reach 4200 neutrons with an extraneous source of 90 neutrons/seconds. For this model, however, six delayed fission products were

modeled rather than one in stochastic system (5.30). The results of 22 experimental times [52] and 500 calculated times [55] using the stochastic differential model are displayed as relative frequency histograms in Fig. 5.8. The figure indicates that the stochastic differential equation model accurately estimates the distribution of times for the Godiva reactor.

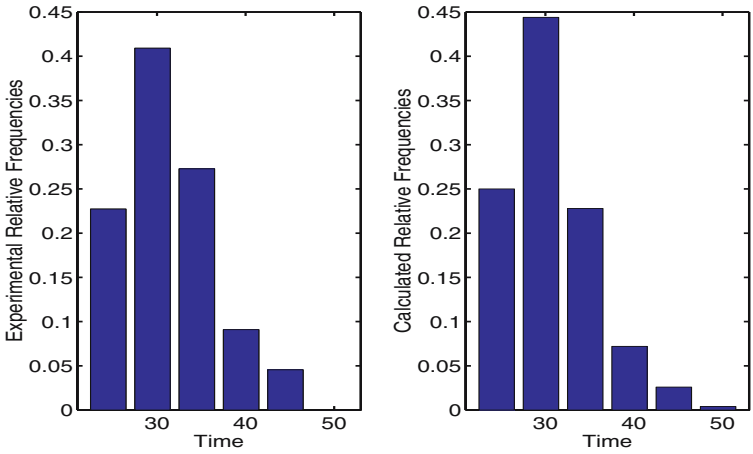

Fig. 5.8. Experimental and calculated times (in seconds) to reach a level of 4200 neutrons in the Godiva reactor

5.3.5 Precipitation

As a fifth physical problem, a simple stochastic model for the rainfall at a certain location over a period of decades is developed. In this situation, a small time interval may be considered as one year. Let $r(t)$ be the total rainfall at time t where $r(0) = 0$ for a starting time $t = 0$. Consider the possible changes in the total rainfall over a very small interval of time Δt. The possible changes in the rainfall for time interval Δt are assumed to be $(\Delta r)_1 = \gamma$ and $(\Delta r)_2 = 0$ with probabilities $p_1 = \lambda \Delta t$ and $p_2 = 1 - \lambda \Delta t$, respectively. Assuming these changes and probabilities, the mean change and the mean square change in the total rainfall are

$$E(\Delta r) = \gamma \lambda \Delta t \quad \text{and} \quad E((\Delta r)^2) = \gamma^2 \lambda \Delta t.$$

Letting $\mu = \gamma \lambda$ and $\sigma = \sqrt{\gamma^2 \lambda}$, the above expectations leads to the stochastic differential equation model for total rainfall $r(t)$ of the form

$$\begin{cases} dr(t) = \mu dt + \sigma dW(t) \\ r(0) = 0. \end{cases} \tag{5.31}$$

To test this model, the annual precipitation data for Lubbock, Texas was obtained from the National Weather Service and is plotted in Fig. 5.9. This data was fit to the stochastic differential equation model using a maximum likelihood estimation procedure discussed in Section 4.9. The values obtained for the parameters μ and σ were $\mu = 18.57$ and $\sigma = 6.11$. Now consider the

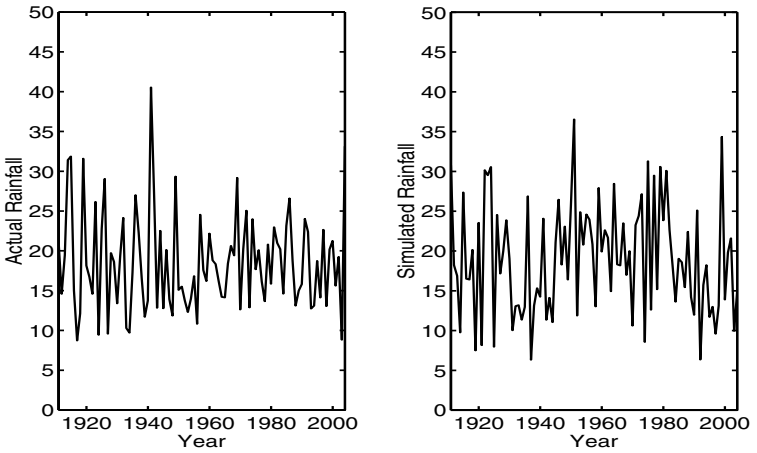

Fig. 5.9. Actual and simulated annual rainfall (in inches) for Lubbock, Texas from 1911 to 2004

annual precipitation p. Note that by (5.31)

$$r(t + \Delta t) - r(t) = \mu \Delta t + \sigma(W(t + \Delta t) - W(t)).$$

Letting $\Delta t = 1$, the annual precipitation for the ith year satisfies the model

$$p_i = \mu + \sigma \eta_i \quad \text{for} \quad i = 1, 2, \ldots, \tag{5.32}$$

where $\eta_i \sim N(0, 1)$ for each i. A simulated precipitation using model (5.32) is given in Fig. 5.9. The simulated yearly rainfall has a similar appearance to the actual yearly rainfall.

5.3.6 Chemical reactions

In this section, chemical reactions between molecules are considered in a stochastic manner. The problem is analogous to the problem described in investigations [43, 44, 45] but it is shown in this section how a stochastic differential equation model can be developed. It is assumed that a fixed volume contains a uniform mixture of n different chemical species that interact through m different chemical reactions. The reaction rates are either proportional to the rates that the molecules collide or, if the reaction is spontaneous, the reaction

rate is just proportional to the number of molecules of the particular chemical species. Given the initial numbers of molecules of the n different chemical species, we wish to find the molecular population levels at a later time.

To illustrate the modeling procedure for chemical reactions, it is useful to consider a specific problem. Therefore, suppose that there are three chemical species S_1, S_2, and S_3 that interact through molecular collisions or spontaneously in the four ways described in Table 5.9. In Table 5.9, μ_1, μ_2, μ_3, and

Table 5.9. Assumed chemical reactions

Reaction	Probability (for small time interval Δt)
$S_1 + S_2 \rightarrow S_3$	$p_1 = \mu_1 X_1 X_2 \Delta t$
$S_3 \rightarrow S_1 + S_2$	$p_2 = \mu_2 X_3 \Delta t$ (spontaneous)
$2S_2 + S_3 \rightarrow 2S_1$	$p_3 = \mu_3 X_2^2 X_3 \Delta t/2$
$2S_1 \rightarrow 2S_2 + S_3$	$p_4 = \mu_4 X_1^2 \Delta t/2$

μ_4 are reaction rate constants and X_1, X_2, and X_3 are the number of molecules of species S_1, S_2, and S_3, respectively. The second reaction is assumed to be spontaneous and so the probability of a reaction only depends on the number of molecules, X_3. For the first reaction, the rate depends on a collision occurring between species S_1 and S_2 and is therefore proportional to the product of X_1 and X_2. The third reaction depends on a collision involving two molecules of S_2 and one molecule of S_3. As there are $X_2(X_2 - 1)/2$ ways to select two molecules from a total of X_2 molecules, the rate of this reaction depends approximately on the product of $X_2^2/2$ with X_3. The fourth reaction depends on two molecules of S_1 interacting and is approximately proportional to $X_1^2/2$. For a thorough discussion of reaction rate dynamics, see [43, 44, 45].

Table 5.10. Possible molecular population changes in small time Δt

Possible Change	Probability
$(\Delta \mathbf{X})_1 = [-1, -1, +1]^T$	$p_1 = \mu_1 X_1 X_2 \Delta t$
$(\Delta \mathbf{X})_2 = [+1, +1, -1]^T$	$p_2 = \mu_2 X_3 \Delta t$
$(\Delta \mathbf{X})_3 = [+2, -2, -1]^T$	$p_3 = \mu_3 X_2^2 X_3 \Delta t/2$
$(\Delta \mathbf{X})_4 = [-2, +2, +1]^T$	$p_4 = \mu_4 X_1^2 \Delta t/2$

To form the stochastic differential equation model, we need $E(\Delta \mathbf{X})$ and $E((\Delta \mathbf{X})(\Delta \mathbf{X})^T)$. To find these, the possible changes for the reactions given in Table 5.9 are listed in Table 5.10 along with their associated probabilities. Then, using the values in this table,

$$E(\Delta\mathbf{X}) = \sum_{i=1}^{4} p_i(\Delta\mathbf{X})_i$$
$$= \begin{bmatrix} -\mu_1 X_1 X_2 + \mu_2 X_3 + \mu_3 X_2^2 X_3 - \mu_4 X_1^2 \\ -\mu_1 X_1 X_2 + \mu_2 X_3 - \mu_3 X_2^2 X_3 + \mu_4 X_1^2 \\ \mu_1 X_1 X_2 - \mu_2 X_3 - \mu_3 X_2^2 X_3/2 + \mu_4 X_1^2/2 \end{bmatrix} = \mathbf{f}(X_1, X_2, X_3)\Delta t$$

and

$$E((\Delta\mathbf{X})(\Delta\mathbf{X})^T) = \sum_{i=1}^{4} p_i(\Delta\mathbf{X})_i(\Delta\mathbf{X})_i^T$$
$$= \begin{bmatrix} a+4b & a-4b & -a-2b \\ a-4b & a+4b & -a+2b \\ -a-2b & -a+2b & a+b \end{bmatrix} = g(X_1, X_2, X_3)\Delta t,$$

where $a = \mu_1 X_1 X_2 + \mu_2 X_3$ and $b = \mu_3 X_2^2 X_3/2 + \mu_4 X_1^2/2$. It follows that the stochastic differential equation model for this example problem has the form:

$$d\mathbf{X}(t) = \mathbf{f}(X_1, X_2, X_3)\, dt + (g(X_1, X_2, X_3))^{1/2}\, d\mathbf{W}(t) \tag{5.33}$$

with $\mathbf{X}(0) = [X_1(0), X_2(0), X_3(0)]^T$ where $\mathbf{W}(t) = [W_1(t), W_2(t), W_3(t)]^T$. Finally, it is useful to notice by Remark 5.4 and the discussion in Section 5.6 that (5.33) can be replaced by the chemical Langevin system [45]:

$$d\mathbf{X}(t) = \mathbf{f}(X_1, X_2, X_3)\, dt + C(X_1, X_2, X_3)\, d\mathbf{W}^*(t)$$

with $\mathbf{X}(0) = [X_1(0), X_2(0), X_3(0)]^T$ where $\mathbf{W}^*(t)$ is a vector $\mathbf{W}^*(t) = [W_1^*(t), W_2^*(t), W_3^*(t), W_4^*(t)]^T$ of four independent Wiener processes and 3×4 matrix C has the form

$$C = \begin{bmatrix} -(\mu_1 X_1 X_2)^{1/2} & (\mu_2 X_3)^{1/2} & 2(\mu_3 X_2^2 X_3/2)^{1/2} & -2(\mu_4 X_1^2/2)^{1/2} \\ -(\mu_1 X_1 X_2)^{1/2} & (\mu_2 X_3)^{1/2} & -2(\mu_3 X_2^2 X_3/2)^{1/2} & 2(\mu_4 X_1^2/2)^{1/2} \\ (\mu_1 X_1 X_2)^{1/2} & -(\mu_2 X_3)^{1/2} & -(\mu_3 X_2^2 X_3/2)^{1/2} & (\mu_4 X_1^2/2)^{1/2} \end{bmatrix}.$$

To test the stochastic differential equation model (5.33), calculational results using the model were compared with those obtained using a Monte Carlo procedure. In the Monte Carlo procedure, the process was checked at each small time interval to see if any reaction occurred. The calculational results are summarized in Table 5.11 and one sample path is plotted in Fig. 5.10. In these calculations, the values of the reaction rate constants were taken as $\mu_1 = 0.02, \mu_2 = 0.4, \mu_3 = 0.001$, and $\mu_4 = 0.03$. The initial numbers of molecules were assumed to be $X_1(0) = X_2(0) = X_3(0) = 100$ and the final time was taken as $t = 1.0$. A total of 5000 sample paths were performed for the stochastic differential equation (SDE) model (5.33) and for the Monte Carlo (MC) procedure. The Monte Carlo computer code and the computer code that solved stochastic differential equation system (5.33) are each listed at the end of this chapter.

Table 5.11. Calculated mean molecular levels and standard deviations at time $t = 1.0$ for the stochastic differential equation (SDE) model and for the Monte Carlo (MC) procedure

Chemical Species	$E(X_i)$ (SDE)	$\sigma(X_i)$ (SDE)	$E(X_i)$ (MC)	$\sigma(X_i)$ (MC)
S_1	79.31	7.62	79.21	7.28
S_2	37.44	6.14	37.61	5.84
S_3	131.17	6.43	131.19	5.54

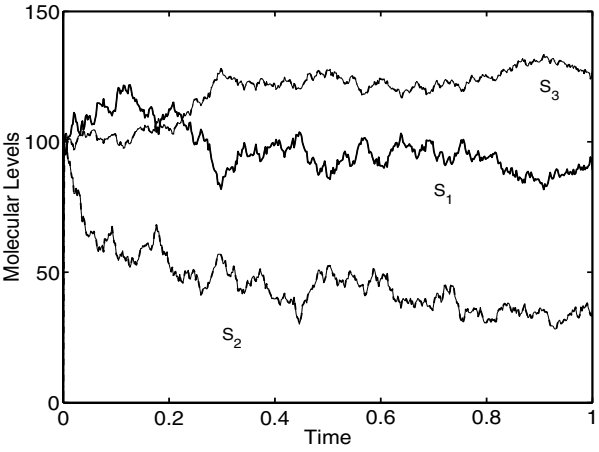

Fig. 5.10. Molecular population levels for one sample path

5.3.7 Cotton fiber breakage

In cotton thread manufacture, the cotton fiber length distribution determines many of the characteristics of the thread. Fiber length is a good indicator of spinning efficiency, yarn strength, and yarn uniformity. Fiber length distribution is affected by breakage during processing [77, 88]. In cotton processing, fiber breakage occurs in ginning and carding. Breakage of the fibers in cotton processing generally results in lower quality yarn.

The development of a stochastic differential equation (SDE) model for fiber-length distributions provides more understanding of the fiber breakage phenomenon and the origination of different fiber-length distributions [107]. By comparing calculations of the stochastic model with fiber-length data, fiber breakage parameters can be estimated and distribution characteristics can be investigated.

In the stochastic model, the fibers are grouped by length. In this manner, the cotton fiber distribution can be considered as a population distribution. The SDE model is derived by carefully considering the population process and breakage possibilities over a short time interval using the stochastic modeling

techniques described in the first section of this chapter. First, a discrete stochastic model is derived where the breakage phenomenon is carefully studied for a short time interval. Then, a system of stochastic differential equations is identified whose probability distribution approximates that of the discrete stochastic model.

Comparisons of calculational results using a stochastic model for cotton fiber breakage with Monte Carlo computational results indicate that an SDE model can accurately estimate fiber-length distributions. In addition, the SDE model generalizes classic deterministic integro-differential equation models for fiber breakage described, for example, in [88]. Furthermore, the SDE model gives information on the variability in fiber-length distributions which deterministic models are unable to provide.

In developing an SDE model, m populations, $\{N_k(t)\}_{k=1}^m$, of fibers having different lengths are considered as functions of time t. Some terminology associated with the stochastic model is required and is introduced as follows.

Let L = fiber length where it is assumed that $0 \le L \le L_{max}$.

Let $L_k = kh$ for $k = 1, 2, \ldots, m$ where $h = L_{max}/m$.

Let $N_k(t)$ = number of fibers of length L_k for $k = 1, 2, \ldots, m$.

Let $q_k\, \Delta t$ = fraction of fibers of length k broken in time Δt.

Let $S_{k,l}$ = fraction of fragments of length L_l formed from breakage of fibers of length L_k. (Note: that $\sum_{l=1}^{k-1} S_{k,l} = 1$ and $S_{k,k-l} = S_{k,l}$.)

Let $p_{k,l}(t)\Delta t = N_k(t) S_{k,l} q_k \Delta t$ = probability of a fragment of length L_l being formed from breakage of a fiber of length L_k in time t to $t + \Delta t$.

To develop the model, the changes in the fiber populations are carefully studied and tabulated for a small time interval Δt. Then, the mean change $E(\Delta \mathbf{N}(t))$ and the covariance in the change $E((\Delta \mathbf{N}(t))(\Delta \mathbf{N}(t))^T)$ for the small time interval are calculated. For example, consider the special case where m=8, that is, there are 8 groups of fibers. Consider a fiber in the 7th group breaking into two fibers, one in group 5 and one in group 2. The change produced is:

$$(\Delta\mathbf{N})^{7,5} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad \text{with probability } \; p_{7,5}(t)\Delta t = N_7(t) S_{7,5} q_7 \Delta t.$$

The value of the expected change $E(\Delta\mathbf{N}(t))$ for the small time interval is calculated by summing the products of the changes with the respective probabilities. In general, for any m, it can be shown that the lth component of $E(\Delta\mathbf{N}(t))$ has the form:

$$\begin{aligned} E(\Delta\mathbf{N}(t))_l &= 2 \sum_{k=l+1}^{m} p_{k,l}(t)\Delta t - \sum_{k=1}^{l-1} p_{l,k}(t)\Delta t \\ &= 2 \sum_{k=l+1}^{m} p_{k,l}(t)\Delta t - N_l(t) q_l \Delta t. \end{aligned}$$

In addition, the covariance matrix, has the form

$$E\left((\Delta\mathbf{N}(t))(\Delta\mathbf{N}(t))^T\right) = \sum_{k=1}^{m} \sum_{l=1}^{k-1} C^{k,l} p_{k,l}(t)\Delta t$$

where $C^{k,l}$ is the appropriate matrix that accounts for a fiber of group k breaking into a fiber of group l and group $k-l$. For example, for the special case where $m = 8$ and a fiber in the 7th group breaks into two fibers, one in group 5 and one in group 2, then the corresponding term produced in the covariance matrix is:

$$C^{7,5} = (\Delta\mathbf{N})^{7,5}(\Delta\mathbf{N})^{7,5})^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now define the expected change and the covariance matrix by:

$$\mathrm{E}(\Delta\mathbf{N}) = \boldsymbol{\beta}(t, \mathbf{N}(t))\Delta t \quad \text{and} \quad \mathrm{E}(\Delta\mathbf{N}\Delta\mathbf{N}^T) = V(t, \mathbf{N}(t))\Delta t.$$

Then, as explained in Section 5.1, the probability distribution $p(t, \mathbf{N})$ of the fiber-length populations with time t approximately satisfies the forward Kolmogorov equation:

$$\frac{\partial p(t,\mathbf{N})}{\partial t} = -\sum_{i=1}^{m} \frac{\partial}{\partial N_i}[\beta_i(t,\mathbf{N})p(t,\mathbf{N})] + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2}{\partial N_i \partial N_j}\left[\sum_{k=1}^{m} v_{i,k}(t,\mathbf{N})v_{j,k}(t,\mathbf{N})p(t,\mathbf{N})\right].$$

The SDE system corresponding to this forward Kolmogorov equation has the form:

$$d\mathbf{N}(t) = \boldsymbol{\beta}(t,\mathbf{N}(t))\,dt + (V(t,\mathbf{N}(t)))^{1/2}\,d\mathbf{W}(t) \tag{5.34}$$

where $\mathbf{N}(t) = [N_1(t), N_2(t), \ldots, N_m(t)]^T$ are the fiber populations in each length group and $\mathbf{W}(t) = [W_1(t), \ldots, W_m(t)]^T$ is an m-dimensional Wiener process. Equation (5.34) is a stochastic differential equation model for the fiber-length populations as a function of time.

