I. Review of probability fundamentals.

A. Probability spaces, random variables. (Shreve, Chapter 1)

In this course, risk-neutral pricing theory is formulated in the language of measure-theoretic probability. To the aspiring quant, measure-theoretic probability might at first appear mysterious, overly abstract, and irrelevant to quantitative analysis. However, it provides a general and powerful way to express both the conceptual ideas and the basic formulas of risk-neutral pricing, and it is the framework employed in the research literature of mathematical finance. For this reason, you should aim to acquire a confidant, intuitive understanding of measure-theoretic language and be able to translate it into practical computational formulae. You do not need to learn advanced, measure-theoretic techniques or proofs.

The basis of measure-theoretic probability is the probability space. This is a triple $(\Omega, \mathcal{F}, P)$, where $\Omega$ is a set called the outcome space, $\mathcal{F}$ is a collection of subsets of $\Omega$, and $P$ is a measure that assigns to each $A \in \mathcal{F}$ a number $P(A)$ in $[0,1]$. It provides a general template for probability modeling. Given a random phenomenon to model, the elements of $\Omega$ represent all its possible outcomes; $P(A)$, where $A$ is a subset of $\Omega$, represents the probability of $A$; and $\mathcal{F}$ is the collection of subsets to which $P$ assigns probabilities. The elements of $\mathcal{F}$ are referred to as events. In order that $(\Omega, \mathcal{F}, P)$ qualify as a probability space, $\mathcal{F}$ and $P$ must satisfy a few basic conditions, called the axioms of probability, which are consistent with intuitive notions of what probability means and with logical coherence. These are: (i) $\mathcal{F}$ is $\sigma$-algebra, meaning it is a non-empty collection that is closed under the set operations of taking complements and of forming countable intersections or unions; (ii) $P(\Omega) = 1$; and (iii) $P$ is countably additive, namely, if $A_1, A_2, \ldots$ is a sequence of disjoint events, then

$$P \left( \bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} P(A_n). \quad (1)$$

The mysteries of $\sigma$-algebras will be discussed more below. For now think of $\mathcal{F}$ as the class of subsets of $\Omega$ to which probabilities are assigned.
A major difference between elementary and measure-theoretic probability is the treatment of random variables. In elementary probability, a random variable is just a representation of the numerical result of a trial with random outcome, and a random variable is modeled by specifying its cumulative distribution function, probability mass function, or density function, depending on the case. Probability spaces do not enter the definition. In the measure-theoretic approach, the underlying model is always a probability space \((\Omega, \mathcal{F}, P)\), and random variables are functions on \(\Omega\). The idea is that \((\Omega, \mathcal{F}, P)\) models an experiment in which some \(\omega\) is chosen at random, and a random variable is a function \(X\) which assigns a numerical attribute \(X(\omega)\) to each outcome \(\omega\) in \(\Omega\); the value of \(X\) is random because the result \(\omega\) of the experiment we are modeling is random.

However, \(X\) cannot be any old function. It must satisfy a criterion called measurability. This is a consequence of the following condition that we impose on \(X\). For any open interval \((a, b)\) we should be able to assign a probability to the event that \(X(\omega)\) falls in \((a, b)\). This event is the following subset of \(\Omega\): \(X^{-1}((a, b)) := \{\omega; X(\omega) \in (a, b)\}\). Thus the probability that \(X\) falls in \((a, b)\) is

\[
P\left(\{\omega; X(\omega) \in (a, b)\}\right).
\]

But for this to have meaning we need that \(\{\omega; X(\omega) \in (a, b)\} \in \mathcal{F}\), because, by definition, probabilities are assigned to subsets in \(\mathcal{F}\) only.

This leads to an important definition, which we state for a general \(\sigma\)-algebra, as we need it more generally: The function \(X\) is said to be \(G\)-measurable if \(\{\omega; X(\omega) \in (a, b)\} \in G\) for all real numbers \(a < b\). In fact, one can show

\[
X \text{ is } G \text{ measurable if and only if } \{\omega; X(\omega) < x\} \in G \text{ for all } x. \quad (2)
\]

(When this is true it follows that \(X^{-1}(U) := \{\omega; X(\omega) \in U\}\) belongs to \(G\) for any Borel set \(U\) of the real line, an important fact to know, but not to prove. Borel sets are defined later.)

Finally, we are lead to the definition of a random variable defined on a probability space \((\Omega, \mathcal{F}, P)\): it is an \(\mathcal{F}\)-measurable function on \((\Omega, \mathcal{F}, P)\). In practice, we do not worry about measurability in this course, except when conditioning, and we will later give explicit heuristic criteria to use and an intuitive interpretation that will be all you need to know.

In elementary probability, we work with random variables by calculating with their cumulative distribution functions or, if they are continuous, their probability density functions. Recall that if \(X\) is a random variable, \(F_X(t)\) denotes the probability that \(X\) takes a value less than or equal to \(t\). In the the measure-theory setting, this is

\[
F_X(t) := P\left(\{\omega; X(\omega) \leq t\}\right).
\]
Thus, the cumulative distribution function of $X$ is determined by the probability measure $\mathbb{P}$ on the underlying probability space. This is very important, because by changing the underlying measure from $\mathbb{P}$ to some other measure $\tilde{\mathbb{P}}$, we will change the distribution of $X$. In fact, change-of-measure is a useful device in mathematical finance.

Let $A$ belong to $\mathcal{F}$. The indicator function of $A$ is the function on $\Omega$, 

$$1_A(\omega) = \begin{cases} 
1, & \text{if } \omega \in A; \\
0, & \text{if } \omega \not\in A.
\end{cases}$$

This is $\mathcal{F}$-measurable and hence a random variable. Its value indicates whether or not $A$ has occurred. Indicator functions appear often in probability theory.

In the measure-theoretic framework, expectation is defined directly on the probability space by the formula

$$E[X] = \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega),$$

where the right-hand side is an integral with respect to the probability measure. This integral is defined first by requiring that $\int 1_A \, d\mathbb{P} = \mathbb{P}(A)$ for any indicator random variable $1_A$, extending by linearity to sums of indicator random variables, and then extending to general $X$ using a limiting procedure; see Chapter 1 in Shreve’s text for details. The definition is mostly of theoretic use; to actually compute $E[X]$ when $X$ has a density $f_X$, we use the formula,

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

In the measure-theoretic framework, this is proved as a consequence of (3); in elementary probability it is a definition. Another consequence of (3) is that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx,$$

whenever the expectation is defined and $f_X$ is the density of $X$.

**Example 1.** Binomial tree model for an asset price. This is a discrete-time model in which, if the price at time $j-1$ is $S$, then the price at time $j$ will be either $Su$ or $Sd$, where $0 < d < u$. We shall assume that $0 < d < 1 < u$ so that if the price goes from $S$ to $Su$ it increases, and if it goes from $S$ to $Sd$ it decreases. Let $T$ be the time horizon of the model. Any price history can be represented as a vector $\omega = (\omega_1, \ldots, \omega_T)$ of 1’s and −1’s, where $\omega_j = 1$ indicates the stock price increases between times $j-1$ and $j$ and $\omega_j = -1$ indicates it decreases. The outcome space $\Omega$ is thus the set of
all such vectors. For $\mathcal{F}$ we may simply take the collection of all subsets of $\Omega$ (this includes the empty set).

There are a number of important random variables for the binomial tree model. The most basic are $X(t)(\omega) = \omega_t$, for $1 \leq t \leq T$; $X(t)(\omega)$ is just the $t^{th}$ coordinate of the vector $\omega$, and it tells us whether the price has moved up or down in the $t^{th}$ time period. Another convenient sequence of random variables is the accumulated sum of price movements, defined as $B(t)(\omega) = \sum_1^t X(j)(\omega) = \sum_1^t \omega_j$; for each $t$, this represents the difference between the number of positive and negative price movements in $\omega = (\omega_1, \ldots, \omega_T)$ up to time $t$. It is not hard to show (exercise) that the number of positive price movements up to time $t$ is $(1/2)[B(t)(\omega)+t]$ and the number of negative price movements is $(1/2)[t-B(t)(\omega)]$. If $S_0$ denotes the price of the asset at time $t=0$, its price at time $t$, when $\omega = (\omega_1, \ldots, \omega_T)$ is the history of price movements, is

$$S(t)(\omega) = S_0 \cdot u^{[B(t)(\omega)+t]/2} \cdot d^{[t-B(t)(\omega)]/2}, \quad (4)$$

Because the price increases by the multiplicative factor $u$ each time it increases, and otherwise decreases by the multiplicative factor $d$. Now in probability theory, once we get going with a model, it is traditional to drop the explicit dependence on $\omega$ from the notation. Thus, we would usually write the last equation as

$$S(t) = S_0 \cdot u^{[B(t)+t]/2} \cdot d^{[t-B(t)]/2}. \quad (5)$$

This is cleaner looking and easier to read, but remember that random variables are always functions on the underlying probability space!

Define $\delta S(t) = S(t+1) - S(t)$ and $\delta B(t) = B(t+1) - B(t)$, which represent the incremental changes of $S$ and $B$ over time interval $t$; observe that $\delta B(t) = X(t+1)$. The student should check that

$$\delta S(t) = \left[-1 + \frac{u+d}{2}\right] S(t) + \frac{u-d}{2} S(t) \delta B(t), \quad 0 \leq t \leq T - 1. \quad (6)$$

This equation is an equivalent way of stating the model for $S(t)$; it has a unique solution, which is the series of random variables defined in (4). In fact, it is a better way to state the model, because it is a dynamical systems model, showing how the price evolves from each time step to the next.

