# Math 622 - Lecture notes I

## TABLE OF CONTENTS

# CHAPTER 1 Probability theory

## 1.1 Probability and Events

### 1.1.1 Events and their properties

Consider an experiment where we toss a coin twice. All the possible outcomes are

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

We call these (elementary) events. The events have the following properties:

a. The union of two events is an event:

$$\{TT\} \cup \{TH\} = \{\text{First toss is } T\}.$$

b. The intersection of two events is an event:

$$\{\text{First toss is } T\} \cap \{\text{Second toss is } T\} = \{TT\}.$$

c. The complement of an event is an event:

$$\{TT\}^c = \{\text{ At least one of the toss is } H\}.$$

Note: In everyday language, union corresponds to OR, intersection corresponds to AND, complement corresponds to NOT.

Suppose we toss a coin $n$ times. It is not difficult to see that that more generally we have the followings:

a'. The union of finitely many events is an event: The event $\{\text{First toss is } T\}$ is the union of finitely many events where each of them has the form $\{T \cdots\}$.

b'. The intersection of finitely many events is an event: The event $\{\text{All tosses are } T\}$ is the intersection of $n$ events where each of them has the form $\{\text{ The nth toss is } T\}$.

Suppose we toss a coin indefinitely. Then we have the followings:

a''. The union of (countably) infinitely many events is an event: The event $\{\text{We eventually see a } T\}$ is the (countable) union of events of the form $\{\text{ The nth toss is } T, n = 1, 2 \cdots\}$.

b''. The intersection of (countably) infinitely many events is an event: The event $\{\text{All the even toss is } T\}$ is the (countable) intersection of events of the form $\{\text{ The nth toss is } T, n = 2, 4, 6 \cdots\}$.

Terminology: When two events have nothing in common (their intersection is $\emptyset$, the empty set) we say they are *mutually exclusive*. For example, the two events $\{\text{First toss is } H\}$ and $\{\text{First toss is } T\}$ are mutally exclusive.

Abstractly, we use capital letters at the beginning of the alphabet: $A, B$ or $E_1, E_2 \cdots$ to denote an event. We also see that in the examples above, an outcome (or an elementary event) is an event that has no sub-event contained in it (in other words, a smallest possible event).

### 1.1.2 Probability

The union of all possible outcomes is an event, (the *universal* event, also called the *sample space*), which we denote by $\Omega$. Then all events are subsets of $\Omega$. We assign a probability, which is a number between 0 and 1, on each event. The probability then is nothing but a mapping from the set of events to the interval $[0, 1]$. Intuitively, this mapping should satisfy the following property:

a. The probability of the union of all outcomes is 1: $P(\Omega) = 1$.

b. The probability of the empty set is 0: $P(\emptyset) = 0$.

c. The probability of the union of two mutually exclusive events is the sum of the individual probability of each event: If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$.

From c, we have the following inclusion - exclusion principle: For any events $A, B$ (not necessarily mutually exclusive)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Exercise: Prove the inclusion - exclusion principle.

Using a,b,c one can come up with more probability identity, for example $P(A^c) = 1 - P(A)$ etc.

When assigning probability, besides a,b,c, we also use the following “commonsense" principle: outcomes that are equally likely have the same probability. For example, if a coin is *fair*, then all outcomes $\{TT\}, \{HH\}, \{TH\}, \{HT\}$ are equally likely. Now applying a and c, we see easily that each of them should have probability equals $1/4$.

### 1.1.3 Examples

**Example 1.1.1.** *We toss a coin twice. The probability that we get at least 1 tail is*

$$P(\{TT\} \cup \{TH\} \cup \{HT\}) = \frac{3}{4}.$$

*The probability that we get no tail is*

$$P(\{HH\}) = \frac{1}{4}.$$

**Example 1.1.2.** *Combinatorics Suppose an urn has 2 white balls and 3 red balls. We pick out (without replacement) 2 balls. What is the probability that the 2 balls are red?*

*Ans: Here we need to see what the sample space is. It is all possible ways we can pick out 2 balls from the urn. What is the event of interest? It is all possible ways we can pick 2* *red balls form the urn. Since each outcome from our pick is equally likely (by equally likely outcome here we mean suppose we number all the balls from 1 to 5, then the possibility we pick out balls 1,2 is the same as the possibiltiy we pick out balls 4,5), the probability of interest is just the ratio of the size of the event with the size of the sample space.*

*Concretely, the number of ways we can pick 2 balls out of 5 balls is* $\binom{5}{2} = 10$*. The number of ways we can pick 2 red balls is* $\binom{3}{2} = 3$*. So the probability is* $\frac{3}{10}$*.*

## 1.2 Conditional probability and independent events

### 1.2.1 Conditional probability

*Motivating example*

Suppose we toss a coin twice. What is the probability that we get 2 tails? From the above, it's $\frac{1}{4}$. Suppose, however, that you know the additional information that the first toss is a tail. We ask the same question: what is the probability that we get 2 tails? Clearly it's no longer $\frac{1}{4}$, because for you, the set of *all possible events* have changed. Namely, the outcomes $\{HH\}, \{HT\}$ are no longer possible.

Concretely, the set of all possible outcomes now are:

$$\{TT\}, \{TH\}.$$

Thus the probability that you get 2 tails is $\frac{1}{2}$. We say: the probability that we get 2 tails, *conditioned on* the first toss being a tail, is $\frac{1}{2}$.

*Conditional probability*

**Definition 1.2.1.** *Let* $A, B$ *be events. If* $P(A) > 0$*, the probability of* $B$ *conditioned on* $A$*, or* $B$ *given* $A$*, denoted* $P(B|A)$*, is defined as:*

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

The interpretation is that we have already had the knowledge that $A$ happened. So the probability of the event B happening, given that $A$ has happened, should be calculated as given in the definition.

**Remark 1.2.2.** *If* $P(A) = 0$ *then we cannot use the above formula to define* $P(B|A)$*. There is a way around it, using the measure theoretic definition of conditional expectation, and the notion of regular conditional probability. We'll discuss this later on in Lecture 1b. See also the discussion on conditional density in Lecture 1b.*

**Example 1.2.3.** *We toss a die. What is the probability that we get a 6, given that we know the toss is even?*

*Ans: Let* $A$ *be the event that we get an even toss,* $B$ *the event that we get a 6 (when you get used to this, you don't have to explicitly name out the events). Then* $P(A) = 1/2$*,* $P(A \cap B) = P(B) = 1/6$*. Thus* $P(B|A) = 1/3$*.*

*Bayes' rule*

From the definition of conditional probability, we have

$$P(B|A)P(A) = P(B \cap A).$$

It is clear that

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

Therefore, we conclude

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

This formula is called the Baye's rule. At first glance this is pure mathematical manipulation. But it has an important implication: that of switching what we conditioned on. An example would illustrate what this means.

It is well-known that medical test is not 100% reliable. That is suppose you test for a disease, which has $1\%$ chance of happening, then even if the test comes out negative, it doesn't mean you have 0% of contracting the disease. Instead, with a very small probability, it could be a false negative. Concretely, suppose that if you indeed have the disease, then there is 98% chance that the test comes out positive, and 2% negative. However, suppose you don't have the disease, there is 95% chance the test comes out negative, and 5% chance it comes out positive. Now you go for the test, and it comes out negative. What is the probability that you contract the disease?

Ans: Let $A$ be the event that you contract the disease and $B$ be the event that the test is positive. Then we have

$$P(B|A) = .98, P(B^c|A) = .02, P(B|A^c) = .05, P(B^c|A^c) = .95.$$

The question asks for $P(A|B^c)$. Thus you see how Bayes' rule is appropriate for the situation. Can you figure out what it is?

### 1.2.2 Independent Events

**Definition 1.2.4.** *Two events $A$ and $B$, are said to be independent if $P(A|B) = P(A)$ and $P(B|A) = P(B)$.*

Remarks: If $P(A|B) = P(A)$ then $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = P(B)$. Thus we actually need one of the two equalities given above for the definition of 2 independent events.

Interpretation: Intuitively, two events are independent if the knowledge of one event already happened does not influence the probability of the other happening, hence the definition.

Alternatively, one can define $A$ and $B$ to be independent if $P(A \cap B) = P(A)P(B)$. You should check that this is equivalent to the condition $P(A|B) = P(A)$ given in the definition. So in fact one have two possible ways to define what it means for 2 events to be independent. The interpretation of the equality $P(A \cap B) = P(A)P(B)$ is not very clear (at least to me) so I prefer to use the other equality for definition of independence.

## 1.3 Random variables

### 1.3.1 Definition

In an experiment, we have (random) outcomes. We can give them names (for example tossing a coin twice, we can get $HH, TT \cdots$). Each of these have some weight attached to them, i.e. their probability ( in the coin toss example, 1/4 for each). However, we cannot do computations with these outcomes unless we give them some numerical values. A *random variable* is a way to *quantify* the random outcomes in a meaningful manner. We use capital letters at the end of the alphabet: $X, Y, Z$, to denote random variables.

Formally, a random variable (*from now on abbreviated as* RV) $X$ is a mapping from the set of outcomes to the real line ($\mathbb{R}$) such that all sets of the form $\{X \in [a, b]\}$ are events. That is, we can assign probability to these sets.

**Example 1.3.1.** *Let $X$ be a random variable corresponding to a coin toss. That is $X = 1$ is the coin turns up $H$ and $X = 0$ if the coin turns up $T$. Then we can see that $P(X = 1) = P(X = 0) = 1/2$.*

Note: There is no reason why $1$ has to be assigned to $H$ and $0$ assigned to $T$. One can assign a different value to these outcomes and get a different variable, as suited one's purpose. For example, the RV $Y$ such that $Y = 1$ if the coin is $H$ nd $Y = -1$ if the coin is $T$ is also an example of a RV.

**Example 1.3.2.** *Let $X$ be a random variable that corresponds to the time one has to wait at the Hill Center's bus stop before one can catch a bus to College Ave. Suppose that the bus arrives every 15 minutes, and they arrive uniformly during any time frame. The we see that $P(a < X < b) = \frac{b-a}{15}$, for $0 \leq a \leq b \leq 15$. Also one should observe that $P(X = a) = 0$ for any $a \in [0, 15]$ (the probability that one waits exactly 7 minutes before the bus arrives is 0).*

### 1.3.2 Discrete versus continuous RVs

In probability theory, one distinguishes between discrete and continuous RVs (note that these are not the only types of RVs there are. One can have a mixed RV as well). Roughly speaking, a discrete RV takes values on a discrete set (for example, the natural numbers is a discrete set, so is $\{1, 2, 3, 4, 5\}$). Moreover, if $X$ is a discrete RV then $P(X = x) > 0$, where $x$ is in the range of $X$. Examples of discrete RVs that you may have learned are: the Binomial, the Geometric, the Hypergeometric, the Poisson.

A continuous RV, on the other hand, takes values on an interval (or several intervals). Moreover, if $X$ is a continuous RV then $P(X = x) = 0$, even if $x$ is in the range of $X$. Examples of continuous RVs that you may have learned are: the Exponential, the Normal, the Uniform, the Gamma, the Cauchy.

### 1.3.3 Probability distribution, pdf, cdf

*Discrete RV*

To characterize a discrete RV, we use the probability distribution function. It gives the formula for the probability that the RV takes some specific value. For example, if $X$ has Bionimial(n,p) distribution, then $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ is the distribution function of $X$.

*Continuous RV*

To characterize a continuous RV, we use the probability density function (pdf). The pdf does not give a probability itself, but it is connected to a probability via the following formula:

$$P(X \leq x) = \int_{-\infty}^{x} f_X(u)du,$$

where $f_X$ above is the pdf of the RV $X$.

### 1.3.4 cdf

Both continuous and discrete RVs can also be described via the cumulative distribution function, which gives the formula for the probability that the RV is less than or equal to some value:

$$F_X(x) = P(X \leq x).$$

Note that if $X$ is a continuous RV, then $F_X$ is differentiable, and its derivative is the density function $f_X$.

### 1.3.5 The moments

*Discrete RV*

Let $X$ be a discrete RV. Then its first moment, the Expectation, is defined as:

$$E(X) = \sum_n nP(X = n),$$

where the sum is understood to be taken over all values in the range of $X$.

It can be showed (note: not a definition) that for any function $f$, the expectation of the RV $f(X)$ is

$$E(f(X)) = \sum_n f(n)P(X = n).$$

In particular, we have the kth moment of $X$ is $E(X^k) = \sum_n n^k P(X = n)$.

*Continuous RV*

For a continuous RV $X$, we define the expectation as:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

More generally, for any function $g$, we have

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

*Variance, covariance, correlation*

Let $X$ be a RV. We then define its variance as

$$Var(X) = E\Big[(X - E(X))^2\Big] = E(X^2) - E^2(X).$$

The variance measures how “spread out" the RV is from its mean.

Let $X, Y$ be RVs. We define their covariance as

$$Cov(X,Y) = E\Big[(X - E(X))(Y - E(Y))\Big] = E(XY) - E(X)E(Y).$$

The covariance measures how “correlated" two RVs are with respect to each other. There is a catch, two different pair of RVs may have the same degree of correlation, but their covariance may be very different. For example, it is clear that

$$Cov(X,X) = Var(X).$$

Intuitively, the degree of "correlation" between $X$ and $X$, versus $100X$ and $100X$ should be the same (they are perfectly correlated in each case). However, you can easily check that $Cov(100X, 100X) = 10000 Cov(X,X)$. Thus we need to introduce another quantity that measures only the correlation and not affected by scaling of the RVs. That is the correlation:

Let $X, Y$ be RVs. We definte their correlation as

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

### 1.3.6 Joint distribution, joint pdf

When we have 2 RVs $X, Y$, besides describing each individual distribution of $X, Y$, we also need to know how they interact together. The joint distribution (in the discrete case) or the joint pdf (in the continuous case) gives us this information. In fact, to calculate $E(XY)$ in the Covariance formula we would need to use the joint distribution of $X, Y$.

a. Discrete: Let $X, Y$ be discrete RVs. Then the joint distribution of $X, Y$ is $P(X = x, Y = y)$.

b. Continuous: Let $X, Y$ be continuous RVs. Then their joint pdf, denoted $f_{X,Y}(x, y)$ is such that

$$P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) du dv.$$

Some elementary properties:

a.

$$\sum_{x,y} P(X = x, Y = y) = 1.$$

b.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u, v) du dv = 1.$$

c. Discrete:

$$E(XY) = \sum_{x,y} xyP(X = x, Y = y).$$

d. Continuous:

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv f_{X,Y}(u, v) du dv.$$

More generally

e. Discrete:

$$E(g(X, Y)) = \sum_{x,y} g(x, y)P(X = x, Y = y).$$

f. Continuous:

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u, v) f_{X,Y}(u, v) du dv.$$

### 1.3.7 Independence

Two random variables $X, Y$ are independent if all events they generated are independent. More specifically, $X, Y$ are independent if for all $x, y$:

$$P(X \le x, Y \le Y) = P(X \le x)Y(\le y).$$

An easier criterion to check is if the joint distribution “splits", i.e.

$$P(X = x, Y = y) = P(X = x)P(Y = y)(\text{ discrete}), \text{ or}$$

$$f_{XY}(x,y) = f_X(x) f_Y(y) \text{ (continous)} .$$

An important property is that if $X, Y$ are independent then $E(XY) = E(X)E(Y)$. Note that, the reverse implication is not generally true. That is $E(XY) = E(X)E(Y)$ does NOT imply that $X, Y$ are independent. See the following example.

