Chapter 1

Introduction to Probability Theory

1.1 The Binomial Asset Pricing Model

The *binomial asset pricing model* provides a powerful tool to understand arbitrage pricing theory and probability theory. In this course, we shall use it for both these purposes.

In the binomial asset pricing model, we model stock prices in discrete time, assuming that at each step, the stock price will change to one of two possible values. Let us begin with an initial positive stock price S_0. There are two positive numbers, d and u, with

$$0 < d < u, \tag{1.1}$$

such that at the next period, the stock price will be either dS_0 or uS_0. Typically, we take d and u to satisfy $0 < d < 1 < u$, so change of the stock price from S_0 to dS_0 represents a *downward* movement, and change of the stock price from S_0 to uS_0 represents an *upward* movement. It is common to also have $d = \frac{1}{u}$, and this will be the case in many of our examples. However, strictly speaking, for what we are about to do we need to assume only (1.1) and (1.2) below.

Of course, stock price movements are much more complicated than indicated by the binomial asset pricing model. We consider this simple model for three reasons. First of all, within this model the concept of arbitrage pricing and its relation to risk-neutral pricing is clearly illuminated. Secondly, the model is used in practice because with a sufficient number of steps, it provides a good, computationally tractable approximation to continuous-time models. Thirdly, within the binomial model we can develop the theory of conditional expectations and martingales which lies at the heart of continuous-time models.

With this third motivation in mind, we develop notation for the binomial model which is a bit different from that normally found in practice. Let us imagine that we are tossing a coin, and when we get a “Head,” the stock price moves up, but when we get a “Tail,” the price moves down. We denote the price at time 1 by $S_1(H) = uS_0$ if the toss results in head (H), and by $S_1(T) = dS_0$ if it

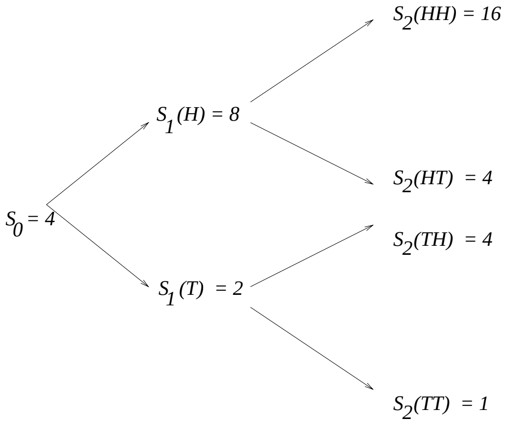

Figure 1.1: *Binomial tree of stock prices with* $S_0 = 4$, $u = 1/d = 2$.

results in tail (T). After the second toss, the price will be one of:

$$S_2(HH) = uS_1(H) = u^2S_0, \quad S_2(HT) = dS_1(H) = duS_0,$$

$$S_2(TH) = uS_1(T) = udS_0, \quad S_2(TT) = dS_1(T) = d^2S_0.$$

After three tosses, there are eight possible coin sequences, although not all of them result in different stock prices at time 3.

For the moment, let us assume that the third toss is the last one and denote by

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

the set of all possible outcomes of the three tosses. The set Ω of all possible outcomes of a random experiment is called the *sample space* for the experiment, and the elements ω of Ω are called *sample points*. In this case, each sample point ω is a sequence of length three. We denote the k-th component of ω by ω_k. For example, when $\omega = HTH$, we have $\omega_1 = H$, $\omega_2 = T$ and $\omega_3 = H$.

The stock price S_k at time k depends on the coin tosses. To emphasize this, we often write $S_k(\omega)$. Actually, this notation does not quite tell the whole story, for while S_3 depends on all of ω, S_2 depends on only the first two components of ω, S_1 depends on only the first component of ω, and S_0 does not depend on ω at all. Sometimes we will use notation such $S_2(\omega_1, \omega_2)$ just to record more explicitly how S_2 depends on $\omega = (\omega_1, \omega_2, \omega_3)$.

Example 1.1 Set $S_0 = 4$, $u = 2$ and $d = \frac{1}{2}$. We have then the binomial "tree" of possible stock prices shown in Fig. 1.1. Each sample point $\omega = (\omega_1, \omega_2, \omega_3)$ represents a path through the tree. Thus, we can think of the sample space Ω as either the set of all possible outcomes from three coin tosses or as the set of all possible paths through the tree.

To complete our binomial asset pricing model, we introduce a *money market* with *interest rate* r; \$1 invested in the money market becomes $\$(1 + r)$ in the next period. We take r to be the interest rate for both *borrowing* and *lending*. (This is not as ridiculous as it first seems, because in a many applications of the model, an agent is either borrowing or lending (not both) and knows in advance which she will be doing; in such an application, she should take r to be the rate of interest for her activity.) We assume that

$$d < 1 + r < u. \tag{1.2}$$

The model would not make sense if we did not have this condition. For example, if $1 + r \geq u$, then the rate of return on the money market is always at least as great as and sometimes greater than the return on the stock, and no one would invest in the stock. The inequality $d \geq 1 + r$ cannot happen unless either r is negative (which never happens, except maybe once upon a time in Switzerland) or $d \geq 1$. In the latter case, the stock does not really go "down" if we get a tail; it just goes up less than if we had gotten a head. One should borrow money at interest rate r and invest in the stock, since even in the worst case, the stock price rises at least as fast as the debt used to buy it.

With the stock as the underlying asset, let us consider a *European call option* with strike price $K > 0$ and expiration time 1. This option confers the right to buy the stock at time 1 for K dollars, and so is worth $S_1 - K$ at time 1 if $S_1 - K$ is positive and is otherwise worth zero. We denote by

$$V_1(\omega) = (S_1(\omega) - K)^+ \overset{\Delta}{=} \max\{S_1(\omega) - K, 0\}$$

the value (payoff) of this option at expiration. Of course, $V_1(\omega)$ actually depends only on ω_1, and we can and do sometimes write $V_1(\omega_1)$ rather than $V_1(\omega)$. Our first task is to compute the *arbitrage price* of this option at time zero.

Suppose at time zero you sell the call for V_0 dollars, where V_0 is still to be determined. You now have an obligation to pay off $(uS_0 - K)^+$ if $\omega_1 = H$ and to pay off $(dS_0 - K)^+$ if $\omega_1 = T$. At the time you sell the option, you don't yet know which value ω_1 will take. You *hedge* your short position in the option by buying Δ_0 shares of stock, where Δ_0 is still to be determined. You can use the proceeds V_0 of the sale of the option for this purpose, and then borrow if necessary at interest rate r to complete the purchase. If V_0 is more than necessary to buy the Δ_0 shares of stock, you invest the residual money at interest rate r. In either case, you will have $V_0 - \Delta_0 S_0$ dollars invested in the money market, where this quantity might be negative. You will also own Δ_0 shares of stock.

If the stock goes up, the value of your portfolio (excluding the short position in the option) is

$$\Delta_0 S_1(H) + (1 + r)(V_0 - \Delta_0 S_0),$$

and you need to have $V_1(H)$. Thus, you want to choose V_0 and Δ_0 so that

$$V_1(H) = \Delta_0 S_1(H) + (1 + r)(V_0 - \Delta_0 S_0). \tag{1.3}$$

If the stock goes down, the value of your portfolio is

$$\Delta_0 S_1(T) + (1 + r)(V_0 - \Delta_0 S_0),$$

and you need to have $V_1(T)$. Thus, you want to choose V_0 and Δ_0 to also have

$$V_1(T) = \Delta_0 S_1(T) + (1 + r)(V_0 - \Delta_0 S_0). \tag{1.4}$$

These are two equations in two unknowns, and we solve them below

Subtracting (1.4) from (1.3), we obtain

$$V_1(H) - V_1(T) = \Delta_0(S_1(H) - S_1(T)), \tag{1.5}$$

so that

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}. \tag{1.6}$$

This is a discrete-time version of the famous "delta-hedging" formula for derivative securities, according to which the number of shares of an underlying asset a hedge should hold is the derivative (in the sense of calculus) of the value of the derivative security with respect to the price of the underlying asset. This formula is so pervasive the when a practitioner says "delta", she means the derivative (in the sense of calculus) just described. Note, however, that my *definition* of Δ_0 is the number of shares of stock one holds at time zero, and (1.6) is a consequence of this definition, not the definition of Δ_0 itself. Depending on how uncertainty enters the model, there can be cases in which the number of shares of stock a hedge should hold is not the (calculus) derivative of the derivative security with respect to the price of the underlying asset.

To complete the solution of (1.3) and (1.4), we substitute (1.6) into either (1.3) or (1.4) and solve for V_0. After some simplification, this leads to the formula

$$V_0 = \frac{1}{1+r}\left[\frac{1+r-d}{u-d}V_1(H) + \frac{u-(1+r)}{u-d}V_1(T)\right]. \tag{1.7}$$

This is the *arbitrage price* for the European call option with payoff V_1 at time 1. To simplify this formula, we define

$$\tilde{p} \triangleq \frac{1+r-d}{u-d}, \quad \tilde{q} \triangleq \frac{u-(1+r)}{u-d} = 1-\tilde{p}, \tag{1.8}$$

so that (1.7) becomes

$$V_0 = \frac{1}{1+r}[\tilde{p}V_1(H) + \tilde{q}V_1(T)]. \tag{1.9}$$

Because we have taken $d < u$, both $\tilde{p}$ and $\tilde{q}$ are defined,i.e., the denominator in (1.8) is not zero. Because of (1.2), both $\tilde{p}$ and $\tilde{q}$ are in the interval $(0,1)$, and because they sum to 1, we can regard them as probabilities of H and T, respectively. They are the *risk-neutral* probabilites. They appeared when we solved the two equations (1.3) and (1.4), and have nothing to do with the actual probabilities of getting H or T on the coin tosses. In fact, at this point, they are nothing more than a convenient tool for writing (1.7) as (1.9).

We now consider a European call which pays off K dollars at time 2. At expiration, the payoff of this option is $V_2 \triangleq (S_2 - K)^+$, where V_2 and S_2 depend on ω_1 and ω_2, the first and second coin tosses. We want to determine the arbitrage price for this option at time zero. Suppose an agent sells the option at time zero for V_0 dollars, where V_0 is still to be determined. She then buys Δ_0 shares of stock, investing $V_0 - \Delta_0 S_0$ dollars in the money market to finance this. At time 1, the agent has a portfolio (excluding the short position in the option) valued at

$$X_1 \stackrel{\Delta}{=} \Delta_0 S_1 + (1+r)(V_0 - \Delta_0 S_0). \tag{1.10}$$

Although we do not indicate it in the notation, S_1 and therefore X_1 depend on ω_1, the outcome of the first coin toss. Thus, there are really two equations implicit in (1.10):

$$\begin{aligned} X_1(H) &\stackrel{\Delta}{=} \Delta_0 S_1(H) + (1+r)(V_0 - \Delta_0 S_0), \\ X_1(T) &\stackrel{\Delta}{=} \Delta_0 S_1(T) + (1+r)(V_0 - \Delta_0 S_0). \end{aligned}$$

After the first coin toss, the agent has X_1 dollars and can readjust her hedge. Suppose she decides to now hold Δ_1 shares of stock, where Δ_1 is allowed to depend on ω_1 because the agent knows what value ω_1 has taken. She invests the remainder of her wealth, $X_1 - \Delta_1 S_1$ in the money market. In the next period, her wealth will be given by the right-hand side of the following equation, and she wants it to be V_2. Therefore, she wants to have

$$V_2 = \Delta_1 S_2 + (1+r)(X_1 - \Delta_1 S_1). \tag{1.11}$$

Although we do not indicate it in the notation, S_2 and V_2 depend on ω_1 and ω_2, the outcomes of the first two coin tosses. Considering all four possible outcomes, we can write (1.11) as four equations:

$$\begin{aligned} V_2(HH) &= \Delta_1(H) S_2(HH) + (1+r)(X_1(H) - \Delta_1(H) S_1(H)), \\ V_2(HT) &= \Delta_1(H) S_2(HT) + (1+r)(X_1(H) - \Delta_1(H) S_1(H)), \\ V_2(TH) &= \Delta_1(T) S_2(TH) + (1+r)(X_1(T) - \Delta_1(T) S_1(T)), \\ V_2(TT) &= \Delta_1(T) S_2(TT) + (1+r)(X_1(T) - \Delta_1(T) S_1(T)). \end{aligned}$$

We now have six equations, the two represented by (1.10) and the four represented by (1.11), in the six unknowns V_0, Δ_0, $\Delta_1(H)$, $\Delta_1(T)$, $X_1(H)$, and $X_1(T)$.