Finally, by using the procedure described in Remark 5.4 or in Section 5.6, it is interesting that an equivalent stochastic differential equation model to system (5.34) can be determined and has the form:

$$d\mathbf{N}(t) = \boldsymbol{\beta}(t,\mathbf{N}(t))\,dt + \sum_{k=1}^{m}\sum_{l=1}^{k-1}(\Delta\mathbf{N})^{k,l}(p_{k,l}(t))^{1/2}\,dW_{k,l}(t),$$

where $p_{k,l}(t) = N_k(t)S_{k,l}q_k$, $W_{k,l}(t)$ for $l = 1, 2, \ldots, k-1$ and $k = 1, 2, \ldots m$ are $m(m-1)/2$ independent Wiener processes, and the ith element of vector $(\Delta\mathbf{N})^{k,l}$ is

$$((\Delta\mathbf{N})^{k,l})_i = \begin{cases} -1, & \text{if } i = k \\ 1, & \text{if } i = l \text{ or } i = k-l \\ 0, & \text{otherwise.} \end{cases}$$

In component form, this model can be written as

$$dN_i(t) = 2\sum_{k=i+1}^{m} N_k(t)S_{k,i}q_k - N_i(t)q_i - \sum_{k=1}^{i-1}(p_{i,k}(t))^{1/2}\,dW_{i,k}(t) + \sum_{k=i+1}^{m}(p_{k,i}(t))^{1/2}\,dW_{k,i}(t) + \sum_{k=i+1}^{m}(p_{k,i}(t))^{1/2}\,dW_{k,k-i}(t).$$

This stochastic differential equation system may be easier to solve computationally than system (5.34) and may offer certain theoretical insights.[1]

To test stochastic model (5.34), the model is compared computationally with Monte Carlo calculations. In the Monte Carlo calculations, for each small time step, each fiber is checked for breakage. If breakage occurs, the fiber is randomly divided. Considered in these calculations is the situation where breakage occurs randomly and the probability for breakage is proportional to the length of the fiber. Under this breakage assumption,

[1] As the number of groups m increases, it appears that a stochastic integro-differential equation for the fiber distribution is obtained which has the form

$$q_k S_{k,j} \Delta t = \mu \left(\frac{h}{L_{max}} \right) \Delta t$$

where μ is a constant which determines the rate of fiber breakage fraction of fibers of length k broken in time Δt and

$$S_{k,j} = \frac{h}{L_{k-1}} = \frac{1}{k-1} \quad \text{for} \quad j = 1, 2, \ldots, k-1, \quad \text{and} \quad k = 2, 3, \ldots, m$$

where $S_{k,j}$ is the fraction of fragments of length L_j formed from breakage of fiber of length L_k. The parameter μ was set equal to unity in the calculations and it was assumed that there were initially 100 fibers each 1 inch in length. The calculational results are compared in Table 5.12. The results indicate very good agreement between the two different procedures.

Table 5.12. Monte Carlo (MC) and stochastic differential equation (SDE) calculational results on fiber lengths at time $t = 1.0$

Avg. Number of Fibers	Standard Dev. in No. of Fibers	Average Fiber Length	Standard Dev. in Fiber Length
200.5 (MC)	10.57 (MC)	0.5001 (MC)	0.0263 (MC)
197.8 (SDE)	11.47 (SDE)	0.5068 (SDE)	0.0265 (SDE)

Additional computations produce fiber-length distributions having a bimodal structure and bimodal distributions are commonly seen in fiber-length data [77]. Results of an example calculation are illustrated in Fig. 5.11. For this calculation, it was assumed that the fibers initially were distributed as $N_k(0) = 2(k-20)$ for $k = 20, 21, \ldots, 35$ and $N_k(0) = 2(50-k)$ for $k = 36, 37, \ldots, 50$ where $N_k(0)$ was the initial number of fibers of length $L_k = 0.02k$.

$$\begin{aligned} \frac{\partial N(t,L)}{\partial t} &= -q(L)N(t,L) + 2\int_L^{L_{max}} S(u,L)q(u)N(t,u)\,du \\ &\quad - \int_0^L p^{1/2}(L,u)\frac{\partial^3 W(L,u,t)}{\partial L \partial u \partial t}\,du + \int_L^{L_{max}} p^{1/2}(u,L)\frac{\partial^3 W(u,L,t)}{\partial u \partial L \partial t}\,du \\ &\quad + \int_L^{L_{max}} p^{1/2}(u,L)\frac{\partial^3 W^*(u,L,t)}{\partial u \partial L \partial t}\,du \end{aligned}$$

where $N_i(t) = \int_{L_{i-1}}^{L_i} N(t,L)\,dL$, $S_{k,i} = \int_{L_{i-1}}^{L_i} S(L_k,L)\,dL$ and where W and W^* are correlated Brownian sheets [3, 113] with the property that $\int_t^{t+\Delta t}\int_{L_{i-1}}^{L_i}\int_{L_{k-1}}^{L_k} \dot{W}^*(u,L,t)\,du\,dL\,dt = \eta_{k,k-i}h(\Delta t)^{1/2}$ and that $\int_t^{t+\Delta t}\int_{L_{i-1}}^{L_i}\int_{L_{k-1}}^{L_k} \dot{W}(u,L,t)\,du\,dL\,dt = \eta_{k,i}h(\Delta t)^{1/2}$ where $\eta_{k,i} \sim N(0,1)$ for $k = i+1, i+2, \ldots, m$ and $i = 1, 2, \ldots, m-1$. Furthermore, the deterministic part of this integro-differential equation model has been previously derived, for example, in reference [88].

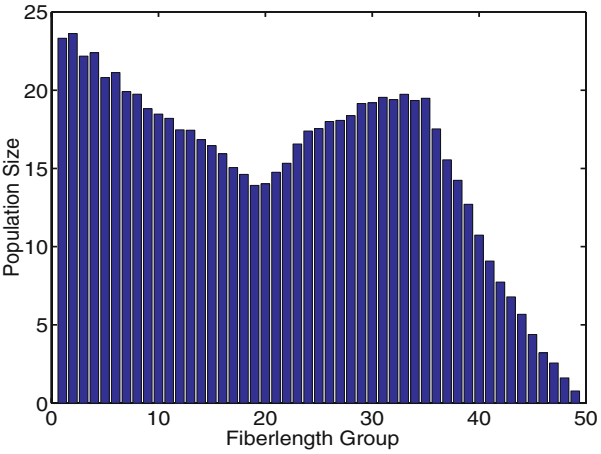

Fig. 5.11. Average fiber length distribution after random breakage for time $t = 1.0$ (Using the SDE model)

5.4 Some Stochastic Finance Models

Models for stock prices, option pricing, and interest rates are three examples studied in this section.

5.4.1 A stock-price model

A stock-price model is developed using the procedure described in the first section of this chapter. It is interesting to apply this procedure to the dynamics of stock prices to see, for example, if a standard stock-price model is obtained. Also, derivation of a stock-price model may lead to a better understanding of the underlying dynamics of stock prices as the parameters in the model are derived from basic assumptions. It is important to note that certain assumptions are made here in the derivation of the model which involve, for example, forms assumed for the probabilities of the possible price changes over a small time step. There is much flexibility in choosing these probabilities and determination of accurate forms undoubtedly requires much testing with stock price data. In addition, large jumps in stock prices caused by sudden major changes in the financial environment are not considered in this model development.

Throughout this section, to simplify notation, two stocks are considered along with a fixed-interest money market account. However, the results can be readily generalized to a system of n stocks. The flow of money (or value) assumed for these two investments is illustrated in Fig. 5.12.

A system of stochastic differential equations is now derived [75] for the two stocks represented in Fig. 5.12. In this model development, the procedure described in the first section of this chapter is followed. First, therefore,

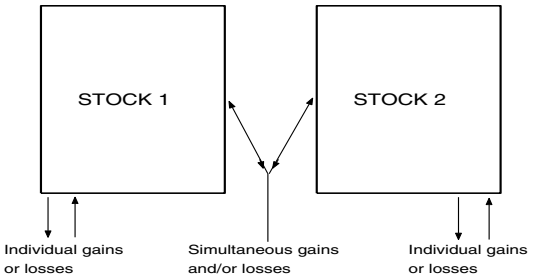

Fig. 5.12. Two stocks illustrated as a stochastic process

changes in the stock prices are considered for a small time interval Δt. It is assumed that this time interval is sufficiently short such that the probability of more than one change in the stock prices is small. Let $\Delta \mathbf{S} = [\Delta S_1, \Delta S_2]^T$ be the change in the two stock prices over a short time step Δt. It is necessary to find the mean and the covariance matrix for the change $\Delta \mathbf{S}$. Neglecting multiple changes in time Δt which have probabilities of order $(\Delta t)^2$, there are nine possibilities for $\Delta \mathbf{S}$ in time Δt. These possibilities are listed in Table 5.13.

Table 5.13. Possible changes in the stock prices with the corresponding probabilities

Change $[\Delta S_1, \Delta S_2]^T$	Probability
$[\Delta S_1, \Delta S_2]_1^T = [1, 0]^T$	$p_1 = b_1 S_1 \Delta t$
$[\Delta S_1, \Delta S_2]_2^T = [-1, 0]^T$	$p_2 = d_1 S_1 \Delta t$
$[\Delta S_1, \Delta S_2]_3^T = [0, 1]^T$	$p_3 = b_2 S_2 \Delta t$
$[\Delta S_1, \Delta S_2]_4^T = [0, -1]^T$	$p_4 = d_2 S_2 \Delta t$
$[\Delta S_1, \Delta S_2]_5^T = [1, 1]^T$	$p_5 = m_{22} S_1 S_2 \Delta t$
$[\Delta S_1, \Delta S_2]_6^T = [1, -1]^T$	$p_6 = m_{21} S_1 S_2 \Delta t$
$[\Delta S_1, \Delta S_2]_7^T = [-1, 1]^T$	$p_7 = m_{12} S_1 S_2 \Delta t$
$[\Delta S_1, \Delta S_2]_8^T = [-1, -1]^T$	$p_8 = m_{11} S_1 S_2 \Delta t$
$[\Delta S_1, \Delta S_2]_9^T = [0, 0]^T$	$p_9 = 1 - \sum_{i=1}^{8} p_i$

Notice, for example, that $\Delta \mathbf{S} = [1, 0]^T$ represents a gain of one unit in the price of stock 1. Also, $\Delta \mathbf{S} = [0, -1]^T$ represents a loss of one unit in the price of stock 2 and $\Delta \mathbf{S} = [1, -1]^T$ represents a simultaneous gain of one unit in the price of stock 1 and a loss of one unit in the price of stock 2. It is assumed that the probability of a change in one stock price is proportional to the stock price. For a simultaneous change in both stock prices, it is assumed that the probability of the change is proportional to the product of the two stock prices.

(To see why this is reasonable, suppose that one stock price is zero. Then the probability of a simultaneous change must be zero.) The probabilities for these changes are also listed in Table 5.13. It is assumed that Δt is sufficiently small so that $p_9 \geq 0$. The parameters $b_1, d_1, b_2, d_2, m_{11}, m_{12}, m_{21}$, and m_{22} define the rates at which stocks experience individual gains or losses or experience simultaneous gains and/or losses. Each parameter may, of course, depend on time t. Considering the above equations for p_i and $\Delta \mathbf{S}_i$ their meaning becomes clear. For example, $b_i S_i \Delta t$ is the probability stock i for $i = 1$ or 2 has a gain of one unit in time interval Δt, $m_{22} S_1 S_2 \Delta t$ is the probability both stocks experience a gain in time interval Δt, and $m_{12} S_1 S_2 \Delta t$ is the probability that stock 1 has a loss and stock 2 has a gain in time interval Δt. Notice that $\sum_{j=1}^{9} p_j = 1$.

Using the above expressions for p_i and $\Delta \mathbf{S}_i$, the expectation vector and the covariance matrix for the change $\Delta \mathbf{S}$ can now be derived. Neglecting terms of order $(\Delta t)^2$, it follows that

$$\begin{aligned} E(\Delta \mathbf{S}) &= \sum_{j=1}^{9} p_j \Delta \mathbf{S}_j = \begin{bmatrix} (b_1 - d_1)S_1 + (m_{22} + m_{21} - m_{12} - m_{11})S_1 S_2 \\ (b_2 - d_2)S_2 + (m_{22} - m_{21} + m_{12} - m_{11})S_1 S_2 \end{bmatrix} \Delta t \\ &= \boldsymbol{\mu}(t, S_1, S_2)\Delta t \end{aligned}$$

and

$$\begin{aligned} E(\Delta \mathbf{S}(\Delta \mathbf{S})^T) &= \sum_{j=1}^{9} p_j \Delta \mathbf{S}_j (\Delta \mathbf{S}_j)^T = \begin{bmatrix} c_1 + c_2 + c_3 & c_2 - c_3 \\ c_2 - c_3 & c_2 + c_3 + c_4 \end{bmatrix} \Delta t \\ &= V(t, S_1, S_2)\Delta t, \end{aligned}$$

where $c_1 = (b_1 + d_1)S_1$, $c_2 = (m_{22} + m_{11})S_1 S_2$, $c_3 = (m_{21} + m_{12})S_1 S_2$, and $c_4 = (b_2 + d_2)S_2$. As the product $E(\Delta \mathbf{S})(E(\Delta \mathbf{S}))^T$ is of order $(\Delta t)^2$, the covariance matrix V is set equal to $E(\Delta \mathbf{S}(\Delta \mathbf{S})^T)/\Delta t$. It is straightforward to show that V is positive definite and hence has a positive definite square root. Denote $B = (V)^{1/2}$. As $\Delta t \to 0$, the probability distribution of the stock prices approximates the probability distribution of solutions to the Itô stochastic differential equation system:

$$d\mathbf{S}(t) = \boldsymbol{\mu}(t, S_1, S_2)\, dt + B(t, S_1, S_2)\, d\mathbf{W}(t) \tag{5.35}$$

with $\mathbf{S}(0) = \mathbf{S}_0$ and where $\mathbf{W}(t)$ is the two-dimensional Wiener process, i.e. $\mathbf{W}(t) = [W_1(t), W_2(t)]^T$. Equation (5.35) is a system of stochastic differential equations that describes the dynamics of the stock prices.

In equation (5.35), $\boldsymbol{\mu}$ and B have the forms (for two stocks):

$$\boldsymbol{\mu}(t, S_1, S_2) = \begin{bmatrix} (b_1 - d_1)S_1 + (m_{22} + m_{21} - m_{12} - m_{11})S_1 S_2 \\ (b_2 - d_2)S_2 + (m_{22} - m_{21} + m_{12} - m_{11})S_1 S_2 \end{bmatrix} \tag{5.36}$$

and

$$B(t, S_1, S_2) = \frac{1}{d} \begin{bmatrix} c_1 + c_2 + c_3 + w & c_2 - c_3 \\ c_2 - c_3 & c_2 + c_3 + c_4 + w \end{bmatrix}, \tag{5.37}$$

where w and d are given by $w = \sqrt{(c_1 + c_2 + c_3)(c_2 + c_3 + c_4) - (c_2 - c_3)^2}$ and $d = \sqrt{c_1 + 2c_2 + 2c_3 + c_4 + 2w}$.

It is interesting that model (5.35) is similar to an affine model [32, 108] for the stock price dynamics. In an affine model, the elements of $\boldsymbol{\mu}$ and $B^T B$ are linear functions of the stock prices. Indeed, for a single stock, model (5.35) simplifies to

$$dS_1 = (b_1 - d_1)S_1\,dt + \sqrt{(b_1 + d_1)S_1}\,dW_1(t) \tag{5.38}$$

which is an affine diffusion model for a single stock. Notice that the form given by equation (5.38) differs from the standard geometric Brownian motion [32, 56, 74, 95] often assumed for stock price where the drift and diffusion coefficients are proportional to stock price S_1. Specifically, stock price S_1 follows geometric Brownian motion if it satisfies a stochastic differential equation of the form:

$$dS_1 = \nu S_1\,dt + \sigma S_1\,dW_1(t),$$

where ν is the drift and σ is the volatility.

Finally, it is supposed that the price per share of the money market is $M(t)$, where it is assumed that

$$\frac{dM}{dt} = r(t)M \tag{5.39}$$

and $r(t)$ is the interest rate.

5.4.2 Option pricing

With the dynamics of the stocks and money market given by equations (5.35) and (5.39), respectively, standard procedures [34, 37, 56, 74, 75, 89, 95] can be applied, for example, to estimate option prices. Considered here will be European call options but other types of options can be treated analogously. For generality, consider a multi-asset option for a mutual fund where the fund consists, per share, of α_1 shares of stock 1 and α_2 shares of stock 2. It is assumed here that $\alpha_1, \alpha_2 \geq 0$. Let $\boldsymbol{\alpha} = [\alpha_1, \alpha_2]^T$. Define $X(t) = \boldsymbol{\alpha}^T \mathbf{S}(t)$ as the price per share of the mutual fund. Let $U(t, S_1(t), S_2(t))$ be the expected value of a European call option on the asset with maturity time T. That is, $U : [0, T] \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ denotes the price of the option at time t when the fund price is $X(t)$. The initial condition is given at time T and is $U(T, S_1(T), S_2(T)) = \max\big(X(T) - K, 0\big)$, where $K > 0$ is the strike price of the option. As shown below, by applying Itô's formula and assuming that the stock prices satisfy an equation of the form (5.35), the following partial differential equation, referred to as the two-dimensional Black-Scholes equation [75, 83], can be derived for the value of the option:

$$\frac{\partial U(t,S_1,S_2)}{\partial t} + r(t)\left(S_1\frac{\partial U(t,S_1,S_2)}{\partial S_1} + S_2\frac{\partial U(t,S_1,S_2)}{\partial S_2}\right)$$
$$+\frac{1}{2}(b_{11}^2+b_{12}^2)\frac{\partial^2 U}{\partial S_1^2} + (b_{11}b_{12}+b_{22}b_{12})\frac{\partial^2 U}{\partial S_1 \partial S_2} + \frac{1}{2}(b_{21}^2+b_{22}^2)\frac{\partial^2 U}{\partial S_2^2}$$
$$= r(t)U(t,S_1,S_2) \tag{5.40}$$

for $0 \le t \le T$ and $0 \le S_1, S_2 < \infty$ with initial condition given at $t = T$ as $U(T, S_1(T), S_2(T)) = \max(X(T) - K, 0)$ and boundary condition $U(t, 0, 0) = 0$. The parameters b_{ij} are the elements of matrix B and are functions of S_1 and S_2. Notice that if $\alpha_1 = 1$ and $\alpha_2 = 0$, then this equation reduces to the Black-Scholes partial differential equation [16, 56] for pricing options for a single stock. This partial differential equation can be readily solved numerically using, for example, finite-difference methods [16, 19] to find $U(0, S_1(0), S_2(0))$.

