Part I

Fundamentals

1

Mathematical Preparations for the Monte Carlo Method

1.1 INTRODUCTION AND OBJECTIVES

In this chapter we introduce a number of mathematical concepts and techniques. The main goal of this chapter is to motivate and define the *Ito integral*. Having laid this foundation we are then in a position to appreciate what stochastic differential equations (SDEs) are. We also give an introduction to random variables as well as the essentials of a number of concepts associated with stochastic analysis, for example sigma algebras, measure theory and probability spaces.

We assume that the reader has some knowledge of probability, statistics and set theory. For an introduction, see Hsu (1997), for example. This chapter is introductory in nature and is meant to make the book more self-contained than it would otherwise be.

1.2 RANDOM VARIABLES

We define what we mean by a random variable and why it is useful in the current context. Please skip to section 1.5 if you already know this material.

1.2.1 Fundamental concepts

In probability we define an *experiment* as something that refers to any process of observation. The result of the observation is called an *outcome* of the experiment. We speak of a *random experiment* when the outcome cannot be predicted. The set of all possible outcomes is called the *sample space* and an element of the sample space is called a *sample point* or an *elementary event*. An *event* is any subset of the sample space.

It is possible to examine the sample space in more detail. A sample space is called *discrete* if it consists of a finite number of sample points or if it has a *countably infinite* number of sample points (saying that a set is *countable* means that it can be put in one-to-one correspondence with the set of integers). Finally, a sample space is said to be *continuous* if the set of sample points forms a continuum.

Let us take a textbook example. To this end, we consider the random experiment of tossing a coin three times. For each throw of the coin, we know that there are two outcomes, namely heads ('H') or tails ('T'). Then, the sample space is given by the set of elementary events:

HHH, HHT, HTH, THH, HTT, THT, TTH, TTT

where we use the notation 'HHT' to mean a head on the first and second throws and a tail on the third throw. The above set represents the set of outcomes of the experiment. It is also possible to define experiments whose outcome is a subset of the set of all outcomes or even some new set whose elements belong to some given space. Some examples are:

- S1: Observe the number of heads in three tosses.
- S2: Observe the elementary events with an even number of heads.

In the first case, the sample space is the set 0, 1, 2, 3 while in the second case the sample space is the set HHT, HTH and THH.

1.2.2 Relative frequency and the beginnings of probability

We consider the case when a random experiment is repeated n times and when we are interested in determining how often a given event takes place, in some average sense. To be more precise, we let n become very large and we define the probability of event A by the limit:

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n} \tag{1.1}$$

and we refer to the quantity $n(A)/n$ as the *relative frequency* of event A, where $n(A)$ is the number of times event A occurs. We thus associate a real number with the events in a sample space and this concept is called the *probability measure*.

Since events are sets we can apply set theoretic operations – such as union, intersection and complementation – to them. Furthermore, we need some axiomatic properties of probability P for a finite sample space S:

- Axiom 1: $P(A) \geq 0$ where A is an event.
- Axiom 2: $P(S) = 1$.
- Axiom 3: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.

In the case of a sample space that is not finite we use:

- Axiom 3: If $A_1, A_2, \ldots$ is an infinite sequence of mutually exclusive events in S, that is $A_i \cap A_j = 0$ for $i \neq j$, then

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P\left(A_j\right).$$

Finally, we say that two events are *(statistically) independent* if and only if

$$P(A \cap B) = P(A)P(B) \tag{1.2}$$

1.2.3 Random variables

We need a way to compute the probabilities of certain events. To this end, we introduce the notion of a single *random variable*. The name can be confusing because it is not a variable in the strict sense but is rather a real-valued function whose domain is the sample space S and whose range is the real line (see Figure 1.1). In other words, a random variable $X(\zeta)$ (we abbreviate it to *r.v.*) assigns a real number (called the value of $X(\zeta)$) to each point ζ in S. Thus, its range is a subset of the set of all real numbers.

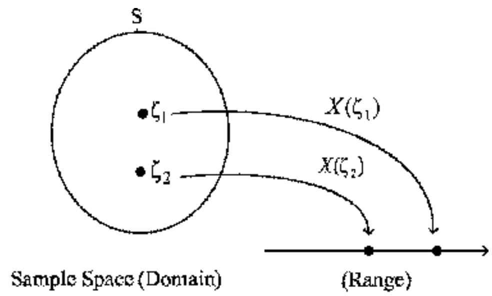

Figure 1.1 Random variable: domain and range

We note that a *r.v.* is *single-valued* because two range values cannot be assigned to the same sample point. Of course, two sample points might have the same value of $X(\zeta)$.

We can associate events and random variables as follows: let X be a *r.v.* and let x be a fixed real number, then we define the event $(X = x)$ as follows:

$$(X = x) = \{\zeta \in S : X(\zeta) = x\}$$

In the same way we can define events based on other operators, for example:

$$\begin{aligned}
&(X \leq x) = \{\zeta : X(\zeta) \leq x\} \\
&(X > x) = \{\zeta : X(\zeta) > x\} \\
&(x_1 < X \leq x_2) = \{\zeta : x_1 < X(\zeta) \leq x_2\}
\end{aligned}$$

The probabilities of these events are

$$\begin{aligned}
&P(X = x) = P\{\zeta \in S : X(\zeta) = x\} \\
&P(X \leq x) = P\{\zeta \in S : X(\zeta) \leq x\} \\
&P(X > x) = P\{\zeta \in S : X(\zeta) > x\} \\
&P(x_1 < X \leq x_2) = P\{\zeta \in S : x_1 < X(\zeta) \leq x_2\}
\end{aligned}$$

We take the example from section 1.2.1 in order to define some random variables. We recall that this is a problem of tossing a coin three times. The complete sample space consists of eight sample points (a set with n elements had 2^n subsets; this is called the *power set*) and we define X to be the $r.v.$ that gives two heads returned. First, we define the event A to be $X = 2$. Thus, using our new notation we see that (Hsu, 1997)

$$A = (X = 2) = \{\zeta \in S : X(\zeta) = 2\} = \{HHT, HTH, THH\}$$

In this case the sample points are equally likely (this means that they all have an equal probability of occurring), from which we can conclude that

$$P(X = 2) = P(A) = \frac{3}{8}$$

Finally, let us consider the event B that is defined by $X < 2$. Then

$$P(X < 2) = \{\zeta \in S : X(\zeta) < 2\} = \{HTT, THT, TTH, TTT\}$$

from which we can deduce that

$$P(X < 2) = \frac{4}{8} = \frac{1}{2}$$

We have now finished the example. An important function in the context of random variables is the *distribution function* (also known as the *cumulative distribution function* or *cdf*) of a *r.v.* defined by

$$F_X(x) = P(X \leq x) \tag{1.3}$$

Much information concerning a random experiment described by a random variable is determined by the behaviour of its cdf. Finally, we can determine probabilities from the cdf, again basing the analysis on equation (1.3). Some examples are

$$\begin{aligned}
&P(a < X \leq b) = F_X(b) - F_X(a)\\
&P(X > a) = 1 - F_X(a)\\
&P(X < b) = F_X(b^-), \text{ where } b^- = \lim_{0 < \epsilon \to 0} b - \epsilon
\end{aligned} \tag{1.4}$$

We now describe two major categories of random variables.