So far we have not specified any probability measure to put on $(\Omega, \mathcal{F})$. This is a problem of modeling. Your choice depends on how you think the economy behaves. One simple choice is to assume that $X(1), \ldots, X(T)$ are independent and that $p = P(X(t) = 1)$ does not change with $t$. Then the probability of seeing outcome $\omega = (\omega_1, \ldots, \omega_T)$ is

$$P(\{\omega\}) = p^{[B(T)(\omega)+T]/2} (1-p)^{[T-B(T)(\omega)]/2}. \quad (4)$$
because 1 appears in the vector $\omega [B(T)(\omega) + T]/2$ times, each time with probability $p$, and $-1$ occurs $[T - B(T)(\omega)]/2$ times, each time with probability $1 - p$. It follows then by the additivity property (1), that for any subset $A$ of $\Omega$,

$$
P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p^{(1/2)[B(T)(\omega) + T]} (1 - p)^{(1/2)[T - B(T)(\omega)]}.
$$

This probability measure turns out to be useful for derivative pricing in the binomial model.

**Example 2. A probability space formulation of the Black-Scholes price model.** In a way, the Black-Scholes model generalizes Example 1 to continuous time. The outcome space of $\Omega$ will consist of all continuous functions $\omega$ on the interval $[0, T]$, with $\omega(0) = 0$. On this outcome space, define the process, $W(t)(\omega) = \omega(t)$; the $W$ process will play the role that the $B$ process of the previous example played. $W$ should be thought of as the random “noise” causing the asset price to fluctuate around its mean rate of growth. Choose the probability measure $P$ so that $W$ is a Brownian motion process. We will not specify how to do this; just accept that it can be done. For the moment we also defer the subtle issue of defining $F$.

The price of the asset at any time $0 \leq t \leq T$, in terms of its price $S_0$ at time $t=0$ and the “noise” input $W$, is defined to be

$$
S(t) = S_0 \exp\{\sigma W + (\alpha - \frac{1}{2}\sigma^2)t\},
$$

where $\alpha$ and $\sigma$ are parameters of the model. Of course, this satisfies the stochastic differential equation

$$
dS(t) = \alpha S(t) dt + \sigma S(t) dW(t),
$$

There is a close connection between this example and the last. Equations (7) and (8) are the continuous-time versions of (5) and (6), respectively. In particular (7) is a dynamical systems model for the asset price, showing how it changes over an infinitesimal interval of time.

**Remark:** In both these examples, the probability spaces are defined very concretely. Usually this is not the case. Instead, a model is defined by specifying stochastic processes on a probability space which is not explicitly constructed, only assumed to exist. For instance, the Black-Scholes model is constructed using a Brownian motion defined on an unspecified probability space. The exact probability space used is typically not important, but must be there in the background in order that we may use the measure-theoretic framework.

**Discussion.** Why do we introduce all these airy abstractions of measure theory, since we revert to the formulas of elementary probability to do calculations anyway?
They are essential to a deep and practical understanding of pricing theory for three reasons:

(1) a proper theory of convergence of sequences of random variables;

(2) a coherent and general theory of conditioning on partial information;

(3) the change of measure technique.

We will not say much about (1), except that convergence is central to probability theory in general and to defining stochastic integrals in particular.

As for (2), the concept of a \( \sigma \)-algebra is crucial to modeling partial information, and the treatment of random variables as functions on a probability space is essential for defining conditional expectation very generally.

Point (3) is important to pricing theory. The measure-theoretic framework allows one to consider the same outcome space and same random variables under different probability measures. Recall that if \( P \) and \( \hat{P} \) are two probability measures on \((\Omega, \mathcal{F})\), we say that \( P \) and \( \hat{P} \) are equivalent (written \( P \sim \hat{P} \)) when

\[
P(A) = 0 \text{ if and only if } \hat{P}(A) = 0;
\]

this is the same as saying \( P(A) > 0 \) if and only if \( \hat{P}(A) > 0 \). Now suppose that \((\Omega, \mathcal{F}, P)\) models a market evolving over time. This model admits arbitrage (in brief, \( P \) admits arbitrage) if there is a self-financing investment strategy that, starting with 0 dollars, yields a payoff \( X \) satisfying \( P(X \geq 0) = 1 \) (it never produces a loss) and \( P(X > 0) > 0 \) (with positive probability it produces a strict gain). Clearly, if \( P \) admits arbitrage, then so does \( \hat{P} \) for every \( \hat{P} \) equivalent to \( P \). Conversely, if one can find a \( \hat{P} \) equivalent to \( P \) that does not admit arbitrage, then neither does \( P \). This is important because you and I may have very different ideas of which probability measure accurately describes the market. But if our personal choices are equivalent measures, meaning we agree what events occur with positive probability, then we can agree on no-arbitrage prices. If they are not equivalent one of us may perceive an arbitrage opportunity where the other does not, and only time will tell who was right.

The first fundamental theorem of asset pricing, to be reviewed later in the lecture, gives a condition on \( P \), called risk-neutrality, implying no arbitrage, for a class of market models based on stochastic integrals. Therefore, if for a given price model we can find an equivalent risk-neutral probability measure, we can price derivatives by the principle of no arbitrage.

B. \( \sigma \)-algebras, measurability, partial information. (Shreve, Chapter 2)

B.1 Definition and examples of \( \sigma \)-algebras.
Recall that a collection $\mathcal{F}$ of subsets of $\Omega$ is a $\sigma$-algebra if (i) $\Omega \in \mathcal{F}$; (ii) whenever $A \in \mathcal{F}$, so also is $A^c = \Omega - A$; (iii) whenever $A_1, A_2, \ldots$ are in $\mathcal{F}$, so also are $\bigcup_1^\infty A_n$ and $\bigcap_1^\infty A_n$. This concept requires some time to get used to. One may ask why it’s important. For example, why not always assign a probability to every subset of $\Omega$ and forget about $\sigma$-algebras? The answer is that when $\Omega$ is uncountable infinite, as even in the simple case when $\Omega$ is the interval $[0, 1]$, it may not be possible to assign a probability to every subset and satisfy the countable additivity assumption (1) at the same time. Therefore, it is necessary to impose restrictions on the class of subsets of $\Omega$ to which probabilities are assigned. However, we want this class of subsets to at least have the properties (i)—(iv) listed above, for otherwise it would be difficult to do even elementary calculations.

A second reason, and the more important one for this course, is that $\sigma$-algebras are the natural framework for modeling partial information and defining conditional expectation. This will be explained in the next section.

The following basic facts about $\sigma$-algebras are important to know.

(a) The family of all subsets (including the empty set) of a set is a $\sigma$-algebra. When the outcome space $\Omega$ of a probability space model is finite or countable, we usually take $\mathcal{F}$ to be this family (see Example 1.)

(b) If $\mathcal{C}$ is any family of subsets of some set $\Omega$, there exists a $\sigma$-algebra, denoted $\sigma(\mathcal{C})$, satisfying (i) $\mathcal{C} \subset \sigma(\mathcal{C})$; (ii) if $\mathcal{G}$ is any $\sigma$-algebra containing $\mathcal{C}$, then $\sigma(\mathcal{C}) \subset \mathcal{G}$. We call $\sigma(\mathcal{C})$ the $\sigma$-algebra generated by $\mathcal{C}$. It is the smallest $\sigma$-algebra containing $\mathcal{C}$.

Here are some often used examples of $\sigma$-algebras defined by the procedure of (b):

(i) Let $\mathcal{C} = \{A_1, A_2, \ldots\}$ be a disjoint partition of a set $\Omega$; this means the $A_i$’s are disjoint and $\bigcup_1^\infty A_n = \Omega$. Then $\sigma(\mathcal{C})$ is the collection of all finite or countable unions of subsets in $\mathcal{C}$ plus the empty set. For example, $A_i \in \sigma(\mathcal{C})$, for any $i A_i \cup A_j \in \sigma(\mathcal{C})$ for any $i$ and $j$, etc. But no element of $\sigma(\mathcal{C})$, except the empty set, is a proper subset of any of the $A_i$’s.

The simplest example occurs when $\mathcal{C} = \{A, A^c\}$ is a partition of $\Omega$ into two sets. In this case, $\sigma(\mathcal{C}) = \{\emptyset, A, A^c, \Omega\}$.

(ii) Let $\mathcal{U}$ the family of all open subintervals of the real line $\mathbb{R}$. The $\sigma$-algebra $\sigma(\mathcal{U})$ is called the Borel $\sigma$-algebra of $\mathbb{R}$.

(iii) (In this example, we do not invoke (b) directly.) Let $X$ be a random variable on $(\Omega, \mathcal{F}, P)$. Let $\sigma(X)$ be the family of all subsets of the form $\{\omega; X(\omega) \in V\}$, where $V$ is a Borel subset of $\mathbb{R}$. This family is in fact a $\sigma$-algebra, and it is called the $\sigma$-algebra generated by $X$; it is the family of all events defined only in terms of the value of $X$. Since $X$ is a random variable, $\sigma(X)$ will be contained in $\mathcal{F}$, but in general it will not be all of $\mathcal{F}$.

(iv) Let $\{X_\alpha; \alpha \in A\}$ be any family of random variables defined on a probability
that X or \( \{ \}

\text{for all } a \in \mathcal{A}; \text{ so } \mathcal{C} \text{ consists of all events that are defined by the value of } X_\alpha \text{ for some } \alpha \in \mathcal{A}. \text{ Then } \sigma(\mathcal{C}) \text{ is called the } \sigma \text{-algebra generated by } \{X_\alpha; \alpha \in \mathcal{A}\} \text{ and is often denoted } \sigma(\{X_\alpha; \alpha \in \mathcal{A}\}). \text{ Again, it is usually a proper subset of } \mathcal{F}.

\text{Examples 1 and 2, continued. For each integer time } t, \text{ let } \mathcal{F}(t) = \sigma(B(1), \ldots, B(t)), \text{ or, equivalently, } \mathcal{F}(t) = \sigma(X(1), \ldots, X(t)). \text{ This is the } \sigma \text{-algebra of all events defined in terms of the history of the binomial tree price up to time } t. \text{ Suppose that } (x_1, \ldots, x_t) \text{ is a sequence of 1’s and } -1 \text{'s. The set of all price histories whose up and down movements up to time } t \text{ are given by } (x_1, \ldots, x_t) \text{ is an event in } \mathcal{F}(t). \text{ It is the event,}

\[ A(x_1, \ldots, x_t) = \{ \omega = (\omega_1, \ldots, \omega_T); \omega_1 = x_1, \ldots, \omega_t = x_t \}. \]

Actually, such sets form a disjoint partition of \( \Omega \) and \( \mathcal{F}(t) \) is the \( \sigma \)-algebra generated by this partition, as in (i) above.

