

Chapter 2

Conditional Expectation

Please see Hull's book (Section 9.6.)

2.1 A Binomial Model for Stock Price Dynamics

Stock prices are assumed to follow this simple binomial model: The initial stock price during the period under study is denoted S_0 . At each time step, the stock price either goes up by a factor of u or down by a factor of d . It will be useful to visualize tossing a coin at each time step, and say that

- the stock price moves up by a factor of u if the coin comes out heads (H), and
- down by a factor of d if it comes out tails (T).

Note that we are not specifying the probability of heads here.

Consider a sequence of 3 tosses of the coin (See Fig. 2.1) The collection of all possible outcomes (i.e. sequences of tosses of length 3) is

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

A typical sequence of Ω will be denoted ω , and ω_k will denote the k th element in the sequence ω . We write $S_k(\omega)$ to denote the stock price at "time" k (i.e. after k tosses) under the outcome ω . Note that $S_k(\omega)$ depends only on $\omega_1, \omega_2, \dots, \omega_k$. Thus in the 3-coin-toss example we write for instance,

$$S_1(\omega) \triangleq S_1(\omega_1, \omega_2, \omega_3) \triangleq S_1(\omega_1),$$

$$S_2(\omega) \triangleq S_2(\omega_1, \omega_2, \omega_3) \triangleq S_2(\omega_1, \omega_2).$$

Each S_k is a *random variable* defined on the set Ω . More precisely, let $\mathcal{F} = \mathcal{P}(\Omega)$. Then \mathcal{F} is a σ -algebra and (Ω, \mathcal{F}) is a measurable space. Each S_k is an \mathcal{F} -measurable function $\Omega \rightarrow \mathbb{R}$, that is, S_k^{-1} is a function $\mathcal{B} \rightarrow \mathcal{F}$ where \mathcal{B} is the Borel σ -algebra on \mathbb{R} . We will see later that S_k is in fact

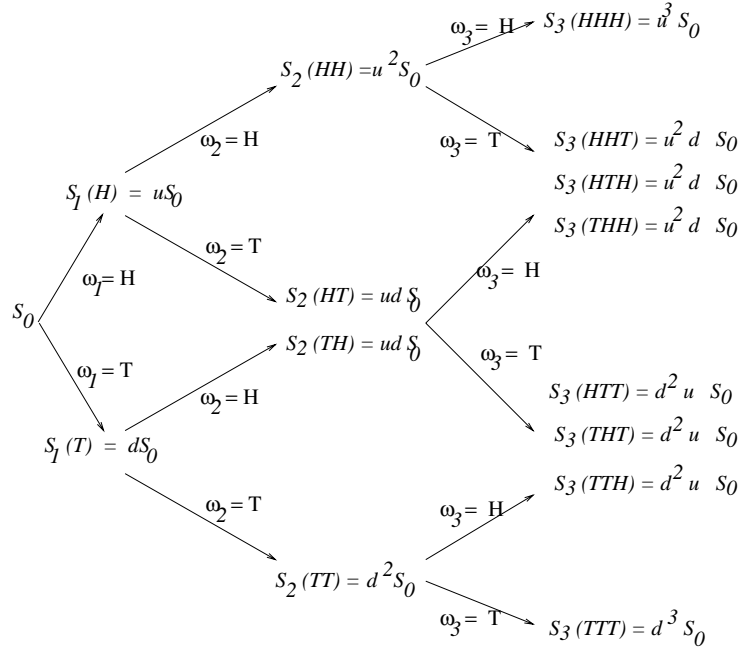


Figure 2.1: A three coin period binomial model.

measurable under a sub- σ -algebra of \mathcal{F} . Recall that the Borel σ -algebra \mathcal{B} is the σ -algebra generated by the open intervals of \mathbb{R} . In this course we will always deal with subsets of \mathbb{R} that belong to \mathcal{B} .

For any random variable X defined on a sample space Ω and any $y \in \mathbb{R}$, we will use the notation:

$$\{X \leq y\} \triangleq \{\omega \in \Omega; X(\omega) \leq y\}.$$

The sets $\{X < y\}$, $\{X \geq y\}$, $\{X = y\}$, etc, are defined similarly. Similarly for any subset B of \mathbb{R} , we define

$$\{X \in B\} \triangleq \{\omega \in \Omega; X(\omega) \in B\}.$$

Assumption 2.1 $u > d > 0$.