To solve these equations, and thereby determine the arbitrage price V_0 at time zero of the option and the hedging portfolio Δ_0, $\Delta_1(H)$ and $\Delta_1(T)$, we begin with the last two

$$\begin{aligned} V_2(TH) &= \Delta_1(T) S_2(TH) + (1+r)(X_1(T) - \Delta_1(T) S_1(T)), \\ V_2(TT) &= \Delta_1(T) S_2(TT) + (1+r)(X_1(T) - \Delta_1(T) S_1(T)). \end{aligned}$$

Subtracting one of these from the other and solving for $\Delta_1(T)$, we obtain the "delta-hedging formula"

$$\Delta_1(T) = \frac{V_2(TH) - V_2(TT)}{S_2(TH) - S_2(TT)}, \tag{1.12}$$

and substituting this into either equation, we can solve for

$$X_1(T) = \frac{1}{1+r}[\tilde{p} V_2(TH) + \tilde{q} V_2(TT)]. \tag{1.13}$$

Equation (1.13), gives the value the hedging portfolio should have at time 1 if the stock goes down between times 0 and 1. We define this quantity to be the *arbitrage value of the option at time* 1 *if* $\omega_1 = T$, and we denote it by $V_1(T)$. We have just shown that

$$V_1(T) \triangleq \frac{1}{1+r}[\tilde{p}V_2(TH) + \tilde{q}V_2(TT)]. \tag{1.14}$$

The hedger should choose her portfolio so that her wealth $X_1(T)$ if $\omega_1 = T$ agrees with $V_1(T)$ defined by (1.14). This formula is analgous to formula (1.9), but postponed by one step. The first two equations implicit in (1.11) lead in a similar way to the formulas

$$\Delta_1(H) = \frac{V_2(HH) - V_2(HT)}{S_2(HH) - S_2(HT)} \tag{1.15}$$

and $X_1(H) = V_1(H)$, where $V_1(H)$ is the value of the option at time 1 if $\omega_1 = H$, defined by

$$V_1(H) \triangleq \frac{1}{1+r}[\tilde{p}V_2(HH) + \tilde{q}V_2(HT)]. \tag{1.16}$$

This is again analgous to formula (1.9), postponed by one step. Finally, we plug the values $X_1(H) = V_1(H)$ and $X_1(T) = V_1(T)$ into the two equations implicit in (1.10). The solution of these equations for Δ_0 and V_0 is the same as the solution of (1.3) and (1.4), and results again in (1.6) and (1.9).

The pattern emerging here persists, regardless of the number of periods. If V_k denotes the value at time k of a derivative security, and this depends on the first k coin tosses $\omega_1, \ldots, \omega_k$, then at time $k-1$, after the first $k-1$ tosses $\omega_1, \ldots, \omega_{k-1}$ are known, the portfolio to hedge a short position should hold $\Delta_{k-1}(\omega_1, \ldots, \omega_{k-1})$ shares of stock, where

$$\Delta_{k-1}(\omega_1, \ldots, \omega_{k-1}) = \frac{V_k(\omega_1, \ldots, \omega_{k-1}, H) - V_k(\omega_1, \ldots, \omega_{k-1}, T)}{S_k(\omega_1, \ldots, \omega_{k-1}, H) - S_k(\omega_1, \ldots, \omega_{k-1}, T)}, \tag{1.17}$$

and the value at time $k-1$ of the derivative security, when the first $k-1$ coin tosses result in the outcomes $\omega_1, \ldots, \omega_{k-1}$, is given by

$$V_{k-1}(\omega_1, \ldots, \omega_{k-1}) = \frac{1}{1+r}[\tilde{p}V_k(\omega_1, \ldots, \omega_{k-1}, H) + \tilde{q}V_k(\omega_1, \ldots, \omega_{k-1}, T)] \tag{1.18}$$

1.2 Finite Probability Spaces

Let Ω be a set with finitely many elements. An example to keep in mind is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \tag{2.1}$$

of all possible outcomes of three coin tosses. Let $\mathcal{F}$ be the set of all subsets of Ω. Some sets in $\mathcal{F}$ are $\emptyset$, $\{HHH, HHT, HTH, HTT\}$, $\{TTT\}$, and Ω itself. How many sets are there in $\mathcal{F}$?

Definition 1.1 A *probability measure* $I\!\!P$ is a function mapping $\mathcal{F}$ into $[0,1]$ with the following properties:

(i) $I\!\!P(\Omega) = 1$,

(ii) If $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$I\!\!P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} I\!\!P(A_k).$$

Probability measures have the following interpretation. Let A be a subset of $\mathcal{F}$. Imagine that Ω is the set of all possible outcomes of some random experiment. There is a certain probability, between 0 and 1, that when that experiment is performed, the outcome will lie in the set A. We think of $I\!\!P(A)$ as this probability.

Example 1.2 Suppose a coin has probability $\frac{1}{3}$ for H and $\frac{2}{3}$ for T. For the individual elements of Ω in (2.1), define

$$\begin{aligned}
I\!\!P\{HHH\} &= \left(\tfrac{1}{3}\right)^3, & I\!\!P\{HHT\} &= \left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right),\\
I\!\!P\{HTH\} &= \left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right), & I\!\!P\{HTT\} &= \left(\tfrac{1}{3}\right)\left(\tfrac{2}{3}\right)^2,\\
I\!\!P\{THH\} &= \left(\tfrac{1}{3}\right)^2\left(\tfrac{1}{3}\right), & I\!\!P\{THT\} &= \left(\tfrac{1}{3}\right)\left(\tfrac{2}{3}\right)^2,\\
I\!\!P\{TTH\} &= \left(\tfrac{1}{3}\right)\left(\tfrac{2}{3}\right)^2, & I\!\!P\{TTT\} &= \left(\tfrac{2}{3}\right)^3.
\end{aligned}$$

For $A \in \mathcal{F}$, we define

$$I\!\!P(A) = \sum_{\omega \in A} I\!\!P\{\omega\}. \tag{2.2}$$

For example,

$$I\!\!P\{HHH, HHT, HTH, HTT\} = \left(\frac{1}{3}\right)^3 + 2\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^2 = \frac{1}{3},$$

which is another way of saying that the probability of H on the first toss is $\frac{1}{3}$.

As in the above example, it is generally the case that we specify a probability measure on only some of the subsets of Ω and then use property (ii) of Definition 1.1 to determine $I\!\!P(A)$ for the remaining sets $A \in \mathcal{F}$. In the above example, we specified the probability measure only for the sets containing a single element, and then used Definition 1.1(ii) in the form (2.2) (see Problem 1.4(ii)) to determine $I\!\!P$ for all the other sets in $\mathcal{F}$.

Definition 1.2 Let Ω be a nonempty set. A σ-algebra is a collection $\mathcal{G}$ of subsets of Ω with the following three properties:

(i) $\emptyset \in \mathcal{G}$,

(ii) If $A \in \mathcal{G}$, then its complement $A^c \in \mathcal{G}$,

(iii) If $A_1, A_2, A_3, \ldots$ is a sequence of sets in $\mathcal{G}$, then $\cup_{k=1}^{\infty} A_k$ is also in $\mathcal{G}$.

Here are some important σ-algebras of subsets of the set Ω in Example 1.2:

$$\begin{aligned}
\mathcal{F}_0 &= \left\{\emptyset, \Omega\right\}, \\
\mathcal{F}_1 &= \left\{\emptyset, \Omega, \{HHH, HHT, HTH, HTT\}, \{THH, THT, TTH, TTT\}\right\}, \\
\mathcal{F}_2 &= \Big\{\emptyset, \Omega, \{HHH, HHT\}, \{HTH, HTT\}, \{THH, THT\}, \{TTH, TTT\}, \\
&\qquad\qquad \text{and all sets which can be built by taking unions of these}\Big\}, \\
\mathcal{F}_3 &= \mathcal{F} = \text{The set of all subsets of } \Omega.
\end{aligned}$$

To simplify notation a bit, let us define

$$\begin{aligned}
A_H &\stackrel{\Delta}{=} \{HHH, HHT, HTH, HTT\} = \{H \text{ on the first toss}\}, \\
A_T &\stackrel{\Delta}{=} \{THH, THT, TTH, TTT\} = \{T \text{ on the first toss}\},
\end{aligned}$$

so that

$$\mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\},$$

and let us define

$$\begin{aligned}
A_{HH} &\stackrel{\Delta}{=} \{HHH, HHT\} = \{HH \text{ on the first two tosses}\}, \\
A_{HT} &\stackrel{\Delta}{=} \{HTH, HTT\} = \{HT \text{ on the first two tosses}\}, \\
A_{TH} &\stackrel{\Delta}{=} \{THH, THT\} = \{TH \text{ on the first two tosses}\}, \\
A_{TT} &\stackrel{\Delta}{=} \{TTH, TTT\} = \{TT \text{ on the first two tosses}\},
\end{aligned}$$

so that

$$\begin{aligned}
\mathcal{F}_2 = \{&\emptyset, \Omega, A_{HH}, A_{HT}, A_{TH}, A_{TT}, \\
&A_H, A_T, A_{HH} \cup A_{TH}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, A_{HT} \cup A_{TT}, \\
&A_{HH}^c, A_{HT}^c, A_{TH}^c, A_{TT}^c\}.
\end{aligned}$$

We interpret σ-algebras as a record of information. Suppose the coin is tossed three times, and you are not told the outcome, but you are told, for every set in $\mathcal{F}_1$ whether or not the outcome is in that set. For example, you would be told that the outcome is not in $\emptyset$ and is in Ω. Moreover, you might be told that the outcome is not in A_H but is in A_T. In effect, you have been told that the first toss was a T, and nothing more. The σ-algebra $\mathcal{F}_1$ is said to contain the "information of the first toss", which is usually called the "information up to time 1". Similarly, $\mathcal{F}_2$ contains the "information of the first two tosses," which is the "information up to time 2." The σ-algebra $\mathcal{F}_3 = \mathcal{F}$ contains "full information" about the outcome of all three tosses. The so-called "trivial" σ-algebra $\mathcal{F}_0$ contains no information. Knowing whether the outcome ω of the three tosses is in $\emptyset$ (it is not) and whether it is in Ω (it is) tells you nothing about ω

Definition 1.3 Let Ω be a nonempty finite set. A *filtration* is a sequence of σ-algebras $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ such that each σ-algebra in the sequence contains all the sets contained by the previous σ-algebra.

Definition 1.4 Let Ω be a nonempty finite set and let $\mathcal{F}$ be the σ-algebra of all subsets of Ω. A random variable is a function mapping Ω into $I\!\!R$.

Example 1.3 Let Ω be given by (2.1) and consider the binomial asset pricing Example 1.1, where $S_0 = 4$, $u = 2$ and $d = \frac{1}{2}$. Then S_0, S_1, S_2 and S_3 are all random variables. For example, $S_2(HHT) = u^2 S_0 = 16$. The "random variable" S_0 is really not random, since $S_0(\omega) = 4$ for all $\omega \in \Omega$. Nonetheless, it is a function mapping Ω into $I\!\!R$, and thus technically a random variable, albeit a degenerate one.

A random variable maps Ω into $I\!\!R$, and we can look at the preimage under the random variable of sets in $I\!\!R$. Consider, for example, the random variable S_2 of Example 1.1. We have

$$\begin{aligned}
&S_2(HHH) = S_2(HHT) = 16,\\
&S_2(HTH) = S_2(HTT) = S_2(THH) = S_2(THT) = 4,\\
&S_2(TTH) = S_2(TTT) = 1.
\end{aligned}$$

Let us consider the interval $[4, 27]$. The preimage under S_2 of this interval is defined to be

$$\{\omega \in \Omega; S_2(\omega) \in [4, 27]\} = \{\omega \in \Omega; 4 \le S_2 \le 27\} = A^c_{TT}.$$

The complete list of subsets of Ω we can get as preimages of sets in $I\!\!R$ is:

$$\emptyset, \Omega, A_{HH}, A_{HT} \cup A_{TH}, A_{TT},$$

and sets which can be built by taking unions of these. This collection of sets is a σ-algebra, called the *σ-algebra generated by the random variable* S_2, and is denoted by $\sigma(S_2)$. The information content of this σ-algebra is exactly the information learned by observing S_2. More specifically, suppose the coin is tossed three times and you do not know the outcome ω, but someone is willing to tell you, for each set in $\sigma(S_2)$, whether ω is in the set. You might be told, for example, that ω is not in A_{HH}, is in $A_{HT} \cup A_{TH}$, and is not in A_{TT}. Then you know that in the first two tosses, there was a head and a tail, and you know nothing more. This information is the same you would have gotten by being told that the value of $S_2(\omega)$ is 4.