1.3 DISCRETE AND CONTINUOUS RANDOM VARIABLES

We classify random variables based on the degree of continuity of their corresponding cumulative distribution functions. If the distribution function of the $r.v.$ X is continuous and has a derivative at all points (with the possible exception of a finite number of points) then we say that X is a *continuous random variable*. However, if the cdf changes values only in jumps and is constant (or undefined) between jumps then we say that X is a *discrete random variable*.

In general, the range of a continuous $r.v.$ is a bounded or unbounded interval on the real line, while for a discrete $r.v.$ the range contains a finite or countably infinite number of points.

We now discuss two functions that are closely related to distribution functions of continuous and discrete random variables. These are called the *probability density function (pdf)* and *probability mass function (pmf)*, respectively. The pdf is the derivative of the cdf, namely:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{1.5}$$

Given the analytical form of the pdf, we can construct the cdf by integration, as follows:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(\xi)d\xi \tag{1.6}$$

Turning now to discrete random variables, let us assume that the cdf of the discrete $r.v.$ X occurs at a discrete set of points:

$$x_1, x_2, \ldots, \text{ with } x_i < x_j, \quad i < j$$

(this set may be finite or countable). Then

$$F_X(x_j) - F_X(x_{j-1}) = P(X \leq x_j) - P(X \leq x_{j-1}) = P(X = x_j)$$

We now define the probability mass function (pmf) as

$$p_x(x) = P(X = x) \tag{1.7}$$

Finally, we define the so-called *mean* and *variance* of discrete and continuous random variables by

$$\mu_X \equiv E(X) = \begin{cases} \Sigma_j x_j p_x(x_j), & X \text{ discrete} \\ \displaystyle\int_{-\infty}^{\infty} x f_X(x)dx, & X \text{ continuous} \end{cases} \tag{1.8}$$

and

$$\sigma_X^2 \equiv Var(X) = E\left\{[X - E(X)]^2\right\}$$

and hence:

$$\sigma_X^2 = \begin{cases} \Sigma_j (x_j - \mu_X)^2 p_x(x_j), & X \text{ discrete} \\ \displaystyle\int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx, & X \text{ continuous} \end{cases} \tag{1.9}$$

We give specific examples of discrete and continuous random variables. First, the *Bernoulli distribution* is a simple discrete distribution having two possible outcomes, namely *success* (with probability p) and *failure* (with probability $1 - p$) where $0 \le p \le 1$. Its pmf is given by

$$p_x(k) = P(X = k) = p^k(1 - p)^{1-k}, \quad k = 0, 1 \tag{1.10}$$

and its cdf is given by

$$F_X(x) = \begin{cases} 0, x < 0 \\ 1 - p, 0 \le x < 1 \\ 1, x \ge 1 \end{cases} \tag{1.11}$$

Finally, the mean and variance are given by

$$\begin{cases} \mu_X = E(x) = p \\ \sigma_X^2 = Var(X) = p(1 - p) \end{cases} \tag{1.12}$$

The Bernoulli distribution is used as a building block for more complicated discrete distributions; for example, individual repetitions of the *binomial distribution* are called *Bernoulli trials*.

Turning our attention to the continuous case, we examine the one-dimensional *uniform distribution* on the interval (a, b) whose pdf is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases} \tag{1.13}$$

and the corresponding cdf is:

$$F_X(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x \ge b \end{cases} \tag{1.14}$$

Finally, the mean and variance are given by

$$\begin{cases} \mu_X = E(x) = (a + b)/2 \\ \sigma_X^2 = Var(X) = (b - a)^2/12 \end{cases} \tag{1.15}$$

We shall encounter the uniform distribution in several places in later chapters of this book because it is used to generate random numbers. In particular, Chapter 22 is devoted to a discussion of discrete and continuous statistical distributions and algorithms for generating random numbers from these distributions. We have provided a C++ class hierarchy for discrete and continuous distributions on the CD.

We conclude this section with the remark that the Boost C++ library has extensive support for about 25 discrete and continuous statistical univariate distributions. A consequence is that developers can use the library in applications without having to create the code themselves. In the later chapters of this book we shall need a number of statistical distributions that the authors developed because the Boost library was not available when they commenced on their projects. We shall discuss Boost distributions in Chapter 3. For now, we give two examples of use to show how easy it is to use the library. We first examine the code in the Boost library that models the Bernoulli distribution. We create an instance of the corresponding class and we calculate some of its essential properties. The code implements the formulae in equations (1.10) and (1.11) and is given by

```
#include <boost/math/distributions/bernoulli.hpp>
//
// Don't forget to tell compiler which namespace
using namespace boost::math;

// Bernoulli distributions
bernoulli_distribution<> myBernoulli(0.4);
cout << "Probability of success: " << myBernoulli.
         success_fraction();

int k = 0;
cout << "pdf of Bernoulli: " << pdf(myBernoulli, k) << endl;
cout << "cdf of Bernoulli : " << cdf(myBernoulli, k) << endl
     << endl;

k = 1;
cout << "pdf of Bernoulli: " << pdf(myBernoulli, k) << endl;
cout << "cdf of Bernoulli : " << cdf(myBernoulli, k) << endl;
```

We now show how the uniform distribution is implemented in C++. In particular, we show the implementation of the formulae in equations (1.13) and (1.14):

```
#include <boost/math/distributions/uniform.hpp>
//

// Don't forget to tell compiler which namespace
using namespace boost::math;
uniform_distribution<> myUniform(0.0, 1.0); // Default type is
                                                'double'
cout << "Lower value: " << myUniform.lower() << ", upper value: "
                        << myUniform.upper() << endl;

// Choose another data type
uniform_distribution<float> myUniform2(0.0, 1.0);
cout << "Lower value: " << myUniform2.lower() << ", upper value: "
                        << myUniform2.upper() << endl;
```

```
// Distributional properties
double x = 0.25;

cout << "pdf of Uniform: " << pdf(myUniform, x) << endl;
cout << "cdf of Uniform: " << cdf(myUniform, x) << endl;
```

This code should give you an idea of how to use Boost for other distributions in your applications. More examples are to be found on the CD.

