MATH3589 Introduction to Financial Mathematics Autumn 2026

Chapter 1: Probability and Stochastic Processes

Lecturer: Kenneth Ng

Preview

In this chapter, we review the basic probability theory needed for the development of discrete-time financial models, with an emphasis on countable probability spaces, discrete random variables, and expectations. We then introduce discrete-time stochastic processes, with particular emphasis on conditional expectations given past information and the Markov property. These concepts form the backbone of the multi-period stock-price models developed in the following chapters.

Key topics in this chapter:

1. Countable probability spaces and discrete random variables;
2. Discrete-time stochastic processes;
3. Conditional expectations of stochastic processes;
4. Markov processes.

1 Probability Spaces

We consider a set Ω, called the *sample space*, which contains all the possible outcomes of a random experiment. A probability is a measure of likelihood that a specific event will occur.

Definition 1.1 A ***probability measure*** $\mathbb{P}$ is a function defined on subsets of the sample space Ω such that the following properties are satisfied:

(i) for any $A \subset \Omega$, $0 \leq \mathbb{P}(A) \leq 1$;

(ii) $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\Omega) = 1$;

(iii) for any *countable* collection of events $\{A\}_{n=1}^{\infty}$ with $A_n \subset \Omega$ and $A_n \cap A_m = \varnothing$ for any $n, m \in \mathbb{N}$, $n \neq m$, $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

The pair $(\Omega, \mathbb{P})$ is called a ***probability space***.

The following useful properties can be deduced from the definition of probability, whose proof are left as an exercise.

1. (Inclusion-Exclusion) For any $A, B \subset \Omega$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
2. (Complementary Event) For the *complementary event* A^c of A, where $A^c := \Omega \backslash A$, it holds that $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
3. (Monotonicity) For any sets $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$.
4. (Law of total probability) Let $\{B_n\}_{n=1}^{\infty}$ be a collection of *mutually exclusive* and *exhaustive* events, i.e., $B_i \cap B_j = \varnothing$ for $i \neq j$, and $\cup_{n=1}^{\infty} B_n = \Omega$. Then, for any $A \subset \Omega$, $\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A \cap B_n)$.

Definition 1.2 A set Ω is said to be *finite* if it contains finitely many elements. It is said to be *countable* if it is either finite or has a one-to-one correspondence with the set of natural numbers $\mathbb{N}$.

Remark 1.1. A finite set is always countable, but the converse may not be true.

When Ω is finite (resp. countable), we call the pair $(\Omega, \mathbb{P})$ a finite (resp. countable) probability space. By the countable additivity of probability measure, the probability of an *event* $A \subseteq \Omega$ in a countable probability space can be computed as:

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

Example 1.1 Consider an experiment in which a coin is tossed three times. Let H (resp. T) denotes the outcome of showing a head (resp. tail) on a single toss. Then, the probability space is given by $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$, which is finite.

If the probability of obtaining a head in a single toss is $p \in [0, 1]$. Then, for $\omega \in \Omega$, $\mathbb{P}(\omega) = p^k(1-p)^{3-k}$, where k represents the number of occurrences of H in ω.

Let A be the event of having at least two heads in three tosses. Then $A = \{HHH, HTH, HHT, THH\}$, and

$$\mathbb{P}(A) = \mathbb{P}(HHH) + \mathbb{P}(HTH) + \mathbb{P}(HHT) + \mathbb{P}(THH) = p^3 + 3p^2(1-p).$$

Example 1.2 Consider an experiment in which a coin is tossed repeatedly until the first head is obtained. Then, the sample space is given by $\Omega = \{H, TH, TTH, TTTH, \dots\}$. The space Ω is countable but not finite.

Using the probability in Example 1.1, for $\omega \in \Omega$ we have $\mathbb{P}(\omega) = p(1-p)^{k-1}$, where k is the length of ω (e.g., $k = 3$ for $\omega = TTH$).

2 Conditional Probability

Conditional probability describes the likelihood of occurrence of an event A given the occurrence of another event B. When the two events A and B are related, the occurrence of B would change the probability of the occurrence of A. For instance, the probability that a 25-year old male who has lung cancer could be as low as 2%. However, if we know that the male is a smoker, the probability could be doubled.

Definition 2.1 Suppose that $A, B \subseteq \Omega$ with $\mathbb{P}(B) > 0$. Then, the ***conditional probability*** of A given B is defined as

$$\boxed{\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.}$$

Given that B has occurred, we need to confine ourselves to the smaller population set B, and searches for the outcome in A therein. Hence, we are using $\mathbb{P}(B)$ as a numeraire in the denominator, and $\mathbb{P}(A \cap B)$ as the occurrence probability within B.

If the occurrence of B does not change the probability of A, then A and B are said to be *independent*.

Definition 2.2 The events A and B are said to be ***independent*** if

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).}$$

Remark 2.1.

1. If $\mathbb{P}(B) > 0$, then A and B are independent if $\mathbb{P}(A|B) = \mathbb{P}(A)$.
2. Any set $A \subset \Omega$ and the empty set $\varnothing$ are independent, since $\mathbb{P}(A)\mathbb{P}(\varnothing) = 0 = \mathbb{P}(A \cap \varnothing)$.

The notion of independence can be generalized to any finite collection of sets.

Definition 2.3 Consider a collection of sets $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$. The collection is said to be

1. ***pairwise independent*** if, for any $i \neq j$,

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j);$$

2. ***mutually independent*** if, for any sub-collection $\{A_{i_1}, \ldots, A_{i_k}\}$ of $\mathcal{A}$, where $k \leq n$,

$$\boxed{\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).} \tag{1}$$

Mutual independence implies pairwise independence, since any sub-collection consisting of two sets ($k = 2$) in $\mathcal{A}$ satisfies (1). However, the converse is in general NOT true.

The following formulas on conditional probabilities are useful:

1. (Law of Total Probability) For any $A \subset \Omega$, and any exhaustive and mutually exclusive events $\{B_i\}_{i=1}^{n}$ with $\mathbb{P}(B_i) > 0$ for $i = 1, 2, \ldots, n$, we have

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_n)\mathbb{P}(B_n). \tag{2}$$

2. (Bayes Formula) For any $A, B \subset \Omega$ with $\mathbb{P}(A), \mathbb{P}(B) > 0$,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{3}$$

Equation (2) can be derived by using the law of total probability and the definition of conditional probability; the Bayes formula (3) is an immediate consequence of the definition of conditional probability.

Example 2.1 A financial institution classifies its corporate borrowers into three credit risk categories: Class 1 (low risk), Class 2 (medium risk), and Class 3 (high risk). The probabilities that a borrower from each class defaults within one year are respectively 0.005, 0.01, and 0.2.

Suppose that the institution's loan portfolio consists of 90% Class 1 borrowers, 7% Class 2 borrowers, and 3% Class 3 borrowers. Given that a randomly selected borrower has defaulted within one year, find the probability that the borrower belongs to Class 2.