**Example 1.3.3.** *Let $X$ have the following distribution: $P(X = 1) = P(X = 0) = P(X = -1) = 1/3$, and let $Y = X^2$. Then it is clear that $X, Y$ are NOT independent (you should try to show this using the definition of independence). However, we can also easily check that*

$$E(XY) = E(X)E(Y) = 0.$$

## 1.4 Conditional expectation

### 1.4.1 Conditional distribution, conditional density

We have discussed conditional probability $P(A|B)$, which is the probability that $A$ happened given the knowledge that $B$ has happened. In a similar way, for 2 RVs $X, Y$, we can talk about the probability that $X$ takes some value $x$ given that we know $Y$ has taken some $y$. If $X$ and $Y$ are correlated in some way, the fact that we have seen $Y$ taking some value should change the probability that $X$ taking value $x$. Formally, we define, for 2 discrete RVs $X, Y$

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

For continuous RVs, we cannot talk about the probablity that $X$ takes some value, given that we have observed $Y$ taking some value. The reason is the probability that $Y$ taking some value is 0, since it is a continuous RV. This poses a slight problem, since in reality, we always observe $Y$ taking some particular value, even if it is a continuous RV (think about the amount of time you wait for the bus to arrive, for example. You always have to wait a particular amount of time until the bus arrives, even if the probability that the continuous random variable representing the time you wait taking that particular value is 0). So for continuous RVs, we talk about the conditional density instead. Formally, we define, for 2 continuous RVs $X, Y$

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

Remark: In the two formulas above, we think of $y$ as *fixed*, and $x$ as taking any possible values in the range of $X$. Thus the conditional distribution, or conditional density, is a function of $x$, given a fixed value $y$. Moreover, for a fixed $y$, the conditional distribution (or probability density), is a probability distribution (or density). That is

$$\sum_x P(X = x|Y = y) = 1; \tag{1.1}$$

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y) dx = 1. \tag{1.2}$$

*Proof.* Left as an exercise.

### 1.4.2 Conditional probability and conditional expectation

*Discrete*

Let $X, Y$ be discrete RVs. The conditional probability $P(X = i|Y = j)$ was defined naturally using the definition of conditional expectation as above. Note that we also have

$$P(X \leq k|Y = j) = \sum_{i \leq k} P(X = i|Y = j).$$

In this way, for every $y$, conditioned on $Y = y$, $P(X < a|Y = y)$ is a proper cumulative distribution function, even though $P(Y = y) = 0$. This is related to the notion of regular conditional probability distribution, discussed below.

We define the conditional expectation of $X$, given $Y = y$ as

$$E(X|Y = y) = \sum_{x} xP(X = x|Y = y).$$

*Continuous*

Let $X, Y$ be continuous RVs. Note that we can NOT define $P(X < a|Y = y)$ using the definition of conditional expectation, because $P(Y = y) = 0$. However, we can define it as followed:

$$P(X < a|Y = y) = \int_{-\infty}^{a} f_{X|Y}(x|y)dx.$$

We define the conditional expectation of $X$, given $Y = y$ as

$$E(X|Y = y) = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx.$$

Interpretation: Besides the fact that conditional expectation is the average (or mean) value of $X$ given $Y = y$, it is also the *best guess* of $X$ given $Y = y$, in some precise sense that we will discuss below.

Remark: Note that in these definitions, $E(X|Y = y)$ is *a real number*. This will be constrasted with $E(X|Y)$, which is a $RV$, the definition of which is given below.

### 1.4.3 Abstract definition of conditional expectation

*Motivation*

The above definitions of $E(X|Y = y)$, while useful, is rather restrictive. It is because we do not have to observe the value of $Y$ to be able to talk about the expectation of $X$

conditioned on $Y$ in a meaningful way. An example will explain. It is clear that the stock price of today depends on the stock price of yesterday (for simplicity let's suppose that stock price only changes discretely from day $n$ to day $n+1$). Suppose we are at day $0$, which is today, and we want to discuss our "expectation", or our best guess, of the stock price on day $n+1$, the guess being made on day $n$. It is clear that on day $n$, we have the knowledge of the stock price of that day, say $S_n$. So what we're asking for is $E(S_{n+1}|S_n)$. Since we are still at day $0$, we do not know what value $S_n$ is, it is a RV to us. However, to discuss our action on day $n$, in anticipation of day $n+1$, it is necessary that we make sense of the notion $E(S_{n+1}|S_n)$. Thus we need an abstract definition of conditional expectation, one that doesn't require us to plug in an observed value for the RV being conditioned on. We will als refer to this as *the measure theoretic definition* of conditional expectation.

*Definition*

**Definition 1.4.1.** *Let $X, Y$ be RVs. The conditional expectation $E(X|Y)$ is a function of $Y$, such that for any function $g$, we have*

$$E\big[E(X|Y)g(Y)\big] = E\big[Xg(Y)\big].$$

Remark: Note that in contrast with the above, as we already said, $E(X|Y)$ is a RV, since it is a function of $Y$ (in some trivial case it could be the constant function, but this does not happen usually). The interpretation of the equality in the definition is that as far as taking expectation with respect to function of $Y$, it does not matter if we use the conditional expectation $E(X|Y)$ or $X$ itself. Thus the conditional expectation $E(X|Y)$ is a guess of $X$, in terms of the random variable $Y$, which satisfies some "indifference" property in terms of expectation.

Perhaps a more satisfactory property of $E(X|Y)$ is that not only it is a guess of $X$ given $Y$, it is *the best* guess of $X$ given $Y$ in the following sense:

**Lemma 1.4.2.** *Let $X, Y$ be RVs. Then for any function $g$ we have*

$$E\Big([E(X|Y) - X]^2\Big) \leq E\Big([g(Y) - X]^2\Big).$$

*Proof.* Left as an exercise.

*Some elementary properties*

The definition (1.4.1) unfortunately does not, most of the time, give us an easy way to compute what $E(X|Y)$ is. So the followings are some elementary properties of conditional expectation that will help us do that. You should try to prove these properties yourself.

a. $E(E(X|Y)) = E(X)$.

b. $E(aX + bY|Z) = a(EX|Z) + bE(Y|Z), a, b$ constant .

c. If $X$ is independent of $Y$ then

$$E(X|Y) = E(X).$$

d. For any function $g$,

$$E(g(Y)X|Y) = g(Y)E(X|Y).$$

e. *The independence lemma:* If $X$ is independent of $Y$ then for any function $g$

$$E\big[g(X,Y)|Y\big] = E\big[g(X,y)\big]|_{y=Y}.$$

*(Properties f and g are not used in this course except in the discussion of regular conditional probability. You can skip these.)*

f. If $X_n \geq 0, X_n \uparrow X$ then $E(X_n|\mathcal{G}) \uparrow E(X|\mathcal{G})$.

g. If $X \in L_2(\Omega, \mathcal{F}, P)$ then $E(X|\mathcal{G})$ is the orthogonal projection of $X$ onto the subspace $L_2(\Omega, \mathcal{G}, P)$ in the Hilbert space $L_2(\Omega, \mathcal{F}, P)$ with inner product $\langle X, Y\rangle := E(XY)$.

Remark: The expression $E\big[g(X,y)\big]|_{y=Y}$ means that we just evaluate $E\big[g(X,y)\big]$ as a *regular expectation* (it is only a random variable in terms of $X$, $y$ is understood to be a constant (or just a dummy variable) here. Note that $E\big[g(X,y)\big]$ is *a function of* $y$. Thus we are free to plug in the random variable $Y$ after we compute what $E\big[g(X,y)\big]$ is.

**Example 1.4.3.** *Let $X$ be a Bernoulli(1/2) random variable and $Y$ has Normal(0,1) distribution, $X$ independent of $Y$. Compute $E(Y^X|Y)$.*

*Ans: We have*

$$E(y^X) = y^0 \cdot \frac{1}{2} + y^1 \cdot \frac{1}{2} = \frac{1}{2}(1+y).$$

*Thus y the independence lemma, $E(Y^X|Y) = \frac{1}{2}(1+Y)$. Note how the distribution of $Y$ is irrelevant in this computation.*

### 1.4.4 Expectation conditional on more than one random variables

In applications, a random variable $X$ may be correlated to not just 1 random variable $Y$, but possibly to $n$ random variables $Y_1, Y_2, \cdots, Y_n$ (it is reasonable to build a model of stock so that the stock price today does not just depend on its performance yesterday, but on its performance in the past month). To discuss the behavior of $X$ given our observations of $Y_1, \cdots Y_n$, we need to extend our notion of conditional expectation to more than 1 random variable. The extension actually is straightforward.

**Definition 1.4.4.** *Let $X, Y_1, Y_2, \cdots, Y_n$ be RVs. The conditional expectation $E(X\big|Y_1, Y_2, \cdots, Y_n)$ is a function of $Y_1, Y_2, \cdots, Y_n$, such that for any function g, we have*

$$E\big[E(X\big|Y_1, Y_2, \cdots, Y_n)g(Y_1, Y_2, \cdots, Y_n)\big] = E\big[Xg(Y_1, Y_2, \cdots, Y_n)\big].$$

Remark: Actually in section (1.4.3), there is no restriction on what the RV $Y$ can be. Thus one could select it to be a *multi-dimensional RV*, effectively making it a random vector with $n$ components

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{bmatrix}.$$

Even more generaly, $X$ itself can also be a multi-dimensional RV,

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_m \end{bmatrix}.$$

Thus we see that we have covered the case of expectation conditional on more than one random variables:

$$E(X_1, X_2, \cdots X_m | Y_1, Y_2, \cdots, Y_n)$$

in section (1.4.3), including the elementary properties. One just needs to interpret the symbol accordingly, for example in $E(aX + bY|Z)$, $a, b$ has to be understood as vector, $aX$ and $bY$ as vector dot products if $X, Y$ are multi-dimensional RV.

### 1.4.5 Probability as an expectation

You may observe that in the abstract definition of conditional expectation, we did not mention about conditional probability. Surely we would want to have a definition for $P(X \le x|Y)$. It turns out that our definition of conditional expectation *already covers conditional probability as a special case.* To be precise, we first need to introduct the following so-called indicator function of an event $E$, denoted as $\mathbf{1}_E$

$$\begin{aligned} \mathbf{1}_E(\omega) &= 1 \text{ if } \omega \in E \\ &= 0 \text{ if } \omega \notin E. \end{aligned}$$

Basically the indicator function is a logical indicator, it's 1 if $E$ happens and 0 if $E$ does not happen. For example $\mathbf{1}_{\{0<1\}} = 1$ and $\mathbf{1}_{\{1+1<3\}} = 0$. But now note that suppose we have a random variable $X$, and say it has a density function $f_X(x)$ then

$$\begin{aligned} E(\mathbf{1}_{\{X \le x\}}) &= \int_{-\infty}^{\infty} \mathbf{1}_{\{y \le x\}}(y) f_X(y) dy \\ &= \int_{-\infty}^{x} f_X(y) dy = P(X \le x). \end{aligned}$$

where the second equality is because $\mathbf{1}_{\{y\leq x\}} = 0$ for all values of $y > x$ so we just stop the integration limit at $x$. Similarly, you can check that

$$\begin{aligned} E(\mathbf{1}_{\{X\geq x\}}) &= P(X \geq x) \\ E(\mathbf{1}_{\{X=x\}}) &= P(X = x). \end{aligned}$$

Thus probability can be expressed as an expectation. More importantly for us, this is still true at the conditional expectation level. More precisely we have the following

**Lemma 1.4.5.** *Let $X, Y$ be random variables. Let $f(Y) = E(\mathbf{1}_{\{X\leq x\}}|Y)$. Then $f(y) = P(X \leq x|Y = y)$ where $P(X \leq x|Y = y)$ is understood in the sense of section (1.4.1). Similarly for $P(X \geq x|Y = y), P(X = x|Y = y)$.*

Remark: For a fixed $x$, the expression $\mathbf{1}_{\{X\leq x\}}$ here is understood as function of $X$. Thus the expression $E(\mathbf{1}_{\{X\leq x\}}|Y)$ is understood in the sense of $E(g(X)|Y)$ where $g(X)$ is just a random variable.

## 1.5 Connection between the measure theoretic and classical definition of conditional expectations

### 1.5.1 Discrete RVs

Let $X, Y$ be two RVs. We have seen that we can define $E(X|Y)$ abstractly via definition (1.4.1). Suppose $X, Y$ are both discrete. Then we also have alternative definitions of $E(X|Y = y)$ via classical probability theory. How are these two connected?

Note that $E(X|Y)$ by definition if a function of $Y$. Thus we can write $E(X|Y) = g(Y)$ for some function $g$. On the other hand, $E(X|Y = y)$ is also clearly a function of $y$. So you can expect that $\forall y$ on the event $\{Y = y\}$

$$E(X|Y) = E(X|Y = y).$$

That is

$$E(X|Y)\mathbf{1}_{Y=y} = E(X|Y = y)\mathbf{1}_{Y=y}.$$

*Proof.* We need to check that for any function $g(Y)$

$$E\Big[E(X|Y = y)\mathbf{1}_{Y=y}g(Y)\Big] = E\Big[X\mathbf{1}_{Y=y}g(Y)\Big].$$

The LHS is equal to

$$\begin{aligned} E(X|Y = y)g(y)P(Y = y) &= \sum_i \frac{iP(X = i, Y = y)}{P(Y = y)} g(y)P(Y = y) \\ &= \sum_i iP(X = i, Y = y)g(y) \\ &= E\Big[Xg(y)\mathbf{1}_{Y=y}\Big] = RHS. \end{aligned}$$

### 1.5.2 Continuous RVs

What about the case when both $X, Y$ are continuous? Here we cannot use the above criterion, as the event $\{Y = y\}$ has probability 0. Rather we will turn it around, and observe that since $E(X|Y = y)$ is a function of $y$, we can also write $E(X|Y = y) = g(y)$. So now we can plug the RV $Y$ into the function $g$ and we claim

$$E(X|Y) = g(Y), P \text{ a.s.},$$

where the a.s. notation means the equality holds outside an event of probability 0 with respect to $P$.

*Proof.*

$$E(X|Y = y) = \int_{-\infty}^{\infty} x \frac{f_{XY}(x, y)}{f_Y(y)} dx.$$

Therefore

$$g(Y) = \int_{-\infty}^{\infty} x \frac{f_{XY}(x, Y)}{f_Y(Y)} dx$$

We need to check that for any function $h(Y)$

$$E\Big(h(Y) \int_{-\infty}^{\infty} x \frac{f_{XY}(x, Y)}{f_Y(Y)} dx\Big) = E\Big(Xh(Y)\Big).$$

But the LHS is equal to

$$\begin{aligned} & \int_{-\infty}^{\infty} h(y) \Big[ \int_{-\infty}^{\infty} x \frac{f_{XY}(x, y)}{f_Y(y)} dx \Big] f_Y(y) dy \\ = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(y) x f_{XY}(x, y) dx dy = E(Xh(Y)) = RHS. \end{aligned}$$

## 1.6 Law of large number

### 1.6.1 The theorem

**Theorem 1.6.1.** *Let $X_1, X_2, \cdots$ be a sequence of independent identically distributed (abbreviated as i.i.d.) RVs such that $E|X_1| < \infty$. Then with probability 1,*

$$\frac{\sum_{i=1}^{n} X_k}{n} \to E(X_1).$$

Notation: It is usually denoted that $S_n = \sum_{i=1}^{n} X_k$, thus one usually sees $\frac{S_n}{n} \to E(X_1)$ in the statement of the law of large number (LLN).

Interpretation: Suppose that you play a game where your winning is random, which is represented by a RV $X$. As you play this game many times, you will find that your average earning (over time) is approximately the expected value of $X$.

