2

Probability Theory on Coin Toss Space

2.1 Finite Probability Spaces

A finite probability space is used to model a situation in which a random experiment with finitely many possible outcomes is conducted. In the context of the binomial model of the previous chapter, we tossed a coin a finite number of times. If, for example, we toss the coin three times, the set of all possible outcomes is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}. \tag{2.1.1}$$

Suppose that on each toss the probability of a head (either actual or risk-neutral) is p and the probability of a tail is $q = 1-p$. We assume the tosses are independent, and so the probabilities of the individual elements ω (sequences of three tosses $\omega = \omega_1\omega_2\omega_3$) in Ω are

$$\begin{aligned} &\mathbb{P}(HHH) = p^3,\ \mathbb{P}(HHT) = p^2q,\ \mathbb{P}(HTH) = p^2q,\ \mathbb{P}(HTT) = pq^2,\\ &\mathbb{P}(THH) = p^2q,\ \mathbb{P}(THT) = pq^2,\ \ \mathbb{P}(TTH) = pq^2,\ \ \mathbb{P}(TTT) = q^3. \end{aligned} \tag{2.1.2}$$

The subsets of Ω are called *events*, and these can often be described in words as well as in symbols. For example, the event

$$\begin{aligned} \text{“The first toss is a head”} &= \{\omega \in \Omega; \omega_1 = H\}\\ &= \{HHH, HHT, HTH, HTT\} \end{aligned}$$

has, as indicated, descriptions in both words and symbols. We determine the probability of an event by summing the probabilities of the elements in the event, i.e.,

$$\begin{aligned} \mathbb{P}(\text{First toss is a head}) &= \mathbb{P}(HHH) + \mathbb{P}(HHT) + \mathbb{P}(HTH) + \mathbb{P}(HTT)\\ &= (p^3 + p^2q) + (p^2q + pq^2)\\ &= p^2(p+q) + pq(p+q) \end{aligned}$$

$$
\begin{aligned}
&= p^2 + pq \\
&= p(p+q) \\
&= p.
\end{aligned}
\tag{2.1.3}
$$

Thus, the mathematics agrees with our intuition.

With mathematical models, it is easy to substitute our intuition for the mathematics, but this can lead to trouble. We should instead verify that the mathematics and our intuition agree; otherwise, either our intuition is wrong or our model is inadequate. If our intuition and the mathematics of a model do not agree, we should seek a reconciliation before proceeding. In the case of (2.1.3), we set out to build a model in which the probability of a head on each toss is p, we proposed doing this by defining the probabilities of the elements of Ω by (2.1.2), and we further defined the probability of an event (subset of Ω) to be the sum of the probabilities of the elements in the event. These definitions force us to carry out the computation (2.1.3) as we have done, and we need to do this computation in order to check that it gets the expected answer. Otherwise, we would have to rethink our mathematical model for the coin tossing.

We generalize slightly the situation just described, first by allowing Ω to be any finite set, and second by allowing some elements in Ω to have probability zero. These generalizations lead to the following definition.

Definition 2.1.1. *A* finite probability space *consists of a* sample space Ω *and a* probability measure $\mathbb{P}$. *The sample space* Ω *is a nonempty finite set and the probability measure* $\mathbb{P}$ *is a function that assigns to each element* ω *of* Ω *a number in* $[0,1]$ *so that*

$$\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1. \tag{2.1.4}$$

An event *is a subset of* Ω*, and we define the probability of an event* A *to be*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega). \tag{2.1.5}$$

As mentioned before, this is a model for some random experiment. The set Ω is the set of all possible outcomes of the experiment, $\mathbb{P}(\omega)$ is the probability that the particular outcome ω occurs, and $\mathbb{P}(A)$ is the probability that the outcome that occurs is in the set A. If $\mathbb{P}(A) = 0$, then the outcome of the experiment is sure not to be in A; if $\mathbb{P}(A) = 1$, then the outcome is sure to be in A. Because of (2.1.4), we have the equation

$$\mathbb{P}(\Omega) = 1, \tag{2.1.6}$$

i.e., the outcome that occurs is sure to be in the set Ω. Because $\mathbb{P}(\omega)$ can be zero for some values of ω, we are permitted to put in Ω even some outcomes of the experiment that are sure not to occur. It is clear from (2.1.5) that if A and B are disjoint subsets of Ω, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B). \tag{2.1.7}$$

2.2 Random Variables, Distributions, and Expectations

A random experiment generally generates numerical data. This gives rise to the concept of a random variable.

Definition 2.2.1. *Let $(\Omega, \mathbb{P})$ be a finite probability space. A* random variable *is a real-valued function defined on Ω. (We sometimes also permit a random variable to take the values $+\infty$ and $-\infty$.)*

Example 2.2.2 (Stock prices). Recall the space Ω of three independent coin-tosses (2.1.1). As in Figure 1.2.2 of Chapter 1, let us define stock prices by the formulas

$$S_0(\omega_1\omega_2\omega_3) = 4 \text{ for all } \omega_1\omega_2\omega_3 \in \Omega_3,$$

$$S_1(\omega_1\omega_2\omega_3) = \begin{cases} 8 & \text{if } \omega_1 = H, \\ 2 & \text{if } \omega_1 = T, \end{cases}$$

$$S_2(\omega_1\omega_2\omega_3) = \begin{cases} 16 & \text{if } \omega_1 = \omega_2 = H, \\ 4 & \text{if } \omega_1 \neq \omega_2, \\ 1 & \text{if } \omega_1 = \omega_2 = T, \end{cases}$$

$$S_3(\omega_1\omega_2\omega_3) = \begin{cases} 32 & \text{if } \omega_1 = \omega_2 = \omega_3 = H, \\ 8 & \text{if there are two heads and one tail,} \\ 2 & \text{if there is one head and two tails,} \\ .50 & \text{if } \omega_1 = \omega_2 = \omega_3 = T. \end{cases}$$

Here we have written the arguments of S_0, S_1, S_2, and S_3 as $\omega_1\omega_2\omega_3$, even though some of these random variables do not depend on all the coin tosses. In particular, S_0 is actually not random because it takes the value 4, regardless of how the coin tosses turn out; such a random variable is sometimes called a *degenerate random variable.* □

It is customary to write the argument of random variables as ω, even when ω is a sequence such as $\omega = \omega_1\omega_2\omega_3$. We shall use these two notations interchangeably. It is even more common to write random variables without any arguments; we shall switch to that practice presently, writing S_3, for example, rather than $S_3(\omega_1\omega_2\omega_3)$ or $S_3(\omega)$.

According to Definition 2.2.1, a random variable is a function that maps a sample space Ω to the real numbers. The *distribution* of a random variable is a specification of the probabilities that the random variable takes various values. *A random variable is not a distribution*, and *a distribution is not a random variable.* This is an important point when we later switch between the actual probability measure, which one would estimate from historical data, and the risk-neutral probability measure. The change of measure will change distributions of random variables but not the random variables themselves. We make this distinction clear with the following example.

Example 2.2.3. Toss a coin three times, so the set of possible outcomes is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Define the random variables

$$X = \text{Total number of heads}, \qquad Y = \text{Total number of tails}.$$

In symbols,

$$\begin{aligned} X(HHH) &= 3, \\ X(HHT) = X(HTH) = X(THH) &= 2, \\ X(HTT) = X(THT) = X(TTH) &= 1, \\ X(TTT) &= 0, \\ \\ Y(TTT) &= 3, \\ Y(TTH) = Y(THT) = Y(HTT) &= 2, \\ Y(THH) = Y(HTH) = Y(HHT) &= 1, \\ Y(HHH) &= 0. \end{aligned}$$

We do not need to know probabilities of various outcomes in order to specify these random variables. However, once we specify a probability measure on Ω, we can determine the distributions of X and Y. For example, if we specify the probability measure $\widetilde{\mathbb{P}}$ under which the probability of head on each toss is $\frac{1}{2}$ and the probability of each element in Ω is $\frac{1}{8}$, then

$$\begin{aligned} &\widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = 0\} = \widetilde{\mathbb{P}}\{TTT\} = \frac{1}{8}, \\ &\widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = 1\} = \widetilde{\mathbb{P}}\{HTT, THT, TTH\} = \frac{3}{8}, \\ &\widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = 2\} = \widetilde{\mathbb{P}}\{HHT, HTH, THH\} = \frac{3}{8}, \\ &\widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = 3\} = \widetilde{\mathbb{P}}\{HHH\} = \frac{1}{8}. \end{aligned}$$

We shorten the cumbersome notation $\widetilde{\mathbb{P}}\{\omega \in \Omega; X(\omega) = j\}$ to simply $\widetilde{\mathbb{P}}\{X = j\}$. It is helpful to remember, however, that the notation $\widetilde{\mathbb{P}}\{X = j\}$ refers to the probability of a subset of Ω, the set of elements ω for which $X(\omega) = j$. Under $\widetilde{\mathbb{P}}$, the probability that X takes the four values 0, 1, 2, and 3 are

$$\begin{aligned} \widetilde{\mathbb{P}}\{X = 0\} = \frac{1}{8},\ \widetilde{\mathbb{P}}\{X = 1\} = \frac{3}{8}, \\ \widetilde{\mathbb{P}}\{X = 2\} = \frac{3}{8},\ \widetilde{\mathbb{P}}\{X = 3\} = \frac{1}{8}. \end{aligned}$$

This table of probabilities where X takes its various values records the *distribution* of X under $\widetilde{\mathbb{P}}$.

The random variable Y is different from X because it counts tails rather than heads. However, under $\widetilde{\mathbb{P}}$, the distribution of Y is the same as the distribution of X:

$$\widetilde{\mathbb{P}}\{Y=0\}=\frac{1}{8},\ \widetilde{\mathbb{P}}\{Y=1\}=\frac{3}{8},$$
$$\widetilde{\mathbb{P}}\{Y=2\}=\frac{3}{8},\ \widetilde{\mathbb{P}}\{Y=3\}=\frac{1}{8}.$$

The point here is that the random variable is a function defined on Ω, whereas its distribution is a tabulation of probabilities that the random variable takes various values. A random variable is not a distribution.

Suppose, moreover, that we choose a probability measure $\mathbb{P}$ for Ω that corresponds to a $\frac{2}{3}$ probability of head on each toss and a $\frac{1}{3}$ probability of tail. Then

$$\mathbb{P}\{X=0\}=\frac{1}{27},\ \mathbb{P}\{X=1\}=\frac{6}{27},$$
$$\mathbb{P}\{X=2\}=\frac{12}{27},\ \mathbb{P}\{X=3\}=\frac{8}{27}.$$

The random variable X has a different distribution under $\mathbb{P}$ than under $\widetilde{\mathbb{P}}$. It is the same random variable, counting the total number of heads, regardless of the probability measure used to determine its distribution. This is the situation we encounter later when we consider an asset price under both the actual and the risk-neutral probability measures.

Incidentally, although they have the same distribution under $\widetilde{\mathbb{P}}$, the random variables X and Y have different distributions under $\mathbb{P}$. Indeed,

$$\mathbb{P}\{Y=0\}=\frac{8}{27},\quad \mathbb{P}\{Y=1\}=\frac{12}{27},$$
$$\mathbb{P}\{Y=2\}=\frac{6}{27},\quad \mathbb{P}\{Y=3\}=\frac{1}{27}. \qquad \square$$

Definition 2.2.4. *Let X be a random variable defined on a finite probability space $(\Omega,\mathbb{P})$. The* expectation *(or* expected value*) of X is defined to be*

$$\mathbb{E}X=\sum_{\omega\in\Omega}X(\omega)\mathbb{P}(\omega).$$

When we compute the expectation using the risk-neutral probability measure $\widetilde{\mathbb{P}}$, we use the notation

$$\widetilde{\mathbb{E}}X=\sum_{\omega\in\Omega}X(\omega)\widetilde{\mathbb{P}}(\omega).$$

The variance *of X is*

$$\mathrm{Var}(X)=\mathbb{E}\Big[(X-\mathbb{E}X^2\Big].$$

It is clear from its definition that expectation is linear: if X and Y are random variables and c_1 and c_2 are constants, then

$$\mathbb{E}(c_1X + c_2Y) = c_1\mathbb{E}X + c_2\mathbb{E}Y.$$

In particular, if $\ell(x) = ax+b$ is a linear function of a dummy variable x (a and b are constants), then $\mathbb{E}[\ell(X)] = \ell(\mathbb{E}X)$. When dealing with convex functions, we have the following inequality.