<u>*Solution:*</u>
Let A be the event that a borrower defaults within one year. We are asked to compute $\mathbb{P}(\text{Class } 2|A)$. By the information given, we have the following table:

Class i	1	2	3
$\mathbb{P}(A\|\text{Class } i)$	0.005	0.01	0.2
$\mathbb{P}(\text{Class } i)$	0.9	0.07	0.03
$\mathbb{P}(A\|\text{Class } i)\mathbb{P}(\text{Class } i)$	0.0045	0.0007	0.006

By the law of total probability,

$$\mathbb{P}(A) = \sum_{i=1}^{3} \mathbb{P}(A|\text{Class } i)\mathbb{P}(\text{Class } i) = 0.0045 + 0.0007 + 0.006 = 0.0112.$$

By Bayes' formula,

$$\mathbb{P}(\text{Class } 2|A) = \frac{\mathbb{P}(A|\text{Class } 2)\mathbb{P}(\text{Class } 2)}{\mathbb{P}(A)} = \frac{0.0007}{0.0112} = 0.0625.$$

3 Random Variables and Expectations

Random variables are mappings from the set of possible outcomes of a random event to a real number for numerical calculations.

Definition 3.1 A function $X : \Omega \mapsto \mathbb{R}$ is called a *random variable*.

Example 3.1 Consider the random experiment of flipping a coin twice, with H denotes getting a head, and T denotes getting a tail. Then, $\Omega = \{\omega_1 = HH, \omega_2 = HT, \omega_3 = TH, \omega_4 = TT\}$. Consider the following two examples of random variables:

1. Let X be the random variable of number of heads obtained. Then $X(\omega_1) = 2$, $X(\omega_2) = X(\omega_3) = 1$, and $X(\omega_4) = 0$.
2. Let Y be the random variable such that $Y = 1$ if there is at least one head in the two flips, and $Y = 0$ otherwise. Then $Y(\omega_1) = Y(\omega_2) = Y(\omega_3) = 1$, and $Y(\omega_4) = 0$. If we let $A \subset \Omega$ be the set containing the outcomes with at least one head. Then $A = \{\omega_1, \omega_2, \omega_3\}$, and we can write Y as

$$Y(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{otherwise.} \end{cases}.$$

Y is also called an *indicator random variable*. This type of variables is very useful in probability theory.

Depending on the support of X, we mainly classify random variables into the following classes in this course:

1. **Discrete Random Variables**
   - X is said to be a discrete random variable if its ***support***[1], $\text{Supp}(X)$, is countable,

[1]Roughly speaking, the support of a distribution is the setSupp(X) such that its pmf/pdf is non-zero.

e.g., $\text{Supp}(X) = \mathbb{N}, \{1, 2, 3\}$ etc.

- The distribution of a discrete random variable X can be expressed by its ***probability mass function (pmf)***:

$$\boxed{p_X(x) = p_x = \mathbb{P}(X = x),\ x \in \text{Supp}(X),}$$

which satisfies

(a) $0 \le p_x \le 1$ for all $x \in \text{Supp}(X)$;

(b) $\sum_{x \in \text{Supp}(X)} p_x = 1$.

- For $A \subseteq \mathbb{R}$, the probability that $X \in A$ is given by

$$\boxed{\mathbb{P}(X \in A) = \sum_{x \in A} p_x.}$$

- Examples: Bernoulli, Binomial, Negative Binomial, Poisson etc.

2. **Continuous Random Variables**

- The support $\text{Supp}(X)$ of a continuous random variable X is uncountable (continuum), e.g., $\text{Supp}(X) = \mathbb{R}, [0, 1]$ etc.

- The distribution of a continuous random variable X can be expressed by its ***probability density function (pdf)***:

$$\boxed{f_X(x) := \frac{d}{dx} F_X(x) = F'_X(x),}$$

where F_X is the cdf of X, such that

(a) $f_X(x) \ge 0$ for all $x \in \text{Supp}(X)$;

(b) $\int_{\text{Supp}(X)} f(x)dx = 1$.

- Notice that $f_X(x)$ does NOT indicate the probability of $X = x$. It only gives the likelihood of X falling in a neighbourhood of x. Looking it in an infinitesimal way, given a small number dx,

$$f_X(x)dx \approx \mathbb{P}(x < X < x + dx).$$

Since $f_X(x)$ itself is NOT a probability, we do not require $f_X(x) \le 1$.

- For $A \subseteq \mathbb{R}$, the probability that $X \in A$ is given by

$$\boxed{\mathbb{P}(X \in A) = \int_A f_X(x)dx.}$$

For this reason, $\mathbb{P}(X = a) = 0$ for any $a \in \mathbb{R}$, since $\int_{\{a\}} f(x)dx = \int_a^a f_X(x)dx = 0$.

- Examples: Uniform, Gamma, Gaussian, Pareto, Exponential, Beta, etc.

Any random variable X is characterized by its ***distribution function***, defined as

$$F_X(x) := \mathbb{P}(X \leq x),\ x \in \mathbb{R}.$$

In other words, if we known F_X, then one can compute $\mathbb{P}(X \in A)$ for any $A \subseteq \mathbb{R}$.

Example 3.2 (**Binomial distribution**) A coin is tossed n times, and the probability of obtaining a head in a single toss is $p \in [0, 1]$. Let X be the random variable representing the number of occurrences of heads in the n tosses. Then, X is a discrete random variable that follows a binomial distribution, whose pmf is given by

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

We shall write $X \sim \text{Bin}(n, p)$, where n denotes the number of trails, and p is referred to as the success probability.

3.1 Expected Values

The expected value of a random variable X gives the average value of X weighted by its distribution. Moments and central moments are expected values of functions of X.

1. **Expected Value:**

   - The expected value of X, denoted by μ or $\mathbb{E}[X]$, is defined as

   $$\boxed{\mu = \mathbb{E}[X] := \begin{cases} \displaystyle\int_{-\infty}^{\infty} x f_X(x) dx, & \text{if } X \text{ is continuous;} \\ \displaystyle\sum_x x p_X(x), & \text{if } X \text{ is discrete.} \end{cases}}$$

   - If $\mathbb{E}[X]$ exists as a finite number, we say that X is *integrable*.
   - Expected value is a *linear operator*: for any $\alpha, \beta \in \mathbb{R}$, and any integrable random variables X, Y,

   $$\boxed{\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y].}$$

   - In general, for any function $g(x)$,

   $$\boxed{\mathbb{E}[g(X)] := \begin{cases} \displaystyle\int_{-\infty}^{\infty} g(x) f_X(x) dx, & \text{if } X \text{ is continuous;} \\ \displaystyle\sum_x g(x) p_X(x), & \text{if } X \text{ is discrete.} \end{cases}}$$

2. **Moments:**

   - For $k = 1, 2, \ldots$, the k-th moment of X, denoted by μ'_k, is the expected value of X^k:
   $$\boxed{\mu'_k := \mathbb{E}[X^k].}$$
   - By *Jensen's inequality*, one can show that X^{k+1} is integrable implies X^k is integrable.