Note that $\mathcal{F}_2$ defined earlier contains all the sets which are in $\sigma(S_2)$, and even more. This means that the information in the first two tosses is greater than the information in S_2. In particular, if you see the first two tosses, you can distinguish A_{HT} from A_{TH}, but you cannot make this distinction from knowing the value of S_2 alone.

Definition 1.5 Let Ω be a nonemtpy finite set and let $\mathcal{F}$ be the σ-algebra of all subsets of Ω. Let X be a random variable on $(\Omega, \mathcal{F})$. The *σ-algebra $\sigma(X)$ generated by X* is defined to be the collection of all sets of the form $\{\omega \in \Omega; X(\omega) \in A\}$, where A is a subset of $I\!R$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. We say that *X is $\mathcal{G}$-measurable* if every set in $\sigma(X)$ is also in $\mathcal{G}$.

Note: We normally write simply $\{X \in A\}$ rather than $\{\omega \in \Omega; X(\omega) \in A\}$.

Definition 1.6 Let Ω be a nonempty, finite set, let $\mathcal{F}$ be the σ-algebra of all subsets of Ω, let $I\!P$ be a probabilty measure on $(\Omega, \mathcal{F})$, and let X be a random variable on Ω. Given any set $A \subseteq I\!R$, we define the *induced measure* of A to be

$$\mathcal{L}_X(A) \triangleq I\!P\{X \in A\}.$$

In other words, the induced measure of a set A tells us the probability that X takes a value in A. In the case of S_2 above with the probability measure of Example 1.2, some sets in $I\!R$ and their induced measures are:

$$\begin{aligned}
&\mathcal{L}_{S_2}(\emptyset) = I\!P(\emptyset) = 0,\\
&\mathcal{L}_{S_2}(I\!R) = I\!P(\Omega) = 1,\\
&\mathcal{L}_{S_2}[0,\infty) = I\!P(\Omega) = 1,\\
&\mathcal{L}_{S_2}[0,3] = I\!P\{S_2 = 1\} = I\!P(A_{TT}) = \left(\frac{2}{3}\right)^2.
\end{aligned}$$

In fact, the induced measure of S_2 places a mass of size $\left(\frac{1}{3}\right)^2 = \frac{1}{9}$ at the number 16, a mass of size $\frac{4}{9}$ at the number 4, and a mass of size $\left(\frac{2}{3}\right)^2 = \frac{4}{9}$ at the number 1. A common way to record this information is to give the *cumulative distribution function* $F_{S_2}(x)$ of S_2, defined by

$$F_{S_2}(x) \triangleq I\!P(S_2 \le x) = \begin{cases} 0, & \text{if } x < 1,\\ \frac{4}{9}, & \text{if } 1 \le x < 4,\\ \frac{8}{9}, & \text{if } 4 \le x < 16,\\ 1, & \text{if } 16 \le x. \end{cases} \tag{2.3}$$

By the *distribution* of a random variable X, we mean any of the several ways of characterizing $\mathcal{L}_X$. If X is discrete, as in the case of S_2 above, we can either tell where the masses are and how large they are, or tell what the cumulative distribution function is. (Later we will consider random variables X which have densities, in which case the induced measure of a set $A \subseteq I\!R$ is the integral of the density over the set A.)

Important Note. In order to work through the concept of a risk-neutral measure, we set up the definitions to make a clear distinction between random variables and their distributions.

A *random variable* is a mapping from Ω to $I\!R$, nothing more. It has an existence quite apart from discussion of probabilities. For example, in the discussion above, $S_2(TTH) = S_2(TTT) = 1$, regardless of whether the probability for H is $\frac{1}{3}$ or $\frac{1}{2}$.

The *distribution* of a random variable is a measure $\mathcal{L}_X$ on $I\!R$, i.e., a way of assigning probabilities to sets in $I\!R$. It depends on the random variable X and the probability measure $I\!P$ we use in Ω. If we set the probability of H to be $\frac{1}{3}$, then $\mathcal{L}_{S_2}$ assigns mass $\frac{1}{9}$ to the number 16. If we set the probability of H to be $\frac{1}{2}$, then $\mathcal{L}_{S_2}$ assigns mass $\frac{1}{4}$ to the number 16. The distribution of S_2 has changed, but the random variable has not. It is still defined by

$$\begin{aligned}
&S_2(HHH) = S_2(HHT) = 16,\\
&S_2(HTH) = S_2(HTT) = S_2(THH) = S_2(THT) = 4,\\
&S_2(TTH) = S_2(TTT) = 1.
\end{aligned}$$

Thus, a random variable can have more than one distribution (a “market” or “objective” distribution, and a “risk-neutral” distribution).

In a similar vein, two *different random variables* can have the *same distribution.* Suppose in the binomial model of Example 1.1, the probability of H and the probability of T is $\frac{1}{2}$. Consider a European call with strike price 14 expiring at time 2. The payoff of the call at time 2 is the random variable $(S_2 - 14)^+$, which takes the value 2 if $\omega = HHH$ or $\omega = HHT$, and takes the value 0 in every other case. The probability the payoff is 2 is $\frac{1}{4}$, and the probability it is zero is $\frac{3}{4}$. Consider also a European put with strike price 3 expiring at time 2. The payoff of the put at time 2 is $(3 - S_2)^+$, which takes the value 2 if $\omega = TTH$ or $\omega = TTT$. Like the payoff of the call, the payoff of the put is 2 with probability $\frac{1}{4}$ and 0 with probability $\frac{3}{4}$. The payoffs of the call and the put are different random variables having the same distribution.

Definition 1.7 Let Ω be a nonempty, finite set, let $\mathcal{F}$ be the σ-algebra of all subsets of Ω, let $I\!P$ be a probabilty measure on $(\Omega, \mathcal{F})$, and let X be a random variable on Ω. The *expected value* of X is defined to be

$$I\!EX \stackrel{\Delta}{=} \sum_{\omega \in \Omega} X(\omega) I\!P\{\omega\}. \tag{2.4}$$

Notice that the expected value in (2.4) is defined to be a sum *over the sample space* Ω. Since Ω is a finite set, X can take only finitely many values, which we label $x_1, \ldots, x_n$. We can partition Ω into the subsets $\{X_1 = x_1\}, \ldots, \{X_n = x_n\}$, and then rewrite (2.4) as

$$\begin{aligned}
I\!EX &\stackrel{\Delta}{=} \sum_{\omega \in \Omega} X(\omega) I\!P\{\omega\}\\
&= \sum_{k=1}^{n} \sum_{\omega \in \{X_k = x_k\}} X(\omega) I\!P\{\omega\}\\
&= \sum_{k=1}^{n} x_k \sum_{\omega \in \{X_k = x_k\}} I\!P\{\omega\}\\
&= \sum_{k=1}^{n} x_k I\!P\{X_k = x_k\}\\
&= \sum_{k=1}^{n} x_k \mathcal{L}_X\{x_k\}.
\end{aligned}$$

Thus, although the expected value is defined as a sum over the sample space Ω, we can also write it as a sum over $I\!\!R$.

To make the above set of equations absolutely clear, we consider S_2 with the distribution given by (2.3). The definition of $I\!\!ES_2$ is

$$\begin{aligned} I\!\!ES_2 &= S_2(HHH)I\!\!P\{HHH\} + S_2(HHT)I\!\!P\{HHT\} \\ &\quad +S_2(HTH)I\!\!P\{HTH\} + S_2(HTT)I\!\!P\{HTT\} \\ &\quad +S_2(THH)I\!\!P\{THH\} + S_2(THT)I\!\!P\{THT\} \\ &\quad +S_2(TTH)I\!\!P\{TTH\} + S_2(TTT)I\!\!P\{TTT\} \\ &= 16 \cdot I\!\!P(A_{HH}) + 4 \cdot I\!\!P(A_{HT} \cup A_{TH}) + 1 \cdot I\!\!P(A_{TT}) \\ &= 16 \cdot I\!\!P\{S_2 = 16\} + 4 \cdot I\!\!P\{S_2 = 4\} + 1 \cdot I\!\!P\{S_2 = 1\} \\ &= 16 \cdot \mathcal{L}_{S_2}\{16\} + 4 \cdot \mathcal{L}_{S_2}\{4\} + 1 \cdot \mathcal{L}_{S_2}\{1\} \\ &= 16 \cdot \frac{1}{9} + 4 \cdot \frac{4}{9} + 4 \cdot \frac{4}{9} \\ &= \frac{48}{9}. \end{aligned}$$

Definition 1.8 Let Ω be a nonempty, finite set, let $\mathcal{F}$ be the σ-algebra of all subsets of Ω, let $I\!\!P$ be a probabilty measure on $(\Omega, \mathcal{F})$, and let X be a random variable on Ω. The *variance* of X is defined to be the expected value of $(X - I\!\!EX)^2$, i.e.,

$$\text{Var}(X) \stackrel{\Delta}{=} \sum_{\omega \in \Omega} (X(\omega) - I\!\!EX)^2 I\!\!P\{\omega\}. \tag{2.5}$$

One again, we can rewrite (2.5) as a sum over $I\!\!R$ rather than over Ω. Indeed, if X takes the values $x_1, \ldots, x_n$, then

$$\text{Var}(X) = \sum_{k=1}^{n} (x_k - I\!\!EX)^2 I\!\!P\{X = x_k\} = \sum_{k=1}^{n} (x_k - I\!\!EX)^2 \mathcal{L}_X(x_k).$$

1.3 Lebesgue Measure and the Lebesgue Integral

In this section, we consider the set of real numbers $I\!\!R$, which is uncountably infinite. We define the *Lebesgue measure* of intervals in $I\!\!R$ to be their length. This definition and the properties of measure determine the Lebesgue measure of many, but not all, subsets of $I\!\!R$. The collection of subsets of $I\!\!R$ we consider, and for which Lebesgue measure is defined, is the collection of *Borel sets* defined below.

We use Lebesgue measure to construct the *Lebesgue integral*, a generalization of the Riemann integral. We need this integral because, unlike the Riemann integral, it can be defined on abstract spaces, such as the space of infinite sequences of coin tosses or the space of paths of Brownian motion. This section concerns the Lebesgue integral on the space $I\!\!R$ only; the generalization to other spaces will be given later.

Definition 1.9 The *Borel σ-algebra*, denoted $\mathcal{B}(\mathbb{R})$, is the smallest σ-algebra containing all open intervals in $\mathbb{R}$. The sets in $\mathcal{B}(\mathbb{R})$ are called *Borel sets*.

Every set which can be written down and just about every set imaginable is in $\mathcal{B}(\mathbb{R})$. The following discussion of this fact uses the σ-algebra properties developed in Problem 1.3.

By definition, every open interval (a,b) is in $\mathcal{B}(\mathbb{R})$, where a and b are real numbers. Since $\mathcal{B}(\mathbb{R})$ is a σ-algebra, every union of open intervals is also in $\mathcal{B}(\mathbb{R})$. For example, for every real number a, the *open half-line*

$$(a,\infty) = \bigcup_{n=1}^{\infty} (a, a+n)$$

is a Borel set, as is

$$(-\infty, a) = \bigcup_{n=1}^{\infty} (a-n, a).$$

For real numbers a and b, the union

$$(-\infty, a) \cup (b, \infty)$$

is Borel. Since $\mathcal{B}(\mathbb{R})$ is a σ-algebra, every complement of a Borel set is Borel, so $\mathcal{B}(\mathbb{R})$ contains

$$[a,b] = \Big((-\infty,a)\cup(b,\infty)\Big)^c.$$

This shows that every closed interval is Borel. In addition, the *closed half-lines*

$$[a,\infty) = \bigcup_{n=1}^{\infty} [a, a+n]$$

and

$$(-\infty, a] = \bigcup_{n=1}^{\infty} [a-n, a]$$

are Borel. Half-open and half-closed intervals are also Borel, since they can be written as intersections of open half-lines and closed half-lines. For example,

$$(a,b] = (-\infty, b] \cap (a, \infty).$$

Every set which contains only one real number is Borel. Indeed, if a is a real number, then

$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n}\right).$$

This means that every set containing finitely many real numbers is Borel; if $A = \{a_1, a_2, \ldots, a_n\}$, then

$$A = \bigcup_{k=1}^{n} \{a_k\}.$$

In fact, every set containing countably infinitely many numbers is Borel; if $A = \{a_1, a_2, \ldots\}$, then

$$A = \bigcup_{k=1}^{n} \{a_k\}.$$

This means that the set of rational numbers is Borel, as is its complement, the set of irrational numbers.

There are, however, sets which are not Borel. We have just seen that any non-Borel set must have uncountably many points.