1.4 MULTIPLE RANDOM VARIABLES

In this section we discuss how to define two or more random variables on the same sample space. We discuss the case in which the range is n-dimensional Euclidean space. Thus, the random variable is a *vector-valued function* (see Figure 1.2). The mapping is formalised as follows:

$$X(\zeta) = (X_1(\zeta), X_2(\zeta), \ldots, X_n(\zeta)), \zeta \in S \tag{1.16}$$

In general terms, we can say that an n-dimensional random variable is an *n-tuple* of one-dimensional random variables. Analogous to the one-dimensional case we now define the generalisations of cdf, pdf and pmf. Let X be an n-variate *r.v.* on a sample space S. Then its *joint cdf* is the generalisation of equation (1.3) and is defined by

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \tag{1.17}$$

The *marginal cdf* is defined by setting one appropriate *r.v.* to ∞ in equation (1.17), for example:

$$F_{X_1,\ldots,X_{n-1}}(x_1, \ldots, x_{n-1}) = F_{X_1,\ldots,X_{n-1}X_n}(x_1, \ldots, x_{n-1}, \infty) \tag{1.18}$$

and

$$F_{X_1X_2}(x_1, x_2) = F_{X_1,\ldots,X_n}(x_1, x_2, \infty, \ldots, \infty)$$

Thus, a marginal cdf is similar to a slice of the joint cdf in the sense that it results in a lower-dimensional space. A continuous n-variate *r.v.* is described by a joint pdf defined by

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{\partial^n F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}{\partial x_1, \ldots, \partial x_n} \tag{1.19}$$

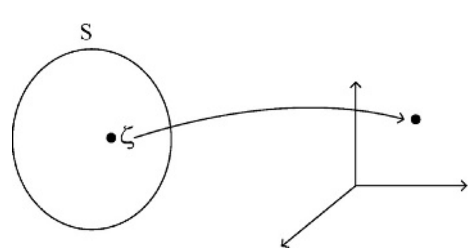

Figure 1.2 Multiple random variables

P1: JYS
c01 JWBK392-Duffy July 27, 2009 11:24 Printer: Yet to come

For a discrete n-variate random variable the joint pmf is given by

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 = x_1,\ldots,X_n = x_n)$$

This is the n-dimensional generalisation of equation (1.7).

We conclude this section by giving an example of an important random variable. This is the n-variate normal distribution. When $n = 2$ we speak of a *bivariate normal* or *Gaussian r.v.* whose joint pdf is given by

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)^{1/2}} \exp(-\frac{1}{2}\, q\, (x, y))$$

where

$$q(x, y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right] \tag{1.20}$$

and

μ_x = mean of X
μ_y = mean of Y
σ_x^2 = variance of X
σ_y^2 = variance of Y
ρ = correlation coefficient of X and Y

The random variables X and Y are independent when $\rho = 0$.

In Chapter 2 we shall discuss random variables in function space.

1.5 A SHORT HISTORY OF INTEGRATION

The theory of integration began more than 2000 years ago. The invention of the *integral calculus* can be attributed to Archimedes (287–212 BC) who used it to compute the areas under two-dimensional curves. But it was not until the 19th century that a rigorous theory was developed by Cauchy and Riemann. In particular, the *Riemann integral* (this is the integral we study at high school) is used in many applications. However, it does have its limitations and a number of mathematicians (most notably Borel and later Lebesgue) succeeded in defining an integral for a wider class of functions than was possible for the Riemann integral. In particular, Lebesgue introduced the concept of *measure* which is a generalisation of the length concept in classical integration theory (Carter and van Brunt, 2000; Rudin, 1970; Spiegel, 1969). Another generalisation by the Dutch mathematician Stieltjes was how to integrate one function with respect to another function. There are many other kinds of integral types, many of them being named after their inventors. A discussion of these topics is outside the scope of this book.

The above integration theories are not suitable for integrals that arise in connection with stochastic differential equations and a new kind of integral is needed. It was introduced by the Japanase mathematician Kiyoshi Ito (see Ito, 1944) and it is the standard approach to defining stochastic integrals in finance. In this chapter we motivate and define this integral because an understanding of it is important for the results and techniques in this book.

Before we embark on the mathematical details of the Ito integral we give a short description of how it came into existence. For a more detailed discussion we refer the reader to Bharucha-Reid (1972) and Tsokos and Padgett (1974). Norbert Wiener introduced the following integral in 1930:

$$\int_a^b g(t)\, dW(t) \tag{1.21}$$

Here $g(t)$ is a deterministic real-valued function and

$$\{W(\tau);\ \tau \in [a,b]\}$$

is a scalar Brownian motion process. Ito generalised this integral in 1944 (Ito, 1944) to support those cases where the integrand g is also a random function:

$$\int_0^t g(\tau;\omega)\, dW(\tau) \tag{1.22}$$

This is now the so-called Ito *stochastic integral*. In this case ω is a sample point from a given sample space. A more general case is the so-called *stochastic integral equation* defined by

$$x(t;\omega) = c + \int_0^t f(\tau, x(\tau;\omega))\, d\tau + \int_0^t g(\tau, x(\tau;\omega))\, dW(\tau) \tag{1.23}$$

where c is some constant, $t \in [0,1]$ and $\{W(t); t \in [0,1]\}$ is a Brownian motion process. Furthermore, the first integral in (1.23) is interpreted as a Lebesgue integral while the second integral is a stochastic integral that we shall presently define. For example, the method of successive approximations can be used to prove the existence and uniqueness of a random solution of equation (1.23), analogous to the way existence and uniqueness of deterministic Fredholm and Volterra integral equations are proved (see Tricomi, 1985). We discuss this topic in more detail in Chapter 2. In much of the literature we see the Ito equation written in differential form:

$$dX(t,\omega) = f(t,X)\, dt + g(t,X)\, dW(t) \tag{1.24}$$

with initial data $X(t_0;\omega) = c$. A remark: some authors denote the Wiener process as depending on two variables, namely time and its sample point dependence:

$$dW(t,\omega) \tag{1.25}$$

We now construct and define the Ito stochastic integral in the following sections. But first, it is necessary to introduce some measure theory and related concepts.