3. **Central Moments:**

   - The k-th central moment of X, denoted by μ_k, is the expected value of $(X-\mu)^k$:
   $$\boxed{\mu_k := \mathbb{E}[(X-\mu)^k].}$$
   - The second central moment of X is also known as *variance*:
   $$\boxed{\mathrm{Var}[X] := \mu_2 = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \mu'_2 - \mu^2.}$$
   For any $\alpha, \beta \in \mathbb{R}$, we have
   $$\boxed{\mathrm{Var}[\alpha X + \beta] = \alpha^2 \mathrm{Var}[X].}$$

Probabilities can be computed as the expected value of an indicator function: for any set $A \subseteq \mathbb{R}$, $\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]$, since
$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega) = \sum_{\omega \in \Omega} \mathbb{1}_A(\omega)\mathbb{P}(\omega) = \mathbb{E}[\mathbb{1}_A].$$

Example 3.3 The expected value and variance of $X \sim \mathrm{Bin}(n,p)$ are given by $\mathbb{E}[X] = np$ and $\mathrm{Var}[X] = np(1-p)$, respectively.

3.2 Multivariate Distributions

Let $\{X_i\}_{i=1}^n = \{X_1, X_2, \ldots, X_n\}$ be a collection of random variables (a.k.a. *random vector*). Depending on the support of the $\{X_i\}_{i=1}^n$, the joint distribution of the random vector can be discussed as follows:

- If the distribution of each X_i is discrete, we can characterize the joint distribution by the *joint pmf*:
$$\boxed{p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n).}$$

- If the distribution of each X_i is continuous, we can characterize the joint distribution by the *joint pdf*:

$$\boxed{f_{X_1,\dots,X_n}(x_1,\dots,x_n).}$$

If X and Y are dependent of each other, the additional knowledge of one variable will change the distribution of the other. This can be described by the *conditional distribution*, which can be defined under a similar spirit as conditional probabilities in Definition 2.1.

The ***conditional distribution of*** X ***given*** Y, denoted by $X|Y$, can be characterized as follows.

1. If X and Y are discrete random variables:
   - The ***conditional pmf*** of X given Y is defined as

$$\boxed{p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}.}$$

2. If X and Y are continuous random variables:
   - the ***conditional pmf*** of X given Y is defined as

$$\boxed{f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.}$$

Given that $Y = y$, we write $X|Y = y$ as the conditional distribution of X given $Y = y$. It is a distribution of X!

Like independence of two events, if the knowledge of Y does not change the distribution of X, we say that X and Y are *independent*.

Definition 3.2 The random variables X and Y are said to be ***independent***, denoted by $X \perp\!\!\!\perp Y$, if

$$\boxed{F_{X,Y}(x,y) = F_X(x)F_Y(y).}$$

If X, Y are both discrete or both continuous, we can also define independence of X and Y as follows:

- If X and Y are both discrete random variables, then $X \perp\!\!\!\perp Y$ iff

$$\boxed{p_{X,Y}(x,y) = p_X(x)p_Y(y),\ \forall x \in \mathrm{Supp}(X), y \in \mathrm{Supp}(Y).}$$

If $p_Y(y) > 0$, this relation can also be written as $p_{X|Y}(x|y) = p_X(x)$, i.e., the conditional pmf of $X|Y = y$ is just p_X, which implies the knowledge of Y does not change the distribution of X.

- If X and Y are both continuous random variables, then $X \perp\!\!\!\perp Y$ iff

$$\boxed{f_{X,Y}(x, y) = f_X(x)f_Y(y),\ \forall x \in \text{Supp}(X), y \in \text{Supp}(Y).}$$

If $f_Y(y) > 0$, this relation can also be written as $f_{X|Y}(x|y) = f_X(x)$.

Proposition 3.1 Let X, Y be two random variables, and f, g be two functions such that $g(X)$ and $h(Y)$ are integrable. If $X \perp\!\!\!\perp Y$,

$$\boxed{\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)].}$$

Proof. Since $X \perp\!\!\!\perp Y$, we have $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Hence,

$$\begin{aligned}\mathbb{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\ &= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy\right) \\ &= \mathbb{E}[g(X)]\mathbb{E}[h(Y)].\end{aligned}$$

□

Example 3.4 Let X be a discrete random variable with pmf $p_X(-1) = 0.4$, $p_X(0) = 0.2$, and $p_X(1) = 0.4$. Define $Y := \mathbb{1}_{\{X=0\}}$.
(a) Find the pmf of Y, and the joint pmf of (X, Y). Is $X \perp\!\!\!\perp Y$?
(b) Compute $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Solution.
(a) Y is a discrete variable with only two values, 1 or 0. $Y = 1$ if $X = 0$ with probability $p_X(0) = 0.2$; $Y = 0$ if $X \neq 0$ with probability $1 - p_X(0) = 0.8$. Hence, the pmf of Y is $p_Y(0) = 0.2$, and $p_Y(1) = 0.8$.
To find the joint pmf of (X, Y), notice that $X = 0$ implies $Y = 1$, and $X \neq 0$ implies $Y = 0$. The only possible outcomes are thus $(X, Y) = (0, 1), (-1, 0)$ and $(1, 0)$. The joint pmf is thus

$$p_{X,Y}(0, 1) = p_X(0) = 0.2,\ p_{X,Y}(-1, 0) = p_X(-1) = 0.4,\ p_{X,Y}(1, 0) = p_X(1) = 0.4.$$

Notice that
$$p_X(1)p_Y(1) = 0.4 \times 0.8 = 0.32 \neq 0 = p_{X,Y}(1, 1).$$
Hence, X and Y are not independent.

(b) Notice that $\mathbb{E}[X] = 0$. Also, $\mathbb{P}(XY = 0) = 1$, which implies $\mathbb{E}[XY] = 0$. Hence, $\mathrm{Cov}(X, Y) = 0$.

$\square$

3.3 Conditional Expectations Given an Event

Conditional expectations is the expected value of a random variable with some additional knowledge. In this section, we shall discuss conditional expectations given an event and given a random variable. In Section 5, we further generalize the notion to conditional expectations given the information to-date.

Definition 3.3 Let X be a random variable, and $A \subset \Omega$ be an event with $\mathbb{P}(A) > 0$. The ***conditional expectation*** of X given A is defined as

$$\boxed{\mathbb{E}[X|A] = \frac{\mathbb{E}[X\mathbb{1}_A]}{\mathbb{P}(A)} = \begin{cases} \dfrac{1}{\mathbb{P}(A)} \displaystyle\sum_{x \in A} x\mathbb{P}(X = x), & \text{if } X \text{ is discrete;} \\ \dfrac{1}{\mathbb{P}(A)} \displaystyle\int_A x f_X(x) dx, & \text{if } X \text{ is continuous.} \end{cases}}$$

From Definition 3.3, we see that $\mathbb{E}[X|A]$ is a *real number*. Since we know that A has occurred, when computing the expected value of X, we are only counting those possible values x which overlaps with A, i.e., $\mathbb{E}[X\mathbb{1}_A]$ in the numerator. The denominator $\mathbb{P}(A)$ is the numeraire that explains the fact that we are confining ourselves to the given set A.

