Chapter 2

Multi-period Model

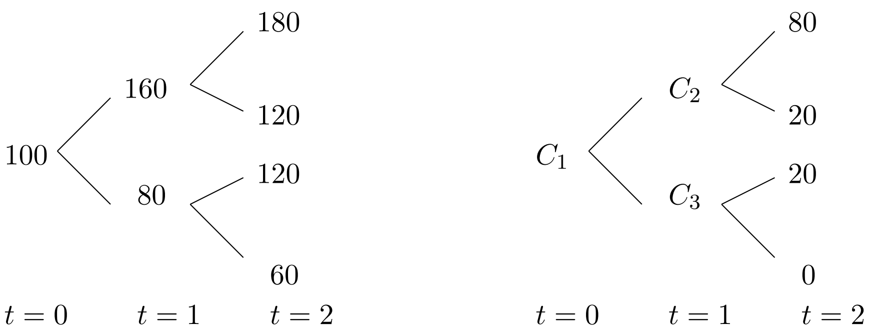
2.1 Multi-period model as a composition of constituent single period models

In Chapter 1, we have looked at the single-period model. In this Chapter, we study the multi-period model. In particular, we begin by analyzing each constituent single-period model and then piecing them together to arrive at the answer. After that, a probabilistic model is presented, which will lead into a coherent theory that will then become the core foundation of the rest of this lecture.

Quiz 3. (Two-period model)
We assume the stock price at time $t = 0$ is 100, and at time $t = 1$ there are only two possibilities that either the stock price will go up to 160 or go down to 80. If the stock price is 160 at $t = 1$, there are only two possibilities that either the stock price will go up to 180 or go down to 120 at $t = 2$. If the stock price is 80 at $t = 1$, there are only two possibilities that either the stock price will go up to 120 or go down to 60 at time $t = 2$. Suppose one has the right, not an obligation, to buy this stock at 100 at time $t = 2$. Then assuming the interest rate is zero, what is the *fair* price of this right that gives no advantage to the buyer or the seller of this right?

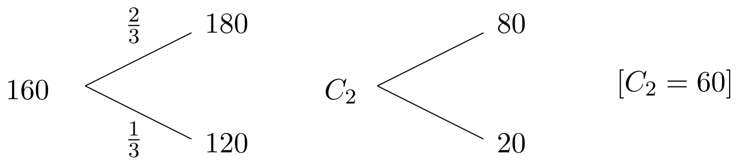
We call this right an option or a contingent claim (of European type). The situation is concisely represented in Figure 2.1, in which C_1, C_2 and C_3 denote the value of this right at each corresponding moment.

Note that we can break this graph into three constituent single-period graphs.(Figure 2.2).

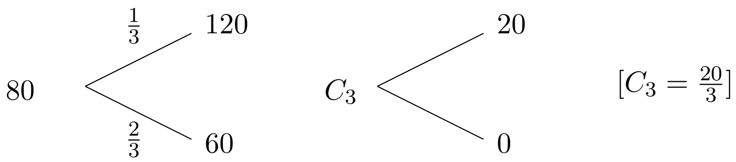


(a) stock price (b) option price

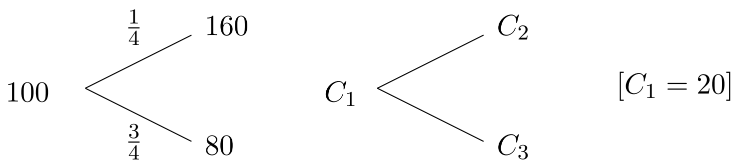
Figure 2.1: The stock price and the value of the option.



(a) constituent single period model
at time $t = 1$ when $S_1 = 160$



(b) constituent single period model
at time $t = 1$ when $S_1 = 80$

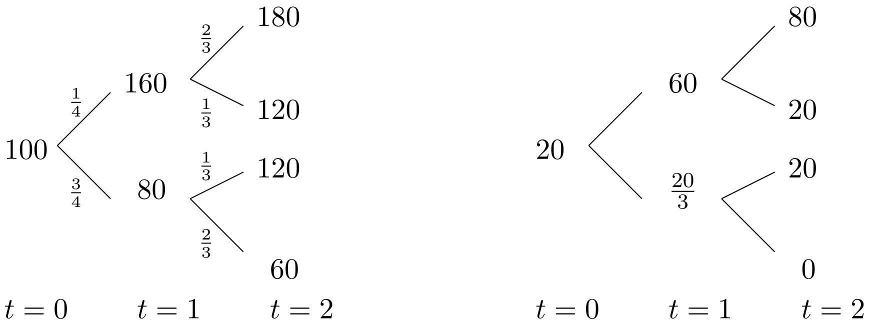


(c) constituent single period model
at time $t = 0$

Figure 2.2: Constituent single-period models.

Applying the method developed in Chapter 1, we can easily find the martingale measures and the value of the option for each single-period. The results are recorded in Figure 2.2 also.

Figure 2.3 shows that the whole picture gotten by piecing together the single-period results.



(a) stock price (b) option price

Figure 2.3: Martingale measure and the option value.

Let us now introduce a probabilistic formalism. First, let w_1, w_2, w_3 and w_4 be paths from $t = 0$ to $t = 2$. Thus they are depicted in Figure 2.4.

We define $\Omega = \{w_1, w_2, w_3, w_4\}$ to be the set of all possible paths in this model. The option then can be thought of as a way of assigning value (at $t = 2$) depending on which path was taken. In other word, the option can be defined as a function $X : \Omega \to \mathbb{R}$ such that $X(w_1) = 80, X(w_2) = 20, X(w_3) = 20$ and $X(w_4) = 0$. In the parlance of probability theory, Ω is called the sample space; each $w \in \Omega$, a sample point; and X a random variable.

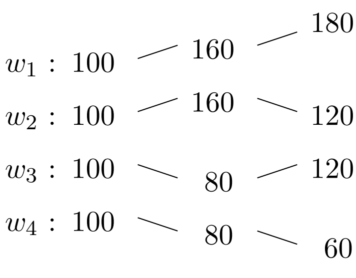


Figure 2.4: Paths from $t = 0$ to $t = 2$.

We now define a new probability measure Q, called the "martingale measure[1]", by multiplying the probabilities of each edge, i.e.

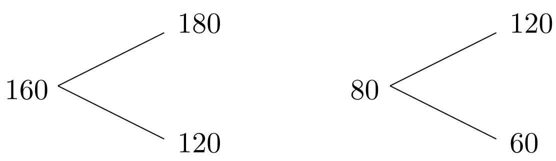
$$Q(w_1) = \frac{1}{4} \times \frac{2}{3} = \frac{2}{12}, \qquad Q(w_2) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12},$$
$$Q(w_3) = \frac{3}{4} \times \frac{1}{3} = \frac{3}{12}, \qquad Q(w_4) = \frac{3}{4} \times \frac{2}{3} = \frac{6}{12}.$$

Let us now rewrite the value C_1 of the option at $t = 0$ by

$$\begin{aligned}
C_1 &= 20 \\
&= \frac{1}{4} \times 60 + \frac{3}{4} \times \frac{20}{3} \\
&= \frac{1}{4} \times \left(\frac{2}{3} \times 80 + \frac{1}{3} \times 20\right) + \frac{3}{4} \times \left(\frac{1}{3} \times 20 + \frac{2}{3} \times 0\right) \\
&= \frac{2}{12} \times 80 + \frac{1}{12} \times 20 + \frac{3}{12} \times 20 + \frac{6}{12} \times 0 \\
&= Q(w_1)X(w_1) + Q(w_2)X(w_2) + Q(w_3)X(w_3) + Q(w_4)X(w_4) \\
&= E_Q[X].
\end{aligned}$$

As we shall be later, this way valuing the option by taking the expectation with respect to martingale measure is a fundamental approach to option pricing.

Remark 2.1. Since two constituent single-period models



have the same price 120 we join the two together to write the whole two period model as

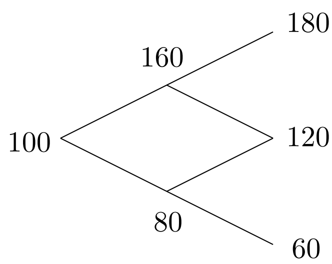


Figure 2.5: Recombinant Tree

The original structure in Figure 2.1 is called a tree while the one in Figure 2.5 is called *recombinant* tree. It should be noted this kind of recombination is due to special price structure of the stock.

[1] We will give more precise definition of martingale measure later in this chapter.

2.2 Replicating Portfolio and Dynamic Hedging

In the previous section, we have constructed and utilized the martingale measure to find the price of option. In this section, let us construct the replicating portfolio for each constituent single-period model.

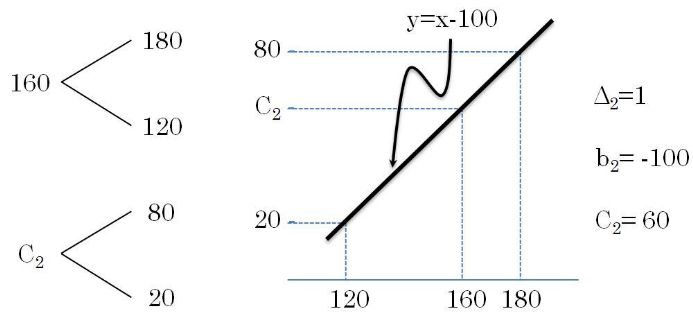


Figure 2.6: The value and the replicating portfolio of the option at time $t = 1$ when $S_1 = 160$.

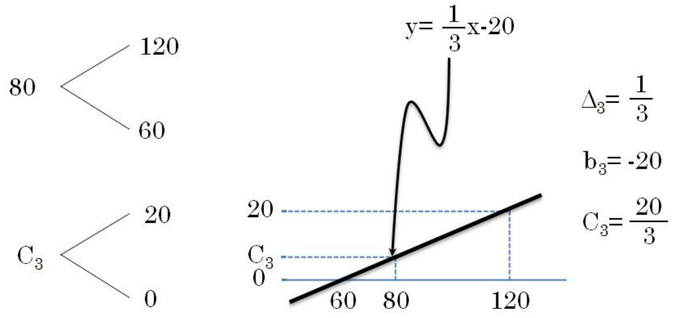


Figure 2.7: The value and the replicating portfolio of the option at time $t = 1$ when $S_1 = 80$.

For the single-period model at time $t = 1$ when $S_1 = 160$ as described in Figure 2.2 (a), we can apply the method in Chapter 1 to obtain the replicating portfolio $(b_2, \Delta_2) = (-100, 1)$, which is depicted in Figure 2.6. Similarly for single-period model at time $t = 1$ when $S_1 = 80$ as in Figure 2.2 (b), we obtain $(b_3, \Delta_3) = (-20, \frac{1}{3})$, which is depicted in Figure 2.7, for the single-period model at time $t = 0$ as in Figure 2.2 (c), $(b_1, \Delta_1) = (-\frac{140}{3}, \frac{2}{3})$, which is depicted in Figure 2.8. Combining them together, we have Figure 2.9. The portfolio constructed and managed this way replicates the option in

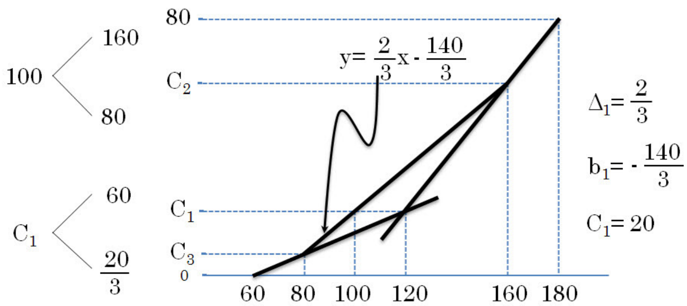


Figure 2.8: The value and the replicating portfolio of the option at time $t = 0$.

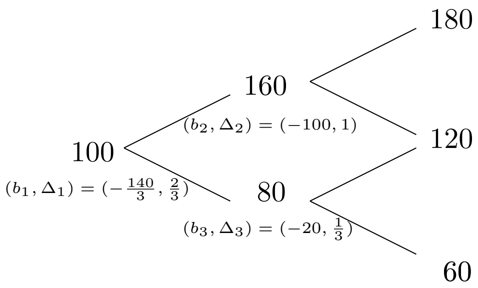


Figure 2.9: Dynamic portfolio.

the following sense. First the option is calculated to be worth 20 at $t = 0$. So an investor, instead of buying the option outright, can do the follows.

(i) At time $t = 0$;

- borrow 140/3 from the bank.
- buy 2/3 shares of stock using the borrowed money and an out-of-pocket cash of 20. [So the initial cash outlay of the investor is 20.]

(ii) Case: $S_1 = 160$ at time $t = 1$;

- increase the total borrowing to 100 by borrowing $\frac{160}{3}$ additionally.
- with it, buy additional $\frac{1}{3}$ share of the stock so that the stock position increases to 1 share.

(iii) Case: $S_1 = 80$ at time $t = 1$;

- sell $\frac{1}{3}$ share of stock to get the sales proceed $\frac{80}{3}$.
- pay the bank $\frac{80}{3}$ to decrease the total borrowing to 20.

(iv) When time $t = 2$;
no matter what happens the value 20 at time $t = 2$ of the portfolio so constructed coincides with that of the option.

This way, this *dynamically* managed portfolio exactly replicates the option at all times under any circumstances. This way of *dynamically* managing portfolio is called the *dynamic hedging* and such dynamically changing portfolio is still called the *portfolio* by dropping the word "dynamic". It is also called a *trading strategy* or a *hedging strategy.*

2.3 Information Structure

Any serious attempt to understand the formalism and the inner workings of mathematical finance presupposes rather sophisticated knowledge of probability theory, in particular, that built on the measure-theoretical foundation.

In this section we begin with a thought experiment as a way of helping the reader grasp intuitively the meaning of information (structure).

Thought Experiment

Let us image the following situation. Suppose there is a room that is divided into two parts. In one part, a die is cast twice and in the other part, there are people who are making bets on the outcomes of the die-casts. However, the gamblers cannot see the result directly. Rather, there is a person in the part of the room where the dice is cast who announces the results in the following manner: after the first cast, that person announces whether the result is even or odd; after the second cast, he announces whether the result is less than or greater than 3.5.

If this is the case, the gamblers can place a wager that pays a certain amount of money if the first outcome is even and the second is less than 3.5. But they cannot place a wager that pays something if the first and the second outcomes are both even as such information is never revealed. Similarly they cannot place a wager that stipulates that the outcomes are both 1 and 1.