Likewise, in Example 2, define \( \mathcal{F}^W(t) = \sigma(W(s), s \leq t) \). This is the \( \sigma \)-algebra generated by the history of the binomial tree price up to time \( t \). In this example, an appropriate \( \sigma \)-algebra for the probability space is \( \mathcal{F} = \mathcal{F}^W(T) \).

\text{B.2 Useful facts about measurability}

Throughout, if \( X \) is a function on \( \Omega \) and \( U \) is a subset of the real line, we use \( \{ X \in U \} \) to denote \( \{ \omega; X(\omega) \in U \} \).

Let \( \mathcal{G} \) be a sub-\( \sigma \)-algebra of \( \mathcal{F} \) in the probability space \((\Omega, \mathcal{F}, P)\). We say that \( X \) is \( \mathcal{G} \)-measurable if \( \{ X \in U \} \) is in \( \mathcal{G} \) for all Borel sets \( U \).

A necessary and sufficient condition that \( X \) be \( \mathcal{G} \)-measurable is that \( \{ X \leq a \} \in \mathcal{G} \) for all \( a \).

The student should know and be able to use the following facts:

a) If \( A \in \mathcal{G} \), then \( 1_A \) is \( \mathcal{G} \) measurable.

b) Consider the simple \( \sigma \)-algebra \( \mathcal{H} = \{ A, A^c, \emptyset, \Omega \} \), where \( A \in \mathcal{F} \) is a non-empty, proper subset of \( \Omega \). (see (ii) above). Let \( X \) be \( \mathcal{H} \)-measurable. Then \( X \) must take the form

\[ X(\omega) = a1_A(\omega) + b1_{A^c}(\omega), \]

that is, it must take a constant value on \( A \) and a (possibly different) constant value on \( A^c \). To see this, let \( \omega_0 \in A \) and define \( a = X(\omega_0) \), and let \( \omega_1 \in A^c \) and \( b = X(\omega_1) \). The subset \( \{1\} \), consisting of the single number 1 is a Borel set and so \( \{ \omega; X(\omega) = 1 \} \) must belong to \( \mathcal{H} \). And since \( \omega_0 \in \{ \omega; X(\omega) = 1 \} \), it follows that either \( \{ \omega; X(\omega) = 1 \} = A \) or \( \{ \omega; X(\omega) = 1 \} = \Omega \); in either case, \( X(\omega) \) for all \( \omega \in A \). The same reasoning shows that \( X(\omega) = b \) for all \( \omega \in A^c \). Thus \( X = a1_A + b1_{A^c} \), as claimed.
c) Let \( Y \) be a random variable that is measurable with respect to \( \sigma(X_1, \ldots, X_N) \). Then one can write \( Y \) in the form \( Y = \psi(X_1, \ldots, X_N) \) for some function \( \psi \). Conversely, if \( Y = \psi(X_1, \ldots, X_N) \), it is \( \sigma(X_1, \ldots, X_N) \)-measurable.

d) Simple arithmetic operations and limiting operations performed on measurable functions lead again to measurable functions.

B.3 Partial information.

Conditioning on partial information is crucial to math finance. Let \( \mathcal{C} \) be a family of events in a probability space modeling an experiment, and suppose the experiment is run and results in some outcome \( \omega \), which we do not know. Let us say that we have observed \( \mathcal{C} \) if for each event \( A \) in \( \mathcal{C} \) we know whether or not \( \omega \) belongs to \( A \). This is equivalent to saying that for each \( A \) in \( \mathcal{C} \), we know the value of the indicator random variable \( 1_A(\omega) \). So, although we may not know \( \omega \) precisely, we have some idea of its whereabouts.

The idea of observing a family of events is a model for partial information. Notice that if we know \( 1_A(\omega) \), then we know also \( 1_{\bar{A}}(\omega) = 1 - 1_A(\omega) \). Likewise, if we know for every \( n \), whether \( \omega \in A_n \) or not, we know whether or not \( \omega \) is in \( \bigcup_{n=1}^\infty A_n \) or \( \bigcap_{n=1}^\infty A_n \). Therefore, it makes sense to suppose that partial information is obtained by observing some \( \sigma \)-algebra of events. We always assume partial information is represented by a \( \sigma \)-algebra.

The following is important and intuitive. Let \( X \) be a random variable. Observing \( \sigma(X) \) is equivalent to knowing the value of \( X(\omega) \). Likewise, observing \( \sigma(X(s), s \leq t) \) is equivalent to knowing the value of \( X(s)(\omega) \) for all \( s \leq t \). More generally, if \( X \) is \( \mathcal{G} \)-measurable and if we have observed \( \mathcal{G} \), then we know the value of \( X(\omega) \), but not necessarily what \( \omega \) is.

The partial information represented by a \( \sigma \)-algebra can be much more complex than the partial information contained in a single random variable, or even a finite set of random variables. This is a major reason for introducing \( \sigma \)-algebras.

B.4 Filtrations.

A filtration is a family \( \{ \mathcal{F}(t); t \geq 0 \} \) of \( \sigma \)-algebras indexed by \( t \geq 0 \) that is increasing, in the sense that \( \mathcal{F}(s) \subset \mathcal{F}(t) \) whenever \( 0 \leq s \leq t \). In math finance, filtrations model market histories; for each \( t \), \( \mathcal{F}(t) \) represents the information contained in the entire history of the market up to time \( t \). When the market is modeled by a stochastic process \( \{ X(t), t \geq 0 \} \), this past history at \( t \) is \( \mathcal{F}^X(t) = \sigma(X(s), s \leq t) \): \( \{ \mathcal{F}^X(t); t \geq 0 \} \) is called the filtration generated by the process \( X \). We have seen such filtrations already for Examples 1 and 2.

C. Conditional expectation (Shreve, Chapter 2)
In undergraduate probability, one defines the notion of the conditional expectation of one random variable \( X \) given that another random variable \( Y \) equals \( y \). If \( X \) and \( Y \) have a joint density \( f(x,y) \), this is

\[
E[h(X)|Y=y] = \int_{-\infty}^{\infty} h(x) \frac{f(x,y)}{f_Y(y)} \, dx, \quad \text{if } f_Y(y) > 0.
\]

(9)

(It is left undefined or set equal to an arbitrary constant if \( f_Y(y) = 0 \).)

This definition is not adequate for our purposes, because partial information can come in much more complicated forms than observation of a single, or even several, random variables. For example, in asset pricing, we would like to condition on observing the whole past history of the market. In general, we want to define the conditional expectation of \( X \) given \( G \), where \( G \) is a sub-\( \sigma \)-algebra of \( \mathcal{F} \), because \( \sigma \)-algebras are the most general way we represent partial information.

The right approach is to define this conditional expectation as a random variable, that is, as a function on \( \Omega \). (By way of contrast, \( E[h(X)|Y=y] \) defines a function of \( y \), not a function on the probability space.) This random variable shall be denoted by \( E[X|G] \). (We should really write \( E[X|G](\omega) \), to show its dependence on \( \omega \), but we shall only do this when extra clarity is needed.) \( E[X|G] \) is defined by imposing two simple conditions:

1. \( E[X|G](\omega) \) must be \( G \)-measurable.
2. For every \( U \in G \), \( E[1_U X] = E[1_U E[X|G]] \).

Where do these conditions come from? Intuitively, the value of \( E[X|G](\omega) \) should be determined from observing \( G \). By the remarks in the previous section, observation of \( G \) determines the value of any \( G \)-measurable random variable. This is why condition (1) is required. Condition (2) is harder to explain, but, it says roughly that \( E[X|G] \) is obtained by averaging over the “smallest” sets in \( G \), and it determines what \( E[X|G] \) must be. To understand the essential point, consider the simple \( \sigma \)-algebra \( G = \{ A, A^c, \emptyset, \Omega \} \) generated by the event \( A \), where \( 0 < \mathbf{P}(A) < 1 \). By condition (1) and the remarks in B.2 (b) above, \( E[X|G] = a1_A + b1_{A^c} \) for some constants \( a \) and \( b \). Now apply condition (2) with \( U = A \). Since \( 1_A 1_A = 1_A \) and \( 1_A 1_{A^c} = 0 \),

\[
E[1_A X] = E[1_A E[X|G]] = E[a1_A 1_A + b1_{A^c}] = aE[1_A] = a\mathbf{P}(A).
\]

It follows that \( a = \frac{E[1_A X]}{\mathbf{P}(A)} \), which is an average of \( X \) over the set \( A \). Similarly, \( b = \frac{E[1_{A^c} X]}{\mathbf{P}(A^c)} \). The nice thing about conditions (1) and (2) is that they generalize the notion of conditional expectation far beyond this simple example.
Theorem 1 If $E[|X|] < \infty$ then there is a unique (up to modification on sets of probability zero) random variable $E[X \mid C]$ satisfying conditions (1) and (2).

This theorem says that the concept of conditional expectation is well-defined, but it does not tell us how to compute the conditional expectation explicitly. For this we usually revert to the formulas of elementary probability theory. You should know the following cases:

- $\mathcal{G} = \{A, A^c, \emptyset, \Omega\}$. This was treated above.

- $\mathcal{G} = \sigma(Y)$. In this case we use $E[Z \mid Y]$ to denote $E[Z \mid \sigma(Y)]$. Since this is $\sigma(Y)$-measurable, it follows from fact (c) in B.2 above that there is a function $\psi$ such that $E[Z \mid \sigma(Y)] = \psi(Y)$. When $Z = h(X)$ and $X$ and $Y$ have a joint density $f$, $\psi(y)$ is precisely the function of $y$ defined by the right-hand side of (9) in the definition of $E[h(X)|Y=y]$. The following formula states this explicitly, but for $Z$ of the more general form $h(X,Y)$:

$$E[h(X,Y) \mid Y](\omega) = \int_{-\infty}^{\infty} h(x,Y(\omega)) \frac{f(x,Y(\omega))}{f_Y(Y(\omega))} \, dx. \quad (10)$$

This formula shows that the measure-theoretic definition of conditional expectation does indeed generalize the elementary definition.