It is interesting to derive partial differential equation (5.40). Denote $U(t, X(S_1, S_2))$ as the value of the option where $X(S_1, S_2)$ is a function of S_1 and S_2. Suppose that $\mathbf{S}(t)$ and $M(t)$ satisfy

$$\begin{cases} d\mathbf{S}(t) = \boldsymbol{\mu}(t)\,dt + B(t)\,d\mathbf{W}(t) \\ dM(t) = r(t)M(t)\,dt \end{cases}$$

where $(\boldsymbol{\mu})_i = \mu_i$ and $(B)_{ij} = b_{ij}$. By Itô's formula,

$$\begin{aligned} \frac{dU}{dt} &= \frac{\partial U}{\partial t} + \mu_1\frac{\partial U}{\partial S_1} + \mu_2\frac{\partial U}{\partial S_2} \\ &\quad + \frac{1}{2}(b_{11}^2+b_{12}^2)\frac{\partial^2 U}{\partial S_1^2} + (b_{11}b_{12}+b_{22}b_{12})\frac{\partial^2 U}{\partial S_1 \partial S_2} + \frac{1}{2}(b_{21}^2+b_{22}^2)\frac{\partial^2 U}{\partial S_2^2} \\ &\quad + \frac{\partial U}{\partial S_1}\left(b_{11}\frac{dW_1}{dt} + b_{12}\frac{dW_2}{dt}\right) + \frac{\partial U}{\partial S_2}\left(b_{21}\frac{dW_1}{dt} + b_{22}\frac{dW_2}{dt}\right). \end{aligned}$$

At time t, an amount $\mathbf{x}(t)$ of the stocks and an amount $y(t)$ of the money market are selected giving a total value of

$$G(t) = x_1(t)S_1(t) + x_2(t)S_2(t) + y(t)M(t).$$

The values of $\mathbf{x}(t)$ and $y(t)$ are chosen so that

$$\begin{cases} G(t) = U(t,X) = U(t, X(S_1,S_2)) \\ \dfrac{dG(t)}{dt} = \dfrac{dU(t,X)}{dt} = \dfrac{dU(t,X(S_1,S_2))}{dt}. \end{cases}$$

Note that

$$\frac{dG(t)}{dt} = x_1(t)\frac{dS_1(t)}{dt} + x_2(t)\frac{dS_2(t)}{dt} + y(t)\frac{dM(t)}{dt} + R(t)$$

where

$$R(t) = \frac{dx_1(t)}{dt} S_1(t) + \frac{dx_2(t)}{dt} S_2(t) + \frac{dy(t)}{dt} M(t).$$

Using

$$d\mathbf{S} = \boldsymbol{\mu}\, dt + B\, d\mathbf{W} \quad \text{and} \quad dM = rM\, dt,$$

then

$$\begin{aligned}\frac{dG}{dt} &= x_1 \left(\mu_1 + b_{11} \frac{dW_1}{dt} + b_{12} \frac{dW_2}{dt} \right) + x_2 \left(\mu_2 + b_{21} \frac{dW_1}{dt} + b_{22} \frac{dW_2}{dt} \right) \\ &\quad + yrM + R.\end{aligned}$$

To satisfy $\dfrac{dG}{dt} = \dfrac{dU}{dt}$, the coefficients of $\dfrac{dW_1}{dt}$ and $\dfrac{dW_2}{dt}$ must match in the preceding expressions. Therefore, we set

$$x_1 = \frac{\partial U}{\partial S_1} \quad \text{and} \quad x_2 = \frac{\partial U}{\partial S_2}.$$

Then, as $G = x_1 S_1 + x_2 S_2 + yM = U$, it follows that

$$y(t) = \frac{1}{M} \left(U - S_1 \frac{\partial U}{\partial S_1} - S_2 \frac{\partial U}{\partial S_2} \right).$$

Substituting this expression into the equation for $\dfrac{dG}{dt}$, then

$$\begin{aligned}\frac{dG}{dt} &= r \left(U - S_1 \frac{\partial U}{\partial S_1} - S_2 \frac{\partial U}{\partial S_2} \right) + \frac{\partial U}{\partial S_1} \left(\mu_1 + b_{11} \frac{dW_1}{dt} + b_{12} \frac{dW_2}{dt} \right) \\ &\quad + \frac{\partial U}{\partial S_2} \left(\mu_2 + b_{21} \frac{dW_1}{dt} + b_{22} \frac{dW_2}{dt} \right) + R.\end{aligned}$$

Recalling that $\mathbf{x}(t)$ and $y(t)$ are chosen so that $\dfrac{dG}{dt} = \dfrac{dU}{dt}$, we set

$$\begin{aligned}r \left(U - S_1 \frac{\partial U}{\partial S_1} - S_2 \frac{\partial U}{\partial S_2} \right) &= \frac{\partial U}{\partial t} + \frac{1}{2}(b_{11}^2 + b_{12}^2) \frac{\partial^2 U}{\partial S_1^2} \\ &\quad + (b_{11} b_{12} + b_{22} b_{12}) \frac{\partial^2 U}{\partial S_1 \partial S_2} + \frac{1}{2}(b_{21}^2 + b_{22}^2) \frac{\partial^2 U}{\partial S_2^2}.\end{aligned}$$

This also yields $R = 0$ since

$$\begin{aligned}R(t) &= S_1 \frac{dx_1}{dt} + S_2 \frac{dx_2}{dt} + M \frac{dy}{dt} \\ &= S_1 \frac{d}{dt} \left(\frac{\partial U}{\partial S_1} \right) + S_2 \frac{d}{dt} \left(\frac{\partial U}{\partial S_2} \right) - \frac{1}{M} \frac{dM}{dt} \left(U - S_1 \frac{\partial U}{\partial S_1} - S_2 \frac{\partial U}{\partial S_2} \right) \\ &\quad + \frac{dU}{dt} - \frac{dS_1}{dt} \frac{\partial U}{\partial S_1} - \frac{dS_2}{dt} \frac{\partial U}{\partial S_2} - S_1 \frac{d}{dt} \left(\frac{\partial U}{\partial S_1} \right) - S_2 \frac{d}{dt} \left(\frac{\partial U}{\partial S_2} \right) \\ &= -r \left(U - S_1 \frac{\partial U}{\partial S_1} - S_2 \frac{\partial U}{\partial S_2} \right) + \frac{dU}{dt} - \frac{dS_1}{dt} \frac{\partial U}{\partial S_1} - \frac{dS_2}{dt} \frac{\partial U}{\partial S_2} \;=\; 0.\end{aligned}$$

As a result, obtained is the partial differential equation

$$\frac{\partial U}{\partial t} + r\left(S_1\frac{\partial U}{\partial S_1} + S_2\frac{\partial U}{\partial S_2}\right) + \frac{1}{2}(b_{11}^2 + b_{12}^2)\frac{\partial^2 U}{\partial S_1^2} + (b_{11}b_{12} + b_{22}b_{12})\frac{\partial^2 U}{\partial S_1 \partial S_2} + \frac{1}{2}(b_{21}^2 + b_{22}^2)\frac{\partial^2 U}{\partial S_2^2} = rU$$

for $0 \le t \le T$, $0 \le S_1, S_2 < \infty$, where $U(t, X(S_1, S_2))$ is the value of the option at time t. Note that the values of $b_{ij}, 1 \le i, j \le 2$ may depend on S_1 and S_2.

Recall that $X(S_1, S_2) = \alpha_1 S_1 + \alpha_2 S_2$ and consider briefly the special case where $\alpha_1, \alpha_2, b_{11}, b_{12}, b_{21}$, and b_{22} are constants. Then

$$\frac{\partial U}{\partial S_1} = \alpha_1\frac{\partial U}{\partial X}, \quad \frac{\partial U}{\partial S_2} = \alpha_2\frac{\partial U}{\partial X},$$

$$\frac{\partial^2 U}{\partial S_1 \partial S_2} = \alpha_1\alpha_2\frac{\partial^2 U}{\partial X^2}, \quad \frac{\partial^2 U}{\partial S_1^2} = \alpha_1^2\frac{\partial^2 U}{\partial X^2}, \quad \frac{\partial^2 U}{\partial S_2^2} = \alpha_2^2\frac{\partial^2 U}{\partial X^2},$$

and the above partial differential equation reduces to

$$rU = \frac{\partial U}{\partial t} + rX\frac{\partial U}{\partial X} + \left(\frac{1}{2}\alpha_1^2(b_{11}^2 + b_{12}^2) + \alpha_1\alpha_2(b_{11}b_{12} + b_{22}b_{12}) + \frac{1}{2}\alpha_2^2(b_{21}^2 + b_{22}^2)\right)\frac{\partial^2 U}{\partial X^2}$$

for $0 \le t \le T$, $0 \le X < \infty$. In addition, if the option is a European call option, for example, then the initial condition is $U(T, X) = \max(X - K, 0)$ with boundary condition $U(t, 0) = 0$. Notice that this equation is similar to a no-arbitrage partial differential equation [16] for a single stock.

5.4.3 Interest rates

As another example of the application of stochastic differential equations in finance, short-term interest rates are considered in this section. Short rates are popular in the market and are useful in pricing derivatives [24]. Two well-known stochastic models for the interest rate are the Vasicek model and the Cox-Ingersoll-Ross model [24]. These two models are derived in this section from basic assumptions on changes in the interest rate over small time intervals.

Let $r(t)$ be the instantaneous interest rate. In each model, it is assumed that there are three possible changes in the interest rate in a short time interval Δt. Specifically, $(\Delta r)_1 = -1, (\Delta r)_2 = 1$, and $(\Delta r)_1 = 0$, i.e., there may be a negative change of one unit, a positive change of one unit, or no change in time interval Δt. For the first model, it is supposed that the probabilities for these changes are those listed in Table 5.14.

Table 5.14. Possible changes in the interest rates with the corresponding probabilities for the first model

Change $(\Delta r)_i$	Probability p_i
$(\Delta r)_1 = -1$	$p_1 = \left[\sigma^2/2 - \alpha(r_e - r)/2\right] \Delta t$
$(\Delta r)_2 = 1$	$p_2 = \left[\sigma^2/2 + \alpha(r_e - r)/2\right] \Delta t$
$(\Delta r)_3 = 0$	$p_3 = 1 - p_1 - p_2$

The $\sigma^2/2$ term in the probabilities accounts for a random change in the interest rate of either plus or minus one unit. The $\mp\alpha(r_e - r)/2$ term in the probabilities models the tendency for the interest rate to move toward an "equilibrium" value r_e. In mathematical finance this is referred to as mean reversion. In particular, if $r(t) > r_e$, then the probability of a negative change in time interval Δt is greater than the probability of a positive change. Furthermore, if $r(t) < r_e$, then the probability of a positive change in time interval Δt is greater than the probability of a negative change.

Using Table 5.14, $E(\Delta r)$ and $E((\Delta r)^2)$ can now be computed giving

$$E(\Delta r) = \sum_{i=1}^{3} p_i (\Delta r)_i = \alpha(r_e - r)\Delta t$$

and

$$E((\Delta r)^2) = \sum_{i=1}^{3} p_i ((\Delta r)_i^2) = \sigma^2 \Delta t.$$

Applying the procedure explained in the first section of this chapter yields the stochastic differential equation model:

$$dr(t) = \alpha(r_e - r(t))\, dt + \sigma\, dW(t) \tag{5.41}$$

with $r(0) = r_0 > 0$. The interest rate model derived here, Eq. (5.41), is the Vasicek interest rate model. Using Itô's formula, the mean and variance for this model can be readily determined as:

$$E(r(t)) = r_e + (r_0 - r_e)e^{-\alpha t} \quad \text{and} \quad \text{Var}(r(t)) = \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha t}\right).$$

Indeed, the exact solution to (5.41) can be found as:

$$r(t) = r_e + (r_0 - r_e)e^{-\alpha t} + e^{-\alpha t} \int_0^t \sigma e^{\alpha s}\, dW(s).$$

In the Vasicek interest rate model, the interest rate $r(t)$ is normally distributed and the mean and variance approach the values r_e and $\dfrac{\sigma^2}{2\alpha}$, respectively, as $t \to \infty$. A disadvantage of the Vasicek model is that the interest rate, $r(t)$, can have negative values.

Now, a second stochastic differential equation model is developed for the instantaneous interest rate. In this model, the possible changes in the interest rate for small time interval Δt are identical to those in the first model. This model differs from the first model in the selection of the probabilities for the changes. The probabilities selected for this model are listed in Table 5.15.

Table 5.15. Possible changes in the interest rates with the corresponding probabilities for the second model

Change $(\Delta r)_i$	Probability p_i
$(\Delta r)_1 = -1$	$p_1 = \left[\sigma^2 r/2 - \alpha(r_e - r)/2\right] \Delta t$
$(\Delta r)_2 = 1$	$p_2 = \left[\sigma^2 r/2 + \alpha(r_e - r)/2\right] \Delta t$
$(\Delta r)_3 = 0$	$p_3 = 1 - p_1 - p_2$

Notice for this model, the mean reversion term, $\mp\alpha(r_e - r)/2$, in probabilities p_1 and p_2 are the same as for the first model. However, for this model, the term accounting for random changes in the interest rate is $\sigma^2 r/2$ rather than $\sigma^2/2$. Recall that this term in the probabilities accounts for a random change in the interest rate of either plus or minus one unit. Thus, in the second model, as the interest rate, $r(t)$, decreases the random behavior also decreases.

Using Table 5.15, $E(\Delta r)$ and $E((\Delta r)^2)$ can now be computed giving

$$E(\Delta r) = \sum_{i=1}^{3} p_i (\Delta r)_i = \alpha(r_e - r)\Delta t$$

and

$$E((\Delta r)^2) = \sum_{i=1}^{3} p_i ((\Delta r)_i^2) = \sigma^2 r \Delta t.$$

Applying the procedure explained in the first section of this chapter yields the stochastic differential equation model:

$$dr(t) = \alpha(r_e - r(t))\, dt + \sigma\sqrt{r(t)}\, dW(t) \tag{5.42}$$

with $r(0) = r_0 > 0$. The interest rate model derived here, Eq. (5.42), is the Cox-Ingersoll-Ross (CIR) interest rate model. As expected from the probabilities p_1 and p_2 in Table 5.15, the diffusion part of Eq. (5.42) decreases as r decreases. Applying Itô's formula, it is straightforward to find that the mean and variance in the interest rate for this model are given by

$$E(r(t)) = r_e + (r_0 - r_e)e^{-\alpha t}$$

and

$$\mathrm{Var}(r(t)) = \frac{\sigma^2 r_e}{2\alpha} + \frac{\sigma^2(r_0 - r_e)}{\alpha} e^{-\alpha t} + \left(\frac{\sigma^2 r_e}{2\alpha} - \frac{\sigma^2 r_0}{\alpha}\right) e^{-2\alpha t}$$

and it is clear, for the CIR model, that as $t \to \infty$, then $E(r(t)) \to r_e$ and $\text{Var}(r(t)) \to \dfrac{\sigma^2 r_e}{2\alpha}$.

An interesting feature of the CIR model is that if $\dfrac{2\alpha r_e}{\sigma^2} \geq 1$, then the interest rate, $r(t)$, is nonnegative with probability one for any $t \geq 0$. Indeed, it can be shown that the probability density of solutions to the stochastic differential equation (5.42) has the form [64]:

$$p(t,r) = c\left(\frac{v}{u}\right)^{q/2} e^{-u-v} I_q(2\sqrt{uv}),$$

where

$$c = \frac{2\alpha}{\sigma^2(1-e^{-\alpha t})}, \quad u = cr_0 e^{-\alpha t}, \quad v = cr, \quad q = \frac{2\alpha r_e}{\sigma^2} - 1,$$

and $I_q(z)$ is the modified Bessel function of order q. That is,

$$I_q(z) = \left(\frac{z}{2}\right)^q \sum_{k=0}^{\infty} \frac{\left(\frac{z^2}{4}\right)^k}{k!\,\Gamma(q+k+1)},$$

where Γ is the gamma function. Using the identity $\int_0^\infty e^{-cr} r^\alpha \, dr = \Gamma(\alpha+1)/c^{\alpha+1}$, it can be shown that $\int_0^\infty p(t,r)\, dr = 1$ for any $t > 0$ assuming that $q = \frac{2\alpha r_e}{\sigma^2} - 1 \geq 0$. That is, $r(t) \geq 0$ with probability one for any time $t \geq 0$ for the CIR interest rate model provided that $\dfrac{2\alpha r_e}{\sigma^2} \geq 1$.

There are several popular interest rate models in addition to the Vasicek and CIR models. A good discussion of interest rate models can be found, for example, in [24].

5.5 A Goodness-of-Fit Test for an SDE Model

In this section, it is assumed that a stochastic differential equation model has been developed for a certain stochastic process. Also, it is assumed that a collection of data is available for the stochastic process. A simple goodness-of-fit test is described in this section to test if there is a lack-of-fit between the stochastic differential equation model and the data. Suppose that the stochastic process is observed at times $t_0, t_1, \ldots, t_{N-1}$ where $t_i = i\Delta t$ for a constant $\Delta t > 0$. Let $x_0, x_1, \ldots, x_{N-1}$ denote the N observations of the process. In addition, suppose that the stochastic differential equation model for the process is

$$dX(t) = f(t,X)\, dt + g(t,X)\, dW(t). \tag{5.43}$$

A goodness-of-fit procedure developed in [22] is now described.