However, from the standpoint of the person who reads and announces the outcome of the two casts, all of three wagers are playable. It means that whether a wager is playable (i.e., makes sense) depends entirely on the information each individual can get at. To describe this situation mathematically let us introduce the following mathematical formalism.

First, define the sample space Ω by

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\}.$$

Thus Ω is a finite set consisting of 36 elements. Define A_1, A_2, A_3 and A_4 as follows:

$$\begin{aligned} A_1 &= \{(x,y) : x \text{ is odd and } y \leq 3\}, \\ A_2 &= \{(x,y) : x \text{ is odd and } y \geq 4\}, \\ A_3 &= \{(x,y) : x \text{ is even and } y \leq 3\}, \\ A_4 &= \{(x,y) : x \text{ is even and } y \geq 4\}. \end{aligned}$$

See Figure 2.10 for illustration.

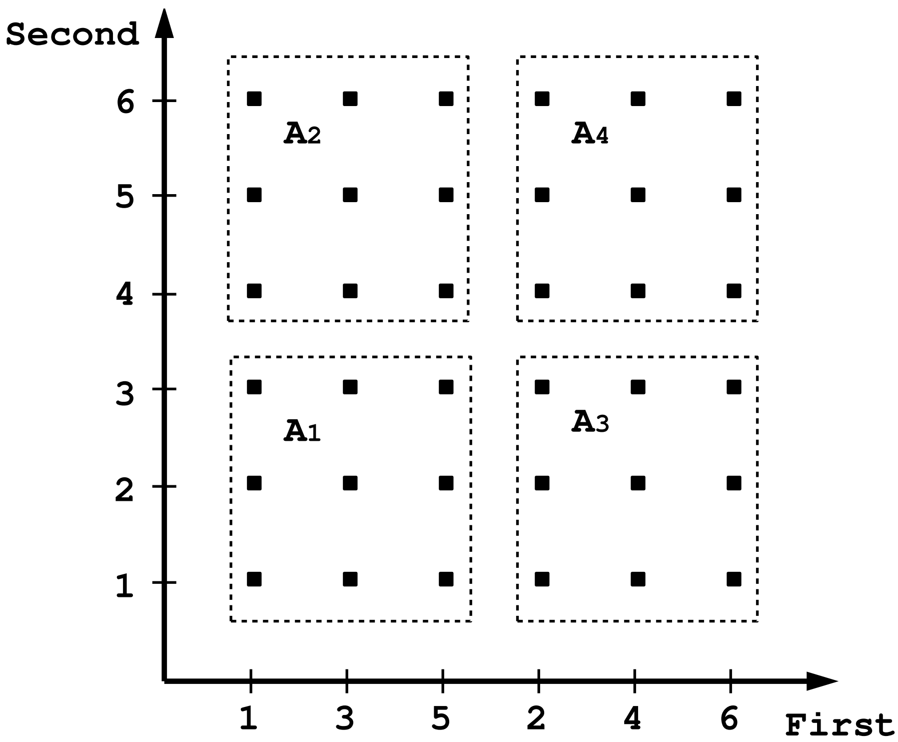


Figure 2.10: the sample space Ω and A_i.

The gamblers will know whether the outcome of two casts belongs to A_1, A_2, A_3 or A_4. But they have no way of knowing any more detailed information. Namely, they can never know if the outcome is such that x is either 1 or 3 and y is 4 or 6.

In this sense, each A_i is an information unit that cannot be further divided as far as the gamblers' knowledge is concerned. On the other hand, the gamblers can make wager on the outcome that the result of the first cast is odd regardless of the outcome of the second cast. It means that they are betting that the outcome of the two casts belongs to $A_1 \cup A_2$. All possible such combinations form a collection $\mathcal{F}$ given below:

$$\begin{aligned} \mathcal{F} = \{ & \phi,\ A_1,\ A_2,\ A_3,\ A_4,\ A_1 \cup A_2,\ A_1 \cup A_3,\ A_1 \cup A_4, \\ & A_2 \cup A_3,\ A_2 \cup A_4,\ A_3 \cup A_4,\ A_1 \cup A_2 \cup A_3, \\ & A_1 \cup A_2 \cup A_4,\ A_1 \cup A_3 \cup A_4,\ A_2 \cup A_3 \cup A_4,\ \Omega\}. \end{aligned}$$

This $\mathcal{F}$ is a collection of subsets of Ω that satisfies the following definition.

Definition 2.2. *Let Ω be any set. (It can be infinite.) A collection $\mathcal{F}$ of subsets of Ω is called a σ-field if*

(i) $\phi, \Omega \in \mathcal{F}$,

(ii) if $A \in \mathcal{F}$, then $A^c = \Omega \backslash A \in \Omega$,

(iii) if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$ then

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

(i.e., a countable union of sets in $\mathcal{F}$ belongs to $\mathcal{F}$.)

(iv) if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$ then

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}.$$

(i.e., a countable intersection of sets in $\mathcal{F}$ belongs to $\mathcal{F}$.)

A subset of Ω belonging to $\mathcal{F}$ is called a measurable set.

Therefore $\mathcal{F}$ constructed above is a σ-field. Let us now pin down the concept of "playability."

Definition 2.3. *Let $X : \Omega \to \mathbb{R}$ be a function and let $\mathcal{F}$ be a σ-field of subset of Ω. We say X is $\mathcal{F}$-measurable ($\mathcal{F}$-random variable) if $\{w \in \Omega : X(w) \in B\} \in \mathcal{F}$ for any interval B of $\mathbb{R}$. In this case, we denote $X \in \mathcal{F}$ by abuse of notation.*

Remark 2.4.

(i) It is customary in probability theory to denote $\{w \in \Omega : X(w) \in B\}$ by $\{X \in B\}$, while in mathematics it is usually denoted by $X^{-1}(B)$.

(ii) If one knows measure theory, this definition can be extended to $\{X \in B\} \in \mathcal{F}$ for any Borel set B. But in most case it suffices to "pretend" that Borel sets are intervals or union of intervals.

Armed with these concepts, let us look at the playable and non-playable wagers.

Example 2.5. *Let X be a wager, i.e. a function $X : \Omega \to \mathbb{R}$, such that*

$$X(w) = \begin{cases} 10, & w \in A_1, \\ 20, & w \in A_2, \\ 30, & w \in A_3, \\ 40, & w \in A_4. \end{cases}$$

Then it is easy to check $X \in \mathcal{F}$, i.e. X is $\mathcal{F}$-measurable or an $\mathcal{F}$-random variable. It is also easy to see that this wager is "playable".

Example 2.6. *Let Y be a wager (i.e. $Y : \Omega \to \mathbb{R}$) such that*

$$Y(x, y) = \begin{cases} 100, & x, y \textit{ are both even}, \\ 200, & \textit{otherwise}. \end{cases}$$

Let $B = (50, 150)$. Then

$$\{Y \in B\} = \{(x, y) : x \textit{ is even, } y \textit{ is even}\}.$$

But $\{Y \in B\} \notin \mathcal{F}$, i.e., Y is not "playable", meaning that the gamblers have no way of determining whether the bet is won or lost.

Let us now turn to the example presented in Section 2.1. The unfolding of events can be easily described as a result of two coin-tossings.

$$\begin{array}{lllll}
w_1 = (H, H): & 100 & \nearrow\ 160 & \nearrow & 180 \\
w_2 = (H, T): & 100 & \nearrow\ 160 & \searrow & 120 \\
w_3 = (T, H): & 100 & \searrow\ 80 & \nearrow & 120 \\
w_4 = (T, T): & 100 & \searrow\ 80 & \searrow & 60
\end{array}$$

$$\Omega = \{w_1, w_2, w_3, w_4\}$$

Figure 2.11: sample space and its sample points.

The sample space Ω in this case is

$$\Omega = \{w_1, w_2, w_3, w_4\}.$$

Obviously the information (knowledge) available to the investors varies as time progresses. At time $t = 0$, nothing is known about the outcome. So the only measurable set must be ϕ or Ω. Thus we define the σ-field $\mathcal{F}_0$ by

$$\mathcal{F}_0 = \{\phi, \Omega\}.$$

At time $t = 1$, the outcome of the first coin-toss becomes known, i.e., a certain amount of information (knowledge) is revealed while the second outcome is still not known yet. If the first outcome is H, then one knows at $t = 1$ the eventually path will be either w_1 or w_2, but still cannot tell which one will be the ultimate outcome. It means that one can say $w \in \{w_1, w_2\}$, but not in any more detail. Similarly if the first outcome is T, one knows $w \in \{w_3, w_4\}$ but no more. This situation is captured by describing a new σ-field $\mathcal{F}_1$ by

$$\mathcal{F}_1 = \{\phi, \{w_1, w_2\}, \{w_3, w_4\}, \Omega\}.$$

At time $t = 2$, more information becomes available. So the corresponding σ-field $\mathcal{F}_2$ is described as

$$\begin{aligned}\mathcal{F}_2 = \{&\phi, \{w_1\}, \{w_2\}, \{w_3\}, \{w_4\}, \{w_1, w_2\}, \{w_1, w_3\}, \{w_1, w_4\},\\ &\{w_2, w_3\}, \{w_2, w_4\}, \{w_3, w_4\}, \{w_1, w_2, w_3\}, \{w_1, w_2, w_4\},\\ &\{w_1, w_3, w_4\}, \{w_2, w_3, w_4\}, \Omega\}\end{aligned}$$

This unfolding of information as time progresses is best described by the family of σ-field parameterized by time called the filtration of σ-fields

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2.$$

This increasing sequence of σ-fields is called a *filtration* or an *information structure.*

Partition and σ-field

Let us now introduce the concept of partition that is a more direct and perhaps more intuitively appealing way of looking at σ-fields.

Definition 2.7. *Let Ω be any non-empty set. A partition $\mathcal{P}$ is a collection of non-empty subsets of Ω such that*

(i) for any $A, B \in \mathcal{P}$, either $A = B$ or $A \cap B = \emptyset$

(ii) the union of all elements of $\mathcal{P}$ is Ω itself.

Remark 2.8. For instance, $\mathcal{P} = \{A_1, A_2, A_3, A_4\}$ in our Thought Experiment is a partition of Ω.

Definition 2.9. *Let $\mathcal{P}$ be a partition of Ω. The σ-field $\sigma(\mathcal{P})$ generated by $\mathcal{P}$ is the set of unions of all possible finite or countable sub-collection of $\mathcal{P}$. (The empty set is the union of empty collection of sets. Therefore $\emptyset \in \mathcal{P}$ and it is also trivial to see that $\sigma(\mathcal{P})$ is a σ-field.)*

The converse is also true as the following proposition shows:

Proposition 2.10. *Let Ω be a finite set and let $\mathcal{F}$ be a σ-field of subsets of Ω. Then there is a partition $\mathcal{P}$ such that $\sigma(\mathcal{P}) = \mathcal{F}$. Furthermore such partition as an unordered collection of disjoint non-empty subsets of Ω is unique.*

Proof. For each $\omega \in \Omega$, define $E(\omega)$ by

$$E(\omega) = \bigcap \{A : \omega \in A \in \mathcal{F}\}.$$

Namely, $E(\omega)$ is the intersection of all measurable sets that contain ω. Then start with some $\omega_1 \in \Omega$, and construct $E(\omega_1)$. Suppose $E(\omega_1) \neq \Omega$. Then there must be $\omega_2 \notin E(\omega_1)$. We claim $E(\omega_1) \cap E(\omega_2) = \emptyset$, for, if not, $E(\omega_2) \setminus E(\omega_1)$ is a measurable set containing ω_2 that is strictly smaller than $E(\omega_2)$. Thus $E(\omega_1)$ and $E(\omega_2)$ are disjoint. If $E(\omega_1) \cup E(\omega_2) \neq \Omega$, we can similarly choose $\omega_3 \notin E(\omega_1) \cup E(\omega_2)$ such that $E(\omega_3)$ is disjoint from $E(\omega_1)$ or $E(\omega_2)$. This process will end in finite steps as Ω is finite, and the resulting collection is the desired partition. The uniqueness easily follows from this construction. □

Definition 2.11. *A subset A of Ω is called a partition element if $A \in \mathcal{P}$.*

The following proposition whose proof is left to the reader is a very handy criterion for measurability of a random variable in case the σ-field is generated by a partition.

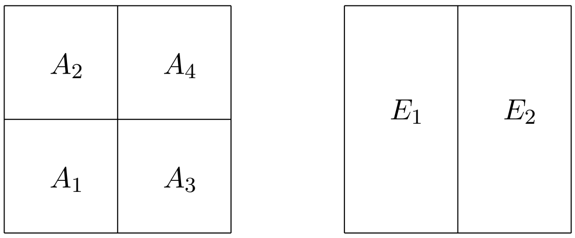
Proposition 2.12. *Let $\mathcal{P}$ be a partition of Ω and let $\mathcal{F} = \sigma(\mathcal{P})$. Let $X : \Omega \to \mathbb{R}$. Then $X \in \mathcal{F}$ if and only if X has a constant value on each partition element of $\mathcal{P}$.*

The partition element is a subset of Ω which loses measurability if it is broken down to a smaller subset; and it is also trivial to see that a subset of Ω is measurable if and only if it is a union of partition elements. Proposition 2.10 means that for *finite* probability space σ-fields and partitions are in one-to-one correspondence.

We now look at how such σ-field related to each other when looked at as partitions.

Definition 2.13. *Let $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ be partitions of Ω. We say $\widetilde{\mathcal{P}}$ is finer than $\mathcal{P}$, if every element of $\mathcal{P}$ is a union of a set of element of $\widetilde{\mathcal{P}}$. In this case $\widetilde{\mathcal{P}}$ is called a refinement of $\mathcal{P}$. We also say that $\mathcal{P}$ is coarser than $\widetilde{\mathcal{P}}$.*

Example 2.14.



$\widetilde{\mathcal{P}} = \{A_1, A_2, A_3, A_4\}$ $\qquad$ $\mathcal{P} = \{E_1, E_2\}$

This picture illustrates $\widetilde{\mathcal{P}}$ as a refinement of P. Note that $E_1 = A_1 \cup A_2$ and $E_2 = A_3 \cup A_4$.

Remark 2.15. Another way of looking at refinement is as follow; Suppose $\widetilde{\mathcal{P}}$ is a refinement of $\mathcal{P}$. Then every $A \in \mathcal{P}$ is broken into one or more elements of $\widetilde{\mathcal{P}}$. In other words, the refinement really means further breaking up element of original partition into a bunch of smaller subsets. This picture will come in handy when we deal with the tree structure associated with filtration or σ-fields (information structure) in Appendix.