### 1.6.2 Application

We'll give one application of the LLN, in pricing of a random game: Suppose there is a game of tossing a fair coin, where if the coin turns up $H$ then you get paid 3 dollars. If it turns up $T$, then you get paid 1 dollars. Question: What is the fair price to charge for this game?

Ans: The fair price to charge for this game is $3\frac{1}{2} + 1\frac{1}{2} = 2$ (dollars). But can you explain why this is the fair price? The reason is the LLN. If this game is played *only once* (an important point, which we'll come back later when we discuss the fair price of financial instrument) then it is not clear that the price is fair. However, the assumption here is that the game will be played *many times,* by potentially many different players. Thus each player's winning is an independent, identically distributed random variable, which takes values 3 and 1 with probability 1/2 each. The total amount of money the house has to pay to these players, after $n$ games have been played, is $\sum_{i=1}^{n} X_i$. By the LLN, this is approximately $nE(X_1)$, which is $2n$, which is the total amount charged by the house. So the house comes out even and this is a fair price for the game.

Remark: This is the main principle behind casino's operation (and profitability). Of course the players are not charged to play the games in the casino. But the game is set up so that the expectation is negative (even if you bet on a roulette table, say on an even number, your chance of winning is still less than 1/2, since there is a 0 and double 0's). Thus by the LLN, with a lot of customers, the casino will have a positive profit. Note that the LLN does allow for an occasional incident where someone plays 1 single game and win big. But if you play a lot of games at the casino, the LLN says that you will lose money eventually.

## 1.7 Central limit theorem

### 1.7.1 The theorem

The LLN gives us an estimate of $\frac{S_n}{n}$ (it is approximately $E(X_1)$). However, for various reasons, we may want a more precise estimate than that. Note that the LLN says nothing about how close to $E(X_1)$ $\frac{S_n}{n}$ is, or (perhaps surprisingly) what distribution we may approximate $\frac{S_n}{n}$ with. It is surprising because we do not have any restriction on the distribution of each individual $X_i$, but it turns out that the approximate distribution of $\frac{S_n}{n}$ is the normal distribution. The precise statement is as followed:

**Theorem 1.7.1.** *Let $X_1, X_2, \cdots$ be i.i.d. RV such that $E|X_1|^2 < \infty$. We will also denote $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$. Then for any real number $x$,*

$$P(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x) \to P(Z \leq x),$$

*where $Z$ has standard normal ($N(0,1)$) distribution.*

### 1.7.2 Application

The central limit theorem is used to estimate probability of the *sum* or the *average* of an i.i.d. sequence of RVs. Determining which case to use requires a close reading into the problem.

**Example 1.7.2.** *The bus arrives at the Hill center according to a uniform* $[0, 12]$ *distribution. Suppose you wait at the Hill center bus stop for 30 days. What is the approximate probability that your* **average** *wait time is more than 5 minutes?*

*Ans: Let* $X_1, X_2, \cdots$ *be i.i.d.* $U[0, 12]$*. Then* $E(X_1) = 6$ *and* $Var(X_1) = 12$*. Thus*

$$P(\frac{S_{30}}{30} \geq 5) = P(\frac{S_{30} - 6 \times 30}{\sqrt{12 \times 30}} \geq \frac{5\sqrt{30}}{\sqrt{12}} - \frac{6\sqrt{30}}{\sqrt{12}}) \approx P(Z \geq -1.58).$$

**Example 1.7.3.** *The earning per day of a casino is distributed as an Exponential(1) RV. (1 here stands for 1 million, we omit the unit). What is the approximate probability that the casino's earning in 1 month is more than 35 millions?*

*Ans: Note that here we're asked for the* **total** *earning. Thus let* $X_1, X_2, \cdots$ *be i.i.d.* $Exp(1)$*. Then* $E(X_1) = 1$ *and* $Var(X_1) = 1$*. Thus*

$$P(S_{30} \geq 35) = P(\frac{S_{30} - 30}{\sqrt{30}} \geq \frac{5}{\sqrt{30}}) \approx P(Z \geq \frac{5}{\sqrt{30}}).$$

## 1.8 Appendix - On regular conditional probability distribution

(This section is not required for the course. You can skip this.)

In the above discussion, you see that we can define, for $X, Y$ jointly continuous

$$P(X < a | Y = y) := \int_{-\infty}^{a} f_{X|Y}(x|y) dx,$$

even though $P(Y = y) = 0$. This is connected with the notion of regular conditional probability distribution (r.c.p.d). We will discuss r.c.p.d. and the connection in this section.

### 1.8.1 R.c.p.d

Let $X : (\Omega, \mathcal{F}, P) \to (S, \mathcal{S})$ be a RV taking values in a general space $S$ with sigma-field $\mathcal{S}$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub sigma-field.

For any set $A \in \mathcal{S}$, the conditional probability of $X \in A$ given $\mathcal{G}$ is given by the conditional expectation

$$P(\omega, A) := E(\mathbf{1}_A(X)|\mathcal{G}). \tag{1.3}$$

As we vary $A$ in $\mathcal{S}$, we obtain a map from $\Omega \times \mathcal{S}$ to $[0, 1]$.

The question is: when we consider simultaneuously all $A \in \mathcal{S}$, that is when we fix a particular $\omega$, can we ensure that for $P$ almost every $\omega$, $P(\omega, \cdot)$ is a probability measure (on $\mathcal{S}$)? The answer is non trivial.

First note that for each $A$, $P(\cdot, A)$ is only defined almostsuredly via (1.3). That is, we are free to modify $P(\cdot, A)$ outside a set of probability 0.

Thus, in particular, for any countable collection of disjoint sets $A_n \in \mathcal{S}$, we can define $P(\omega, A_n)$ and $P(\omega, \cup_n A_n)$ such that

$$P(\omega, \cup_n A_n) = \sum_n P(\omega, A_n)$$

holds outside an exceptional set $N^{\{A_n\}}$ of probability 0. The notation $N^{\{A_n\}}$ is used to emphasize the dependence of the exceptional set on the sequence $\{A_n\}$.

We desire a version of $P(\cdot, \cdot)$ (that is a definition of $P(\omega, A)$ for any $A \in \mathcal{S}$ outside a set $N \subseteq \Omega$ of probability 0. This set $N$ is defined **independent of** all the sets A), such that $P(\omega, \cdot)$ satisfies the above countable additivity property for **any** infinite collection of disjoint sets.

Since there are possibly uncountable number of such collections of sets, the corresponding exceptional sets $N^{\{A_n\}}$ of probability 0 could add up to a set with positive probability, or even become non-measurabale.

**Definition 1.8.1.** *Let $(\Omega, \mathcal{F}, P), \mathcal{G}$ and $X : (\Omega, \mathcal{F}, P) \to (S, \mathcal{S})$ be as above. A family of probability distributions on $(S, \mathcal{S})$, denoted by $P(\omega, \cdot)_{\omega \in \Omega}$ is called a r.c.d. of $X$ given $\mathcal{G}$ if for each $A \in \mathcal{S}$, $P(\cdot, A) = E(\mathbf{1}_A(X)|\mathcal{G})$-a.s. . When $(S, \mathcal{S}) = (\Omega, \mathcal{F})$ and $X(\omega) = \omega$, $P(\omega, \cdot), \omega \in \Omega$ is called a r.c.p on $\mathcal{F}$ given $\mathcal{G}$.*

Remark: The above definition captures the two desirable properties of a r.c.p.d that we discussed above: For each $\omega$, $P(\omega, \cdot)$ is a probability on $\mathcal{S}$. Second, it agrees with the conditional expectation given $\mathcal{G}$ almost surely. Note that the family is given *a priori* to the discussion about its agreement with the conditional expectation. In this way it is defined for all $\omega \in O$. The countable additivity property follows from the definition because

$$P(\cdot, A_n) = E(\mathbf{1}_{A_n}(X)|\mathcal{G})\text{-a.s.}$$
$$P(\cdot, \cup_n A_n) = E(\mathbf{1}_{\cup_n A_n}(X)|\mathcal{G})\text{-a.s.}$$

Thus we can choose an exepctional set $N^{\{A_n\}}$ of probability 0 such that outside of $N^{\{A_n\}}$

$$P(\cdot, \cup_n A_n) = E(\mathbf{1}_{\cup_n A_n}(X)|\mathcal{G}) = \sum_n E(\mathbf{1}_{A_n}(X)|\mathcal{G}) = \sum_n P(\cdot, A_n),$$

where the second equality follows from the property that if $X_n \geq 0, X_n \uparrow X$ then $E(X_n|\mathcal{G}) \uparrow E(X|\mathcal{G})$.

If $X$ has a r.c.d given $\mathcal{G}$ then conditional expectations of $X$ given $\mathcal{G}$ can be expressed as integrals over the r.c.d.

**Proposition 1.8.2.** *Let $P(\omega,\cdot)_{\omega\in\Omega}$ be a r.c.d of $X$ given $\mathcal{G}$. Then for any Borel measurable funcion $f:(S,\mathcal{S})\to(\mathbb{R},\mathcal{B})$ with $E(f(X))<\infty$, we have*

$$E(f(X)|\mathcal{G})(\omega)=\int f(x)P(\omega,dx),\ \textit{for } P \textit{ a.e. } \omega.$$

*Examples of r.c.p.d., discrete RV*

Let $X,Y$ have discrete distribution. In this case WLOG we can consider $\Omega=\mathbb{Z}^2$, $X,Y$ are coordinate mappings: for $\omega=(\omega_1,\omega_2)$:

$$\begin{aligned} X(\omega) &= \omega_1 \\ Y(\omega) &= \omega_2. \end{aligned}$$

We let $\mathcal{F}=\mathcal{F}^{X,Y}, \mathcal{G}=\mathcal{F}^{Y}$. We define a probability measure on $\mathcal{F}$ simply by

$$P(\omega)=P(X=\omega_1,Y=\omega_2).$$

(In this way, if $\omega$ is such that $X$ cannot take value $\omega_1$ or $Y$ cannot take value $\omega_2$, we simply assign $P(\omega)=0$).

For any $\omega\in\Omega$ and any $A\in\mathcal{B}(\mathbb{R})$ we define

$$\begin{aligned} P(\omega,A) &:= P(X\in A|\mathcal{G})(\omega):=\sum_{\tilde\omega:\tilde\omega_1\in A,\tilde\omega_2=\omega_2}\frac{P(X=\widetilde{\omega}_1,Y=\omega_2)}{P(Y=\omega_2)} \\ &= \frac{\sum_{\tilde\omega:\tilde\omega_1\in A,\tilde\omega_2=\omega_2}P(X=\widetilde{\omega}_1,Y=\omega_2)}{P(Y=\omega_2)} \\ &= \frac{\sum_{\tilde\omega:\tilde\omega_1\in A,\tilde\omega_2=\omega_2}P(\tilde\omega)}{\sum_{\tilde\omega:\tilde\omega_2=\omega_2}P(\tilde\omega)} \\ &= P(X\in A|Y=\omega_2). \end{aligned}$$

We claim that $P(\omega,A)$ is the r.c.d. of $X$ given $\mathcal{G}$.

First, for any fixed $\omega$, it is clear that $P(\omega,A)$ is a probability measure. To check that $P(\omega,A)=E(\mathbf{1}_A(X)|\mathcal{G})(\omega)$, for every $\omega$, we need to check that for any function $g(Y)$;

$$\int P(\omega,A)g(Y)(\omega)dP(\omega)=\int \mathbf{1}_A(X)(\omega)g(Y)(\omega)dP(\omega).$$

The LHS is equal to

$$\sum_{\omega}\frac{\sum_{\tilde\omega:\tilde\omega_1\in A,\tilde\omega_2=\omega_2}P(\tilde\omega)}{\sum_{\tilde\omega:\tilde\omega_2=\omega_2}P(\tilde\omega)}(\omega)g(Y)(\omega)P(\omega).$$

Observe that for any $\omega$, $\frac{\sum_{\tilde\omega:\tilde\omega_1\in A,\tilde\omega_2=\omega_2}P(\tilde\omega)}{\sum_{\tilde\omega:\tilde\omega_2=\omega_2}P(\tilde\omega)}(\omega)$ and $g(Y)(\omega)$ only depend on $\omega_2$. That is, if $\omega$ and $\omega'$ are such that $\omega_2=\omega_2'$ then

$$\frac{\sum_{\tilde\omega:\tilde\omega_1\in A,\tilde\omega_2=\omega_2}P(\tilde\omega)}{\sum_{\tilde\omega:\tilde\omega_2=\omega_2}P(\tilde\omega)}(\omega)=\frac{\sum_{\tilde\omega:\tilde\omega_1\in A,\tilde\omega_2=\omega_2}P(\tilde\omega)}{\sum_{\tilde\omega:\tilde\omega_2=\omega_2}P(\tilde\omega)}(\omega')$$

and $g(Y)(\omega) = g(Y)(\omega')$. Therefore the LHS is actually equal to

$$\begin{aligned}
& \sum_{\omega} \left( \frac{\sum_{\tilde{\omega}:\tilde{\omega}_1 \in A, \tilde{\omega}_2 = \omega_2} P(\tilde{\omega})}{\sum_{\tilde{\omega}:\tilde{\omega}_2 = \omega_2} P(\tilde{\omega})} (\omega) g(Y)(\omega) \right) \sum_{\tilde{\omega}:\tilde{\omega}_2 = \omega_2} P(\tilde{\omega}) \\
= & \sum_{\omega} \left( \sum_{\tilde{\omega}:\tilde{\omega}_1 \in A, \tilde{\omega}_2 = \omega_2} P(\tilde{\omega}) \right) (\omega) g(Y)(\omega) \\
= & \int \mathbf{1}_A(X)(\omega) g(Y)(\omega) dP(\omega).
\end{aligned}$$

*Examples of r.c.p.d., continuous RV*

Let $X, Y$ have continuous distribution. In this case we consider $\Omega = \mathbb{R}^2$, $X, Y$ are coordinate mappings: for $\omega = (\omega_1, \omega_2)$:

$$\begin{aligned}
X(\omega) &= \omega_1 \\
Y(\omega) &= \omega_2.
\end{aligned}$$

We let $\mathcal{F} = \mathcal{F}^{X,Y}, \mathcal{G} = \mathcal{F}^Y$. For a set $E \in \mathcal{F}$ we define

$$P(E) = \int_E f_{XY}(x, y) dx dy.$$

(Again, in this way if there is a set $E \in \mathcal{B}(\mathbb{R}^2)$ such that $X, Y$ cannot take values in $E$, we simply assign $P(E) = 0$).

For any $\omega \in \Omega$ and any $A \in \mathcal{B}(\mathbb{R})$ we define

$$\begin{aligned}
P(\omega, A) &:= P(X \in A | \mathcal{G})(\omega) := \frac{\int_A f_{XY}(x, \omega_2) dx}{f_Y(\omega_2)} \\
&= \frac{\int_A f_{XY}(x, \omega_2) dx}{\int_{-\infty}^{\infty} f_{XY}(x, \omega_2) dx} = P(X \in A | Y = \omega_2).
\end{aligned}$$

We claim that $P(\omega, A)$ is the r.c.d. of $X$ given $\mathcal{G}$.