In this procedure, M simulations of (5.43) are calculated from time t_{i-1} until time t_i starting at x_{i-1}. For example, if Euler's method is used with K steps then

$$\begin{aligned} X_{j+1,i}^{(m)} &= X_{j,i}^{(m)} + f\left(t_{i-1} + \frac{j\Delta t}{K}, X_{j,i}^{(m)}\right)\frac{\Delta t}{K} \\ &\quad + g\left(t_{i-1} + \frac{j\Delta t}{K}, X_{j,i}^{(m)}\right)\sqrt{\frac{\Delta t}{K}}\,\eta_{j,i}^{(m)} \end{aligned} \tag{5.44}$$

for $j = 0, 1, \ldots, K-1$ and $m = 1, 2, \ldots, M$ with $X_{0,i}^{(m)} = x_{i-1}$ and where $\eta_{j,i}^{(m)} \sim N(0,1)$ for each i, j and m.

Let $X_i^{(m)} = X_{K,i}^{(m)}$ be the mth simulated value at t_i for $m = 1, 2, \ldots, M$ and for $i = 1, 2, \ldots, N-1$. Now define

$$s_i^{(m)} = \begin{cases} 1, & \text{if} \quad x_i \geq X_i^{(m)} \\ 0, & \text{if} \quad x_i < X_i^{(m)} \end{cases}$$

and let

$$r_i = 1 + \sum_{m=1}^{M} s_i^{(m)} \quad \text{for} \quad i = 1, 2, \ldots, N-1.$$

Then, r_i is the rank of value x_i as compared with the M simulated values $X_i^{(m)}, 1 \leq m \leq M$, for $i = 1, 2, \ldots, N-1$. Notice that $1 \leq r_i \leq M+1$ for $i = 1, 2, \ldots, N-1$.

The null hypothesis is that model (5.43) describes the stochastic process. Under the null hypothesis, the ranks r_i have equally likely values between 1 and $M+1$. A χ^2 goodness-of-fit test is used to test this hypothesis. To perform this test, the observed and expected frequencies are needed. Let

$$I_{i,q} = \begin{cases} 1, & \text{if} \quad r_i = q \\ 0, & \text{if} \quad r_i \neq q \end{cases}$$

for $i = 1, 2, \ldots, N-1$ and let

$$\Omega(q) = \sum_{i=1}^{N-1} I_{i,q}$$

for $q = 1, 2, \ldots, M+1$. Notice that $\Omega(q)$ is the observed frequency that the rank equals the value q so, for example, $\sum_{q=1}^{M+1} \Omega(q) = N-1$. The expected frequency under the null hypothesis is $\dfrac{N-1}{M+1}$. The test statistic is

$$Q_M = \sum_{q=1}^{M+1} \frac{\left(\Omega(q) - \dfrac{N-1}{M+1}\right)^2}{\left(\dfrac{N-1}{M+1}\right)},$$

which under the null hypothesis, is approximately distributed as a chi square random variable with M degrees of freedom. A large value of Q_M indicates a lack-of-fit between the stochastic model and the data. Specifically, if $P(\chi^2(M) \geq Q_M)$ is smaller than a preset level of significance, then the null hypothesis is rejected indicating a lack-of-fit of the stochastic differential equation model (5.43) with the data. The chi square approximation fails, however, if the expected frequencies under the null hypothesis are small. Consequently, the rule-of-thumb often applied is that the expected frequencies should be no less than 5. Applying this rule gives $(N-1)/(M+1) \geq 5$ which implies that $M \leq (N-6)/5$. For example, if $N = 100$, then the number of simulations M should be no more than 18.

Example 5.5. Testing two SDE models

Consider the Aransas-Wood Buffalo population of whooping cranes described in Section 4.9. For this example, there are $N = 47$ values of the population size over the years 1939-1985 which are listed in Table 4.5 and graphed in Fig. 4.2. Goodness-of-fit tests for two different stochastic differential equation models for this data are described in this example. The computer program that performed the goodness-of-fit tests is listed at the end of this chapter.

First, suppose a stochastic differential equation model of the form

$$dX(t) = \frac{\theta_1}{X^2(t)}\,dt + \frac{\theta_2}{X^2(t)}\,dW(t), \quad X(0) = 18, \tag{5.45}$$

where $X(t)$ is population size and the parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ are determined using the maximum likelihood procedure described in Section 4.9. Using this procedure, the values of θ_1 and θ_2 are estimated as 3510.0 and 13500.0, respectively. Performing $M = 8$ simulations of (5.45) at each data point for $t = t_1, t_2, \ldots, t_{46}$ and applying the goodness-of-fit test of this section, the value $Q_8 = 18.6$ is calculated. Since two parameters were estimated using the data, the number of degrees of freedom is adjusted to $M - 2 = 6$. The probability of having a value of χ^2 with 6 degrees of freedom this large is less than 0.005, that is, $P(\chi^2(6) \geq 18.6) < 0.005$. This indicates a lack-of-fit of the stochastic differential equation model (5.45) with the data.

Second, suppose a stochastic differential equation model of the form

$$dX(t) = \theta_1 X(t)\,dt + \sqrt{\theta_2 X(t)}\,dW(t), \quad X(0) = 18, \tag{5.46}$$

where $X(t)$ is population size and the parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ are determined using the maximum likelihood procedure described in Section 4.9. For this model, the values of θ_1 and θ_2 are estimated as 0.0361 and 0.609, respectively. For model (5.46), the goodness-of-fit test for $M = 8$ simulations gives the value $Q_8 = 4.09$. The probability that $\chi^2(6)$ has a value this large is greater than 0.66 which indicates that the null hypothesis cannot be rejected. That is, it cannot be rejected that the stochastic differential equation model (5.46) describes the whooping crane population size.

5.6 Alternate Equivalent SDE Models

Equivalent stochastic differential equation models, referred to in Remark 5.4, are discussed in this section. In the modeling procedure described in this chapter, a discrete stochastic model is developed by studying changes in the process states over a small time interval. Similarities between the forward Kolmogorov equations satisfied by the probability distributions of discrete- and continuous-time stochastic models infer an Itô stochastic differential equation model from the discrete stochastic model. This approach is a natural extension of the procedure used for many years in modeling deterministic dynamical processes in physics and engineering, where changes in the system are studied over a small interval of time and a differential equation is obtained as the time interval approaches zero. In this procedure, the number of Wiener processes never exceeds the number of states in the system.

In an alternate modeling procedure, the dynamical system is carefully studied to determine all the different independent random changes that occur in the system. Appropriate terms are determined for these changes in developing a discrete-time stochastic model which is then approximated by a stochastic differential equation system [43, 44, 45, 50]. In this procedure, the number of different random changes may possibly far exceed the number of states and a stochastic differential equation model is obtained where the number of Wiener processes may exceed the number of equations. This procedure is straightforward and yields, for example, stochastic differential equation systems that are generally easy to solve numerically.

In this section, it is shown that the two procedures produce stochastic differential equation systems that are structurally different yet have identical probability distributions [9]. Indeed, Euler's method for solving the different stochastic differential equation systems converge strongly in the mean-square sense to the same solution. As the stochastic models can be interchanged, conceptual or computational advantages possessed by either model can be employed in any particular problem.

To study equivalence of two stochastic differential equation systems, let

$$\mathbf{X}(t) = [X_1(t), X_2(t), \ldots, X_d(t)]^T,$$

$$\mathbf{W}(t) = [W_1(t), W_2(t), \ldots, W_m(t)]^T,$$

$$\mathbf{f} : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d,$$

and

$$G : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times m},$$

where $W_i(t)$, $i = 1, \ldots, m$ are independent Wiener processes and $m \geq d$. Consider the system of Itô stochastic differential equations of the form

$$d\mathbf{X}(t) = \mathbf{f}(t, \mathbf{X}(t))\, dt + G(t, \mathbf{X}(t))\, d\mathbf{W}(t). \tag{5.47}$$

In component form, system (5.47) can be expressed as

$$X_i(t) = X_i(0) + \int_0^t f_i(s, \mathbf{X}(s))\, ds + \int_0^t \sum_{j=1}^{m} g_{i,j}(s, \mathbf{X}(s))\, dW_j(s) \tag{5.48}$$

for $i = 1, 2, \ldots, d$, where f_i is the ith element of $\mathbf{f}$ and $g_{i,j}$ is the i, j entry of the $d \times m$ matrix G.

Itô's formula can be applied to (5.47). Let

$$\mathbf{F} : [0, T] \times \mathbb{R}^d \to \mathbb{R}^k \quad \text{and} \quad \mathbf{Y}(t) = \mathbf{F}(t, \mathbf{X}(t)). \tag{5.49}$$

For system (5.47), it follows from Itô's formula that the pth component of $\mathbf{Y}(t)$ satisfies

$$\begin{aligned} dY_p(t) = &\left[\frac{\partial F_p}{\partial t} + \sum_{i=1}^{d} f_i \frac{\partial F_p}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{l=1}^{m} g_{i,l} g_{j,l} \frac{\partial^2 F_p}{\partial x_i \partial x_j} \right] dt \\ &+ \sum_{i=1}^{d} \frac{\partial F_p}{\partial x_i} \sum_{l=1}^{m} g_{i,l}\, dW_l(t) \end{aligned} \tag{5.50}$$

for $p = 1, 2, \ldots, k$ where all the terms are evaluated at $(t, \mathbf{X}(t))$.

Furthermore, the forward Kolmogorov equation or Fokker-Planck equation for the probability density function $p(t, \mathbf{x})$ associated with the stochastic differential system (5.47) has the form

$$\begin{aligned} \frac{\partial p(t, \mathbf{x})}{\partial t} = &\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left[p(t, \mathbf{x}) \sum_{l=1}^{m} g_{i,l}(t, \mathbf{x}) g_{j,l}(t, \mathbf{x}) \right] \\ &- \sum_{i=1}^{d} \frac{\partial \big[p(t, \mathbf{x}) f_i(t, \mathbf{x}) \big]}{\partial x_i}, \end{aligned} \tag{5.51}$$

where, if $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^d$ and $\mathbf{z}_1 \le \mathbf{z}_2$, then

$$P(\mathbf{z}_1 \le \mathbf{X}(t) \le \mathbf{z}_2) = \int_{z_{1,d}}^{z_{2,d}} \int_{z_{1,d-1}}^{z_{2,d-1}} \cdots \int_{z_{1,1}}^{z_{2,1}} p(t, \mathbf{x})\, dx_1\, dx_2, \ldots, dx_d.$$

Consider the Euler-Maruyama (or Euler's) method for numerical solution of (5.47). Euler's method for system (5.47) has the form

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \mathbf{f}(t_n, \mathbf{X}_n) \Delta t + G(t_n, \mathbf{X}_n) \Delta \mathbf{W}_n \tag{5.52}$$

for $n = 0, 1, 2, \ldots, N-1$, where $\mathbf{X}_n \approx \mathbf{X}(t_n)$, $\Delta t = T/N$, and $\Delta \mathbf{W}_n = \mathbf{W}(t_{n+1}) - \mathbf{W}(t_n)$. In component form, Euler's method is

$$X_{i,n+1} = X_{i,n} + f_i(t_n, \mathbf{X}_n) \Delta t + \sum_{j=1}^{m} g_{i,j}(t_n, \mathbf{X}_n) \Delta W_{j,n} \tag{5.53}$$

for $i = 1, 2, \ldots, d$, where $\Delta W_{j,n} \sim N(0, \Delta t)$ are independent normally distributed random numbers for $n = 0, 1, \ldots$ and $j = 1, \ldots, d$.

Now define the $d \times d$ symmetric positive semidefinite matrix $V = GG^T$. Matrix V has entries

$$v_{i,j}(t, \mathbf{X}) = \sum_{l=1}^{m} g_{i,l}(t, \mathbf{X}) g_{j,l}(t, \mathbf{X}). \tag{5.54}$$

for $i, j = 1, \ldots, d$. Define the $d \times d$ symmetric positive semidefinite matrix $B = (b_{i,j}(t, \mathbf{X})) = V^{1/2}$. Then Euler's method given in (5.53) can be written equivalently as

$$X_{i,n+1} = X_{i,n} + f_i(t_n, \mathbf{X}_n)\Delta t + \sum_{j=1}^{d} b_{i,j}(t_n, \mathbf{X}_n)\Delta W^*_{j,n}, \tag{5.55}$$

where $\sum_{j=1}^{d} b_{i,j}(t_n, \mathbf{X}_n)\Delta W^*_{j,n} = \sum_{j=1}^{m} g_{i,j}(t_n, \mathbf{X}_n)\Delta W_{j,n}$ and $\Delta W^*_{j,n} \sim N(0, \Delta t)$ are independent normally distributed random numbers for $n = 0, 1, \ldots$ and $j = 1, \ldots, d$. That is, $B\Delta \mathbf{W}^*_n = G\Delta \mathbf{W}_n$.

To verify that $B\Delta \mathbf{W}^*_n = G\Delta \mathbf{W}_n$, an argument employing singular value decompositions can be employed. The singular value decomposition of G is $G = PDQ$ where P and Q are orthogonal matrices of sizes $d \times d$ and $m \times m$, respectively, and D is a $d \times m$ matrix with $r \leq d$ positive diagonal entries. It follows that $V = GG^T = PDD^TP^T$, where $B = P(DD^T)^{1/2}P^T$. Now, given B, G and $\Delta \mathbf{W}_n$, the following computation shows that there exists $\Delta \mathbf{W}^*_n$, a vector of d linearly independent normally distributed random numbers, such that $B\Delta \mathbf{W}^*_n = G\Delta \mathbf{W}_n$. Note that $P^T\Delta \mathbf{W}^*_n = ((DD^T)^{1/2})^+ DQ\Delta \mathbf{W}_n + \Delta \mathbf{W}^{**}_n$ and

$$\begin{aligned}
&E(P^T\Delta \mathbf{W}^*_n(\Delta \mathbf{W}^*_n)^T P) \\
&= E\Big(((DD^T)^{1/2})^+ DQ\Delta \mathbf{W}_n(\Delta \mathbf{W}_n)^T Q^T D^T((DD^T)^{1/2})^+ \Delta \mathbf{W}^{**}_n(\Delta \mathbf{W}^{**}_n)^T\Big) \\
&= I_d(\Delta t),
\end{aligned}$$

where I_d is the $d \times d$ identity matrix, $\Delta \mathbf{W}^{**}_n$ is a vector of length d with the first r entries equal to 0 and the next $d - r$ entries independent normally distributed random variables, and where $((DD^T)^{1/2})^+$ is the $d \times d$ pseudoinverse of $(DD^T)^{1/2}$ [93]. Conversely, given B, G, and $\Delta \mathbf{W}^*_n$, the following computation shows that there exists $\Delta \mathbf{W}_n$, a vector of m linearly independent normally distributed random numbers, such that $B\Delta \mathbf{W}^*_n = G\Delta \mathbf{W}_n$. Note that $Q\Delta \mathbf{W}_n = D^+(DD^T)^{1/2}P^T\Delta \mathbf{W}^*_n + \Delta \mathbf{W}^{***}_n$ and

$$\begin{aligned}
&E(Q\Delta \mathbf{W}_n(\Delta \mathbf{W}_n)^T Q^T) \\
&= E\Big(D^+(DD^T)^{1/2}P^T\Delta \mathbf{W}^*_n(\Delta \mathbf{W}^*_n)^T P(DD^T)^{1/2}(D^+)^T + \Delta \mathbf{W}^{***}_n(\Delta \mathbf{W}^{***}_n)^T\Big) \\
&= I_m(\Delta t),
\end{aligned}$$

where $\Delta \mathbf{W}_n^{***}$ is a vector of length m with the first r entries equal to 0 and the next $m-r$ entries independent normally distributed random variables and where D^+ is the $m \times d$ pseudoinverse of D.

Notice that system (5.55) is Euler's method for the stochastic system

$$X_i^*(t) = X_i^*(0) + \int_0^t f_i(s, \mathbf{X}^*(s))\, ds + \int_0^t \sum_{j=1}^{d} b_{i,j}(s, \mathbf{X}^*(s))\, dW_j^*(s) \quad (5.56)$$

for $i = 1, \ldots, d$. Because Euler's method converges in the mean square sense, (5.47) and (5.56) are equivalent stochastic systems in the sense that they share the same sample paths. Also notice that Itô's formula for (5.56) with $\mathbf{F}$ and $\mathbf{Y}$ defined in (5.49) satisfies

$$\begin{aligned} dY_p(t,\omega) = &\left[\frac{\partial F_p}{\partial t} + \sum_{i=1}^{d} f_i \frac{\partial F_p}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} v_{i,j} \frac{\partial^2 F_p}{\partial x_i \partial x_j} \right] dt \\ &+ \sum_{i=1}^{d} \frac{\partial F_p}{\partial x_i} \sum_{j=1}^{d} b_{i,j} dW_j^*(t) \end{aligned} \quad (5.57)$$

for $p = 1, \ldots, k$. In addition, the forward Kolmogorov equation corresponding to (5.56) is

$$\begin{aligned} \frac{\partial p(t,\mathbf{x})}{\partial t} = &- \sum_{i=1}^{d} \frac{\partial \left[p(t,\mathbf{x}) f_i(t,\mathbf{x}) \right]}{\partial x_i} \\ &+ \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left[p(t,\mathbf{x}) v_{i,j}(t,\mathbf{x}) \right]. \end{aligned} \quad (5.58)$$

In effect, system (5.47) with $m \geq d$ Wiener processes can be written as an equivalent system with d Wiener processes

$$d\mathbf{X}(t) = \mathbf{f}(t, \mathbf{X}(t))\, dt + B(t, \mathbf{X}(t))\, d\mathbf{W}^*(t), \quad (5.59)$$

where $\mathbf{W}^*(t) = [W_1^*(t), W_2^*(t) \ldots, W_d^*(t)]^T$ and the $d \times d$ matrix B satisfies $B^2 = GG^T$.

Consider now a stochastic modeling problem that involves the d states $S_1, S_2, \ldots, S_d$ with a total of $m \geq d$ possible random changes to these states at each time step Δt. Suppose that the probabilities of the changes are $p_j \Delta t \equiv p_j(t, \mathbf{S})\Delta t$ for $j = 1, 2, \ldots, m$, where the jth change alters the ith state by the amount $\lambda_{j,i}$ for $i = 1, \ldots, d$. A standard deterministic model for this problem is the following system of ordinary differential equations:

$$d\mathbf{S}(t) = \mathbf{f}(t, \mathbf{S}(t))\, dt, \quad (5.60)$$

where the ith element of the vector $\mathbf{f}$ is

$$f_i(t, \mathbf{S}(t)) = \sum_{j=1}^{m} p_j(t, \mathbf{S}(t))\lambda_{j,i} \tag{5.61}$$

for $i = 1, 2, \ldots, d$. For a small time step Δt, problem (5.60) can be approximated using Euler's method by the formula

$$S_{n+1,i} = S_{n,i} + f_i(t_n, \mathbf{S}_n)\Delta t, \tag{5.62}$$

where $t_n = n\Delta t$ and $S_{n,i} \approx S_i(t_n)$ for $i = 1, \ldots, d$ and $n = 0, 1, \ldots$.