Theorem 2.2.5 (Jensen's inequality). *Let X be a random variable on a finite probability space, and let $\varphi(x)$ be a convex function of a dummy variable x. Then*

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}X).$$

Proof: We first argue that a convex function is the maximum of all linear functions that lie below it; i.e., for every $x \in \mathbb{R}$,

$$\varphi(x) = \max\{\ell(x); \ell \text{ is linear and } \ell(y) \leq \varphi(y) \text{ for all } y \in \mathbb{R}\}. \tag{2.2.1}$$

Since we are only considering linear functions that lie below φ, it is clear that

$$\varphi(x) \geq \max\{\ell(x); \ell \text{ is linear and } \ell(y) \leq \varphi(y) \text{ for all } y \in \mathbb{R}\}.$$

On the other hand, let x be an arbitrary point in $\mathbb{R}$. Because φ is convex, there is always a linear function ℓ that lies below φ and for which $\varphi(x) = \ell(x)$ for this particular x. This is called a *support line of φ at x* (see Figure 2.2.1). Therefore,

$$\varphi(x) \leq \max\{\ell(x); \ell \text{ is linear and } \ell(y) \leq \varphi(y) \text{ for all } y \in \mathbb{R}\}.$$

This establishes (2.2.1). Now let ℓ be a linear function lying below φ. We have

$$\mathbb{E}\left[\varphi(X)\right] \geq \mathbb{E}\left[\ell(X)\right] = \ell(\mathbb{E}X).$$

Since this inequality holds for every linear function ℓ lying below φ, we may take the maximum on the right-hand side over all such ℓ and obtain

$$\begin{aligned}\mathbb{E}\left[\varphi(X)\right] &\geq \max\{\ell(\mathbb{E}X); \ell \text{ is linear and } \ell(y) \leq \varphi(y) \text{ for all } y \in \mathbb{R}\}\\ &= \varphi(\mathbb{E}X).\end{aligned}$$

□

One consequence of Jensen's inequality is that

$$\mathbb{E}\left[X^2\right] \geq (\mathbb{E}X)^2.$$

We can also obtain this particular consequence of Jensen's inequality from the formula

$$0 \leq \mathbb{E}\Big[(X - \mathbb{E}X^2)\Big] = \mathbb{E}\Big[X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2\Big] = \mathbb{E}[X^2] - (\mathbb{E}X)^2.$$

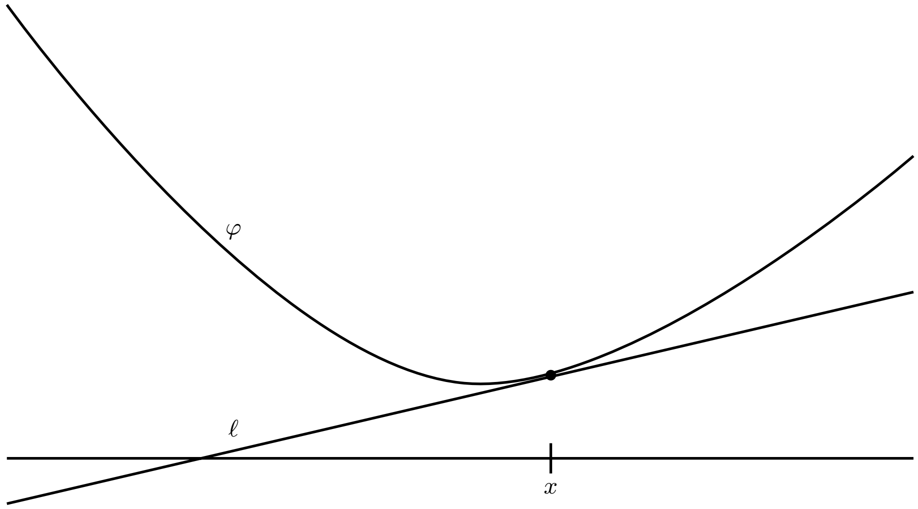

Fig. 2.2.1. Support line of φ at x.

2.3 Conditional Expectations

In the binomial pricing model of Chapter 1, we chose risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$ by the formula (1.1.8), which we repeat here:

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}. \tag{2.3.1}$$

It is easily checked that these probabilities satisfy the equation

$$\frac{\tilde{p}u + \tilde{q}d}{1+r} = 1. \tag{2.3.2}$$

Consequently, at every time n and for every sequence of coin tosses $\omega_1 \dots \omega_n$, we have

$$S_n(\omega_1 \dots \omega_n) = \frac{1}{1+r}\Big[\tilde{p}S_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q}S_{n+1}(\omega_1 \dots \omega_n T)\Big] \tag{2.3.3}$$

(i.e., the stock price at time n is the discounted weighted average of the two possible stock prices at time $n+1$, where $\tilde{p}$ and $\tilde{q}$ are the weights used in the averaging). To simplify notation, we define

$$\widetilde{\mathbb{E}}_n[S_{n+1}](\omega_1 \dots \omega_n) = \tilde{p}S_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q}S_{n+1}(\omega_1 \dots \omega_n T) \tag{2.3.4}$$

so that we may rewrite (2.3.3) as

$$S_n = \frac{1}{1+r}\widetilde{\mathbb{E}}_n[S_{n+1}], \tag{2.3.5}$$

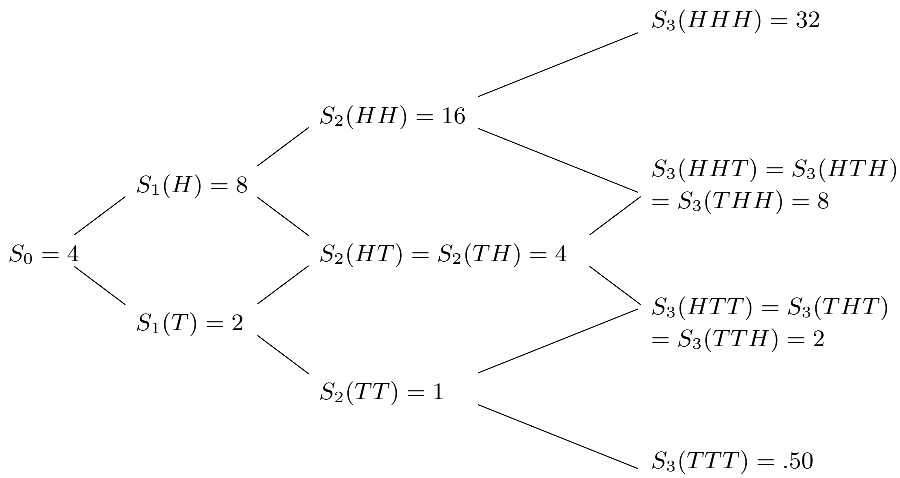

Fig. 2.3.1. A three-period model.

and we call $\widetilde{\mathbb{E}}_n[S_{n+1}]$ the *conditional expectation of* S_{n+1} *based on the information at time* n. The conditional expectation can be regarded as an estimate of the value of S_{n+1} based on knowledge of the first n coin tosses.

For example, in Figure 2.3.1 and using the risk-neutral probabilities $\tilde{p} = \tilde{q} = \frac{1}{2}$, we have $\widetilde{\mathbb{E}}_1[S_2](H) = 10$ and $\widetilde{\mathbb{E}}_1[S_2](T) = 2.50$. When we write simply $\widetilde{\mathbb{E}}_1[S_2]$ without specifying whether the first coin toss results in head or tail, we have a quantity whose value, not known at time zero, will be determined by the random experiment of coin tossing. According to Definition 2.2.1, such a quantity is a random variable.

More generally, whenever X is a random variable depending on the first N coin tosses, we can estimate X based on information available at an earlier time $n \leq N$. The following definition generalizes (2.3.4).

Definition 2.3.1. *Let n satisfy $1 \leq n \leq N$, and let $\omega_1 \dots \omega_n$ be given and, for the moment, fixed. There are 2^{N-n} possible continuations $\omega_{n+1} \dots \omega_N$ of the sequence fixed $\omega_1 \dots \omega_n$. Denote by $\#H(\omega_{n+1} \dots \omega_N)$ the number of heads in the continuation $\omega_{n+1} \dots \omega_N$ and by $\#T(\omega_{n+1} \dots \omega_N)$ the number of tails. We define*

$$\begin{aligned} &\widetilde{\mathbb{E}}_n[X](\omega_1 \dots \omega_n) \\ &\quad = \sum_{\omega_{n+1} \dots \omega_N} \tilde{p}^{\#H(\omega_{n+1} \dots \omega_N)} \tilde{q}^{\#T(\omega_{n+1} \dots \omega_N)} X(\omega_1 \dots \omega_n \omega_{n+1} \dots \omega_N) \end{aligned} \tag{2.3.6}$$

and call $\widetilde{\mathbb{E}}_n[X]$ the conditional expectation of X based on the information at time n.

Based on what we know at time zero, the conditional expectation $\widetilde{\mathbb{E}}_n[X]$ is random in the sense that its value depends on the first n coin tosses, which we do not know until time n. For example, in Figure 2.3.1 and using $\tilde{p} = \tilde{q} = \frac{1}{2}$, we obtain

$$\widetilde{\mathbb{E}}_1[S_3](H) = 12.50, \quad \widetilde{\mathbb{E}}_1[S_3](T) = 3.125,$$

so $\widetilde{\mathbb{E}}_1[S_3]$ is a random variable.

Definition 2.3.1 continued *The two extreme cases of conditioning are $\widetilde{\mathbb{E}}_0[X]$, the conditional expectation of X based on no information, which we define by*

$$\widetilde{\mathbb{E}}_0[X] = \widetilde{\mathbb{E}}X, \tag{2.3.7}$$

and $\widetilde{\mathbb{E}}_N[X]$, the conditional expectation of X based on knowledge of all N coin tosses, which we define by

$$\widetilde{\mathbb{E}}_N[X] = X. \tag{2.3.8}$$

The conditional expectations above have been computed using the risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$. This is indicated by the $\widetilde{\ }$ appearing in the notation $\widetilde{\mathbb{E}}_n$. Of course, conditional expectations can also be computed using the actual probabilities p and q, and these will be denoted by $\mathbb{E}_n$.

Regarded as random variables, conditional expectations have five fundamental properties, which we will use extensively. These are listed in the following theorem. We state them for conditional expectations computed under the actual probabilities, and the analogous results hold for conditional expectations computed under the risk-neutral probabilities.

Theorem 2.3.2 (Fundamental properties of conditional expectations). *Let N be a positive integer, and let X and Y be random variables depending on the first N coin tosses. Let $0 \leq n \leq N$ be given. The following properties hold.*

(i) ***Linearity of conditional expectations****. For all constants c_1 and c_2, we have*

$$\mathbb{E}_n[c_1X + c_2Y] = c_1\mathbb{E}_n[X] + c_2\mathbb{E}_n[Y].$$

(ii) ***Taking out what is known****. If X actually depends only on the first n coin tosses, then*

$$\mathbb{E}_n[XY] = X \cdot \mathbb{E}_n[Y].$$

(iii) ***Iterated conditioning****. If $0 \leq n \leq m \leq N$, then*

$$\mathbb{E}_n\Big[\mathbb{E}_m[X]\Big] = \mathbb{E}_n[X].$$

In particular, $\mathbb{E}\Big[\mathbb{E}_m[X]\Big] = \mathbb{E}X$.

(iv) ***Independence****. If X depends only on tosses $n+1$ through N, then*

$$\mathbb{E}_n[X] = \mathbb{E}X.$$

(v) ***Conditional Jensen's inequality****. If $\varphi(x)$ is a convex function of the dummy variable x, then*

$$\mathbb{E}_n[\varphi(X)] \geq \varphi(\mathbb{E}_n[X]).$$

The proof of Theorem 2.3.2 is provided in the appendix. We illustrate the first four properties of the theorem with examples based on Figure 2.3.1 using the probabilities $p = \frac{2}{3}$, $q = \frac{1}{3}$. The fifth property, the conditional Jensen's inequality, follows from linearity of conditional expectations in the same way that Jensen's inequality for expectations follows from linearity of expectations (see the proof of Theorem 2.2.5).

