Chapter 3

BASIC PROBABILITY THEORY

1. Overview

In this chapter, we briefly review some basic probability theory which we shall need later. This includes the background material for defining stochastic processes in the next chapter. In particular, we focus on multivariate normal distributions and conditional expectations, since most models for financial asset prices used in derivative pricing are conditionally heteroscedastic normal models. Familiarity with these two subjects is required for the remaining of the book.

To facilitate the understanding of new models, we shall illustrate them repeatedly on the following same example on interest rate:

**Interest Model**

Let \( h \) be a unit of time such as second, minute or hour measured in terms of year and let \( t = nh, n = 0, 1, \ldots, N \). For example, if the unit is a day, \( h = 1/365 \). By \( r_n \) we denote an interest rate at \( t = nh \). If \( h = 90/360 \), \( r_n \) can be used to indicate a 3 month LIBOR (London Interbank Offered Rate) traded in London Interbank market. Then a change of the interest rates from time \( n - 1 \) to time \( n \) can be described by

\[
\Delta r_n = r_n - r_{n-1}. \tag{1.1}
\]

Let \( N(\mu, \sigma^2) \) denote a normal (or Gaussian) distribution with mean \( \mu \) and variance \( \sigma^2 \) (see definition in (2.6)). Assume that

\[
\Delta r_n = a(b - r_{n-1})h + c\sqrt{h}\epsilon_n, \quad \epsilon_n \sim N(0, 1). \tag{1.2}
\]

The notation \( \epsilon_n \sim N(0, 1) \) means that the probability distribution of \( \epsilon_n \) is a normal distribution with mean 0 and variance 1. The term \( \epsilon_n \) is a random factor (variable) associated with time \( n \) and is often called an innovation.
model described in (1.2) is a time-discretized version of the well known Vasicek model for interest rates. In view of (1.1), (1.2) is equivalent to

\[ r_n = r_{n-1} + a(b - r_{n-1})h + c\sqrt{h}\epsilon_n, \quad \epsilon_n \sim N(0, 1). \]  

(1.3)

As \( r_{n-1} \) is realized at \((n - 1)\), \( r_n \) in the equation (1.3) is expressed by the right hand side of that equation as \( \{r_{n-1} + a(b - r_{n-1})h\} \) is realized at \((n - 1)\), and \( c\sqrt{h}\epsilon_n \) is newly realized at \( n \). The reason why \( \epsilon_n \) is multiplied by \( \sqrt{h} \) will become clear later. Let \( \eta_n = c\sqrt{h}\epsilon_n \). Then the distribution of \( \eta_n \) is normal whose mean is 0 and variance \( c^2h \), i.e., \( \eta_n = c\sqrt{h}\epsilon_n \sim N(0, c^2h) \). Hence the model (1.3) can be expressed as

\[ r_n = A + Br_{n-1} + \eta_n \quad A = abh, \quad B = 1 - ah, \quad \eta_n \sim N(0, c^2h). \]  

(1.4)

This is the so-called autoregressive model of order 1 in the analysis of time series. Note that both \( A \) and \( B \) depend on the time unit \( h \). If \( h \to 0 \), then \( A \to 0, B \to 1 \) and \( c^2h \) (the variance of \( \eta_n \)) tends to zero, which in turn implies \( r_n \to r_{n-1} \). The reason for multiplying \( \epsilon_n \) by \( \sqrt{h} \) is that it guarantees \( c^2h \to 0 \) as \( h \to 0 \), and thus \( r_n \to r_{n-1} \). If stated as the conditional distribution of \( r_n \) given \( r_{n-1} \), the model (1.4) can be expressed as

\[ r_n \text{ given } r_{n-1} \sim N(A + Br_{n-1}, c^2h). \]  

2. Conditional Distributions and Conditional Expectations

2.1 One-Dimensional (Univariate) Distributions

A random variable \( Z \) is said to be continuous if \( Z \) can assume any value on the real line. This notion of continuity has nothing to do with the continuity of time, and hence should not be confused with the notion of "continuous time models". Most of the random variables we deal with in this book are continuous. The probability that a continuous random variable \( Z \) falls into an interval \((a, b)\) is expressed as

\[ P(a < Z \leq b) = \int_a^b g(z)dz, \]  

(2.1)

where \( g(z) \) is a probability density function (pdf) which satisfies

\[ g(z) \geq 0 \quad \text{for all } z, \quad \text{and} \quad \int_{-\infty}^{\infty} g(z)dz = 1. \]  

(2.2)

The continuity of \( Z \) does not necessarily imply the continuity of \( g(z) \). However, for simplicity, we assume that \( g \) is continuous on the set \( \{z|g(z) > 0\} \). This set is called the support of \( Z \) or of \( g(\cdot) \). For a bounded interval \([a, b] \), we