Example 1.4 (The Cantor set.) *This example gives a hint of how complicated a Borel set can be. We use it later when we discuss the sample space for an infinite sequence of coin tosses.*

Consider the unit interval $[0,1]$, *and remove the middle half, i.e., remove the open interval*

$$A_1 \triangleq \left(\frac{1}{4}, \frac{3}{4}\right).$$

The remaining set

$$C_1 = \left[0, \frac{1}{4}\right] \cup \left[\frac{3}{4}, 1\right]$$

has two pieces. From each of these pieces, remove the middle half, i.e., remove the open set

$$A_2 \triangleq \left(\frac{1}{16}, \frac{3}{16}\right) \bigcup \left(\frac{13}{16}, \frac{15}{16}\right).$$

The remaining set

$$C_2 = \left[0, \frac{1}{16}\right] \bigcup \left[\frac{3}{16}, \frac{1}{4}\right] \bigcup \left[\frac{3}{4}, \frac{13}{16}\right] \bigcup \left[\frac{15}{16}, 1\right].$$

has four pieces. Continue this process, so at stage k, *the set* C_k *has* 2^k *pieces, and each piece has length* $\frac{1}{4^k}$. *The* Cantor set

$$C \triangleq \bigcap_{k=1}^{\infty} C_k$$

is defined to be the set of points not removed at any stage of this nonterminating process.

Note that the length of A_1, *the first set removed, is* $\frac{1}{2}$. *The "length" of* A_2, *the second set removed, is* $\frac{1}{8} + \frac{1}{8} = \frac{1}{4}$. *The "length" of the next set removed is* $4 \cdot \frac{1}{32} = \frac{1}{8}$, *and in general, the length of the* k*-th set removed is* 2^{-k}. *Thus, the total length removed is*

$$\sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

and so the Cantor set, the set of points not removed, has zero "length."

Despite the fact that the Cantor set has no "length," there are lots of points in this set. In particular, none of the endpoints of the pieces of the sets $C_1, C_2, \ldots$ *is ever removed. Thus, the points*

$$0, \frac{1}{4}, \frac{3}{4}, 1, \frac{1}{16}, \frac{3}{16}, \frac{13}{16}, \frac{15}{16}, \frac{1}{64}, \ldots$$

are all in C. *This is a countably infinite set of points. We shall see eventually that the Cantor set has uncountably many points.* ◇

Definition 1.10 Let $\mathcal{B}(\mathbb{R})$ be the σ-algebra of Borel subsets of $\mathbb{R}$. A *measure on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a function μ mapping $\mathcal{B}$ into $[0,\infty]$ with the following properties:

(i) $\mu(\emptyset) = 0$,

(ii) If $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{B}(\mathbb{R})$, then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

Lebesgue measure is defined to be the measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which assigns the measure of each interval to be its length. Following Williams's book, we denote Lebesgue measure by μ_0.

A measure has all the properties of a probability measure given in Problem 1.4, except that the total measure of the space is not necessarily 1 (in fact, $\mu_0(\mathbb{R}) = \infty$), one no longer has the equation

$$\mu(A^c) = 1 - \mu(A)$$

in Problem 1.4(iii), and property (v) in Problem 1.4 needs to be modified to say:

(v) If $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{B}(\mathbb{R})$ with $A_1 \supseteq A_2 \supseteq \cdots$ *and* $\mu(A_1) < \infty$, then

$$\mu\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} \mu(A_n).$$

To see that the additional requirment $\mu(A_1) < \infty$ is needed in (v), consider

$$A_1 = [1,\infty), A_2 = [2,\infty), A_3 = [3,\infty), \ldots.$$

Then $\cap_{k=1}^{\infty} A_k = \emptyset$, so $\mu_0(\cap_{k=1}^{\infty} A_k) = 0$, but $\lim_{n\to\infty} \mu_0(A_n) = \infty$.

We specify that the Lebesgue measure of each interval is its length, and that determines the Lebesgue measure of all other Borel sets. For example, the Lebesgue measure of the Cantor set in Example 1.4 must be zero, because of the "length" computation given at the end of that example.

The Lebesgue measure of a set containing only one point must be zero. In fact, since

$$\{a\} \subseteq \left(a - \frac{1}{n}, a + \frac{1}{n}\right)$$

for every positive integer n, we must have

$$0 \le \mu_0\{a\} \le \mu_0\left(a - \frac{1}{n}, a + \frac{1}{n}\right) = \frac{2}{n}.$$

Letting $n \to \infty$, we obtain

$$\mu_0\{a\} = 0.$$

The Lebesgue measure of a set containing countably many points must also be zero. Indeed, if $A = \{a_1, a_2, \ldots\}$, then

$$\mu_0(A) = \sum_{k=1}^{\infty} \mu_0\{a_k\} = \sum_{k=1}^{\infty} 0 = 0.$$

The Lebesgue measure of a set containing uncountably many points can be either zero, positive and finite, or infinite. We may not compute the Lebesgue measure of an uncountable set by adding up the Lebesgue measure of its individual members, because there is no way to add up uncountably many numbers. The integral was invented to get around this problem.

In order to think about Lebesgue integrals, we must first consider the functions to be integrated.

Definition 1.11 Let f be a function from $I\!\!R$ to $I\!\!R$. We say that f is *Borel-measurable* if the set $\{x \in I\!\!R; f(x) \in A\}$ is in $\mathcal{B}(I\!\!R)$ whenever $A \in \mathcal{B}(I\!\!R)$. In the language of Section 2, we want the *σ-algebra generated by* f to be contained in $\mathcal{B}(I\!\!R)$.

Definition 3.4 is purely technical and has nothing to do with keeping track of information. It is difficult to conceive of a function which is not Borel-measurable, and we shall pretend such functions don't exist. Hencefore, "function mapping $I\!\!R$ to $I\!\!R$" will mean "Borel-measurable function mapping $I\!\!R$ to $I\!\!R$" and "subset of $I\!\!R$" will mean "Borel subset of $I\!\!R$".

Definition 1.12 An *indicator function* g from $I\!\!R$ to $I\!\!R$ is a function which takes only the values 0 and 1. We call

$$A \stackrel{\Delta}{=} \{x \in I\!\!R; g(x) = 1\}$$

the set *indicated* by g. We define the *Lebesgue integral* of g to be

$$\int_{I\!\!R} g \, d\mu_0 \stackrel{\Delta}{=} \mu_0(A).$$

A *simple function* h from $I\!\!R$ to $I\!\!R$ is a linear combination of indicators, i.e., a function of the form

$$h(x) = \sum_{k=1}^{n} c_k g_k(x),$$

where each g_k is of the form

$$g_k(x) = \begin{cases} 1, & \text{if } x \in A_k, \\ 0, & \text{if } x \notin A_k, \end{cases}$$

and each c_k is a real number. We define the *Lebesgue integral* of h to be

$$\int_R h \, d\mu_0 \stackrel{\Delta}{=} \sum_{k=1}^{n} c_k \int_{I\!\!R} g_k d\mu_0 = \sum_{k=1}^{n} c_k \mu_0(A_k).$$

Let f be a nonnegative function defined on $I\!\!R$, possibly taking the value ∞ at some points. We define the *Lebesgue integral* of f to be

$$\int_{I\!\!R} f \, d\mu_0 \stackrel{\Delta}{=} \sup \left\{ \int_{I\!\!R} h \, d\mu_0; h \text{ is simple and } h(x) \leq f(x) \text{ for every } x \in I\!\!R \right\}.$$

It is possible that this integral is infinite. If it is finite, we say that f *is integrable*.

Finally, let f be a function defined on $\mathbb{R}$, possibly taking the value ∞ at some points and the value $-\infty$ at other points. We define the *positive* and *negative parts* of f to be

$$f^+(x) \overset{\Delta}{=} \max\{f(x), 0\}, \;\; f^-(x) \overset{\Delta}{=} \max\{-f(x), 0\},$$

respectively, and we define the *Lebesgue integral* of f to be

$$\int_{\mathbb{R}} f \, d\mu_0 \overset{\Delta}{=} \int_{\mathbb{R}} f^+ \, d\mu_0 - - \int_{\mathbb{R}} f^- \, d\mu_0,$$

provided the right-hand side is not of the form $\infty - \infty$. If both $\int_{\mathbb{R}} f^+ \, d\mu_0$ and $\int_{\mathbb{R}} f^- \, d\mu_0$ are finite (or equivalently, $\int_{\mathbb{R}} |f| \, d\mu_0 < \infty$, since $|f| = f^+ + f^-$), we say that f is *integrable*.

Let f be a function defined on $\mathbb{R}$, possibly taking the value ∞ at some points and the value $-\infty$ at other points. Let A be a subset of $\mathbb{R}$. We define

$$\int_A f \, d\mu_0 \overset{\Delta}{=} \int_{\mathbb{R}} \mathbb{I}_A f \, d\mu_0,$$

where

$$\mathbb{I}_A(x) \overset{\Delta}{=} \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A, \end{cases}$$

is the *indicator function of* A.

The Lebesgue integral just defined is related to the Riemann integral in one very important way: if the Riemann integral $\int_a^b f(x)dx$ is defined, then the Lebesgue integral $\int_{[a,b]} f \, d\mu_0$ agrees with the Riemann integral. The Lebesgue integral has two important advantages over the Riemann integral. The first is that the Lebesgue integral is defined for more functions, as we show in the following examples.

Example 1.5 Let Q be the set of rational numbers in $[0,1]$, and consider $f \overset{\Delta}{=} \mathbb{I}_Q$. Being a countable set, Q has Lebesgue measure zero, and so the Lebesgue integral of f over $[0,1]$ is

$$\int_{[0,1]} f \, d\mu_0 = 0.$$

To compute the Riemann integral $\int_0^1 f(x)dx$, we choose partition points $0 = x_0 < x_1 < \cdots < x_n = 1$ and divide the interval $[0,1]$ into subintervals $[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]$. In each subinterval $[x_{k-1}, x_k]$ there is a rational point q_k, where $f(q_k) = 1$, and there is also an irrational point r_k, where $f(r_k) = 0$. We approximate the Riemann integral from above by the *upper sum*

$$\sum_{k=1}^n f(q_k)(x_k - x_{k-1}) = \sum_{k=1}^n 1 \cdot (x_k - x_{k-1}) = 1,$$

and we also approximate it from below by the *lower sum*

$$\sum_{k=1}^n f(r_k)(x_k - x_{k-1}) = \sum_{k=1}^n 0 \cdot (x_k - x_{k-1}) = 0.$$

No matter how fine we take the partition of $[0,1]$, the upper sum is always 1 and the lower sum is always 0. Since these two do not converge to a common value as the partition becomes finer, the Riemann integral is not defined. $\diamond$

Example 1.6 Consider the function

$$f(x) \triangleq \begin{cases} \infty, & \text{if } x = 0, \\ 0, & \text{if } x \neq 0. \end{cases}$$

This is not a simple function because simple function cannot take the value ∞. Every simple function which lies between 0 and f is of the form

$$h(x) \triangleq \begin{cases} y, & \text{if } x = 0, \\ 0, & \text{if } x \neq 0, \end{cases}$$

for some $y \in [0, \infty)$, and thus has Lebesgue integral

$$\int_{\mathbb{R}} h \, d\mu_0 = y\mu_0\{0\} = 0.$$

It follows that

$$\int_{\mathbb{R}} f \, d\mu_0 = \sup \left\{ \int_{\mathbb{R}} h \, d\mu_0 ; h \text{ is simple and } h(x) \leq f(x) \text{ for every } x \in \mathbb{R} \right\} = 0.$$

Now consider the Riemann integral $\int_{-\infty}^{\infty} f(x) \, dx$, which for this function f is the same as the Riemann integral $\int_{-1}^{1} f(x) \, dx$. When we partition $[-1, 1]$ into subintervals, one of these will contain the point 0, and when we compute the upper approximating sum for $\int_{-1}^{1} f(x) \, dx$, this point will contribute ∞ times the length of the subinterval containing it. Thus the upper approximating sum is ∞. On the other hand, the lower approximating sum is 0, and again the Riemann integral does not exist. $\diamond$

The Lebesgue integral has all *linearity* and *comparison* properties one would expect of an integral. In particular, for any two functions f and g and any real constant c,

$$\begin{aligned} \int_{\mathbb{R}} (f+g) \, d\mu_0 &= \int_{\mathbb{R}} f \, d\mu_0 + \int_{\mathbb{R}} g \, d\mu_0, \\ \int_{\mathbb{R}} cf \, d\mu_0 &= c \int_{\mathbb{R}} f \, d\mu_0, \end{aligned}$$

and whenever $f(x) \leq g(x)$ for all $x \in \mathbb{R}$, we have

$$\int_{\mathbb{R}} f \, d\mu_0 \leq \int_{\mathbb{R}} gd \, d\mu_0.$$

Finally, if A and B are disjoint sets, then

$$\int_{A \cup B} f \, d\mu_0 = \int_A f \, d\mu_0 + \int_B f \, d\mu_0.$$

There are three *convergence theorems* satisfied by the Lebesgue integral. In each of these the situation is that there is a sequence of functions $f_n, n = 1, 2, \ldots$ converging *pointwise* to a limiting function f. *Pointwise convergence* just means that

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ for every } x \in I\!\!R.$$

There are no such theorems for the Riemann integral, because the Riemann integral of the limiting function f is too often not defined. Before we state the theorems, we given two examples of pointwise convergence which arise in probability theory.