Example 2.3.3 (Linearity of conditional expectations). With $p = \frac{2}{3}$ and $q = \frac{1}{3}$ in Figure 2.3.1, we compute

$$\begin{aligned}\mathbb{E}_1[S_2](H) &= \frac{2}{3}\cdot 16 + \frac{1}{3}\cdot 4 \;=\; 12,\\ \mathbb{E}_1[S_3](H) &= \frac{4}{9}\cdot 32 + \frac{2}{9}\cdot 8 + \frac{2}{9}\cdot 8 + \frac{1}{9}\cdot 2 \;=\; 18,\end{aligned}$$

and consequently $\mathbb{E}_1[S_2](H) + \mathbb{E}_1[S_3](H) = 12 + 18 = 30$. But also

$$\mathbb{E}_1[S_2+S_3](H) = \frac{4}{9}(16+32) + \frac{2}{9}(16+8) + \frac{2}{9}(4+8) + \frac{1}{9}(4+2) = 30.$$

A similar computation shows that

$$\mathbb{E}_1[S_2+S_3](T) = 7.50 = \mathbb{E}_1[S_2](T) + \mathbb{E}_1[S_3](T).$$

In conclusion, regardless of the outcome of the first coin toss,

$$\mathbb{E}_1[S_2+S_3] = \mathbb{E}_1[S_2] + E_1[S_3].$$

Example 2.3.4 (Taking out what is known). We first recall from Example 2.3.3 that

$$\mathbb{E}_1[S_2](H) = \frac{2}{3}\cdot 16 + \frac{1}{3}\cdot 4 = 12.$$

If we now want to estimate the product S_1S_2 based on the information at time one, we can factor out the S_1, as seen by the following computation:

$$\mathbb{E}_1[S_1S_2](H) = \frac{2}{3}\cdot 128 + \frac{1}{3}\cdot 32 = 96 = 8\cdot 12 = S_1(H)\mathbb{E}_1[S_2](H).$$

A similar computation shows that

$$E_1[S_1S_2](T) = 6 = S_1(T)\mathbb{E}_1[S_2](T).$$

In conclusion, regardless of the outcome of the first toss,

$$\mathbb{E}_1[S_1S_2] = S_1\mathbb{E}_1[S_2].$$

Example 2.3.5 (Iterated conditioning). We first estimate S_3 based on the information at time two:

$$\begin{aligned}
\mathbb{E}_2[S_3](HH) &= \frac{2}{3}\cdot 32 + \frac{1}{3}\cdot 8 = 24,\\
\mathbb{E}_2[S_3](HT) &= \frac{2}{3}\cdot 8 + \frac{1}{3}\cdot 2 \ = 6,\\
\mathbb{E}_2[S_3](TH) &= \frac{2}{3}\cdot 8 + \frac{1}{3}\cdot 2 \ = 6,\\
\mathbb{E}_2[S_3](TT) &= \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot \frac{1}{2} \ = 1.50.
\end{aligned}$$

We now estimate the estimate, based on the information at time one:

$$\begin{aligned}
\mathbb{E}_1\Big[\mathbb{E}_2[S_3]\Big](H) &= \frac{2}{3}\cdot \mathbb{E}_2[S_3](HH) + \frac{1}{3}\cdot \mathbb{E}_2[S_3](HT)\\
&= \frac{2}{3}\cdot 24 + \frac{1}{3}\cdot 6 \ = \ 18,\\
\mathbb{E}_1\Big[\mathbb{E}_2[S_3]\Big](T) &= \frac{2}{3}\cdot \mathbb{E}_2[S_3](TH) + \frac{1}{3}\mathbb{E}_2[S_3](TT)\\
&= \frac{2}{3}\cdot 6 + \frac{1}{3}\cdot 1.50 \ = \ 4.50.
\end{aligned}$$

The estimate of the estimate is an average of averages, and it is not surprising that we can get the same result by a more comprehensive averaging. This more comprehensive averaging occurs when we estimate S_3 directly based on the information at time one:

$$\begin{aligned}
E_1[S_3](H) &= \frac{4}{9}\cdot 32 + \frac{2}{9}\cdot 8 + \frac{2}{9}\cdot 8 + \frac{1}{9}\cdot 2 = 18,\\
E_1[S_3](T) &= \frac{4}{9}\cdot 8 + \frac{2}{9}\cdot 2 + \frac{2}{9}\cdot 2 + \frac{1}{9}\cdot \frac{1}{2} \ = 4.50.
\end{aligned}$$

In conclusion, regardless of the outcome of the first toss, we have

$$\mathbb{E}_1\Big[\mathbb{E}_2[S_3]\Big] = \mathbb{E}_1[S_3].$$

Example 2.3.6 (Independence). The quotient $\frac{S_2}{S_1}$ takes either the value 2 or $\frac{1}{2}$, depending on whether the second coin toss results in head or tail, respectively. In particular, $\frac{S_2}{S_1}$ does not depend on the first coin toss. We compute

$$\begin{aligned}
\mathbb{E}_1\left[\frac{S_2}{S_1}\right](H) &= \frac{2}{3}\cdot \frac{S_2(HH)}{S_1(H)} + \frac{1}{3}\cdot \frac{S_2(HT)}{S_1(H)}\\
&= \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot \frac{1}{2} \ = \ \frac{3}{2},\\
\mathbb{E}_1\left[\frac{S_2}{S_1}\right](T) &= \frac{2}{3}\cdot \frac{S_2(TH)}{S_1(T)} + \frac{1}{3}\cdot \frac{S_2(TT)}{S_1(T)}\\
&= \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot \frac{1}{2} \ = \ \frac{3}{2}.
\end{aligned}$$

We see that $\mathbb{E}_1\left[\frac{S_2}{S_1}\right]$ does not depend on the first coin toss (is not really random) and in fact is equal to

$$\mathbb{E}\frac{S_2}{S_1} = \frac{2}{3}\cdot 2 + \frac{1}{3}\cdot\frac{1}{2} = \frac{3}{2}.$$

2.4 Martingales

In the binomial pricing model of Chapter 1, we chose risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$ so that at every time n and for every coin toss sequence $\omega_1 \dots \omega_n$ we have (2.3.3). In terms of the notation for conditional expectations introduced in Section 2.3, this fact can be written as (2.3.5). If we divide both sides of (2.3.5) by $(1+r)^n$, we get the equation

$$\frac{S_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{S_{n+1}}{(1+r)^{n+1}}\right]. \tag{2.4.1}$$

It does not matter in this model whether we write the term $\frac{1}{(1+r)^{n+1}}$ inside or outside the conditional expectation because it is constant (see Theorem 2.3.2(i)). In models with random interest rates, it would matter; we shall follow the practice of writing this term inside the conditional expectation since that is the way it would be written in models with random interest rates.

Equation (2.4.1) expresses the key fact that under the risk-neutral measure, for a stock that pays no dividend, the *best estimate based on the information at time n of the value of the discounted stock price at time $n+1$ is the discounted stock price at time n*. The risk-neutral probabilities are chosen to enforce this fact. Processes that satisfy this condition are called *martingales*. We give a formal definition of martingale under the actual probabilities p and q; the definition of martingale under the risk-neutral probabilities $\tilde{p}$ and $\tilde{q}$ is obtained by replacing $\mathbb{E}_n$ by $\widetilde{\mathbb{E}}_n$ in (2.4.2).

Definition 2.4.1. *Consider the binomial asset-pricing model. Let $M_0, M_1, \dots, M_N$ be a sequence of random variables, with each M_n depending only on the first n coin tosses (and M_0 constant). Such a sequence of random variables is called an* adapted stochastic process.

(i) If

$$M_n = \mathbb{E}_n[M_{n+1}],\ n = 0, 1, \dots, N-1, \tag{2.4.2}$$

we say this process is a martingale.

(ii) If

$$M_n \le \mathbb{E}_n[M_{n+1}],\ n = 0, 1, \dots, N-1,$$

we say the process is a submartingale *(even though it may have a tendency to increase);*

(iii) If

$$M_n \geq \mathbb{E}_n[M_{n+1}],\ n = 0, 1, \ldots, N-1,$$

we say the process is a supermartingale *(even though it may have a tendency to decrease).*

Remark 2.4.2. The martingale property in (2.4.2) is a "one-step-ahead" condition. However, it implies a similar condition for any number of steps. Indeed, if $M_0, M_1, \ldots, M_N$ is a martingale and $n \leq N-2$, then the martingale property (2.4.2) implies

$$M_{n+1} = \mathbb{E}_{n+1}[M_{n+2}].$$

Taking conditional expectations on both sides based on the information at time n and using the iterated conditioning property (iii) of Theorem 2.3.2, we obtain

$$\mathbb{E}_n[M_{n+1}] = \mathbb{E}_n\Big[\mathbb{E}_{n+1}[M_{n+2}]\Big] = E_n[M_{n+2}].$$

Because of the martingale property (2.4.2), the left-hand side is M_n, and we thus have the "two-step-ahead" property

$$M_n = \mathbb{E}_n[M_{n+2}].$$

Iterating this argument, we can show that whenever $0 \leq n \leq m \leq N$,

$$M_n = \mathbb{E}_n[M_m]. \tag{2.4.3}$$

One might call this the "multistep-ahead" version of the martingale property.

Remark 2.4.3. The expectation of a martingale is constant over time, i.e., if $M_0, M_1, \ldots, M_N$ is a martingale, then

$$M_0 = \mathbb{E}M_n,\ n = 0, 1, \ldots, N. \tag{2.4.4}$$

Indeed, if $M_0, M_1, \ldots, M_N$ is a martingale, we may take expectations on both sides of (2.4.2), using Theorem 2.3.2(iii), and obtain $\mathbb{E}M_n = \mathbb{E}[M_{n+1}]$ for every n. It follows that

$$\mathbb{E}M_0 = \mathbb{E}M_1 = \mathbb{E}M_2 = \cdots = \mathbb{E}M_{N-1} = \mathbb{E}M_N.$$

But M_0 is not random, so $M_0 = \mathbb{E}M_0$, and (2.4.4) follows. □

In order to have a martingale, the equality in (2.4.2) must hold for all possible coin toss sequences. The stock price process in Figure 2.3.1 would be a martingale if the probability of an up move were $\hat{p} = \frac{1}{3}$ and the probability of a down move were $\hat{q} = \frac{2}{3}$ because, at every node in the tree in Figure 2.3.1, the stock price shown would then be the average of the two possible subsequent stock prices averaged with these weights. For example,

$$S_1(T) = 2 = \frac{1}{3} \cdot S_2(TH) + \frac{2}{3} \cdot S_2(TT).$$

A similar equation would hold at all other nodes in the tree, and therefore we would have a martingale under these probabilities.

A martingale has no tendency to rise or fall since the average of its next period values is always its value at the current time. Stock prices have a tendency to rise and, indeed, should rise on average faster than the money market in order to compensate investors for their inherent risk. In Figure 2.3.1 more realistic choices for p and q are $p = \frac{2}{3}$ and $q = \frac{1}{3}$. With these choices, we have

$$\mathbb{E}_n[S_{n+1}] = \frac{3}{2} S_n$$

at every node in the tree (i.e., on average, the next period stock price is 50% higher than the current stock price). This growth rate exceeds the 25% interest rate we have been using in this model, as it should. In particular, with $p = \frac{2}{3}$, $q = \frac{1}{3}$, and $r = \frac{1}{4}$, the discounted stock price has a tendency to rise. Note that when $r = \frac{1}{4}$, we have $\frac{1}{1+r} = \frac{4}{5}$, so the discounted stock price at time n is $\left(\frac{4}{5}\right)^n S_n$. We compute

$$\mathbb{E}_n\left[\left(\frac{4}{5}\right)^{n+1} S_{n+1}\right] = \left(\frac{4}{5}\right)^{n+1} \mathbb{E}_n[S_{n+1}] = \left(\frac{4}{5}\right)^n \cdot \frac{4}{5} \cdot \frac{3}{2} \cdot S_n \geq \left(\frac{4}{5}\right)^n S_n.$$

The discounted stock price is a submartingale under the actual probabilities $p = \frac{2}{3}$, $q = \frac{1}{3}$. This is typically the case in real markets.

The risk-neutral probabilities, on the other hand, are chosen to make the discounted stock price be a martingale. In Figure 2.3.1 with $\tilde{p} = \tilde{q} = \frac{1}{2}$, one can check that the martingale equation

$$\widetilde{\mathbb{E}}_n\left[\left(\frac{4}{5}\right)^{n+1} S_{n+1}\right] = \left(\frac{4}{5}\right)^n S_n \tag{2.4.5}$$

holds at every node. The following theorem shows that this example is representative.

Theorem 2.4.4. *Consider the general binomial model with* $0 < d < 1+r < u$. *Let the risk-neutral probabilities be given by*

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}.$$

Then, under the risk-neutral measure, the discounted stock price is a martingale, i.e., equation (2.4.1) holds at every time n *and for every sequence of coin tosses.*

We give two proofs of this theorem, an elementary one, which does not rely on Theorem 2.3.2, and a deeper one, which does rely on Theorem 2.3.2. The second proof will later be adapted to continuous-time models.