Assuming that Δt is a small but fixed time interval, a discrete-time stochastic model can be formulated by considering the random changes at each time step. Let $\mathbf{r}_j$ represent a random change of the jth kind to the state vector. Then $\mathbf{r}_j$, to order $O((\Delta t)^2)$, is defined as follows:

$$\mathbf{r}_j = \begin{cases} [\lambda_{j,1}, \lambda_{j,2}, \ldots, \lambda_{j,d}]^T & \text{with probability } \; p_j \Delta t \\ [0, 0, \ldots, 0]^T & \text{with probability } \; 1 - p_j \Delta t. \end{cases}$$

For Δt small, $(\mathbf{r}_j)_i$ has approximate mean $\lambda_{j,i} p_j \Delta t$ and variance $\lambda_{j,i}^2 p_j \Delta t$. A discrete stochastic model for $\mathbf{S}_{n+1}$, given the vector $\mathbf{S}_n$, is

$$\mathbf{S}_{n+1} = \mathbf{S}_n + \sum_{j=1}^{m} \mathbf{r}_j \tag{5.63}$$

for $n = 0, 1, \ldots$. In component form, (5.63) becomes

$$S_{n+1,i} = S_{n,i} + \sum_{j=1}^{m} (\mathbf{r}_j)_i \tag{5.64}$$

for $i = 1, \ldots, d$ and $n = 0, 1, \ldots$.

In the alternate modeling procedure, the random changes are approximated using m independent normal random variables and Eq. (5.64) for small Δt takes the form

$$S_{n+1,i} = S_{n,i} + f_i(t_n, \mathbf{S}_n)\Delta t + \sum_{j=1}^{m} \lambda_{j,i} p_j^{1/2} (\Delta t)^{1/2} \eta_j \tag{5.65}$$

for $n = 0, 1, \ldots$, where f_i is defined in (5.61) and $\eta_j \sim N(0, 1)$ for each $j = 1, \ldots, m$. Notice the similarity between the deterministic equation (5.62) and the stochastic equation (5.65). The discrete stochastic model (5.65) converges strongly (in the mean-square sense) as $\Delta t \to 0$ to the SDE system

$$\begin{cases} d\mathbf{S}(t) = \mathbf{f}(t, \mathbf{S}(t))\, dt + G(t, \mathbf{S}(t))\, d\mathbf{W}(t), \\ \mathbf{S}(0) = \mathbf{S}_0, \end{cases} \tag{5.66}$$

where the i, j entry in the matrix G is $g_{i,j} = \lambda_{j,i} p_j^{1/2}$ for $i = 1, 2, \ldots, d$, $j = 1, 2, \ldots, m$, and $\mathbf{W}(t)$ is a vector of m independent Wiener processes. Notice that SDE system (5.66) has m Wiener processes and the $d \times d$ matrix $V = GG^T$ has entries

$$(V)_{i,l} = (GG^T)_{i,l} = \sum_{j=1}^{M} g_{i,j} g_{l,j} = \sum_{j=1}^{M} p_j \lambda_{ji} \lambda_{jl} = v_{i,l} \tag{5.67}$$

for $i, l = 1, \ldots, d$. The entries in G are easy to write down given the probabilities of the different changes based on the discrete-time Markov chain (5.63). It is interesting to note that there are other SDE systems equivalent to (5.66) that can be generated from the probabilities of the changes. For example, if matrix G is replaced by $-G$ in (5.66), this alternate system generates the same sample paths. In general, a matrix H can be used instead of G if $HH^T = V$, where $V = GG^T$.

In the modeling procedure discussed in this chapter, if the changes are small and Δt is small, then the probability distribution associated with the discrete-time stochastic system (5.63) can be approximated by the solution to the forward Kolmogorov equation

$$\begin{aligned} \frac{\partial p(t, \mathbf{x})}{\partial t} &= -\sum_{i=1}^{d} \frac{\partial \left[p(t, \mathbf{x}) f_i(t, \mathbf{x}) \right]}{\partial x_i} \\ &\quad + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left[p(t, \mathbf{x}) v_{i,j}(t, \mathbf{x}) \right], \end{aligned} \tag{5.68}$$

where $v_{i,j}$ is the i, jth entry of $d \times d$ matrix V. The probability distribution $p(t, x_1, x_2, \ldots, x_d)$ that solves (5.68) is identical to the distribution of solutions corresponding to the stochastic differential equation system

$$\begin{cases} d\mathbf{S}(t) = \mathbf{f}(t, \mathbf{S}(t))\, dt + B(t, \mathbf{S}(t))\, d\mathbf{W}^*(t) \\ \mathbf{S}(0) = \mathbf{S}_0, \end{cases} \tag{5.69}$$

where the $d \times d$ matrix $B = V^{1/2}$ and $\mathbf{W}^*(t)$ is a vector of d independent Wiener processes. Therefore, the discrete stochastic model (5.63) is closely related to a stochastic differential equation model (5.69). Specifically, the probability distribution of solutions to (5.63) is approximately the same as the probability distribution of solutions to (5.69). In addition, the drift vector and diffusion matrix, $\mathbf{f}$ and B, respectively, of the stochastic differential equation model are equal to the expected change divided by Δt and the square root of the covariance matrix of the change divided by Δt. Specifically, letting $\boldsymbol{\lambda}_j = [\lambda_{j,1}, \lambda_{j,2}, \ldots, \lambda_{j,n}]^T$, then the expected change in $\mathbf{S}$ and the covariance in the change are

$$\begin{cases} E(\Delta \mathbf{S}) = \sum_{j=1}^{m} p_j \boldsymbol{\lambda}_j \Delta t = \mathbf{f} \Delta t \\ E(\Delta \mathbf{S} (\Delta \mathbf{S})^T) = \sum_{j=1}^{m} p_j \boldsymbol{\lambda}_j (\boldsymbol{\lambda}_j)^T \Delta t = V \Delta t. \end{cases} \tag{5.70}$$

Notice that the $d \times m$ matrix G satisfies $V = GG^T$ and the stochastic differential equation system (5.69) can be modified to the system (5.66) by the argument in this section. Indeed, Itô's formula and the forward Kolmogorov equation are identical for both stochastic differential equation systems (5.66) and (5.69). Finally, notice that system (5.69) is generally more complicated than (5.66) as the $d \times d$ matrix B is the square root of V even though G is a $d \times m$ matrix. Consequently, system (5.66) is generally easier to solve computationally. However, if the number of changes, m, is much greater than the number of states, d, then equation (5.66) loses much of its computational advantages.

In summary, in the alternate procedure for constructing an Itô stochastic differential equation model for a dynamical system consisting of d states with $m \geq d$ different independent random changes, each independent random change is explicitly included. This procedure is in contrast to the modeling procedure described in this chapter where means and covariances of the random changes are calculated which, in turn, determine the stochastic differential equation model.

To illustrate the alternate modeling procedure, the chemical reaction problem described in Section 5.3.6 is modeled using the alternate procedure. As described in this example, suppose that there are three chemical species S_1, S_2, and S_3 that interact through molecular collisions or spontaneously in the four ways described in Table 5.9.

Using the alternate modeling procedure for this example gives the SDE model:

$$d\mathbf{X}(t) = \mathbf{f}(X_1, X_2, X_3)\, dt + C(X_1, X_2, X_3)\, d\mathbf{W}^*(t) \tag{5.71}$$

with $\mathbf{X}(0) = [X_1(0), X_2(0), X_3(0)]^T$ and where $\mathbf{W}^*(t)$ is a vector $\mathbf{W}^*(t) = [W_1^*(t), W_2^*(t), W_3^*(t), W_4^*(t)]^T$ of four independent Wiener processes and 3×4 matrix C has the form

$$C = \begin{bmatrix} -(\mu_1 X_1 X_2)^{1/2} & (\mu_2 X_3)^{1/2} & 2(\mu_3 X_2^2 X_3/2)^{1/2} & -2(\mu_4 X_1^2/2)^{1/2} \\ -(\mu_1 X_1 X_2)^{1/2} & (\mu_2 X_3)^{1/2} & -2(\mu_3 X_2^2 X_3/2)^{1/2} & 2(\mu_4 X_1^2/2)^{1/2} \\ (\mu_1 X_1 X_2)^{1/2} & -(\mu_2 X_3)^{1/2} & -(\mu_3 X_2^2 X_3/2)^{1/2} & (\mu_4 X_1^2/2)^{1/2} \end{bmatrix}.$$

To test the stochastic differential equation model (5.71), calculational results using the two SDE models were compared. In these calculations, the values of the reaction rate constants were taken as $\mu_1 = 0.02, \mu_2 = 0.4, \mu_3 = 0.001$, and $\mu_4 = 0.03$. The initial numbers of molecules were assumed to be $X_1(0) = X_2(0) = X_3(0) = 100$ and the final time was taken as $t = 1.0$. The SDE models were numerically solved using the Euler-Maruyama method with 5000 sample paths. The results using these two SDE models are compared in Table 5.16. Also, a sample path using model (5.33) is plotted in Fig. 5.10 and a sample path using model (5.71) is plotted in Fig. 5.13. Notice the good agreement between the two different SDE models.

Table 5.16. Calculated mean molecular levels and standard deviations at time $t = 1.0$ using SDE models (5.33) and (5.71)

Chemical Species	$E(X_i)$ (5.33)	$\sigma(X_i)$ (5.33)	$E(X_i)$ (5.71)	$\sigma(X_i)$ (5.71)
S_1	79.31	7.62	79.39	7.69
S_2	37.44	6.14	37.47	6.13
S_3	131.17	6.43	131.09	5.85

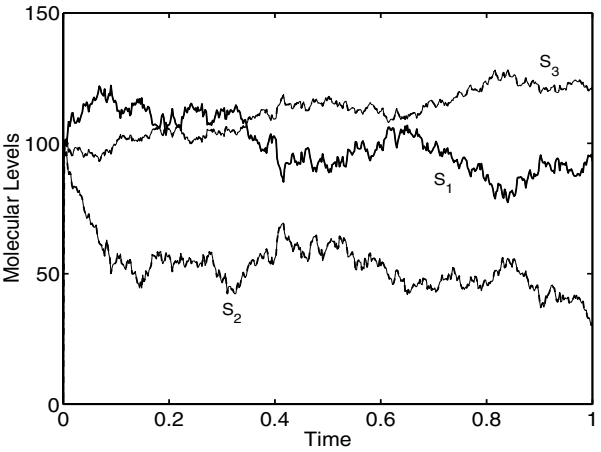

Fig. 5.13. Molecular population levels for one sample path of SDE (5.71)

Exercises

5.1. In the model described schematically in Fig. 5.1, suppose that $m_{12} = a_1 S_1$, $m_{21} = a_2 S_2$, $b_1 = b_2 = d_1 = d_2 = m_{22} = m_{11} = 0$, and $\lambda_1 = \lambda_2$.
(a) Derive a stochastic differential equation model for this system.
(b) Show that $d(S_1(t) + S_2(t)) = 0$ so that $S_1(t) + S_2(t)$ is constant with time.

5.2. In the model described schematically in Fig. 5.1, suppose that $m_{12} = a_1 S_1$, $m_{21} = a_2 S_2$, $b_1 = r_1 S_1$, and $b_2 = d_1 = d_2 = m_{22} = m_{11} = 0$.
(a) Find the matrix $V = E(\Delta\mathbf{S}(\Delta\mathbf{S})^T)/\Delta t$ for this problem.
(b) Use Remark 5.4 to find a 2×3 matrix C such that $CC^T = V$.

5.3. Show that the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ has no matrix square root while the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}^2$$

has an infinite number of matrix square roots.

5.4. Suppose that a cricket begins walking at the origin and proceeds along the positive x-axis. At each time step Δt, the cricket walks $3\Delta t$ units to the right, i.e., the cricket's speed is 3. In addition, at each time step Δt, the cricket jumps back one unit with probability $\Delta t/2$.
(a) Derive a stochastic differential equation model that approximately describes the position of the cricket.
(b) For your stochastic differential equation model, show that $E(X(t)) = 5t/2$ and $\mathrm{Var}(X(t)) = t/2$ where $X(t)$ is the cricket's position at time t.

5.5. Consider three populations S_1, S_2, and S_3 that compete with each other. For a small time interval Δt, suppose that the probabilities of a birth in each of the three populations are, respectively, $b_1 S_1 \Delta t, b_2 S_2 \Delta t$, and $b_3 S_3 \Delta t$ and the probabilities of a death in each of the three populations are, respectively, $(a_1 S_1 + u_{11} S_1^2 + u_{12} S_1 S_2 + u_{13} S_1 S_3)\Delta t$, $(a_2 S_2 + u_{21} S_1 S_2 + u_{22} S_2^2 + u_{23} S_2 S_3)\Delta t$, and $(a_3 S_3 + u_{31} S_1 S_3 + u_{32} S_2 S_3 + u_{33} S_3^2)\Delta t$. Derive a stochastic differential equation model for the population sizes, $[S_1(t), S_2(t), S_3(t)]^T$, with time t.

5.6. Suppose that $b = d = 1$, $X(0) = 100$, and the maximum population size is $K = 200$.
(a) Find the mean persistence time, $T(100)$, of the population. Note that the mean persistence time $T(y)$ with initial population size y satisfies the second order differential equation

$$y \frac{d^2 T(y)}{dy^2} = -1 \quad \text{with} \quad T(0) = 0, \frac{dT(200)}{dy} = 0.$$

(b) Compare the mean persistence time obtained in part (a) with the persistence time for the corresponding deterministic population model $dX(t) = (b - d)X(t)\, dt, X(0) = 100$.

5.7. Consider an SIR model of an epidemic where the births and deaths are assumed to be negligible. In this model, $S(t), I(t)$, and $R(t)$ are the susceptible, infected, and removed population sizes, respectively, and $N = S(t) + I(t) + R(t)$ is constant. Individuals are removed after recovering from the disease and are assumed to be immune thereafter. Suppose that the probability of one individual becoming infected is $\beta S I \Delta t / N$ during a small time interval Δt and the probability of one individual recovering is $\gamma I \Delta t$. Derive a stochastic differential equation for this randomly varying system.

5.8. Consider finding the mean exit time of trajectories for the stochastic differential equation

$$dy(t) = -2y(t)\, dt + \sqrt{4y(t)}\, dW(t), \; y(0) = 10,$$

where there is no maximum population size and exit occurs when the trajectory reaches zero. Therefore, for N trajectories, $T \approx \sum_{m=1}^{N} t_m/N$ where t_m is the time for the mth trajectory to reach zero, i.e., $y(t_m) = 0$. The MATLAB program listed below estimates exit times for trajectories of a stochastic differential equation. In using the program to estimate the mean exit time, vary the number of trajectories, the step length used in Euler's method, the final time, and the maximum size of $y(t)$ to obtain the mean exit time to within 10% of the exact mean exit time. Compare your calculated result with the exact exit time

$$T = \int_0^{10} \frac{1 - e^{-z}}{2z}\, dz + \int_0^{\infty} \frac{(1 - e^{-10})e^{-z}}{2z + 20}\, dz.$$

```
% A program for first exit calculations
 clear
 nsamp=20;
% nsamp paths are calculated
% statistical error decreases as nsamp increases
 t=1;
% The proportion exiting up to time t are calculated
% t must be sufficiently large to estimate mean exit time
 nt=fix(.001+t/.005);
 % method error decreases as nt increases
 y0=10;
 ymax=40;
% y0 is the initial position
% ymax is maximum position; paths reflect at y=ymax
% Exits occur at y=yexit
 yexit=0;
 alp=-2;
 bet=4;
 h=t/nt;
 hs=sqrt(h);
 randn('state',2)
 for ncase=1:3
 sumex=0;
 te=zeros(nsamp,1);
 for jj=1:nsamp
   y=y0;
   r=randn(nt,1);
   nchk=0;
   n=0;
   t=0;
   while (n < nt)
   n=n+1;
```

```
    t=t+h;
    f=alp*y;
    g=sqrt(bet*y);
    y=y+h*f+hs*g*r(n);
% This is Euler's approximation to the SDE
    if(y > ymax)
    y=2*ymax-y;
    end
    if (y <= yexit)
     nchk=1;
     te(jj)=t;
     n=nt;
    end
    end
     sumex=sumex+nchk;
    end
    tp=sum(te)/sumex;
    p=sumex/nsamp;
% tp is mean exit time for paths that exit
% p is proportion exiting
disp((sprintf(' %12.0f %12.6f %12.6f', ncase, tp, p)));
end
```

5.9. Consider two groups of particles revolving around a circular path. Let $S_1(t)$ be the number of particles moving clockwise and $S_2(t)$ be the number of particles moving counterclockwise. Suppose at each time interval Δt, the particles scatter with the medium and one clockwise particle changes direction with probability $\gamma S_1(t)\Delta t$. Likewise, at each time interval Δt, one counterclockwise particle changes direction with probability $\gamma S_2(t)\Delta t$. Find a stochastic differential equation system for $S_1(t)$ and $S_2(t)$. (A check on your model is that the total number of particles should be constant for all time, i.e., $d(S_1(t) + S_2(t)) = 0$ for all t.)

5.10. Consider the particle transport system of the previous exercise. Suppose now, however, in addition to scattering with the medium there are probabilities $\beta S_1(t)\Delta t$ and $\beta S_2(t)\Delta t$ of one particle being absorbed by the medium for the clockwise-moving group and of one particle being absorbed by the medium for the counterclockwise-moving group, respectively, during each time interval Δt.
(a) Construct a stochastic differential equation model for this new system.
(b) Determine $E(S_1(t) + S_2(t))$ for $t \geq 0$. (In this problem, the total number of particles is not constant.)