2.2 Information

Definition 2.1 (Sets determined by the first k tosses.) We say that a set $A \subset \Omega$ is *determined by the first k coin tosses* if, knowing only the outcome of the first k tosses, we can decide whether the outcome of *all* tosses is in A . In general we denote the collection of sets determined by the first k tosses by \mathcal{F}_k . It is easy to check that \mathcal{F}_k is a σ -algebra.

Note that the random variable S_k is \mathcal{F}_k -measurable, for each $k = 1, 2, \dots, n$.

Example 2.1 In the 3 coin-toss example, the collection \mathcal{F}_1 of sets determined by the first toss consists of:

1. $A_H \triangleq \{HHH, HHT, HTH, HTT\}$,
2. $A_T \triangleq \{THH, THT, TTH, TTT\}$,
3. ϕ ,
4. Ω .

The collection \mathcal{F}_2 of sets determined by the first two tosses consists of:

1. $A_{HH} \triangleq \{HHH, HHT\}$,
2. $A_{HT} \triangleq \{HTH, HTT\}$,
3. $A_{TH} \triangleq \{THH, THT\}$,
4. $A_{TT} \triangleq \{TTH, TTT\}$,
5. The complements of the above sets,
6. Any union of the above sets (including the complements),
7. ϕ and Ω .

■

Definition 2.2 (Information carried by a random variable.) Let X be a random variable $\Omega \rightarrow \mathbb{R}$. We say that a set $A \subset \Omega$ is *determined by the random variable X* if, knowing only the value $X(\omega)$ of the random variable, we can decide whether or not $\omega \in A$. Another way of saying this is that for every $y \in \mathbb{R}$, either $X^{-1}(y) \subset A$ or $X^{-1}(y) \cap A = \phi$. The collection of subsets of Ω determined by X is a σ -algebra, which we call the σ -algebra generated by X , and denote by $\sigma(X)$.

If the random variable X takes finitely many different values, then $\sigma(X)$ is generated by the collection of sets

$$\{X^{-1}(X(\omega)) \mid \omega \in \Omega\};$$

these sets are called the *atoms* of the σ -algebra $\sigma(X)$.

In general, if X is a random variable $\Omega \rightarrow \mathbb{R}$, then $\sigma(X)$ is given by

$$\sigma(X) = \{X^{-1}(B); B \in \mathcal{B}\}.$$

Example 2.2 (Sets determined by S_2) The σ -algebra generated by S_2 consists of the following sets:

1. $A_{HH} = \{HHH, HHT\} = \{\omega \in \Omega; S_2(\omega) = u^2 S_0\}$,
2. $A_{TT} = \{TTH, TTT\} = \{S_2 = d^2 S_0\}$,
3. $A_{HT} \cup A_{TH} = \{S_2 = ud S_0\}$,
4. Complements of the above sets,
5. Any union of the above sets,
6. $\phi = \{S_2(\omega) \in \phi\}$,
7. $\Omega = \{S_2(\omega) \in \mathbb{R}\}$.

■

2.3 Conditional Expectation

In order to talk about conditional expectation, we need to introduce a probability measure on our coin-toss sample space Ω . Let us define

- $p \in (0, 1)$ is the probability of H ,
- $q \triangleq (1 - p)$ is the probability of T ,
- the coin tosses are *independent*, so that, e.g., $\mathbb{P}(HHT) = p^2q$, etc.
- $\mathbb{P}(A) \triangleq \sum_{\omega \in A} \mathbb{P}(\omega)$, $\forall A \subset \Omega$.

Definition 2.3 (Expectation.)

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

If $A \subset \Omega$ then

$$I_A(\omega) \triangleq \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

and

$$\mathbb{E}(I_A X) = \int_A X d\mathbb{P} = \sum_{\omega \in A} X(\omega) \mathbb{P}(\omega).$$

We can think of $\mathbb{E}(I_A X)$ as a *partial average* of X over the set A .

2.3.1 An example

Let us estimate S_1 , given S_2 . Denote the estimate by $\mathbb{E}(S_1|S_2)$. From elementary probability, $\mathbb{E}(S_1|S_2)$ is a random variable Y whose value at ω is defined by

$$Y(\omega) = \mathbb{E}(S_1|S_2 = y),$$

where $y = S_2(\omega)$. Properties of $\mathbb{E}(S_1|S_2)$:

- $\mathbb{E}(S_1|S_2)$ should depend on ω , i.e., it is a *random variable*.
- If the value of S_2 is known, then the value of $\mathbb{E}(S_1|S_2)$ should also be known. In particular,
 - If $\omega = HHH$ or $\omega = HHT$, then $S_2(\omega) = u^2 S_0$. If we know that $S_2(\omega) = u^2 S_0$, then even without knowing ω , we know that $S_1(\omega) = u S_0$. We define

$$\mathbb{E}(S_1|S_2)(HHH) = \mathbb{E}(S_1|S_2)(HHT) = u S_0.$$

- If $\omega = TTT$ or $\omega = TTH$, then $S_2(\omega) = d^2 S_0$. If we know that $S_2(\omega) = d^2 S_0$, then even without knowing ω , we know that $S_1(\omega) = d S_0$. We define

$$\mathbb{E}(S_1|S_2)(TTT) = \mathbb{E}(S_1|S_2)(TTH) = d S_0.$$

- If $\omega \in A = \{HTH, HTT, THH, THT\}$, then $S_2(\omega) = udS_0$. If we know $S_2(\omega) = udS_0$, then we do not know whether $S_1 = uS_0$ or $S_1 = dS_0$. We then take a weighted average:

$$\mathbb{P}(A) = p^2q + pq^2 + p^2q + pq^2 = 2pq.$$

Furthermore,

$$\begin{aligned} \int_A S_1 d\mathbb{P} &= p^2quS_0 + pq^2uS_0 + p^2qdS_0 + pq^2dS_0 \\ &= pq(u+d)S_0 \end{aligned}$$

For $\omega \in A$ we define

$$\mathbb{E}(S_1|S_2)(\omega) = \frac{\int_A S_1 d\mathbb{P}}{\mathbb{P}(A)} = \frac{1}{2}(u+d)S_0.$$

Then

$$\int_A \mathbb{E}(S_1|S_2) d\mathbb{P} = \int_A S_1 d\mathbb{P}.$$

In conclusion, we can write

$$\mathbb{E}(S_1|S_2)(\omega) = g(S_2(\omega)),$$

where

$$g(x) = \begin{cases} uS_0 & \text{if } x = u^2S_0 \\ \frac{1}{2}(u+d)S_0 & \text{if } x = udS_0 \\ dS_0 & \text{if } x = d^2S_0 \end{cases}$$

In other words, $\mathbb{E}(S_1|S_2)$ is random *only through dependence on* S_2 . We also write

$$\mathbb{E}(S_1|S_2 = x) = g(x),$$

where g is the function defined above.

The random variable $\mathbb{E}(S_1|S_2)$ has two fundamental properties:

- $\mathbb{E}(S_1|S_2)$ is $\sigma(S_2)$ -measurable.
- For every set $A \in \sigma(S_2)$,

$$\int_A \mathbb{E}(S_1|S_2) d\mathbb{P} = \int_A S_1 d\mathbb{P}.$$

2.3.2 Definition of Conditional Expectation

Please see Williams, p.83.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let \mathcal{G} be a sub- σ -algebra of \mathcal{F} . Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mathbb{E}(X|\mathcal{G})$ is defined to be any random variable Y that satisfies:

- (a) Y is \mathcal{G} -measurable,

(b) For every set $A \in \mathcal{G}$, we have the “partial averaging property”

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}.$$

Existence. There is always a random variable Y satisfying the above properties (provided that $\mathbb{E}|X| < \infty$), i.e., conditional expectations always exist.

Uniqueness. There can be more than one random variable Y satisfying the above properties, but if Y' is another one, then $Y = Y'$ almost surely, i.e., $\mathbb{P}\{\omega \in \Omega; Y(\omega) = Y'(\omega)\} = 1$.

Notation 2.1 For random variables X, Y , it is standard notation to write

$$\mathbb{E}(X|Y) \triangleq \mathbb{E}(X|\sigma(Y)).$$

Here are some useful ways to think about $\mathbb{E}(X|\mathcal{G})$:

- A random experiment is performed, i.e., an element ω of Ω is selected. The value of ω is partially but not fully revealed to us, and thus we cannot compute the exact value of $X(\omega)$. Based on what we know about ω , we compute an estimate of $X(\omega)$. Because this estimate depends on the partial information we have about ω , it depends on ω , i.e., $\mathbb{E}[X|Y](\omega)$ is a function of ω , although the dependence on ω is often not shown explicitly.
- If the σ -algebra \mathcal{G} contains finitely many sets, there will be a “smallest” set A in \mathcal{G} containing ω , which is the intersection of all sets in \mathcal{G} containing ω . The way ω is partially revealed to us is that we are told it is in A , but not told which element of A it is. We then define $\mathbb{E}[X|Y](\omega)$ to be the average (with respect to \mathbb{P}) value of X over this set A . Thus, for all ω in this set A , $\mathbb{E}[X|Y](\omega)$ will be the same.