The following Proposition whose proof is rather trivial nonetheless is a useful device.

Proposition 2.16. *Let $\mathcal{F}$ be a σ-field generated by a partition $\mathcal{P}$ and $\widetilde{\mathcal{F}}$ be a σ-field generated by a partition $\widetilde{\mathcal{P}}$. Then $\mathcal{F}$ is a sub σ-field of $\widetilde{\mathcal{F}}$ if and only if $\widetilde{\mathcal{P}}$ is a refinement of $\mathcal{P}$.*

2.4 More Probability Theory

In Section 2.3, we have introduced the sample space Ω and the σ-field $\mathcal{F}$, a set of subsets of Ω. The pair $(\Omega, \mathcal{F})$ is usually called a *measure* space. We are now ready to give a formal definition of probability measure.

Definition 2.17. *A probability measure P defined on the measure space $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ such that*

(i) $P(\phi) = 0$; $P(\Omega) = 1$,

(ii) If $\{A_i : i = 1, 2, \ldots\}$ is a collection of mutually disjoint measurable sets (i.e. $A_i \in \mathcal{F}$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i),$$

(iii) If $A \in \mathcal{F}$, then $P(A^c) = P(\Omega \backslash A) = 1 - P(A)$.

Remark 2.18.

(i) The triple $(\Omega, \mathcal{F}, P)$ is usually called a probability space.

(ii) In our discrete model, we always assume $P(A) > 0$, for all non-empty measurable set A, unless stated otherwise.

(iii) Once a probability space $(\Omega, \mathcal{F}, P)$ is given, we can integrate any random variable (i.e. measurable function) $X \in \mathcal{F}$. The definition of integral $\int_\Omega X dP = \int_\Omega X(w) dP(w)$ requires a modicum of knowledge of measure theory, which an interested reader can easily pick up from any textbook. Thus we will use $\int_\Omega X dP$ without giving proper definition by simply appealing to the intuition of the reader.

(iv) When, however, Ω is a finite set, and if every point set is measurable, i.e., $\{w\} \in \mathcal{F}$, $\forall w \in \Omega$, the integral is simply the sum. i.e.,

$$\int_\Omega X dP = \sum_{w \in \Omega} X(w) P(w).$$

Definition 2.19. *(Induced measure) Let $(\Omega, \mathcal{F}, P)$ be probability space, and let $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-random variable. we define a new probability measure P^X on $(\mathbb{R}, \mathcal{B})$ as follow:*

For any $B \in \mathcal{B}$,

$$P^X(B) = P\big(\{X \in B\}\big),$$

where $\mathcal{B}$ is the Borel σ-field, and $B \in \mathcal{B}$.

Remark 2.20.

(i) Technically, Borel σ-field is the smallest σ-field containing all intervals. But for practical purpose, the reader may pretend that B is an interval or a union of intervals.

(ii) P^X is called the probability measure induced by X. If there is a function $f_X(x)$ on $\mathbb{R}$ such that

$$P^X(B) = \int_B f_X(x)dx, \quad \forall B \in \mathcal{B},$$

$f_X(x)$ is called the probability density function of P^X. (In the language of measure theory, $f_X(x)$ exists as the Radon-Nikodym derivative $\dfrac{dP^X}{dx}$, if P^X is absolutely continuous with respect to the standard Lebesgue measure dx on $\mathbb{R}$.)

Intuitive introduction to Lebesgue integration

In the early part of 20^{th} century, Lebesgue introduced a new approach to integration, which revolutionized many parts of mathematics. Although we do not need it in full details, its idea nonetheless is very useful in transcribing the integral over Ω to that over $\mathbb{R}$.

Let $f : [a, b] \to \mathbb{R}$ be a function. For the sake of simplicity, let us assume that f is a continuous function. Let $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_N = b\}$ be a partition of $[a, b]$, and let μ be the usual measure on $[a, b]$. Then the Riemann integral $\displaystyle\int_{[a,b]} f d\mu$ $\Big($usually written as $\displaystyle\int_a^b f(t)dt\Big)$ is given as the limit

$$\int_{[a,b]} f d\mu = \lim_{|\Delta\mathcal{P}|\to 0} \sum_i f(t_i^*)\Delta t_i,$$

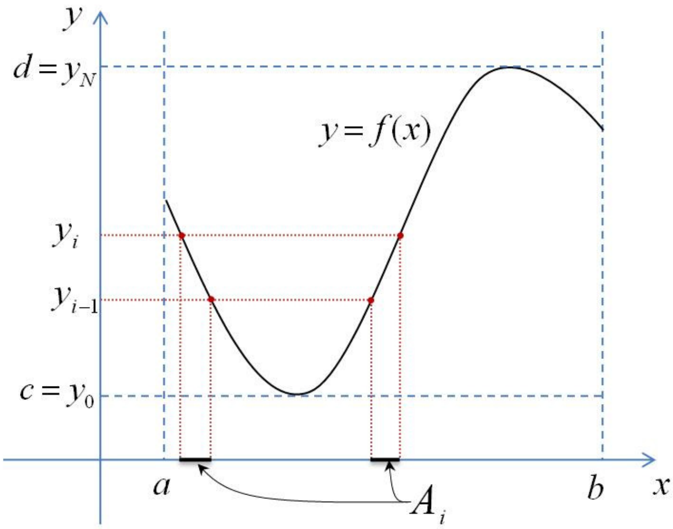
where $\Delta t_i = t_i - t_{i-1}$, $t_i^* \in [t_{i-1}, t_i]$ and $|\Delta\mathcal{P}| = \max_i\{\Delta t_i\}$.

Lesbegue, on the other hand, devised a new way of looking at this integral. He, instead of partitioning the domain, partitioned the range and form a similar sum, and then took its limit. Let $[c, d] = f([a, b])$ be the range of f, and let $\mathcal{P} = \{c = y_0 < y_1 < \cdots < y_N = d\}$ be a partition of $[c, d]$. Define

$$A_i = \{t \in [a, b] : f(t) \in (y_{i-1}, y_i]\}.$$

Then form the sum

$$\sum_i y_i^* \mu(A_i), \tag{2.1}$$



where $y_i^* \in [y_{i-1}, y_i]$ and $\mu(A_i)$ is the measure in $\mathbb{R}$ of A_i with respect to the standard measure of $\mathbb{R}$. In other words, $\mu(A_i)$ is the "total length" of A_i, which, in our example, is a union of intervals. The Lebesgue integral is the limit of (2.1) as $|\Delta\mathcal{P}| \to 0$. It is well known that for any "reasonable" (in particular, continuous) f, the Lebesgue integral coincides with the Riemann integral. i.e.,

$$\int_{[a,b]} f d\mu = \lim_{|\Delta\mathcal{P}|\to 0} \sum_i y_i^* \mu(A_i). \tag{2.2}$$

If we apply the terminology of induced measure as introduced above, we can write

$$\mu(A_i) = \mu^f(\Delta y_i),$$

where μ^f is the measure on the range induced by f, and $\Delta y_i = (y_{i-1}, y_i]$. Therefore (2.2) can be written as

$$\int_{[a,b]} f d\mu = \lim_{|\Delta\mathcal{P}|\to 0} \sum_i y_i^* \mu^f(\Delta y_i). \tag{2.3}$$

The right hand side of (2.3) is symbolically written as

$$\int_{[c,d]} y\mu^f(dy),$$

which is really

$$\int_{[c,d]} y d\mu^f(y).$$

To summarize, we have

$$\int_{[a,b]} f(x)d\mu(x) = \int_{[c,d]} y d\mu^f(y). \tag{2.4}$$

Remark 2.21. This formula (2.4) is of utmost significance in probability theory in that the integral over the domain (LHS) is expressed as the integral over the range (RHS) via the induced measure μ^f.

Expectations written in terms of the induced measure

Formula (2.4) has a direct bearing on the (measure-theoretic) integral of random variables. The argument leading to (2.4) can be rigorously justified in terms of measure theory, although we wouldn't go in there. Namely, the following is true:

Proposition 2.22. *Let $\{\Omega, \mathcal{F}, P\}$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-random variable. Then*

(1) $E_P[X] = \int_\Omega X dP = \int_\mathbb{R} x dP^X(x) = \int_\mathbb{R} x P^X(dx)$.

(2) For any continuous (in fact, Borel measurable) function $\varphi : \mathbb{R} \to \mathbb{R}$,

$$E_P[\varphi(X)] = \int_\Omega \varphi(X) dP = \int_\mathbb{R} \varphi(x) dP^X(x) = \int_\mathbb{R} \varphi(x) P^X(dx).$$

(3) If the probability density function $f_X(x)$ of X exists, then

$$E_P[\varphi(X)] = \int_\mathbb{R} \varphi(x) f_X(x) dx.$$

(4) For any Borel set B of $\mathbb{R}$,

$$\begin{aligned} \int_{\{X \in B\}} \varphi(X) dP &= \int_B \varphi(x) dP^X(x) \\ &= \int_B \varphi(x) f_X(x) dx. \end{aligned}$$

Proof. (1) is the consequence of the argument alluded just above this proposition. In particular, in this context, (2.4) means $\int_\Omega X(\omega) dP(\omega) = \int_\mathbb{R} x dP^X(x)$. (2) can be proved by approximating φ as a sum of simple functions and use (1) then by passing to the limit. (3) follows the definition of the probability density function. (4) can be easily proved if we replace X with $1_{\{X \in B\}} X$ and applying (2) and (3). Here 1_A is the indicator function, meaning $1_A(w) = 1$ if $w \in A$, and 0, otherwise. $\square$

Independence

We need the following concepts of independence.

Definition 2.23. *Let $A, B \in \mathcal{F}$. Then A and B are called independent if $P(A \cap B) = P(A)P(B)$.*

Definition 2.24. *Let $(\Omega, \mathcal{F}, P)$ be a given probability space. We say $A_1, \cdots, A_k \in \mathcal{F}$ are* **independent** *if $P(A_1 \cap \cdots \cap A_k) = \prod_{i=1}^{k} P(A_i)$.*

Definition 2.25. *As in above, let $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_k$ be sub σ-fields of $\mathcal{F}$. Then $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_k$ are called* **independent σ-fields***, if for any $A_i \in \mathcal{F}_i$, $i = 1, \cdots, k$, $P(\cap_{i=1}^{k} A_i) = \prod_{i=1}^{k} P(A_i)$.*

Definition 2.26. *Let X be a random variable. Define $\mathcal{F}(X) = \mathcal{F}_X$ to be the smallest σ-field with respect to which X is measurable. $\mathcal{F}(X)$ is called the σ-field generated by X.*

For illustration of $\mathcal{F}(X)$, refer to the discussion following Question 2.1.

Exercise 2.1. *Let X be a random variable defined on $(\Omega, \mathcal{F})$ as given in the "Thought Experiment". Assume that*

$$X(\omega) = \begin{cases} 10, & \text{if } \omega \in A_1 \cup A_2 \cup A_3; \\ 20, & \text{if } \omega \in A_4. \end{cases}$$

Describe $\mathcal{F}(X)$.

Definition 2.27. *Let $X_1, X_2, \cdots, X_n$ are random variables. $X_1, X_2, \cdots, X_n$ are independent, if $\mathcal{F}(X_1), \cdots, \mathcal{F}(X_n)$ are independent σ-fields.*

Definition 2.28. *A random variable X and a σ-field $\mathcal{G}$ are called independent if $\mathcal{F}_X$ and $\mathcal{G}$ are independent in the sense given above.*

Definition 2.29. *Let X be a random variable and let $F_X(a) = P(X < a)$, which is called* **a cumulative distribution***. Similarly, multi-dimensional cumulative distribution is defined by*

$$F_{X_1,\cdots,X_k}(a_1, \cdots, a_k) = P(X_1 < a_1, X_2 < a_2, \cdots, X_k < a_k),$$

where $X_1, \cdots, X_k$ are random variables.

Exercise 2.2. *Prove that random variables, $X_1, \cdots, X_k$ are independent if and only if*

$$\begin{aligned} F_{X_1,\cdots,X_k}(a_1, \cdots, a_k) &= P(X_1 < a_1, X_2 < a_2, \cdots, X_k < a_k) \\ &= F_{X_1}(a_1)F_{X_2}(a_2) \cdots F_{X_k}(a_k) \end{aligned}$$

for all $a_i \in \mathbb{R}$, $i = 1, \cdots, k$.

Conditional expectation

Let us begin with a simple example. Let $(\Omega, \mathcal{F}, P)$ be a probability such that

$$\Omega = \{w_1, w_2, w_3, w_4\}$$

and $\mathcal{F}$ is the set of all subsets of Ω. We assume

$$P(w_i) = \frac{1}{4}$$

for $i = 1, 2, 3, 4$. Let X and Z be random variables such that

$$Z(w) = \begin{cases} 10, & w = w_1, \\ 20, & w = w_2, \\ 40, & w = w_3, \\ 80, & w = w_4, \end{cases}$$

while

$$X(w) = \begin{cases} 75, & w = w_1 \text{ or } w_2, \\ 10, & w = w_3 \text{ or } w_4. \end{cases}$$

The situation is simply illustrated in following Figure 2.12.

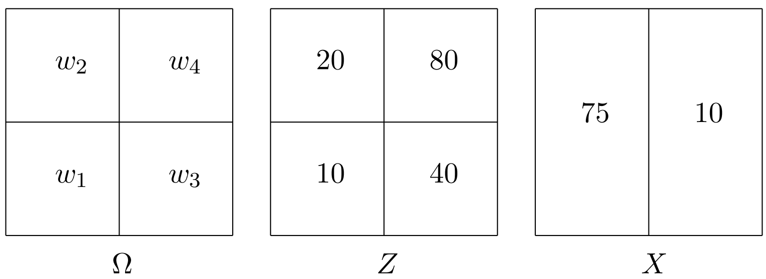


Figure 2.12: Ω, Z and X.

Question 2.1. What is the expected value of Z when we know that $X = 75$?

Since we know $X = 75$, the instance w_3 or w_4 cannot occur. Thus it is reasonable to assume that w_1 or w_2 occurs with equal probability. Thus the expected value of Z in this case must be

$$\frac{1}{2} \times 10 + \frac{1}{2} \times 20 = 15$$

which is written as $E[Z|X = 75]$.