- The following is a particularly important formula for us. Suppose that $X$ is independent of the $\sigma$-algebra $\mathcal{G}$ (this means every event in $\sigma(X)$ is independent of every event in $\mathcal{G}$). Suppose that $Y$ is $\mathcal{G}$-measurable (and hence independent of $X$). Then

$$E[h(X,Y) \mid \mathcal{G}] = E[h(X,Y) \mid Y(\omega)] = E[h(X,y)] \bigg|_{y=Y(\omega)}. \quad (11)$$

(This right-hand-side means we evaluate the function $y \to E[h(X,y)]$ at $y = Y(\omega)$.) This formula is stated and explained in Lemma 2.3.4 in Shreve. Memorize it!

Two fundamental properties of conditional expectation, used over and over, are

If $Z$ is $\mathcal{G}$-measurable, $E[ZX \mid \mathcal{G}] = ZE[X \mid \mathcal{G}]$; \hfill (Tower property)

If $\mathcal{G} \subset \mathcal{H}$, $E[E[X \mid \mathcal{H}] \mid \mathcal{G}] = E[X \mid \mathcal{G}]$. \hfill (13)

D. Martingales.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let $\{\mathcal{F}(t); t \geq 0\}$ be a filtration in this space. A stochastic process $M = \{M(t); t \geq 0\}$ defined on this probability space is said to be a martingale with respect to $\{\mathcal{F}(t); t \geq 0\}$ and $\mathcal{P}$ if
For every $t \geq 0$, $E[|M(t)|] < \infty$.

For every $t$, $M(t)$ is measurable with respect to $\mathcal{F}(t)$ and for all $0 \leq s < t$,

$$E[M(t) \mid \mathcal{F}(s)] = M(s) \quad (14)$$

Equation (14) is called the martingale property. Its validity depends on the filtration $\{\mathcal{F}(t); t \geq 0\}$ and the probability $P$. However, when the probability measure is fixed and not subject to change, we shall just say that $M$ is a martingale with respect to $\{\mathcal{F}(t); t \geq 0\}$ or an $\{\mathcal{F}(t); t \geq 0\}$-martingale.

Observe that if $M$ is a martingale, it is constant in expectation: $E[M(t)] = E[M(0)]$ for all $t \geq 0$.

Suppose that $M$ represents your fortune at time $t$ in an ongoing game of chance. If $s$ is the current time and $t > s$ a future time, (14) says that your expected fortune at time $t$, conditioned on everything that has happened up to time $s$, is just what you presently have, $M(s)$. Thus, the game is fair, at least as measured by expected value.

The basic examples for us of martingales are Brownian motion and stochastic integrals with respect to Brownian motion. Let $W$ be a Brownian motion and let $\{\mathcal{F}(t); t \geq 0\}$ be a filtration for $W$; see Shreve, page 51 for the definition. Then if $t > s$, we know $W(t) - W(s)$ is independent of $\mathcal{F}(s)$ and has zero mean. Thus $E[W(t) \mid \mathcal{F}(s)] - W(s) = E[W(t) - W(s) \mid \mathcal{F}(s)] = E[W(t) - W(s)] = 0$, and the martingale property follows immediately.

Consider a game in which you are allowed to place adapted bets on the increments of a $\{\mathcal{F}(t); t \geq 0\}$-martingale $M$. This means that if you place a bet of $\alpha$ at time $s$ and hold it until time $t$, you will earn the amount $\alpha(M(t) - M(s))$. To say the bet is adapted means that if $\alpha$ is bet at time $s$, it must be $\mathcal{F}(s)$ measurable; this is a way of saying you are not able to look into the future when deciding how much to bet. The expected gain of this bet is $E[\alpha(M(t) - M(s)) \mid \mathcal{F}(s)] = \alpha E[M(t) - M(s) \mid \mathcal{F}(s)] = 0$. Thus, the game is fair no matter how you bet. This observation leads to an important heuristic principle: let $X(t)$ be the gain at $t$ from betting on the increments of a martingale using adapted bets, or let $X(t)$ be a limit of such gains for a sequence of betting schemes. Then $X$ is also a martingale with respect to $\{\mathcal{F}(t); t \geq 0\}$. Of course, to make this into a theorem requires a more precise formulation and further technical conditions. Stochastic integrals with respect to a Brownian motion provide one example of such a precise formulation. Shreve shows how to define the stochastic integral

$$\int_0^t \alpha(s) dW(s), \quad t \leq T$$

for a process $\alpha(s)$ satisfying: (i) $\alpha(s)$ is $\mathcal{F}(s)$-measurable for all $s$; and (ii) $E[\int_0^T \alpha^2(s) ds] < \infty$. It is very useful for intuition to think of $\alpha(\cdot)$ as a betting strategy and to think
of the stochastic integral as the earnings gained by using \( \alpha(\cdot) \) to bet on the increments of \( W \); formally, \( \alpha(s)\,dW(s) \) is the gain from betting \( \alpha(s) \) on the increment \( dW(s) = W(s + ds) - W(s) \) and the stochastic integral just adds these gains up. A main result of stochastic integration theory, one we use over and over, is that if \( E[\int_0^T \alpha^2(s)\,ds] < \infty \), the process \( \int_0^t \alpha(s)\,dW(s) \) is a martingale up to time \( T \).

**Example 3:** A martingale in the model of Example 1. In the binomial tree model, let \( p = \frac{1-d}{u-d} \). Recall that \( p \) is the probability that the asset price increases by a factor of \( u \) as opposed to decreasing by a factor of \( d \), where \( 0 < d < 1 < u \). We claim that then, \( \{S(t); 0 \leq t \leq T\} \) is a martingale with respect to \( \mathcal{F}(t) = \sigma(X(1), \ldots, X(t)) \). To see this, we will use equation (6), the identity \( B(t+1) - B(t) = X(t+1) \), and the fact that under the measure \( \mathbb{P} \) defined for Example 1, \( X(1), \ldots, X(T) \) are independent and identically distributed with \( \mathbb{P}(X = 1) = p, \mathbb{P}(X = -1) = 1-p \). It follows that \( X(t+1) \) is independent of \( \mathcal{F}(t) \) and hence that \( E[X(t+1)|\mathcal{F}(t)] = E[X(t+1)] = p - (1-p) = 2p - 1 = 2(1-d)/(u-d) - 1 \). Hence, by conditioning both sides of (6) on \( \mathcal{F}(t) \), and using property (12) of conditional expectation,

\[
E\left[S(t+1) - S(t) \mid \mathcal{F}(t)\right] = S(t)E\left[-1 + \frac{u+d}{2} + \frac{u-d}{2}X(t+1) \mid \mathcal{F}(t)\right] = S(t)[-1 + \frac{u+d}{2} + \frac{u-d}{2}(2p-1)] = 0.
\]

\[\diamondsuit\]

**Example 4.** A martingale in the model of Example 2. Consider Example 2. The price of the asset at time \( t \), discounted by its growth rate \( \alpha \), is

\[e^{-\alpha t}S(t) = S_0 \exp\{\sigma W(t) - (1/2)\sigma^2 t\} \].

Itô’s rule shows this to be a martingale. With \( f(t, x) = S_0 e^{-\sigma^2 t/2} e^{\sigma x} \), Itô implies that

\[
d[e^{-\alpha t} S(t)] = df(t, W(t)) = [f_t + (1/2)f_{xx}](W(t))\,dt + f_x(t, W(t))\,dW(t)
\]

\[= \sigma e^{-\alpha t} S(t)\,dB(t) \]

It follows that \( e^{-\alpha t} S(t) = S_0 + \int_0^t \sigma e^{-\alpha s} S(s)\,dW(s) \), which is a martingale.

**E. Brownian motion, stochastic integration, Itô’s rule.**

This section outlines the bare essentials of what you need to know about Brownian motion and stochastic integrals.

A standard Brownian motion \( W = \{W(t); t \geq 0\} \) is a stochastic processes (defined on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), of course) satisfying: (i) \( W(0) = 0 \); (ii) the increments of \( W \) are independent—that is, for any \( 0 < t_1 < t_2 < \cdots < t_n \),
\( W(t_1), W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1}) \) are independent; (iii) for any 0 ≤ s < t, \( W(t) - W(s) \) is normally distributed with mean 0 and variance \( t - s \); (iv) for each \( \omega \) the sample path, \( t \rightarrow W(t)(\omega) \), is continuous in \( t \).

A direct consequence of condition (iii) in the definition of Brownian motion is that \( E[(W(t) - W(s))^2] = t - s \). This is related to the central fact about Brownian motion: its quadratic variation up to time \( T \) equals \( T \) itself. This means the following: for each \( n \), suppose \( \Pi^{(n)} \) is a partition \( t_0^{(n)} < t_1^{(n)} < \cdots < t_m^{(n)} = t \) of \([0, T]\), and suppose \( \lim_{n \to \infty} \max_j [t_j^{(n)} - t_{j-1}^{(n)}] = 0 \). Then

\[
\sum_j \left[ W(t_j^{(n)}) - W(t_{j-1}^{(n)}) \right]^2 \text{ converges in probability to } T \text{ as } n \to \infty. \tag{15}
\]

There is an informal way to think about these statements. Let \( dt \) represent an “infinitesimal” change in \( t \), and let \( dW(t) = [W(t + dt) - W(t)] \) be the corresponding increment in \( dt \). We know \( E[(dW(t))^2] = dt \), but equation (15) says in addition that if we “add” up the values \( (dW(t))^2 \) for \( t \) ranging from 0 to \( T \) we get the deterministic value \( T \). This suggests we can think of \((dW(t))^2 \) informally as being equal to \( dt \). This idea leads non-rigorously to Itô’s rule and stochastic calculus.

It is simple to verify that a Brownian motion \( W \) is a martingale, and also that \( W^2(t) - t \) is a martingale. The converse is true: Levy’s characterization of Brownian motion says that a process \( W \) with continuous paths for which \( W(0) = 0 \) and both \( W \) and \( W^2(t) - t \) are martingales, must be a Brownian motion.