Quite often, we are interested in the expected value of X given the value of another related random variable, that is, $\mathbb{E}[X|A]$, where $A = \{Y = y\}$. If X, Y are discrete, we have $\mathbb{P}(A) = \mathbb{P}(Y = y)$, and $\mathbb{E}[X\mathbb{1}_{\{Y=y\}}] = \sum_x x p_{X,Y}(x, y)$ (show that!). However, when Y is continuous, $\mathbb{P}(Y = y) = 0$. In that case, we define the conditional expectation using the conditional density function.

Definition 3.4 Let X and Y be random variables. The ***conditional expectation*** of X given $Y = y$ is defined as

$$\boxed{\mathbb{E}[X|Y = y] = \begin{cases} \displaystyle\sum_x x p_{X|Y}(x|y), & \text{if } X, Y \text{ are discrete;} \\ \displaystyle\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx, & \text{if } X, Y \text{ are continuous.} \end{cases}}$$

Example 3.5 Consider the random variables X and Y in Example 3.4. Find $\mathbb{E}[X|Y=y]$ for $y=0$ and 1.

Solution:
If $Y=1$, we must have $X=0$. Hence, $\mathbb{E}[X|Y=1]=0$. If $Y=0$, we know that X has equal probability of being -1 and 1. Hence, $\mathbb{E}[X|Y=0]=-1\times 0.5+1\times 0.5=0$.

3.4 Conditional Expectations Given a Random Variable

In Definition 3.4, we defined the conditional expectation $\mathbb{E}[X|Y=y]$, which evaluates the average value of X given that $Y=y$. As the possible outcome y of Y varies, the value $\mathbb{E}[X|Y=y]$ will also change accordingly. Hence, one can view the mapping $\mathbb{E}[X|Y=y]$ as a *function of* y. Let $h(y):=\mathbb{E}[X|Y=y]$. Then, the *random variable* $h(Y)$ is called the *conditional expectation* of X given Y:

Definition 3.5 Let X,Y be random variables. Define $h(y):=\mathbb{E}[X|Y=y]$. Then, the ***conditional expectation*** of X given Y is the random variable

$$\boxed{\mathbb{E}[X|Y]:=h(Y).}$$

Remark 3.2.

1. $\mathbb{E}[X|Y]=h(Y)$ is a **random variable**.
2. $\mathbb{E}[X|Y]=h(Y)$ is a **function of** Y, instead of X.
3. $\mathbb{E}[X|Y]$ is the *"best estimate"* of the random variable X based on Y in the $\mathcal{L}^2$-sense.
4. If $X \perp\!\!\!\perp Y$, $\mathbb{E}[X|Y=y]=\mathbb{E}[X]$. Hence $h(y)=\mathbb{E}[X]$ is a constant function, and $\mathbb{E}[X|Y]=\mathbb{E}[X]$.
5. If $X=g(Y)$, i.e., X is some transformation of Y, then $\mathbb{E}[X|Y]=g(Y)=X$. In other words, once we know Y, the best estimate of X is just X itself, since we can readily compute X by $X=g(Y)$.
6. In credibility theory, $\mathbb{E}[X|Y]$ is also called the *hypothetical mean*.

The following is a very important property concerning conditional expectations.

Theorem 3.3 (Law of iterated expectations/Tower property) For any random variables X and Y, we have

$$\boxed{\mathbb{E}\left[\mathbb{E}[X|Y]\right]=\mathbb{E}[X].}$$

Remark 3.4. In some textbooks, the law of iterated expectations is sometimes written as

$$\mathbb{E}[X] = \mathbb{E}_Y \left[\mathbb{E}_X[X|Y]\right].$$

Proof. Using the definition of conditional expectations and Fubini's theorem,

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[h(Y)] &= \int_{-\infty}^{\infty} h(y) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \mathbb{E}[X|Y=y] f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \right) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x,y) dx dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_X(x) dx = \mathbb{E}[X].
\end{aligned}$$

□

Example 3.6 Suppose that the one-period return R (in decimal form) of a risky asset depends on the prevailing market regime Θ. Conditional on $\Theta = \theta$, it is known that the return R has the pdf

$$f_{R|\Theta}(r|\theta) = \frac{1}{2\theta} e^{-\frac{r}{2\theta}}, \ r > 0.$$

The market regime Θ is a discrete random variable with

$$\mathbb{P}(\Theta = 1) = 0.6, \qquad \mathbb{P}(\Theta = 5) = 0.3, \qquad \mathbb{P}(\Theta = 10) = 0.1.$$

(a) Find the pmf of $\mathbb{E}[R \mid \Theta]$.
(b) Calculate $\mathbb{E}[R]$.

Solution.
(a) For $\theta = 1, 5$ and 10, we know that

$$\mathbb{E}[R|\Theta = \theta] = \int_0^{\infty} \frac{r}{2\theta} e^{-\frac{r}{2\theta}} dx = 2\theta.$$

Hence, the pmf of $\mathbb{E}[R|\Theta]$ is given by

$$\mathbb{P}\left(\mathbb{E}[R|\Theta] = y\right) = \begin{cases} 0.6, & \text{if } y = 2; \\ 0.3, & \text{if } y = 10; \\ 0.1, & \text{if } y = 20. \end{cases}$$

(b) By the law of iterated expectation,

$$\begin{aligned} \mathbb{E}[R] &= \mathbb{E}\left[\mathbb{E}[R|\Theta]\right] \\ &= \mathbb{E}[R|\Theta = 1]p_\Theta(1) + \mathbb{E}[R|\Theta = 5]p_\Theta(5) + \mathbb{E}[R|\Theta = 10]p_\Theta(10) \\ &= 2 \times 0.6 + 10 \times 0.3 + 20 \times 0.1 = 6.2. \end{aligned}$$

□

4 Discrete Time Stochastic Processes

A ***discrete-time stochastic process*** is a sequence of random variables $\{X_k\}_{k=0}^n$ indexed by an ordered sequence of indices, where

- X_k represents the value of a quantity at the corresponding time (index) k;
- X_0 is deterministic (initial condition), which may be excluded depending on the application.

The chronological ordering means that X_k is determined by the information available up to time index k and does not depend on future information. Such a process is said to be *adapted* (or *non-anticipating*); we will not discuss this concept formally in this course.

The underlying probability space $(\Omega, \mathbb{P})$ for a discrete-time stochastic process $\{X_k\}_{k=0}^n$ consists of sample points $\omega \in \Omega$ which represents one complete sequence of the random outcomes over the entire time horizon, known as a ***sample path***. It can be viewed as a sequence $\omega = (\omega_1, \dots, \omega_n)$, where ω_k represents the outcome at time k. Thus, $X_k(\omega)$ represents the value of the process at time k along the sample path ω.