Note in Theorem 2.4.4 that the stock does not pay a dividend. For a dividend-paying stock, the situation is described in Exercise 2.10.

First Proof: Let n and $\omega_1 \ldots \omega_n$ be given. Then

$$\begin{aligned}
&\widetilde{\mathbb{E}}_n \left[\frac{S_{n+1}}{(1+r)^{n+1}} \right] (\omega_1 \ldots \omega_n) \\
&\quad = \frac{1}{(1+r)^n} \cdot \frac{1}{1+r} \left[\tilde{p} S_{n+1}(\omega_1 \ldots \omega_n H) + \tilde{q} S_{n+1}(\omega_1 \ldots \omega_n T) \right] \\
&\quad = \frac{1}{(1+r)^n} \cdot \frac{1}{1+r} \left[\tilde{p} u S_n(\omega_1 \ldots \omega_n) + \tilde{q} d S_n(\omega_1 \ldots \omega_n) \right] \\
&\quad = \frac{S_n(\omega_1 \ldots \omega_n)}{(1+r)^n} \cdot \frac{\tilde{p}u + \tilde{q}d}{1+r} \\
&\quad = \frac{S_n(\omega_1 \ldots \omega_n)}{(1+r)^n}.
\end{aligned}$$

Second Proof: Note that $\frac{S_{n+1}}{S_n}$ depends only on the $(n+1)$st coin toss. Using the indicated properties from Theorem 2.3.2, we compute

$$\begin{aligned}
\widetilde{\mathbb{E}}_n \left[\frac{S_{n+1}}{(1+r)^{n+1}} \right] &= \widetilde{\mathbb{E}}_n \left[\frac{S_n}{(1+r)^{n+1}} \cdot \frac{S_{n+1}}{S_n} \right] \\
&= \frac{S_n}{(1+r)^n} \widetilde{\mathbb{E}}_n \left[\frac{1}{1+r} \cdot \frac{S_{n+1}}{S_n} \right] \\
&\qquad \text{(Taking out what is known)} \\
&= \frac{S_n}{(1+r)^n} \cdot \frac{1}{1+r} \widetilde{\mathbb{E}} \frac{S_{n+1}}{S_n} \\
&\qquad \text{(Independence)} \\
&= \frac{S_n}{(1+r)^n} \frac{\tilde{p}u + \tilde{q}d}{1+r} \\
&= \frac{S_n}{(1+r)^n}.
\end{aligned}$$

□

In a binomial model with N coin tosses, we imagine an investor who at each time n takes a position of Δ_n shares of stock and holds this position until time $n+1$, when he takes a new position of Δ_{n+1} shares. The portfolio rebalancing at each step is financed by investing or borrowing, as necessary, from the money market. The "portfolio variable" Δ_n may depend on the first n coin tosses, and Δ_{n+1} may depend on the first $n+1$ coin tosses. In other words, the *portfolio process* $\Delta_0, \Delta_1, \ldots, \Delta_{N-1}$ is *adapted*, in the sense of Definition 2.4.1. If the investor begins with initial wealth X_0, and X_n denotes his wealth at each time n, then the evolution of his wealth is governed by the *wealth equation* (1.2.14) of Chapter 1, which we repeat here:

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n), \; n = 0, 1, \ldots, N-1. \tag{2.4.6}$$

Note that each X_n depends only on the first n coin tosses (i.e., the *wealth process* is adapted).

We may inquire about the average rate of growth of the investor's wealth. If we mean the average under the actual probabilities, the answer depends on the portfolio process he uses. In particular, since a stock generally has a higher average rate of growth than the money market, the investor can achieve a rate of growth higher than the interest rate by taking long positions in the stock. Indeed, by borrowing from the money market, the investor can achieve an arbitrarily high *average* rate of growth. Of course, such leveraged positions are also extremely risky.

On the other hand, if we want to know the average rate of growth of the investor's wealth under the risk-neutral probabilities, the portfolio the investor uses is irrelevant. Under the risk-neutral probabilities, the average rate of growth of the stock is equal to the interest rate. No matter how the investor divides his wealth between the stock and the money market account, he will achieve an average rate of growth equal to the interest rate. Although some portfolio processes are riskier then others under the risk-neutral measure, they all have the same average rate of growth. We state this result as a theorem, whose proof is given in a way that we can later generalize to continuous time.

Theorem 2.4.5. *Consider the binomial model with N periods. Let $\Delta_0, \Delta_1, \dots, \Delta_{N-1}$ be an adapted portfolio process, let X_0 be a real number, and let the wealth process $X_1, \dots, X_N$ be generated recursively by (2.4.6). Then the discounted wealth process $\frac{X_n}{(1+r)^n}$, $n = 0, 1, \dots, N$, is a martingale under the risk-neutral measure; i.e.,*

$$\frac{X_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n \left[\frac{X_{n+1}}{(1+r)^{n+1}} \right], \ n = 0, 1, \dots, N-1. \tag{2.4.7}$$

PROOF: We compute

$$\begin{aligned}
\widetilde{\mathbb{E}}_n \left[\frac{X_{n+1}}{(1+r)^{n+1}} \right] &= \widetilde{\mathbb{E}}_n \left[\frac{\Delta_n S_{n+1}}{(1+r)^{n+1}} + \frac{X_n - \Delta_n S_n}{(1+r)^n} \right] \\
&= \widetilde{\mathbb{E}}_n \left[\frac{\Delta_n S_{n+1}}{(1+r)^{n+1}} \right] + \widetilde{\mathbb{E}}_n \left[\frac{X_n - \Delta_n S_n}{(1+r)^n} \right] \\
&\qquad \text{(Linearity)} \\
&= \Delta_n \widetilde{\mathbb{E}}_n \left[\frac{S_{n+1}}{(1+r)^{n+1}} \right] + \frac{X_n - \Delta_n S_n}{(1+r)^n} \\
&\qquad \text{(Taking out what is known)} \\
&= \Delta_n \frac{S_n}{(1+r)^n} + \frac{X_n - \Delta_n S_n}{(1+r)^n} \\
&\qquad \text{(Theorem 2.4.4)} \\
&= \frac{X_n}{(1+r)^n}.
\end{aligned}$$

□

Corollary 2.4.6. *Under the conditions of Theorem 2.4.5, we have*

$$\widetilde{\mathbb{E}}\frac{X_n}{(1+r)^n} = X_0,\ n = 0, 1, \dots, N. \tag{2.4.8}$$

PROOF: The corollary follows from the fact that the expected value of a martingale cannot change with time and so must always be equal to the time-zero value of the martingale (see Remark 2.4.3). Applying this fact to the $\widetilde{\mathbb{P}}$-martingale $\frac{X_n}{(1+r)^n}$, $n = 0, 1, \dots, N$, we obtain (2.4.8). □

Theorem 2.4.5 and its corollary have two important consequences. The first is that there can be no arbitrage in the binomial model. If there were an arbitrage, we could begin with $X_0 = 0$ and find a portfolio process whose corresponding wealth process $X_1, X_2, \dots, X_N$ satisfied $X_N(\omega) \geq 0$ for all coin toss sequences ω and $X_N(\overline{\omega}) > 0$ for at least one coin toss sequence $\overline{\omega}$. But then we would have $\widetilde{\mathbb{E}}X_0 = 0$ and $\widetilde{\mathbb{E}}\frac{X_N}{(1+r)^N} > 0$, which violates Corollary 2.4.6.

In general, if we can find a risk-neutral measure in a model (i.e., a measure that agrees with the actual probability measure about which price paths have zero probability, and under which the discounted prices of all primary assets are martingales), then there is no arbitrage in the model. This is sometimes called the *First Fundamental Theorem of Asset Pricing*. The essence of its proof is contained in the preceding paragraph: under a risk-neutral measure, the discounted wealth process has constant expectation, so it cannot begin at zero and later be strictly positive with positive probability unless it also has a positive probability of being strictly negative. The First Fundamental Theorem of Asset Pricing will prove useful for ruling out arbitrage in term-structure models later on and thereby lead to the Heath-Jarrow-Morton no-arbitrage condition on forward rates.

The other consequence of Theorem 2.4.5 is the following version of the *risk-neutral pricing formula*. Let V_N be a random variable (derivative security paying off at time N) depending on the first N coin tosses. We know from Theorem 1.2.2 of Chapter 1 that there is an initial wealth X_0 and a replicating portfolio process $\Delta_0, \dots, \Delta_{N-1}$ that generates a wealth process $X_1, \dots, X_N$ satisfying $X_N = V_N$, no matter how the coin tossing turns out. Because $\frac{X_n}{(1+r)^n}$, $n = 0, 1, \dots, N$, is a martingale, the "multistep ahead" property of Remark 2.4.2 implies

$$\frac{X_n}{(1+r)^n} = \mathbb{E}_n\left[\frac{X_N}{(1+r)^N}\right] = \mathbb{E}_n\left[\frac{V_N}{(1+r)^N}\right]. \tag{2.4.9}$$

According to Definition 1.2.3 of Chapter 1, we define the price of the derivative security at time n to be X_n and denote this price by the symbol V_n. Thus, (2.4.9) may be rewritten as

$$\frac{V_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^N}\right] \tag{2.4.10}$$

or, equivalently,

$$V_n = \widetilde{\mathbb{E}}_n \left[\frac{V_N}{(1+r)^{N-n}} \right]. \tag{2.4.11}$$

We summarize with a theorem.

Theorem 2.4.7 (Risk-neutral pricing formula). *Consider an N-period binomial asset-pricing model with $0 < d < 1 + r < u$ and with risk-neutral probability measure $\widetilde{\mathbb{P}}$. Let V_N be a random variable (a derivative security paying off at time N) depending on the coin tosses. Then, for n between 0 and N, the price of the derivative security at time n is given by the* risk-neutral pricing formula *(2.4.11). Furthermore, the discounted price of the derivative security is a martingale under $\widetilde{\mathbb{P}}$; i.e.,*

$$\frac{V_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n \left[\frac{V_{n+1}}{(1+r)^{n+1}} \right],\ n = 0, 1, \ldots, N-1. \tag{2.4.12}$$

The random variables V_n defined by (2.4.11) are the same as the random variable V_n defined in Theorem 1.2.2.

The remaining steps in the proof of Theorem 2.4.7 are outlined in Exercise 2.8. We note that we chose the risk-neutral measure in order to make the discounted stock price a martingale. According to Theorem 2.4.7, a consequence of this is that discounted derivative security prices under the risk-neutral measure are also martingales.

So far, we have discussed only derivative securities that pay off on a single date. Many securities, such as coupon-paying bonds and interest rate swaps, make a series of payments. For such a security, we have the following pricing and hedging formulas.

Theorem 2.4.8 (Cash flow valuation). *Consider an N-period binomial asset pricing-model with $0 < d < 1 + r < u$, and with risk-neutral probability measure $\widetilde{\mathbb{P}}$. Let $C_0, C_1, \ldots, C_N$ be a sequence of random variables such that each C_n depends only on $\omega_1 \ldots \omega_n$. The price at time n of the derivative security that makes payments $C_n, \ldots, C_N$ at times $n, \ldots, N$, respectively, is*

$$V_n = \widetilde{\mathbb{E}}_n \left[\sum_{k=n}^{N} \frac{C_k}{(1+r)^{k-n}} \right],\ n = 0, 1, \ldots, N. \tag{2.4.13}$$

The price process V_n, $n = 0, 1, \ldots, N$, satisfies

$$C_n(\omega_1 \ldots \omega_n) = V_n(\omega_1 \ldots \omega_n) - \frac{1}{1+r} \Big[\tilde{p} V_{n+1}(\omega_1 \ldots \omega_n H) + \tilde{q} V_{n+1}(\omega_1 \ldots \omega_n T) \Big]. \tag{2.4.14}$$

We define

$$\Delta_n(\omega_1 \ldots \omega_n) = \frac{V_{n+1}(\omega_1 \ldots \omega_n H) - V_{n+1}(\omega_1 \ldots \omega_n T)}{S_{n+1}(\omega_1 \ldots \omega_n H) - S_{n+1}(\omega_1 \ldots \omega_n T)}, \tag{2.4.15}$$

where n ranges between 0 and $N-1$. If we set $X_0 = V_0$ and define recursively forward in time the portfolio values $X_1, X_2, \dots, X_N$ by

$$X_{n+1} = \Delta_n S_{n+1} + (1+r)(X_n - C_n - \Delta_n S_n), \tag{2.4.16}$$

then we have

$$X_n(\omega_1 \dots \omega_n) = V_n(\omega_1 \dots \omega_n) \tag{2.4.17}$$

for all n and all $\omega_1 \dots \omega_n$.