First, for any fixed $\omega$, it is clear that $P(\omega, A)$ is a probability measure. To check that $P(\omega, A) = E(\mathbf{1}_A(X)|\mathcal{G})(\omega)$, for every $\omega$, we need to check that for any function $g(Y)$;

$$\int P(\omega, A) g(Y)(\omega) dP(\omega) = \int \mathbf{1}_A(X)(\omega) g(Y)(\omega) dP(\omega).$$

The LHS is equal to (in the following $\omega = (x, y) = (\omega_1, \omega_2)$)

$$\int_{\mathbb{R}^2} P(\omega, A) g(Y)(\omega) f_{XY}(\omega) dx dy = \int_{\mathbb{R}^2} \frac{\int_A f_{XY}(x, \omega_2) dx}{\int_{-\infty}^{\infty} f_{XY}(x, \omega_2) dx} (\omega) g(Y)(\omega) f_{XY}(\omega) dx dy.$$

Again note that both $\frac{\int_A f_{XY}(x,\omega_2)dx}{\int_{-\infty}^{\infty} f_{XY}(x,\omega_2)dx}(\omega)$ and $g(Y)(\omega)$ only depend on $\omega_2$. Thus the LHS is equal to

$$\begin{aligned}
& \int_{\mathbb{R}} \left( \frac{\int_A f_{XY}(x,\omega_2)dx}{\int_{-\infty}^{\infty} f_{XY}(x,\omega_2)dx}(\omega) g(Y)(\omega) \right) \left( \int_{\mathbb{R}} f_{XY}(\omega)dx \right) dy \\
= \; & \int_{\mathbb{R}} \left( \int_A f_{XY}(x,\omega_2)dx \right) (\omega) g(Y)(\omega) dy \\
= \; & \int_{\mathbb{R}} \int_A f_{XY}(x,y) g(Y) dx dy = \int \mathbf{1}_A(X)(\omega) g(Y)(\omega) dP(\omega).
\end{aligned}$$

*Existence of r.cp.d.*

It is clear from our above discussion that we do not always have a family of r.c.p.d. The following gives a sufficient condition for the existence of r.c.p.d.

**Theorem 1.8.3.** *Let $X : (\Omega, \mathcal{F}, P) \to (S, \mathcal{S})$ be a RV taking values in a general space $S$ with sigma-field $\mathcal{S}$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub sigma-field. If $S$ is a* **complete separable metric space** *with Borel sigma field $\mathcal{S}$ then there exists a r.c.p.d. family $P(\omega, \cdot)_{\omega \in \Omega}$ of $X$ given $\mathcal{G}$.*

Remark: From our examples, in the discrete case it is clear that the r.c.p.d. always exists. The key there is because $X, Y$ takes values on a space with *countably many* values, so that we can re-cast the probability space as $\mathbb{Z}^2$. This demonstrates the condition requiring $S$ being a separable metric space

In the jointly continuous case the r.c.p.d also exists, but it is not clear how the separability comes into play, as we used the density to define the r.c.p.d. there. One can suspect that the existence of the density $f_{XY}$, or the Radon-Nikodym derivative of the distribution of $X, Y$ with respect to the Lebesgue measure on $\mathbb{R}^2$, has to do with the separability of $\mathbb{R}^2$.

## 1.9 Exercises

1. Our friend John tells we that he has two daughters. we also know that he has three children in total. What is the probability that his youngest child is a girl (assuming that a boy and a girl are equally likely)?

2. The arrival time of shuttles at a bus stop from 5:00pm to 5:30pm on a weekday is uniformly distributed. In other words, let $X$ be the waiting time (in minutes) after 5:00 pm until the arrival of a particular shuttle, then $X$ has the p.d.f

$$f_X(x) = \frac{1}{30}, 0 \leq x \leq 30. \tag{1.4}$$

Suppose there are 30 shuttles arriving at the bus stop from 5:00 pm to 5:30 pm, and their distribution are i.i.d. What, then, is the approximate probability that their average arrival time on a particular weekday is before 5:12 pm?

3. Studying probability and statistics has a positive effect on students' job placement. A student who has succesfully completed these classes has a $70\%$ probabilty of landing a job with Goldman Sachs. A student who did not succesfully complete the program, however, only has a $40\%$ probability of landing such a job. our friend, Tom, just got a position with Goldman Sach. Suppose the probability of a student succesfully completing the Financial Math program at Rutgers is $80\%$. What is the probability that Tom succesfully finished his Financial Math program at Rutgers?

4. The breakdown time of the Apple Ipad is an exponential(5) random variable. In other words, let $X$ be the time until break down (in years) of a particular Ipad, then $X$ has the p.d.f

$$f_X(x) = \frac{1}{5}e^{-\frac{1}{5}x}, 0 \leq x < \infty. \tag{1.5}$$

Suppose an Apple store has 30 Ipads, and the distribution of their break down times are i.i.d. What, then, is the approximate probability that the average break down time of these Ipads is before 2 years?

5. There are 120 students in the Introduction to Probability class. Suppose that each student has a probability of .3 of getting an A in this class, and the students' performance is independent of one another. What, then, is the approximate probability that the class will have at least 20 students getting an A ?

6. In this problem we will verify that the conditional expectation $E(X|Y)$ is the best guess of $X$ given $Y$ in the following sense

$$E\big[(X - E(X|Y))^2\big] \leq E\big[(X - g(Y))^2\big], \text{ for all } g(Y). \tag{1.6}$$

a. Show that

$$E\big[E(X|Y)X\big] = E\big[E(X|Y)^2\big].$$

Hint:

$$E\big[E(X|Y)X\big] = E\Big[E\Big(E(X|Y)X\Big|Y\Big)\Big].$$

Proceed using properties of conditional expectation.
b. Use the result of part a to prove (1.6).

7. Let $X, Y$ be independent random variables, $Y$ having Normal(0,1) distribution and $X$ has distribution $P(X = 1.5) = P(X = 0.5) = 1/2$. Let $Z = XY$. Compute
a. $E(Z|Y)$.
b. $E(Z^2|Y)$.
8.
a. Let $X$ have distribution Uniform[0,$Y$] distribution, where $Y$ has Exp(1) distribution. Compute the joint distribution of $X, Y$ and $E(X)$.
b.Let $X$ have distribution Exponential($Y$) distribution where $Y$ has Uniform[1,2] distribution. Compute the joint distribution of $X, Y$ and $E(X)$.
9. Prove the inequalities in (1.4.2).
10. Prove equations (1.1) and (1.2).

# CHAPTER 2 Derivative pricing in one period model

In this Chapter and the following ones, we apply the probabability theory described in Chapter (1) to study the problem of pricing a financial derivative. A financial derivative is simply a contract between two parties which specifies agreements on the buying and / or selling of a certain commodity or security. We refer to commodity or the security on which the contract is based as the underlying. Since the contract derives its value from the underlying, we call such contract the derivatives. We start with a brief description of the various derivatives that will be studied in this book. The central question for this chapter is *the fair price* of contracts such as the one described below. More importantly, we want to clarify the sense in which we call it a fair price.

## 2.1 The financial derivatives

### 2.1.1 Forward contracts

Suppose we run a factory, and we know that in November we will need a large amount of oil (say 10,000 barrels). Suppose the price of oil now is 100 dollars per barrel; but for certain reason we do not want to purchase 10,000 barrels right now. This could be because we did not have the cash available; or the cost of inventory will be very high from now until November when we actually need the oil. But waiting until November to purchase is also risky, since the price may jump to 115 dollars per barrel. So what we do is entering into a contract to lock down a price for the future purchase of oil at say, 105 dollars per barrel. In this way, we have entered into a *forward contract* with *expiration date* on Nov 1, for a barrel of oil with *strike price* 105 dollars. For simplicity we suppose each contract is just for 1 barrel. To lock in the price for 10,000 barrels we simply buy 10,000 contracts.

### 2.1.2 European call option

One feature of the forward contract is that once we enter the contract, we must purchase the product on the expiration date. Suppose in November the oil price drops to 95 dollars per barrel. If we already are in a forward contract with strike equals 105 dollars, we're in an undesirable situation. Another type of financial derivative, which offers a more flexible agreement than the forward contract is the so-called European option. A European call option on a certain underlying with expiration date $T$ and strike price $K$ is a contract that gives we the right, but *not* the obligation, to purchase one unit of the underlying at time $T$ at price $K$.

**Remark 2.1.1.** *An example of a European call option is a feature that is offered in several airlines, where with a small fee we can lock in the price for a ticket for one day, two days or a week. We do not have the obligation to buy the ticket at the locked-in price. Indeed if the price for the ticket falls before the expiration of the fare-lock, we can simply ignore the fare-lock and purchase the ticket at a lower price. This is exactly a call option on the airplane ticket with strike equals the locked-in price and expiration being one day, two days or a week.*

### 2.1.3 European put option

Suppose we are operating a farm that grows corn and the current time is July. We do not harvest our crops until October; but we would want to lock in a price for our crop as an insurance against price drop. If additionally we also want the flexibility to sell if in October the actual price for our crop is higher than the locked-in price, then we want to enter into a European put option. An European put option can be viewed as the exact oppposite of a European call option. That is, a European put option on a certain underlying with expiration date $T$ and strike price $K$, is a contract that gives we the right, but not the obligation to sell one unit of the underlying at time $T$ at price $K$.

### 2.1.4 Other types of derivatives

There are many other derivatives besides forward contract, European call and put option. A few examples include the American, look back, Asian, Bermudan, Barrier options. We will discuss these later in Chapter (3).

## 2.2 Different approaches to pricing

When we have a product (be it a tangible product, like a laptop, or a service, like a legal consulting, or a random game, like a casino game), we would like to know what its fair price is. Here are some possible approaches we may take:

By cost of components

We may decide how much the components of our product cost, and the price is simply the sum of these costs (our labor or however much we contribute into the making of the product is also considered into these cost factors). This approach only works, obviously, when our product can be decomposed into components, and each has a cost.

By supply and demand

An alternative approach to find the fair price is by supply or demand. Another way to say it is let the market decides what the price is. This answer has limitations of course. First, no one knows for sure how the supply and demand curves look like. So to know the price precisely, we would need to model these curves, a non-simple task. Second, letting the market decides the price works for a product already on the market. If we have a completely new product (say we just came up with a new financial derivative), looking into the market for a price is not helpful.

By the Law of Large Number

This approach works for a random game, as mentioned in the previous lecture. The basis for it is that we expect *a large number of* people to play the game, thus our revenue from selling the tickets to the game will balance out the profits and losses we make from each individual game. Note that there have to be *two* factors to make this approach works: a lot of incidents of the product (the game) and they are *independent, identically* distributed. Without either, pricing by the Law of Large Number will not work.

**Remark 2.2.1.** *We may refer to this approach as pricing by expectation. Here we refer to it as pricing by the Law of Large Number to avoid confusion, since later on we will see another approach of pricing by expectation, which is completely different from this one and not based on the Law of Large Number.*

### 2.2.1 Pricing of financial derivatives

We observe that none of the approach above applies to the pricing of the financial derivatives we mentioned in Section 2.1. Actually, the reason why the the Law of Large Number approach does not work is a bit subtle. Even if we have a probabilistic model for the underlying, the Law of Large Number is still not the right way to price. A more detailed explanation will be given when we come to the one period model.

We make the following observation: in any financial derivative, there is a transfer of risk from the buyer to the seller of the contract. By selling us a financial derivative, the contract-seller assumes the risk of facing the ups an downs of the market price of the underlying. Therefore, the fair price of the financial derivative must properly reflect this risk that the seller takes. This philosophy will lead to the pricing by portfolio replication approach, discussed in Section 2.3.

On the other hand, a financial derivative is traded on the market. One can observe as a fact that the price of the underlying is strongly correlated with that of the derivative. Since we can buy and sell a derivative together with the underlying, a mis-price of the derivative might lead to some risk-free profit. We require then the price of the derivative is such that this risk-free profit cannot happen. This philosophy will lead to the pricing by no-arbitrage principle approach, discussed in Section 2.4.

## 2.3 Pricing by portfolio replication

### 2.3.1 Hedging portfolio

Let us revisit the example of the forward contract in section (2.1.1) from the point of view of the contract seller. How much would we charge for such a contract? A better question is:

what would we do with the money we obtained from selling the contract, keeping in mind that we have the obligation to sell 1 barrel of oil at 105 USD in November? Clearly we would like to fulfill the contract at the expiration time without any additional money out of our pocket. Therefore, we should invest the money received from the contract so that our position is covered. Since the oil price and the forward contract based on oil are positively correlated, we should invest some money into the money market (i.e. a saving account with interest) and some money into the oil market. If our investment strategy is right, the value of our portfolio would be equal to the payout we have to make independent of the outcome of the underlying price. Thus we have completely hedged our risk. We call the portfolio in this case a hedging portfolio or a replicating portfolio, from the fact that our portfolio 'replicates the value of the financial derivative at the expiration time.

One implicit assumption we made in the above scenario is that our portfolio is self-financing. That is we can re-adjust our position in between the sale time and the expiration time, but we cannot do this using additional funding from outside, nor can we withdraw money from the portfolio. Clearly if the portfolio is not self-financing, then its risk-hedging property becomes meaningless. This suggests that the fair price of a financial derivative is how much it takes to set up the hedging portfolio at the initial time.

**Remark 2.3.1.** *The existence of a hedging portfolio is not guaranteed. In the case of the non-existence of a hedging portfolio, we will have to find a different way to define a fair price, see Section* (2.4). *But if there is a hedging portfolio, then its initial value must be the price we charge for the contract. Otherwise there will be an arbitrage opportunity, which we will discuss next.*

### 2.3.2 Arbitrage opportunity

An arbitrage opportunity is an opportunity to earn profit without risk. This is clearly an undesirable situation for the economy. One of the central principles in derivative pricing is that arbitrage opportunity does not exist. We call this the no-arbitrage principle. Our first application of the no-arbitrage principle is to show the equality between the price of a derivative and its hedging portfolio.

**Lemma 2.3.2.** *Suppose the no-arbitrage principle holds. If there exists a hedging portfolio for a financial derivative then the portfolio's initial value must be the price of the derivative.*

*Proof.* Suppose the price $x$ for the derivative is higher than the value $y$ of the hedging portfolio. At time $0$, we sell the contract for $x$ and use $y$ to finance our portfolio. Thus we gain $x - y > 0$ dollars at the beginning, which we can put into a saving account. At the expiration time, because the portfolio is replicating, our position is completely cover. Thus we have made a riskless profit; the hedging portfolio gives us an arbitrage opportunity. A similar argument holds when $x < y$.

### 2.3.3 Price of a forward contract

We will leave the example of the oil forward contract in Section (2.1.1) and work with an abstract setting. Suppose we have a forward contract with expiration $T$ with strike $K$ on an underlying $S$. That is at time $T$, we can obtain one share of $S$ for $K$ dollars. We will denote the price of $S$ at time $t$ as $S_t$. Thus the value of this forward contract at time $T$ is $S_T - K$. Suppose also that the interest rate is $r$. What is the price for the contract at the current time $t = 0$?

**Lemma 2.3.3.** *The price for a forward contract with strike $K$ and expiration $T$ on an underlying $S$ at time $t = 0$ is $S_0 - Ke^{-rT}$.*

*Proof.* At time 0, we purchase 1 share of $S$ and borrow $Ke^{-rT}$ from the bank. Since we receive $S_0 - Ke^{-rT}$ dollars from selling the forward contract, our initial position is completely balanced. At time $T$, one share of $S$ is worth $S_T$ and we owe the bank $Ke^{-rT}e^{rT}$ due to accumulating interest. Thus the value of our portfolio is $S_T - Ke^{-rT}e^{rT} = S_T - K$, which is exactly the value of the forward contract at time $T$. We have constructed a hedging portfolio whose initial value is $S_0 - Ke^{-rT}$. By the discussion above, the price for the forward contract is $S_0 - Ke^{-rT}$.

### 2.3.4 Finding the hedging portfolio for a forward contract

The price $S_0 - Ke^{-rT}$ gives us an idea of how to construct a hedging portfolio. But if we do not know this price, how can we proceed? We will just assume that our portfolio has $x$ shares of $S$ and $y$ dollars in the money market at the beginning. If we can construct $x$ and $y$ so that at the expiration we have

$$xS_T + ye^{rT} = S_T - K, \tag{2.1}$$

then by the no arbitrage principle the price of the forward contract would be $xS_0 + y$.