The chronological nature of the process means that $X_k(\omega)$ is determined only by the outcomes $\omega_1, \dots, \omega_k$ observed up to time k and does not depend on the future outcomes $\omega_{k+1}, \dots, \omega_n$. Hence, we may also write $X_k(\omega) = X_k(\omega_1, \dots, \omega_k)$ in the sequel.

Example 4.1 (**Coin tosses**) Suppose a coin is tossed 8 times. Then, each sample point is a sequence $\omega = (\omega_1, \dots, \omega_8)$, $\omega_k \in \{H, T\}$. For example, $\omega = (H, T, H, T, H, T, T, T)$ represents one possible sequence of outcomes. If the process $\{X_k\}_{k=1}^8$ be a stochastic process with X_k representing the number of heads observed up to the k-th toss.

$$X_3(\omega) = X_3\left((H,T,H,T,H,T,T,T)\right) = 2,$$
$$X_5(\omega) = X_5\left((H,T,H,T,H,T,T,T)\right) = 3.$$

Note that $X_k(\omega)$ only uses the first k outcomes of the sample path, even though ω specifies the entire sequence.

Suppose that for each coin toss, the probability of showing H is $p \in (0,1)$, and we want to find $\mathbb{P}(X_3 < 2)$. The paths $\omega \in \Omega$ that satisfy $X_3(\omega) \leq 2$ are given by

$$(T,T,T,\dots), \quad (T,T,H,\dots), \quad (T,H,T,\dots,), \quad (H,T,T,\dots).$$

Hence, $\mathbb{P}(X_3 \leq 2) = (1-p)^3 + 3(1-p)^2 p$.
In general, the probability of that the outcome up to the k-th trial is $(\omega_1,\dots,\omega_k)$ is given by

$$\mathbb{P}((\omega_1,\dots,\omega_k)) = \sum_{\substack{\widetilde{\omega}\in\Omega:\\ (\widetilde{\omega}_1,\dots,\widetilde{\omega}_k)=(\omega_1,\dots,\omega_k)}} \mathbb{P}(\widetilde{\omega}).$$

Example 4.2 (**Non-adapted process**) Continued from Example 4.1, let Y_k denote the number of heads from the k-th through the 8-th toss. Thus, for $\omega = (\omega_1,\dots,\omega_8)$, $Y_k(\omega)$ depends on the outcomes $\omega_k,\dots,\omega_8$, including outcomes after the k-th toss. Hence, $\{Y_k\}_{k=1}^8$ is not adapted, since Y_k depends on future information.

Random walks in the following examples are fundamental stochastic processes.

Example 4.3 (**Additive random walk**) Let $\{X_k\}_{k=0}^n$ be a stochastic process defined by

$$X_0 := 0. \quad X_k = \sum_{i=1}^k \xi_i, \qquad k = 1,\dots,n,$$

where $\{\xi_k\}_{k=1}^n$ are ***independent and identically distributed*** (i.i.d.) random variables taking values $+1$ and -1 with probabilities p and $1-p$, respectively. Equivalently, we can express X_k as

$$X_k = \sum_{i=1}^{k-1} \xi_i + \xi_k = X_{k-1} + \xi_k.$$

The process moves up or down by a fixed amount at each step and is called an *additive random walk*.

Example 4.4 (**Multiplicative random walk**) Let $\{S_k\}_{k=0}^n$ be defined by

$$S_0 := 1, \quad S_k = \prod_{i=1}^{k} \xi_i, \qquad k = 1, \ldots, n,$$

where $\{\xi_k\}_{k=1}^n$ are i.i.d. random variables taking values $u > 1$ and $0 < d < 1$ with probabilities p and $1-p$, respectively. Equivalently,

$$S_k = \prod_{i=1}^{k} \xi_i \times \xi_k = S_{k-1}\xi_k.$$

The process changes by a fixed multiplicative factor at each step and is called a *multiplicative random walk*. This provides the basic structure of the binomial stock-price model.

5 Conditional Expectations

In Subsections 3.3–3.4, we discussed the notion of conditional expectations given an event A. This section extends the notion and defines conditional expectations *given the information up to time k*.

Let $(\Omega, \mathbb{P})$ be a probability space for a stochastic process $\{X_k\}_{k=0}^n$. For a (sequence of) outcome $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$, the first k entries, $\omega_1, \ldots, \omega_k$, encodes the information of the outcome ω for time $1, \ldots, k$, which is known at time k. The collection

$$\Omega_k := \{(\omega_1, \ldots, \omega_k) : (\omega_1, \ldots, \omega_k, \ldots, \omega_n) \in \Omega\}$$

encodes all possibilities up to time k.

Let Y be a random variable. The ***conditional expectation of Y given the information up to time k*** is the expected value of Y based on the first k outcomes $(\omega_1, \ldots, \omega_k)$. It therefore varies with the observed history and is itself a random variable. Mathematically, it is defined as follows:

Definition 5.1 The ***conditional expectation*** of Y given the information up to time t_k, denoted by $\mathbb{E}_k[Y]$, is defined as a random variable on Ω_k:

$$\boxed{\mathbb{E}_k[Y](\omega_1, \ldots, \omega_k) := \sum_{\substack{\widetilde{\omega} \in \Omega: \\ (\widetilde{\omega}_1, \ldots, \widetilde{\omega}_k) = (\omega_1, \ldots, \omega_k)}} Y(\widetilde{\omega})\, \mathbb{P}(\widetilde{\omega} | (\omega_1, \ldots, \omega_k)),} \tag{4}$$

provided that $\mathbb{P}((\omega_1, \ldots, \omega_k)) > 0$.

Remark 5.1.

1. If $k = 0$, $\mathbb{E}_k[Y] = \mathbb{E}[Y]$ be definition.
2. We can interpret $\mathbb{E}_k[Y]$ as the "best estimate" of Y given the information up to time k.
3. The sequence $\{\mathbb{E}_k[Y]\}_{k=0}^n$ forms a stochastic process.

Intutively, the RHS of (4) is the probability-weighted sum of all possible values of Y among the outcomes provided that its history up to time k is $(\omega_1, \ldots, \omega_k)$.

Example 5.1 (**Conditional expectation: Coin tosses**) Continuing from Example 4.1, and let $Y = X_8$ denote the total number of heads in the 8 tosses.
(a) Compute $\mathbb{E}_3[Y](H, T, H)$,
(b) In general, compute $\mathbb{E}_3[Y](\omega_1, \omega_2, \omega_3)$.