Similarly, it is easy to check that $E[Z|X = 10] = 60$. If we use the notation $E[Z|X]$ by dropping the specific value X takes, $E[Z|X]$ can be regarded as a random variable defined by

$$E[Z|X](w) = \begin{cases} 15, & \text{if } w = w_1 \text{ or } w_2 \\ 60, & \text{if } w = w_3 \text{ or } w_4. \end{cases}$$

Suppose Y is another random variable whose value structure is:

30	80

Y

Then clearly $E[Z|Y=30]=15$ and $E[Z|Y=80]=60$. But $E[Z|Y]$ as a random variable on Ω must coincide with $E[Z|X]$. A moment's ponderance leads us to see that $E[Z|X]$ depends only on the "information structure" of X, not on any specific values X takes on. This information structure of X is in fact $\mathcal{F}(X)$ defined above. Namely, $\mathcal{F}(X)$ is the smallest σ-field with respect to which X is measurable, so in this case

$$\mathcal{G} = \mathcal{F}(X) = \mathcal{F}(Y) = \{\phi, \{w_1, w_2\}, \{w_3, w_4\}, \Omega\}.$$

Also this argument suggests that it is more instructive to use the notation $E[Z|\mathcal{G}]$, which is exactly what we will use from now on. To get a handle on $E[Z|\mathcal{G}]$, let us look at the following question.

Question 2.2. Let $(\Omega, \mathcal{F}, P)$ be as before. Assume X and Z are random variables whose value are given as

50	130
50	80

X

20	80
10	40

Z

What is $E[Z|0<X<100]$?

Then it is clear that w_4 cannot occur and w_1, w_2, w_3 occurs with equal probability $\frac{1}{3}$. Let $B=(0,100)$. Then

$$\begin{aligned} E[Z|0<X<100] &= \frac{1}{3}\times 10 + \frac{1}{3}\times 20 + \frac{1}{3}\times 40 \\ &= \frac{\frac{1}{4}\times 10 + \frac{1}{4}\times 20 + \frac{1}{4}\times 40}{\frac{3}{4}}. \end{aligned}$$

And

$$\begin{aligned}
\text{the numerator} &= \frac{1}{4}\times 10 + \frac{1}{4}\times 20 + \frac{1}{4}\times 40 = \int_{\{X\in B\}} ZdP \\
&= \frac{1}{2}\left(\frac{1}{2}\times 10 + \frac{1}{2}\times 20\right) + \frac{1}{4}\times 40 \\
&= P(X=50)E[Z|X=50] + P(X=80)E[Z|X=80] \\
&= \int_B E[Z|X=x]dP^X(x) \\
&= \int_{\{X\in B\}} E[Z|X]dP.
\end{aligned}$$

The last equality can be seen as follow:
Let $\phi(x) = E[Z|X=x]$. Then

$$\begin{aligned}
\int_B E[Z|X=x]dP^X(x) &= \int_B \phi(x)dP^X(x) \\
&= \int_{\{X\in B\}} \phi(X)dP \\
&= \int_{\{X\in B\}} E[Z|X]dP.
\end{aligned}$$

To summarize, we have

$$\int_{\{X\in B\}} E[Z|X]dP = \int_{\{X\in B\}} ZdP. \tag{2.5}$$

Replace $E[Z|X]$ with $E[Z|\mathcal{G}]$, where $\mathcal{G} = \mathcal{F}(X)$. It is well known (easy to prove) that any measurable set of $\mathcal{G}$ is of the form $\{X \in B\}$. Thus (2.5) can be rephrased as

$$\int_D E[Z|\mathcal{G}]dP = \int_D ZdP$$

$\forall D \in \mathcal{G}$. This motivates the following definition.

Definition 2.30. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $Z : \Omega \to \mathbb{R}$ be a random variable such that $E_p[|Z|] < \infty$. Suppose $\mathcal{G}$ is a sub σ-field of $\mathcal{F}$. Then the conditional expectation $E[Z|\mathcal{G}]$ of Z with respect to $\mathcal{G}$ is defined by*

(i) $E[Z|\mathcal{G}]$ is a $\mathcal{G}$-random variable.

(ii) For any $D \in \mathcal{G}$,

$$\int_D E[Z|\mathcal{G}]dP = \int_D ZdP.$$

Example 2.31. *Let $(\Omega, \mathcal{F}, P)$ be a probability space discussed in Question 2.1. Let $\mathcal{G}$ be the sub σ-field such that*

$$\mathcal{G} = \{\phi, \{w_1, w_2\}, \{w_3, w_4\}, \Omega\}$$

Let Z be a random variable whose values are given by

30	*200*
10	*100*

Z

Let us compute $E[Z|\mathcal{G}]$. First, since $E[Z|\mathcal{G}] \in \mathcal{G}$ and $\mathcal{G}$ cannot distinguish w_1 and w_2, $E[Z|\mathcal{G}]$ must be constant on $D_1 = \{w_1, w_2\}$. Similarly $E[Z|\mathcal{G}]$ must be constant on $D_2 = \{w_3, w_4\}$. Let c_1 and c_2 be its values

c_1	c_2

$E[Z|\mathcal{G}]$

Applying (ii) to D_1, we have

$$\int_{D_1} E[Z|\mathcal{G}]dP = c_1 P(D_1) = \frac{1}{2}c_1$$
$$= \int_{D_1} ZdP = \frac{1}{4} \times 10 + \frac{1}{4} \times 30 = 10$$

Thus $c_1 = 20$. Similarly we can check $c_2 = 150$. Thus $E[Z|\mathcal{G}]$ must have values

20	150

$E[Z|\mathcal{G}]$

Note that $E[Z|\mathcal{G}]$ respects the "information structure" of $\mathcal{G}$.

Properties of conditional expectation

Various properties of conditional expectations listed in Theorem 2.35 are key tools we will use throughout this lecture. Before we proceed we need a few more definitions.

Definition 2.32. *A random variable X is called an integrable random variable if $E[|X|] < \infty$.*

Definition 2.33. *A property is said to hold almost surely (a.s.) if the set of $\omega \in \Omega$ at which this property does not hold (i) is measurable and (ii) has measure zero. For instance, we say two random variables X and Y are equal a.s. (written as $X = Y$, a.s.), if $\{\omega : X(\omega) \neq Y(\omega)\}$ is a measurable set of measure zero.*

The following lemma is a good illustration of the above "almost sure" property; and it will be use in many contexts.

Lemma 2.34. *Let X and Y be two random variables such that*

$$\int_D X dP = \int_D Y dP$$

for any *measurable set $D \in \mathcal{F}$. Then $X = Y$ almost surely.*

Proof. Let $A = \{\omega \in \Omega : X(\omega) \neq Y(\omega)\}$. Then A can be written as

$$A = \bigcup_{n=1}^{\infty} (E_n \cup G_n),$$

where

$$E_n = \left\{\omega \in \Omega : X(\omega) - Y(\omega) > \frac{1}{n}\right\},$$

$$G_n = \left\{\omega \in \Omega : X(\omega) - Y(\omega) < -\frac{1}{n}\right\}.$$

Obviously E_n are G_n are measurable for $n = 1, 2, \cdots$. Thus A is also measurable as it is a countable union of measurable sets. Assume some of them has positive measure, say $P(E_k) > 0$ for some k. Then

$$\int_{E_k} X dP \geq \frac{1}{k} P(E_k) + \int_{E_k} Y dP > \int_{E_k} Y dP.$$

Therefore we can conclude that $P(E_n) = 0$ for any n and we can similarly assert that $P(G_n) = 0$ for all n. Therefore $P(A) = 0$. $\square$

We now list the properties of conditional expectation.

Theorem 2.35. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and Z, Z_1, Z_2, and Y be integrable random variables. Then the following are true where equalities are held to be true in the "almost sure" sense.*

(1) $E[\alpha_1 Z_1 + \alpha_2 Z_2|\mathcal{G}] = \alpha_1 E[Z_1|\mathcal{G}] + \alpha_2 E[Z_2|\mathcal{G}]$, *where α_1 and α_2 are constants.*

(2) *If $Z \geq 0$, then $E[Z|\mathcal{G}] \geq 0$.*

(3) *For σ-fields $\mathcal{D} \subseteq \mathcal{G} \subseteq \mathcal{F}$, $E[E[Z|\mathcal{G}]|\mathcal{D}] = E[Z|\mathcal{D}]$.*

(4) *If Z is independent of $\mathcal{G}$, then $E[Z|\mathcal{G}] = E[Z]$.*

(5) *If $Z \in \mathcal{G}$, then $E[Z|\mathcal{G}] = Z$.*

(6) *If $Z \in \mathcal{G}$ and $Y \in \mathcal{F}$, then $E[ZY|\mathcal{G}] = ZE[Y|\mathcal{G}]$.*

(7) *If $\mathcal{G} = \{\phi, \Omega\}$ (i.e., trivial σ-field), then $E[Z|\mathcal{G}] = E[Z]$.*

Remark 2.36. If $(\Omega, \mathcal{F}, P)$ is a finite probability space such that $P(A) > 0$ for any non-empty measurable set, then one can replace the "almost sure" equalities with genuine equalities.

Sketch of Proof:
We shall prove (3), (4), (5), (6) and (7) and leave the others to reader.

(3) Let $D \in \mathcal{D}$. Then we have,

$$\begin{aligned}
\int_D E[E[Z|\mathcal{G}]|\mathcal{D}]dP &= \int_D E[Z|\mathcal{G}]dP \\
&\qquad \text{(by the definition of } E[\cdot|\mathcal{D}]) \\
&= \int_D Z dP \\
&\qquad \text{(Since } D \in \mathcal{D} \subset \mathcal{G}\text{, use the} \\
&\qquad \text{definition of } E[\cdot|\mathcal{G}]) \\
&= \int_D E[Z|\mathcal{D}]dP. \\
&\qquad \text{(by the definition of } E[\cdot|\mathcal{D}])
\end{aligned}$$

Since this holds for any $D \in \mathcal{D}$, the proof follows from the Lemma 2.34.

(4) If D $\in \mathcal{G}$, then

$$\begin{aligned}
\int_D Z dP &= \int_\Omega \mathbf{1}_D Z dP = E[\mathbf{1}_D Z] \\
&= E[\mathbf{1}_D]E[Z] \quad (\because \text{independence}) \\
&= P(D)E[Z] \\
&= \int_D E[Z]dP.
\end{aligned}$$

So, by the definition of $E[\cdot|\mathcal{G}]$,

$$E[Z|\mathcal{G}] = E[Z].$$

(5) It is an easy consequence of (6) if we let $Y \equiv 1$ in (6), because $E[1|\mathcal{F}] \equiv 1$.

(6) First try $Z = \mathbf{1}_A$, where $A \in \mathcal{G}$. Then, for $D \in \mathcal{G}$

$$\begin{aligned}
\int_D ZY dP &= \int_D \mathbf{1}_A Y dP = \int_{D\cap A} Y dP \\
&= \int_{D\cap A} E[Y|\mathcal{G}] dP \\
&\qquad (\because \text{ by the definition of } E[\cdot|\mathcal{G}]) \\
&= \int_D \mathbf{1}_A E[Y|\mathcal{G}] dP.
\end{aligned}$$

Therefore $E[ZY|\mathcal{G}] = ZE[Y|\mathcal{G}]$ by the definition of $E[\cdot|\mathcal{G}]$. If Z is a sum of simple functions, linearity implies that

$$E[ZY|\mathcal{G}] = ZE[Y|\mathcal{G}].$$

Then, by passing to the limit, we get the desired result.

(7) We have only to show that any $\mathcal{G}$-measurable random variable is a constant. But since Ω is the only non-empty measurable set $E[Z|\mathcal{G}]$ must be constant. Let $E[Z|\mathcal{G}] = c$. Now observe that:

$$c = cP(\Omega) = \int_\Omega E[Z|\mathcal{G}] dP = \int_\Omega Z dP = E[Z].$$

Therefore $E[Z|\mathcal{G}] = E[Z]$.

2.5 Formal Presentation of Discrete Multi-Period Model

In this section, we present a formal mathematical model that will serve as the basic prototype of all subsequent, more sophisticated continuous model. It relies rather heavily on the probabilistic framework so far developed.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. In this section, we always assume Ω is a finite set unless stated otherwise. We also assume that for any non-empty measurable set A has positive measure. Let us first list the basic ingredients of the model.

- **Time**

Time is modeled as discrete integers that runs from 0 to T, i.e. $t = 0, 1, \ldots, T$. Time $t = 0$ represents the present; Time $t = T$ represents the end time of this model. In particular, $t = T$ is the expiry of whatever European open option we will consider.

- **Information structure**

The information unfolds as time progresses, which is modeled as a filtration of sub σ-fields of $\mathcal{F}$

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_t \subset \cdots \subset \mathcal{F}_T$$

where $\mathcal{F}_0 = \{\phi, \Omega\}$. This filtration is usually denoted by $(\mathcal{F}_t)$, or $(\mathcal{F}_t)_{t=0,\cdots,T}$.

- **Stochastic process**

Definition 2.37.

(1) A stochastic process is a family of random variables X_t for each $t = 0, 1, \ldots, T$.

(2) X_t is called an adapted process, if $X_t \in \mathcal{F}_t$ for each $t = 0, 1, \ldots, T$.

(3) An adapted process X_t is called a predictable (previsible) process if $X_t \in \mathcal{F}_{t-1}$ for $t = 1, 2, \ldots, T$.

- **Martingale**

Let Q be any measure defined on $(\Omega, \mathcal{F})$.

Definition 2.38. *Let X_t ba an adapted process. X_t is called a Q-martingale if*

(i) $E_Q[|X_t|] < \infty$ for $t = 0, 1, \ldots, T$ and

(ii) $E_Q[X_{t+1}|\mathcal{F}_t] = X_t$.

X_t is super Q-martingale (resp. sub Q-martingale) if (i) and the following $(ii)'$ satisfied

$(ii)'$ $E_Q[X_{t+1}|\mathcal{F}_t] \leq X_t$ (resp. $E_Q[X_{t+1}|\mathcal{F}_t] \geq X_t$).

- **Bank account (cash bond) process**

Bank account is modeled as a predictable stochastic process B_t such that

(i) $B_0 = 1$ and

(ii) $B_t \geq B_{t-1}$ for $t = 1, \ldots, T$.

- **Assets**

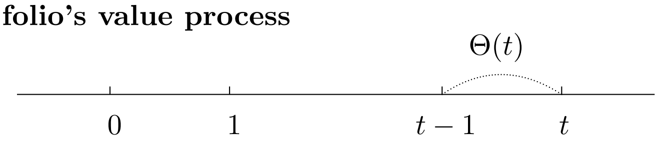
There are N primary assets (stocks) each of which is represented by an adapted stochastic process $S_i(t)$, for $t = 1, 2, \ldots, T$. Also, we usually set $S_0(t) = B_t$.