1.6 σ-ALGEBRAS, MEASURABLE SPACES AND MEASURABLE FUNCTIONS

We now introduce several major concepts:

- σ-algebras.
- Measurable space and measurable sets.
- Probability spaces.
- Toplogical spaces and measurable functions.

Let X be a set. Then a collection $\mathcal{F}$ of subsets of X is said to be a σ-algebra if $\mathcal{F}$ has the following properties (Rudin, 1970, page 8):

1. The *empty set* $\emptyset$ is a member of $\mathcal{F}$.
2. If $F \in \mathcal{F}$, then the complement F^c is also in $\mathcal{F}$, $F^c \equiv X - F$.
3. If $A = \cup_{n=1}^{\infty} A_n$ and if $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, then $A \in \mathcal{F}$.

In other words the σ-algebra is closed under complementation and countable unions of members of the σ-algebra. We note that X is also a member of $\mathcal{F}$ because of 1 and 2. In general, a given set can have many σ-algebras. Let us take a simple example. Define the set $X = \{1, 2, 3, 4\}$; we see that the set $\{\{1, 2\}, \{3, 4\}, \{1, 2, 3, 4\}\}$ is a σ-algebra because:

1. The empty set is a member.
2. The complement of each subset is a member.
3. The union of any number of subsets is a member.

We now discuss the concept of *measure*. In general terms, measure describes the size, area or volume of a set. It is a generalisation of the length of an interval or the area of a closed curve that we are accustomed to in Riemann integration. A measure on a set X is a function that assigns some real number to subsets of X. If $\mathcal{F}$ is a σ-algebra on X then X is called a *measurable space* and the members of $\mathcal{F}$ are called the *measurable sets* in X. We sometimes use the notation $(X, \mathcal{F})$ to denote the measurable space.

We define some specific functions that map measurable sets to the real numbers. An important case is the *probability measure* P defined on the sample space Ω by the mapping:

$$\begin{aligned}
&P : \mathcal{F} \to [0, 1] \\
&\text{such that} \\
&\text{(i)} \quad P(\emptyset) = 0, \quad P(\Omega) = 1 \\
&\text{(ii)} \quad \text{If } A_1, A_2, \ldots, \in \mathcal{F} \text{ and } A_i \cap A_j = \emptyset \\
&\text{for } i \neq j, \text{ then} \\
&P\left(\cup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \\
&\text{(iii)} \ P(A^c) = 1 - P(A), \text{ where } A^c = \Omega - A
\end{aligned} \tag{1.26}$$

We call the triple $(\Omega, \mathcal{F}, P)$ a *probability space*.

1.7 PROBABILITY SPACES AND STOCHASTIC PROCESSES

The concept of probability space was introduced by Kolmogorov and it is the foundation for probability theory. In order to align the notation with standard literature on stochastic differential equations we use the symbol Ω instead of the symbol X for a set. In particular, Ω will denote the sample space as already discussed. Then a probability space is a triple consisting of

- A sample space Ω.
- A σ-algebra $\mathcal{F}$ of subsets of Ω.
- A probability measure mapping $\mathcal{F}$ to the real numbers.

This triple is written as $(\Omega, \mathcal{F}, P)$

We now come to the important definition. A *stochastic process* (also called a random process) is a parametrised collection of random variables: $\{X(t); t \in T\}$ defined on a given probability space and indexed by the parameter t in the index set T. We assume that the variable t (sometimes called *time*) is always non-negative. In particular, we shall see in this book that the interval $[a, b]$ in which t is defined has the canonical form $[0, T]$ where the constant $T > 0$ represents the expiry time. Thus, the stochastic process is a function of t on the one hand and it depends on sample space on the other hand. We make this statement precise by noting that, for a fixed value of t we get a *random variable* defined by

$$\omega \to X_t(\omega), \quad \text{where } \omega \in \Omega \tag{1.27}$$

and then for a fixed $\omega \in \Omega$ is called a *path* of the stochastic process.

Intuitively, we consider t to be 'time' and each sample point ω to be a 'particle' in sample space. Thus, we can view the stochastic process X as a function of two variables:

$$X : [a, b] \times \Omega \to \mathbb{R}$$

In later chapters we shall encounter a number of examples of stochastic processes.

Finally, we introduce the last topics before we can construct the Ito integral. As before, let us assume that we have a probability space $(\Omega, \mathcal{F}, P)$; then we say that the function

$$\begin{aligned}&f : \Omega \to \mathbb{R}^n \text{ is } \mathcal{F}\text{-measurable if the set}\\ &f^{-1}(U) = \{\omega \in \Omega; f(\omega) \in U\} \in \mathcal{F} \text{ for all open sets } U \subset \mathbb{R}^n\end{aligned}$$

(an open set U is one where any point x in U can be changed by a small amount in any direction and still be inside U). A random variable X is an $\mathcal{F}$-measurable function $\Omega \to \mathbb{R}^n$.

Every random variable has a probability measure μ_X on $\mathbb{R}^n$ defined by If $\int_\Omega |X(\omega)| dP(\omega) < \infty$, then the number $E[X] \equiv \int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}^n} x d\mu_X(x)$ is called the expectation of X with respect to P.