Example 1.7 Consider a sequence of normal densities, each with variance 1 and the n-th having mean n:

$$f_n(x) \stackrel{\Delta}{=} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-n)^2}{2}}.$$

These converge pointwise to the function

$$f(x) = 0 \text{ for every } x \in I\!\!R.$$

We have $\int_{I\!\!R} f_n d\mu_0 = 1$ for every n, so $\lim_{n\to\infty} \int_{I\!\!R} f_n d\mu_0 = 1$, but $\int_{I\!\!R} f\, d\mu_0 = 0$. ◇

Example 1.8 Consider a sequence of normal densities, each with mean 0 and the n-th having variance $\frac{1}{n}$:

$$f_n(x) \stackrel{\Delta}{=} \sqrt{\frac{n}{2\pi}}\, e^{-\frac{x^2}{2n}}.$$

These converge pointwise to the function

$$f(x) \stackrel{\Delta}{=} \begin{cases} \infty, & \text{if } x = 0, \\ 0, & \text{if } x \neq 0. \end{cases}$$

We have again $\int_{I\!\!R} f_n d\mu_0 = 1$ for every n, so $\lim_{n\to\infty} \int_{I\!\!R} f_n d\mu_0 = 1$, but $\int_{I\!\!R} f\, d\mu_0 = 0$. The function f is not the Dirac delta; the Lebesgue integral of this function was already seen in Example 1.6 to be zero. ◇

Theorem 3.1 (Fatou's Lemma) *Let $f_n, n = 1, 2, \ldots$ be a sequence of nonnegative functions converging pointwise to a function f. Then*

$$\int_{I\!\!R} f\, d\mu_0 \leq \liminf_{n\to\infty} \int_{I\!\!R} f_n\, d\mu_0.$$

If $\lim_{n\to\infty} \int_{I\!\!R} f_n\, d\mu_0$ is defined, then Fatou's Lemma has the simpler conclusion

$$\int_{I\!\!R} f\, d\mu_0 \leq \lim_{n\to\infty} \int_{I\!\!R} f_n\, d\mu_0.$$

This is the case in Examples 1.7 and 1.8, where

$$\lim_{n\to\infty} \int_{I\!\!R} f_n\, d\mu_0 = 1,$$

while $\int_{I\!\!R} f\,d\mu_0 = 0$. We could modify either Example 1.7 or 1.8 by setting $g_n = f_n$ if n is even, but $g_n = 2f_n$ if n is odd. Now $\int_{I\!\!R} g_n\,d\mu_0 = 1$ if n is even, but $\int_{I\!\!R} g_n\,d\mu_0 = 2$ if n is odd. The sequence $\{\int_{I\!\!R} g_n\,d\mu_0\}_{n=1}^{\infty}$ has two cluster points, 1 and 2. By definition, the smaller one, 1, is $\liminf_{n\to\infty} \int_{I\!\!R} g_n\,d\mu_0$ and the larger one, 2, is $\limsup_{n\to\infty} \int_{I\!\!R} g_n\,d\mu_0$. Fatou's Lemma guarantees that even the smaller cluster point will be greater than or equal to the integral of the limiting function.

The key assumption in Fatou's Lemma is that all the functions take only nonnegative values. Fatou's Lemma does not assume much but it is is not very satisfying because it does not conclude that

$$\int_{I\!\!R} f\,d\mu_0 = \lim_{n\to\infty} \int_{I\!\!R} f_n\,d\mu_0.$$

There are two sets of assumptions which permit this stronger conclusion.

Theorem 3.2 (Monotone Convergence Theorem) *Let $f_n, n = 1, 2, \ldots$ be a sequence of functions converging pointwise to a function f. Assume that*

$$0 \leq f_1(x) \leq f_2(x) \leq f_3(x) \leq \cdots \text{ for every } x \in I\!\!R.$$

Then

$$\int_{I\!\!R} f\,d\mu_0 = \lim_{n\to\infty} \int_{I\!\!R} f_n\,d\mu_0,$$

where both sides are allowed to be ∞.

Theorem 3.3 (Dominated Convergence Theorem) *Let $f_n, n = 1, 2, \ldots$ be a sequence of functions, which may take either positive or negative values, converging pointwise to a function f. Assume that there is a nonnegative integrable function g (i.e., $\int_{I\!\!R} g\,d\mu_0 < \infty$) such that*

$$|f_n(x)| \leq g(x) \text{ for every } x \in I\!\!R \text{ for every } n.$$

Then

$$\int_{I\!\!R} f\,d\mu_0 = \lim_{n\to\infty} \int_{I\!\!R} f_n\,d\mu_0,$$

and both sides will be finite.

1.4 General Probability Spaces

Definition 1.13 A *probability space* $(\Omega, \mathcal{F}, I\!\!P)$ consists of three objects:

(i) Ω, a nonempty set, called the *sample space*, which contains all possible outcomes of some random experiment;

(ii) $\mathcal{F}$, a σ-algebra of subsets of Ω;

(iii) $I\!\!P$, a probability measure on $(\Omega, \mathcal{F})$, i.e., a function which assigns to each set $A \in \mathcal{F}$ a number $I\!\!P(A) \in [0, 1]$, which represents the probability that the outcome of the random experiment lies in the set A.

Remark 1.1 We recall from Homework Problem 1.4 that a probability measure $I\!\!P$ has the following properties:

(a) $I\!\!P(\emptyset) = 0.$

(b) (Countable additivity) If $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$I\!\!P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} I\!\!P(A_k).$$

(c) (Finite additivity) If n is a positive integer and $A_1, \ldots, A_n$ are disjoint sets in $\mathcal{F}$, then

$$I\!\!P(A_1 \cup \cdots \cup A_n) = I\!\!P(A_1) + \cdots + I\!\!P(A_n).$$

(d) If A and B are sets in $\mathcal{F}$ and $A \subseteq B$, then

$$I\!\!P(B) = I\!\!P(A) + I\!\!P(B \setminus A).$$

In particular,

$$I\!\!P(B) \geq I\!\!P(A).$$

(d) (Continuity from below.) If $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$ with $A_1 \subseteq A_2 \subseteq \cdots$, then

$$I\!\!P\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} I\!\!P(A_n).$$

(d) (Continuity from above.) If $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$ with $A_1 \supseteq A_2 \supseteq \cdots$, then

$$I\!\!P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} I\!\!P(A_n).$$

We have already seen some examples of finite probability spaces. We repeat these and give some examples of infinite probability spaces as well.

Example 1.9 Finite coin toss space.
Toss a coin n times, so that Ω is the set of all sequences of H and T which have n components. We will use this space quite a bit, and so give it a name: Ω_n. Let $\mathcal{F}$ be the collection of all subsets of Ω_n. Suppose the probability of H on each toss is p, a number between zero and one. Then the probability of T is $q \stackrel{\Delta}{=} 1 - p$. For each $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ in Ω_n, we define

$$I\!\!P\{\omega\} \stackrel{\Delta}{=} p^{Number\ of\ H\ in\ \omega} \cdot q^{Number\ of\ T\ in\ \omega}.$$

For each $A \in \mathcal{F}$, we define

$$I\!\!P(A) \stackrel{\Delta}{=} \sum_{\omega \in A} I\!\!P\{\omega\}. \tag{4.1}$$

We can define $I\!\!P(A)$ this way because A has only finitely many elements, and so only finitely many terms appear in the sum on the right-hand side of (4.1). ◇

Example 1.10 Infinite coin toss space.
Toss a coin repeatedly without stopping, so that Ω is the set of all nonterminating sequences of H and T. We call this space Ω_∞. This is an uncountably infinite space, and we need to exercise some care in the construction of the σ-algebra we will use here.

For each positive integer n, we define $\mathcal{F}_n$ to be the σ-algebra determined by the first n tosses. For example, $\mathcal{F}_2$ contains four basic sets,

$$\begin{aligned}
A_{HH} &\stackrel{\Delta}{=} \{\omega = (\omega_1, \omega_2, \omega_3, \ldots); \omega_1 = H, \omega_2 = H\}\\
&= \text{The set of all sequences which begin with } HH,\\
A_{HT} &\stackrel{\Delta}{=} \{\omega = (\omega_1, \omega_2, \omega_3, \ldots); \omega_1 = H, \omega_2 = T\}\\
&= \text{The set of all sequences which begin with } HT,\\
A_{TH} &\stackrel{\Delta}{=} \{\omega = (\omega_1, \omega_2, \omega_3, \ldots); \omega_1 = T, \omega_2 = H\}\\
&= \text{The set of all sequences which begin with } TH,\\
A_{TT} &\stackrel{\Delta}{=} \{\omega = (\omega_1, \omega_2, \omega_3, \ldots); \omega_1 = T, \omega_2 = T\}\\
&= \text{The set of all sequences which begin with } TT.
\end{aligned}$$

Because $\mathcal{F}_2$ is a σ-algebra, we must also put into it the sets $\emptyset$, Ω, and all unions of the four basic sets.

In the σ-algebra $\mathcal{F}$, we put every set in every σ-algebra $\mathcal{F}_n$, where n ranges over the positive integers. We also put in every other set which is required to make $\mathcal{F}$ be a σ-algebra. For example, the set containing the single sequence

$$\{HHHHH\cdots\} = \{H \text{ on every toss}\}$$

is not in any of the $\mathcal{F}_n$ σ-algebras, because it depends on all the components of the sequence and not just the first n components. However, for each positive integer n, the set

$$\{H \text{ on the first } n \text{ tosses}\}$$

is in $\mathcal{F}_n$ and hence in $\mathcal{F}$. Therefore,

$$\{H \text{ on every toss}\} = \bigcap_{n=1}^{\infty} \{H \text{ on the first } n \text{ tosses}\}$$

is also in $\mathcal{F}$.

We next construct the probability measure $I\!\!P$ on $(\Omega_\infty, \mathcal{F})$ which corresponds to probability $p \in [0,1]$ for H and probability $q = 1 - p$ for T. Let $A \in \mathcal{F}$ be given. If there is a positive integer n such that $A \in \mathcal{F}_n$, then the description of A depends on only the first n tosses, and it is clear how to define $I\!\!P(A)$. For example, suppose $A = A_{HH} \cup A_{TH}$, where these sets were defined earlier. Then A is in $\mathcal{F}_2$. We set $I\!\!P(A_{HH}) = p^2$ and $I\!\!P(A_{TH}) = qp$, and then we have

$$I\!\!P(A) = I\!\!P(A_{HH} \cup A_{TH}) = p^2 + qp = (p+q)p = p.$$

In other words, the probability of a H on the second toss is p.

Let us now consider a set $A \in \mathcal{F}$ for which there is no positive integer n such that $A \in \mathcal{F}$. Such is the case for the set $\{H$ on every toss$\}$. To determine the probability of these sets, we write them in terms of sets which are in $\mathcal{F}_n$ for positive integers n, and then use the properties of probability measures listed in Remark 1.1. For example,

$$\begin{aligned} \{H \text{ on the first toss}\} &\supseteq \{H \text{ on the first two tosses}\} \\ &\supseteq \{H \text{ on the first three tosses}\} \\ &\supseteq \cdots, \end{aligned}$$

and

$$\bigcap_{n=1}^{\infty} \{H \text{ on the first } n \text{ tosses}\} = \{H \text{ on every toss}\}.$$

According to Remark 1.1(d) (continuity from above),

$$I\!\!P\{H \text{ on every toss}\} = \lim_{n\to\infty} I\!\!P\{H \text{ on the first } n \text{ tosses}\} = \lim_{n\to\infty} p^n.$$

If $p = 1$, then $I\!\!P\{H$ on every toss$\} = 1$; otherwise, $I\!\!P\{H$ on every toss$\} = 0$.