- **Portfolio(trading strategy)**

For each subinterval $[t-1, t]$, the number of units of i-th asset ($i = 0, 1, \ldots, N$) this portfolio carries is fixed at time $t-1$, and is denoted by $\theta_i(t)$. Thus in our language $\theta_i(t) \in \mathcal{F}_{t-1}$, i.e., $\theta_i(t)$ is a predictable process. The collection of these are denoted by $\Theta(t)$, i.e.,

$$\Theta(t) = (\theta_0(t), \theta_1(t), \ldots, \theta_N(t)).$$

- **Portfolio's value process**



If one determines $\theta_i(t)$ for $i = 0, 1, \ldots, N$ at time $t-1$, the portfolio's value at the end of the holding period $[t-1, t]$, i.e., at t, becomes $\sum_{i=0}^{N} \theta_i(t) S_i(t)$. And at time t the investor then chooses new units $\theta_i(t+1)$ for each $i = 0, 1, \ldots, N$, to be carried throughout the holding period $[t, t+1]$. To do so, he/she needs total sum of money equal to $\sum_{i=0}^{N} \theta_i(t+1) S_i(t)$ at time t. If there is no extra money further brought in or taken out, these two values(sums) must be equal. To formalize this idea, let us use the following definitions

$$V_0 = V(0) = \sum_{i=0}^{N} \theta_i(1) S(0).$$

$$V_t = V(t-) = \sum_{i=0}^{N} \theta_i(t) S_i(t), \quad \text{for } t = 1, \ldots, T.$$

$$V(t+) = \sum_{i=0}^{N} \theta_i(t+1) S_i(t), \quad \text{for } t = 0, \ldots, T-1.$$

Note that $V(t-)$ is the value of the portfolio at time t as a consequence of the position held for the time interval $[t-1, t]$, while $V(t+)$ is the value at time t as a consequence of the new choices made at time t.

Definition 2.39. *The portfolio $\Theta(t)$ is called self-financing if $V(t-) = V(t+)$ for $t = 1, \ldots, (T-1)$.*

Let us now look at the portfolio's value change,

$$\Delta V_t = V_t - V_{t-1} \quad \text{for } t = 1, \ldots, T.$$

For $t = 2, \ldots, T$,

$$\begin{aligned} \Delta V_t &= V_t - V_{t-1} \\ &= \sum_{i=0}^{N} \theta_i(t) S_i(t) - \sum_{i=0}^{N} \theta_i(t-1) S_i(t-1), \end{aligned}$$

and for $t = 1$

$$\begin{aligned} \Delta V_1 &= V_1 - V_0 \\ &= \sum_{i=0}^{N} \theta_i(1) S_i(1) - \sum_{i=0}^{N} \theta_i(1) S_i(0). \end{aligned}$$

If this portfolio is self-financing,

$$V((t-1)-) = V((t-1)+), \text{ for } t = 2, \ldots, T.$$

Thus

$$\begin{aligned} \Delta V_t &= \sum_{i=0}^{N} \theta_i(t) S_i(t) - \sum_{i=0}^{N} \theta_i(t) S_i(t-1) \\ &= \sum_{i=0}^{N} \theta_i(t) \Delta S_i(t), \end{aligned} \tag{2.6}$$

for $t = 2, \ldots, T$, where

$$\Delta S_i(t) = S_i(t) - S_i(t-1).$$

For $t = 1$, we also have

$$\Delta V_1 = \sum_{i=0}^{N} \theta_i(1) \Delta S_i(1).$$

Therefore (2.6) holds for a self-financing portfolio for $t = 1, \ldots, T$. The converse is equally easy to prove so that we have the following result.

Proposition 2.40. *The portfolio is self-financing if and only if*

$$\Delta V_t = \sum_{i=0}^{N} \theta_i(t) \Delta S_i(t)$$

for $t = 1, \ldots, T$.

Exercise 2.3. *Prove proposition 2.40.*

We now discuss the discounted version.

• Discounted asset prices

We define the discounted asset prices by $S_i^*(t) = S_i(t)/B_t = S_i(t)/S_0(t)$ for $i = 0, 1, \cdots N$. Thus in particular, $S_0^*(t) \equiv 1$, and $S_i^*(0) = S_i(0)$.

• Discounted portfolio value

We define discounted portfolio value V_t^* by :

$$\begin{aligned} V_0^* = V_0 &= \sum_{i=0}^{N} \theta_i(1) S_i(0), \\ V_t^* = \frac{V_t}{B_t} &= \sum_{i=0}^{N} \theta_i(t) S_i^*(t) \quad \text{for } t = 1, \ldots, T, \\ V^*(t+) &= \sum_{i=0}^{N} \theta_i(t+1) S_i^*(t) \quad \text{for } t = 1, \ldots, T-1. \end{aligned}$$

The following proposition is a discounted version of Proposition 2.40. Its proof is almost verbatim the same and is hence left to the reader.

Proposition 2.41. *The portfolio is self-financing if and only if*

$$\Delta V_t^* = \sum_{i=1}^{N} \theta_i(t) \Delta S_i^*(t),$$

for $t = 1, \ldots, T$, *where* $\Delta V_t^* = V_t^* - V_{t-1}^*$.

Note that as $\Delta S_0^*(t) \equiv 0$ in the above Proposition the sum can be taken only for $i = 1$ to N.

Exercise 2.4. *Prove proposition 2.41.*

Definition 2.42. *A portfolio is an arbitrage if*

(i) it is self-financing,

(ii) $V_T^* \geq V_0$ *as a random variable,*

(iii) $E_P[V_T^*] \gneqq V_0$.

Thus the presence of an arbitrage means that there is a way of starting with nothing and ending up with no possibility of losing any money at T no matter what happens and there is non-zero probability of making money at T under some favorable circumstances. (Note that, by the assumption on P stated at the beginning of this section, (iii) means that there exists an even A (i.e, $A \in \mathcal{F}$) such that $P(A) > 0$ and $V_T^*(\omega) > V_0$ for $\omega \in A$.) While it is nice to encounter such cases if you are a trader, it entails all kinds of paradoxes in theory and it is not realistic to assume one is consistently lucky to have such opportunity in the free market. So we always assume that the market has no arbitrage.

• Martingale measure

Let us introduce the martingale measure, which is one of the most fundamental tools in finance.

Definition 2.43. *A probability measure Q on a finite probability space $(\Omega,\ \mathcal{F})$ is a martingale measure, if*

(i) $E_Q[\mid S_i(t) \mid] < \infty$, *for all i and t,*

(ii) $E_Q[S_i^*(t+1) \mid \mathcal{F}_t] = S_i^*(t)$ *for each* $i = 1, \cdots, N$,

(iii) $Q(A) > 0$, *for any non-empty* $A \in \mathcal{F}$.

Note that (ii) immediately implies that $E_Q[S_i^*(t+s) \mid \mathcal{F}_t] = S_i^*(t)$ for any $s \geq 1$, which can be easily seen by repeatedly applying the conditional expectation with respect to $\mathcal{F}_{t+s-1}, \mathcal{F}_{t+s-2}, \cdots, \mathcal{F}_t$. The following theorem is one of the most fundamental in finance. Its proof is not hard but needs the machinery of the tree structure to break the multi-period model into a family of constituent single period models. The detail of proof is given in Appendix II.

Theorem 2.44. *There is no arbitrage in the market if and only if there exists a martingale measure.*

Proposition 2.45. *Assume there is no arbitrage in the market. Let Θ be a self-financing portfolio. Then its discounted value process V_t^* is a Q-martingale* for any *martingale measure Q.*

Proof. Since Ω is finite, the integrability of V_t^* is obvious. Now

$$V_{t+1}^* = \sum_{i=0}^{N} \theta_i(t+1) S_i^*(t+1).$$

Thus,

$$\begin{aligned}
E_Q[V_{t+1}^* \mid \mathcal{F}_t] &= E_Q\left[\sum_{i=0}^{N} \theta_i(t+1) S_i^*(t+1) \mid \mathcal{F}_t\right] \\
&= \sum_{i=0}^{N} \theta_i(t+1) E_Q[S_i^*(t+1) \mid \mathcal{F}_t] \qquad (\because \theta_i(t+1) \in \mathcal{F}_t) \\
&= \sum_{i=0}^{N} \theta_i(t+1) S_i^*(t) \qquad (\because S_i^*(t) \text{ is a } Q\text{-martingale.}) \\
&= V_t^*(t+) \\
&= V_t^*. \qquad (\because \text{self-financing condition.})
\end{aligned}$$

$\square$

Definition 2.46.

(1) A European option (contigent claim) with expiry T is a random variable $X \in \mathcal{F}_T$.

(2) A European option is attainable (replicable or marketable) if there is a self-financing portfolio that replicates X at time $t = T$, i.e., the value V_T of the replicating portfolio at $t = T$ coincides with X as random variables. Such portfolio is called a replicating portfolio.

If there is a portfolio replicating a European option X, then the *fair* value of X must be V_0. For, otherwise, one can engage in an arbitrage involving X and the portfolio. Namely, if X is traded at a price less than V_0, then at $t = 0$ one can buy X and sell short the replicating portfolio. If one manages the portfolio dynamically according to the prescription given by $\Theta(t) = (\theta_0(t), \cdots, \theta_N(t))$, one ends up with V_T that is identical to X as random variables. Therefore the original difference in price between V_0 and the market price of X at $t = 0$ becomes the riskless profit. Similarly, if X is traded at a price greater than V_0, then one can sell X and go long with the replicating portfolio. One then similarly end up with a riskless profit that is the difference between the two prices at t.

This way of valuing X is certainly logical. But a trouble may arise if there are several replicating portfolios and their initial values (V_0) may not coincide with each other. The following proposition says that such is not the case.

Proposition 2.47. *Suppose the market has no arbitrage. Let X be a European option with expiry T. Let $\Theta(t)$ and $\widetilde{\Theta}(t)$ replicating portfolios. Then their value at time t ($t = 0, 1, \cdots, T$) coincide, i.e.,*

$$V_t = \widetilde{V}_t,$$

where V_t (resp., $\widetilde{V}_t$) is the value of portfolio $\Theta(t)$ (resp., $\widetilde{\Theta}(t)$).

Proof. Choose any martingale measure Q. Since, by Proposition 2.45, V_t^* and $\widetilde{V}_t^*$ are both Q-martingales, and $V_T^* = X^* = \widetilde{V}_T^*$.

$$\begin{aligned} V_t^* &= E_Q[V_T^* \mid \mathcal{F}_t] = E_Q[X^* \mid \mathcal{F}_t] \\ \widetilde{V}_t^* &= E_Q[\widetilde{V}_T^* \mid \mathcal{F}_t] = E_Q[X^* \mid \mathcal{F}_t]. \end{aligned}$$

Therefore $V_t^* = \widetilde{V}_t^*$, which implies $V_t = \widetilde{V}_t$. $\square$

The arbitrage argument presented above applies at any time t, thus we have the following fundamental principle:

Theorem 2.48 (Martingale (Risk Neutral) Valuation Principle). *Assume the market has no arbitrage. Let X be an attainable European option. Then its value at time t is given by*

$$B_t E_Q \left[\frac{X}{B_T} \mid \mathcal{F}_t \right]$$

for any martingale measure Q. In particular its value at time $t = 0$ is

$$E_Q \left[\frac{X}{B_T} \right].$$

Proof. We have already proved that the value of X at t is given by V_t for any replicating portfolio. In the course of the proof of the above proposition, we also proved that

$$V_t^* = E_Q[X^* \mid \mathcal{F}_t]. \tag{2.7}$$

Obviously V_t^* has nothing to do with any particular Q. Therefore (2.7) must hold for any martingale measure Q. The proof is complete with rewriting (2.7) with $X^* = X/B_T$ and $V_t^* = V_t/B_t$. Since $\mathcal{F}_0 = \{\phi,\ \Omega\}$, and $B_0 = 1$, we also have

$$V_0 = E_Q \left[\frac{X}{B_T} \right].$$

□

Definition 2.49. *The market is complete if* any *European option is attainable.*

The following theorem is also fundamental. Like the proof of theorem 2.44, its proof requires the breaking up of the multi-period tree structure into a family of constituent simple family models. Its proof is given in Appendix II.

Theorem 2.50. *The market is complete if and only if there exists a unique martingale measure.*

2.A Appendix I: Information and Tree

Let $(\Omega, \mathcal{F}, P)$ be a finite probability space, and let

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T \subset \mathcal{F}.$$

be a filtration of σ-fields, i.e., an information structure. This information structure gives rise to a data structure called tree which is more intuitive and demands less of the machinery of probability space.

To help the reader easily grasp the key ideas, let us first use the example we have used at the beginning of Chapter 2. So let Ω be a finite set consisting of four elements

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}.$$

In Section 2.3, we have also defined the information structure $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$ where $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_1 = \big\{\emptyset, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\big\}$ and $\mathcal{F}_2$ is the set of all subsets of Ω. To describe them using partition, let

$$\begin{aligned} \mathcal{P}_0 &= \{A_1\}, \\ \mathcal{P}_1 &= \{B_1, B_2\}, \\ \mathcal{P}_2 &= \{C_1, C_2, C_3, C_4\}, \end{aligned}$$

where $A_1 = \Omega$, $B_1 = \{\omega_1, \omega_2\}$, $B_2 = \{\omega_3, \omega_4\}$, $C_1 = \{\omega_1\}$, $C_2 = \{\omega_2\}$, $C_3 = \{\omega_3\}$, and $C_4 = \{\omega_4\}$. Note also that $\mathcal{F} = \mathcal{F}_2$. Then by Definition 2.9, it is trivial to check that

$$\begin{aligned} \mathcal{F}_0 &= \sigma(\mathcal{P}_0), \\ \mathcal{F}_1 &= \sigma(\mathcal{P}_1), \\ \mathcal{F}_2 &= \sigma(\mathcal{P}_2). \end{aligned}$$

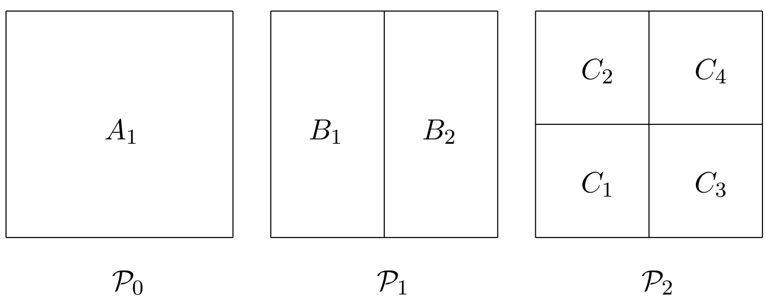


Figure 2.13: Partitions $\mathcal{P}_0, \mathcal{P}_1$ and $\mathcal{P}_2$.