1.8 THE ITO STOCHASTIC INTEGRAL

We are now in a position to motivate what it means when we speak of an Ito stochastic integral:

$$\int_a^b f(t, \omega) dW_t(\omega) \text{ or, equivalently} \int_a^b f(t, \omega) dW(t, \omega) \tag{1.28}$$

defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in $\mathbb{R}^n$. In general, we shall define this integral for a simple class of functions (sometimes called *step functions*) and we define the integral for arbitrary functions that are approximated by these step functions. This is similar to how the Lebesgue integral is motivated and we follow some of the steps from Øksendal (1998). First of all, we define the *indicator function* for a closed interval $[a, b]$ as follows:

$$\chi_{[a,b]}(t) = \begin{cases} 1, \text{ if } t \in [a, b] \\ 0, \text{ if } t \notin [a, b] \end{cases}$$

It is possible to define this function for open and half-open/half-closed intervals in a similar manner as above. We approximate the integral (1.28) by the following discrete sum:

$$I(f)(\omega) \equiv \sum_{n=0}^{N-1} f(t_n^*, \omega) \left\{ W_{t_{n+1}}(\omega) - W_{t_n}(\omega) \right\} \tag{1.29}$$

P1: JYS
c01 JWBK392-Duffy July 27, 2009 11:24 Printer: Yet to come

In contrast to the Riemann integral – where it does not make any difference where we choose the mesh-point argument for the function f – with stochastic integrals the following options have become the most popular:

$$\begin{array}{l}(a)\ t_n^* = t_n \Rightarrow \text{Ito integral}\\ (b)\ t_n^* = (t_n + t_{n+1})/2 \Rightarrow \text{Stratonovich integral}\end{array} \tag{1.30}$$

We summarise the construction of the integral. We define an elementary function as one that has the representation:

$$\phi(t,\omega) = \sum_{n=0}^{N-1} \alpha_n(\omega)\chi_{[t_n,t_{n+1}]}(t) \tag{1.31}$$

Then its stochastic integral is defined as follows:

$$\int_a^b \phi(t,\omega)dW = \Sigma_{n\geq 0}\alpha_n(\omega)[W_{t_{n+1}} - W_{t_n}](\omega) \tag{1.32}$$

where

$$t_n = t_n^{(N)} = \begin{cases} \frac{n}{2^N} \text{ if } a \leq \frac{n}{2^N} \leq b \\ a \text{ if } \frac{k}{2^N} < a \\ b \text{ if } \frac{k}{2^N} > b \end{cases}$$

and

$\{t_0, \ldots, t_{N-1}\}$ is a partition of the interval (a, b)

We conclude this section with the definition of the Ito integral and the main result is

Let f be some ‘smooth enough’ function on $[a, b]$

Then the Ito integral of f on $[a, b]$ is defined by

$$\int_a^b f(t,\omega)dW_t(\omega) = \lim_{n\to\infty}\int_a^b \phi_n(t,\omega)dW_t(\omega) \tag{1.33}$$

where

$\{\phi_n\}$ is a sequence of elementary functions such that

$$E\left[\int_a^b (f(t,\omega) - \phi_n(t,\omega)^2 dt\right] \to 0 \text{ as } n \to \infty$$

Theorem Let f be defined on $[a, b]$ and let $f_n(t,\omega)$ be a sequence of functions such that

$$E\left[\int_a^b (f_n(t,\omega) - f(t,\omega)^2 dt\right] \to 0 \text{ as } n \to \infty$$

Then

$$\int_a^b f(t,\omega)dW_t(\omega) = \lim_{n\to\infty}\int_a^b f_n(t,\omega)dW_t(\omega) \text{ in the } L^2 \text{ sense}$$

1.9 APPLICATIONS OF THE LEBESGUE THEORY

In this section we discuss a number of theorems and results based on the Lebesgue integration theory. They are important because we can use them in proving existence and uniqueness results for many kinds of equations in particular when we study stochastic equations and their applications to finance. Furthermore, we give some examples on how to apply these results. For an introduction, we refer the reader to Rudin (1964), Rudin (1970) and (the more introductory level) Spiegel (1969).

1.9.1 Convergence theorems

Lemma 1.1 (Fatou's Lemma). Let $\{f_n\}$ be a sequence of non-negative measurable functions defined on a set X and suppose that

$$\lim_{n\to\infty} f_n(x) = f(x) \ \forall\ x \in X$$

Then

$$\liminf_{n\to\infty} \int_X f_n(x)dx \geq \int_X f\,dx$$

where lim inf of a sequence $\{a_n\}$ is defined as the number a such that infinitely many terms of the sequence are less than $a+\epsilon$ while only a finite number of terms are less than $a-\epsilon$ for $\epsilon > 0$.

We note that this theorem also holds when $\lim_{n\to\infty} f_n(x) = f(x)$ almost everywhere (that is, everywhere except on a set of measure zero).

Theorem 1.1 (Monotone Convergence Theorem, MCT). Let $\{f_n\}$ be a sequence of measurable functions in a set X and suppose that

$$\begin{aligned}&\text{(a) } 0 \leq f_1(x) \leq f_2(x) \leq \ldots \leq \infty \ \forall\ x \in X\\ &\text{(b) } \lim_{n\to\infty} f_n(x) = f(x), \quad \forall\ x \in X\end{aligned}$$

Then f is measurable and we have the following convergence result:

$$\lim_{n\to\infty} \int_X f_n\,dx = \int_X f\,dx$$

This is also called the *Beppo Levi theorem*. This result allows us to exchange integration and limits. We now wish to prove similar results under milder and different assumptions concerning the integrands.

Theorem 1.2 (Bounded Convergence). Let $\{f_n\}$ be a sequence of measurable functions on a set X such that $\lim_{n\to\infty} f_n(x) = f(x)$. If the sequence is uniformly bounded, that is

$$\exists\ M > 0 \text{ such that } |f_n(x)| \leq M,\ \forall\ n \geq 0$$

then we have

$$\lim_{n\to\infty} \int_X f_n(x)dx = \int_X \lim_{n\to\infty} f_n(x)dx = \int_X f(x)dx$$

Theorem 1.3 (Dominated Convergence Theorem, DCT). Let $\{f_n\}$ be a sequence of measurable functions on a set X such that

$$\lim_{n\to\infty} f_n(x) = f(x)$$

Then if there exists an integrable function $M(x)$ such that $|f_n(x)| \leq M(x)\ \forall\ n = 0, 1, 2, \ldots$ then we have:

$$\lim_{n\to\infty} \int_X f_n(x)dx = \int_X \lim_{n\to\infty} f_n(x)dx = \int_X f(x)dx$$

We now state the variant of Theorem 3 for series of functions.