5.11. Suppose that dust particles of mass m and altitude $y(t)$ are drifting upward with speed $v(t)$ under the influence of a randomly varying upward air

current of speed $v_c(t)$. The particles experience a gravitational force downward, $-mg$, and a frictional force upward, $kv_a(t)$, where k is a constant and $v_a(t) = v_c(t) - v(t)$ is the air speed at the particle surface. The air current speed $v_c(t)$ is randomly varying with mean speed v_e. Assume that during any small time interval Δt, the probability of an increase of magnitude α in $v_c(t)$ is $(\lambda + \beta(v_e - v_c(t))\Delta t$ and the probability of a decrease of $-\alpha$ in $v_c(t)$ is $(\lambda - \beta(v_e - v_c(t))\Delta t$. Derive a stochastic differential equation system for $v_c(t)$, $y(t)$, and $v(t)$.

5.12. Suppose that in a certain chemically reacting system, the two chemical species S_1 and S_2 are reacting in the three ways:

$$S_1 + 3S_2 \to 3S_1 + S_2, \quad S_1 + S_2 \to 2S_2, \quad \text{and} \quad S_1 \to S_2.$$

Furthermore, during a small time interval Δt, the probabilities of these three reactions are $\beta_1 X_1(t) X_2^3(t)\Delta t/6$, $\beta_2 X_1(t) X_2(t)\Delta t$, and $\beta_3 X_1(t)\Delta t$, respectively, where $X_1(t)$ and $X_2(t)$ are the number of molecules of species S_1 and S_2, respectively.
(a) Construct a stochastic differential equation model for the reacting system of the form $d\mathbf{X} = \boldsymbol{\mu}\, dt + V^{1/2}\, d\mathbf{W}(t)$ where V is a 2×2 symmetric positive semidefinite matrix.
(b) Find a stochastic differential equation model for the system of the form $d\mathbf{X} = \boldsymbol{\mu}\, dt + C\, d\mathbf{W}^*(t)$ where C is a 2×3 matrix.

5.13. For a single stock whose price satisfies the stochastic differential equation $dS = \mu S\, dt + \sigma S\, dW(t)$ with interest rate r, the Black-Scholes equation is:

$$\frac{\partial U(t,S)}{\partial t} + rS\frac{\partial U(t,S)}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 U(t,S)}{\partial S^2} = rU(t,S).$$

For a European call option, U satisfies at time T the condition $U(T,S) = \max\big(S - K, 0\big)$, where K is the strike price of the option. Show that the Black-Scholes formula:

$$U(t,S) = SN(x_1) - Ke^{-r(T-t)}N(x_2),$$

where $x_1 = \big(\ln(S/k)(T-t)^{-1/2} + (r+\sigma^2/2)(T-t)^{1/2}\big)/\sigma$, $x_2 = x_1 - \sigma(T-t)^{1/2}$, and $N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\, dy$ satisfies the Black-Scholes equation and also satisfies the required condition at time T. Thus,

$$U(0,S) = S(0)N(x_1(0)) - Ke^{-rT}N(x_2(0))$$

is the price of the option at time $t = 0$ where $x_1(0) = \big(\ln(S(0)/k)T^{-1/2} + (r + \sigma^2/2)T^{1/2}\big)/\sigma$, and $x_2(0) = x_1(0) - \sigma T^{1/2}$. For this problem, to simplify the verification, note that

$$\frac{\partial N(x)}{\partial z} = N'(x)\frac{\partial x}{\partial z} \quad \text{and} \quad \frac{\partial^2 N(x)}{\partial z^2} = N'(x)\frac{\partial^2 x}{\partial z^2} - xN'(x)\left(\frac{\partial x}{\partial z}\right)^2.$$

5.14. Consider a stock whose price is modeled by the stochastic differential equation $dS(t) = \mu S(t)\,dt + \sigma(t)S(t)\,dW(t)$ where the volatility $\sigma(t)$ is varying randomly with time t about a mean value σ_e. Suppose that during a small time interval Δt, $\sigma(t)$ changes the amount $\gamma\sigma(t)$ with probability $\big(\alpha + \beta(\sigma_e - \sigma(t))\big)\Delta t$ and changes the amount $-\gamma\sigma(t)$ with probability $\big(\alpha - \beta(\sigma_e - \sigma(t))\big)\Delta t$. Construct a stochastic differential equation for the volatility $\sigma(t)$.

5.15. (Project) Consider two species that are in competition. The sizes of the two populations are $y_1(t)$ and $y_2(t)$, respectively, where t is time in years. Assume that the probability of a birth in a small time interval Δt for the first population is $\frac{5}{6}y_1(t)\Delta t$ and the the probability of a birth in the second population is $\frac{9}{10}y_2(t)\Delta t$. Also assume that the probability of a death in the first population is $\big(\frac{2}{5}y_1(t) + \frac{1}{100}y_1^2(t) + \frac{1}{45}y_1(t)y_2(t)\big)\,\Delta t$ and the probability of a death in the second population is $\big(\frac{3}{4}y_2(t) + \frac{1}{150}y_2^2(t) + \frac{1}{200}y_1(t)y_2(t)\big)\,\Delta t$. Furthermore, assume that the initial population sizes are $y_1(0) = y_2(0) = 15$.
(a) Construct a stochastic differential equation model for the population sizes $y_1(t)$ and $y_2(t)$.
(b) Modify the computer program in Exercise 5.8 to calculate exit times for a stochastic differential equation system so that exit occurs when either $y_1 \le 0$ or $y_2 \le 0$. Calculate to within 0.1 year the mean exit time for these two competing populations. Also, calculate the proportion of population trajectories where the first population reaches zero before the second population and the proportion of population trajectories where the second population first reaches zero. Notice that statistical and method errors in the computations need to be studied to ensure that the calculational results are accurate.
(c) Compare the results of the stochastic differential equation model with the results predicted by the deterministic competition model. In particular, computationally solve the deterministic system and determine whether one population drives the other population to extinction.
(d) Hand in a listing of your computer program for part (b) along with explanations and results for parts (a), (b), and (c). (A check on your model in parts (a) and (c) is that an equilibrium for the population sizes in the deterministic model occurs at $y_1^{(e)} = 10$ and $y_2^{(e)} = 15$. A check on your computations for part (c) is that the calculated mean exit time should be between 14.4 and 15.4 years.)

5.16. (Project) A lifeboat was launched at position $(0, 0)$ in the ocean and has been drifting for 168 hours. It is necessary to estimate the lifeboat's position $(x(t), y(t))$. It is known that ocean currents are affecting the velocity $v\cos(\theta)\mathbf{i} + v\sin(\theta)\mathbf{j}$ of the lifeboat. Assume that the speed v and the angle θ are independently randomly varying. In particular, assume that for a small time interval Δt, the probability of an increase in speed of .1 kilometer per

hour is $(8+1.5(5-v(t)))\Delta t$ and the probability of a decrease in speed of $-.1$ kilometer per hour is $(8-1.5(5-v(t)))\Delta t$. For a small time interval Δt, the probability of a change in the angle of .1 radians is $(2+.5(\theta_e(t)-\theta(t)))\Delta t$ and the probability of a change in angle of $-.1$ radians is $(2-.5(\theta_e(t)-\theta(t)))\Delta t$ where $\theta_e(t) = \frac{\pi}{2}\left(\frac{x(t)}{x(t)+300}\right)$. Furthermore, assume that the initial values are $v(0) = \theta(0) = x(0) = y(0) = 0$.
(a) Construct a stochastic differential equation model for the quantities $v(t)$, $\theta(t)$, $x(t)$, $y(t)$ noting that $\frac{dx}{dt} = v\cos(\theta)$ and $\frac{dy}{dt} = v\sin(\theta)$.
(b) Modify, for example, the computer program in Exercise 2.10, to calculate trajectories of the stochastic differential equation system. Calculate to within 1 kilometer the mean values of the x and y coordinates of the lifeboat's position at 168 hours. Also, calculate the standard deviations in the x and y coordinates at 168 hours. Notice that statistical and method errors in the computations need to be studied to ensure that the calculational results are accurate.
(c) Plot two lifeboat trajectories and the average trajectory. Specifically, make two-dimensional plots of the lifeboat's position, $(x(t), y(t))$, with time t for $0 \le t \le 168$.
(d) Hand in a listing of your computer program for part (b) along with explanations, results, and plots for parts (a), (b), and (c). (A check on your model and computations is that the x coordinate of the lifeboat's position at 168 hours has mean value 558 kilometers and a standard deviation of 64 kilometers.)

Computer Programs

Program 5.1. Program to solve an SDE model of a SIS problem

This Fortran program solves a stochastic SIS system using Euler's method with nt equally spaced steps in time. The number of sample paths calculated is specified as nrun. The mean, mean square, and standard deviation in the solution at the final time are output. The parameter y1 is the susceptible population size and y2 is the infected population size. The program is set up to solve the stochastic differential equation system described in Section 5.2.2 Output of the program is given following the program listing.

```
      real*8 xx
      nt=2000
      xx=981177.
c This program solves a stochastic SIS problem.
      nrun=10000
      arun=nrun
      sm1=0.0
```

```
      sq1=0.0
      sm2=0.0
      sq2=0.0
      ant=nt
      arun=nrun
      do 75 jj=1,nrun
      y1=950.0
      y2=50.0
      time=100.0
      h=time/ant
      hs=sqrt(h)
      t=0.0
      do 600 i=1,nt
      call random(xx,rand1,rand2)
      call random(xx,rand3,rand4)
      call fts(t,y1,y2,f1,f2,g1,g2,g3,g4)
      t=t+h
      y1=y1+h*f1+hs*rand1*(g1)+rand2*hs*(g2)
      y2=y2+h*f2+hs*rand1*(g3)+rand2*hs*(g4)
600   continue
      sm1=sm1+y1/arun
      sq1=sq1+y1*y1/arun
      sm2=sm2+y2/arun
      sq2=sq2+y2*y2/arun
75    continue
      sqr1=sqrt(sq1-sm1*sm1)
      sqr2=sqrt(sq2-sm2*sm2)
      write(6,162) h,time
      write(6,162) sm1,sq1,sqr1
      write(6,162) sm2,sq2,sqr2
162   format(5x,,3(f12.2,3x))
      stop
      end
      subroutine random(xx,rand1,rand2)
      real*8 xx,a,b,d,rng(2)
      a=16807.
      ib=2147483647
      b=ib
      do 55 i=1,2
      id=a*xx/b
      d=id
      xx=a*xx-d*b
55    rng(i)=xx/b
      pi=3.141592654
      u1=rng(1)
```

```
      u2=rng(2)
      hlp=sqrt(-2.0*alog(u1))
      rand1=hlp*cos(pi*2.0*u2)
      rand2=hlp*sin(pi*2.0*u2)
      return
      end
      subroutine fts(t,y1,y2,f1,f2,g1,g2,g3,g4)
      alp=.04
      gam=.01
      am12=alp*y2/(y1+y2)
      am21=gam
      f1=-am12*y1+am21*y2
      f2=am12*y1-am21*y2
      g1=sqrt((am12*y1+am21*y2)/2.0)
      g2=-g1
      g3=-g1
      g4=g1
      return
      end
      0.05          100.00
      561.66       317186.25          41.49
      438.34       193861.39          41.49
```

Program 5.2. Monte Carlo program of an SIS problem

This Fortran program solves a stochastic SIS system using a Monte Carlo method. In each time step of length h, every individual in the population is checked to determine whether recovery or infection occurs. The number of sample paths calculated is specified as nrun. The mean, mean square, and standard deviation of the population sizes at the final time are output. The parameter nx1 is the susceptible population size and nx2 is the infected population size. The program is set up to solve the SIS problem described in Section 5.2.2. Output of the program is given following the program listing.

```
      real*8 xx
      xx=3404658.
c This is a Monte Carlo program to simulate an SIS epidemic.
      nrun=10000
c The value of nrun is the number of sample paths.
      x1av=0.0
      x2av=0.0
      x1x1av=0.0
      x2x2av=0.0
      do 400 nr=1,nrun
      am12=.04
```

```
      am21=.01
      nx1=950
      nx2=50
      ntot=nx1+nx2
      nt=500
      ant=nt
      time=100.0
      h=time/ant
      do 100 i=1,nt
      if(nx1.lt.0) nx1=0
      if(nx2.lt.0) nx2=0
      mx1=0
      mx2=0
      do 200 j=1,nx1
      call random(xx,r)
      antot=ntot
      anx2=nx2
      am12=.04*anx2/antot
      if(am12*h < r) goto 200
      mx2=mx2+1
      mx1=mx1-1
200   continue
      do 300 j=1,nx2
      call random(xx,r)
      if(am21*h < r) goto 300
      mx2=mx2-1
      mx1=mx1+1
300   continue
      nx1=nx1+mx1
      nx2=nx2+mx2
100   continue
      anx1=nx1
      anx2=nx2
      anrun=nrun
      x1av=x1av+anx1/anrun
      x2av=x2av+anx2/anrun
      x1x1av=x1x1av+anx1*anx1/anrun
      x2x2av=x2x2av+anx2*anx2/anrun
400   continue
      x1sd=sqrt(x1x1av-x1av*x1av)
      x2sd=sqrt(x2x2av-x2av*x2av)
      write(6,550) nrun,h,time,am12,am21
550   format(2x,i5,6(2x,f8.3))
      write(6,560) x1av,x1x1av,x1sd
      write(6,560) x2av,x2x2av,x2sd
```

```
560  format(2x,3(2x,f12.2))
     stop
     end
     subroutine random(xx,r)
     real*8 xx,a,b,d
     a=16807.
     ib=2147483647
     b=ib
     id=a*xx/b
     d=id
     xx=a*xx-d*b
     r=xx/b
     return
     end
10000  0.200  100.000  0.017  0.010
562.28     317837.06         40.98
437.72     193279.28         40.98
```

Program 5.3. To plot solutions of an SDE model of an epidemic

This MATLAB program solves a stochastic SIS system using Euler's method with nt equally spaced steps in time. The number of sample paths calculated is specified as nrun. The parameter ys is the susceptible population size and yi is the infected population size. Plots are produced of individual susceptible and infected population trajectories with time along with average population sizes. The program is set up to solve the SIS stochastic differential equation system described in Section 5.2.2 for producing Fig. 5.3.

```
% Stochastic SIS model is programmed in MATLAB
% random.m file is needed
% rand=random(xx) returns with three numbers
% rand(1), rand(2) are normally distributed
% xx is set equal to rand(3) for next call to random.m
% average of nrun paths and individual paths are plotted
clear
 clf
 xx =98945;
% xx starts the random number sequence
 nt=400;
 nrun=100;
 time=200;
 % nt is number of time intervals, time is total time
 % nrun is the number of different paths
 h=time/nt;
 hs=sqrt(h);
```

```
tt=linspace(0,time,nt+1);
for i=1:nt+1
   sms(i)=0;
   smi(i)=0;
  paths(i)=0;
pathi(i)=0;
end
for jj=1:nrun
   ys=950;
   yi=50;
   sms(1)=ys;
   smi(1)=yi;
  paths(1)=ys;
  pathi(1)=yi;
  t=-h;
for i=1:nt
  t=t+h;
  rand=random(xx);
  xx=rand(3);
  % need xx=rand(3) for next call to random.m
  m12= .04*yi/(ys+yi);
  m21=.01;
  hlp1=-m12*ys+m21*yi;
  hlp2=sqrt((m12*ys+m21*yi)/2);
  ys=ys+h*hlp1+hlp2*hs*(rand(1)-rand(2));
  yi=yi-h*hlp1+hlp2*hs*(-rand(1)+rand(2));
  sms(i+1)=sms(i+1)+ys/nrun;
  smi(i+1)=smi(i+1)+yi/nrun;
  paths(i+1)=ys;
  pathi(i+1)=yi;
end
end
set(gca,'fontsize',18,'linewidth',1.5);
plot(tt,paths,'k-','linewidth',1.5)
axis([0,100,0,1000]);
hold on
plot(tt,pathi,'k-','linewidth',1.5)
plot(tt,sms,'k--','linewidth',1.5)
plot(tt,smi,'k--','linewidth',1.5)
xlabel('Time')
ylabel('Population Sizes')
set(gca,'linewidth',1.5)
hold off
% Listed next is function program random.m
% Uses congruential generator xx=16807*xx mod(2^31-1)
```

```
% Box-Muller method converts to normal random numbers
% xx=rand(3) is input to the generator
function rand = random(xx)
       a=16807;
       b=2147483647;
   for i=1:2
       d=fix(a*xx/b);
       xx=a*xx-d*b;
       rng(i)=xx/b;
   end
       p=3.141592654;
       u1=rng(1);
       u2=rng(2);
       hlp=sqrt(-2.0*log(u1));
       rand(1)=hlp*cos(p*2.0*u2);
       rand(2)=hlp*sin(p*2.0*u2);
       rand(3)=xx;
```

Program 5.4. A Monte Carlo program for molecular levels

This Fortran program solves a stochastic chemical reacting system using a Monte Carlo method. In the Monte Carlo procedure, the process is checked at each time interval of length h to see if any reaction occurs. The number of sample paths calculated is specified as nrun. The mean, mean square, and standard deviation of the molecular levels at the final time are output. The parameters nx1, nx2, and nx3 are the molecular population levels for a reacting system that involves three different chemical species. The program is set up to solve the reacting problem described in Section 5.3.6. Output of the program is given following the program listing.

```
       real*8 xx
       xx=3404658.
c Monte Carlo program to calculate chemical molecular levels.
       nrun=5000
       x1av=0.0
       x2av=0.0
       x3av=0.0
       x1x1av=0.0
       x2x2av=0.0
       x3x3av=0.0
       do 400 nr=1,nrun
       am1=.02
       am2=.4
       am3=.001
       am4=.03
```

```
      nx1=100
      nx2=100
      nx3=100
      ntot=nx1+nx2+nx3
      nt=1000
      ant=nt
      time=1.0
      h=time/ant
      do 100 i=1,nt
      if(nx1.lt.0) nx1=0
      if(nx2.lt.0) nx2=0
      if(nx3.lt.0) nx3=0
      mx1=0
      mx2=0
      mx3=0
      anx1=nx1
      anx2=nx2
      anx3=nx3
      call random(xx,r)
      p1=am1*anx1*anx2*h
      if(r.gt.p1) goto 200
      mx2=mx2-1
      mx1=mx1-1
      mx3=mx3+1
200   continue
      call random(xx,r)
      p2=am2*anx3*h
      if(r.gt.p2) goto 300
      mx3=mx3-1
      mx1=mx1+1
      mx2=mx2+1
300   continue
      call random(xx,r)
      p3=am3*anx3*anx2*anx2*h/2.0
      if(r.gt.p3) goto 310
      mx3=mx3-1
      mx1=mx1+2
      mx2=mx2-2
310   continue
      call random(xx,r)
      p4=am4*anx1*anx1*h/2.0
      if(r.gt.p4) goto 320
      mx3=mx3+1
      mx1=mx1-2
      mx2=mx2+2
```

```
320  continue
     nx1=nx1+mx1
     nx2=nx2+mx2
     nx3=nx3+mx3
100  continue
     anx1=nx1
     anx2=nx2
     anx3=nx3
     anrun=nrun
     x1av=x1av+anx1/anrun
     x2av=x2av+anx2/anrun
     x3av=x3av+anx3/anrun
     x1x1av=x1x1av+anx1*anx1/anrun
     x2x2av=x2x2av+anx2*anx2/anrun
     x3x3av=x3x3av+anx3*anx3/anrun
400  continue
     x1sd=sqrt(x1x1av-x1av*x1av)
     x2sd=sqrt(x2x2av-x2av*x2av)
     x3sd=sqrt(x3x3av-x3av*x3av)
     write(6,550) nrun,h,time,am1,am2,am3,am4
550  format(2x,i5,6(2x,f8.3))
     write(6,560) x1av,x1x1av,x1sd
     write(6,560) x2av,x2x2av,x2sd
     write(6,560) x3av,x3x3av,x3sd
560  format(2x,3(2x,f12.2))
     stop
     end
     subroutine random(xx,r)
     real*8 xx,a,b,d
     a=16807.
     ib=2147483647
     b=ib
     id=a*xx/b
     d=id
     xx=a*xx-d*b
     r=xx/b
     return
     end
5000  0.001  1.000   0.020   0.400   0.001   0.030
79.21       6327.89          7.28
37.61       1448.60          5.84
131.19      17240.56         5.54
```