Now using these partitions, let us create a tree in the following manner. For $t = 0$, place one "node" a_1 that represents the only element A_1 of $\mathcal{P}_0$; for $t = 1$, place two "nodes" b_1 and b_2 that represent two elements B_1 and B_2 of $\mathcal{P}_1$, respectively; for $t = 2$, place four "nodes" c_1, c_2, c_3 and c_4 that represent four elements C_1, C_2, C_3 and C_4 of $\mathcal{P}_2$, respectively. They are drawn as in Figure 2.14:

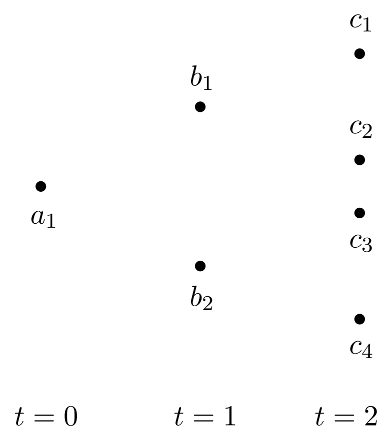


Figure 2.14: Nodes of the Tree.

Next draw an edge (branch, arc) between two nodes using the rule given below:

- If the time difference between two nodes is not exactly one, no edge is drawn
- If the time difference between them is exactly one, an edge between them is drawn if and only if the set in the partition representing one node is a subset of the set in the other partition representing the other node.

For example, in the above picture $C_2 = \{\omega_2\}$ is a subset of $B_1 = \{\omega_1, \omega_2\}$ and the time corresponding to C_2 is 2 while the time for B_1 is 1. Therefore an edge between Nodes b_1 and c_2 is drawn. But since $C_3 = \{\omega_3\}$ is not subset of $B_1 = \{\omega_1, \omega_2\}$, there is no edge connecting b_1 and c_3. Note also that there should be no edge between a_1 and any of the node representing C_1, C_2, C_3, or C_4 because the time difference is 2. If all possible edges are drawn according to this rule, we come up with the tree depicted in Figure 2.15.

The procedure described above is the prescription of creating a tree out of given information structure. Since we always assume $\mathcal{F}_0 = \{\emptyset, \Omega\}$, i.e., $\mathcal{P}_0 = \{\Omega\}$, there is only one node for time $t = 0$. Let us mark it as a special node. In the parlance of graph theory, such node is called the root node. Note also that the nodes corresponding

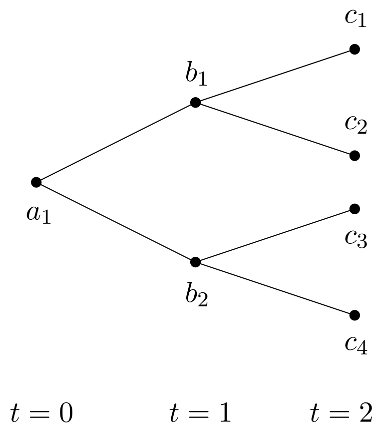


Figure 2.15: Tree corresponding to the Information Structure.

to time $t = T$ are called the leaf nodes, meaning that there is only one edge connected to them.

Conversely, suppose a tree is given. We assume further that there is one particular node designated as the root node. Then we can create a finite measure space $(\Omega, \mathcal{F})$ and an information structure as described below:

Step 1: Arrange node according to time.

Place the root node as the unique node at the level corresponding to $t = 0$. Next, place the nodes of edge distance one from the root node as nodes at the level corresponding to up time t. Suppose we have placed nodes corresponding to up time t. We want to place more nodes corresponding to time $t + 1$ in the following manner: Find the nodes that have edge distance one from some node at level t, discard those that are already placed at level $t - 1$, and place at level $t + 1$ those that are not already at level $t - 1$. This procedure must end since tree is always assumed to have finitely many nodes. As the edges are also already given in the tree in the first place, all we have accomplished so far is the rearrangement of nodes according to the time levels.

For the purpose of illustration, suppose we have a graph as in Figure 2.16, where the node marked with the letter 'r' is taken as the root node. Since a tree is a graph with no cycle it is really a tree. If we rearrange this graph according to the recipe in Step 1, we have a rearranged tree as in Figure 2.17.

One should note that there is no guarantee that all leaf nodes

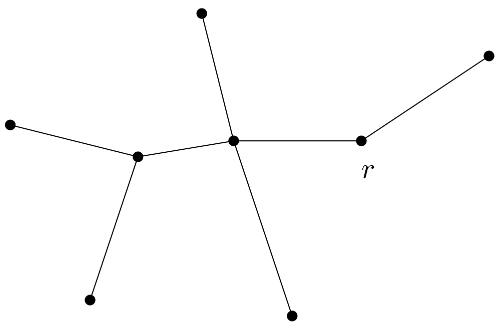
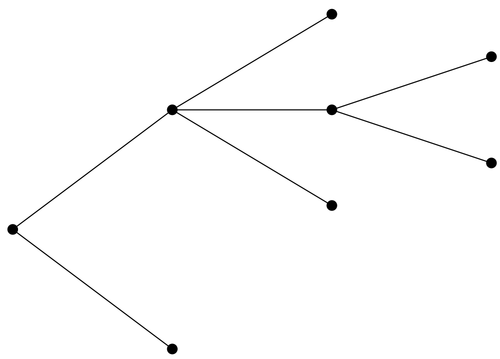


Figure 2.16: Tree graph.



$t = 0$ $t = 1$ $t = 2$ $t = 3(= T)$

Figure 2.17: Rearranged Tree.

may lie at the same time level. Let T be the biggest time level at which some leaf node lies. In Figure 2.17, $T = 3$. We then enlarge the tree in the following manner: suppose n is a leaf node that lies at time level $t < T$. Then add one node to each time level $t + 1, \cdots, T$ and connect them by edges from n all the way to the new leaf node at level T. If we do this for every such leaf node, we have an enlarged tree in which all leaf nodes are placed at time T. For instance, the enlarged tree of the tree in Figure 2.17 would look like the one in Figure 2.18.

To formalize what we have done, we need the following definition.

Definition 2.51. *A tree is a graph in which there is no closed path. A tree is called a* rooted tree*, if one node is designated as a special node, which we call the root or the root node. A rooted tree is called a* normalized rooted tree *if every path from the root node to any leaf node that is not the root node has the same length.*

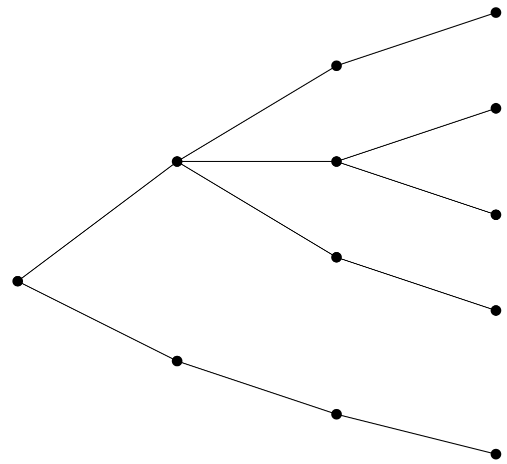


Figure 2.18: Normalized Rooted Tree

In what follows, we assume our tree is always a normalized rooted tree unless stated otherwise.

Step 2: Construction of $(\Omega, \mathcal{F})$.

Let us fix some terminologies. First, by a leaf or a leaf node we always mean a leaf node that is not a root node. By path, we mean a sequence of adjacent edges. We define the sample space Ω to be *the set of all paths from the root node to the leaf nodes*, and the σ-field $\mathcal{F}$ to be the set of all possible subsets of Ω. Together $(\Omega, \mathcal{F})$ defines a measure space.

Step 3: Construction of Information Structure.

For each node, define the *subset of* Ω *corresponding to that node* as the set of paths passing through that node. For reasons that become clear in our subsequent discussion, we call such subset of Ω a partition element corresponding to that node. (For instance, in the example depicted in Figure 2.15, the partition element corresponding to Node b_1 is $B_1 = \{\omega_1, \omega_2\}$, which, according to our definition here, is the set of paths passing through b_1.)

Let $\mathcal{P}_t$ be the set of all such partition element at time level t. In other words, if we let $E(n)$ be *the node set* which is defined to be the set of paths (from the root to the leave nodes) passing through the node n, $\mathcal{P}_t$ is given as

$$\mathcal{P}_t = \{E(n) : n \text{ is a node at time level } t\}.$$

It is easy to see that $\mathcal{P}_t$ is a partition of Ω. For, $E(n) \cap E(m) = \emptyset$ if n and m are distinct node at the same time level t; and since every path passes through some node at time level t, thus

$$\bigcup \{E(n) : n \text{ is a node at time level } t\} = \Omega.$$

Let us now look at the relation between $\mathcal{P}_t$ and $\mathcal{P}_{t+1}$. Let n be a node at time level t, and let $E(n)$ be the node set corresponding to n, i.e., the set of paths passing through n; and let m be a node at time level $t+1$ with $E(m)$ the corresponding node set. Suppose n and m are connected by an edge. Then any path passing through m must pass through n. Thus $E(m)$ must be subset of $E(n)$. On the other hand, suppose there is no edge between n and m. Then no path passing through m can pass through n. Thus $E(m)$ and $E(n)$ are disjoint. Therefore we can conclude that $E(n)$ at time level t is broken up into node sets at time level $t+1$ which again correspond to nodes at time level $t+1$ that are connected to n by edges. In other words, $\mathcal{P}_{t+1}$ is a refinement of $\mathcal{P}_t$. Thus by Proposition 2.16, $\mathcal{F}_t = \sigma(\mathcal{P}_t)$ is a sub σ-field of $\mathcal{F}_{t+1} = \sigma(\mathcal{P}_{t+1})$.

The two procedures outlined above of creating a normalized rooted tree out of finite measure space with information structure and of constructing out of a normalized rooted tree a measure space and information structure define, roughly speaking, a one-to-one correspondence between the set of finite measure space with an information structure and the set of trees. But this correspondence does not hold in the literal sense unless we make proper "qualification".

Obviously what is missing in the whole discussion is $\mathcal{F}$ itself, as our discussion stops at $\mathcal{F}_T$. If $\mathcal{F}_T$ is different from $\mathcal{F}$, we have not prescribed what to do with $\mathcal{F}$. One obvious remedy is to increase artificially the time level by one to, define $\mathcal{F}_{T+1} = \mathcal{F}$, and apply the above procedures with the understanding that the nodes at the last time level (i.e., $T+1$) correspond to the partition elements of $\mathcal{F} = \mathcal{F}_{T+1}$ and the nodes at the penultimate (next to the last) time level (i.e., T) correspond to the partition elements of $\mathcal{F}_T$. For example, look at the case of Thought Experiment. In there Ω is a set of 36 elements. Suppose we define $\mathcal{P}_0 = \{\Omega\}$; $\mathcal{P}_1 = \{E_1, E_2\}$, where $E_1 = A_1 \cup A_2$ and $E_2 = A_3 \cup A_4$; and $\mathcal{P}_2 = \{A_1, A_2, A_3, A_4\}$. Then the above procedure gives rise to the normalized rooted tree in Figure 2.19:

Obviously the node sets corresponding to the leaf nodes are not singleton sets. Each of them contains nine sample points. For in-

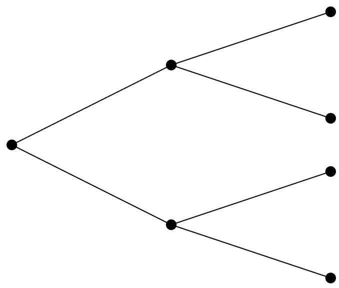


Figure 2.19: Normalized Rooted Tree corresponding to Thought Experiment.

stance

$$A_1 = \{(1,1),(1,2),(1,3),(3,1),(3,2),(3,3),(5,1),(5,2),(5,3)\}.$$

If one adds one more artificial time level, say 3, and define $\mathcal{F}_3 = \mathcal{F}$, the normalized rooted tree in Figure 2.19 changes to:

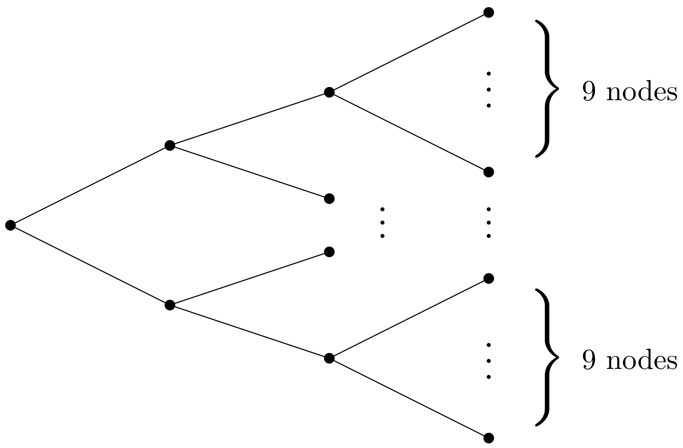


Figure 2.20: Normalized Rooted Tree with $\mathcal{F}_3$.

But if $\mathcal{F}$ itself is not a power set, i.e., not every singleton set is measurable, this remedy of adding one more time level still does not resolve the whole problem. For instance, suppose $A \in \mathcal{F}$ is a partition element that is not a singleton set. Then the normalized rooted tree with time level up to $T+1$ has a leaf node, say n, corresponding to A. Obviously there is only one path from the root to n. Therefore the resulting sample space will have one element (path) representing n, while A itself contains more than one element in the original formalism. In other words, the measure space gotten out of the normalized rooted tree still is not the same as the original measure space even with one more time level $(T+1)$ added.

However these problems do not create any essential difficulty in representing and handling the measure space $(\Omega, \mathcal{F})$ as a tree. First of all, as far as the time (horizon) of the financial problem is concerned, we do not really care about what happens after T. What we are interested in is what happens only up to time T. For this reason, it does not create any essential problem by "pretending" $\mathcal{F}$ is in fact $\mathcal{F}_T$. Second, if there is a partition element of $\mathcal{F}_T$ that is not a singleton set, one cannot distinguish the points in that set by any means of measurability. So we might as well merge all elements into in each partition element of $\mathcal{F}_T$ one without changing any measure-theoretic (probabilistic) character of the setting. These two observations motivate the following:

Definition 2.52. *A finite measure space* $(\Omega, \mathcal{F})$ *with information structure* $(\mathcal{F}_t)_{t=0,\cdots,T}$ *is called in reduced form, if*

(i) $\mathcal{F}_T = \mathcal{F}$.