Theorem 1.4 (Dominated Convergence for Infinite Series). Let $\{f_n\}$ be a sequence of measurable functions on a set X. If there exists an integrable function $M(x)$ such that $|s_m(x)| \leq M(x)$ where $s_m(x) = \sum_{n=1}^m f_n(x)$ and $\lim_{m=\infty} s_m(x) = s(x)$, then

$$\int_X \sum_{n=1}^{\infty} f_n(x)dx = \sum_{n=1}^{\infty} \int_X f_n(x)dx$$

We take an example to show how these theorems can be applied to calculating integrals. The first example will be calculated using the infinite series variant of the Dominated Convergence Theorem. The integral to be computed is

$$I \equiv \int_0^{\infty} \frac{e^{-x}}{1 - e^{-x}} x^{\alpha-1} dx, \quad \alpha > 1$$

Using the fact that

$$\frac{e^{-x}}{1 - e^{-x}} = \sum_{n=1}^{\infty} e^{-nx}$$

and that each partial sum of this series is bounded we get

$$I = \int_0^{\infty} \sum_{n=1}^{\infty} e^{-nx} x^{\alpha-1} dx = \sum_{n=1}^{\infty} \int_0^{\infty} e^{-nx} x^{\alpha-1} dx$$

Now using the change of variables

$$y = nx$$

we get

$$I = \sum_{n=1}^{\infty} \int_0^{\infty} e^{-y} y^{\alpha-1} n^{-\alpha} dy = \sum_{n=1}^{\infty} n^{-\alpha} \Gamma(\alpha)$$

where we express the integral in terms of the *gamma function*:

$$\Gamma(\alpha) \equiv \int_0^{\infty} e^{-y} y^{\alpha-1} dy$$

Finally, we can write the integral in the form:

$$I = \Gamma(\alpha) \sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}$$

The second example shows how to use the theorems that exchange integration and limits. Define:

$$f_n(x) = \frac{n}{x+n}, \quad x \in [0,1], \quad n \geq 1$$

We use the following properties of the sequence of functions:

$$0 \leq f_n(x) \leq 1 \text{ (DCT)}$$
$$0 \leq f_n(x) \leq f_{n+1}(x) \text{ (MCT)}$$

Then using either MCT or DCT we can calculate the integral as follows:

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx = \int_0^1 \lim f_n(x)dx = \int_0^1 dx = 1$$

1.9.2 Fubini's theorem

We now discuss the Lebesgue integral in the plane and in particular the *Lebesgue double* integral. The technique is based on *iterated integrals*.

Theorem 1.5 (Fubini's Theorem). Let $f(x, y)$ be a measurable function in the set

$$R = \{(x, y) : a \leq x \leq b, \quad c \leq y \leq d\}$$

Then

$$\int\int_R f(x,y)dxdy = \int_c^d \left\{ \int_a^b f(x,y)dx \right\} dy = \int_a^b \left\{ \int_c^d f(x,y)dy \right\} dx$$

Basically, this theorem allows us to compute a two-dimensional integral by an iteration of two one-dimensional integrals.

Finally, we note that Fubini's theorem can be applied to two-dimensional stochastic integration in which one integral is a Lebesgue integral and the other an Ito integral (for example, Karatzas and Shreve 1991).

1.10 SOME USEFUL INEQUALITIES

We introduce a number of useful results that relate to inequalities between functions. They are useful in many kinds of applications.

The first inequality is named after Gronwall and it is used to prove the stability of solutions to differential, integral and integro-differential equations, for example.

Theorem 1.6 (Gronwall's inequality, continuous case). Let u and β be continuous functions on $[a, b]$ such that $\beta(t) \geq 0 \ \forall \ t \in [a, b]$. Assume the inequality:

$$u(t) \leq K + \int_a^t \beta(s)u(s)ds \text{ where } K \text{ is a constant}$$

Then $u(t) \leq K \exp\left(\int_a^t \beta(s)ds\right), t \in [a, b]$.

The following inequality is the discrete analogue of Theorem 1.6. It is used to prove stability of finite difference schemes and their accuracy.

Theorem 1.7 (Gronwall's inequality, discrete case). Let $a_n,\ n = 0, \ldots, N$ be a sequence of real numbers such that $|a_n| \leq K + hM \sum_{j=0}^{n-1} |a_n|, n = 1, \ldots, N$ where K and M are positive constants, then $|a_n| \leq (hM|a_0| + K)\exp(Mnh), n = 1, \ldots, N$.

A generalisation of this discrete theorem is

$$\text{Let } |a_n| \leq K + \sum_{r=0}^{n-1} g(r)|a_r| \text{ with } g(r) \geq 0,\ \text{then } |a_n| \leq K \exp\left(\sum_{r=0}^{n-1} g(r)\right)$$

We now discuss convex and concave functions. A real function f defined on the interval $[a, b]$ where $-\infty \leq a < b \leq \infty$ is called *convex* if

$$\begin{aligned}&f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)\\&\text{for } x, y \in (a, b) \text{ and } 0 \leq \lambda \leq 1\end{aligned}$$

A function f is said to be *concave* if the function $F = -f$ is convex, in other words:

$$f((1-\lambda)x + \lambda y) \geq (1-\lambda)f(x) + \lambda f(y)$$

We now introduce Jensen's inequality that relates the integral of a convex function to the convex function of an integral.

Theorem 1.8 (Jensen's Inequality) (Rudin, 1970). Let μ be a positive measure on a σ-algebra in Ω such that $\mu(\Omega) = 1$. If $f \in L^1(\mu)$ and $a < f(x) < b\ \forall x \in \Omega$ then if φ is convex on (a, b), we have the inequality:

$$\varphi\left(\int_\Omega f d\mu\right) \leq \int_\Omega (\varphi \circ f) d\mu$$

Jensen's inequality has the following *finite form*: let $\left\{a_j\right\}_{j=1}^n > 0$ and $\left\{x_j\right\}_{j=1}^n$ be two sequences and let φ be a real convex function. Then

$$\varphi\left(\frac{\sum_{j=1}^n a_j x_j}{\sum_{j=1}^n a_j}\right) \leq \frac{\sum_{j=1}^n a_j \varphi(x_j)}{\sum_{j=1}^n a_j}$$

1.11 SOME SPECIAL FUNCTIONS

We introduce a class of functions that is important in stochastic analysis; these are functions whose powers are Lebesgue integrable on some set in n-dimensional space (Adams, 1975). We motivate this class of functions by examining real-valued functions of a single variable.

Let $p \geq 1$ be an integer. The space of all functions for which

$$\int_a^b |f(x)|^p dx < \infty$$

is denoted by $L^p[a, b]$, or L^p for short. In words, these are functions whose powers are Lebesgue integrable on the interval $[a, b]$. A special case is when $p = 2$; we then speak of the class of *square-integrable* functions, thus

$$L^2[a, b] = \left\{ f : \int_a^b |f(x)|^2 dx < \infty \right\}$$

Another important case is when $p = \infty$. We shall discuss both of these cases in the context of convergence.