Equation (2.1) contains two unknowns $x, y$. The key to solving it is to note that this equation has to hold for all outcomes of $S_T$. Also observe that in this equation only $S_T$ is a random variable, while the rest of $x, y, K, e^{rT}$ are constants. We rewrite the equation as

$$(x - 1)S_T = -ye^{rT} - K.$$

The left hand side is a random variable, the right hand side is a constant. They can only be equal if the left hand side is also a constant, from which we conclude $x = 1$. It follows that $y = -Ke^{-rT}$.

**Remark 2.3.4.** *Observe that the price of a forward contract we found is model independent. That is we did not make any assumption about the distribution of $S_T$ or the behavior of $S_t$ from time 0 to $T$. This is due to the fact that the value of the forward contract $S_T - K$ is a linear function of $S_T$. A similar approach will not hold when the derivative value is no longer a linear function of the underlying asset price. For example, we will need to build models for $S_T$ to find the price of the European options.*

### 2.3.5 Forward price

For the forward contract with strike $K$, there is a special value of $K_0$ that makes the value of the contract at the initial time $t = 0$ equal to zero. That is the buyer does not have to pay anything to enter the contract. Note that this does not imply the contract value remains zero after time $t = 0$. In fact at time $T$ its value is still $S_T - K$. Because $S_T$ is random, it cannot be the case that $S_T - K = 0$ with probability 1.

From the above discussion, we can easily see that this value of $K_0$ is $S_0 e^{rT}$. This is called the forward price of the underlying $S$ (for expiration time $T$). In words, the forward price (of an asset with expiration $T$) is the strike price of a forward contract on the same asset with expiration $T$ such that the contract costs no money to enter. We should distinguish this concept from the price of a forward contract, which is typically not zero at $t = 0$.

## 2.4 Pricing by the no-arbitrage principle

We discussed in a preliminary way the no-arbitrage principle in Section (2.3.2). We concluded that if there is a hedging portfolio the price of the financial derivative must be the same as the value of the portfolio. In many scenarios, a hedging portfolio does not exist. One common reason is because there are more random sources than the number of financial asset so in a sense we cannot hedge away all randomness.

However, we can still discuss about the fair price of a financial derivative in this case. Assuming the derivative $V$ is traded in the market, one can form a portfolio consisting of $V$, the underlying $S$ and the money market account that earns interest $r$. If the price of $V$ is such that that it allows for an arbitrage-portfolio, then it cannot be the right price.

So we can simply define the price of $V$ is such that no matter how one combines $V, S$ and the money market account, no arbitrage-portfolio can arise. This turns out to be an acceptable concept to arrive at a price for $V$. We refer to it as the pricing by no-arbitrage principle.

**Remark 2.4.1.** *Pricing by the no-arbitrage principle does not seem to lead to a concrete price. As we shall see, no-arbitrage price is derived via the risk-neutral pricing technique. Another observation to keep in mind is the pricing by no-arbitrage principle also allows for many possible prices of $V$. In this case, the market will ultimately decide which price to pick for the derivative.*

## 2.5 Introduction to the one period model

### 2.5.1 Value of a European option

Consider a European call option on an asset $S$ with expiration $T$ and strike price $K$. From now on, we will denote the value of a financial product at time $t$ as $V_t$. We first discuss the mathematical expression of $V_T$.

Clearly at time $T$ two things can happens: Either $S_T > K$ or $S_T \leq K$. If $S_T > K$, the holder would exercise the option to buy 1 share of $S$ at price $K$. Since the asset is actually worth $S_T$, he has made a gain of $S_T - K > 0$. If $S_T \leq K$ he simply does not exercise the option. In that case $V_T = 0$. Thus we have

$$V_T = \max(S_T - K, 0).$$

We introduce some notation to re-write $\max(S_T - K, 0)$. For a real number $x$, we denote $\max(x, 0)$ as $x^+$ and $\max(-x, 0)$ as $x^-$. For example, $5^+ = 5, (-5)^+ = 0, (-5)^- = 5, 5^- = 0$. With this notation, we see that the value of a European call option at time $T$ is $V_T = (S_T - K)^+$.

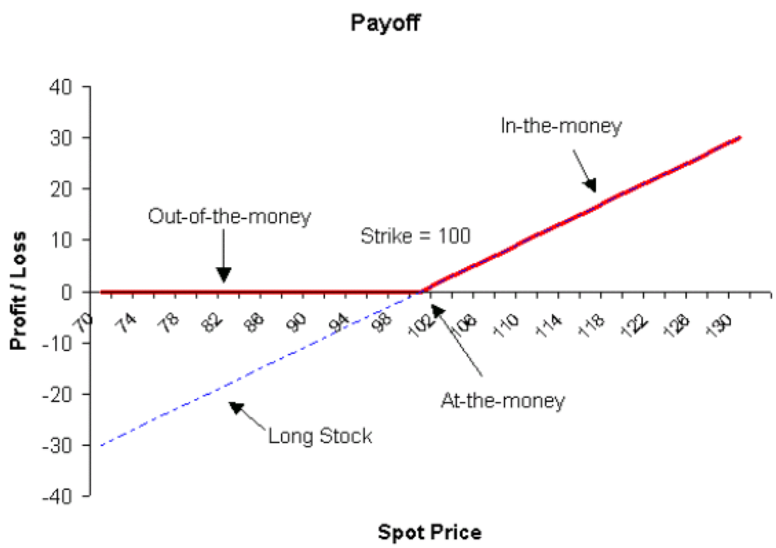



Figure 2.1: Payoff of a call-option

Recall that a European put option gives the owner the right, but not the obligation to sell one share of the underlying $S$ at price $K$ at time $T$. Thus the value of a corresponding European put option is $V_T^{put} = (S_T - K)^- = (K - S_T)^+$.

### 2.5.2 An attempt to construct a hedging portfolio for European options

Suppose we sell one share of a European call option on $S$ with strike $K$ and expiration $T$. To hedge our risk, we invest in the underlying $S$ and the money market . Specifically suppose at time $t = 0$ we buy $x$ shares of $S$ and hold $y$ dollars in the money market. Can we construct a portfolio that replicates the value of a European call option at time $T$? The equation that the portfolio value has to satisfy is

$$xS_T + ye^{rT} = (S_T - K)^+; \tag{2.2}$$

and it has to hold true no matter what value $S_T$ takes. However here we cannot repeat the approach in Section (2.3.3) to solve for $x$ and $y$ because the function $(S_T - K)^+$ is not linear in $S_T$.

The system (2.2) is in general an over-determined system for $x$ and $y$. If there are $n$ different possible outcomes or $S_T$ then the system (2.2) is a linear system (in $x, y$) with two unknowns and $n$ equations. Thus to have a hope to solve for such as system we need to limit ourselves to the case $n = 2$. That is we posit that $S_T$ has only two possible outcomes: $u$ for up and $d$ for down. This is the one period binomial model that we discuss in the next Section.

## 2.6 The one period binomial model

### 2.6.1 Model specification

The one period binomial model is a model about the underlying asset $S$. We posit the followings: there are 2 stages of action, the inital time $t = 0$ and the expiration time $t = T$. We purchase (or sell) the option at $t = 0$ and exercise the option (or fulfill the option's terms) at $t = T$. At any other time $0 < t < T$ we do not take any action to buy or sell any financial asset.

The value of the asset at $t = 0$ is a constant $S_0$. Its value at $t = T$ is a random variable $S_T$ that can take on two values: $S_T = uS_0$ or $S_T = dS_0$ where $u$ and $d$ are positive real numbers. We will also assume the interest rate is $r > 0$. We think of $d, u, r$ as parameters of the model that we can estimate and plug into the model using information about the asset and the interest rate.

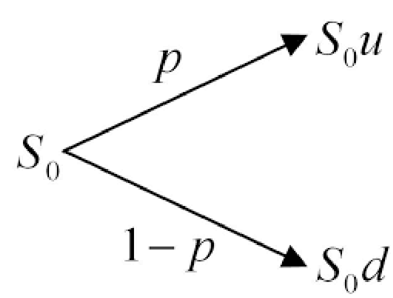



Figure 2.2: One period binomial model

### 2.6.2 Restriction on $d, u, r$

Whenever we build a model, we should check whether our model violates any fundamental principle. In the one period model we need to have

$$d < e^{rT} < u,$$

otherwise there is an arbitrage opportunity. Indeed, suppose $e^{rT} \leq d < u$. This implies that the underlying $S$ is guaranteed to have a higher return than the money market account. If this is the case, we can borrow $S_0$ from the money market and purchase one share of $S$. At $t = T$ we will be guaranteed a non-negative profit. A similar argument holds when $u \leq e^{rT}$.

### 2.6.3 Pricing of a financial product using the hedging portfolio

Consider a financial derivative $V$ based on an underlying $S$. At the expiration time $T$, if the asset goes up ($S_T = uS_0$) then $V_T = V_u$ and if the asset goes down ($S_T = dS_0$) $V_T = V_d$. Here $V_u$ and $V_d$ are 2 constants that depends on the specific derivative in consideration.

**Example 2.6.1.** *Let $V^{call}$ (respectively $V^{put}, V^{forward}$) be the value of the European call option (respectively European put option, forward contract) based on $S$ with strike $K$ and expiration $T$. Then*

$$\begin{aligned} V_u^{call} &= (uS_0 - K)^+, V_d^{put} = (dS_0 - K)^+; \\ V_u^{put} &= (K - uS_0)^+, V_d^{put} = (K - dS_0)^+; \\ V_u^{forward} &= uS_0 - K, V_d^{forward} = dS_0 - K. \end{aligned}$$

Note that in general we do not have $V_d < V_u$ even if $dS_0 < uS_0$ by construction. For example if $dS_0 < K < uS_0$ then $V_u^{put} = 0$ and $V_d^{put} = K - dS_0 > 0$.

We construct a replicating portfolio for $V$ that consists of $x$ shares of $S$ and $y$ dollars in the money market at $t = 0$. The replication condition requires:

$$\begin{aligned} xuS_0 + ye^{rT} &= V_u \\ xdS_0 + ye^{rT} &= V_d. \end{aligned}$$

The system can be written in a matrix form as

$$\begin{bmatrix} uS_0 & e^{rT} \\ dS_0 & e^{rT} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} V_u \\ V_d \end{bmatrix}. \tag{2.3}$$

The inverse of the matrix $\begin{bmatrix} uS_0 & e^{rT} \\ dS_0 & e^{rT} \end{bmatrix}$ is $\frac{1}{e^{rT}S_0(u-d)} \begin{bmatrix} e^{rT} & -e^{rT} \\ -dS_0 & uS_0 \end{bmatrix}$. Note that $\frac{1}{e^{rT}S_0(u-d)} \neq 0$ because of our requirement that $d < u$.

The solution for the system (2.3) is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{e^{rT}S_0(u-d)} \begin{bmatrix} e^{rT} & -e^{rT} \\ -dS_0 & uS_0 \end{bmatrix} \begin{bmatrix} V_u \\ V_d \end{bmatrix}.$$

Then easily we see that $x = \frac{V_u - V_d}{S_0(u-d)}$ and $y = \left[V_u - \frac{V_u - V_d}{u-d}u\right]e^{-rT}$. Thus the price of the financial derivative $V$ is

$$V_0 = xS_0 + y = \frac{V_u - V_d}{u-d} + \left[V_u - \frac{V_u - V_d}{u-d}u\right]e^{-rT}. \tag{2.4}$$

### 2.6.4 Some remarks

1. The formula (2.4) gives the price for any financial derivative based on $S$ in the one period model. The only requirement is the correlation $V_T = V_u$ when $S_T = uS_0$ and $V_T = V_d$ when $S_T = dS_0$. This correlation exactly captures the fact that the value of $V$ is derived from the value of $S$. Without this correlation, the pricing formula (2.4) becomes meaningless.

2. Formula (2.4) relies heavily on the fact that there are only two possible outcomes at $t = T$, namely $S_T = uS_0$ or $S_T = dS_0$. If we were to posit a more general model where there are pore than two possible outcomes then the system (2.3) may not be solvable.

**Example 2.6.2.** *Suppose there are three possible out comes* $\omega_1, \omega_2, \omega_3$ *at* $t = T$*. Let* $S_T(\omega_1) = S_T(\omega_2) = uS_0, S_T(\omega_3) = dS_0$ *but* $V_T(\omega_1) = V_u$ *and* $V_T(\omega_2) = V_T(\omega_3) = V_d$*. The matrix equation* (2.3) *becomes*

$$\begin{bmatrix} uS_0 & e^{rT} \\ uS_0 & e^{rT} \\ dS_0 & e^{rT} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} V_u \\ V_d \\ V_d \end{bmatrix}.$$

*We can easily see that this system has no solution since the first two equations*

$$\begin{aligned} xuS_0 + ye^{rT} &= V_u \\ xuS_0 + ye^{rT} &= V_d \end{aligned}$$

*are in consistent if* $V_u \neq V_d$*.*

3. $S_T$ is a random variable. However we did not require its probability distribution to obtain formula (2.4). What if we do have information that allows us to plug in some probability distribution for $S_T$? For example when $S$ is a stock we can observe the of the past performance of the company to estimate the distribution of $S_T$. Would this information play some role in determining the price? This is an important question which we address in the next section.

## 2.7 Pricing by Expectation

### 2.7.1 An example

Consider a company whose stock price is $S_0 = 90$ today. Suppose tomorrow, with probability $0.9$ the stock price is $S_1 = 100$ and with probability $0.1$ the stock price is $S_1 = 80$. Also suppose that the interest rate $r = 0$. Consider a contract that will deliver a share of $S$ to the holder tomorrow, but the payment has to be made today. This is an example of a forward contract with strike price $0$. What is the price of such a contract?

An intuitive answer is that since $E(S_1) = (.9)100 + (.1)80 = 98$ and $r = 0$, the price for the contract is 98 dollars. However this answer is incorrect. Let us recall that the price of a general forward contract with strike $K$ is $S_0 - Ke^{-rT}$. Since $K = 0$ the price is $S_0 = 90$.

We show in (2.3.3) that $90$ dollars is the no-arbitrage price of the forward contract with $K = 0$. Hence if the contract is priced at $98$ dollars, there will be an arbitrage opportunity. Indeed we would simply sell such a contract at $98$ dollars, and use the money to buy one share of stock at $90$. Tomorrow we give the share of stock to the contract holder to close our position, and make an $8$ dollar profit without risk.

We want to understand why giving the price by $E(S_1)$ is incorrect. Recall from Section (2.2) that pricing by Expectation is an implicit appeal to the Law of Large Number. However, the Law of Large Number does not apply in the current situation. Indeed, observe that in this case we only have one instance of $S_1$. That is there is only one variable $S_1$ that we receive the random payment from. This clearly does not fall into the large number of independent random variables context of the the Law of Large Number. What if we sell the contracts to many buyers to create an instance of a large number of random variables? Unfortunately here the Law of Large Number still does not apply; since all of these buyers will face the same $S_1$. That is every one has to observe the same stock price tomorrow from the same company. We can say there are many identical copies of a random variable; but this does not fulfill the *independent* requirement of the the Law of Large Number.

We remark that there is an approach of pricing financial derivative by taking expectation. However, it is an expectation that is taken under a risk neutral probability, which we have yet defined. We will present this approach in (2.7.3).

### 2.7.2 A discussion of risk versus expectation

In the example of the previous Section, the probability that the stock price goes up is very high at 90 %. So we may wonder if in a realistic situation, a buyer would be willing to pay a little more for than $90$ dollars for $S$. Maybe we would not pay $98$ dollars for the contract; but $90$ dollars seems like a too low value for such a “good" stock.