Solution.
(a) If there are j Heads among the remaining five tosses, then the total number of Heads is $2 + j$. Note that $Y = 2 + j$ given the first 3 tosses is (H, T, H) iff the rest of the 5 tosses yields 5 heads. Hence,

$$\mathbb{P}(Y = 2 + j | (H, T, H)) = \binom{5}{j} p^j (1 - p)^j.$$

Therefore,

$$\begin{aligned}
\mathbb{E}_3[Y](H, T, H) &= \sum_{j=0}^{5} \binom{5}{j} (2 + j) p^j (1 - p)^{5-j} \\
&= 2 \sum_{j=0}^{5} \binom{5}{j} p^j (1 - p)^{5-j} + \sum_{j=0}^{5} \binom{5}{j} j\, p^j (1 - p)^{5-j} \\
&= 2 + 5p.
\end{aligned}$$

Alternatively, we can interpret

$$\begin{aligned}
\mathbb{E}_3[Y](H, T, H) &= 2 + \text{expected number of heads in the remaining 5 tosses} \\
&= 2 + 5p.
\end{aligned}$$

(b) From the calculations in (a), we see that $\mathbb{E}_3[Y](\omega_1, \omega_2, \omega_3)$ is essentially the number of heads in the first 3 tosses, i.e., $X_3 = j$, plus the expected number of heads in the remaining 5 tosses. Therefore, if there are j-heads in $(\omega_1, \omega_2, \omega_3)$, i.e., $X_3 = j$,

$$\mathbb{E}_3[Y](\omega_1, \omega_2, \omega_3) = j + 5p = X_3 + 5p.$$

□

Example 5.2 (**Conditional expectation: Gambling wealth**) Suppose a gambler starts with wealth $X_0 = 100$. At each round, the gambler wins with probability p and loses with probability $1-p$, independently across rounds. If the gambler wins, the wealth is multiplied by 1.2, while if the gambler loses, the wealth is multiplied by 0.9. Thus,

$$X_{k+1} = \begin{cases} 1.2X_k, & \text{with probability } p, \\ 0.9X_k, & \text{with probability } 1-p. \end{cases}$$

Consider a sample path

$$\omega = (W, W, L, \omega_4, \omega_5, \ldots),$$

where W and L denote a win and a loss, respectively. Then

$$X_3(\omega) = 100(1.2)(1.2)(0.9) = 129.6.$$

Let $Y = X_5$ be the gambler's wealth after five rounds.

(a) Compute the conditional expectation $\mathbb{E}_3[Y](\omega)$.

(b) Derive a general expression for $\mathbb{E}_3[Y](\omega)$ in terms of $X_3(\omega)$.

Solution.

(a) Since the history up to time t_3 is already known, only the outcomes of the next two rounds are uncertain. The four possible outcomes are

$$(W, W), \quad (W, L), \quad (L, W), \quad (L, L).$$

Their corresponding wealth values after 5 rounds are

$$1.2^2 \times 129.6, \qquad (1.2)(0.9) \times 129.6, \qquad (0.9)(1.2) \times 129.6, \qquad 0.9^2 \times 129.6,$$

with probabilities p^2, $p(1-p)$, $p(1-p)$, and $(1-p)^2$, respectively. Therefore,

$$\mathbb{E}_3[Y](\omega) = 129.6 \left[p^2(1.2)^2 + 2p(1-p)(1.2)(0.9) + (1-p)^2(0.9)^2\right].$$

(b) More generally, for any sample path ω, once the history up to and including the 3rd round is known, only the next two outcomes matter. Hence,

$$\begin{aligned} \mathbb{E}_3[Y](\omega) &= X_3(\omega) \left[p^2(1.2)^2 + 2p(1-p)(1.2)(0.9) + (1-p)^2(0.9)^2\right] \\ &= X_3(\omega) \left[1.2p + 0.9(1-p)\right]^2 \\ &= X_3(\omega)(0.9 + 0.3p)^2. \end{aligned}$$

□

Example 5.3 (**Conditional expectation: Dependent coin tosses**) Suppose a coin is tossed 5 times. The first toss is fair:

$$\mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2}.$$

For each subsequent toss, the probability of repeating the previous outcome is 2/3, while the probability of obtaining a different outcome is 1/3.

Let X_k denote the number of heads observed up to the k-th toss, with $X_0 = 0$, and $Y = X_5$. Suppose that the history up to the third toss is $\omega = (H, T, H, \omega_4, \omega_5)$. Thus, $X_3(\omega) = 2$ and the most recent outcome is H.
Compute $\mathbb{E}_3[Y](\omega)$.

Solution. Given the history (H, T, H), only the outcomes of the fourth and fifth tosses remain uncertain. Since the fourth toss follows an H, we have

$$\mathbb{P}((\omega_4, \omega_5) = (H, H)|H, T, H) = \frac{4}{9}, \quad \mathbb{P}((\omega_4, \omega_5) = (H, T)|H, T, H) = \frac{2}{9},$$
$$\mathbb{P}((\omega_4, \omega_5) = (T, H)|H, T, H) = \frac{1}{9}, \quad \mathbb{P}((\omega_4, \omega_5) = (T, T)|H, T, H) = \frac{2}{9}.$$

Therefore, by the definition of conditional expectation,

$$\mathbb{E}_3[Y](\omega) = 4\left(\frac{4}{9}\right) + 3\left(\frac{2}{9}\right) + 3\left(\frac{1}{9}\right) + 2\left(\frac{2}{9}\right) = \frac{29}{9}.$$

□

5.1 Properties of Conditional Expectations

Theorem 5.2 Let X and Y be random variables on the probability space $(\Omega, \mathbb{P})$, where Ω contains outcomes for a sequence of n events as above. For $k = 0, \ldots, n$, the conditional expectation satisfies the following properties:

(i) **Linearity.** For any $a, b \in \mathbb{R}$,

$$\mathbb{E}_k[aX + bY] = a\mathbb{E}_k[X] + b\mathbb{E}_k[Y].$$

(ii) **Taking out what is known.** If X only depends on the first k outcomes, i.e., $X(\omega_1, \ldots, \omega_n) = X(\omega_1, \ldots, \omega_k)$, then

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

(iii) **Tower property.** for any $0 \le j \le k \le n$,

$$\mathbb{E}_k\left[\mathbb{E}_j[X]\right] = \mathbb{E}_j\left[\mathbb{E}_k[X]\right] = \mathbb{E}_j[X].$$

In particular, $\mathbb{E}[\mathbb{E}_k[X]] = \mathbb{E}_0[\mathbb{E}_k[X]] = \mathbb{E}[X]$.

(iv) **Independence.** If X is independent of the first k outcomes, i.e., $X(\omega_1, \ldots, \omega_n) = X(\omega_{k+1}, \ldots, \omega_n)$, then

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

Theorem 5.2 can be proven using the formula (4), whose proof is omitted. Below, we give the interpretation of each of the properties.

(i) **Linearity:** like (unconditional) expectation, conditional expectation is linear.

(ii) **Taking out what is known:** if X can be determined by the first k outcomes, then the best estimate of X given the information up to time index k (i.e., $\mathbb{E}_k[X]$) is just X itself, since it is completely known.

(iii) **Tower property:** If $j < k$, then the information available at time index j is less detailed than that available at time index k. Thus, if we first form the best estimate of X using the information at time k and then use the earlier information at time j to estimate it, we obtain the same result as directly forming the best estimate of X using the information at time j.

(iv) **Independence:** If X has nothing to do with the first k outcomes, knowing the information up to time index k will not provide any useful information in estimating X. The best estimate given the information up to time index k is just its mean, $\mathbb{E}[X]$.