A filtration \( \{\mathcal{F}(t); t \geq 0\} \) is a called filtration for the Brownian motion \( W \) if: (i) \( W(t) \) is measurable w.r.t. \( \mathcal{F}(t) \) for each \( t \); (ii) for any 0 ≤ s < t, \( W(t) - W(s) \) is independent of \( \mathcal{F}(s) \). If \( W \) is a Brownian motion, the filtration it generates by the definition \( \mathcal{F}^W(t) = \sigma(W(s), s \leq t) \) is a filtration for \( W \). But often we need to work with filtrations for \( W \) which are strictly larger, in order to model other sources in the flow of historical information. If \( \{\mathcal{F}(t); t \geq 0\} \) is a filtration for the Brownian motion \( W \), then \( W \) is a martingale with respect to this filtration.

The following fact is used over and over again. Let \( \{\mathcal{F}(t); t \geq 0\} \) be a filtration for the Brownian motion \( W \). For any fixed \( s \), the process \( \{W(s+t) - W(s); t \geq 0\} \) is a Brownian motion independent of \( \mathcal{F}(s) \).

A stochastic process \( X = \{X(t); t \geq 0\} \) is said to be adapted to a filtration \( \{\mathcal{F}(t); t \geq 0\} \) if \( X(t) \) is \( \mathcal{F}(t) \)-measurable for all \( t \). Recall that this implies that for each \( t \), if we have observed \( \mathcal{F}(t) \) then we know the value of \( X(t)(\omega) \). Adaptedness is an important concept in finance. Typically, \( \mathcal{F}(t) \) represents the history of the market up to time \( t \). A portfolio process \( \Pi = \{\pi(t); t \geq 0\} \) is a process representing the ongoing investment choices made by an investor. We assume that investors cannot look into the future (the world would be far different if they could!) Their choice of portfolio at time \( t \) can only depend on the information available to them, namely,
the market history $\mathcal{F}(t)$. Mathematically, this means that portfolio processes must be adapted to the market history filtration.

Let $W$ be a Brownian motion and let $\{\mathcal{F}(t); t \geq 0\}$ be a filtration for $W$. When $\alpha$ is adapted to this filtration and

$E[\int_0^T \alpha^2(s) \, ds] < \infty,$

we can define the stochastic integral process,

$$\int_0^t \alpha(s) \, dW(s), \quad 0 \leq t \leq T.$$  

This integral has the following properties.

(a) It is has continuous paths.

(b) It is a mean-zero martingale with respect to $\{\mathcal{F}(t); t \geq 0\}$.

(c) The quadratic variation of this process on $[0, t]$ is $\int_0^t \alpha^2(s) \, ds$.

(d) $E \left[ \left( \int_0^t \alpha(s) \, dw(s) \right)^2 \right] = E \left[ \int_0^t \alpha^2(s) \, ds \right]$.

The stochastic integral is defined first on simple process, that is, adapted processes of the form

$$\alpha(t) = \sum_{j=1}^n \alpha_j 1_{[t_{j-1}, t_j)}(t),$$

where $t_0 < t_1 < \cdots < t_n = T$ is a partition of $[0, t]$, and, for each $j$, $\alpha_j$ is $\mathcal{F}_{t_{j-1}}$-measurable. For such a process (here $t \wedge s = \min\{t, s\}$)

$$\int_0^t \alpha(s) \, dW(s) = \sum_{j=1}^n \alpha_j \left[ W(t \wedge t_j) - W(t \wedge t_{j-1}) \right].$$

As we discussed above, this should be interpreted as the total amount earned up to time $t$ by betting $\alpha(t_j) = \alpha_t$ at time $t_{j-1}$ on the increment $[W(t \wedge t_j) - W(t \wedge t_{j-1})]$.

The definition in the general case is carried out by approximating a general adapted process satisfying (16) by simple processes. The integral in this case can be thought of as the total amount earned up to time $t$ by betting $\alpha(t)$ on each infinitesimal increment $dW(t)$.

(In the interests of full disclosure, it is necessary to also assume what is called a joint-measurability condition on $\alpha$. This is a technical detail we do not need to worry about.)
Let \( Y(t) = Y_0 + \int_0^t \beta(s) \, ds + \int_0^t \alpha(s) \, dW(s) \), where \( \beta \) is integrable and adapted to \( \{ \mathcal{F}(t); t \geq 0 \} \). Such a process is called an Itô process. It is common to express the definition of an Itô process using differential notation. For example, \( dY(t) = \beta(t) \, dt + \alpha(t) \, dW(t) \), \( Y(0) = Y_0 \). Next, let \( f(t, x) \) be a function which is at least once continuously differentiable in \( t \) and twice continuously differentiable in \( x \). Then Itô’s rule says that

\[
\begin{align*}
    f(t, Y(t)) &= f(0, Y_0) + \int_0^t \left[ f_t(s, Y(s)) + \frac{1}{2} \alpha^2(s) f_{xx}(s, Y(s)) \right] \, ds + \int_0^t f_x(s, Y(s)) \, dY(s) \\
    &= f(0, Y_0) + \int_0^t \left[ f_t(s, Y(s)) + f_x(s, Y(s)) \beta(s) + \frac{1}{2} \alpha^2(s) f_{xx}(s, Y(s)) \right] \, ds \\
    &\quad + \int_0^t f_x(s, Y(s)) \alpha(s) \, dW(s).
\end{align*}
\]

This is often also expressed in differential notation:

\[
d[f(t, Y(t))] = \left[ f_t(t, Y(t)) + \frac{1}{2} \alpha^2(t) f_{xx}(t, Y(t)) \right] \, dt + f_x(t, Y(t)) \, dY(t).
\]

Itô’s rule is informally derived by applying Taylor’s expansion to order 2 and replacing \((dW(t))^2\) by \(dt\), as discussed above, and replacing \(dW(t)\,dt\) and \((dt)^2\) by 0.

There is a technical detail to stating Itô’s formula that is glossed over in Shreve’s text. By the theory developed so far, the last integral in (17) is defined for \( t \leq T \) only if \( \int_0^T [f_x(s, Y(s)) \alpha(s)]^2 \, ds < \infty \). To have to verify this whenever Itô’s rule is applied would be tedious. Fortunately, stochastic integrals can be generalized to a larger class of integrands that includes that of (17) in all cases. In fact, \( \int_0^T \gamma(s) \, dW(s) \) can be defined for any adapted process \( \gamma \) satisfying only

\[
P \left( \int_0^T \gamma^2(s) \, ds < \infty \right) = 1.
\]

**F. Markov processes.**

Let \( X \) be a stochastic process. Let \( \{ \mathcal{F}(t); t \geq 0 \} \) be a filtration to which \( X \) is adapted. We say that \( X \) is a Markov process with respect to \( \{ \mathcal{F}(t); t \geq 0 \} \) if for any \( 0 \leq s \leq t \) and any Borel function \( h \) for which \( E[|h(X(t))|] < \infty \),

\[
    E [h(X(t)) \mid \mathcal{F}(s)] = E [h(X(t)) \mid X(s)].
\]

The same definition applies to vector-valued processes \( X(t) = (X_1(t), \ldots, X_d(t)) \). In this context the right hand side of (19) must be interpreted as \( E [h(X(t)) \mid X_1(s), \ldots, X_d(s)] \).
One can check that (19) is true by showing that

\[ E[h(X(t)) \mid \mathcal{F}(s)] = \psi(s, X(s)) \]  

for some function \( \psi(s, x) \). Indeed, suppose this is true. Since \( X \) is adapted to \( \{\mathcal{F}(t); t \geq 0\} \) by assumption, \( \sigma(X(s)) \subset \mathcal{F}(s) \) and the tower property of conditional expectation implies

\[
E[H(X(T)) \mid X(s)] = E[H(X(T)) \mid \sigma(X(s))] = E\left[ E[H(X(T)) \mid \mathcal{F}(s)] \mid \sigma(X(s)) \right]
\]

\[
= E\left[ \psi(s, X(s)) \mid \sigma(X(s)) \right] = \psi(s, X(s)) = E[h(X(t)) \mid \mathcal{F}(s)]
\]

The Markov property (19) then follows.

Markovian models are very useful in finance. Derivative prices in the risk-neutral theory are expressed as conditional expectations of the form, \( E[H(Y(T)) \mid \mathcal{F}(t)] \). To compute this in general we have to know the distribution of \( Y(T) \) based on the whole past history \( \mathcal{F}(t) \), and there is no simple, general approach to deduce such a potentially complicated conditional distribution. However suppose that \( Y \) is a Markov process; then the price reduces to the conditional expectation \( E[H(Y(T)) \mid Y(t)] \), and this we know how to compute, for example from formula (10), if we know the joint density of \( (Y(t), Y(T)) \). More generally, suppose that \( Y(t) \) is one component of a vector-valued Markov process—in other words, there is a Markov process \( X(t) = (Y(t), X_2(t), \ldots, X_d(t)) \). Then \( E[H(Y(T)) \mid \mathcal{F}(t)] = E[H(Y(T) \mid Y(t), X_2(t), \ldots, X_d(t))] \), and again this can be computed from knowing the joint density of \( X(t) \) and \( X(T) \). The great simplification afforded by the Markov property is a major impetus for developing Markov models in finance.

**Example 5.** The Black-Scholes price process of Example 2 is a Markov process. This is easy to derive from the following identity, which follows directly from the definition of \( S(t) \) in equation (7): if \( s < t \)

\[ S(t) = S(s) \exp\{\sigma(B(t) - B(s)) + [\alpha - \sigma^2/2](t - s)\}. \]

Recall that \( B \) is a Brownian motion. Let \( \{\mathcal{F}(t); t \geq 0\} \) be a filtration for \( B \). Since a Brownian motion has independent increments and the only randomness in \( S(s) \) is contributed by \( B(s) \), \( S(s) \) and \( B(t) - B(s) \) are independent. Thus, using the conditional expectation formula of (11),

\[
E\left[ h(S(t)) \mid \mathcal{F}(s) \right] = E\left[ h\left(S(s)e^{\sigma(B(t) - B(s)) + [\alpha - \sigma^2/2](t - s)}\right) \mid \mathcal{F}(s) \right]
\]

\[
= E\left[ h\left(x e^{\sigma(B(t) - B(s)) + [\alpha - \sigma^2/2](t - s)}\right) \bigg| x = S(s) \right]
\]
This is a function of \(S(s)\) alone, which according to (20) shows that the price process \(S\) is Markovian.