To understand this question, let us suppose we are offered a game where with 1 % probability we can win 1000 dollars but with 99 % we will lose 9 dollars. We can only play the game once. Would we be willing to play? The expectation of this game is slightly more than 1 dollars. So if we use the expectation as a standard of judgment, then we would definitely play this game. However, in real life many would hesitate to play this game. Those who hesitate can be described as *risk averse.* Since the game can be played only once, there is a very big chance we would lose money. The ones who are willing to play the game can be described as *risk seeking.* The bottom line is the decision to play the game (or to purchase the stock at a price higher than 90 dollars) is not based on expectation. Expectation is not the appropriate standard to make the decision in a game we play only once.

The price of 98 dollars we found for the forward contract in the previous Section is a no-arbitrage price. That is it is a price that allows for no arbitrage opportunity based on the contract. In terms of risk preference, a person who accepts the price of 98 dollars today in exchange for a share of the asset tomorrow is said to be *risk neutral.* He or she is indifferent (or neutral) about receiving a riskless payment today versus a risky payment tomorrow in terms of the asset. This is reflected in his or her acceptance to exchange 98 dollars (the riskless price of the asset today) for a random (risky) payment tomorrow.

### 2.7.3 Pricing by expectation under the risk neutral distribution

Recall the setting in Section (2.6.3): a derivative $V$ based on $S$ that pays $V_u$ when $S_T = uS_0$ and $V_d$ when $S_T = dS_0$. We found the price $V_0$ to be

$$V_0 = \frac{V_u - V_d}{u - d} + \left[V_u - \frac{V_u - V_d}{u - d}u\right]e^{-rT}. \tag{2.5}$$

By collecting the terms that involve $V_u$ and the terms that involve $V_d$, we have

$$V_0 = e^{-rT}\left[V_u\frac{e^{rT} - d}{u - d} + V_d\frac{u - e^{rT}}{u - d}\right]. \tag{2.6}$$

Recall the specification $d < e^{rT} < u$ of the parameters $u, d, r$ in Section (2.6.2). It follows that if we denote

$$q = \frac{e^{rT} - d}{u - d}$$

then

$$1 - q = \frac{u - e^{rT}}{u - d}$$

and $0 < q < 1$. Thus $\{q, 1 - q\}$ is a probability distribution. The formula (2.6) can be re-writen as

$$V_0 = E^Q(e^{-rT}V_T),$$

where $E^Q$ is understood as taking expectation under a distribution $Q$ that puts weight $q$ on the event $u$ that the $S$ goes up and $1 - q$ on the event $d$ that $S$ goes down. Correspondingly, the distribution $Q$ puts weight $q$ on $V_u$ and $1 - q$ on $V_d$ for the random variable $V_T$. The probability distribution $Q$ is called a risk neutral distribution for the one period model.

**Remark 2.7.1.** *We can view the formula* (2.6) *as a mathematical representation of* $V_0$ *(the other representation is formula* (2.5)*). In other words, there does not have to be a physical interpretation of taking expectation under the risk neutral distribution. We will discuss more aspects of the riskneutral distribution in the next section. There is no appeal to the Law of Large Number in formula* (2.6) *since the probabiliy* $Q$ *is not necessarily the real world probability* $P$ *that* $S_T$ *will go up or down. For example, in Section* (2.7.1) *these real world probabilities are* $0.9$ *and* $0.1$ *respectively. Indeed most of the time* $Q$ *and* $P$ *are very different distributions.*

### 2.7.4 Remarks about the risk neutral measure

- In the formula (2.6), if we use $V_u = uS_0, V_d = dS_0$, then $V_0 = S_0$. This is exactly the same result as we've discussed in section (2.7.1). In this context, the risk neutral measure has the interpretation that

$$E^Q(e^{-rT}S_T) = S_0. \tag{2.7}$$

  That is, if a person puts the weight of a risk neutral probability on $S$ then he is indifferent (neutral) between the payoff today and the discounted expectation of the future payoff based on $S$.

- We do not necessarily live in a risk neutral world. That is the real life probability of the stock increasing or decreasing its value may be different from the risk neutral distribution. In fact this is often the case. We call the probability distribution of a stock in the physical world the objective probability and denote it by $P$. Taking expectation under the risk neutral measure should be looked at as a mathematical tool to obtain the *no arbitrage price* for the financial derivative.

- Two probability distributions $P$ and $Q$ are *equivalent* if for any event $E$, $P(E) = 0$ if and only if $Q(E) = 0$. That is an *improbable* event under one distribution cannot be *probable* under the other distribution. We require a risk neutral probability $Q$ to be equivalent to the objective probability $P$. Indeed, the likelihood of the no arbitrage event only makes sense under the objective probability $P$. When pricing derivative, we operate under the risk neutral framework. The equivalence between $P$ and $Q$ allows us connect back to the real world to assert that the price we found is indeed a no-arbitrage price.

- We will use a variation of (2.7) as a definition of the risk neutral measure. Under the risk neutral measure, the asset satisfies the property that the expectation of the discounted future value is the present value. We say a probability distribution $Q$ is an equivalent risk neutral probability if it is equivalent to the objective probability $P$ and the discounted asset price if a *martingale* under $Q$. The notion of a martingale is discussed in Chapter (4).

- The objective probability is not directly connected to pricing a financial product, except for the requirement that the risk neutral distribution has to be equivalent to it. We usually just compute the risk neutral distribution without specifying the objective probability in derivative pricing.

### 2.8 The fundamental theorems of asset pricing in 1 period model

#### 2.8.1 Mathematical definition of arbitrage opportunity

**Definition 2.8.1.** *An arbitrage opportunity is a* **self-financing** *portfolio* $\pi$ *such that*

$$\begin{aligned} \pi_0 &= 0, \\ P(\pi_T \geq 0) &= 1 \\ P(\pi_T > 0) &> 0. \end{aligned} \tag{2.8}$$

Remark:

1. Self-financing means the portfolio's funding is its own money: it cannot have outside source of funding nor can one withdraw (consume) money from the portfolio.

2. Note that the probability used in the definition is the **physical probability**. It makes sense, since intuitively an arbitrage opportunity is a chance to make money without risk. The risk should be measured via the real world probability.

3. Note, however, that since the risk neutral probability $P^Q$ is equivalent to $P$, $P(\pi_T > 0) = 1$ if and only if $P^Q(\pi_T > 0) = 1$.

4. The portfolio can consist of the underlying asset and the saving account and the financial derivative in consideration.

5. Actually a portfolio does not have to start out at value $\pi_0 = 0$. We can use the following equivalent definition for an arbitrage opportunity:

**Definition 2.8.2.** *An arbitrage opportunity is a* **self-financing** *portfolio* $\pi$ *such that*

$$\begin{aligned} P(e^{-rT}\pi_T \geq \pi_0) &= 1 \\ P(e^{-rT}\pi_T > \pi_0) &> 0. \end{aligned} \tag{2.9}$$

Exercise: Prove that these definitions are equivalent.

#### 2.8.2 The fundamental theorems of asset pricing

1. There is no arbitrage opportunity if and only if a risk neutral measure exists.

Let's use the binomial 1 period model. We have seen that this model is always arbitrage free, since we can find the replicating portfolio, *if and only if the condition* $d < e^{rT} < u$ *is satisfied.* We now show this condition is equivalent to the risk neutral measure exists.

Note that

$$E^Q(e^{-rT}S_T) = e^{-rT}S_0(uq + d(1-q)) = S_0$$

and this leads to

$$q = \frac{e^{rT} - d}{u - d}$$

as we mentioned before. Note that $0 < q < 1$ if and only if $d < e^{rT} < u$ in this case. ($q$ cannot be 0 or 1 since in this case, *the risk neutral probability cannot be equivalent to the physical probability*).

2. In an arbitrage free market, the replicating portfolio for **any financial derivative** exists if and only if the risk neutral measure is unique. In this case we say *the market is complete.*

Remark:

a. The key word in the 2nd fundamental theorem of asset pricing is **any derivative**. A consequence of this is if the risk neutral measure is *not* unique, then there must be *some* financial derivative based on the asset that we cannot priced. We give an example where the risk neutral measure is not unique, and the replicating portfolio does not exist: the trinomial model.

b. We need the no arbitrage condition to be satisfied first before we can discuss the completeness of the market. This is mainly used in verifying *the underying asset itself cannot create an arbitrage opportunity,* (as manifested in the condition $d < e^{rT} < u$ in the 1 period binomial model).

### 2.8.3 A trinomial model example

Suppose $S_T$ have three possible outcomes: $u, m, d$ (for up, middle and down). Also for simplicity let $m = e^{rT}$. We will impose $d < e^{rT} < u$ as our usual **necessary condition** for no arbitrage, but *there might be some additional conditions that are required for no arbitrage*. Additional investigation of the model is needed to completely determine what the set of no arbitrage conditions are, in general. Here it turns out that the condition $d < e^{rT} < u$ **is also sufficient**, because of our choice $m = e^{rT}$. For other choices of $m$, we can see that the no arbitrage conditions need further modification.

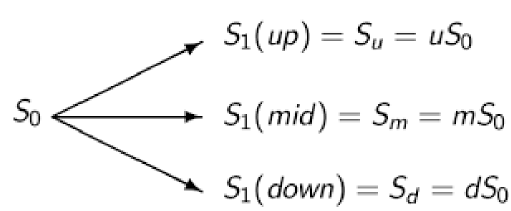



Figure 2.3: One period trinomial model

And suppose $\min(P(S_T = u), P(S_T = m), P(S_T = d)) > 0$. Then the risk neutral probability $Q$ needs to put positive weight on all these outcomes. So let

$$P^Q(S_T = u) = q_1, P^Q(S_T = e^{rT}) = q_2, P^Q(S_T = d) = 1 - q_1 - q_2.$$

We then require

$$E^Q(e^{-rT}S_T) = e^{-rT}S_0(uq_1 + e^{rT}q_2 + d(1 - q_1 - q_2)) = S_0.$$

That is

$$q_1(u-d) + q_2(e^{rT} - d) = e^{rT} - d.$$

There are several choices for $q_1, q_2$.
a. If we choose $q_1 = q_2$ then we have

$$q_1 = q_2 = \frac{e^{rT} - d}{u - d + e^{rT} - d}.$$

It follows automatically that $0 < q_1 = q_2 < 1$.
We also need $q_1 + q_2 < 1$ therefore we require

$$\frac{e^{rT} - d}{u - d + e^{rT} - d} < \frac{1}{2}.$$

This is equivalent to $e^{rT} - d < u - d$ or $e^{rT} < u$ as we already have.
b. If we choose $q_2 = 2q_1$ then

$$q_1 = \frac{e^{rT} - d}{u - d + 2(e^{rT} - d)}.$$

This automatically implies that $q_1, q_2 > 0$. We need to guarantee $0 < q_1, q_2, q_3 < 1$. It is enough to require $q_1 + q_2 < 1$ or equivalently $3q_1 < 1$. This is equivalent to

$$\frac{e^{rT} - d}{u - d + 2(e^{rT} - d)} < \frac{1}{3}.$$

But this is equivalent to, again $e^{rT} - d < u - d$ or $e^{rT} < u$.

So we see that **the no arbitrage condition in our 1 period trionomial model is** $d < e^{rT} < u$. However, the risk neutral measure **is not unique** as we have showed there are at least 2 choices for it.

Observe also that **the replicating portfolio does not exist** as we remarked, since the system

$$\begin{aligned} xuS_0 + ye^{rT} &= V_u \\ xe^{rT}S_0 + ye^{rT} &= V_m \\ xdS_0 + ye^{rT} &= V_d \end{aligned}$$

is over-determined. The matrix

$$\begin{bmatrix} uS_0 & e^{rT} & V_u \\ e^{rT}S_0 & e^{rT} & V_m \\ dS_0 & e^{rT} & V_d \end{bmatrix}$$

has REF

$$\left[\begin{array}{ccc} uS_0 & e^{rT} & V_u \\ S_0(u-e^{rT}) & 0 & V_u - V_m \\ 0 & 0 & \frac{(V_u-V_m)(u-d)}{u-e^{rT}} - (V_u - V_d) \end{array}\right],$$

so unless

$$V_u - V_m = (V_u - V_d)\frac{u-e^{rT}}{u-d}, \tag{2.10}$$

the system cannot have a solution. (Note that for the forward contract $V_T = S_T - K$ we always have a replicating portfolio. It is trivial to check that the forward contract satisfies (2.10)).

### 2.8.4 No-arbitrage pricing using risk neutral measure

An important consequence of the first fundamental theorem of asset pricing is that if a risk neutral measure $Q_1$ exists, then **a candidate** for no arbitrage price of a financial derivative that pays $V_T$ at time $T$ is $V_0^1 = E^{Q_1}(e^{-rT}V_T)$. We say a candidate because there may exist another risk neutral measure $Q_2$ and it is sufficient, for the no-arbitrage condition to hold, to charge $V_0^2 = E^{Q_2}(e^{-rT}V_T)$ for the derivative as well. Thus there is no unique price for the financial product in this case, if our criterion is only the no-arbitrage condition.

To demonstrate, we show that a portfolio consisting of just the financial derivative itself cannot be an arbitrage opportunity. Note that the no arbitrage condition requires that the two following conditions **cannot hold together**

$$\begin{array}{rcl} P(e^{-rT}V_T \geq V_0) & = & 1 \\ P(e^{-rT}V_T > V_0) & > & 0. \end{array}$$

But this is equivalent to requires that the two following conditions **cannot hold together**

$$\begin{array}{rcl} P^{Q_1}(e^{-rT}V_T \geq V_0) & = & 1 \\ P^{Q_1}(e^{-rT}V_T > V_0) & > & 0. \end{array}$$

(or for that matter, any measure $\tilde{P}$ equivalent with $P$ the above two cannot hold together under $\tilde{P}$).

we can verify that the condition $V_0 = E^{Q_1}(e^{-rT}V_T)$ implies that

$$\begin{array}{rcl} P^{Q_1}(e^{-rT}V_T \geq V_0) & = & 1 \\ P^{Q_1}(e^{-rT}V_T > V_0) & > & 0. \end{array}$$

Note that the condition $S_0 = E^{Q_1}(e^{-rT}V_T)$ implies that a portfolio consisting of the underlying asset and the derivative cannot be an arbitrage opportunity. Clearly a saving account satisfies $y_0 = e^{-rT}(e^{rT}y_0)$. Thus we see how the existence of an equivalent risk neutral measure implies the no arbitrage condition.

### 2.9 Market with more than 1 asset

Observe that in the trinomial model above, the reason why we cannot replicate certain financial product is because we do not *have enough financial assets.* It makes sense that as the financial product "becomes more complex" (in the sense that it has more outcomes), we need more underlying assets to replicate it. In particular, we can imagine a market with 2 underlying assets $S^1, S^2$. They can go up, stay neutral or go down at time $T$. Specifically, there are 3 possible outcomes for $S^i(T) : S^i(T) = u_i S^i(0), S^i(T) = e^{rT} S^i(0), S^i(T) = d_i S^i(0), i = 1, 2$.

For simplicity we can assume that $S_1, S_2$ move in "synchrony", that is if $S^1$ goes up, stays neutral or goes down, then $S^2$ would also go up, stay neutral or go down (This is not so innocent as it seems, *the synchronicity of* $S^1, S^2$ *can cause the non existence of the equivalent risk neutral measure*, see examples below).

What are the conditions on $u_i, d_i$ and the synchronicity of $S^1, S^2$ that would make the model arbitrage free?