The p-integrable function class becomes a *metric space* (we discuss this in more detail in Chapter 2) and the corresponding *norm* is defined by

$$d(f, g) \equiv \|f - g\|_p = \left\{ \int_a^b |f(x) - g(x)|^p \right\}^{1/p}$$

For example, the *triangle inequality* holds for this space:

$$\|f - g\|_p \leq \|f - h\|_p + \|h - g\|_p \text{ where } f, g, h \in L^p[a, b]$$

We now state some useful inequalities (Rudin, 1970):

- Hölder's inequality:

$$\left| \int_a^b f(x)g(x)dx \right| \leq \|f\|_p \|g\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$$

- Schwarz's inequality (special case of Hölder's inequality with $p = q = 2$):

$$\left| \int_a^b f(x)g(x)dx \right| \leq \left\{ \int_a^b |f(x)|^2 dx \right\}^{1/2} \left\{ \int_a^b |g(x)|^2 \right\}^{1/2} = \|f\|_2 \|g\|_2$$

- Minkowski's inequality:

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p, \quad 1 \leq p \leq \infty$$

These inequalities and the above theory generalise to general sets and measures.

We take an example of proving that a function's power is integrable. We claim:

$$f \in L^1[0, 8] \text{ where } f(x) = \tfrac{1}{\sqrt[3]{x}}$$

To this end, define the function:

$$[f(x)]_p = \begin{cases} \frac{1}{\sqrt[3]{x}}, & \frac{1}{\sqrt[3]{x}} \leq p \text{ or equivalently } x \geq \frac{1}{p^3} \\ p, & \frac{1}{\sqrt[3]{x}} > p \text{ or equivalently } x < \frac{1}{p^3} \end{cases}$$

Then the following sequence of steps leads to the desired conclusion:

$$\begin{aligned}\int_0^8 \frac{dx}{\sqrt[3]{x}} &= \lim_{p\to\infty}\int_0^8 [f(x)]_p dx \\ &= \lim_{p\to\infty}\left[\int_0^{1/p^3} p dx + \int_{1/p^3}^8 \frac{dx}{\sqrt[3]{x}}\right] \\ &= \lim_{p\to\infty}\left[[px]_0^{1/p^3} + \left[\frac{3}{2}x^{2/3}\right]_{1/p^3}^8\right] \\ &= \lim_{p\to\infty}\left[\frac{1}{p^2} + 6 - \frac{3}{2p^2}\right] = 6\end{aligned}$$

We are finished. We note that this integral exists as an *improper Riemann integral*:

$$\lim_{\epsilon\to 0}\int_\epsilon^8 \frac{dx}{\sqrt[3]{x}} = \lim_{\epsilon\to 0}\left[\frac{3}{2}x^{2/3}\right]_\epsilon^8 = 6$$

In this case the Riemann and Lebesgue integrals have the same value.
We now discuss the relationship between L^p spaces and stochastic theory.

1.12 CONVERGENCE OF FUNCTION SEQUENCES

A sequence of functions $\{f_n(x)\}$ belonging to $L^p[a, b]$ is said to be a *Cauchy sequence* if

$$\lim_{\substack{m\to\infty\\ n\to\infty}}\int_a^b |f_m(x) - f_n(x)|^p dx = 0$$

or, if given $\epsilon > 0$, there exists a number $n_0 > 0$ such that

$$\|f_m - f_n\|_p \equiv \left(\int_a^b |f_m(x) - f_n(x)|^p dx\right)^{1/p} < \epsilon$$

when $m > n_0, n > n_0$

Furthermore, if there exists a function $f(x)$ in L^p such that

$$\lim_{n\to\infty}\int_a^b |f_n(x) - f(x)|^p dx = 0$$

we then say that $\{f_n(x)\}$ converges in the mean (or is mean convergent) to $f(x)$ in the space L^p. We write this property in the form:

$$\lim_{n\to\infty} f_n(x) = f(x)$$

We state that the space L^p is complete if every Cauchy sequence in the space converges in the mean to a function in the space.

Theorem 1.9 (Riesz-Fischer). Any L^p space is complete. Thus, if $\{f_n(x)\}$ is a sequence of functions in L^p and

$$\lim_{\substack{m\to\infty\\ n\to\infty}}\int_a^b |f_m(x) - f_n(x)|^p dx = 0$$

then there exists a function $f \in L^p$ such that f_n converges to f in the mean, that is $\lim_{n\to\infty} \int_a^b |f_n(x) - f(x)|^p dx < \infty$.

Now let $\{f_n(x)\}$ be a sequence of measurable functions defined almost everywhere. Then $\{f_n(x)\}$ converges in measure to $f(x)$ if $\lim_{n\to\infty} \mu \{x : |f_n(x) - f(x)| \geq \sigma\} = 0 \; \forall\, \sigma > 0$.

Theorem 1.10 (Convergence of a sequence of functions).

- If a sequence converges almost everywhere, then it converges in measure.
- If a sequence converges in the mean, then it converges in measure.
- If a sequence converges in measure, then there exists a subsequence that converges almost everywhere.

1.13 APPLICATIONS TO STOCHASTIC ANALYSIS

The above discussion leads naturally to the definition of some results in stochastic analysis. In this case we are interested in probability measures.

Let (Ω, A, P) be a probability measure space, that is

$\Omega =$ non-empty abstract set.
$\mathcal{F} = \sigma$-algebra of subsets of Ω.
$P =$ complete probability measure on A.

We denote the space $L^p(\Omega, \mathcal{F}, P)$, $1 \leq p < \infty$ as the set of measurable functions defined on $\mathbb{R}^1_+ = \{x \in \mathbb{R}^1; x > 0\}$ such that the inequality $\forall\, t \in \mathbb{R}^1_+$ is satisfied and we have $\int_\Omega |x(t;\omega)|^p dP(\omega) < +\infty$.

The norm in this space is

$$\|x(t;\omega)\|_p = \left\{ \int_\Omega |x(t;\omega)|^p dP(\omega) \right\}^{1/p} < \infty$$

If $x(t;\omega)$ is a vector-valued function with n components, then the norm becomes

$$\|x(t;\omega)\| = \left(\Sigma_{j=1}^n \|x_j(t;\omega)\|^2 \right)^{1/2}$$

Let X be a continuous random variable; then its *mean value* or *expected value* is given by:

$$E(X) = \int_{\mathbb{R}} xp(x)dx$$

where $p(x)$ is its probability density function.