For later application to the Black-Scholes formula, observe that since \(B(t) - B(s)\) is normal with mean 0 and variance \(t - s\), the conditional expectation in the previous formula is

\[
E[h(S(t)) \mid S(s) = x] = \int_{-\infty}^{\infty} h\left(x e^{\sigma y + \left[\alpha - \sigma^2/2\right](t-s)} e^{-y^2/2(t-s)} \right) \frac{dy}{\sqrt{2\pi(t-s)}}
\]

(21)

The following result is also important to know. Let \(\beta(t, x)\) and \(\gamma(t, x)\) be continuous and continuously differentiable in \(x\) and suppose that \(\beta_x(t, x)\) and \(\gamma_x(t, x)\) are uniformly bounded. Let \(W\) be a Brownian motion and let \(\{F(t); t \geq 0\}\) be a filtration for \(W\). Then the stochastic differential equation,

\[
dX(t) = \beta(t, X(t)) \, dt + \gamma(t, X(t)) \, dW(t), \quad X(0) = X_0,
\]

where \(W\) is a Brownian motion, has a unique solution which is a Markov process. The basic reason for the Markov property is that for any \(t > s \geq 0\), \(X(t)\) can be written as a function of \(X(s)\) and the process \(W(s+\tau) - W(s)\), \(0 \leq \tau \leq t-s\), but \(\{W(s+\tau) - W(s); \tau \geq 0\}\) is independent of \(F(s)\), so all the information in \(F(s)\) relevant to computing conditional expectations is contained in the value of \(X(s)\).

More general stochastic differential equations for asset prices than the Black-Scholes equation are used in math finance for local volatility models.

G. Change of measure and expectations.

One advantage of always working with an underlying probability space is the ease of applying the technique of changing probability measure.

Let \((\Omega, \mathcal{F}, P)\) be a probability space. Let \(Z\) be a non-negative random variable such that \(E_P[Z] = 1\). For every event \(A\) in \(\mathcal{F}\), define

\[
\hat{P}(A) = E_P[1_A Z] = \int_{\Omega} 1_A(\omega) Z(\omega) \, dP(\omega).
\]

(22)

Then \(\hat{P}\) is a measure, and, since \(\hat{P}(\Omega) = E_P[1_\Omega Z] = E_P[Z] = 1\), it is a probability measure.

When \(\hat{P}\) and \(P\) are related as in (22), we often use \(\frac{d\hat{P}}{dP}\) to denote \(Z\). Given two probability measure \(P\) and \(Q\) it is useful to know when \(\frac{dQ}{dP}\) exists, meaning we can write \(Q(A) = E_{\hat{P}}[1_A dQ/d\hat{P}]\). We say that \(Q\) is absolutely continuous with respect to \(P\)
if \( Q(A) = 0 \) whenever \( P(A) = 0 \), and in this case we write \( Q \ll P \). The Radon-Nikodym theorem says that \( Q \ll P \) is a necessary and sufficient condition for \( \frac{dQ}{dP} \) to exist.

Two probability measure \( Q \) and \( P \) are said to be equivalent if \( Q \ll P \) and \( P \ll Q \); this means that the class of events with positive probability is the same for both measures. The idea of equivalence is conceptually important to derivative pricing. The idea is that you and I might model the stochastic behavior of the market using different probability measures. But if our measures are equivalent, we should agree on the price of any derivative that can be replicated by the underlyings. Also, if my model does not admit arbitrage, neither will yours and vice-versa. So, from the point of view of pricing, modeling only makes a difference up to equivalence of probability measures. Put another way, if you assign a positive probability to events which I think cannot happen, we will likely come up with different prices.

Let \( \frac{d\hat{P}}{dP} = Z \). It is important to be able to express expectations with respect to \( \hat{P} \) in terms of expectations with respect to \( P \) and vice-versa. By definition, \( E_{\hat{P}}[1_A] = \hat{P}(A) \), so another way to write (20) is \( E_{\hat{P}}[1_A] = E_P[1_A Z] \). Therefore if \( X \) is a simple random variable of the form \( X = \sum_i^n c_i 1_{A_i} \),

\[
E_{\hat{P}}[X] = \sum_{i=1}^n c_i E_{\hat{P}}[1_A] = \sum_{i=1}^n c_i E_P[1_A Z] = E_P[XZ].
\]

This generalizes, leading to a formula expressing expectation with respect to \( \hat{P} \) in terms of expectation with respect to \( P \): if \( E_P[|XZ|] < \infty \), then

\[
E_{\hat{P}}[X] = E_P[XZ]. \tag{23}
\]

We can go in the reverse direction if \( P(Z > 0) = 1 \). Then \( Z^{-1} \) is defined on a set of \( P \) probability 1, and

\[
E_P[X] = E_P[(XZ^{-1})Z] = E_{\hat{P}}[XZ^{-1}]. \tag{24}
\]

Example 6; baby version of Girsanov’s theorem.

Let \( X \) be a random variable. We shall write \( X \sim n(\mu, \sigma^2) \) to indicate that \( X \) is a normal random variable with mean \( \mu \) and variance \( \sigma^2 \); hence its density is

\[
f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.
\]

We shall denote the standard normal cumulative distribution function by \( N \):

\[
N(x) = \int_{-\infty}^{x} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \, dz.
\]

For the following example we need to know the following basic fact:
• $X \sim n(\mu, \sigma^2)$ if and only if its moment generating function is $M_X(\lambda) = E[e^{\lambda X}] = e^{\mu \lambda + \sigma^2 \lambda^2 / 2}$, $-\infty < \lambda < \infty$.

**Theorem 2** Let $X$ be a random variable on $(\Omega, \mathcal{F}, P)$ and assume $X \sim n(0, \tau)$. Define

$$\tilde{P}(A) = e^{-\tau a^2 / 2} E_P[1_A e^{aX}].$$

Then $\tilde{P}$ is a probability measure and, as a random variable on $(\Omega, \mathcal{F}, \tilde{P})$, $X$ is normal with mean $\tau a$ and variance $\tau$.

This theorem is, in a sense, a baby version of Girsanov’s theorem. To prove it, notice that $\tilde{P}(A) = E_P[1_A Z]$ where

$$Z = \frac{e^{aX}}{e^{\tau a^2 / 2}}.$$

If $X \sim N(0, \tau)$, $E_P[e^{aX}] = e^{\tau a^2 / 2}$, hence $E_P[Z] = 1$, and so $\tilde{P}$ is a probability measure.

The moment generating function of $X$, as a random variable on $(\Omega, \mathcal{F}, \tilde{P})$, is, using formula (21) for expressing expectation under a change of measure,

$$E_{\tilde{P}}[e^{\lambda X}] = e^{-\tau a^2 / 2} E_P[e^{\lambda X} e^{aX}] = e^{-\tau a^2 / 2} e^{\tau (a + \lambda)^2 / 2}.$$

The last step just uses fact (i) above and the assumption that $X \sim n(0, \tau)$, as a random variable on $(\Omega, \mathcal{F}, \tilde{P})$. By simplifying the last expression:

$$E_{\tilde{P}}[e^{\lambda X}] = e^{\tau a \lambda + \tau \lambda^2 / 2}.$$

But this is precisely the moment generating function of a normal random variable with mean $\tau a$ and variance $\tau$.

The following consequence of Theorem 2 is used in deriving option pricing formulas for the standard Black-Scholes model.

**Corollary 1** Let $X \sim n(0, \tau)$ on $(\Omega, \mathcal{F}, P)$. Then

$$E_P[e^{aX} 1_{\{X \geq b\}}] = e^{\tau a^2 / 2} N\left(\frac{a \tau - b}{\sqrt{\tau}}\right) \quad (25)$$

To prove this, observe that, using the notation of Theorem 2 and its proof,

$$E_P[e^{aX} 1_{\{X \geq b\}}] = e^{\tau a^2 / 2} E_{\tilde{P}}[1_{\{X \geq b\}}].$$
But since $X \sim n(\tau a, \tau)$ as a random variable on $(\Omega, \mathcal{F}, \hat{P})$,

$$e^{\tau a^2/2} E_{\hat{P}}[1_{\{X \geq b\}}] = e^{\tau a^2/2} \int_b^\infty \frac{e^{-(x-\tau a)^2/2\tau}}{\sqrt{2\pi \tau}} \, dx.$$

Make the change of variable $z = -(x - \tau a)/\sqrt{\tau}$ in the integral. Then you get,

$$E_{\hat{P}}[e^{\alpha X} 1_{\{X \geq b\}}] = e^{\tau a^2/2} \int_{-\infty}^{(\tau a-b)/\sqrt{\tau}} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \, dz = e^{\tau a^2/2} N\left(\frac{\tau a - b}{\sqrt{\tau}}\right). \diamond$$

**G. Girsanov’s theorem.**

Girsanov’s theorem is a change of measure technique for Brownian motion.

Let $W$ be a Brownian motion on $(\Omega, \mathcal{F}, P)$ and let $\{\mathcal{F}(t); t \geq 0\}$ be a filtration for $W$. Let $\theta(\cdot)$ be a stochastic process adapted to $\{\mathcal{F}(t); t \geq 0\}$, and set

$$\hat{W}(t) = \int_0^t \theta(s) \, ds + W(t), \quad t \geq 0.$$

Girsanov’s theorem tells us how to define a new probability measure $\hat{P}$ such that $\hat{W}$ is a Brownian motion on $(\Omega, \mathcal{F}, \hat{P})$, at least up to some fixed time.