A similar argument shows that if $0 < p < 1$ so that $0 < q < 1$, then every set in Ω_∞ which contains only one element (nonterminating sequence of H and T) has probability zero, and hence very set which contains countably many elements also has probabiliy zero. We are in a case very similar to Lebesgue measure: every point has measure zero, but sets can have positive measure. Of course, the only sets which can have positive probabilty in Ω_∞ are those which contain uncountably many elements.

In the infinite coin toss space, we define a sequence of random variables $Y_1, Y_2, \ldots$ by

$$Y_k(\omega) \triangleq \begin{cases} 1 & \text{if } \omega_k = H, \\ 0 & \text{if } \omega_k = T, \end{cases}$$

and we also define the random variable

$$X(\omega) = \sum_{k=1}^{n} \frac{Y_k(\omega)}{2^k}.$$

Since each Y_k is either zero or one, X takes values in the interval $[0,1]$. Indeed, $X(TTTT\cdots) = 0$, $X(HHHH\cdots) = 1$ and the other values of X lie in between. We define a "dyadic rational number" to be a number of the form $\frac{m}{2^k}$, where k and m are integers. For example, $\frac{3}{4}$ is a dyadic rational. Every dyadic rational in (0,1) corresponds to two sequences $\omega \in \Omega_\infty$. For example,

$$X(HHTTTTT\cdots) = X(HTHHHHH\cdots) = \frac{3}{4}.$$

The numbers in (0,1) which are not dyadic rationals correspond to a single $\omega \in \Omega_\infty$; these numbers have a unique binary expansion.

Whenever we place a probability measure $I\!\!P$ on $(\Omega, \mathcal{F})$, we have a corresponding induced measure $\mathcal{L}_X$ on $[0,1]$. For example, if we set $p = q = \frac{1}{2}$ in the construction of this example, then we have

$$\begin{aligned}
\mathcal{L}_X\left[0, \frac{1}{2}\right] &= I\!\!P\{\text{First toss is } T\} = \frac{1}{2}, \\
\mathcal{L}_X\left[\frac{1}{2}, 1\right] &= I\!\!P\{\text{First toss is } H\} = \frac{1}{2}, \\
\mathcal{L}_X\left[0, \frac{1}{4}\right] &= I\!\!P\{\text{First two tosses are } TT\} = \frac{1}{4}, \\
\mathcal{L}_X\left[\frac{1}{4}, \frac{1}{2}\right] &= I\!\!P\{\text{First two tosses are } TH\} = \frac{1}{4}, \\
\mathcal{L}_X\left[\frac{1}{2}, \frac{3}{4}\right] &= I\!\!P\{\text{First two tosses are } HT\} = \frac{1}{4}, \\
\mathcal{L}_X\left[\frac{3}{4}, 1\right] &= I\!\!P\{\text{First two tosses are } HH\} = \frac{1}{4}.
\end{aligned}$$

Continuing this process, we can verify that for any positive integers k and m satisfying

$$0 \le \frac{m-1}{2^k} < \frac{m}{2^k} \le 1,$$

we have

$$\mathcal{L}_X\left[\frac{m-1}{2^k}, \frac{m}{2^k}\right] = \frac{1}{2^k}.$$

In other words, the $\mathcal{L}_X$-measure of all intervals in $[0,1]$ whose endpoints are dyadic rationals is the same as the Lebesgue measure of these intervals. The only way this can be is for $\mathcal{L}_X$ to be Lebesgue measure.

It is interesing to consider what $\mathcal{L}_X$ would look like if we take a value of p other than $\frac{1}{2}$ when we construct the probability measure $I\!\!P$ on Ω.

We conclude this example with another look at the Cantor set of Example 3.2. Let Ω_{pairs} be the subset of Ω in which every even-numbered toss is the same as the odd-numbered toss immediately preceding it. For example, $HHTTTTHH$ is the beginning of a sequence in Ω_{pairs}, but HT is not. Consider now the set of real numbers

$$C' \triangleq \{X(\omega); \omega \in \Omega_{pairs}\}.$$

The numbers between $(\frac{1}{4}, \frac{1}{2})$ can be written as $X(\omega)$, but the sequence ω must begin with either TH or HT. Therefore, none of these numbers is in C'. Similarly, the numbers between $(\frac{1}{16}, \frac{3}{16})$ can be written as $X(\omega)$, but the sequence ω must begin with $TTTH$ or $TTHT$, so none of these numbers is in C'. Continuing this process, we see that C' will not contain any of the numbers which were removed in the construction of the Cantor set C in Example 3.2. In other words, $C' \subseteq C$. With a bit more work, one can convince onself that in fact $C' = C$, i.e., by requiring consecutive coin tosses to be paired, we are removing exactly those points in $[0,1]$ which were removed in the Cantor set construction of Example 3.2. $\diamond$

In addition to tossing a coin, another common random experiment is to pick a number, perhaps using a random number generator. Here are some probability spaces which correspond to different ways of picking a number at random.

Example 1.11
Suppose we choose a number from $I\!\!R$ in such a way that we are sure to get either 1, 4 or 16. Furthermore, we construct the experiment so that the probability of getting 1 is $\frac{4}{9}$, the probability of getting 4 is $\frac{4}{9}$ and the probability of getting 16 is $\frac{1}{9}$. We describe this random experiment by taking Ω to be $I\!\!R$, $\mathcal{F}$ to be $\mathcal{B}(I\!\!R)$, and setting up the probability measure so that

$$I\!\!P\{1\} = \frac{4}{9},\ I\!\!P\{4\} = \frac{4}{9},\ I\!\!P\{16\} = \frac{1}{9}.$$

This determines $I\!\!P(A)$ for every set $A \in \mathcal{B}(I\!\!R)$. For example, the probability of the interval $(0, 5]$ is $\frac{8}{9}$, because this interval contains the numbers 1 and 4, but not the number 16.

The probability measure described in this example is $\mathcal{L}_{S_2}$, the measure induced by the stock price S_2, when the initial stock price $S_0 = 4$ and the probability of H is $\frac{1}{3}$. This distribution was discussed immediately following Definition 2.8. $\diamond$

Example 1.12 Uniform distribution on $[0, 1]$.
Let $\Omega = [0, 1]$ and let $\mathcal{F} = \mathcal{B}([0, 1])$, the collection of all Borel subsets containined in $[0, 1]$. For each Borel set $A \subseteq [0, 1]$, we define $I\!\!P(A) = \mu_0(A)$ to be the Lebesgue measure of the set. Because $\mu_0[0, 1] = 1$, this gives us a probability measure.

This probability space corresponds to the random experiment of choosing a number from $[0, 1]$ so that every number is “equally likely” to be chosen. Since there are infinitely mean numbers in $[0, 1]$, this requires that every number have probabilty zero of being chosen. Nonetheless, we can speak of the probability that the number chosen lies in a particular set, and if the set has uncountably many points, then this probability can be positive. $\diamond$

I know of no way to design a physical experiment which corresponds to choosing a number at random from $[0, 1]$ so that each number is equally likely to be chosen, just as I know of no way to toss a coin infinitely many times. Nonetheless, both Examples 1.10 and 1.12 provide probability spaces which are often useful approximations to reality.

Example 1.13 Standard normal distribution.
Define the standard normal density

$$\varphi(x) \stackrel{\Delta}{=} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Let $\Omega = I\!\!R$, $\mathcal{F} = \mathcal{B}(I\!\!R)$ and for every Borel set $A \subseteq I\!\!R$, define

$$I\!\!P(A) \stackrel{\Delta}{=} \int_A \varphi \, d\mu_0. \tag{4.2}$$

If A in (4.2) is an interval $[a,b]$, then we can write (4.2) as the less mysterious Riemann integral:

$$I\!\!P[a,b] \triangleq \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx.$$

This corresponds to choosing a point at random on the real line, and every single point has probability zero of being chosen, but if a set A is given, then the probability the point is in that set is given by (4.2). $\diamond$

The construction of the integral in a general probability space follows the same steps as the construction of Lebesgue integral. We repeat this construction below.

Definition 1.14 Let $(\Omega, \mathcal{F}, I\!\!P)$ be a probability space, and let X be a random variable on this space, i.e., a mapping from Ω to $I\!\!R$, possibly also taking the values $\pm\infty$.

- If X is an *indicator*, i.e,

  $$X(\omega) = I\!\!I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \in A^c, \end{cases}$$

  for some set $A \in \mathcal{F}$, we define

  $$\int_\Omega X\,dI\!\!P \triangleq I\!\!P(A).$$

- If X is a *simple function*, i.e,

  $$X(\omega) = \sum_{k=1}^n c_k I\!\!I_{A_k}(\omega),$$

  where each c_k is a real number and each A_k is a set in $\mathcal{F}$, we define

  $$\int_\Omega X\,dI\!\!P \triangleq \sum_{k=1}^n c_k \int_\Omega I\!\!I_{A_k}\,dI\!\!P = \sum_{k=1}^n c_k I\!\!P(A_k).$$

- If X is *nonnegative* but otherwise general, we define

  $$\begin{aligned}&\int_\Omega X\,dI\!\!P \\ &\triangleq \sup\left\{\int_\Omega Y\,dI\!\!P; Y \text{ is simple and } Y(\omega) \le X(\omega) \text{ for every } \omega \in \Omega\right\}.\end{aligned}$$

  In fact, we can always construct a sequence of simple functions $Y_n,\ n = 1, 2, \ldots$ such that

  $$0 \le Y_1(\omega) \le Y_2(\omega) \le Y_3(\omega) \le \ldots \text{ for every } \omega \in \Omega,$$

  and $Y(\omega) = \lim_{n\to\infty} Y_n(\omega)$ for every $\omega \in \Omega$. With this sequence, we can define

  $$\int_\Omega X\,dI\!\!P \triangleq \lim_{n\to\infty} \int_\Omega Y_n\,dI\!\!P.$$

- If X is *integrable*, i.e,

  $$\int_\Omega X^+ \, d\mathbb{P} < \infty, \quad \int_\Omega X^- \, d\mathbb{P} < \infty,$$

  where

  $$X^+(\omega) \stackrel{\Delta}{=} \max\{X(\omega), 0\}, \quad X^-(\omega) \stackrel{\Delta}{=} \max\{-X(\omega), 0\},$$

  then we define

  $$\int_\Omega X \, d\mathbb{P} \stackrel{\Delta}{=} \int_\Omega X^+ \, d\mathbb{P} - - \int_\Omega X^- \, d\mathbb{P}.$$

If A is a set in $\mathcal{F}$ and X is a random variable, we define

$$\int_A X \, d\mathbb{P} \stackrel{\Delta}{=} \int_\Omega \mathbb{I}_A \cdot X \, d\mathbb{P}.$$

The *expectation* of a random variable X is defined to be

$$\mathbb{E}X \stackrel{\Delta}{=} \int_\Omega X \, d\mathbb{P}.$$

The above integral has all the linearity and comparison properties one would expect. In particular, if X and Y are random variables and c is a real constant, then

$$\begin{aligned} \int_\Omega (X+Y) \, d\mathbb{P} &= \int_\Omega X \, d\mathbb{P} + \int_\Omega Y \, d\mathbb{P}, \\ \int_\Omega cX \, d\mathbb{P} &= c \int_\Omega X \, dP, \end{aligned}$$

If $X(\omega) \le Y(\omega)$ for every $\omega \in \Omega$, then

$$\int_\Omega X \, d\mathbb{P} \le \int_\Omega Y \, d\mathbb{P}.$$

In fact, we don't need to have $X(\omega) \le Y(\omega)$ for *every* $\omega \in \Omega$ in order to reach this conclusion; it is enough if the set of ω for which $X(\omega) \le Y(\omega)$ has probability one. When a condition holds with probability one, we say it holds *almost surely*. Finally, if A and B are disjoint subsets of Ω and X is a random variable, then

$$\int_{A \cup B} X \, d\mathbb{P} = \int_A X \, d\mathbb{P} + \int_B X \, d\mathbb{P}.$$

We restate the Lebesgue integral convergence theorem in this more general context. We acknowledge in these statements that conditions don't need to hold for every ω; almost surely is enough.