In Theorem 2.4.8, V_n is the so-called *net present value* at time n of the sequence of payments $C_n, \dots, C_N$. It is just the sum of the value $\widetilde{\mathbb{E}}_n \left[\frac{C_k}{(1+r)^{(n-k)}}\right]$ of each of the payments C_k to be made at times $k = n, k = n+1, \dots, k = N$. Note that the payment at time n is included. This payment C_n depends on only the first n tosses and so can be taken outside the conditional expectation $\widetilde{\mathbb{E}}_n$, i.e.,

$$V_n = C_n + \widetilde{\mathbb{E}}_n \left[\sum_{k=n+1}^{N} \frac{C_k}{(1+r)^{k-n}}\right], n = 0, 1, \dots, N-1. \tag{2.4.18}$$

In the case of $n = N$, (2.4.13) reduces to

$$V_N = C_N. \tag{2.4.19}$$

Consider an agent who is short the cash flows represented by $C_0, \dots, C_n$ (i.e., an agent who must make the payment C_n at each time n). (We allow these payments to be negative as well as positive. If a payment is negative, the agent who is short actually receives cash.) Suppose the agent in the short position invests in the stock and money market account, so that, at time n, before making the payment C_n, the value of his portfolio is X_n. He then makes the payment C_n. Suppose he then takes a position Δ_n in stock. This will cause the value of his portfolio at time $n+1$ before making the payment C_{n+1} to be X_{n+1}, given by (2.4.16). If this agent begins with $X_0 = V_0$ and chooses his stock positions Δ_n by (2.4.15), then (2.4.17) holds and, in particular, $X_N = V_N = C_N$ (see (2.4.17), and (2.4.19)). Then, at time N he makes the final payment C_N and is left with 0. He has perfectly hedged the short position in the cash flows. This is the justification for calling V_n the value at time n of the future cash flows, including the payment C_n to be made at time n.

PROOF OF THEOREM 2.4.8: To prove (2.4.17), we proceed by induction on n. The induction hypothesis is that $X_n(\omega_1 \dots \omega_n) = V_n(\omega_1 \dots \omega_n)$ for some $n \in \{0, 1, \dots, N-1\}$ and all $\omega_1 \dots \omega_n$. We need to show that

$$X_{n+1}(\omega_1 \dots \omega_n H) = V_{n+1}(\omega_1 \dots \omega_n H), \tag{2.4.20}$$

$$X_{n+1}(\omega_1 \dots \omega_n T) = V_{n+1}(\omega_1 \dots \omega_n T). \tag{2.4.21}$$

We prove (2.4.20); the proof of (2.4.21) is analogous.

From (2.4.18) and iterated conditioning (Theorem 2.3.2(iii)), we have

$$\begin{aligned} V_n &= C_n + \widetilde{\mathbb{E}}_n \left[\frac{1}{1+r} \widetilde{\mathbb{E}}_{n+1} \left[\sum_{k=n+1}^{N} \frac{C_k}{(1+r)^{k-(n+1)}} \right] \right] \\ &= C_n + \widetilde{\mathbb{E}}_n \left[\frac{1}{1+r} V_{n+1} \right], \end{aligned}$$

where we have used (2.4.13) with n replaced by $n+1$ in the last step. In other words, for all $\omega_1 \dots \omega_n$, we have

$$\begin{aligned} & V_n(\omega_1 \dots \omega_n) - C_n(\omega_1 \dots \omega_n) \\ &= \frac{1}{1+r} \Big[\tilde{p} V_{n+1}(\omega_1 \dots \omega_n H) + \tilde{q} V_{n+1}(\omega_1 \dots \omega_n T) \Big]. \end{aligned}$$

Since $\omega_1 \dots \omega_n$ will be fixed for the rest of the proof, we will suppress these symbols. For example, the last equation will be written simply as

$$V_n - C_n = \frac{1}{1+r} \Big[\tilde{p} V_{n+1}(H) + \tilde{q} V_{n+1}(T) \Big].$$

We compute

$$\begin{aligned} X_{n+1}(H) &= \Delta_n S_{n+1}(H) + (1+r)(X_n - C_n - \Delta_n S_n) \\ &= \frac{V_{n+1}(H) - V_{n+1}(T)}{S_{n+1}(H) - S_{n+1}(T)} \big(S_{n+1}(H) - (1+r) S_n \big) \\ &\qquad + (1+r)(V_n - C_n) \\ &= \frac{V_{n+1}(H) - V_{n+1}(T)}{(u-d) S_n} \big(u S_n - (1+r) S_n \big) \\ &\qquad + \tilde{p} V_{n+1}(H) + \tilde{q} V_{n+1}(T) \\ &= \big(V_{n+1}(H) - V_{n+1}(T) \big) \frac{u-1-r}{u-d} + \tilde{p} V_{n+1}(H) + \tilde{q} V_{n+1}(T) \\ &= \big(V_{n+1}(H) - V_{n+1}(T) \big) \tilde{q} + \tilde{p} V_{n+1}(H) + \tilde{q} V_{n+1}(T) \\ &= (\tilde{p} + \tilde{q}) V_{n+1}(H) \;=\; V_{n+1}(H). \end{aligned}$$

This is (2.4.20). □

2.5 Markov Processes

In Section 1.3, we saw that the computational requirements of the derivative security pricing algorithm of Theorem 1.2.2 can often be substantially reduced by thinking carefully about what information needs to be remembered as we go from period to period. In Example 1.3.1 of Section 1.3, the stock price was relevant, but the path it followed to get to its current price was not. In Example 1.3.2 of Section 1.3, the stock price and the maximum value it had achieved up to the current time were relevant. In this section, we formalize the procedure for determining what is relevant and what is not.

Definition 2.5.1. *Consider the binomial asset-pricing model. Let* $X_0, X_1, \dots, X_N$ *be an adapted process. If, for every* n *between* 0 *and* $N-1$ *and for every function* $f(x)$*, there is another function* $g(x)$ *(depending on* n *and* f*) such that*

$$\mathbb{E}_n[f(X_{n+1})] = g(X_n), \tag{2.5.1}$$

we say that $X_0, X_1, \dots, X_N$ *is a* Markov process.

By definition, $\mathbb{E}_n[f(M_{n+1})]$ is random; it depends on the first n coin tosses. The Markov property says that this dependence on the coin tosses occurs through X_n (i.e., the information about the coin tosses one needs in order to evaluate $\mathbb{E}_n[f(X_{n+1})]$ is summarized by X_n). We are not so concerned with determining a formula for the function g right now as we are with asserting its existence because its mere existence tells us that if the payoff of a derivative security is random only through its dependence on X_N, then there is a version of the derivative security pricing algorithm in which we do not need to store path information (see Theorem 2.5.8). In examples in this section, we shall develop a method for finding the function g.

Example 2.5.2 (Stock price). In the binomial model, the stock price at time $n+1$ is given in terms of the stock price at time n by the formula

$$S_{n+1}(\omega_1 \dots \omega_n \omega_{n+1}) = \begin{cases} uS_n(\omega_1 \dots \omega_n), \text{ if } \omega_{n+1} = H, \\ dS_n(\omega_1 \dots \omega_n), \text{ if } \omega_{n+1} = T. \end{cases}$$

Therefore,

$$\mathbb{E}_n[f(S_{n+1})](\omega_1 \dots \omega_n) = pf(uS_n(\omega_1 \dots \omega_n)) + qf(dS_n(\omega_1 \dots \omega_n)),$$

and the right-hand side depends on $\omega_1 \dots \omega_n$ only through the value of $S_n(\omega_1 \dots \omega_n)$. Omitting the coin tosses $\omega_1 \dots \omega_n$, we can rewrite this equation as

$$\mathbb{E}_n[f(S_{n+1})] = g(S_n),$$

where the function $g(x)$ of the dummy variable x is defined by $g(x) = pf(ux) + qf(dx)$. This shows that the stock price process is Markov.

Indeed, the stock price process is Markov under either the actual or the risk-neutral probability measure. To determine the price V_n at time n of a derivative security whose payoff at time N is a function v_N of the stock price S_N (i.e., $V_N = v_N(S_N)$), we use the risk-neutral pricing formula (2.4.12), which reduces to

$$V_n = \frac{1}{1+r}\widetilde{\mathbb{E}}_n[V_{n+1}], \ n = 0, 1, \dots, N-1.$$

But $V_N = v_N(S_N)$ and the stock price process is Markov, so

$$V_{N-1} = \frac{1}{1+r}\widetilde{\mathbb{E}}_{N-1}[v_N(S_N)] = v_{N-1}(S_{N-1})$$

for some function v_{N-1}. Similarly,

$$V_{N-2} = \frac{1}{1+r}\widetilde{\mathbb{E}}_{N-2}[v_{N-1}(S_{N-1})] = v_{N-2}(S_{N-2})$$

for some function v_{N-2}. In general, $V_n = v_n(S_n)$ for some function v_n. Moreover, we can compute these functions recursively by the algorithm

$$v_n(s) = \frac{1}{1+r}\Big[\tilde{p}v_{n+1}(us) + \tilde{q}v_{n+1}(ds)\Big],\ n = N-1, N-2, \ldots, 0. \tag{2.5.2}$$

This algorithm works in the binomial model for any derivative security whose payoff at time N is a function only of the stock price at time N. In particular, we have the same algorithm for puts and calls. The only difference is in the formula for $v_N(s)$. For the call, we have $v_N(s) = (s-K)^+$; for the put, we have $v_N(s) = (K-s)^+$. □

The martingale property is the special case of (2.5.1) with $f(x) = x$ and $g(x) = x$. In order for a process to be Markov, it is necessary that for *every* function f there must be a corresponding function g such that (2.5.1) holds. Not every martingale is Markov. On the other hand, even when considering the function $f(x) = x$, the Markov property requires only that $\mathbb{E}_n[M_{n+1}] = g(M_n)$ for some function g; it does not require that the function g be given by $g(x) = x$. Not every Markov process is a martingale. Indeed, Example 2.5.2 shows that the stock price is Markov under both the actual and the risk-neutral probability measures. It is typically not a martingale under either of these measures. However, if $pu + qd = 1$, then the stock price is both a martingale and a Markov process under the actual probability measure.

The following lemma often provides the key step in the verification that a process is Markov.

Lemma 2.5.3 (Independence). *In the N-period binomial asset pricing model, let n be an integer between 0 and N. Suppose the random variables $X^1, \ldots, X^K$ depend only on coin tosses 1 through n and the random variables $Y^1, \ldots, Y^L$ depend only on coin tosses $n+1$ through N. (The superscripts $1, \ldots, K$ on X and $1, \ldots, L$ on Y are superscripts, not exponents.) Let $f(x^1, \ldots, x^K, y^1, \ldots, y^L)$ be a function of dummy variables $x^1, \ldots, x^K$ and $y^1, \ldots, y^L$, and define*

$$g(x^1, \ldots, x^K) = \mathbb{E}f(x^1, \ldots, x^K, Y^1, \ldots, Y^L). \tag{2.5.3}$$

Then

$$\mathbb{E}_n[f(X^1, \ldots, X^K, Y^1, \ldots, Y^L)] = g(X^1, \ldots, X^K). \tag{2.5.4}$$

For the following discussion and proof of the lemma, we assume that $K = L = 1$. Then (2.5.3) takes the form

$$g(x) = \mathbb{E}f(x, Y) \tag*{(2.5.3)$'$}$$

and (2.5.4) takes the form

$$\mathbb{E}_n[f(X,Y)] = g(X), \tag{2.5.4$'$}$$

where the random variable X is assumed to depend only on the first n coin tosses, and the random variable Y depends only on coin tosses $n+1$ through N.

This lemma generalizes the property "taking out what is known" of Theorem 2.3.2(ii). Since X is "known" at time n, we want to "take it out" of the computation of the conditional expectation $\mathbb{E}_n[f(X,Y)]$. However, because X is inside the argument of the function f, we cannot simply factor it out as we did in Theorem 2.3.2(ii). Therefore, we hold it constant by replacing the random variable X by an arbitrary but fixed dummy variable x. We then compute the conditional expectation of the random variable $f(x,Y)$, whose randomness is due only to the dependence of Y on tosses $n+1$ through N. Because of Theorem 2.3.2(iv), this conditional expectation is the same as the unconditional expectation in $(2.5.3)'$. Finally, we recall that $E_n[f(X,Y)]$ must depend on the value of the random variable X, so we replace the dummy variable x by the random variable X after g is computed.