We can come up with a financial derivative based on $S^1, S^2$, for example $V_T = (S^1_T + S^2_T - K)^+$. Can we find a replicating portfolio for **any** $V_T$ (not just this particular example)? The answer is yes, via solving the system

$$\begin{aligned} x_1 u_1 S^1(0) + x_2 u_2 S^2(0) + y e^{rT} &= V_u \\ x_1 e^{rT} S^1(0) + x_2 e^{rT} S^2(0) + y e^{rT} &= V_n \\ x_1 d_1 S^1(0) + x_2 d_2 S^2(0) + y e^{rT} &= V_d \end{aligned}$$

Finally, we have the following principle for asset pricing in discrete model: Suppose the market has $n$ risky assets $S_1, S_2, \cdots, S_n$, and each $S_k$ has $m$ possible outcomes. Then the market is complete if $n \geq m$.

#### 2.9.1 Some examples

Suppose $r = 0$. Let $S^1_0 = 200, S^2_0 = 300$ and $S^1_T$ can take values $400, 200, 100$, $S^2_T$ can take values $400, 300, 100$. Below we will consider two examples with these set up, just changing the synchronicity of $S^1, S^2$. we'll see both markets are complete, but one is abitrage free and one is not.

*Arbitrage free and complete market*

There are 3 possible outcomes $\omega_1, \omega_2, \omega_3$. Then we specify that $S^1_0 = 200$

$$\begin{aligned} S^1_T(\omega_1) &= 400 \\ S^1_T(\omega_2) &= 200 \\ S^1_T(\omega_3) &= 100 \end{aligned}$$

and $S_0^2 = 300$

$$\begin{aligned} S_T^2(\omega_1) &= 300 \\ S_T^2(\omega_2) &= 400 \\ S_T^2(\omega_3) &= 100. \end{aligned}$$

Note that when $S^1$ goes up, $S^2$ states neutral and vice versa. We claim that this model is arbitrage free and complete. Indeed we can check that the unique risk neutral probability is

$$P^Q(\omega_1) = \frac{1}{7}; P^Q(\omega_2) = \frac{4}{7}; P^Q(\omega_3) = \frac{2}{7}.$$

Let's apply this to pricing a call option that pays $(S_T^1 + S_T^2 - 400)^+$ at time $T$. If we hold $x_i$ shares of $S^i$ and $y$ dollars in cash then the matrix system for $x_1, x_2, y$ is

$$\begin{bmatrix} 400 & 300 & 1 & 300 \\ 200 & 400 & 1 & 200 \\ 100 & 100 & 1 & 0 \end{bmatrix},$$

which has REF

$$\begin{bmatrix} 1 & 0 & 0 & 5/7 \\ 0 & 1 & 0 & 3/7 \\ 0 & 0 & 1 & -800/7 \end{bmatrix}.$$

Therefore the price for this option is

$$x_1 S_0^1 + x_2 S_0^2 + y = \frac{5}{7}200 + \frac{3}{7}300 - \frac{800}{7} = \frac{1100}{7},$$

which agrees with the price we get via expectation under risk neutral probability:

$$\frac{1}{7}V_u + \frac{4}{7}V_n + \frac{2}{7}V_d = \frac{1}{7}300 + \frac{4}{7}200 = \frac{1100}{7}.$$

Actually because the left hand side of the matrix is always the identity, we can see that any financial product is replicable.

*Complete but not arbitrage free market*

This time we change the outcomes of $S^1, S^2$ to $S_0^1 = 200$

$$\begin{aligned} S_T^1(\omega_1) &= 400 \\ S_T^1(\omega_2) &= 200 \\ S_T^1(\omega_3) &= 100 \end{aligned}$$

and $S_0^2 = 300$

$$\begin{aligned} S_T^2(\omega_1) &= 400 \\ S_T^2(\omega_2) &= 300 \\ S_T^2(\omega_3) &= 100. \end{aligned}$$

Note that now $S^2$ goes up and stays neutral whenever $S^1$ goes up and stays neutral and vice versa. we can check that the only risk neutral measure we can find is

$$P^Q(\omega_1) = P^Q(\omega_3) = 0; P^Q(\omega_2) = 1.$$

But this is *not equivalent* to the physical measure, assuming that the physical measure puts positive weights on all three outcomes. Thus this means there is arbitrage opportunity for this model. Can we find it?

Surprisingly, this model is still complete. For simplicity again let's consider again the call option that pays $(S_T^1 + S_T^2 - 400)^+$ at time $T$. If we hold $x_i$ shares of $S^i$ and $y$ dollars in cash then the matrix system for $x_1, x_2, y$ is

$$\begin{bmatrix} 400 & 400 & 1 & 400 \\ 200 & 300 & 1 & 100 \\ 100 & 100 & 1 & 0 \end{bmatrix},$$

which has REF

$$\begin{bmatrix} 1 & 0 & 0 & 5/3 \\ 0 & 1 & 0 & -1/3 \\ 0 & 0 & 1 & -400/3 \end{bmatrix}.$$

The thing to note is the left hand side of the matrix again is the identity. Thus any asset is replicable, yet the market is NOT arbitrage free!

### 2.10 Asset that pays dividend

In many cases, an asset can pay dividend, either in the form of cash or shares. Then the payment can be made at discrete time points (lump sum payment) or continuously. Eitherway, the stock price will decrease after dividend payment is made. We describe these situations and the effect of dividend payment on a portfolio value.

#### 2.10.1 Stock dividend

The first form of dividend payment is in a percentage of the stock price, that is *reinvested in to the stock*. This has the overall effect of increasing the number of shares a stock holder have in his portfolio.

Suppose the dividend payment rate is 30 % per annumn. Also suppose that this is a one time payment made at time $t$. To clarify the situation, we will use $t^-$ to describe the moment right before time $t$. Then the company makes a one time payment of 30 % $S_{t-}$ in dividend at time $t$. The stock price $S_t$ becomes $S_t = .7S_{t-}$. If you hold $x$ shares of $S$ then your portfolio value at time $t$ (after the dividend payment) is

$$\pi_t = x(.7S_{t-}) + x(.3S_{t-}) = xS_{t-} = \pi_{t-},$$

so it hasn't changed in value.

Now suppose you use this money to reinvest into the stock, then the number of shares you hold at time $t$ is

$$\tilde{x} = \frac{\pi_t}{S_t} = \frac{xS_{t-}}{0.7S_{t-}} = \frac{x}{.7}.$$

So another way to look at your portfolio value at time $t$ is

$$\pi_t = \frac{x}{0.7}S_t = \frac{1}{1-0.3}xS_t.$$

The above calculation is just an illustration. In the situation that the dividend payment is stock dividend, the reinvesment into the stock *is automatically made.*

Now suppose that the dividend is paid $n$ times over the time interval $[0, T]$, ($T$ is in year), still at the same rate 30% per annum. Then at each time $\frac{0.3T}{n}$ % of the stock price is paid in dividend. You can verify that

$$\pi_T = \left(\frac{1}{1-\frac{0.3T}{n}}\right)^n xS_T.$$

As $n \to \infty$ this approaches

$$\pi_T = e^{0.3T}xS_T.$$

Thus we say if the dividend payment rate is $q$ and the stock is continuously reinvested, then a portfolio consisting of one share of the stock at time $0$ is worth $e^{qT}S_T$ at time $T$. More precisely, 1 share of $S$ has grown to $e^{qT}$ shares of $S$ at time $T$.

### 2.10.2 Cash dividend

The second form of dividend payment is in cash, *that is not automatically reinvested into the stock.* More specifically, suppose a cash amount $d$ is paid out a time $t$. Then the stock price $S_t$ becomes

$$S_t = S_{t-} - d.$$

A portfolio consisting of $x$ shares of $S$ stays the same in value after the dividend payment:

$$\pi_t = x(S_{t-} - d) + xd = xS_{t-} = \pi_{t-}.$$

Since the cash payment is not automatically reinvested into the stock, the value of this particular portfolio at time $T$ is

$$\pi_T = xS_T + xde^{r(T-t)}.$$

Remark: 1. Technically the cash dividend can also be reinvested into the stock at the discretion of the portfolio holder. But in this case it will come back to the stock dividend case discussed above, just not continuously re-invested. Since we are discussing the one period model, we will assume the portfolio holder does not rebalance their portfolio in between time $0$ and time $T$. The re-balancing situation will be discussed later in the multi-period model.

2. The difference between stock dividend and cash dividend, besides the re-investment issue, is the denomination. Stock dividend is denoted in percentage of stock, and cash is just in dollars.

### 2.10.3 Stock price has to decrease after dividend payment

You may wonder if there is a situation where the stock price stays the same after the dividend is paid out (in either form). The answer is no. The reason is there will be arbitrage if this is the case. Suppose a cash dividend $d$ is made at time $t$. Then if the stock price stays the same, one can simply borrow $S_t$ from the bank to buy 1 share of the asset right before time $t$, collect the dividend payment $d$ at time $t$ and sell the asset to pay back the loan to the bank. This way one made $d$ dollars in riskless profit. But if the asset price decreases to $S_t - d$ after the dividend payment, then the situation described above won't happen.

## 2.11 Exercises

**1.** (a) Suppose Apple is currently trading at \$500.00. Alice believes Apple will go up in a major way and wants to bet \$60,000 on this speculation. Right now, suppose a January call at strike \$600 costs \$6.00. Therefore Alice can buy 10,000 calls with her money; since Apple options are exchange-traded in standardized contracts for the purchase or sale of 100 shares, this means she can buy 100 contracts. Suppose she buys and holds the option until expiry. Call this investment strategy I. Strategy II is to simply invest the entire \$60,000 today in Apple stock and sell in January at the same date the options expire. (For simplicity assume the interest rate is 0.) If the stock price on the date of expiry is \$620 per share, calculate her profit from each different strategy.

**Comment:** we should find in (a) that the profit for strategy I is significantly greater; this is an example of how an option may provide leverage for speculation. Of course, strategy I also magnifies possible losses. If the option expires worthless, then Alice will lose her entire \$60,000. She would lose everything in strategy II only if the stock price went to 0, which is very unlikely.

(b) For each strategy in (a), determine the profit or loss as a function of the price $S_T$ of Apple stock at the expiration date. Determine that price $S_T^*$ at which both strategies produce the same profit. Alice would have to strongly believe that $S_T$ will be larger than $S_T^*$ to prefer strategy I.

**2.** A hedger would buy a put on XYZ stock to protect against a decrease in XYZ stock price. A speculator would buy a put on XYZ stock to try to profit from a decrease in XYZ stock price. Explain.

**3.** Suppose we enter a 3 month forward contract on a non-dividend paying stock. The price of the stock today is \$ 25 and the risk-free borrowing and lending rate is 6% per annum. What is the forward price?

A month later the price of the stock is \$ 30 and the interest rate is the same. What is the above contract worth at this time?

**4.** In this question, we will find the forward price for a forward contract on a stock that pays dividend. Assume the interest rate is for the money market is $r$.

a. Suppose a stock pays a one time dividend worth $d$ dollars at time $t$, where $0 < t < T$. That is, at time $t$, the holder of 1 share of stock receives $d$ dollars (and only at this time). Shows that the forward price is

$$F(0,T) = [S_0 - e^{-rt}d]e^{rT}.$$

b. Suppose a stock pays dividend continuously at a rate $r_d > 0$, called the dividend yield, which is continuously reinvested in the stock. That is, an investment in one share held at time 0 will increase to become $e^{r_d T} S_T$ at time $T$. Show that the forward price is

$$F(0,T) = S_0 e^{(r-r_d)T}.$$

**5.** Consider the one period model: we have a stock whose price at time $k = 0$ is $S_0$ and

$$\begin{aligned} S_T &= uS_0 \text{ with probability } q \\ &= dS_0 \text{ with probability } 1 - q, \end{aligned}$$

where $u \leq e^{rT} \leq d$. We add an additional assumption that this stock pays dividend at rate $r_d$ that is reinvested: one share of $S$ at time $k = 0$ is worth $e^{r_d T} S_T$ at time $T$.

a. What is the no arbitrage condition for this model?

b. What is the no arbitrage price of a financial derivative that is worth $V_u$ when $S_T = uS_0$ and worth $V_d$ when $S_T = dS_0$? (our answer will be in terms of $q, r, T, S_0, u, d, V_u, V_d$).

c. Is this model arbitrage free and is the market complete in this model?

**6.** Show that in the 1 period binomial model, there are infinitely many choices for $x$ and $y$ in the game theory portfolio so that $\pi_T(\omega)$ is a constant in $\omega$. Moreover, for any of those choice, $V_0$ is the same, and equals to the value we found in class.

**7.**

a. Show that if the price of the forward contract on an asset $S$ with expiration $T$ and strike $K$ is not $S_0 - Ke^{-rT}$ then there is an arbitrage opportunity.

b. Show that in the game theory portfolio, if we don't have $\pi_T = \pi_0 e^{rT}$ then there is an arbitrage opportunity.

**8.**

a. Suppose the stock of a company follows the following model: $S_0 = 300$, $S_T = 350$ with probability $99\%$ and $S_t = 250$ with probability $1\%$. Suppose $T = 1$ year and $r = 0$. What is the no arbitrage price of a forward contract on $S$ with expiration $T$ and strike $280$?

b. Answer the same question as a, with the difference that $P(S_T = 350) = 1$.

**9.** Show that the trinomial model we described in Section 5.2 of the lecture note is arbitrage free.

**10.** Show that the one period, trinomial market with 2 assets in the trinomial model we described in Section (2.9) is complete. What conditions do we need for this conclusion? What condition do we need for the market to be arbitrage free?

# CHAPTER 3 Derivative pricing in multi-period model

## 3.1 Description of the multi-period binomial model

To make our model richer, we'll transition from the 1-period model to the multi-period binomial model. Specifically we'll have:

### 3.1.1 Notations

The present time, denoted at $n = 0$ and the expiration time, denoted at $n = N$. The lower case letter $n, k$ (and occasionally $i, j, m$) will be used to denote the time variable. The notation $S_k$ (or $S_n, S_i, S_j \cdots$) denotes the value of the stock (the underlying) at time $k$ (or time $n, i, j \cdots$). An important convention we'll use is that $S_0$ *will always be a constant*, that is the present value of the stock is always known. For any $k \geq 1$, $S_k$ is *a random variable.* The specific distribution of $S_k$ will be discussed below.

Similarly, we'll denote $V_k$ to be the value of a specific financial product at time $k$. In particular, if $V$ is the European call option with strike $K$ and expiration $N$, then $V_N = (S_N - K)^+$. We'll also denote $\pi_k$ to be the value of a specific portfolio at time $k$. In particular, if the portfolio is *replicating* then $V_N = \pi_N$. Also note that $V_0, \pi_0$ are also constants, and for $k \geq 1$, $V_k, \pi_k$ are *random variables*.

We will suppose that the time interandom variableals between any two discrete moments $k, k+1$ are the same, denoted as $\Delta T$. Thus the expiration time can also be written as $T = N\Delta T$.

For a replicating portfolio, we will denote the number of shares of $S$ we hold at a particular time as $\Delta_k$ (do not confuse this with the interandom variableal length $\Delta T$. In general, $\Delta_k$ will also be *a random variable* (which is easy to understand, as the number of shares we hold at time $k$ will depend on the actual value of $S_k$ at that time).

The interest rate will be denoted as $r$.