Theorem 4.4 (Fatou's Lemma) *Let $X_n, n = 1, 2, \ldots$ be a sequence of almost surely nonnegative random variables converging almost surely to a random variable X. Then*

$$\int_\Omega X \, d\mathbb{P} \le \liminf_{n \to \infty} \int_\Omega X_n \, d\mathbb{P},$$

or equivalently,

$$\mathbb{E}X \le \liminf_{n \to \infty} \mathbb{E}X_n.$$

Theorem 4.5 (Monotone Convergence Theorem) *Let $X_n, n = 1, 2, \ldots$ be a sequence of random variables converging almost surely to a random variable X. Assume that*

$$0 \leq X_1 \leq X_2 \leq X_3 \leq \cdots \textit{ almost surely.}$$

Then

$$\int_\Omega X \, d\mathbb{P} = \lim_{n\to\infty} \int_\Omega X_n d\mathbb{P},$$

or equivalently,

$$\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}X_n.$$

Theorem 4.6 (Dominated Convergence Theorem) *Let $X_n, n = 1, 2, \ldots$ be a sequence of random variables, converging almost surely to a random variable X. Assume that there exists a random variable Y such that*

$$|X_n| \leq Y \textit{ almost surely for every } n.$$

Then

$$\int_\Omega X \, d\mathbb{P} = \lim_{n\to\infty} \int_\Omega X_n \, d\mathbb{P},$$

or equivalently,

$$\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}X_n.$$

In Example 1.13, we constructed a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by integrating the standard normal density. In fact, whenever φ is a nonnegative function defined on R satisfying $\int_{\mathbb{R}} \varphi \, d\mu_0 = 1$, we call φ a *density* and we can define an associated probability measure by

$$\mathbb{P}(A) \triangleq \int_A \varphi \, d\mu_0 \text{ for every } A \in \mathcal{B}(\mathbb{R}). \tag{4.3}$$

We shall often have a situation in which two measure are related by an equation like (4.3). In fact, the market measure and the risk-neutral measures in financial markets are related this way. We say that φ in (4.3) is the *Radon-Nikodym derivative* of $d\mathbb{P}$ with respect to μ_0, and we write

$$\varphi = \frac{d\mathbb{P}}{d\mu_0}. \tag{4.4}$$

The probability measure $\mathbb{P}$ weights different parts of the real line according to the density φ. Now suppose f is a function on $(R, \mathcal{B}(\mathbb{R}), \mathbb{P})$. Definition 1.14 gives us a value for the abstract integral

$$\int_{\mathbb{R}} f \, d\mathbb{P}.$$

We can also evaluate

$$\int_{\mathbb{R}} f\varphi \, d\mu_0,$$

which is an integral with respec to Lebesgue measure over the real line. We want to show that

$$\int_{\mathbb{R}} f \, d\mathbb{P} = \int_{\mathbb{R}} f\varphi \, d\mu_0, \tag{4.5}$$

an equation which is suggested by the notation introduced in (4.4) (substitute $\frac{d\mathbb{P}}{d\mu_0}$ for φ in (4.5) and "cancel" the $d\mu_0$). We include a proof of this because it allows us to illustrate the concept of the *standard machine* explained in Williams's book in Section 5.12, page 5.

The standard machine argument proceeds in four steps.

Step 1. Assume that f is an *indicator function*, i.e., $f(x) = \mathbb{I}_A(x)$ for some Borel set $A \subseteq \mathbb{R}$. In that case, (4.5) becomes

$$\mathbb{P}(A) = \int_A \varphi\, d\mu_0.$$

This is true because it is the definition of $\mathbb{P}(A)$.

Step 2. Now that we know that (4.5) holds when f is an indicator function, assume that f is a *simple function*, i.e., a linear combination of indicator functions. In other words,

$$f(x) = \sum_{k=1}^{n} c_k h_k(x),$$

where each c_k is a real number and each h_k is an indicator function. Then

$$\begin{aligned}
\int_{\mathbb{R}} f\, d\mathbb{P} &= \int_{\mathbb{R}} \left[\sum_{k=1}^{n} c_k h_k\right] d\mathbb{P} \\
&= \sum_{k=1}^{n} c_k \int_{\mathbb{R}} h_k\, d\mathbb{P} \\
&= \sum_{k=1}^{n} c_k \int_{\mathbb{R}} h_k \varphi\, d\mu_0 \\
&= \int_{\mathbb{R}} \left[\sum_{k=1}^{n} c_k h_k\right] \varphi\, d\mu_0 \\
&= \int_{\mathbb{R}} f \varphi\, d\mu_0.
\end{aligned}$$

Step 3. Now that we know that (4.5) holds when f is a simple function, we consider a general nonnegative function f. We can always construct a sequence of nonnegative simple functions $f_n, n = 1, 2, \ldots$ such that

$$0 \le f_1(x) \le f_2(x) \le f_3(x) \le \ldots \text{ for every } x \in \mathbb{R},$$

and $f(x) = \lim_{n\to\infty} f_n(x)$ for every $x \in \mathbb{R}$. We have already proved that

$$\int_{\mathbb{R}} f_n\, d\mathbb{P} = \int_{\mathbb{R}} f_n \varphi\, d\mu_0 \text{ for every } n.$$

We let $n \to \infty$ and use the Monotone Convergence Theorem on both sides of this equality to get

$$\int_{\mathbb{R}} f\, d\mathbb{P} = \int_{\mathbb{R}} f \varphi\, d\mu_0.$$

Step 4. In the last step, we consider an *integrable* function f, which can take both positive and negative values. By *integrable*, we mean that

$$\int_{I\!\!R} f^+ \, dI\!\!P < \infty, \quad \int_{I\!\!R} f^- \, dI\!\!P < \infty.$$

¿From Step 3, we have

$$\begin{aligned} \int_{I\!\!R} f^+ \, dI\!\!P &= \int_{I\!\!R} f^+ \varphi \, d\mu_0, \\ \int_{I\!\!R} f^- \, dI\!\!P &= \int_{I\!\!R} f^- \varphi \, d\mu_0. \end{aligned}$$

Subtracting these two equations, we obtain the desired result:

$$\begin{aligned} \int_{I\!\!R} f \, dI\!\!P &= \int_{I\!\!R} f^+ \, dI\!\!P - \int_{I\!\!R} f^- \, dI\!\!P \\ &= \int_{I\!\!R} f^+ \varphi \, d\mu_0 - \int_{I\!\!R} f^- \varphi \, d\mu_0 \\ &= \int_{R} f \varphi \, d\mu_0. \end{aligned}$$

1.5 Independence

In this section, we define and discuss the notion of independence in a general probability space $(\Omega, \mathcal{F}, I\!\!P)$, although most of the examples we give will be for coin toss space.

1.5.1 Independence of sets

Definition 1.15 We say that two sets $A \in \mathcal{F}$ and $B \in \mathcal{F}$ are *independent* if

$$I\!\!P(A \cap B) = I\!\!P(A) I\!\!P(B).$$

Suppose a random experiment is conducted, and ω is the outcome. The probability that $\omega \in A$ is $I\!\!P(A)$. Suppose you are not told ω, but you are told that $\omega \in B$. Conditional on this information, the probability that $\omega \in A$ is

$$I\!\!P(A|B) \triangleq \frac{I\!\!P(A \cap B)}{I\!\!P(B)}.$$

The sets A and B are independent if and only if this conditional probability is the uncondidtional probability $I\!\!P(A)$, i.e., knowing that $\omega \in B$ does not change the probability you assign to A. This discussion is symmetric with respect to A and B; if A and B are independent and you know that $\omega \in A$, the conditional probability you assign to B is still the unconditional probability $I\!\!P(B)$.

Whether two sets are independent depends on the probability measure $I\!\!P$. For example, suppose we toss a coin twice, with probability p for H and probability $q = 1 - p$ for T on each toss. To avoid trivialities, we assume that $0 < p < 1$. Then

$$I\!\!P\{HH\} = p^2, \;\; I\!\!P\{HT\} = I\!\!P\{TH\} = pq, \;\; I\!\!P\{TT\} = q^2. \tag{5.1}$$

Let $A=\{HH,HT\}$ and $B=\{HT,TH\}$. In words, A is the set "H on the first toss" and B is the set "one H and one T." Then $A\cap B=\{HT\}$. We compute

$$\begin{aligned}
I\!\!P(A)&=p^2+pq=p,\\
I\!\!P(B)&=2pq,\\
I\!\!P(A)I\!\!P(B)&=2p^2q,\\
I\!\!P(A\cap B)&=pq.
\end{aligned}$$

These sets are independent if and only if $2p^2q=pq$, which is the case if and only if $p=\frac{1}{2}$.

If $p=\frac{1}{2}$, then $I\!\!P(B)$, the probability of one head and one tail, is $\frac{1}{2}$. If you are told that the coin tosses resulted in a head on the first toss, the probability of B, which is now the probability of a T on the second toss, is still $\frac{1}{2}$.

Suppose however that $p=0.01$. By far the most likely outcome of the two coin tosses is TT, and the probability of one head and one tail is quite small; in fact, $I\!\!P(B)=0.0198$. However, if you are told that the first toss resulted in H, it becomes very likely that the two tosses result in one head and one tail. In fact, conditioned on getting a H on the first toss, the probability of one H and one T is the probability of a T on the second toss, which is 0.99.

1.5.2 Independence of σ-algebras

Definition 1.16 Let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$. We say that $\mathcal{G}$ and $\mathcal{H}$ are *independent* if every set in $\mathcal{G}$ is independent of every set in $\mathcal{H}$, i.e,

$$I\!\!P(A\cap B)=I\!\!P(A)I\!\!P(B)\text{ for every }A\in\mathcal{H},\;\;B\in\mathcal{G}.$$

Example 1.14 Toss a coin twice, and let $I\!\!P$ be given by (5.1). Let $\mathcal{G}=\mathcal{F}_1$ be the σ-algebra determined by the first toss: $\mathcal{G}$ contains the sets

$$\emptyset,\Omega,\{HH,HT\},\{TH,TT\}.$$

Let $\mathcal{H}$ be the σ-albegra determined by the second toss: $\mathcal{H}$ contains the sets

$$\emptyset,\Omega,\{HH,TH\},\{HT,TT\}.$$

These two σ-algebras are independent. For example, if we choose the set $\{HH,HT\}$ from $\mathcal{G}$ and the set $\{HH,TH\}$ from $\mathcal{H}$, then we have

$$\begin{aligned}
&I\!\!P\{HH,HT\}I\!\!P\{HH,TH\}=(p^2+pq)(p^2+pq)=p^2,\\
&I\!\!P\Big(\{HH,HT\}\cap\{HH,TH\}\Big)=I\!\!P\{HH\}=p^2.
\end{aligned}$$

No matter which set we choose in $\mathcal{G}$ and which set we choose in $\mathcal{H}$, we will find that the product of the probabilties is the probability of the intersection.

Example 1.14 illustrates the general principle that when the probability for a sequence of tosses is defined to be the product of the probabilities for the individual tosses of the sequence, then every set depending on a particular toss will be independent of every set depending on a different toss. We say that the different tosses are independent when we construct probabilities this way. It is also possible to construct probabilities such that the different tosses are not independent, as shown by the following example.

Example 1.15 Define $I\!\!P$ for the individual elements of $\Omega = \{HH, HT, TH, TT\}$ to be

$$I\!\!P\{HH\} = \frac{1}{9},\ I\!\!P\{HT\} = \frac{2}{9},\ I\!\!P\{TH\} = \frac{1}{3},\ I\!\!P\{TT\} = \frac{1}{3},$$

and for every set $A \subseteq \Omega$, define $I\!\!P(A)$ to be the sum of the probabilities of the elements in A. Then $I\!\!P(\Omega) = 1$, so $I\!\!P$ is a probability measure. Note that the sets $\{H$ on first toss$\} = \{HH, HT\}$ and $\{H$ on second toss$\} = \{HH, TH\}$ have probabilities $I\!\!P\{HH, HT\} = \frac{1}{3}$ and $I\!\!P\{HH, TH\} = \frac{4}{9}$, so the product of the probabilities is $\frac{4}{27}$. On the other hand, the intersection of $\{HH, HT\}$ and $\{HH, TH\}$ contains the single element $\{HH\}$, which has probability $\frac{1}{9}$. These sets are not independent.

1.5.3 Independence of random variables

Definition 1.17 We say that two random variables X and Y are *independent* if the σ-algebras they generate $\sigma(X)$ and $\sigma(Y)$ are independent.

In the probability space of three independent coin tosses, the price S_2 of the stock at time 2 is independent of $\frac{S_3}{S_2}$. This is because S_2 depends on only the first two coin tosses, whereas $\frac{S_3}{S_2}$ is either u or d, depending on whether the *third* coin toss is H or T.

Definition 1.17 says that for independent random variables X and Y, every set defined in terms of X is independent of every set defined in terms of Y. In the case of S_2 and $\frac{S_3}{S_2}$ just considered, for example, the sets $\{S_2 = udS_0\} = \{HTH, HTT\}$ and $\left\{\frac{S_3}{S_2} = u\right\} = \{HHH, HTH, THH, TTH\}$ are indepedent sets.