(ii) $\mathcal{F}$ *is the set of all subsets of* Ω.

Let us now see how to create a reduced-form measure space with information structure out of a given measure space. Suppose $(\Omega, \mathcal{F})$ is a finite measure space and let $(\mathcal{F}_t)_{t=0,\cdots,T}$ be an information structure. Define an equivalence relation $\sim$ on Ω by $\omega_1 \sim \omega_2$ if ω_1 and ω_2 belong to the same partition element of $\mathcal{F}_T$. Then let $\widetilde{\Omega} = \Omega / \sim$ and for any $A \in \mathcal{F}$, define $\widetilde{A} = A / \sim$. Similarly for any sub σ-field $\mathcal{G}$ of $\mathcal{F}$, define

$$\widetilde{\mathcal{G}} = \{\widetilde{A} : \widetilde{A} = A / \sim \text{ where } A \in \mathcal{F}\}.$$

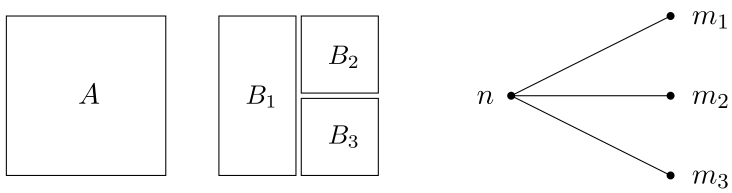
Lastly, define $\mathcal{F} = \mathcal{F}_T$. This way we can define a new finite measure space $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ with new information structure $(\widetilde{\mathcal{F}}_t)_{t=0,\cdots,T}$.

Then $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ with $(\widetilde{\mathcal{F}}_t)_{t=0,\cdots,T}$ is the desired reduced-form measure space. Note also that this reduction procedure can be used to define the equivalence relation among measure spaces with information structure. Similarly, let $X : \Omega \to \mathbb{R}$ be a random variable such that $X \in \mathcal{F}$. Since X is constant on every partition element of $\mathcal{F}$, it is an easy exercise to create out of X a new random variable $\widetilde{X} : \widetilde{\Omega} \to \mathbb{R}$ with $\widetilde{X} \in \widetilde{\mathcal{F}}$ that retains all the essential characters of X.

Armed with all these preliminaries, we are now ready to conclude the following important theorem.

Theorem 2.53. *The set of all finite measure space with information structure in reduced-form is in one-to-one correspondence with the set of all normalized rooted tree by the procedures prescribed above.*

Let us now look at the probability measures from various view points. First, let P be a probability measure on a reduced-form measure space $(\Omega, \mathcal{F})$ with the corresponding normalized rooted tree $\mathcal{T}$. By Theorem 2.53, we can now identify each element of Ω with a path from the root to the leaf node. Similarly, let n be a node of $\mathcal{T}$ at time level t and let A be a partition element in $\mathcal{P}_t$ corresponding to n in the tree representation of $(\Omega, \mathcal{F})$, then A can be identified with the node set $E(n)$ which is defined to be the set of paths from the root to the leaf nodes passing through n. Suppose A is partition element of $\mathcal{F}_t$ that is broken into partition elements $B_1, \cdots, B_k$ of $\mathcal{F}_{t+1}$. Let n be the node at time level t in $\mathcal{T}$ corresponding to A and m_i be the node at time level $t+1$ corresponding to B_i for $i = 1, \cdots, k$. The situation is depicted as in Figure 2.21. We assign the probability of



(a) A and B_i (b) part of $\mathcal{T}$

Figure 2.21: Partition element A broken into B_1, B_2 and B_3.

the node as the probability of the corresponding partition element. For example, suppose in Figure 2.21 that $P(A) = 1/4$, $P(B_1) = 1/8$, $P(B_2) = 1/16$ and $P(B_3) = 1/16$.

Obviously $B_i \cap B_j = \emptyset$ for $i \neq j$ and $A = \bigcup_{i=1}^{k} B_i$. Therefore

$$P(A) = \sum_{i=1}^{k} P(B_i). \tag{2.8}$$

Then the assignment of probability to each node is as in Figure 2.22. Repeating this way, the probability of each node is assigned in such a way that the probability of a node is equal to the sum of the probabilities of its child nodes (the child nodes are the nodes at the next time level nodes connected to it.).

Recall the conditional probability $P(B_i|A)$ is given by

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i)}{P(A)}$$

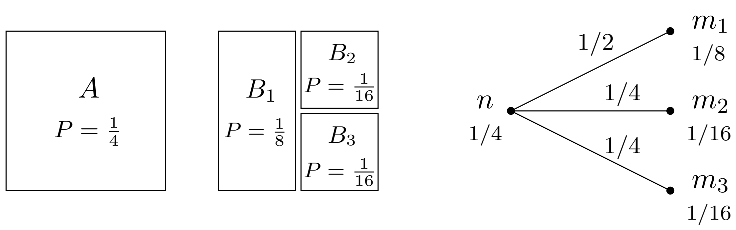


Figure 2.22: Probability assignment.

since $B_i \subset A$. We define the probability of the edge connecting n to m_i to be $P(B_i|A)$. For example, in Figure 2.22, the number $1/2$ on the edge connecting n to m_1 indicates that $P(m_1|n) = 1/2$. Note also that (2.8) implies

$$\sum_{i=1}^{k} P(m_i|n) = 1. \tag{2.9}$$

This relation is true for every node n. Namely, given any node, the sum of the probabilities of the edges connecting it to its child nodes is 1. For instance, check that in Figure 2.22, the sum of the probabilities of the three edges is equal to 1. These facts motivates the following definition.

Definition 2.54. *Let $\mathcal{T}$ be a normalized rooted tree. A number assigned to an edge is called an edge probability if it is positive and is less than or equal to* 1. *The set of edge probabilities is called* consistent *if for every node n, the sum of the edge probabilities of its child nodes is* 1. *i.e., (2.9) holds.*

In the above, we have shown that the probability on $(\Omega, \mathcal{F})$ gives rise to a consistent set of edge probabilities. Let us see the converse is also true. So suppose $\mathcal{T}$ is a normalized rooted tree. Assume that a consistent set of edge probabilities is given. Let n be any node at time t and let m_i be one of its child node. Denote by $P(m_i|n)$ the edge probability of the edge connecting n to m_i. By (2.9), $P(m_i|n)$ can be regarded as the conditional probability. On the other hand, let C be a partition element in $\mathcal{P}_{t+1}$ corresponding to a node that is not connected to n by an edge. It means that $C \cap A = \emptyset$. Therefore $P(C|A) = 0$. Namely, the conditional probability is 0 if two nodes are not connected by an edge. Therefore we have

$$\sum_{u} P(u|n) = 1, \tag{2.10}$$

where the sum is taken over all nodes at time $t+1$. Given such consistent set of edge probabilities, define the probability on $\mathcal{T}$ as follows. For any leaf node, let $e_1, \cdots, e_T$ be the sequence of edges that constitute the path from the root node to that leaf node. Then define the probability of that leaf node to be

$$\prod_{j=1}^{T} P(e_j),$$

where $P(e_j)$ is the edge probability of edge e_j. To see it is indeed a probability, we need to check that sum of all such probabilities is equal to 1. For the convenience of notation, let N_t be the number of nodes at time level t and let i_t be the index that runs from 1 to N_t. We also denote by $n^t_{i_t}$ a typical (generic) node at time level t, while n_0 denotes the root node. It is then easily seen that the sum of the probabilities of the paths is equal to

$$\sum_{i_1,\cdots,i_T} P\left(n^1_{i_1}|n_0\right) P\left(n^2_{i_2}|n^1_{i_1}\right) \cdots P\left(n^{T-1}_{i_{T-1}}\middle|n^{T-2}_{i_{T-2}}\right) P\left(n^T_{i_T}\middle|n^{T-1}_{i_{T-1}}\right). \tag{2.11}$$

In view of (2.10), we have

$$\sum_{i_T=1}^{N_T} P\left(n^T_{i_T}|n^{T-1}_{i_{T-1}}\right) = 1.$$

Therefore (2.11) is equal to

$$\sum_{i_1,\cdots,i_{T-1}} P\left(n^1_{i_1}|n_0\right) P\left(n^2_{i_2}|n^1_{i_1}\right) \cdots P\left(n^{T-1}_{i_{T-1}}\middle|n^{T-2}_{i_{T-2}}\right).$$

Upon repeating, it is seen to be equal to 1.

It is easy to express the probability of each node in terms of edge probabilities. Fix a node at time level t. Recall that according to our construction, the probability $P(n)$ of node n is the sum of probabilities of the paths passing through n, which can be reformulated as follows:

Lemma 2.55. *Let n be a node at time level t, and let $e_1, \cdots, e_t$ be edges connecting the root n_0 to n. Then*

$$P(n) = \prod_{j=1}^{t} P(e_j).$$

Proof. Denote by n_{j-1} and n_j the nodes at the end of the edge e_j. (So in this notation, $n = n_t$.) Then by the construction of the probabilities of $\mathcal{T}$ it is obvious that

$$P(n) = P(n_t) =$$
$$\sum_{i_{t+1},\cdots,i_T} P(n_1|n_0)\cdots P\left(n_t|n_{t-1}\right) P\left(n_{i_{t+1}}^{t+1}\middle|n_t\right)\cdots P\left(n_{i_T}^T\middle|n_{i_{T-1}}^{T-1}\right).$$

Now by (2.10)

$$\sum_{i_T=1}^{N_T} P\left(n_{i_T}^T|n_{i_{T-1}}^{T-1}\right) = 1.$$

Therefore

$$P(n) =$$
$$\sum_{i_{t+1},\cdots,i_{T-1}} P(n_1|n_0)\cdots P\left(n_t|n_{t-1}\right) P\left(n_{i_{t+1}}^{t+1}\middle|n_t\right)\cdots P\left(n_{i_{T-1}}^{T-1}\middle|n_{i_{T-2}}^{T-2}\right).$$

Repeating, we get

$$P(n) = P(n_1|n_0)\cdots P\left(n_t|n_{t-1}\right).$$

□

In summary, we have the following:

Theorem 2.56. *Let $(\Omega, \mathcal{F})$ be a measure space in reduced form and let $\mathcal{T}$ be the corresponding normalized rooted tree.*

(i) A probability measure on $(\Omega, \mathcal{F})$, equivalently a probability measure on the set of paths of $\mathcal{T}$, gives rise to a consistent set of edge probabilities, as prescribed above; conversely

(ii) a consistent set of edge probabilities gives rise to a probability measure on $(\Omega, \mathcal{F})$, or equivalently a probability measure on the set of of $\mathcal{T}$, as prescribed above.

Furthermore (i) and (ii) are inverses to each other.

2.B Appendix II: Proofs of Theorem 2.44 and Theorem 2.50

We present proofs of Theorem 2.44 and Theorem 2.50. In fact, we prove them in the setting of normaized rooted trees, which is equivalent to proving them for reduced-form measure (probability) spaces. In view of "essential" equivalence of the general probability spaces with the reduced-form ones as expounded in Appendix I, whatever modification needed to push our arguments to the general probability space setting is trivial although writing it down is a tedium of work; hence this last modification step is left to the reader.

In what follows, we assume $\mathcal{T}$ is a normalized rooted tree with time level $t = 0, \cdots, T$. Figure 2.23 shows a typical node n with its child nodes $m_1, \cdots, m_k$. Suppose some constituent single period

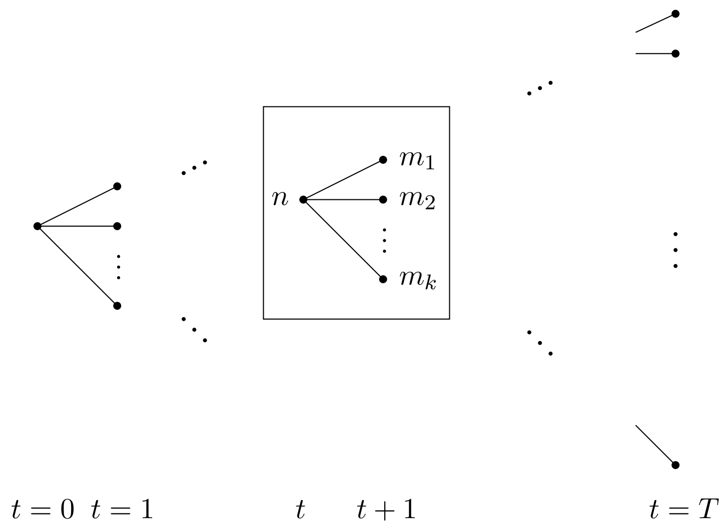


Figure 2.23: A typical node and its child nodes.

of the multi-period model has an arbitrage; say, let us assume that the single period model in the box of Figure 2.23 has an arbitrage. One can then create an arbitrage trading strategy (portfolio) for the whole multi-period model as follows: one does nothing until the path comes to node n where one constructs an arbitrage portfolio for that single period and settle it for cash (bank account) at the next time level. Afterwards, one does nothing until time T. It is certainly an arbitrage portfolio for the entire multi-period model. Before we proceed to the converse, let us look at the single period, say the one in the box of Figure 2.23. Let $V_t^*(n)$ be its discounted value at node n

and let $V^*_{t+1}(m_j)$ be its discounted value at node m_j for $j = 1, \cdots, k$. It is trivial to see that *this* single period portfolio has no arbitrage if and only if

$$\begin{aligned} &\min\{V^*_{t+1}(m_j) : j = 1, \cdots, k\} \\ &\qquad \le V^*_t(n) \le \max\{V^*_{t+1}(m_j) : j = 1, \cdots, k\} \end{aligned} \tag{2.12}$$

Let us now suppose that none of the constituent single period model has an arbitrage. We want to prove that the multi-period model has no arbitrage. To prove that, let us fix *any* portfolio for the multi-period model. Obviously this multi-period portfolio gives rise to single period portfolio for every constituent single period model. Let us look at a single period model in the box of Figure 2.26. Its node at $T-1$ is marked as b; and its nodes at T are marked as $a_1, \cdots, a_k$. Let $V^*_{T-1}(b)$ be the discounted value of the portfolio at node b and $V^*_T(a_j)$ the discounted value at node a_j for $j = 1, \cdots, k$. Since every constituent single period model is assumed to have no arbitrage, by (2.12), we have

$$\begin{aligned} &\min\{V^*_T(a_j) : j = 1, \cdots, k\} \\ &\qquad \le V^*_{T-1}(b) \le \max\{V^*_T(a_j) : j = 1, \cdots, k\}. \end{aligned}$$

This kind of inequality holds for every node at time $T-1$. Repeating the same argument, the same kind of inequality holds for every node at time $T-2$ and so on. Therefore it is rather easy to conclude that

$$\begin{aligned} &\min\{V^*_T(n) : n \text{ is a node at } T\} \\ &\qquad \le V_0 \le \max\{V^*_T(n) : n \text{ is a node at } T\}, \end{aligned}$$

which again mean that the multi-period portfolio itself is not an arbitrage. Therefore, we have the following:

Theorem 2.57. *The multi-period model has no arbitrage if and only if every constituent single period model has no arbitrage.*

Let us now look at the martingale measure. Figure 2.24 depicts a typical constituent single period model. Let $q(e_1), \cdots, q(e_k)$ be edge probabilities. They define a single period martingale measure if and only if

$$S^*_i(t)(n) = \sum_{j=1}^{k} q(e_j) S^*_i(t+1)(m_j), \tag{2.13}$$

where $S^*_i(t)(n)$ is the constant value of $S^*_i(t)$ on the partition element, i.e., the node set $E(n)$, and likewise $S^*_i(t+1)(m_j)$ is the value of $S^*_i(t+1)$ on $E(m_j)$, for any $i = 1, \cdots, N$.