2.3.3 Further discussion of Partial Averaging

The partial averaging property is

$$\int_A \mathbb{E}(X|\mathcal{G}) d\mathbb{P} = \int_A X d\mathbb{P}, \forall A \in \mathcal{G}. \quad (3.1)$$

We can rewrite this as

$$\mathbb{E}[I_A \cdot \mathbb{E}(X|\mathcal{G})] = \mathbb{E}[I_A \cdot X]. \quad (3.2)$$

Note that I_A is a \mathcal{G} -measurable random variable. In fact the following holds:

Lemma 3.10 *If V is any \mathcal{G} -measurable random variable, then provided $\mathbb{E}|V \cdot \mathbb{E}(X|\mathcal{G})| < \infty$,*

$$\mathbb{E}[V \cdot \mathbb{E}(X|\mathcal{G})] = \mathbb{E}[V \cdot X]. \quad (3.3)$$

Proof: To see this, first use (3.2) and linearity of expectations to prove (3.3) when V is a *simple* \mathcal{G} -measurable random variable, i.e., V is of the form $V = \sum_{k=1}^n c_k I_{A_k}$, where each A_k is in \mathcal{G} and each c_k is constant. Next consider the case that V is a nonnegative \mathcal{G} -measurable random variable, but is not necessarily simple. Such a V can be written as the limit of an increasing sequence of simple random variables V_n ; we write (3.3) for each V_n and then pass to the limit, using the Monotone Convergence Theorem (See Williams), to obtain (3.3) for V . Finally, the general \mathcal{G} -measurable random variable V can be written as the difference of two nonnegative random-variables $V = V^+ - V^-$, and since (3.3) holds for V^+ and V^- it must hold for V as well. Williams calls this argument the “standard machine” (p. 56). ■

Based on this lemma, we can replace the second condition in the definition of a conditional expectation (Section 2.3.2) by:

(b’) For every \mathcal{G} -measurable random-variable V , we have

$$\mathbb{E}[V \cdot \mathbb{E}(X|\mathcal{G})] = \mathbb{E}[V \cdot X]. \quad (3.4)$$

2.3.4 Properties of Conditional Expectation

Please see Williams p. 88. Proof sketches of some of the properties are provided below.

(a) $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$.

Proof: Just take A in the partial averaging property to be Ω .

The conditional expectation of X is thus an unbiased estimator of the random variable X .

(b) If X is \mathcal{G} -measurable, then

$$\mathbb{E}(X|\mathcal{G}) = X.$$

Proof: The partial averaging property holds trivially when Y is replaced by X . And since X is \mathcal{G} -measurable, X satisfies the requirement (a) of a conditional expectation as well.

If the information content of \mathcal{G} is sufficient to determine X , then the best estimate of X based on \mathcal{G} is X itself.

(c) (Linearity)

$$\mathbb{E}(a_1 X_1 + a_2 X_2|\mathcal{G}) = a_1 \mathbb{E}(X_1|\mathcal{G}) + a_2 \mathbb{E}(X_2|\mathcal{G}).$$

(d) (Positivity) If $X \geq 0$ almost surely, then

$$\mathbb{E}(X|\mathcal{G}) \geq 0.$$

Proof: Take $A = \{\omega \in \Omega; \mathbb{E}(X|\mathcal{G})(\omega) < 0\}$. This set is in \mathcal{G} since $\mathbb{E}(X|\mathcal{G})$ is \mathcal{G} -measurable. Partial averaging implies $\int_A \mathbb{E}(X|\mathcal{G}) d\mathbb{P} = \int_A X d\mathbb{P}$. The right-hand side is greater than or equal to zero, and the left-hand side is strictly negative, unless $\mathbb{P}(A) = 0$. Therefore, $\mathbb{P}(A) = 0$.

(h) (Jensen's Inequality) If $\phi : R \rightarrow R$ is convex and $\mathbb{E}|\phi(X)| < \infty$, then

$$\mathbb{E}(\phi(X)|\mathcal{G}) \geq \phi(\mathbb{E}(X|\mathcal{G})).$$

Recall the usual Jensen's Inequality: $\mathbb{E}\phi(X) \geq \phi(\mathbb{E}(X))$.

(i) (Tower Property) If \mathcal{H} is a sub- σ -algebra of \mathcal{G} , then

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H}).$$

\mathcal{H} is a sub- σ -algebra of \mathcal{G} means that \mathcal{G} contains more information than \mathcal{H} . If we estimate X based on the information in \mathcal{G} , and then estimate the estimator based on the smaller amount of information in \mathcal{H} , then we get the same result as if we had estimated X directly based on the information in \mathcal{H} .