Program 5.5. Solving an SDE model for chemical molecular levels

This Fortran program solves a stochastic differential model of a chemical reacting system using Euler's method with nt equally spaced steps in time. The number of sample paths calculated is specified as nrun. The mean, mean square, and standard deviation in the solution at the final time are output. In addition, a MATLAB plotting program is produced in file sdechem.m. The program is set up to solve the stochastic differential equation system described in Section 5.3.6 and is used to produce Fig. 5.10. The parameters y1, y2, and y3 are the molecular population levels for a reacting system that involves three different chemical species. Output of the program is given following the program listing.

```
      real*4 rand(10),v(10,10),pp(10),y(10),y2(10)
      real*4 yy1(5001),yy2(5001),yy3(5001),f(10)
      real*4 av(10),av2(10),sd(10)
      real*8 xx
      open(unit=9,file='sdechem.m',status='unknown')
c A MATLAB plotting program is produced in file sdechem.m.
      xx=870686.
c This program solves for chemical molecular levels.
      time = 1.0
      nt=500
c nt is the number of time steps.
      ant=nt
      dt=time/ant
      nrun=5000
c nrun is the number of sample paths.
      anrun=nrun
      nchem=3
      am1=.02
      am2=.4
      am3=.001
      am4=.03
c Initial values are input here.
      do 100 j=1,nrun
      y(1)=100.
      y(2)=100.
      y(3)=100.
      yy1(1)=y(1)
      yy2(1)=y(2)
      yy3(i)=y(3)
      tt=0.0
      do 200 i = 1,nt
      tt=tt+dt
c The drift term is calculated here.
```

```
        hlp1=am1*y(1)*y(2)
        hlp2=am2*y(3)
        hlp3=am3*y(2)*y(2)*y(3)/2.0
        hlp4=am4*y(1)*y(1)/2.0
        f(1) = -hlp1+hlp2+2.0*hlp3-2.0*hlp4
        f(2) = -hlp1+hlp2-2.0*hlp3+2.0*hlp4
        f(3) = hlp1-hlp2-hlp3+hlp4
c The covariance matrix is calculated here.
        do 105 m=1,nchem
        do 105 n=1,nchem
  105   v(m,n)=0.0
        v(1,1) =  hlp1+hlp2+4.0*hlp3+4.0*hlp4
        v(1,2) =  hlp1+hlp2-4.0*hlp3-4.0*hlp4
        v(1,3) =  -hlp1-hlp2-2.0*hlp3-2.0*hlp4
        v(2,1) =  v(1,2)
        v(2,2) =  hlp1+hlp2+4.0*hlp3+4.0*hlp4
        v(2,3) =  -hlp1-hlp2+2.0*hlp3+2.0*hlp4
        v(3,1) =  v(1,3)
        v(3,2) =  v(2,3)
        v(3,3) =  hlp1+hlp2+hlp3+hlp4
        call random(xx,rand)
        call sqrtm(3,v,rand,pp)
        dtsq=sqrt(dt)
        do 87 k=1,nchem
  87    y(k) = y(k)+dt*f(k)+ pp(k)*dtsq
        yy1(i+1)=y(1)
        yy2(i+1)=y(2)
        yy3(i+1)=y(3)
  200   continue
        do 89 k=1,nchem
        av(k)=av(k)+y(k)/anrun
        av2(k)=av2(k)+y(k)*y(k)/anrun
  89    continue
  100   continue
        do 91 k=1,nchem
  91    sd(k)=sqrt(av2(k)-av(k)*av(k))
        write(6,780) nrun,nt,time,am1,am2,am3,am4
  780   format(2x,i5,2x,i5,6(2x,f8.2))
        do 795 k=1,3
        write(6,790) av(k),av2(k),sd(k)
  790   format(2x,5(2x,f10.2))
  795   continue
       nt2=nt+1
       write(9,370)
 370   format(' clear all')
```

```
      write(9,371)
  371 format(' close')
      write(9,330) dt
  330 format(' dt = ', f12.6)
      write(9,335) time
  335 format(' time = ',f12.6)
      write(9,170)
  170 format(' y1=[')
      do 171 i=1,nt2
  171 write(9,174) yy1(i)
  174 format(2x,f11.6)
      write(9,176)
  176 format(' ];')
      write(9,172)
  172 format(' y2=[')
      do 173 i=1,nt2
  173 write(9,174) yy2(i)
      write(9,176)
      write(9,182)
  182 format(' y3=[')
      do 183 i=1,nt2
  183 write(9,174) yy3(i)
      write(9,176)
      write(9,206) nt2
  206 format(' n = ', i6)
      write(9,207)
  207 format(' t=linspace(0,1,n)')
      write(9,208)
  208 format(' plot(t,y1,t,y2,t,y3)')
      stop
      end
      subroutine random(xx,rand)
      real*8 xx,a,b,d,rng(2)
      real*4 rand(10)
      a=16807.
      ib=2147483647
      b=ib
      do 56 j=1,5
      do 55 i=1,2
      id=a*xx/b
      d=id
      xx=a*xx-d*b
   55 rng(i)=xx/b
      pi=3.141592654
      u1=rng(1)
```

```
      u2=rng(2)
      hlp=sqrt(-2.0*alog(u1))
      rand(2*j-1)=hlp*cos(pi*2.0*u2)
      rand(2*j)=hlp*sin(pi*2.0*u2)
 56   continue
      return
      end
      subroutine sqrtm(m,b,c,p)
      real*4 b(10,10),c(10),p(10),y1(10),y2(10),s
      real*4 y3(10),y4(10),r(10,10),h,anh,s1,aa(10,10)
      real*4 a(10,10),f(10),x(10),t
c This subroutine calculates p=sqrt(b)c by solving an IVP.
      s=0.0
      do 110 i=1,m
      do 110 j=1,m
 110  s=s + b(i,j)*b(i,j)
      s=sqrt(s)
      do 115 i=1,m
      do 115 j=1,m
      a(i,j)=b(i,j)/(s*2.0)
      aa(i,j)=-a(i,j)
      if(i.eq.j) aa(i,j)=1.0+aa(i,j)
 115  continue
c The number of steps depends on integer nh > 1.
      nh=2
      do 120 i=1,m
      y1(i)=c(i)
      y2(i)=c(i)
      y3(i)=c(i)
      y4(i)=c(i)
 120  continue
      do 125 mh=1,4
      nh=nh*2
      anh=nh
      h=1.0/anh
      do 220 i=1,m
      y1(i)=y2(i)
      y2(i)=y3(i)
      y3(i)=y4(i)
 220  y4(i)=c(i)
      t=-h
      do 125 k=1,nh
      t=t+h
      do 135 i=1,m
      s1=0.0
```

```
      do 130 j=1,m
      r(i,j)=-aa(i,j)*t
      if(i.eq.j) r(i,j)=1.0+r(i,j)
      s1=s1-.5*aa(i,j)*y4(j)
130   continue
135   f(i)=s1
      call linsys(m,f,r,x)
      do 140 i=1,m
140   y4(i)=y4(i)+h*x(i)
125   continue
c A fourth-order extrapolation method is applied.
      do 150 i=1,m
      hlp=64.0*y4(i)-56.0*y3(i)+14.0*y2(i)-y1(i)
150   p(i)=hlp*sqrt(2.0*s)/21.0
      return
      end
      subroutine linsys(m,f,r,x)
c This subroutine solves the linear system rx=f.
      real*4 f(10),r(10,10),x(10),hlp,ff(10),rr(10,10)
c Gaussian elimination is used.
      do 235 i=1,m
      ff(i)=f(i)
      do 235 j=1,m
235   rr(i,j)=r(i,j)
      if(m.eq.1) goto 237
      do 205 k=1,m-1
      do 210 i=k+1,m
      hlp=rr(i,k)/rr(k,k)
      ff(i)=ff(i)-hlp*ff(k)
      do 210 j=k,m
      rr(i,j)=rr(i,j)-hlp*rr(k,j)
210   continue
205   continue
237   continue
      x(m)=ff(m)/rr(m,m)
      if(m.eq.1) goto 238
      do 215 kk=1,m-1
      k=m-kk
      hlp=0.0
      do 220 j=k+1,m
220   hlp=hlp+rr(k,j)*x(j)
      x(k)=(ff(k)-hlp)/rr(k,k)
215   continue
238   continue
      return
```

```
      end
5000   500     1.00     0.02     0.40     0.00     0.03
       79.31     6348.07        7.62
       37.44     1439.18        6.14
      131.17    17246.00        6.43
```

Program 5.6. A computer program to test goodness-of-fit

This Fortran program tests goodness-of-fit between a stochastic differential equation model and data. The method of Bak, Nielsen, and Madsen [22] is used. The program is set up to test two different stochastic differential equation models for the Aransas-Wood Buffalo whooping crane population data described in Section 4.9. The program calculates Q_M values for the two stochastic differential equation models. For this example, there are n = 47 values of the population size over the years 1939-1985 which are listed in Table 4.5 and graphed in Fig. 4.2. Output of the program is given following the program listing.

```
      real*4 x(50),xc(50,20),omega(50)
      integer ir(50)
      real*8 xx
c This code tests goodness of fit.
      n=47
c The method of Bak, Nielsen, and Madsen is used.
      data (x(i), i=1,47)/ 18, 22, 26, 16, 19, 21, 18, 22,
     * 25, 31, 30, 34, 31, 25, 21, 24, 21, 28, 24, 26, 32,
     * 33, 36, 39, 32, 33, 42, 44, 43, 48, 50, 56, 57, 59,
     * 51, 49, 49, 57, 69, 72, 75, 76, 78, 73, 73, 75, 86/
      do 999 icase=1,2
c Parameter icase =1 or 2 denotes SDE model 1 or 2.
      xx=102038.
      m=8
      h=1.0
      do 10 j=1,m+1
 10   omega(j)=0.0
      kk=4
      akk=kk
      h=h/akk
      do 202 i=2,n
      xs=x(i-1)
      xe=x(i)
      do 202 j=1,m
      xk=xs
      do 252 k=1,kk
      call functs(icase,xk,f,g)
```

```
      call random(xx,rand1,rand2)
252   xk=xk+h*f+sqrt(h)*g*rand1
      xc(i,j)=xk
202   continue
      do 402 i=2,n
      irr=1
      do 302 j=1,m
      xe=x(i)
      xcalc=xc(i,j)
      if(xe.gt.xcalc) irr=irr+1
302   continue
402   ir(i)=irr
      do 502 i=2,n
      irr=ir(i)
      omega(irr)=omega(irr)+1.0
502   continue
      chi2=0.0
      an=n
      am=m
      hlp=(an-1.0)/(am+1.0)
      do 602 j=1,m+1
602   chi2=chi2+(omega(j)-hlp)**2/hlp
      write(6,100) icase,chi2
100   format(5x,i7,5x,f9.2)
999   continue
      stop
      end
      subroutine functs(icase,x,f,g)
      th1=3510.0
      th2=13500.0
      f=th1/(x*x)
      g=th2/(x*x)
      if(icase.eq.1) goto 17
      th1=.0361
      th2=.6090
      f=th1*x
      g=sqrt(th2*x)
 17   continue
      return
      end
      subroutine random(xx,rand1,rand2)
      real*8 xx,a,b,d,rng(2)
      a=16807.
      ib=2147483647
      b=ib
```

```
      do 55 i=1,2
      id=a*xx/b
      d=id
      xx=a*xx-d*b
55    rng(i)=xx/b
      pi=3.141592654
      u1=rng(1)
      u2=rng(2)
      hlp=sqrt(-2.0*alog(u1))
      rand1=hlp*cos(pi*2.0*u2)
      rand2=hlp*sin(pi*2.0*u2)
      return
      end
    1         18.57
    2          4.09
```

References

1. Abukhaled, M.I., Allen, E.J.: A class of second-order Runge-Kutta methods for numerical solution of stochastic differential equations. Stochastic Analysis and Applications, **16**, 977–991 (1998)
2. Allen, E.J., Victory, H.D.: A computational investigation of the random particle method for numerical solution of the kinetic Vlasov-Poisson-Fokker-Planck equations. Physica A, **209**, 318–346 (1994)
3. Allen, E.J., Novosel, S.J., Zhang, Z.: Finite element and difference approximation of some linear stochastic partial differential equations. Stochastics and Stochastics Reports, **64**, 117–142 (1998)
4. Allen, E.J.: Stochastic differential equations and persistence time of two interacting populations. Dynamics of Continuous, Discrete, and Impulsive Systems, **5**, 271–281 (1999)
5. Allen, E.J., Baglama, J., Boyd, S.K.: Numerical approximation of the product of the square root of a matrix with a vector. Linear Algebra and its Applications, **310**, 167–181 (2000)
6. Allen, E.J.: Random selection of 3-digit numbers. Mathematical Spectrum, **33**, 8–10 (2000/2001)
7. Allen, E.J., Victory, H.D.: Modelling and simulation of a schistosomiasis infection with biological control. Acta Tropica, **87**, 251–267 (2003)
8. Allen, E.J., Allen, L.J.S., Schurz, H.: A comparison of persistence-time estimation for discrete and continuous population models that include demographic and environmental variability. Mathematical Biosciences, **196**, 14–38 (2005)
9. Allen, E.J., Allen, L.J.S., Arciniega, A., Greenwood, P.: Construction of equivalent stochastic differential equation models. manuscript, (2007)
10. Allen, L.J.S.: Some discrete-time SI, SIR, and SIS epidemic models. Mathematical Biosciences, **124**, 83–105 (1994)
11. Allen, L.J.S.: An Introduction To Stochastic Processes With Applications to Biology. Pearson Education Inc., Upper Saddle River, New Jersey (2003)
12. Allen, L.J.S., Allen, E.J.: A comparison of three different stochastic population models with regard to persistence time. Theoretical Population Biology, **68**, 439–449 (2003)
13. Allen, L.J.S., Kirupaharan, N., Wilson, S.M.: SIS epidemic models with multiple pathogen strains, Journal of Difference Equations and Applications, **10**, 53–75 (2004)

14. Allen, L.J.S.: Personal communication, (2006)
15. Allen, L.J.S.: An Introduction to Mathematical Biology. Pearson Prentice Hall, Upper Saddle River, New Jersey (2006)
16. Anderson, L.B.G., Rupert,B.-R.: The equity option volatility smile: an implicit finite-difference approach. The Journal of Computational Finance, **1**, 5–37 (1998)
17. Anderson, S.L.: Random number generation on vector supercomputers and other advanced architectures. SIAM Review, **32**, 221–251 (1990)
18. Arciniega, A., Allen, E.J.: Rounding error in numerical solution of stochastic differential equations. Stochastic Analysis and Applications, **21**, 281–300 (2003)
19. Arciniega, A., Allen, E.J.: Extrapolation of difference methods in option valuation. Applied Mathematics and Computation, **153**, 165–186 (2004)
20. Arnold, L.: Stochastic Differential Equations: Theory and Applications. John Wiley and Sons, New York (1974)
21. Bailey, N.T.J.: The Elements of Stochastic Processes With Applications to the Natural Sciences. John Wiley and Sons, New York (1964)
22. Bak, J.: Nonparametric Methods in Finance. MS Thesis, IMM-EKS-1998-34, Department of Mathematical Modelling, Technical University of Demark, Lyngby (1998)
23. Barrett, J.H., Bradley, J.S.: Ordinary Differential Equations. International Textbook Company, Scranton, Pennsylvania (1972)
24. Baxter, M., Rennie, A.: Financial Calculus, An Introduction to Derivative Pricing. Cambridge University Press, Cambridge (1996)
25. Bell, G.I., Glasstone S.: Nuclear Reactor Theory. Van Nostrand Reinhold Company, New York (1970)
26. Box, G.E.P., Muller, M.E.: A note on the generation of random normal deviates. The Annals of Mathematical Statistics, **29**, 610–611 (1958)
27. Boyce, M.S.: Time-series analysis and forecasting of the Aransas-Wood Buffalo whooping crane population. In: Lewis, J.C., Ziewitz, J.W. (eds) Proceedings of the 1985 Crane Workshop, Platte River Whooping Crane Habitat Maintenance Trust and USFWS, Grand Island, Nebraska (1987)
28. Burden, R.L., Faires, J.D., Reynolds, A.C.: Numerical Analysis. Prindle, Weber, Schmidt, Inc., Boston, Massachusetts (1978)
29. Capasso, V., Bakstein, D.: An Introduction to Continuous-Time Stochastic Processes: Theory, Models, and Applications to Finance, Biology, and Medicine. Birkhäuser, Boston (2005)
30. Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B.: An empirical comparison of alternative models of the short-term interest rate. Journal of Finance, **47**, 1209–1227 (1992)
31. Chang, C.C.: Random vortex methods for the Navier-Stokes equations. Journal of Computational Physics, **76**, 281–300 (1988)
32. Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G.: Alternative models for stock price dynamics. Journal of Econometrics, **116**, 225–257 (2003)
33. Cushing, J.M.: Integrodifferential Equations and Delay Models in Population Dynamics. Springer-Verlag, Berlin (1977)
34. Cvitanić, J., Goukasian, L., Zapatero, F.: Monte Carlo computation of optimal portfolios in complete markets. J. Econom. Dynam. Control, **27**, 971–986 (2003)
35. Davis, P.J., Rabinowitz, P.: Numerical Integration. Blaisdell Publishing Company, Waltham, Massachusetts (1967)