In the context of Example 2.5.2, we can take $X = S_n$, which depends only on the first n coin tosses, and take $Y = \frac{S_{n+1}}{S_n}$, which depends only on the $(n+1)$st coin toss, taking the value u if the $(n+1)$st toss results in a head and taking the value d if it results in a tail. We are asked to compute

$$\mathbb{E}_n[f(S_{n+1})] = \mathbb{E}_n[f(XY)].$$

We replace X by a dummy variable x and compute

$$g(x) = \mathbb{E}f(xY) = pf(ux) + qf(dx).$$

Then $\mathbb{E}_n[f(S_{n+1}] = g(S_n)$.

Proof of Lemma 2.5.3: Let $\omega_1 \ldots \omega_n$ be fixed but arbitrary. By the definition (2.3.6) of conditional expectation,

$$\begin{aligned}&\mathbb{E}_n[f(X,Y)](\omega_1 \ldots \omega_n)\\ &= \sum_{\omega_{n+1}\ldots\omega_N} f(X(\omega_1 \ldots \omega_n), Y(\omega_{n+1} \ldots \omega_N))p^{\#H(\omega_{n+1}\ldots\omega_N)}q^{\#T(\omega_{n+1}\ldots\omega_N)},\end{aligned}$$

whereas

$$\begin{aligned}g(x) &= \mathbb{E}f(x,Y)\\ &= \sum_{\omega_{n+1}\ldots\omega_N} f(x, Y(\omega_{n+1} \ldots \omega_N))p^{\#H(\omega_{n+1}\ldots\omega_N)}q^{\#T(\omega_{n+1}\ldots\omega_N)}.\end{aligned}$$

It is apparent that

$$\mathbb{E}_n[f(X,Y)](\omega_1 \ldots \omega_N) = g(X(\omega_1 \ldots \omega_N)). \qquad \square$$

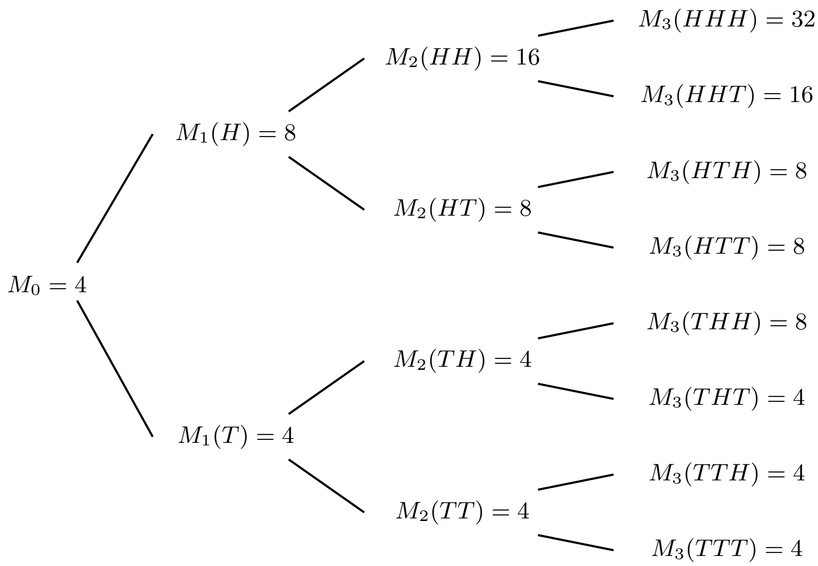

Fig. 2.5.1. The maximum stock price to date.

Example 2.5.4 (Non-Markov process). In the binomial model of Figure 2.3.1, consider the maximum-to-date process $M_n = \max_{0 \le k \le n} S_k$, shown in Figure 2.5.1. With $p = \frac{2}{3}$ and $q = \frac{1}{3}$, we have

$$\mathbb{E}_2[M_3](TH) = \frac{2}{3} M_3(THH) + \frac{1}{3} M_3(THT) = \frac{16}{3} + \frac{4}{3} = 6\frac{2}{3},$$

but

$$\mathbb{E}_2[M_3](TT) = \frac{2}{3} M_3(TTH) + \frac{1}{3} M_3(TTT) = \frac{8}{3} + \frac{4}{3} = 4.$$

Since $M_2(TH) = M_2(TT) = 4$, there cannot be a function g such that $\mathbb{E}_3[M_3](TH) = g(M_2(TH))$ and $\mathbb{E}_3[M_3](TT) = g(M_2(TT))$. The right-hand sides would be the same, but the left-hand sides would not. The maximum-to-date process is not Markov because recording only that the value of the maximum-to-date at time two is 4, without recording the value of the stock price at time two, neglects information relevant to the evolution of the maximum-to-date process after time two. □

When we encounter a non-Markov process, we can sometimes recover the Markov property by adding one or more so-called *state variables*. The term "state variable" is used because if we can succeed in recovering the Markov property by adding these variables, we will have determined a way to describe the "state" of the market in terms of these variables. This approach to recovering the Markov property requires that we generalize Definition 2.5.1 to multidimensional processes.

Definition 2.5.5. *Consider the binomial asset-pricing model. Let* $\{(X_n^1, \dots, X_n^K); n = 0, 1, \dots, N\}$ *be a* K*-dimensional adapted process; i.e.,* K *one-dimensional adapted processes. If, for every* n *between* 0 *and* $N-1$ *and for every function* $f(x^1, \dots, x^K)$*, there is another function* $g(x^1, \dots, x^K)$ *(depending on* n *and* f*) such that*

$$\mathbb{E}_n[f(X_{n+1}^1, \dots, X_{n+1}^K)] = g(X_n^1, \dots, X_n^K), \tag{2.5.5}$$

we say that $\{(X_n^1, \dots, X_n^K); n = 0, 1, \dots, N\}$ *is a* K-dimensional Markov process.

Example 2.5.6. In an N-period binomial model, consider the two-dimensional adapted process $\{(S_n, M_n); n = 0, 1, \dots, N\}$, where S_n is the stock price at time n and $M_n = \max_{0 \le k \le n} S_k$ is the stock price maximum-to-date. We show that this two-dimensional process is Markov. To do that, we define $Y = \frac{S_{n+1}}{S_n}$, which depends only on the $(n+1)$st coin toss. Then

$$S_{n+1} = S_n Y$$

and

$$M_{n+1} = M_n \vee S_{n+1} = M_n \vee (S_n Y),$$

where $x \vee y = \max\{x, y\}$. We wish to compute

$$\mathbb{E}_n[f(S_{n+1}, M_{n+1})] = \mathbb{E}_n[f(S_n Y, M_n \vee (S_n Y))].$$

According to Lemma 2.5.3, we replace S_n by a dummy variable s, replace M_n by a dummy variable m, and compute

$$g(s, m) = \mathbb{E} f(sY, m \vee (sY)) = pf(us, m \vee (us)) + qf(ds, m \vee (ds)).$$

Then

$$\mathbb{E}_n[f(S_{n+1}, M_{n+1})] = g(S_n, M_n).$$

Since we have obtained a formula for $\mathbb{E}_n[f(S_{n+1}, M_{n+1})]$ in which the only randomness enters through the random variables S_n and M_n, we conclude that the two-dimensional process is Markov. In this example, we have used the actual probability measure, but the same argument shows that $\{(S_n, M_n); n = 0, 1, \dots, N\}$ is Markov under the risk-neutral probability measure $\widetilde{\mathbb{P}}$. □

Remark 2.5.7. The Markov property, in both the one-dimensional form of Definition 2.5.1 and the multidimensional form of Definition 2.5.5, is a "one-step-ahead" property, determining a formula for the conditional expectation of X_{n+1} in terms of X_n. However, it implies a similar condition for any number of steps. Indeed, if $X_0, X_1, \dots, X_N$ is a Markov process and $n \le N-2$, then the "one-step-ahead" Markov property implies that for every function h there is a function f such that

$$\mathbb{E}_{n+1}[h(X_{n+2})] = f(X_{n+1}).$$

Taking conditional expectations on both sides based on the information at time n and using the iterated conditioning property (iii) of Theorem 2.3.3, we obtain

$$\mathbb{E}_n[h(X_{n+2})] = \mathbb{E}_n\Big[\mathbb{E}_{n+1}[h(X_{n+2})]\Big] = \mathbb{E}_n[f(X_{n+1})].$$

Because of the "one-step-ahead" Markov property, the right-hand side is $g(X_n)$ for some function g, and we have obtained the "two-step-ahead" Markov property

$$\mathbb{E}_n[h(X_{n+2})] = g(X_n).$$

Iterating this argument, we can show that whenever $0 \le n \le m \le N$ and h is any function, then there is another function g such that the "multi-step-ahead" Markov property

$$\mathbb{E}_n[h(X_m)] = g(X_n) \tag{2.5.6}$$

holds. Similarly, if $\{(X_n^1, \dots, X_n^K); n = 1, 2, \dots, N\}$ is a K-dimensional Markov process, then whenever $0 \le n \le m \le N$ and $h(x^1, \dots, x^K)$ is any function, there is another function $g(x^1, \dots, x^K)$ such that

$$\mathbb{E}_n[h(X_m^1, \dots, X_m^K)] = g(X_n^1, \dots, X_n^K). \tag{2.5.7}$$

□

In the binomial pricing model, suppose we have a Markov process $X_0, X_1, \dots, X_N$ under the risk-neutral probability measure $\widetilde{\mathbb{P}}$, and we have a derivative security whose payoff V_N at time N is a function v_N of X_N, i.e., $V_N = v_N(X_N)$. The difference between V_N and v_N is that the argument of the former is $\omega_1 \dots \omega_N$, a sequence of coin tosses, whereas the argument of the latter is a real number, which we will sometimes denote by the dummy variable x. In particular, there is nothing random about $v_N(x)$. However, if in place of the dummy variable x we substitute the random variable X_N (actually $X_N(\omega_1 \dots \omega_N)$), then we have a random variable. Indeed, we have

$$V_N(\omega_1 \dots \omega_N) = v_N(X_N(\omega_1 \dots \omega_N)) \text{ for all } \omega_1 \dots \omega_N.$$

The risk-neutral pricing formula (2.4.11) says that the price of this derivative security at earlier times n is

$$V_n(\omega_1 \dots \omega_n) = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^{N-n}}\right](\omega_1 \dots \omega_n) \text{ for all } \omega_1 \dots \omega_n.$$

On the other hand, the "multi-step-ahead" Markov property implies that there is a function v_n such that

$$\widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^{N-n}}\right](\omega_1 \dots \omega_n) = v_n\big(X_n(\omega_1 \dots \omega_n)\big) \text{ for all } \omega_1 \dots \omega_n.$$

Therefore, the price of the derivative security at time n is a function of X_n, i.e.,

$$V_n = v_n(X_n).$$

Instead of computing the random variables V_n, we can compute the functions v_n, and this is generally much more manageable computationally. In particular, when the Markov process $X_0, X_1, \dots, X_N$ is the stock price itself, we get the algorithm (2.5.2).

The same idea can be used for multidimensional Markov processes under $\widetilde{\mathbb{P}}$. A case of this was Example 1.3.2 of Section 1.3, in which the payoff of a derivative security was $V_3 = M_3 - S_3$, the difference between the stock price at time three and its maximum between times zero and three. Because only the stock price and its maximum-to-date appear in the payoff, we can use the two-dimensional Markov process $\{(S_n, M_n); n = 0, 1, 2, 3\}$ to treat this problem, which was done implicitly in that example.