(j) (Taking out what is known) If Z is \mathcal{G} -measurable, then

$$\mathbb{E}(ZX|\mathcal{G}) = Z.\mathbb{E}(X|\mathcal{G}).$$

When conditioning on \mathcal{G} , the \mathcal{G} -measurable random variable Z acts like a constant.

Proof: Let Z be a \mathcal{G} -measurable random variable. A random variable Y is $\mathbb{E}(ZX|\mathcal{G})$ if and only if

- (a) Y is \mathcal{G} -measurable;
- (b) $\int_A Y d\mathbb{P} = \int_A ZX d\mathbb{P}, \forall A \in \mathcal{G}$.

Take $Y = Z.\mathbb{E}(X|\mathcal{G})$. Then Y satisfies (a) (a product of \mathcal{G} -measurable random variables is \mathcal{G} -measurable). Y also satisfies property (b), as we can check below:

$$\begin{aligned} \int_A Y d\mathbb{P} &= \mathbb{E}(I_A.Y) \\ &= \mathbb{E}[I_A Z \mathbb{E}(X|\mathcal{G})] \\ &= \mathbb{E}[I_A Z.X] \text{ ((b') with } V = I_A Z) \\ &= \int_A ZX d\mathbb{P}. \end{aligned}$$

(k) (Role of Independence) If \mathcal{H} is independent of $\sigma(\sigma(X), \mathcal{G})$, then

$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X|\mathcal{G}).$$

In particular, if X is independent of \mathcal{H} , then

$$\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(X).$$

If \mathcal{H} is independent of X and \mathcal{G} , then nothing is gained by including the information content of \mathcal{H} in the estimation of X .

2.3.5 Examples from the Binomial Model

Recall that $\mathcal{F}_1 = \{\phi, A_H, A_T, \Omega\}$. Notice that $\mathbb{E}(S_2|\mathcal{F}_1)$ must be constant on A_H and A_T .

Now since $\mathbb{E}(S_2|\mathcal{F}_1)$ must satisfy the partial averaging property,

$$\begin{aligned}\int_{A_H} \mathbb{E}(S_2|\mathcal{F}_1) d\mathbb{P} &= \int_{A_H} S_2 d\mathbb{P}, \\ \int_{A_T} \mathbb{E}(S_2|\mathcal{F}_1) d\mathbb{P} &= \int_{A_T} S_2 d\mathbb{P}.\end{aligned}$$

We compute

$$\begin{aligned}\int_{A_H} \mathbb{E}(S_2|\mathcal{F}_1) d\mathbb{P} &= \mathbb{P}(A_H) \cdot \mathbb{E}(S_2|\mathcal{F}_1)(\omega) \\ &= p \mathbb{E}(S_2|\mathcal{F}_1)(\omega), \forall \omega \in A_H.\end{aligned}$$

On the other hand,

$$\int_{A_H} S_2 d\mathbb{P} = p^2 u^2 S_0 + p q u d S_0.$$

Therefore,

$$\mathbb{E}(S_2|\mathcal{F}_1)(\omega) = p u^2 S_0 + q u d S_0, \forall \omega \in A_H.$$

We can also write

$$\begin{aligned}\mathbb{E}(S_2|\mathcal{F}_1)(\omega) &= p u^2 S_0 + q u d S_0 \\ &= (p u + q d) u S_0 \\ &= (p u + q d) S_1(\omega), \forall \omega \in A_H\end{aligned}$$

Similarly,

$$\mathbb{E}(S_2|\mathcal{F}_1)(\omega) = (p u + q d) S_1(\omega), \forall \omega \in A_T.$$

Thus in both cases we have

$$\mathbb{E}(S_2|\mathcal{F}_1)(\omega) = (p u + q d) S_1(\omega), \forall \omega \in \Omega.$$

A similar argument one time step later shows that

$$\mathbb{E}(S_3|\mathcal{F}_2)(\omega) = (p u + q d) S_2(\omega).$$

We leave the verification of this equality as an exercise. We can verify the Tower Property, for instance, from the previous equations we have

$$\begin{aligned}\mathbb{E}[\mathbb{E}(S_3|\mathcal{F}_2)|\mathcal{F}_1] &= \mathbb{E}[(p u + q d) S_2|\mathcal{F}_2] \\ &= (p u + q d) \mathbb{E}(S_2|\mathcal{F}_1) \quad (\text{linearity}) \\ &= (p u + q d)^2 S_1.\end{aligned}$$

This final expression is $\mathbb{E}(S_3|\mathcal{F}_1)$.