Example 5.4 Referring to Example 5.2, we can write the gambler's wealth at time k as

$$X_k = \prod_{j=1}^{k} \xi_j,$$

where $\{\xi_j\}_{j=1}^{\infty}$ is a sequence of i.i.d. random variables with ξ_j representing the outcome of the j-th bet, with

$$\xi_j = \begin{cases} 1.2, & \text{with probability } p; \\ 0.9, & \text{with probability } 1-p. \end{cases}$$

Let $Y = X_5$.

(a) Show that $\mathbb{E}_k[X_{k+1}] = X_k(0.9 + 0.3p)$.

(b) Using (a) and the properties of conditional expectations, verify that for any $0 \le n < m$, $\mathbb{E}_n[X_m] = X_n(0.9 + 0.3p)^{m-n}$.

Solution.

(a) Note that the outcomes $\xi_1, \ldots, \xi_k$ are known at time index k, while ξ_{k+1} is independent of the first k outcomes. Hence,

$$\begin{aligned}
\mathbb{E}_k[X_{k+1}] &= \mathbb{E}_k\left[\prod_{j=1}^{k+1} \xi_j\right] \\
&= \mathbb{E}_k\left[\prod_{j=1}^{k} \xi_j \times \xi_{k+1}\right] \\
&= \left(\prod_{j=1}^{k} \xi_j\right) \mathbb{E}_k[\xi_{k+1}] && \text{(taking out what is known)} \\
&= X_k \mathbb{E}_k[\xi_{k+1}] \\
&= X_k \mathbb{E}[\xi_{k+1}] && \text{(independence)} \\
&= X_k[1.2p + 0.9(1-p)] \\
&= X_k(0.9 + 0.3p).
\end{aligned}$$

(b) Using (a), we have $\mathbb{E}_{m-1}[X_m] = X_{m-1}(0.9 + 0.3p)$. Now, considering taking conditional expectations $\mathbb{E}_{m-2}[\cdot]$ on both sides, the LHS becomes

$$\mathbb{E}_{m-2}\left[\mathbb{E}_{m-1}[X_m]\right] = \mathbb{E}_{m-2}[X_m] \quad \text{(tower property)}.$$

On the other hand, the RHS reads

$$\mathbb{E}_{m-2}\left[X_{m-1}(0.9 + 0.3p)\right] = (0.9+0.3p)\mathbb{E}_{m-2}[X_{m-1}] = X_{m-2}(0.9+0.3p)^2 \quad \text{(by (a))}.$$

Hence, we arrive at $\mathbb{E}_{m-2}[X_m] = X_{m-2}(0.9+0.3p)^2$. By taking conditional expectations $\mathbb{E}_{m-3}[\cdot], \ldots, \mathbb{E}_n[\cdot]$ successively, we deduce that $\mathbb{E}_n[X_m] = X_n(0.9 + 0.3p)^{m-n}$. □

Example 5.5 (**Verifying the tower property**) Consider the dependent coin-toss model from Example 5.3. Suppose that the first three tosses are $(\omega_1, \omega_2, \omega_3) = (H, T, H)$, so that $X_3 = 2$. Let $Y = X_5$.

(a) Compute $\mathbb{E}_4[Y](\omega)$, $\omega = (H, T, H, \omega_4)$, for each possible outcome of the fourth toss ω_4.

(b) Using the probabilities of the fourth toss given the history (H, T, H), compute $\mathbb{E}_3[\mathbb{E}_4[Y]](\omega)$, and thus verify the tower property.

Solution.

(a) Since the third toss is H, the fourth toss is H with probability $2/3$ and T with probability $1/3$.
If $\omega_4 = H$, then $X_4 = 3$. Since the fourth toss is H, the fifth toss is H with

probability 2/3 and T with probability 1/3. Therefore,

$$\mathbb{E}_4[Y](H,T,H,H) = 4\left(\frac{2}{3}\right) + 3\left(\frac{1}{3}\right) = \frac{11}{3}.$$

If $\omega_4 = T$, then $X_4 = 2$. Since the fourth toss is T, the fifth toss is H with probability 1/3 and T with probability 2/3. Therefore,

$$\mathbb{E}_4[Y](H,T,H,T) = 3\left(\frac{1}{3}\right) + 2\left(\frac{2}{3}\right) = \frac{7}{3}.$$

(b) Given the history (H,T,H), the fourth toss is H with probability 2/3 and T with probability 1/3. Hence,

$$\mathbb{E}_3[\mathbb{E}_4[Y]]\,(H,T,H) = \frac{2}{3}\left(\frac{11}{3}\right) + \frac{1}{3}\left(\frac{7}{3}\right) = \frac{29}{9}.$$

In Example 5.3, we have computed that $\mathbb{E}_3[Y] = \frac{29}{9}$. Hence, the calculations show that $\mathbb{E}_3[\mathbb{E}_4[Y]] = \mathbb{E}_3[Y]$, verifying the tower property.

□

6 Markov Processes

A ***Markov process*** is a stochastic process $\{X_k\}_{k=0}^n$, where the distribution of X_{k+1} depends only on the most recent knowledge X_k if we are given the entire history up to time index k. Mathematically, a Markov process is defined as follows.

Definition 6.1 (Markov processes) We say that the stochastic process $\{X_k\}_{k=0}^n$ is a ***Markov process*** if, for any $k = 0, \ldots, n-1$ and any function $f(x)$, there exists another function $g(x)$ such that

$$\mathbb{E}_k[f(X_{k+1})] = g(X_k).$$

The conditional expectation $\mathbb{E}_k[f(X_{k+1})]$ represents the best estimate of $f(X_{k+1})$ given the information available up to time k. The fact that it can be written as $g(X_k)$ means that the current value X_k contains all the information from the past that is relevant for estimating the next state. In other words, once X_k is known, the earlier history $X_0, \ldots, X_{k-1}$ provides no additional information for predicting X_{k+1}.

The Markov property is particularly relevant when pricing European-type derivatives. In the multi-period binomial model, the derivative value at time index k depends only on the current stock price at that time, rather than the entire historic prices.

Example 6.1 Show that the following processes are Markov:

(a) The process $\{X_k\}_{k=0}^{8}$ that denotes the number of heads observed in Example 4.1.

(b) The gambler's wealth process $\{X_k\}_{k=0}^{\infty}$ in Example 5.2.

Solution.

(a) For any function f and $k = 0, \ldots, 7$,

$$\begin{aligned}\mathbb{E}_k[f(X_{k+1})](\omega_1, \ldots, \omega_k) &= f\left(X_{k+1}(\omega_1, \ldots, \omega_k, H)\right) \mathbb{P}((\omega_1, \ldots, \omega_k, H)|(\omega_1, \ldots, \omega_k)) \\ &\quad + f\left(X_{k+1}(\omega_1, \ldots, \omega_k, T)\right) \mathbb{P}((\omega_1, \ldots, \omega_k, T)|(\omega_1, \ldots, \omega_k)) \\ &= pf(X_k(\omega_1, \ldots, \omega_k) + 1) \\ &\quad + (1-p)f(0.9X_k(\omega_1, \ldots, \omega_k)),\end{aligned}$$

or simply

$$\mathbb{E}_k[f(X_{k+1})] = pf(X_k + 1) + (1-p)f(X_k).$$

Define $g(x) := pf(x+1) + (1-p)f(x)$. Then, we have $\mathbb{E}_k[f(X_{k+1})] = g(X_k)$, and hence the process is Markov.