We now define what we mean by convergence in this context; to this end, let X_1, $X_2, \ldots, X_n, \ldots$ be a sequence of random variables. We are interested in asymptotic behaviour of this sequence. In other words, does a random variable X that is the limit of X_n as $n \to \infty$ converge in some sense? The answer depends on the norm that we use (Kloeden, Platen and Schurz, 1997, page 27):

- *Mean-square convergence*:

  $$E(X_n^2) < \infty, \quad n = 1, 2, \ldots, \quad E(X^2) < \infty$$

  and

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

- *Strong convergence*:

  $$E(|X_n|) < \infty, \quad n = 1, 2, \ldots \text{ with } E(|X|) < \infty$$

  and

  $$\lim_{n\to\infty} E(|X_n - X|) = 0$$

  This is a generalisation of deterministic convergence; in particular, it is used when investigating convergence of numerical schemes.
- *Weak convergence*:

  $$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)dF_{X_n}(x) = \int_{\mathbb{R}} f(x)dF(x)$$

  for all *test functions* $f(x)$ with compact support.
- *Convergence in distribution*:

  $$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ for all points of } F_X \text{ where it it continuous}$$

- *Convergence with probability one* (w.p.1):

  $$P\left(\left\{\omega \in \Omega : \lim_{n\to\infty} |X_n(\omega) - X(\omega)|\right\} = 0\right) = 1$$

  This is called *almost sure convergence*.

1.14 SUMMARY AND CONCLUSIONS

We have given an account of a number of important mathematical concepts that underpin the topics in succeeding chapters. We have attempted to be as complete as possible while at the same time avoiding those topics that are not absolutely essential to a full understanding of the material.

1.15 EXERCISES AND PROJECTS

1. (**) Find the sample spaces corresponding to the following random experiments:
   - Toss a die and observe the number that shows on top.
   - Toss a coin four times and observe the total number of heads obtained.
   - Toss a coin four times and observe the sequence of heads and tails observed.
   - A transistor is placed in an oven at high temperature and its operational lifetime is observed (this process is called *destructive testing*).
   - A launched missile's position is observed at *n* points in time. At each point the missile's height above the ground is recorded.
2. (**) Consider the experiment of tossing two dice. Answer the following questions:
   - What is the sample space?

– Find the event that the sum of dots on the two dice equals 7.
– Find the event that the sum of dots on the two dice is greater than 10.
– Find the event that the sum of dots on the two dice is greater than 12.

3. (***) Prove that the number of events in a sample space S with n elementary events is 2^n (this set is sometimes called the *power set* of S). Design an algorithm to compute the power set in C++.
4. (*) Based on the definition of the cumulative distribution function in equation (1.3) prove or deduce the following properties:

$$\begin{aligned}
&\text{(a) } 0 \le F_X(x) \le 1\\
&\text{(b) } F_X(x_1) \le F_X(x_2), \quad x_1 \le x_2\\
&\text{(c) } \lim_{x\to\infty} F_X(x) = 1\\
&\text{(d) } \lim_{x\to-\infty} F_X(x) = 0\\
&\text{(e) } \lim_{x\to a^+} F_X(x) = F_X(a^+) = F_X(a) \text{ where } a^+ = \lim_{0\,<\,\epsilon\to\,0} a+\epsilon
\end{aligned}$$

   Remark: property (b) means that the cdf is a nondecreasing function while property (e) states that it is *continuous from the right*. Property (a) is obvious because the cdf is a probability function.
5. (**) A random variable X is called an *exponential r.v.* with parameter $\lambda > 0$ if its pdf is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases}$$

   Find the cdf as well as the mean and variance for this distribution.
6. (**) A random variable X is called a *Poisson r.v.* with parameter $\lambda > 0$ if its pmf is given by

$$p_X(k) = P(x = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

   Find the cdf as well as the mean and variance for this distribution. This distribution is important because it is a *building block* for stochastic processes with jumps. We shall discuss this topic in more detail in later chapters.
7. (***) The *covariance* of two random variables X and Y is defined as

$$Cov(X, Y) = \sigma_{XY} = E(XY) - E(X)E(Y)$$

   The *correlation coefficient* is defined as

$$\rho(X, Y) = \rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

   where σ and σ are the standard deviations of X and Y, respectively. Prove that

$$|\rho_{XY}| \le 1 \text{ or } -1 \le \rho_{XY} \le 1$$

8. (**) Let X be a set. Prove that the following are σ-algebras over X:

– The *trivial σ-algebra* consisting of X and the empty set.
– The *power set* of X (this is the set of all subsets).
– The collection of all subsets of X that are countable or whose complements are countable.

9. (***) Use *De Morgan's laws* to prove that a σ-algebra is closed under countable intersections. The laws are

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

where A^c is the complement of A.

10. (**) Define the function

$$g_n(x) \equiv ae^{-nax} - be^{-nbx} \quad (0 < a < b),\ n = 0, 1, \ldots$$

Is the sum of the integral equal to the integral of the sum?

$$\sum_{n=0}^{\infty} \int_{\mathbb{R}} g_n(x)dx = \int_{\mathbb{R}} \sum_{n=0}^{\infty} g_n(x)dx$$

11. (**) Prove (by any means) that

$$\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n x^{\alpha-1}dx = \int_0^{\infty} e^{-x}x^{\alpha-1}dx \ (\alpha > 1)$$

(We know that $\lim_{n\to\infty}(1 - \frac{x}{n})^n = e^{-x}$.)

12. (**) Prove that $f(x) = \frac{1}{\sqrt[3]{x}}$ does not belong to $L^3[0, 8]$.
13. (*) Prove that the functions $f(x) = x^2$ and $f(x) = |x|$ are convex.
14. (**) Let $f(x) = |x|^p$. For which value of p is this function convex? Prove that $f(x) = \sqrt{x}$ is not convex. Prove that the affine transformation $f(x) = ax + b$, where $a, b, x \in \mathbb{R}^2$ is convex.
15. (**) Compute $\lim_{n\to\infty} \int_0^n \left(1 \pm \frac{x}{n}\right)^n e^{-x/2}dx$ and $\lim_{n\to\infty} \int_0^n \left(1 \pm \frac{x}{n}\right)^n e^{-2x}dx$.