Here we generalize Example 1.3.2 to an N-period binomial model with a derivative security whose payoff at time N is a function $v_N(S_N, M_N)$ of the stock price and the maximum stock price. (We do not mean that v_N is necessarily a function of *both* S_N and M_N but rather that these are the *only* random variables on which V_N depends. For example, we could have $V_N = (M_N - K)^+$. Even though the stock price does not appear in this particular V_N, we would need it to execute the pricing algorithm (2.5.9) below because the maximum-to-date process is not Markov by itself.) According to the "multi-step-head" Markov property, for any n between zero and N, there is a (nonrandom) function $v_n(s, m)$ such that the price of the option at time n is

$$V_n = v_n(S_n, M_n) = \widetilde{\mathbb{E}}_n\left[\frac{v_N(S_N, M_N)}{(1+r)^{N-n}}\right].$$

We can use the Independence Lemma 2.5.3 to derive an algorithm for computing the functions v_n. We always have the risk-neutral pricing formula (see (2.4.12))

$$V_n = \frac{1}{1+r}\widetilde{\mathbb{E}}_n[V_{n+1}]$$

relating the price of a derivative security at time n to its price at time $n+1$. Suppose that for some n between zero and $N-1$, we have computed the function v_{n+1} such that $V_{n+1} = v_{n+1}(S_{n+1}, M_{n+1})$. Then

$$\begin{aligned}
V_n &= \frac{1}{1+r}\widetilde{\mathbb{E}}_n[V_{n+1}] \\
&= \frac{1}{1+r}\widetilde{\mathbb{E}}_n\left[v_{n+1}(S_{n+1}, M_{n+1})\right] \\
&= \frac{1}{1+r}\widetilde{\mathbb{E}}_n\left[v_{n+1}\left(S_n \cdot \frac{S_{n+1}}{S_n}, M_n \vee \left(S_n \cdot \frac{S_{n+1}}{S_n}\right)\right)\right].
\end{aligned}$$

To compute this last expression, we replace S_n and M_n by dummy variables s and m because they depend only on the first n tosses. We then take the unconditional expectation of $\frac{S_{n+1}}{S_n}$ because it does not depend on the first n tosses, i.e., we define

$$\begin{aligned} v_n(s,m) &= \frac{1}{1+r}\widetilde{\mathbb{E}}_n\left[v_{n+1}\left(s\cdot\frac{S_{n+1}}{S_n}, m\vee\left(s\cdot\frac{S_{n+1}}{S_n}\right)\right)\right] \\ &= \frac{1}{1+r}\Big[\tilde{p}v_{n+1}(us, m\vee(us)) + \tilde{q}v_{n+1}(ds, m\vee(ds))\Big]. \end{aligned} \tag{2.5.8}$$

The Independence Lemma 2.5.3 asserts that $V_n = v_n(S_n, M_n)$.

We will only need to know the value of $v_n(s,m)$ when $m \geq s$ since $M_n \geq S_n$. We can impose this condition in (2.5.8). But when $m \geq s$, if $d \leq 1$ as it usually is, we have $m \vee (ds) = m$. Therefore, we can rewrite (2.5.8) as

$$v_n(s,m) = \frac{1}{1+r}\Big[\tilde{p}v_{n+1}(us, m\vee(us)) + \tilde{q}v_{n+1}(ds, m)\Big], \tag{2.5.9}$$
$$m \geq s > 0,\ n = N-1, N-2, \ldots, 0.$$

This algorithm works for any derivative security whose payoff at time N depends only on the random variables S_N and M_N.

In Example 1.3.2, we were given that $V_3 = v_3(s,m)$, where $v_3(s,m) = m - s$. We used (2.5.9) to compute v_2, then used it again to compute v_1, and finally used it to compute v_0. These steps were carried out in Example 1.3.2.

In continuous time, we shall see that the analogue of recursive equations (2.5.9) are partial differential equations. The process that gets us from the continuous-time analogue of the risk-neutral pricing formula to these partial differential equations is the *Feynman-Kac Theorem.*

We summarize this discussion with a theorem.

Theorem 2.5.8. *Let $X_0, X_1, \ldots, X_N$ be a Markov process under the risk-neutral probability measure $\widetilde{\mathbb{P}}$ in the binomial model. Let $v_N(x)$ be a function of the dummy variable x, and consider a derivative security whose payoff at time N is $v_N(X_N)$. Then, for each n between 0 and N, the price V_n of this derivative security is some function v_n of X_n, i.e.,*

$$V_n = v_n(X_n),\ n = 0, 1, \ldots, N. \tag{2.5.10}$$

There is a recursive algorithm for computing v_n whose exact formula depends on the underlying Markov process $X_0, X_1, \ldots, X_N$. Analogous results hold if the underlying Markov process is multidimensional.

2.6 Summary

This chapter sets out the view of probability that begins with a random experiment having outcome ω. The collection of all possible outcomes is called the *sample space* Ω, and on this space we have a probability measure $\mathbb{P}$. When Ω is finite, we describe $\mathbb{P}$ by specifying for each $\omega \in \Omega$ the probability $\mathbb{P}(\omega)$ assigned to ω by $\mathbb{P}$. A random variable is a function X from Ω to $\mathbb{R}$, and the expectation of the random variable X is $\mathbb{E}X = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$. If we have a second probability measure $\widetilde{\mathbb{P}}$ on Ω, then we will have another way of computing the expectation, namely $\widetilde{\mathbb{E}}X = \sum_{\omega \in \Omega} X(\omega)\widetilde{\mathbb{P}}(\omega)$. The random variable X is the same in both cases, even though the two expectations are different. The point is that the random variable should not be thought of as a distribution. When we change probability measures, distributions (and hence expectations) will change, but random variables will not.

In the binomial model, we may see coin tosses $\omega_1 \dots \omega_n$ and, based on this information, compute the conditional expectation of a random variable X that depends on coin tosses $\omega_1 \dots \omega_n \omega_{n+1} \dots \omega_N$. This is done by averaging over the possible outcomes of the "remaining" coin tosses $\omega_{n+1} \dots \omega_N$. If we are computing the conditional expectation under the risk-neutral probabilities, this results in the formula

$$\begin{aligned} &\widetilde{\mathbb{E}}_n[X](\omega_1 \dots \omega_n) \qquad (2.3.6) \\ &\quad = \sum_{\omega_{n+1} \dots \omega_N} \tilde{p}^{\#H(\omega_{n+1} \dots \omega_N)} \tilde{q}^{\#T(\omega_{n+1} \dots \omega_N)} X(\omega_1 \dots \omega_n \omega_{n+1} \dots \omega_N). \end{aligned}$$

This conditional expectation is a random variable because it depends on the first n coin tosses $\omega_1 \dots \omega_n$. Conditional expectations have five fundamental properties, which are provided in Theorem 2.3.2.

In a multiperiod binomial model, a *martingale* under the risk-neutral probability measure $\widetilde{\mathbb{P}}$ is a sequence of random variables $M_0, M_1, \dots, M_N$, where each M_n depends on only the first n coin tosses, and

$$M_n(\omega_1 \dots \omega_n) = \widetilde{\mathbb{E}}_n[M_{n+1}](\omega_1 \dots \omega_n)$$

no matter what the value of n and no matter what the coin tosses $\omega_1 \dots \omega_n$ are. A martingale has no tendency to rise or fall. Conditioned on the information we have at time n, the expected value of the martingale at time $n+1$ is its value at time n.

Under the risk-neutral probability measure, the discounted stock price is a martingale, as is the discounted value of any portfolio that trades in the stock and money markets account. In particular, if X_n is the value of a portfolio at time n, then

$$\frac{X_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{X_N}{(1+r)^N}\right],\ 0 \le n \le N.$$

If we want to have X_N agree with the value V_N of a derivative security at its expiration time N, then we must have

$$\frac{X_n}{(1+r)^n} = \widetilde{\mathbb{E}}_n\left[\frac{X_N}{(1+r)^N}\right] = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^N}\right] \qquad (2.4.9)$$

at all times $n = 0, 1, \dots, N$. When a portfolio does this, we define the value V_n of the derivative security at time n to be X_n, and we thus have the risk-neutral pricing formula

$$V_n = \widetilde{\mathbb{E}}_n \left[\frac{V_N}{(1+r)^{N-n}} \right]. \tag{2.4.11}$$

A *Markov process* is a sequence of random variables $X_0, X_1, \ldots, X_N$ with the following property. Suppose n is a time between 0 and $N-1$, we have observed the first n coin tosses $\omega_1 \ldots \omega_n$, and we want to estimate either a function of X_{n+1} or, more generally, a function of X_{n+k} for some k between 1 and $N-n$. We know both the individual coin tosses $\omega_1 \ldots \omega_n$ and the resulting value $X_n(\omega_1 \ldots \omega_n)$ and can base our estimate on this information. For a Markov process, knowledge of the individual coin tosses (the "path") does not provide any information relevant to this estimation problem beyond that information already contained in our knowledge of the value $X_n(\omega_1 \ldots \omega_n)$.

Consider an underlying asset-price process $X_0, X_1, \ldots, X_N$ that is Markov under the risk-neutral measure and a derivative security payoff at time N that is a function of this asset price at time N; i.e., $V_N = v_N(X_N)$. The price of the derivative security at all times n prior to expiration is a function of the underlying asset price at those times; i.e.,

$$V_n = v_n(X_n), \; n = 0, 1, \ldots, N. \tag{2.4.11}$$

In this notation, V_n is a random variable depending on the coin tosses $\omega_1 \ldots \omega_n$. It is potentially path-dependent. On the other hand, $v_n(x)$ is a function of a real number x. When we replace x by the random variable X_n, then $v_n(X_n)$ also becomes random, but in a way that is guaranteed not to be path-dependent. Equation (2.4.11) thus guarantees that the price of the derivative security is not path-dependent.

2.7 Notes

The sample space view of probability theory dates back to Kolmogorov [29], who developed it in a way that extends to infinite probability spaces. We take up this subject in Chapters 1 and 2 of Volume II. Martingales were invented by Doob [13], who attributes the idea and the name "martingale" to a gambling strategy discussed by Ville [43].

The risk-neutral pricing formula is due to Harrison and Kreps [17] and Harrison and Pliska [18].

2.8 Exercises

Exercise 2.1. Using Definition 2.1.1, show the following.

(i) If A is an event and A^c denotes its complement, then $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(ii) If $A_1, A_2, \dots, A_N$ is a finite set of events, then

$$\mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) \le \sum_{n=1}^{N} \mathbb{P}(A_n). \tag{2.8.1}$$

If the events $A_1, A_2, \dots, A_N$ are disjoint, then equality holds in (2.8.1).

Exercise 2.2. Consider the stock price S_3 in Figure 2.3.1.

(i) What is the distribution of S_3 under the risk-neutral probabilities $\tilde{p} = \frac{1}{2}$, $\tilde{q} = \frac{1}{2}$.
(ii) Compute $\widetilde{\mathbb{E}}S_1$, $\widetilde{\mathbb{E}}S_2$, and $\widetilde{\mathbb{E}}S_3$. What is the average rate of growth of the stock price under $\widetilde{\mathbb{P}}$?
(iii) Answer (i) and (ii) again under the actual probabilities $p = \frac{2}{3}$, $q = \frac{1}{3}$.

Exercise 2.3. Show that a convex function of a martingale is a submartingale. In other words, let $M_0, M_1, \dots, M_N$ be a martingale and let φ be a convex function. Show that $\varphi(M_0), \varphi(M_1), \dots, \varphi(M_N)$ is a submartingale.

Exercise 2.4. Toss a coin repeatedly. Assume the probability of head on each toss is $\frac{1}{2}$, as is the probability of tail. Let $X_j = 1$ if the jth toss results in a head and $X_j = -1$ if the jth toss results in a tail. Consider the stochastic process $M_0, M_1, M_2, \dots$ defined by $M_0 = 0$ and

$$M_n = \sum_{j=1}^{n} X_j, \ n \ge 1.$$

This is called a *symmetric random walk*; with each head, it steps up one, and with each tail, it steps down one.

(i) Using the properties of Theorem 2.3.2, show that $M_0, M_1, M_2, \dots$ is a martingale.
(ii) Let σ be a positive constant and, for $n \ge 0$, define

$$S_n = e^{\sigma M_n}\left(\frac{2}{e^{\sigma} + e^{-\sigma}}\right)^n.$$

Show that $S_0, S_1, S_2, \dots$ is a martingale. Note that even though the symmetric random walk M_n has no tendency to grow, the "geometric symmetric random walk" $e^{\sigma M_n}$ does have a tendency to grow. This is the result of putting a martingale into the (convex) exponential function (see Exercise 2.3). In order to again have a martingale, we must "discount" the geometric symmetric random walk, using the term $\frac{2}{e^{\sigma}+e^{-\sigma}}$ as the discount rate. This term is strictly less than one unless $\sigma = 0$.

Exercise 2.5. Let $M_0, M_1, M_2, \dots$ be the symmetric random walk of Exercise 2.4, and define $I_0 = 0$ and

$$I_n = \sum_{j=0}^{n-1} M_j(M_{j+1} - M_j), \ n = 1, 2, \dots .$$

(i) Show that
$$I_n = \frac{1}{2}M_n^2 - \frac{n}{2}.$$

(ii) Let n be an arbitrary nonnegative integer, and let $f(i)$ be an arbitrary function of a variable i. In terms of n and f, define another function $g(i)$ satisfying
$$E_n\big[f(I_{n+1})\big] = g(I_n).$$
Note that although the function $g(I_n)$ on the right-hand side of this equation may depend on n, the only random variable that may appear in its argument is I_n; the random variable M_n may not appear. You will need to use the formula in part (i). The conclusion of part (ii) is that the process $I_0, I_1, I_2, \ldots$ is a Markov process.