Define,

$$Z(t) = \exp\{-\int_0^t \theta(s) \, dW(s) - \frac{1}{2} \int_0^t \theta^2(s) \, ds\}.$$

(Assume enough regularity on $\theta$ so that all integrals in this expression are defined.)

**Theorem 3** Assume that $E_P\left[\int_0^T \theta^2(s) Z^2(s) \, ds\right] < \infty$. Then $Z(t)$, $0 \leq t \leq T$ is a martingale with respect to $\{\mathcal{F}(t); t \geq 0\}$ and $P$, and so $E_P[Z(T)] = E_P[Z(0)] = 1$. If $\hat{P}$ is the probability measure, $\hat{P}(A) = E_P[1_A Z(T)]$, then $\hat{W}(t)$, $0 \leq t \leq T$ is a Brownian motion (up to time $T$) as a stochastic process on $(\Omega, \mathcal{F}, \hat{P})$, and $\{\mathcal{F}(t); t \geq 0\}$ is a filtration for $\hat{W}$.

Girsanov’s theorem is used to find risk-neutral measures for Black-Scholes price models and is used to do explicit computations of expectations.

**Example 7. Girsanov’s theorem and the Black-Scholes formula.** Let $r$ be a constant—when we come to pricing $r$ will denote the risk free interest rate. We shall rewrite the Black-Scholes models defined in Example 2. As usual $(\Omega, \mathcal{F}, P)$ denotes the underlying probability space. Thus,

$$dS(t) = S(t) [\alpha \, dt + \sigma \, dW(t)].$$
By Girsanov’s theorem, if
\[
\frac{d\tilde{P}}{dP} = \exp\left\{ \int_0^T \left( \frac{(r-\alpha)}{\sigma} \right) dt - \frac{1}{2} \int_0^T \left( \frac{(r-\alpha)}{\sigma} \right)^2 dt \right\} = e^{(r-\alpha)/\sigma W(T) - \frac{1}{2}(r-\alpha)^2 T},
\]
then \( \tilde{W}(t) = \frac{(\alpha-r)}{\sigma} t + W(t), 0 \leq t \leq T, \) is a Brownian motion on \((\Omega, \mathcal{F}, \tilde{P})\). By an easy computation, on this probability space,
\[
dS(t) = S(t) \left[ r \ dt + \sigma \ d\tilde{W}(t) \right].
\]
We will see later that this is the risk-neutral model for the Black-Scholes price process.

II. Risk-neutral pricing.

A. Market models.

A model for a market with one riskless asset and \(n\) risky assets will consist of the following:

(a) A probability space \((\Omega, \mathcal{F}, P)\) modeling future states of the economy, together with a filtration \(\{\mathcal{F}(t)\}\); for each \(t\), \(\mathcal{F}(t)\) represents the information available to all investors at time \(t\).

(b) A non-negative process \(R(\cdot) = \{R(t)\}\) that is adapted to \(\{\mathcal{F}(t); t \geq 0\}\); \(R(t)\) models the risk-free interest rate (compounded continuously) at time \(t\). The discount process associated to \(R(\cdot)\) is \(D(t) = \exp\{-\int_0^t R(s) \ ds\}\).

(c) Stochastic processes \(\{S_1(t)\}, \{S_2(t)\}, \ldots, \{S_n(t)\}\), each of which is adapted to \(\{\mathcal{F}(t); t \geq 0\}\); \(S_i(t)\) represents the price of risky asset \(i\) at time \(t\).

We shall say that the model is a generalized Black-Scholes model if there are \(d\) independent Brownian motions \(W_1, \ldots, W_d\) defined on \((\Omega, \mathcal{F}, P)\) such that

(i) \(\{\mathcal{F}(t); t \geq 0\}\) is a filtration for each \(W_i\);

(ii) There are adapted processes \(\alpha_i(\cdot)\) and \(\sigma_{ij}(\cdot)\) such that for each \(i\), \(S_i\) satisfies the stochastic differential equation,
\[
dS_i(t) = \alpha_i(t) S_i(t) \ dt + S_i(t) \sum_{j=1}^d \sigma_{ij}(t) \ dW_j(t)
\]
This is the type of model introduced in Chapter 5 of Shreve and the one that we will study until we introduce jump processes.
We shall use the phrase *standard Black-Scholes model* to describe the special case in which: (i) there is only one risky asset \( S \); (ii) \( \alpha, \sigma, \) and \( R(t) = r \) are constants; and (iii),
\[
dS(t) = \alpha S(t) \, dt + \sigma S(t) \, dW(t)
\]  
\[(28)\]

**B. Risk-neutral models.**

Consider a market model as defined in the previous section. We say that the model is risk-neutral, and that \( \mathbf{P} \) is a risk-neutral measure, if for each risky asset \( i \), \( 1 \leq i \leq n \),
\[
\{D(t)S_i(t); t \geq 0\} \text{ is a martingale with respect to } \{\mathcal{F}(t); t \geq 0\} \text{ and } \mathbf{P}.
\]  
\[(29)\]

When the model is risk-neutral, we shall use \( \tilde{\mathbf{P}} \) to denote the risk-neutral measure and \( \tilde{E} \) to denote expectation with respect to \( \tilde{\mathbf{P}} \).

*Example 8: Black-Scholes continued.* The Black-Scholes price process on the probability space \( (\Omega, \mathcal{F}, \tilde{\mathbf{P}}) \) constructed in Example 7 is risk-neutral, if the interest rate process is the constant \( r \). In this case \( D(t) = e^{-rt} \) and so the discounted asset price is \( e^{-rt}S(t) \). Now, \( S \) satisfies the equation (26), where \( \tilde{W} \) is a Brownian motion. So it follows directly from Example 4 in section I.D, with \( r \) in place of \( \alpha \) and \( \tilde{W} \) in place of \( W \), that \( e^{-rt}S(t) \) is a martingale.

*Example 9. The risk-neutral form of the generalized Black-Scholes model.* If model (33) admits a risk-neutral measure, \( \tilde{\mathbf{P}} \), then it has the following standard form, which you should know automatically:
\[
dS_i(t) = R(t)S_i(t) \, dt + S_i(t) \sum_{j=1}^{d} \sigma_{ij}(t) \, d\tilde{W}_j(t), \quad 1 \leq i \leq m,
\]  
\[(30)\]

where \( \tilde{W}_j, 1 \leq j \leq d \) are independent Brownian motions under \( \tilde{\mathbf{P}} \).

The motivation for introducing risk-neutral models is the following heuristic principle. Suppose we have a price model set up on \( (\Omega, \mathcal{F}, \mathbf{P}) \); if there is a measure \( \tilde{\mathbf{P}} \) which is equivalent to \( \mathbf{P} \) such that the model is risk-neutral on \( (\Omega, \mathcal{F}, \tilde{\mathbf{P}}) \), then the original model, the risk-neutral model, and any model equivalent to these, will not admit arbitrage. Here is the idea. We have already discussed in these notes, but it bears repeating. An arbitrage exists if there is some admissible trading strategy which produces no loss with probability one and a positive gain with some positive probability. Admissible strategies are those which are adapted to the filtration modeling the information available to the investor. Now, the discounted gain \( M(t) \) of any such trading strategy is essentially the yield obtained by betting on the increments of the discounted asset prices of the model. As explained above in section D,
if the discounted assets are martingales under $\tilde{\mathbb{P}}$, so also is $M(t)$. In particular, if $M(0) = 0$, then $\tilde{E}[M(t)] = 0$ for any $t \geq 0$. This means that $M$ cannot represent an arbitrage. For it it did, then at some time $t$ we must have $\tilde{\mathbb{P}}(M(t) \geq 0) = 1$ and $\tilde{\mathbb{P}}(M(t) > 0) > 0$; but this would imply that $\tilde{E}[M(t)] > 0$, which contradicts $\tilde{E}[M(t)] = 0$. Now, if $Q$ is any probability measure equivalent to $\tilde{\mathbb{P}}$, for example the original $\mathbb{P}$, an arbitrage under $Q$ would be an arbitrage under $\tilde{\mathbb{P}}$ since an event has positive probability under $Q$ if and only if it has positive probability under $\tilde{\mathbb{P}}$. Thus no model with a measure equivalent to $\tilde{\mathbb{P}}$ admits arbitrage if $\tilde{\mathbb{P}}$ is risk-neutral.

To turn this heuristic argument into a precise theorem requires a more precise formulation and technical conditions. These are provided by Shreve in Chapter 5 for the generalized Black-Scholes price model defined above in (33). The final result is stated as the First Fundamental Theorem of Asset Pricing on page 231.

Consider the generalized Black-Scholes model and let $\mathbb{P}$ denote its underlying probability measure. Shreve also gives conditions under which, using Girsanov’s theorem, one can show that there is a measure $\tilde{\mathbb{P}}$ such that $\tilde{\mathbb{P}} \sim \mathbb{P}$ and for which the model is risk neutral. He also proves the Second Fundamental Theorem of Asset Pricing: if the market is complete, that is, if it is possible to hedge every contingent claim (derivative security) based on the information in the filtration $\{\mathcal{F}(t); t \geq 0\}$, then the risk-neutral measure is unique. Moreover, he gives sufficient conditions for completeness for the generalized Black-Scholes model.

Example 7 shows that the Black-Scholes model admits a risk-neutral measure. In fact it is unique.