(b) For any function f and $k > 0$,

$$\begin{aligned}\mathbb{E}_k[f(X_{k+1})](\omega_1, \ldots, \omega_k) &= f\left(X_{k+1}(\omega_1, \ldots, \omega_k, W)\right) \mathbb{P}((\omega_1, \ldots, \omega_k, W)|(\omega_1, \ldots, \omega_k)) \\ &\quad + f\left(X_{k+1}(\omega_1, \ldots, \omega_k, L)\right) \mathbb{P}((\omega_1, \ldots, \omega_k, L)|(\omega_1, \ldots, \omega_k)) \\ &= pf(1.2X_k(\omega_1, \ldots, \omega_k)) \\ &\quad + (1-p)f(0.9X_{k+1}(\omega_1, \ldots, \omega_k)),\end{aligned}$$

or simply

$$\mathbb{E}_k[f(X_{k+1})] = pf(1.2X_k) + (1-p)f(0.9X_k).$$

Define $g(x) := pf(1.2x) + (1-p)f(0.9x)$. Then, we have $\mathbb{E}_k[f(X_{k+1})] = g(X_k)$, and hence the process is Markov.

□

We also present some processes that are non-Markovian.

Example 6.2 Show that the process $\{X_k\}_{k=0}^{5}$ that denotes the number of heads observed in Example 5.3 is non-Markovian.

Solution. Let f be a function. For any $k = 2, \ldots, 4$, the distribution of X_{k+1} depends not only on X_k, but also on whether the k-th toss is a head or not. This information is not encoded in the value of X_k alone; rather, it is given in $X_k - X_{k-1}$, which represents the outcome in the k-th toss.

We thus have two cases depending on the values of $X_k - X_{k-1}$. Suppose that

$(\omega_1, \ldots, \omega_k)$ is that $X_k - X_{k-1} = 1$, i.e., the k-th toss is a head. Then,

$$\begin{aligned}\mathbb{E}_k[f(X_{k+1})](\omega_1, \ldots, \omega_k) &= f(X_{k+1}(\omega_1, \ldots, \omega_k, H))\mathbb{P}(H|X_k - X_{k-1} = 1) \\ &\quad + f(X_{k+1}(\omega_1, \ldots, \omega_k, T))\mathbb{P}(T|X_k - X_{k-1} = 1) \\ &= \frac{2f(X_k(\omega_1, \ldots, \omega_k) + 1)}{3} + \frac{f(X_k(\omega_1, \ldots, \omega_k))}{3}\end{aligned}$$

On the other hand, if $X_k - X_{k-1} = 0$, i.e., the k-th toss is a tail. Then,

$$\begin{aligned}\mathbb{E}_k[f(X_{k+1})](\omega_1, \ldots, \omega_k) &= f(X_{k+1}(\omega_1, \ldots, \omega_k, H))\mathbb{P}(H|X_k - X_{k-1} = 0) \\ &\quad + f(X_{k+1}(\omega_1, \ldots, \omega_k, T))\mathbb{P}(T|X_k - X_{k-1} = 0) \\ &= \frac{f(X_k(\omega_1, \ldots, \omega_k) + 1)}{3} + \frac{2f(X_k(\omega_1, \ldots, \omega_k))}{3}\end{aligned}$$

Combining the two cases, we have

$$\mathbb{E}_k[f(X_{k+1})] = \begin{cases} \frac{2}{3}f(X_k + 1) + \frac{1}{3}f(X_k), & \text{if } X_k - X_{k-1} = 1; \\ \frac{1}{3}f(X_k + 1) + \frac{2}{3}f(X_k), & \text{if } X_k - X_{k-1} = 0. \end{cases}$$

However, the expressions depend on the actual values of $X_k - X_{k-1}$, rather than solely on X_k. Hence, we cannot find a g such that $\mathbb{E}_k[f(X_{k+1})] = g(X_k)$. Consequently, the process is non-Markovian. □

Example 6.3 (**Running maximum**) Consider the gambler's wealth process $\{X_k\}_{k=0}^\infty$ in Example 5.2, where

$$X_{k+1} = \begin{cases} 1.2X_k, & \text{with probability } \frac{1}{4}, \\ 0.8X_k, & \text{with probability } \frac{3}{4}. \end{cases}$$

Let $\{M_k\}_{k=0}^\infty$ be the *running maximum* process of the wealth, i.e., $M_k := \max_{0 \le j \le k} X_k$. We claim that the process $\{M_k\}_{k=0}^\infty$ is non-Markovian.

For instance, consider

$$\begin{aligned}\mathbb{E}_2[M_3](L, W) &= \frac{M_3(W, L, W)}{4} + \frac{3M_3(W, L, L)}{4}, \\ \mathbb{E}_2[M_3](L, L) &= \frac{M_3(L, L, W)}{4} + \frac{3M_3(L, L, L)}{4}.\end{aligned}$$

Note that

$$M_3(L, W, W) = \max\{100, 100(0.8), 100(0.8)(1.2), 100(0.8)(1.2)^2\} = 115.2,$$

$$
\begin{aligned}
M_3(L,W,L) &= \max\{100, 100(0.8), 100(0.8)(1.2), 100(0.8)^2(1.2)\} = 100,\\
M_3(L,L,W) &= \max\{100, 100(0.8), 100(0.8)^2, 100(0.8)^2(1.2)\} = 100,\\
M_3(L,L,L) &= \max\{100, 100(0.8), 100(0.8)^2, 100(0.8)^3\} = 100.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}_2[M_3](L,W) &= \frac{115.2}{4} + \frac{3(100)}{4} = 103.8,\\
\mathbb{E}_2[M_3](L,L) &= \frac{100}{4} + \frac{3(100)}{4} = 100.
\end{aligned}
$$

Note that $M_2(L,W) = M_2(L,L) = 100$. However, there does not exist a function such that $103.8 = \mathbb{E}_2[M_3](L,W) = g(M_2(L,W)) = g(100)$, but at the same time, $100 = \mathbb{E}_2[M_3](L,L) = g(M_2(L,L)) = g(100)$. Hence, $\{M_k\}_{k=0}^{\infty}$ is not Markov.

The processes in the above two examples are non-Markov since the distribution of the process at the time $k+1$ depends not only on the most recent value at time k, but also the specific outcome at k. For instance, in Example 6.2, X_{k+1} depends on both X_k and the outcome of the k-th toss, $X_k - X_{k-1}$; in Example 6.3, the distribution of M_{k+1} will depend on both M_k and the outcome fo the k-th game. To obtain a Markovian structure, one needs to augment the process by introducing an additional variable. We shall discuss in details when pricing path-dependent options/derivatives.