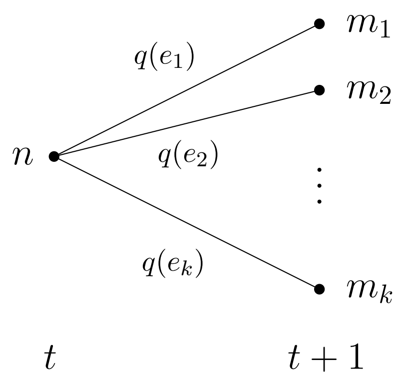


Figure 2.24: Constituent single period model with its martingale measure.

On the other hand, let Q be a multi-period martingale measure. It means precisely that for any node n, and for any $i = 1, \cdots, N$,

$$\int_{E(n)} E_Q\left[S_i^*(t+1)|\mathcal{F}_t\right] dQ = \int_{E(n)} S_i^*(t)dQ. \tag{2.14}$$

However, $S_i^*(t)$ has the constant value $S_i^*(t)(n)$ over $E(n)$; thus the right hand side becomes

$$\int_{E(n)} S_i^*(t)dQ = Q(n)S_i^*(t)(n), \tag{2.15}$$

where $Q(n) = Q(E(n))$. Now

$$E(n) = \bigcup_{j=1}^{k} E(m_j).$$

Therefore, since $S_i^*(t+1)$ has the constant value $S_i^*(t+1)(m_j)$ over $E(m_j)$,

$$\begin{aligned}\int_{E(n)} E_Q\left[S_i^*(t+1)|\mathcal{F}_t\right] dQ &= \int_{E(n)} S_i^*(t+1)dQ \\ &\quad (\because \text{ by definition of} \\ &\quad\ \ \text{conditional expectation}) \\ &= \sum_{j=1}^{k} \int_{E(m_j)} S_i^*(t+1)dQ \\ &= \sum_{j=1}^{k} Q(m_j)S_i^*(t+1)(m_j). \quad (2.16)\end{aligned}$$

Now, the edge probability of the edge e_j connecting n to m_j is

$$q(e_j) = q(m_j|n) = \frac{Q(m_j)}{Q(n)}. \tag{2.17}$$

Therefore, equating (2.15) and (2.16) and using (2.17), we see that (2.14) becomes (2.13). Hence, we have the following:

Theorem 2.58. *Let Q be a probability measure on the normalized rooted tree. It is a multi-period martingale measure if and only if the corresponding single period measure is a martingale measure for every constituent single period. Furthermore, this correspondence from one to the other is inverse to each other.*

• Proof of Theorem 2.44

Assume the multi-period model has no arbitrage. Then by Theorem 2.57, every constituent single-period model has no arbitrage, which in turn implies by Theorem 1.7 that every constituent single-period model has a martingale measure. These single-period martingale measures define a consistent set of edge probabilities in the sense expounded in Appendix I, which in turn defines a measure on the normalized rooted tree representing the multi-period model. Upon invoking Theorem 2.58, this multi-period measure is seen to be a martingale measure.

Conversely, assume there is a martingale measure for the multi-period model. Then Theorem 2.58 implies that it defines a martingale measure for every constituent single period model. Theorem 1.7 then implies that none of the constituent single-period model has arbitrage. Therefore we can conclude that the multi-period model has no arbitrage by invoking Theorem 2.57.

• Proof of Theorem 2.50

Here we assume of course the market has no arbitrage. Suppose the multi-period model is complete. We claim then that every constituent single-period model is complete. Choose any constituent single period model, say the one in the box of Figure 2.23. Let X be any contingent claim for this single period model. It means in particular that we are specifying the value $X(m_j)$ of X at every child node m_j of n. Let $c_j = X^*(m_j)$ for $j = 1, \cdots, k$. On the other hand, the multi-period model is assumed to have no arbitrage, therefore we can choose a martingale measure Q for the multi-period model itself. In view of what we have proved earlier, this martingale measure gives rise to a martingale measure (edge probabilities) for every constituent single-period model, which will be used below. Let us now extend X to every descendant node of n. Let $d_1, d_2, \cdots, d_l$ be the child nodes of m_j as depicted in Figure 2.25. Assign the value

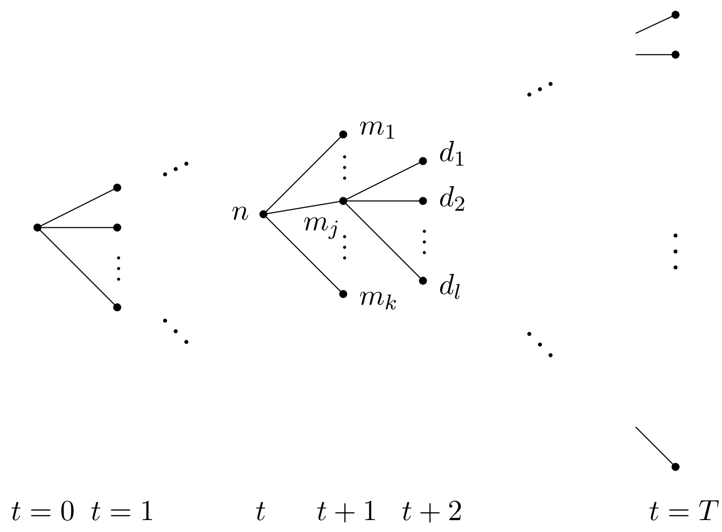


Figure 2.25: Child nodes of m_j.

c_j to every child node of m_j; and then assign the value c_j to every descendant of m_j. We repeat this for each m_j, $j = 1, \cdots, k$. Furthermore, assign the value 0 to every node at time $t+1$ that is not a child node of n; and assign the value 0 to every descendant of such nodes. This assignment in particular defined the value at every node at time T in such a way that the node at time T has value c_j if it is a descendant of some m_j, $j = 1, \cdots, k$; and the value of the node at time T is zero if it is not a descendant of any of m_j, $j = 1, \cdots, k$. This value assignment to the nodes at time T defines a European option $Y \in \mathcal{F}_T$ such that its discounted value Y^* satisfies that

$$\begin{aligned} Y^*(\text{descendant of } m_j) &= c_j, \quad \text{for } j = 1, \cdots, k, \\ Y^*(\text{other leaf node}) &= 0. \end{aligned}$$

By the assumption, Y is an attainable contingent claim, and let $\Theta(t)$ be a replicating portfolio (for the multi-period model). Let us now examine the value of Y^* at the nodes at time $T-1$. Say, look at the single period model in the box of Figure 2.26. Assume first that the node b is a descendant of m_j. Then, by our value assignment

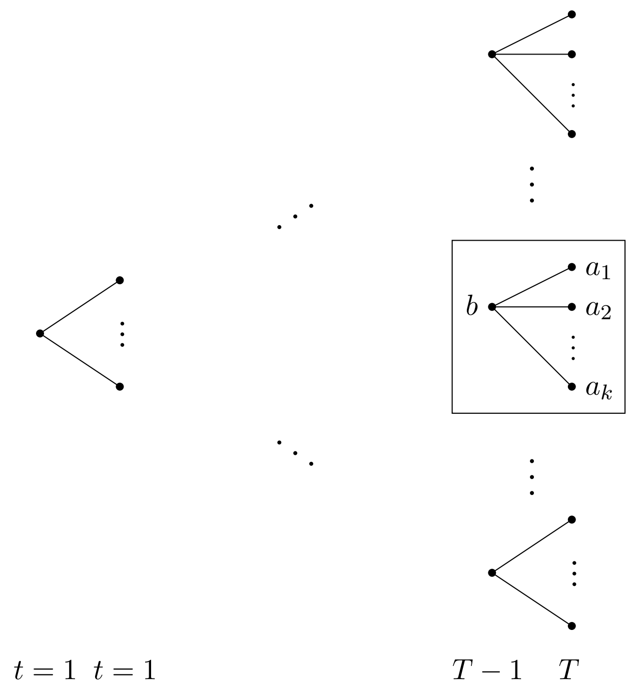


Figure 2.26: Constituent single period models from time $T - 1$ to time T.

scheme, $Y^*(a_i) = c_j$ for any child node a_i of b. Therefore

$$\begin{aligned} Y^*(b) &= \sum_{i=1}^{k} q(a_i|b)Y^*(a_i) \\ &= c_j, \end{aligned}$$

where $q(a_i|b)$ is the edge probability gotten out of Q. Therefore $Y^*(b) = X^*(b)$. If on the other hand b is not a descendant of any of m_j, then b has the value 0 and so does any of its child, say a_i, $i = 1, \cdots, k$. Therefore $Y^*(b) = 0 = X^*(b)$. In other words, we have shown that $X^* = Y^*$ at every node at time $T - 1$. Repeating the same argument, it is easy to see that

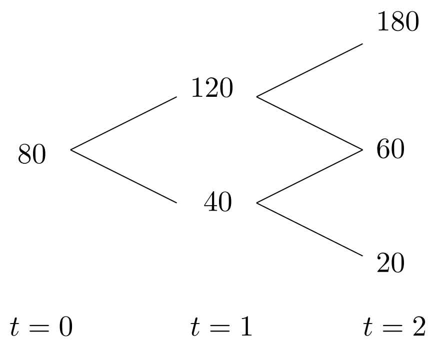
$$X = Y$$

at every node at any time $s = t{+}1, \cdots, T$. In particular, if we restrict X and Y to the nodes m_j, $j = 1, \cdots, k$, of the single period model in the box of Figure 2.23, which of course is the one in question, they are identical as contingent claims. Now we know that Y has a replicating portfolio $\Theta(t)$, which, if restricted to this single period model in question, is a replicating portfolio of Y restricted to this single period model. Therefore the original contingent claim X for the single period model in question also is attainable.

Conversely, suppose the martingale measure for the multi-period model is unique. By Theorem 2.56, the martingale measure for every constituent single period model is unique. Therefore any contingent claim for any constituent single-period model can be replicated, which in turn easily implies that any contingent claim for the multi-period model can be replicated.

Exercises

2.1. Consider a two-period binomial model with the time index t running from $t = 0$ to $t = 2$. This model has two underlying assets, a bank account process B_t s.t. $B_0 = 1, B_{t+1} = (1 + R)B_t$ and stock with price process S_t. Assume interest rate $R = \frac{1}{6}$.



(a) Describe the path space Ω

(b) Describe the martingale measure Q.

(c) Consider a contingent claim X

$$X = \begin{cases} 10, & \text{if } S_2 = 180 \\ 20, & \text{if } S_2 = 60 \\ 30, & \text{if } S_2 = 20 \end{cases}$$

on the underlying stock with maturity $T = 2$.
Find the price of this contingent claim at $t = 0$.

2.2. Let Ω be a finite set consisting of nine elements, i.e.,

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9\}.$$

Pictorially Ω is given as below:

ω_3	ω_6	ω_9
ω_2	ω_5	ω_8
ω_1	ω_4	ω_7

Let $\mathcal{F}$ be the σ-field consisting of all subsets of Ω and let P be probability measure such that $P(\omega_i) = \frac{1}{9}$ for $i = 1, 2, \ldots, 9$. Let $\mathcal{P}$ be a partition such that

$$\mathcal{P} = \{A_1, A_2\}$$

where $A_1 = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $A_2 = \{\omega_5, \omega_6, \omega_7, \omega_8, \omega_9\}$.

(a) Describe the σ-field $\mathcal{D}$ generated by $\mathcal{P}$, i.e., $\mathcal{D} = \sigma(\mathcal{P})$.

(b) Let $X : \Omega \to \mathbb{R}$ be a random variable whose value is given as

20	30	30
20	30	40
10	10	40

Is $X \in \mathcal{D}$? Justify your answer.

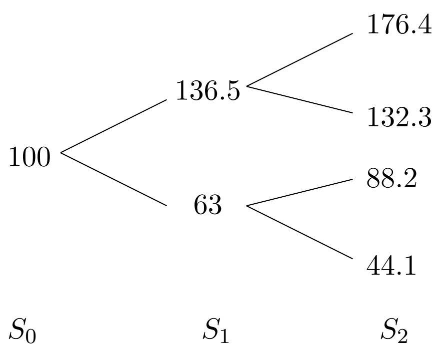
(c) Let $\mathcal{G} = \mathcal{F}_X$. Describe its generating partition $\mathcal{P}_X$ such that $\mathcal{G} = \sigma(\mathcal{P}_X)$.

(d) Let Y be a random variable whose value is given pictorially as follows:

300	600	900
200	500	800
100	400	700

Find $E_P[Y|\mathcal{G}]$ and $E_P[Y|\mathcal{D}]$.

2.3. Let S_t be a stock price process given as below:



Assume that the interest rate for each time interval (i.e., from $t = 0$ to $t = 1$ and $t = 1$ to $t = 2$) is 0.05. Let X be a European option given by

$$X = \begin{cases} 0, & \text{when } S_2 = 176.4; \\ 4.41, & \text{when } S_2 = 132.3; \\ 13.23, & \text{when } S_2 = 88.2; \\ 8.82, & \text{when } S_2 = 44.1. \end{cases}$$

(a) Describe the martingale measure Q.

(b) Compute the value of X at $t = 0$.

(c) Find the replicating portfolio.

(d) Describe $B_1 E_Q[X/B_2|\mathcal{F}_1]$ and explain what it means. (B_t is the bank account process, and $\mathcal{F}_1 = \sigma(S_1)$.)