Suppose X and Y are independent random variables. We defined earlier the measure induced by X on $I\!\!R$ to be

$$\mathcal{L}_X(A) \stackrel{\Delta}{=} I\!\!P\{X \in A\},\ A \subseteq I\!\!R.$$

Similarly, the measure induced by Y is

$$\mathcal{L}_Y(B) \stackrel{\Delta}{=} I\!\!P\{Y \in B\},\ B \subseteq I\!\!R.$$

Now the pair (X, Y) takes values in the plane $I\!\!R^2$, and we can define the measure induced by the pair

$$\mathcal{L}_{X,Y}(C) = I\!\!P\{(X, Y) \in C\},\ C \subseteq I\!\!R^2.$$

The set C in this last equation is a subset of the plane $I\!\!R^2$. In particular, C could be a "rectangle", i.e, a set of the form $A \times B$, where $A \subseteq I\!\!R$ and $B \subseteq I\!\!R$. In this case,

$$\{(X, Y) \in A \times B\} = \{X \in A\} \cap \{Y \in B\},$$

and X and Y are independent if and only if

$$\begin{aligned} \mathcal{L}_{X,Y}(A\times B) &= I\!P\Big(\{X\in A\}\cap\{Y\in B\}\Big) \\ &= I\!P\{X\in A\}I\!P\{Y\in B\} \qquad (5.2) \\ &= \mathcal{L}_X(A)\mathcal{L}_Y(B). \end{aligned}$$

In other words, for independent random variables X and Y, the *joint distribution* represented by the measure $\mathcal{L}_{X,Y}$ factors into the product of the *marginal distributions* represented by the measures $\mathcal{L}_X$ and $\mathcal{L}_Y$.

A *joint density* for (X,Y) is a nonnegative function $f_{X,Y}(x,y)$ such that

$$\mathcal{L}_{X,Y}(A\times B)=\int_A\int_B f_{X,Y}(x,y)\,dx\,dy.$$

Not every pair of random variables (X,Y) has a joint density, but if a pair does, then the random variables X and Y have *marginal densities* defined by

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,\eta)\,d\eta,\quad f_Y(y)\int_{-\infty}^{\infty} f_{X,Y}(\xi,y)\,d\xi.$$

These have the properties

$$\begin{aligned} \mathcal{L}_X(A) &= \int_A f_X(x)\,dx,\ A\subseteq I\!R, \\ \mathcal{L}_Y(B) &= \int_B f_Y(y)\,dy,\ B\subseteq I\!R. \end{aligned}$$

Suppose X and Y have a joint density. Then X and Y are independent variables if and only if the joint density is the product of the marginal densities. This follows from the fact that (5.2) is equivalent to independence of X and Y. Take $A=(-\infty,x]$ and $B=(-\infty,y]$, write (5.1) in terms of densities, and differentiate with respect to both x and y.

Theorem 5.7 *Suppose X and Y are independent random variables. Let g and h be functions from $I\!R$ to $I\!R$. Then $g(X)$ and $h(Y)$ are also independent random variables.*

PROOF: Let us denote $W=g(X)$ and $Z=h(Y)$. We must consider sets in $\sigma(W)$ and $\sigma(Z)$. But a typical set in $\sigma(W)$ is of the form

$$\{\omega;W(\omega)\in A\}=\{\omega:g(X(\omega))\in A\},$$

which is defined in terms of the random variable X. Therefore, this set is in $\sigma(X)$. (In general, we have that every set in $\sigma(W)$ is also in $\sigma(X)$, which means that X contains at least as much information as W. In fact, X can contain strictly more information than W, which means that $\sigma(X)$ will contain all the sets in $\sigma(W)$ and others besides; this is the case, for example, if $W=X^2$.)

In the same way that we just argued that every set in $\sigma(W)$ is also in $\sigma(X)$, we can show that every set in $\sigma(Z)$ is also in $\sigma(Y)$. Since every set in $\sigma(X)$ is independent of every set in $\sigma(Y)$, we conclude that every set in $\sigma(W)$ is independent of every set in $\sigma(Z)$. ◇

Definition 1.18 Let $X_1, X_2, \ldots$ be a sequence of random variables. We say that these random variables are *independent* if for every sequence of sets $A_1 \in \sigma(X_1), A_2 \in \sigma(X_2), \ldots$ and for every positive integer n,

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \cdots \mathbb{P}(A_n).$$

1.5.4 Correlation and independence

Theorem 5.8 *If two random variables X and Y are independent, and if g and h are functions from $\mathbb{R}$ to $\mathbb{R}$, then*

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}g(X) \cdot \mathbb{E}h(Y),$$

provided all the expectations are defined.

PROOF: Let $g(x) = \mathbb{I}_A(x)$ and $h(y) = \mathbb{I}_B(y)$ be indicator functions. Then the equation we are trying to prove becomes

$$\mathbb{P}\Big(\{X \in A\} \cap \{Y \in B\}\Big) = \mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\},$$

which is true because X and Y are independent. Now use the standard machine to get the result for general functions g and h. $\diamond$

The *variance* of a random variable X is defined to be

$$\text{Var}(X) \stackrel{\Delta}{=} \mathbb{E}[X - \mathbb{E}X]^2.$$

The covariance of two random variables X and Y is defined to be

$$\begin{aligned} \text{Cov}(X,Y) &\stackrel{\Delta}{=} \mathbb{E}\Big[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\Big] \\ &= \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y. \end{aligned}$$

According to Theorem 5.8, for independent random variables, the covariance is zero. If X and Y both have positive variances, we define their *correlation coefficient*

$$\rho(X,Y) \stackrel{\Delta}{=} \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

For independent random variables, the correlation coefficient is zero.

Unfortunately, two random variables can have zero correlation and still not be independent. Consider the following example.

Example 1.16 Let X be a standard normal random variable, let Z be independent of X and have the distribution $\mathbb{P}\{Z = 1\} = \mathbb{P}\{Z = -1\} = 0$. Define $Y = XZ$. We show that Y is also a standard normal random variable, X and Y are uncorrelated, but X and Y are not independent.

The last claim is easy to see. If X and Y were independent, so would be X^2 and Y^2, but in fact, $X^2 = Y^2$ almost surely.

We next check that Y is standard normal. For $y \in I\!\!R$, we have

$$\begin{aligned} I\!\!P\{Y \leq y\} &= I\!\!P\{Y \leq y \text{ and } Z=1\} + I\!\!P\{Y \leq y \text{ and } Z=-1\} \\ &= I\!\!P\{X \leq y \text{ and } Z=1\} + I\!\!P\{-X \leq y \text{ and } Z=-1\} \\ &= I\!\!P\{X \leq y\}I\!\!P\{Z=1\} + I\!\!P\{-X \leq y\}I\!\!P\{Z=-1\} \\ &= \frac{1}{2}I\!\!P\{X \leq y\} + \frac{1}{2}I\!\!P\{-X \leq y\}. \end{aligned}$$

Since X is standard normal, $I\!\!P\{X \leq y\} = I\!\!P\{X \leq -y\}$, and we have $I\!\!P\{Y \leq y\} = I\!\!P\{X \leq y\}$, which shows that Y is also standard normal.

Being standard normal, both X and Y have expected value zero. Therefore,

$$\text{Cov}(X,Y) = I\!\!E[XY] = I\!\!E[X^2 Z] = I\!\!EX^2 \cdot I\!\!EZ = 1 \cdot 0 = 0.$$

Where in $I\!\!R^2$ does the measure $\mathcal{L}_{X,Y}$ put its mass, i.e., what is the distribution of (X,Y)?

We conclude this section with the observation that for independent random variables, the variance of their sum is the sum of their variances. Indeed, if X and Y are independent and $Z = X+Y$, then

$$\begin{aligned} \text{Var}(Z) &\stackrel{\Delta}{=} I\!\!E\Big[(Z - I\!\!EZ)^2\Big] \\ &= I\!\!E\Big(X + Y - I\!\!EX - I\!\!EY)^2\Big] \\ &= I\!\!E\Big[(X - I\!\!EX)^2 + 2(X - I\!\!EX)(Y - I\!\!EY) + (Y - I\!\!EY)^2\Big] \\ &= \text{Var}(X) + 2I\!\!E[X - I\!\!EX]I\!\!E[Y - I\!\!EY] + \text{Var}(Y) \\ &= \text{Var}(X) + \text{Var}(Y). \end{aligned}$$

This argument extends to any finite number of random variables. If we are given independent random variables $X_1, X_2, \ldots, X_n$, then

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n). \tag{5.3}$$

1.5.5 Independence and conditional expectation.

We now return to property (k) for conditional expectations, presented in the lecture dated October 19, 1995. The property as stated there is taken from Williams's book, page 88; we shall need only the second assertion of the property:

(k) If a random variable X is independent of a σ-algebra $\mathcal{H}$, then

$$I\!\!E[X|\mathcal{H}] = I\!\!EX.$$

The point of this statement is that if X is independent of $\mathcal{H}$, then the best estimate of X based on the information in $\mathcal{H}$ is $I\!\!EX$, the same as the best estimate of X based on no information.

To show this equality, we observe first that $I\!\!E X$ is $\mathcal{H}$-measurable, since it is not random. We must also check the partial averaging property

$$\int_A I\!\!E X \, dI\!\!P = \int_A X \, dI\!\!P \text{ for every } A \in \mathcal{H}.$$

If X is an indicator of some set B, which by assumption must be independent of $\mathcal{H}$, then the partial averaging equation we must check is

$$\int_A I\!\!P(B) \, dI\!\!P = \int_A I\!\!I_B \, dI\!\!P.$$

The left-hand side of this equation is $I\!\!P(A) I\!\!P(B)$, and the right hand side is

$$\int_\Omega I\!\!I_A I\!\!I_B \, dI\!\!P = \int_\Omega I\!\!I_{A \cap B} \, dI\!\!P = I\!\!P(A \cap B).$$

The partial averaging equation holds because A and B are independent. The partial averaging equation for general X independent of $\mathcal{H}$ follows by the standard machine.

1.5.6 Law of Large Numbers

There are two fundamental theorems about sequences of independent random variables. Here is the first one.

Theorem 5.9 (Law of Large Numbers) *Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables, each with expected value μ and variance σ^2. Define the sequence of averages*

$$Y_n \triangleq \frac{X_1 + X_2 + \cdots + X_n}{n}, \quad n = 1, 2, \ldots.$$

Then Y_n converges to μ almost surely as $n \to \infty$.

We are not going to give the proof of this theorem, but here is an argument which makes it plausible. We will use this argument later when developing stochastic calculus. The argument proceeds in two steps. We first check that $I\!\!E Y_n = \mu$ for every n. We next check that Var$(Y_n) \to 0$ as $n \to 0$. In other words, the random variables Y_n are increasingly tightly distributed around μ as $n \to \infty$.

For the first step, we simply compute

$$I\!\!E Y_n = \frac{1}{n}[I\!\!E X_1 + I\!\!E X_2 + \cdots + I\!\!E X_n] = \frac{1}{n} \underbrace{[\mu + \mu + \cdots + \mu]}_{n \text{ times}} = \mu.$$

For the second step, we first recall from (5.3) that the variance of the sum of independent random variables is the sum of their variances. Therefore,

$$\text{Var}(Y_n) = \sum_{k=1}^n \text{Var}\left(\frac{X_k}{n}\right) = \sum_{k=1}^n \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

As $n \to \infty$, we have Var$(Y_n) \to 0$.

1.5.7 Central Limit Theorem

The Law of Large Numbers is a bit boring because the limit is nonrandom. This is because the denominator in the definition of Y_n is so large that the variance of Y_n converges to zero. If we want to prevent this, we should divide by $\sqrt{n}$ rather than n. In particular, if we again have a sequence of independent, identically distributed random variables, each with expected value μ and variance σ^2, but now we set

$$Z_n \triangleq \frac{(X_1-\mu)+(X_2-\mu)+\cdots+(X_n-\mu)}{\sqrt{n}},$$

then each Z_n has expected value zero and

$$\text{Var}(Z_n) = \sum_{k=1}^{n} \text{Var}\left(\frac{X_k-\mu}{\sqrt{n}}\right) = \sum_{k=1}^{n} \frac{\sigma^2}{n} = \sigma^2.$$

As $n \to \infty$, the distributions of all the random variables Z_n have the same degree of tightness, as measured by their variance, around their expected value 0. The Central Limit Theorem asserts that as $n \to \infty$, the distribution of Z_n approaches that of a normal random variable with mean (expected value) zero and variance σ^2. In other words, for every set $A \subset I\!R$,

$$\lim_{n\to\infty} I\!P\{Z_n \in A\} = \frac{1}{\sigma\sqrt{2\pi}} \int_A e^{-\frac{x^2}{2\sigma^2}} dx.$$