C. Risk-neutral pricing

Start with a market model as defined in section II.A, with probability measure $\mathbb{P}$. Consider a derivative security with payoff $V(T)$ written on the risky assets. Let $V(t)$ denote the price at which this claim is traded at any time before $t$. We shall derive a formula for the price using the principle of no arbitrage. The basic assumption is that there exists a risk-neutral measure $\tilde{\mathbb{P}}$ such that $\tilde{\mathbb{P}} \sim \mathbb{P}$. As we saw, this assumption implies that the underlying asset prices afford no arbitrage assumption. Assume also that there is a replicating portfolio, that is, the market is complete. This means there is a way to invest in the underlying assets so that the value of the portfolio at the time of expiry is the derivative payoff $V(T)$. Let $X(t)$ denote the value of this replicating portfolio for earlier times $t$; $X(t)$ is the amount of money we need to invest in order to replicate the payoff. There will be an arbitrage opportunity unless $V(t) = X(t)$. For example if $V(t) > X(t)$, we can short the derivative. This will give us $V(t)$ and we can invest $X(t)$ of this according to the replicating strategy to meet our obligations and invest the remainder $V(t) = X(t)$ at the risk free rate. We will end up at time $t$ with the profit $(V(t) - X(t))D(t)/D(T)$ and no loss. Now, by the heuristic principle that gains of trading on a martingale constitute a martingale $D(t)V(t) = D(t)X(t)$.
must be a martingale under the risk-neutral measure. Therefore
\[ V(t) = D^{-1}(t) \tilde{E}[D(T)V(T) | \mathcal{F}(t)] \] (31)

This is the risk-neutral pricing formula.

This formula provides one possible price for the derivative that does not allow arbitrage. In Chapter 5, Shreve shows by a hedging argument that, for a complete market for the generalized Black-Scholes model, there is a unique risk-neutral measure and a unique no-arbitrage price.

(An alternative argument which does not involve hedging is to think of the derivative as another asset being added to the market. Then \( D(t)V(t) \) must also be a martingale under the risk-neutral measure to avoid arbitrage. This argument does not invoke a replicating portfolio, but in the absence of such a portfolio, it is not clear what the price of a derivative should be from the point of view of the seller, because he or she will not be able to hedge it perfectly.)

In summary, we have arrived at a procedure for pricing derivatives. Begin with a market model. This should be constructed to capture the statistical properties of the actual underlying assets. Then look for a risk-neutral equivalent measure. If the market model is complete (basically, if the derivative payoffs depend only on the underlying asset values), then formula (31) provides the no-arbitrage price for any contingent claim.

To see how Girsanov’s theorem is used to define an equivalent risk-neutral model, consider the following one-dimensional set up. We are given a probability space \((\Omega, \mathcal{F}, \mathbf{P})\) and a Brownian motion \(W\) defined on this probability space. We let \(\{\mathcal{F}(t); t \geq 0\}\) denote the filtration generated by this Brownian motion; from the heuristic point of view, \(W\) will be the input responsible for all the randomness in the model. We are given also a risk-free rate process \(r(t), t \geq 0\), which may be random, but must be adapted to the filtration \(\{\mathcal{F}(t); t \geq 0\}\). The associated discount process is \(D(t) = \exp\{-\int_0^t r(s) \, ds\}\). Define a price process model by the usual equation,
\[ dS(t) = \alpha(t)S(t) \, dt + \sigma(t)S(t) \, dW(t), \]
where \(\alpha\) and \(\sigma\) are also process adapted to \(\{\mathcal{F}(t); t \geq 0\}\).

The discounted price process is \(D(t)S(t)\), and an application of Itô’s rule gives
\[ d[D(t)S(t)] = (\alpha(t) - r(t))[D(t)S(t)] \, dt + \sigma(t)[D(t)S(t)] \, dW(t). \]

Unless \(\alpha(t) = r(t)\) for all \(t\), \(D(t)S(t)\) is not a martingale, because the coefficient of \(dt\) is then not identically zero. Thus, the model we have set up is not risk-neutral.
Now suppose we define $\tilde{W}(t) = W(t) + \int_0^t \theta(s) \, ds$, for some process $\theta$ to be determined. Then $W(t) = \tilde{W}(t) - \int_0^t \theta(s) \, ds$, hence $dW(t) = d\tilde{W}(t) - \theta(t) \, dt$, and therefore, by substitution into the equation above,

$$d[D(t)S(t)] = (\alpha(t) - r(t) - \theta(t) \sigma(t))[D(t)S(t)] \, dt + \sigma(t)[D(t)S(t)] d\tilde{W}(t).$$

Now choose $\theta$ so that the coefficient of the first term is identically zero. By some algebra, this is done by setting

$$\theta(t) = \frac{\alpha(t) - r(t)}{\sigma(t)}. \tag{32}$$

Assume this makes sense (that is $\sigma(t)(\omega) \neq 0$ for all $t \geq 0$ and all $\omega$). Then

$$d[D(t)S(t)] = \sigma(t)[D(t)S(t)] d\tilde{W}(t).$$

But while there is now no $dt$ term, $\tilde{W}$ is not a Brownian motion, as a process on the original probability space. Here is where Girsanov’s Theorem comes to the rescue. It says that if we define

$$\tilde{P}(A) = E_P\left[1_A \exp\left\{-\int_0^T \theta(s) \, dW(s) - \int_0^T \theta^2(s) \, ds\right\}\right] \tag{33}$$

and if

$$E_P\left[\exp\left\{-\int_0^T \theta(s) \, dW(s) - \int_0^T \theta^2(s) \, ds\right\}\right] = 1 \tag{34}$$

then $\tilde{W}(t), 0 \leq t \leq T$ is a Brownian motion if $\tilde{P}$ is used as the underlying measure rather than $P$. And then, on the probability space $(\Omega, \mathcal{F}, \tilde{P})$, $D(t)S(t)$ is a martingale and $\tilde{P}$ defines a risk neutral measure up to time $T$. Then we can use (19) as the basis for no-arbitrage pricing of derivatives written on $S$ and expiring by time $T$.

Of course one needs to verify condition (34). This is not an issue we worry much about in Math Finance I, except to note that one simple condition for it to hold, which you will see in Shreve, is that $\theta$, as defined in (32), satisfies

$$|\theta(t)(\omega)| \leq K \quad \text{for all } \omega \text{ and for all } t \leq T,$$

where $K$ is some constant.

D. The Black-Scholes formula

As an example of risk-neutral pricing we derive the Black-Scholes formula for the price of a call option. This is the formula one derives by applying the risk-neutral pricing paradigm to the standard Black-Scholes price model. As in Example 7, let $r$ be the constant, risk-free interest rate, so that $D(t) = e^{-rt}$. As usual $\{\mathcal{F}(t); t \geq 0\}$
is the filtration for \( W \). Then, the risk-neutral version of the Black-Scholes model is that derived in Example 7 (see Example 8), and

\[ S(t) = S(0) \exp\{\sigma \tilde{W}(t) + (r - \frac{1}{2} \sigma^2)t\}. \] (35)

The notation \( \tilde{E} \) shall denote expectation with respect to the risk-neutral measure \( \tilde{P} \).

Let \( c(t) \) denote the price of a European call with payoff \( (S(T) - K)^+ \) at expiry \( T \). The risk-neutral pricing formula (31) says that

\[ V(t) = e^{-r(T-t)} \tilde{E}\left[ (S(T) - K)^+ \mid F(t) \right] \] (36)

This is a conditional expectation precisely of the type we saw in equation (21). We need only replace \( \alpha \) in this formula by \( r \), \( t \) by \( T \) and \( s \) by \( t \), and set \( h(y) = (y - K)^+ \). Rewriting the formula as an expectation shows that the option price is \( V(t) = c(t, S(t)) \), where

\[ c(t, x) = e^{-r(T-t)} \tilde{E}\left[ (S(T) - K)^+ \mid S(t) \right] = e^{-r(T-t)} \tilde{E}\left[ (xe^{\sigma Y + (r - \sigma^2/2)(T-t)} - K)^+ \mid S(t) \right], \] (37)

and \( Y \) is a normal random variable with mean 0 and variance \( T-t \).

Let \( b = \sigma^{-1} \ln \left( \frac{K}{x} \right) - (r - \sigma^2/2)(T - t) \). Then \( xe^{\sigma Y(t) + (r - \sigma^2/2)(T-t)} \geq K \) if and only if \( Y \geq b \), and hence

\[ \left( xe^{\sigma Y(t) + (r - \sigma^2/2)(T-t)} - K \right)^+ = \left( xe^{\sigma Y(t) + (r - \sigma^2/2)(T-t)} - K \right) 1_{\{Y \geq b\}} \]

Plug this into (37) and arrange terms: the result, after some minor simplifications and cancellations, is

\[ c(t, x) = e^{-(\sigma^2/2)(T-t)} \tilde{E}\left[ xe^{\sigma Y} 1_{\{Y \geq b\}} \right] - e^{-r(T-t)} K \tilde{E}\left[ 1_{\{Y \geq b\}} \right]. \]

Since the density of \( Y \) is \( \frac{e^{-y^2/2(T-t)}}{\sqrt{2\pi (T-t)}} \),

\[ \tilde{E}\left[ 1_{\{Y \geq b\}} \right] = \int_b^\infty \frac{e^{-x^2/2(T-t)}}{\sqrt{2\pi (T-t)}} \, dx = \int_{-\infty}^{-b/\sqrt{T-t}} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \, dz = N(-b/\sqrt{T-t}). \]

Here \( N \) is the standard normal cumulative distribution function, and the second equality is derived by the substitution \( z = -x/\sqrt{T-t} \).

As for the other term, we showed how to calculate \( \tilde{E}[e^{\sigma Y} 1_{\{Y \geq b\}}] \) in Corollary 1, equation (25). In this case, \( \tau \) in that formula equals \( T - t \). Hence,

\[ \tilde{E}\left[ xe^{\sigma Y} 1_{\{Y \geq b\}} \right] = xe^{(\sigma^2/2)(T-t)} N\left( \frac{\sigma(T-t) - b}{\sqrt{T-t}} \right). \]
Notice that these formulas depend on time only through the parameter \( \tau = T - t \).

In the notation of Shreve,

\[
\frac{\sigma(T-t) - b}{\sqrt{T-t}} = \frac{\ln(\frac{x}{K}) + (r + \sigma^2/2)(T-t)}{\sigma \sqrt{T-t}} = d_+(T-t, x)
\]

\[
-\frac{b}{\sqrt{T-t}} = \frac{\ln(\frac{x}{K}) + (r - \sigma^2/2)(T-t)}{\sigma \sqrt{T-t}} = d_-(T-t, x).
\]

Thus

\[
c(t, x) = xN(d_+(T-t, x)) - Ke^{-r(T-t)} N(d_-(T-t, x)).
\]