Exercise 2.6 (Discrete-time stochastic integral). Suppose $M_0, M_1, \ldots, M_N$ is a martingale, and let $\Delta_0, \Delta_1, \ldots, \Delta_{N-1}$ be an adapted process. Define the *discrete-time stochastic integral* (sometimes called a *martingale transform*) $I_0, I_1, \ldots, I_N$ by setting $I_0 = 0$ and
$$I_n = \sum_{j=0}^{n-1} \Delta_j (M_{j+1} - M_j),\ n = 1, \ldots, N.$$
Show that $I_0, I_1, \ldots, I_N$ is a martingale.

Exercise 2.7. In a binomial model, give an example of a stochastic process that is a martingale but is not Markov.

Exercise 2.8. Consider an N-period binomial model.

(i) Let $M_0, M_1, \ldots, M_N$ and $M_0', M_1', \ldots, M_N'$ be martingales under the risk-neutral measure $\widetilde{\mathbb{P}}$. Show that if $M_N = M_N'$ (for every possible outcome of the sequence of coin tosses), then, for each n between 0 and N, we have $M_n = M_n'$ (for every possible outcome of the sequence of coin tosses).

(ii) Let V_N be the payoff at time N of some derivative security. This is a random variable that can depend on all N coin tosses. Define recursively $V_{N-1}, V_{N-2}, \ldots, V_0$ by the algorithm (1.2.16) of Chapter 1. Show that
$$V_0, \frac{V_1}{1+r}, \ldots, \frac{V_{N-1}}{(1+r)^{N-1}}, \frac{V_N}{(1+r)^N}$$
is a martingale under $\widetilde{\mathbb{P}}$.

(iii) Using the risk-neutral pricing formula (2.4.11) of this chapter, define
$$V_n' = \widetilde{\mathbb{E}}_n\left[\frac{V_N}{(1+r)^{N-n}}\right],\ n = 0, 1, \ldots, N-1.$$
Show that
$$V_0', \frac{V_1'}{1+r}, \ldots, \frac{V_{N-1}'}{(1+r)^{N-1}}, \frac{V_N}{(1+r)^N}$$
is a martingale.

(iv) Conclude that $V_n = V'_n$ for every n (i.e., the algorithm (1.2.16) of Theorem 1.2.2 of Chapter 1 gives the same derivative security prices as the risk-neutral pricing formula (2.4.11) of Chapter 2).

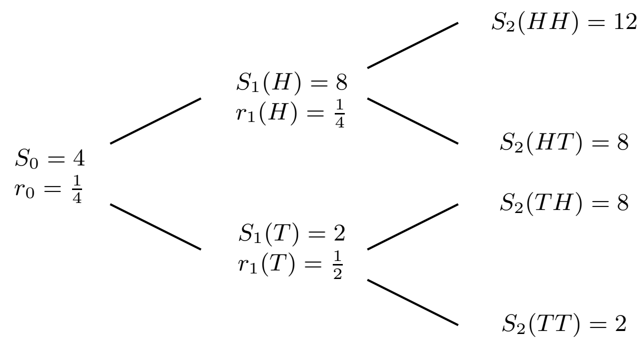

Fig. 2.8.1. A stochastic volatility, random interest rate model.

Exercise 2.9 (Stochastic volatility, random interest rate). Consider a two-period stochastic volatility, random interest rate model of the type described in Exercise 1.9 of Chapter 1. The stock prices and interest rates are shown in Figure 2.8.1.

(i) Determine risk-neutral probabilities

$$\widetilde{\mathbb{P}}(HH),\ \widetilde{\mathbb{P}}(HT),\ \widetilde{\mathbb{P}}(TH),\ \widetilde{\mathbb{P}}(TT),$$

such that the time-zero value of an option that pays off V_2 at time two is given by the risk-neutral pricing formula

$$V_0 = \widetilde{\mathbb{E}}\left[\frac{V_2}{(1+r_0)(1+r_1)}\right].$$

(ii) Let $V_2 = (S_2 - 7)^+$. Compute V_0, $V_1(H)$, and $V_1(T)$.

(iii) Suppose an agent sells the option in (ii) for V_0 at time zero. Compute the position Δ_0 she should take in the stock at time zero so that at time one, regardless of whether the first coin toss results in head or tail, the value of her portfolio is V_1.

(iv) Suppose in (iii) that the first coin toss results in head. What position $\Delta_1(H)$ should the agent now take in the stock to be sure that, regardless of whether the second coin toss results in head or tail, the value of her portfolio at time two will be $(S_2 - 7)^+$?

Exercise 2.10 (Dividend-paying stock). We consider a binomial asset pricing model as in Chapter 1, except that, after each movement in the stock price, a dividend is paid and the stock price is reduced accordingly. To describe this in equations, we define

$$Y_{n+1}(\omega_1 \dots \omega_n \omega_{n+1}) = \begin{cases} u, \text{ if } \omega_{n+1} = H, \\ d, \text{ if } \omega_{n+1} = T. \end{cases}$$

Note that Y_{n+1} depends only on the $(n+1)$st coin toss. In the binomial model of Chapter 1, $Y_{n+1}S_n$ was the stock price at time $n+1$. In the dividend-paying model considered here, we have a random variable $A_{n+1}(\omega_1 \dots \omega_n \omega_{n+1})$, taking values in $(0,1)$, and the dividend paid at time $n+1$ is $A_{n+1}Y_{n+1}S_n$. After the dividend is paid, the stock price at time $n+1$ is

$$S_{n+1} = (1 - A_{n+1})Y_{n+1}S_n.$$

An agent who begins with initial capital X_0 and at each time n takes a position of Δ_n shares of stock, where Δ_n depends only on the first n coin tosses, has a portfolio value governed by the wealth equation (see (2.4.6))

$$\begin{aligned} X_{n+1} &= \Delta_n S_{n+1} + (1+r)(X_n - \Delta_n S_n) + \Delta_n A_{n+1} Y_{n+1} S_n \\ &= \Delta_n Y_{n+1} S_n + (1+r)(X_n - \Delta_n S_n). \end{aligned} \tag{2.8.2}$$

(i) Show that the discounted wealth process is a martingale under the risk-neutral measure (i.e., Theorem 2.4.5 still holds for the wealth process (2.8.2)). As usual, the risk-neutral measure is still defined by the equations

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-1-r}{u-d}.$$

(ii) Show that the risk-neutral pricing formula still applies (i.e., Theorem 2.4.7 holds for the dividend-paying model).

(iii) Show that the discounted stock price is not a martingale under the risk-neutral measure (i.e., Theorem 2.4.4 no longer holds). However, if A_{n+1} is a constant $a \in (0,1)$, regardless of the value of n and the outcome of the coin tossing $\omega_1 \dots \omega_{n+1}$, then $\frac{S_n}{(1-a)^n(1+r)^n}$ is a martingale under the risk-neutral measure.

Exercise 2.11 (Put–call parity). Consider a stock that pays no dividend in an N-period binomial model. A European call has payoff $C_N = (S_N - K)^+$ at time N. The price C_n of this call at earlier times is given by the risk-neutral pricing formula (2.4.11):

$$C_n = \widetilde{\mathbb{E}}_n \left[\frac{C_N}{(1+r)^{N-n}} \right], \ n = 0, 1, \dots, N-1.$$

Consider also a put with payoff $P_N = (K - S_N)^+$ at time N, whose price at earlier times is

$$P_n = \widetilde{\mathbb{E}}_n \left[\frac{P_N}{(1+r)^{N-n}} \right], \; n = 0, 1, \ldots, N-1.$$

Finally, consider a *forward contract* to buy one share of stock at time N for K dollars. The price of this contract at time N is $F_N = S_N - K$, and its price at earlier times is

$$F_n = \widetilde{\mathbb{E}}_n \left[\frac{F_N}{(1+r)^{N-n}} \right], \; n = 0, 1, \ldots, N-1.$$

(Note that, unlike the call, the forward contract requires that the stock be purchased at time N for K dollars and has a negative payoff if $S_N < K$.)

(i) If at time zero you buy a forward contract and a put, and hold them until expiration, explain why the payoff you receive is the same as the payoff of a call; i.e., explain why $C_N = F_N + P_N$.

(ii) Using the risk-neutral pricing formulas given above for C_n, P_n, and F_n and the linearity of conditional expectations, show that $C_n = F_n + P_n$ for every n.

(iii) Using the fact that the discounted stock price is a martingale under the risk-neutral measure, show that $F_0 = S_0 - \frac{K}{(1+r)^N}$.

(iv) Suppose you begin at time zero with F_0, buy one share of stock, borrowing money as necessary to do that, and make no further trades. Show that at time N you have a portfolio valued at F_N. (This is called a *static replication* of the forward contract. If you sell the forward contract for F_0 at time zero, you can use this static replication to hedge your short position in the forward contract.)

(v) The *forward price* of the stock at time zero is defined to be that value of K that causes the forward contract to have price zero at time zero. The forward price in this model is $(1+r)^N S_0$. Show that, at time zero, the price of a call struck at the forward price is the same as the price of a put struck at the forward price. This fact is called *put–call parity.*

(vi) If we choose $K = (1+r)^N S_0$, we just saw in (v) that $C_0 = P_0$. Do we have $C_n = P_n$ for every n?

Exercise 2.12 (Chooser option). Let $1 \le m \le N-1$ and $K > 0$ be given. A *chooser option* is a contract sold at time zero that confers on its owner the right to receive either a call or a put at time m. The owner of the chooser may wait until time m before choosing. The call or put chosen expires at time N with strike price K. Show that the time-zero price of a chooser option is the sum of the time-zero price of a put, expiring at time N and having strike price K, and a call, expiring at time m and having strike price $\frac{K}{(1+r)^{N-m}}$. (Hint: Use put–call parity (Exercise 2.11).)

Exercise 2.13 (Asian option). Consider an N-period binomial model. An *Asian option* has a payoff based on the average stock price, i.e.,

$$V_N = f\left(\frac{1}{N+1}\sum_{n=0}^{N} S_n\right),$$

where the function f is determined by the contractual details of the option.

(i) Define $Y_n = \sum_{k=0}^{n} S_k$ and use the Independence Lemma 2.5.3 to show that the two-dimensional process (S_n, Y_n), $n = 0, 1, \ldots, N$ is Markov.

(ii) According to Theorem 2.5.8, the price V_n of the Asian option at time n is some function v_n of S_n and Y_n; i.e.,

$$V_n = v_n(S_n, Y_n),\ \ n = 0, 1, \ldots, N.$$

Give a formula for $v_N(s, y)$, and provide an algorithm for computing $v_n(s, y)$ in terms of v_{n+1}.

Exercise 2.14 (Asian option continued). Consider an N-period binomial model, and let M be a fixed number between 0 and $N-1$. Consider an Asian option whose payoff at time N is

$$V_N = f\left(\frac{1}{N-M}\sum_{n=M+1}^{N} S_n\right),$$

where again the function f is determined by the contractual details of the option.

(i) Define

$$Y_n = \begin{cases} 0, & \text{if } 0 \le n \le M, \\ \sum_{k=M+1}^{n} S_k, & \text{if } M+1 \le n \le N. \end{cases}$$

Show that the two-dimensional process (S_n, Y_n), $n = 0, 1, \ldots, N$ is Markov (under the risk-neutral measure $\widetilde{\mathbb{P}}$).

(ii) According to Theorem 2.5.8, the price V_n of the Asian option at time n is some function v_n of S_n and Y_n, i.e.,

$$V_n = v_n(S_n, Y_n),\ \ n = 0, 1, \ldots, N.$$

Of course, when $n \le M$, Y_n is not random and does not need to be included in this function. Thus, for such n we should seek a function v_n of S_n alone and have

$$V_n = \begin{cases} v_n(S_n), & \text{if } 0 \le n \le M, \\ v_n(S_n, Y_n), & \text{if } M+1 \le n \le N. \end{cases}$$

Give a formula for $v_N(s, y)$, and provide an algorithm for computing v_n in terms of v_{n+1}. Note that the algorithm is different for $n < M$ and $n > M$, and there is a separate transition formula for $v_M(s)$ in terms of $v_{M+1}(\cdot, \cdot)$.

Springer

http://www.springer.com/978-0-387-40100-3

Stochastic Calculus for Finance I
The Binomial Asset Pricing Model
Shreve, S.
2004, XV, 187 p., Hardcover
ISBN: 978-0-387-40100-3