36. Dennis, B., Munholland, P.L., and Scott, J.M.: Estimation of growth and extinction parameters for endangered species. Ecological Monographs, **61**, 115–143 (1991)
37. Dupire, B. (ed): Monte Carlo Methodologies and Applications for Pricing and Risk Management. Risk Books, London (1998)
38. Edelstein-Keshet, L.: Mathematical Models in Biology. Random House, New York (1987)
39. Emmert, K.E., Allen, L.J.S.: Population extinction in deterministic and stochastic discrete-time epidemic models with periodic coefficients with applications to amphibians. Natural Resource Modeling, **19**, 117–164 (2006)
40. Friedman, A.: Stochastic Differential Equations and Applications, Volume 1. Academic Press, New York (1975)
41. Gard, T.C.: Introduction to Stochastic Differential Equations. Marcel Decker, New York (1987)
42. Gikhman, I.I., Skorokhod, A.V.: Stochastic Differential Equations. Springer-Verlag, New York (1972)
43. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics, **22**, 403–434 (1976)
44. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry, **81**, 2340–2361 (1977)
45. Gillespie, D.T.: Approximate accelerated stochastic simulation of chemically reacting systems. Journal of Chemical Physics, **115**, 1716–1733 (2001)
46. Golec, J., Sathananthan, S.: Sample path approximation of stochastic integro-differential equations. Stochastic Analysis and Applications, **17**, 579–588 (1999)
47. Golec, J., Sathananthan, S.: Strong approximations of stochastic integro-differential equations. Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms, **8**, 139–151 (2001)
48. Golec, J., Sathananthan, S.: Stability analysis of a stochastic logistic model. Mathematical and Computer Modelling, **38**, 585–593 (2003)
49. Gopalsamy, K.: Stability and Oscillations in Delay Differential Equations of Population Dynamics. Kluwer Academic Publishers Group, Dordrecht (1992)
50. Greenwood, P., Gordillo, L.F., and Kuske, R.: Autonomous stochastic resonance produces epidemic oscillations of fluctuating size. Proceedings of Prague Stochastics (2006)
51. Grimmet, G., Stirzaker, D.: Probability and Random Processes. Clarendon Press, Oxford (1982)
52. Hansen, G.E.: Assembly of fissionable material in the presence of a weak neutron source. Nuclear Science and Engineering, **8**, 709–719 (1960)
53. Harvey, C.R.: The real term structure and consumption growth. Journal of Financial Economics, **22**, 305–333 (1988)
54. Havlak, K.J., Victory, H.D.: The numerical analysis of random particle methods applied to Vlasov-Poisson-Fokker-Planck kinetic equations. SIAM Journal on Numerical Analysis, **33**, 291–317 (1996)
55. Hayes, J.G., Allen, E.J.: Stochastic point-kinetics equations in nuclear reactor dynamics. Annals of Nuclear Energy, **32**, 572–587 (2005)
56. Heath, D.C., Swindle, G. (eds): Introduction to Mathematical Finance. Proceedings of Symposia in Applied Mathematics, Volume 57. American Mathematical Society, Providence, Rhode Island (1997)

57. Hethcote, H.W., van den Driessche, P.: Two SIS epidemiologic models with delays. Journal of Mathematical Biology, **40**, 3–26 (2000)
58. Hetrick, D.L.: Dynamics of Nuclear Reactors. The University of Chicago Press, Chicago (1971)
59. Higham, D.J.: Mean-square and asymptotic stability of the stochastic theta method. SIAM J. Numer. Anal., **38**, 753–769 (2000)
60. Higham, D.J.: An algorithmic introduction to numerical simulation of stochastic differential equations. SIAM Review, **43**, 525–546 (2001)
61. Higham, N.J.: Newton's method for the matrix square root. Mathematics of Computation, **46**, 537–549 (1986)
62. Hogg, R.V., Craig, A.T.: An Introduction To Mathematical Statistics. Third Edition, The Macmillan Company, New York (1965)
63. Hsu, H.P.: Probability, Random Variables, and Random Processes. Schaum's Outline Series, McGraw-Hill, New York (1996)
64. Hurn, A.S., Lindsay, K.A., Martin, V.L.: On the efficacy of simulated maximum likelihood for estimating the parameters of stochastic differential equations. Journal of Time Series Analysis, **24**, 45–63 (2003)
65. Hutson, V., Pym, J.S.: Applications of Functional Analysis and Operator Theory. Academic Press, London (1980)
66. Ivanov, M.F., Shvets, V.F.: The method of stochastic differential equations for computing the kinetics of a plasma with collisions., U.S.S.R. Comput. Maths. Math. Phys., **20**, 146–155 (1980)
67. Jacob, J., Protter, P.: Probability Essentials. Second Edition, Springer-Verlag, Berlin (2003)
68. Kirupaharan, N., Allen, L.J.S.: Coexistence of multiple pathogen strains in stochastic epidemic models with density-dependent mortality. Bulletin of Mathematical Biology, **66**, 841–864 (2004)
69. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer-Verlag, New York (1992)
70. Kloeden, P.E., Platen, E., Schurz, H.: Numerical Solution of SDE Through Computer Experiments. Springer-Verlag, Berlin (1994)
71. Kloeden, P.E., Platen, E., Schurz, H., Sørenson, M.: On the effects of discretization on estimators of diffusions. J. Appl. Probab., **33**, 1061–1076 (1996)
72. Kloeden, P.E.: The systematic derivation of higher order numerical schemes for stochastic differential equations. Milan Journal of Mathematics, **70**, 187–207 (2002)
73. Kolmogorov, A.N., Fomin, S.V.: Introductory Real Analysis. Dover Publications, Inc., New York (1970)
74. Korn, R., Korn, E.: Option Pricing and Portfolio Optimization: Modern Methods in Financial Mathematics. Graduate Studies in Mathematics, Volume 31, American Mathematical Society, Providence, Rhode Island (2001)
75. Koskodan, R.: Extrapolation of an implicit method for numerical solution of stochastic differential equations and development of a stock-price model with application to multi-asset option pricing. Ph.D. Thesis, Texas Tech University, Lubbock, Texas (2006)
76. Koskodan, R., Allen, E.J.: Extrapolation of the stochastic theta numerical method for stochastic differential equations. Stochastic Analysis and Applications, **24**, 475–487 (2006)
77. Krifa, M.: Fiber length distribution in cotton processing: dominant features and interaction effects. Textile Res. J., **76**, 426–435 (2006)

78. Kurtz, T.G.: Limit theorems for sequences of jump Markov processes approximating ordinary differential equations. Journal of Applied Probability, **8**, 344–356 (1971)
79. Ladde, G.S., Laksmikantham, V.: Random Differential Inequalities. Academic Press, New York (1980)
80. Lakić, S., Petković, M.S.: On the matrix square root, Z. Angew. Math. Mech. (ZAMM), **78**, 173–182 (1998)
81. Langtangen, H.P.: Numerical Solution of First Passage Problems in Random Vibration. SIAM Journal of Scientific Computing, **15**, 977–996 (1994)
82. Leader, J.J.: Numerical Analysis and Scientific Computation. Pearson Education, Inc., Boston (2004)
83. Lötstedt, P., Persson, J., von Sydow, L., Tysk, J.: Space-time adaptive finite difference method for European multi-asset options. Technical Report 2004-055, Department of Information Technology, Uppsala Universitet (2005)
84. Madras, N.: Lectures on the Monte Carlo Method. American Mathematical Society, Providence, Rhode Island (2002)
85. McCabe B., Tremayne, A.: Elements of Modern Asymptotic Theory With Statistical Applications. Manchester University Press, Manchester (1992)
86. McCormack, R.K., Allen, L.J.S.: Disease emergence in deterministic and stochastic models for host and pathogen. Applied Mathematics and Computation, **168**, 1281–1305 (2005)
87. Melsa, J.L., Sage, A.P.: An Introduction to Probability and Stochastic Processes. Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1973)
88. Meyer, R., Almin, K.E., Steenberg, B.: Length Reduction of Fibres Subject to Breakage. Brit. J. Appl. Phys., **17**, 409–416 (1966)
89. Mikosch, T.: Elementary Stochastic Calculus with Finance in View. World Scientific, Singapore (1998)
90. Milstein, G.N., Tretyakov, M.V.: Stochastic Numerics for Mathematical Physics. Springer-Verlag, Berlin (2004)
91. Morton, K.W., Mayers, D.F.: Numerical Solution of Partial Differential Equations. Cambridge University Press, Cambridge (1994)
92. Øksendal, B.: Stochastic Differential Equations. Second Edition, Springer-Verlag, New York (1980)
93. Ortega, J.M.: Matrix Theory: A Second Course. Plenum Press, New York (1987)
94. Roberts, J.B.: First-passage probabilities for randomly excited systems: diffusion methods. Probabilistic Engineering Mechanics, **1**, 66–81 (1986)
95. Ross, S.M.: An Introduction to Mathematical Finance. Cambridge University Press, Cambridge (1999)
96. Rozanov, Y.A.: Introduction to Random Processes. Springer-Verlag, Berlin (1987)
97. Ryashko, L.B., Schurz, H.: Mean square stability analysis of some linear stochastic systems. Dynamic Systems and Applications, **6**, 165–190 (1997)
98. Schurz H.: Asymptotical mean square stability of an equilibrium point of some linear numerical solutions with multiplicative noise. Stochastic Analysis and Applications, **14**, 313–353 (1996)
99. Schurz, H.: Stability, Stationarity, and Boundedness of Some Implicit Numerical Methods for Stochastic Differential Equations and Applications. Logos-Verlag, Berlin (1997)

100. Schurz, H.: Numerical analysis of stochastic differential equations without tears. In: Kannan, D., Lakshmikantham, V. (eds) Handbook of Stochastic Analysis and Applications. Marcel-Dekker, New York (2002)
101. Schurz, H.: General theorems for numerical approximation of stochastic processes on the Hilbert space $H_2([0,T],\mu,\mathbb{R}^d)$. Electronic Transactions of Numerical Analysis, **16**, 50–69 (2003)
102. Schurz, H.: Stability of numerical methods for ordinary stochastic differential equations along Lyapanov-type and other functions with variable step sizes. Electronic Transactions on Numerical Analysis, **20**, 27–49 (2005)
103. Schurz, H.: On stochastic Liénard equations. Preprint m-05-008, Department of Mathematics, Southern Illinois University (2006)
104. Sharp, W.D., Allen, E.J.: Numerical solution of first passage time problems using an approximate Chapman-Kolmogorov relation. Probabilistic Engineering Mechanics, **13**, 233–241 (1998)
105. Sharp, W.D., Allen, E.J.: Stochastic neutron transport equations for rod and plane geometries. Annals of Nuclear Energy, **27**, 99–116 (2000)
106. Shiryaev, A.N.: Probability. Second Edition, Springer-Verlag, New York (1996)
107. Simsek, H.: Stochastic models for cotton fiber breakage, MS Thesis, Texas Tech University, Lubbock, Texas (2007)
108. Singleton, K.J.: Estimation of affine asset pricing models using the empirical characteristic function. Journal of Econometrics, **102**, 111–141 (2001)
109. Stanton, R.: A nonparametric model of term structure dynamics and the market price of interest rate risk. Journal of Finance, **52**, 1973–2002 (1997)
110. Talay, D.: Simulation and numerical analysis of stochastic differential equations. In: Kree, P., Wedig, W. (eds) Effective Stochastic Analysis. Springer-Verlag, New York (1988)
111. Talay, D., Tubaro, L.: Expansion of the global error for numerical schemes solving stochastic differential equations. Stochastic Analysis and Applications, **8**, 483–509 (1990)
112. Turelli, M.: A reexamination of stability in randomly varying environments with comments on the stochastic theory of limiting similarity. Theoretical Population Biology, **13**, 244–267 (1978)
113. Walsh, J.B.: An Introduction to Stochastic Partial Differential Equations. Lecture Notes in Mathematics 1180, Springer-Verlag, Berlin (1986)
114. Volkov, Yu.A., Polyudov, A.N.: A discrete model of the Fokker-Planck equation. U.S.S.R. Comput. Maths. Math. Phys., **27**, 71–80 (1987)

Basic Notation

ω	event, outcome of a random experiment
Ω	sample space, set of all possible outcomes
$\mathcal{A}$	set of events
P	probability measure
$(\Omega, \mathcal{A}, P)$	probability space
X or $X(\omega)$	random variable
$F_X(x)$	probability distribution of X, i.e., $P(\{\omega \in \Omega : X(\omega) \leq x\})$
$X \sim U[a, b]$	X is uniformly distributed on $[a, b]$
$X \sim N(\mu, \sigma^2)$	X is normally distributed with mean μ and variance σ^2
$E(X)$	Expectation of X
$\mathrm{Var}(X)$	Variance of X
$I_A(\omega)$	indicator function
(X, Y)	inner product of X with Y
H_{RV}	Hilbert space of random variables
$\|X\|_{RV}$	norm in H_{RV}, i.e., $(E(\|X\|^2))^{1/2}$
H_{SP}	Hilbert space of stochastic processes
$\|f\|_{SP}$	norm in H_{SP}, i.e., $\left(\int_0^T (E(\|f(t)\|^2)\, dt\right)^{1/2}$

$\mathbb{R}^n$	n-dimensional real space
$\mathbf{X}$	a vector, e.g., $[X_1, X_2, \ldots, X_n]^T$
$\Delta\mathbf{X}$	incremental change $[\Delta X_1, \Delta X_2, \ldots, \Delta X_n]^T$
w.p.1	with probability one
$p(t, x)$	probability density
$p(y, t, x, s)$	transition probability density
$f(t)$ or $f(t, \omega)$	stochastic process
$W(t)$ or $W(t, \omega)$	Wiener process
$J(f)$ or $J(f)(\omega)$	$\int_a^b f(s, \omega)\, ds$
$J(f)(t)$ or $J(f)(t, \omega)$	$\int_a^t f(s, \omega)\, ds$
$I(f)$ or $I(f)(\omega)$	$\int_a^b f(s, \omega)\, dW(s)$
$I(f)(t)$ or $I(f)(t, \omega)$	$\int_a^t f(s, \omega)\, dW(s)$
$V^{1/2}$	matrix square root, i.e. $V^{1/2}V^{1/2} = V$
MLE	Maximum Likelihood Estimation
SIS	Susceptible Infected Susceptible epidemic
SIR	Susceptible Infected Removed epidemic
MC	Monte Carlo
IVP	Initial Value Problem
SDE	Stochastic Differential Equation

Index

MATHEMATICAL MODELLING:
Theory and Applications

---

1. M. Křížek and P. Neittaanmäki: *Mathematical and Numerical Modelling in Electrical Engineering*. Theory and Applications. 1996
   ISBN 0-7923-4249-6
2. M.A. van Wyk and W.-H. Steeb: *Chaos in Electronics*. 1997
   ISBN 0-7923-4576-2
3. A. Halanay and J. Samuel: *Differential Equations, Discrete Systems and Control*. Economic Models. 1997
   ISBN 0-7923-4675-0
4. N. Meskens and M. Roubens (eds.): *Advances in Decision Analysis*. 1999
   ISBN 0-7923-5563-6
5. R.J.M.M. Does, K.C.B. Roes and A. Trip: *Statistical Process Control in Industry*. Implementation and Assurance of SPC. 1999
   ISBN 0-7923-5570-9
6. J. Caldwell and Y.M. Ram: *Mathematical Modelling*. Concepts and Case Studies. 1999
   ISBN 0-7923-5820-1
7. 1. R. Haber and L. Keviczky: *Nonlinear System Identification - Input-Output Modeling Approach*. Volume 1: Nonlinear System Parameter Identification. 1999
   ISBN 0-7923-5856-2; ISBN 0-7923-5858-9 Set
   2. R. Haber and L.Keviczky: *Nonlinear System Identification - Input-Output Modeling Approach*. Volume 2: Nonlinear System Structure Identification. 1999
   ISBN 0-7923-5857-0; ISBN 0-7923-5858-9 Set
8. M.C. Bustos, F. Concha, R. Bürger and E.M. Tory: *Sedimentation and Thickening*. Phenomenological Foundation and Mathematical Theory. 1999
   ISBN 0-7923-5960-7
9. A.P. Wierzbicki, M. Makowski and J. Wessels (eds.): *Model-Based Decision Support Methodology with Environmental Applications*. 2000
   ISBN 0-7923-6327-2
10. C. Rocşoreanu, A. Georgescu and N. Giurgiţeanu: *The FitzHugh-Nagumo Model*. Bifurcation and Dynamics. 2000
    ISBN 0-7923-6427-9
11. S. Aniţa: *Analysis and Control of Age-Dependent Population Dynamics*. 2000
    ISBN 0-7923-6639-5
12. S. Dominich: *Mathematical Foundations of Informal Retrieval*. 2001
    ISBN 0-7923-6861-4
13. H.A.K. Mastebroek and J.E. Vos (eds.): *Plausible Neural Networks for Biological Modelling*. 2001
    ISBN 0-7923-7192-5

MATHEMATICAL MODELLING:
Theory and Applications

14. A.K. Gupta and T. Varga: *An Introduction to Actuarial Mathematics*. 2002
ISBN 1-4020-0460-5

15. H. Sedaghat: *Nonlinear Difference Equations*. Theory with Applications to Social Science Models. 2003
ISBN 1-4020-1116-4

16. A. Slavova: *Cellular Neural Networks: Dynamics and Modelling*. 2003
ISBN 1-4020-1192-X

17. J.L. Bueso, J.Gómez-Torrecillas and A. Verschoren: *Algorithmic Methods in Non-Commutative Algebra*. Applications to Quantum Groups. 2003
ISBN 1-4020-1402-3

18. A. Swishchuk and J. Wu: *Evolution of Biological Systems in Random Media: Limit Theorems and Stability*. 2003
ISBN 1-4020-1554-2

19. K. van Montfort, J. Oud and A. Satorra (eds.): *Recent Developments on Structural Equation Models*. Theory and Applications. 2004
ISBN 1-4020-1957-2

20. M. Iglesias, B. Naudts, A. Verschoren and C. Vidal: *Foundations of Generic Optimization*. Volume 1: A Combinatorial Approach to Epistasis. 2005
ISBN 1-4020-3666-3

21. G. Marinoschi: *Functional Approach to Nonlinear Models of Water Flow in Soils*. 2006
ISBN 978-1-4020-4879-1

22. E. Allen: *Modeling with Itô Stochastic Differential Equations*. 2007
ISBN 978-1-4020-5952-0