### 3.1.2 The evolution of the stock

At any time $k$, there are two possibilities for the stock to evolve: either jump up to $S_{k+1} = uS_k$ or jump down to $S_{k+1} = dS_k$ for some $d < e^{r\Delta T} < u$. Moreover, the distribution of $S_{k+1}$ *only depends on* $S_k$ *and not any further history of* $S$. In mathematical notation we write:

$$P(S_{k+1} = x | S_0, S_1, \cdots, S_k) = P(S_{k+1} = x | S_k),$$

for any real value $x$. That is, as far as deciding the behavior of $S_{k+1}$, the only information we need is the behavior of $S_k$. Any further information about its past history is irrelevant. We call this property of a random process (which $S_k, k = 1, \cdots N$ is) *the Markov property*.

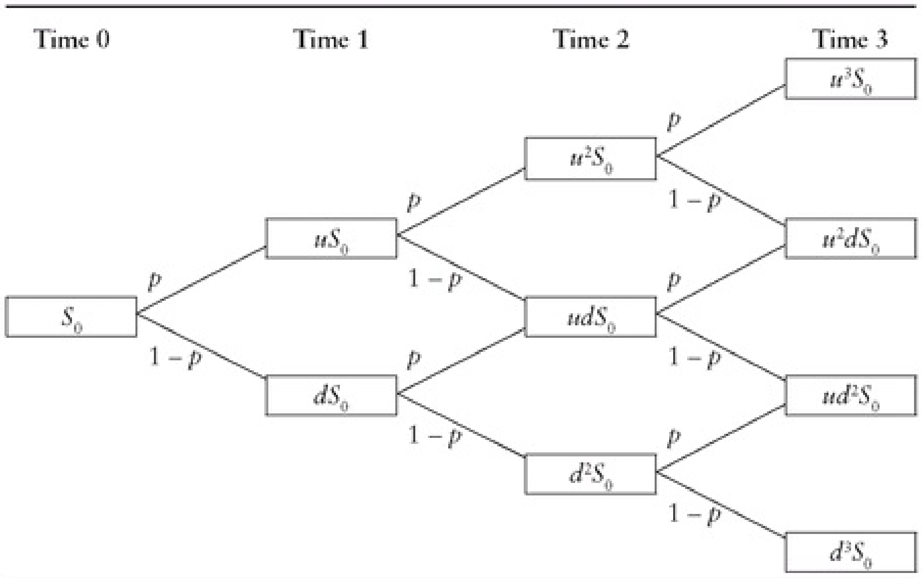



Figure 3.1: Multi-period binomial model

An example of something that *is not* a Markov chain is any process whose evolution depends on more than its immediate history. For example, you *may* argue that the distribution of tomorrow's weather (whether it's gonna rain, shine, snow, be windy etc.) does not just depend on today's weather but also on the last couple of days. I.e. if it has been raining for the last 2 days then the chance of its continuing raining is higher than if it's only been raining today. In this sense, the day by day weather is not a Markov process. But one can consider the weather of a two or three days in a row, and then it may be a Markov process. This is a technique called changing the state space of a process. This particular discussion is only for your information. In this class we'll only be dealing with process that is already Markov to start with.

### 3.1.3 A mathematical construct of $S_k$

We will construct our multi-period model for $S_k$ out of *independent, identically distributed random variables that represent jump size.* This will be a very important idea, that we may see again in the construction of general Markov processes, or in Brownian motion, an important ingredient in our continous model later on.

More specifically, let $X_1, X_2, \cdots, X_N$ be i.i.d random variables with distribution

$$P(X_1 = u) = p; P(X_1 = d) = 1 - p,$$

for some $0 < p < 1$. (Note: since these random variables are *identical*, we only need to prescribe the distribution of $X_1$ and the rest would have the same distribution).

Then we simply define

$$S_k := S_0 X_1 X_2 \cdots X_k = S_0 \prod_{i=1}^{k} X_i.$$

Note that in this way we have a recursion relationship between $S_i$ and $S_j, i < j$ which comes in handy when we compute conditional expectation (discussed later on in this lecture)

$$S_j = S_i X_{i+1} X_{i+2} \cdots X_j = S_i \prod_{k=i+1}^{j} X_k.$$

We claim that this way we recover the description of the evolution of $S_k$ given above. Indeed, it is easy to see the probability space and state space are the same. The only thing to check is the Markovian property of $S_k$. We will give a rigorous justification when we discuss conditional expectation in the multi-period model. For now, let's just give some intuition why it is true. We have

$$P(S_{k+1} = x | S_1, S_2, \cdots, S_k) \quad = \quad P(S_k X_{k+1} = x | S_1, S_2, \cdots, S_k)$$

And we see that $X_{k+1}$ is *independent of* $S_1, S_2, \cdots, S_k$ by construction. Thus conditioning on $S_1, S_2, \cdots S_k$ is the same as *not conditioning* as far as $X_{k+1}$ is concerned (recall that if $X$ is independent of $Y$ then $E(X|Y) = E(X)$). On the other hand, the only information we need to determine $S_k$ is $S_k$ itself, and we don't need $S_1, S_2, \cdots, S_{k-1}$. Putting these two facts together, we can believe that

$$P(S_k X_{k+1} = x | S_1, S_2, \cdots, S_k) = P(S_k X_{k+1} | S_k),$$

which is the Markov property.

This description will be important when we use the probabilistic approach (or expectation approach) to pricing a financial product. One important remark here is that *the replicating portfolio approach still works in multi-period model.* The only drawback is it is computationally intensive (we'll see why). Thus if all we are interested in is pricing, then the expectation approach is more efficient. We'll treat the replicating portfolio approach in the in the final section of this section.

### 3.1.4 Some preliminary discussion about Markov chain

As we see in the previous section, $S_k$ is a Markov chain if conditioning on the whole past history of $S_k$ gives us as much information as conditioning on the most recent past event. In this sense, any process whose evolution depends on more than its immediate history is not a Markov chain. For example, we may argue that the distribution of tomorrow's weather (whether it's gonna rain, shine, snow, be windy $\cdots$) does not just depend on today's weather but also on the last couple of days. In other words if it has been raining for the last two days then the chance of its continuing raining is higher than if it's only been raining today. In this sense, the day by day weather status is not a Markov chain. However one can consider the weather status of a two days in a row, and then vector-valued process may be a Markov chain. This is a technique called changing the state space of a process.

Mathematically, a process has Markov property if it has independent increments. This increment may be in an additive form, as in a random walk, or in a multiplicative form, as in our multi-period binomial model. For example,let $X_i$ be i.i.d. random variables where

$$P(X_i = 1) = P(X_i = -1) = 1/2,$$

and $S_k = \sum_{i=1}^k X_i$. Then $S_k$ is a one-dimensional symmetric random walk. It has the interpretation that at any time $k$, we flip a coin and move to the right one unit if the coin is head and move to the left one unit otherwise.

We are already familiar with the construction of the multiplicative independent increment from the multi-period stock model. We remark that by taking the log, the multiplicative increment structure transforms into the additive increment structure: if

$$S_k = \prod_{i=1}^k X_i$$

then

$$\log(S_k) = \log_{i=1}^k \log(X_i).$$

We now show that the independent increment property implies the Markov property. It is clear that $S_k$ satisfies the Markov property if and only if $\log(S_k)$ is also Markov. Therefore, by the above remark, we only need to consider the process $S_k$ with additive independent increment structure. Thus let

$$S_k = \sum_{i=1}^k X_i,$$

where $X_i' s$ are i.i.d. random variables. We need to show

$$E(S_{k+1}|S_0, S_1, \cdots, S_k) = E(S_{k+1}|S_k). \tag{3.1}$$

Note that

$$\begin{aligned} E(S_{k+1}|S_0, S_1, \cdots, S_k) &= E(S_k|S_0, S_1, \cdots, S_k) + E(X_{k+1}|S_0, S_1, \cdots, S_k) \\ &= S_k + E(X_{k+1}), \end{aligned}$$

since $X_{k+1}$ is independent of $S_0, S_1, \cdots, S_k$. Also

$$E(S_{k+1}|S_k) = E(S_k|S_k) + E(X_{k+1}|S_k) = S_k + E(X_{k+1}).$$

Thus (3.1) is verified.

### 3.1.5 The probability space and the state space

we can easily see that for a given $k$, there are only $k+1$ values $S_k$ can take. Namely

$$S_k = S_0 u^i d^{k-i}, i = 0, 1, \cdots, k.$$

It is convenient to have notations to refer to this event. We'll give an example when $N = 3$.

It is clear that when $N = 3$, there are only 8 possible outcomes, namely

$$\Omega = \Big\{uuu, uud, udu, duu, ddu, dud, udd, ddd\Big\},$$

which corresponds to, for example

$$S_3(uuu) = S_0 u^3, S_3(dud) = S_0 d^2 u, S_2(uud) = S_0 u^2, S_0(udu) = S_0 ud \cdots$$

From this list we can create other events, such as

$$\{uu\} = \{uuu\} \cup \{uud\}, \{ud\} = \{udu\} \cup \{udd\} \cdots$$

(and events at the time $k = 1$ level etc)

Note that these events are outcomes for $S_2$ but *not outcome* for $S_3$. Thus it makes sense to say $S_2(ud) = S_0 ud$ but not $S_3(ud)(=?)$.

We'll refer to $\Omega$ as our *probability space* and the values $S_k$ can take as our *state space* for $S_k$. Note that for a particular $N$, there are $2^N$ outcomes in $\Omega$, and there are $k+1$ members in the state space of $S_k$.

## 3.2 The distribution of $S_k$

We have the following result:

**Lemma 3.2.1.** *In the above construction, $S_k$ has a binomial distribution. Namely*

$$P(S_k = S_0 u^i d^{k-i}) = \binom{k}{i} p^i (1-p)^{k-i}.$$

Remark: This is not the typical $Bin(k, p)$ distribution that we're used to, *as far as the values $S_k$ takes is concerned.* But if we consider any time $S$ goes up as a success, and count how many times it goes up until time $k$, then indeed we get a Binomial distribution in the traditional sense.

*Proof.* Recall that $S_k = S_0 X_1 X_2 \cdots X_k$. Clearly $S_k = S_0 u^i d^{k-i}$ if and only if $i$ of $X's$ take value $u$ and $k-i$ of them take value $d$. Because they are independent, the probability for a particular arrangement to happen is $p^i(1-p)^{k-i}$. There are $\binom{k}{i}$ such arrangements, since we just choose $i$ of them among $k$ total to take value $u$.

### 3.3 Financial products in multi-period model

The multiperiod model allows for a richer variety of financial products, namely the products that can depend on the *past history* of the stock. We'll describe the financial products we'll encounter in this course for the multi-period model below.

#### 3.3.1 The European-style options

A financial product that makes a payment $f(S_N)$ for some deterministic function $f$ at time $N$ is referred to as a European-style option. For example, if $f(x) = (x - K)^+$ then we have the European-call and $f(x) = (K - x)^+$ then we have the European-put option. But we can also choose $f(x) = x^2, f(x) = \sin(x)$ or $f(x) = (x^2 - K)^+$. Of course the question about the financial interpretation of these products can be raised (what kind of advantage to the buyers do they offer?). But mathematically, we can treat all of these the same ways as we treat the European-put and European-call options. Moreover, these additional examples allow for more tractable mathemtical problems to be worked on. To recap, the distinguishing features of a European-style option is that the *exercise time* is only at the expiration time $N$ and *the dependence of the payoff is only on the value of the stock at the expiration time* $S_N$.

#### 3.3.2 Exotic options

Now that we're in multi-period model, it makes sense to talk about a history of $S$ on the time interandom variableal $0, 1, \cdots, N$. A financial product that makes payment $f(S_0, S_1, \cdots, S_N)$ for some deterministic function $f$ at time $N$ is referred to as an exotic-option. Thus like a European-style option, the exercise time of an exotic option is still the expiration time $N$, but the pay off can be dependent on the past history of $S$.

Remark: we may argue whether the word *option* has any meaning in these contexts, since we do not have the interpretation of a choice here. This is correct, but simply the word option here can be understood as a financial product. It is also the terminology employed in the finanical literature, so we'll follow the convention here.

Some examples of exotic options: The **look back** option:

$$f(x_1, x_2, \cdots, x_N) = \max_{i=1,\cdots,N} x_i.$$

We call it a look back option since we get the *best* value of the stock in its past history as the payment.

On the other hand, if the payment is the average of the stock value in its history, then we get the **Asian** option:

$$f(x_1, x_2, \cdots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

We also have the **barrier options**: if the stock ever get below or above some threshold, then the option is activated or the option becomes worthless. The first scenario is referred to as **down (or up) and in options**. The second is referred to as **down (or up) and out options.** The barrier options can in turn be call or put option. It means if the option *ever gets activated* (or *never knocked out*) then its payment at the expiration time will be the same as a European call or put option. In particular, let's consider a barrier call option with strike $K$ and barrier $L$ ($L$ is a constant). The function $f(x_1, x_2, \cdots, x_N)$ in this case takes the following form:

Up and in:

$$f(x_1, x_2, \cdots, x_N) = \mathbf{1}_{\{\max_{i=1,\cdots,N} x_i \geq L\}}(x_N - K)^+.$$

Up and out:

$$f(x_1, x_2, \cdots, x_N) = \mathbf{1}_{\max_{\{i=1,\cdots,N} x_i \leq L\}}(x_N - K)^+.$$

Down and in:

$$f(x_1, x_2, \cdots, x_N) = \mathbf{1}_{\min_{\{i=1,\cdots,N} x_i \leq L\}}(x_N - K)^+.$$

Down and out:

$$f(x_1, x_2, \cdots, x_N) = \mathbf{1}_{\{\max_{i=1,\cdots,N} x_i \geq L\}}(x_N - K)^+.$$

Where, for a real number $x$, we have

$$\begin{aligned} \mathbf{1}_{\{x \geq L\}} &= 1 \text{ if } x \geq L \\ &= 0 \text{ if } x < L. \end{aligned}$$

### 3.3.3 American options

The last product we'll encounter is the American option, which is similar to a European option, with the additional feature that it gives the holder *the right to choose the time to exercise* the option with a fixed strike price $K$. More specifically, for an American *call* option, the option holder can choose a time between $0$ and $N$ to exercise and *pay* $K$ dollars for a share of $S$. For an American call option, the option holder can choose a time between $0$ and $N$ to exercise and *sell* a share of $S$ for $K$ dollars.

## 3.4 Pricing by the replicating portfolio approach

### 3.4.1 Self-financing portfolio

The idea of pricing using the replicating portfolio approach is the same: we would like to construct a portfolio, initially with $\Delta_0$ shares of stock and $y$ dollars in the money market such that at the expiration time

$$\Delta_N S_N + y_N = V_N, \tag{3.2}$$

where $V_N$ is the value of the financial product and $y_N$ is the amount of cash in the portfolio at time $N$. Notice that here we use $\Delta_N$ and $y_N$ versus the construction in the 1 period model, where the number of shares of stocks remains the same and the money market account only becomes $ye^{rT}$. The reason is in the multi-period model, we can't expect to select our portfolio at time $0$ and leave it unattended, hoping that equation (3.2) will hold at time $N$. Such approach is called the **buy and hold** approach for a portfolio; and obviously we can't generally use such approach for pricing in multi-period model. (One reason is we can visualize, with the buy and hold approach, we're treating the multi-period model as one big one-period model. However, now at the terminal time $T$, $S_N$ can take $N + 1$ values, in stead of just 2 values as before. Thus one cannot solve for $\Delta_0$ and $y$ in a over-determined system, as discussed in the previous lecture).

Thus we need to re-balance our portfolio at each time period $1, 2, \cdots, N$. How we select the portfolio will be addressed next; but first we need to note there is one obvious constraint, if the portfolio is replicating. That is the portfolio has to be *self-financing*, i.e. one cannot put additional funding into the portfolio, nor can one withdraw the cash from it. Letting $y_k$ be the amount of cash in the portfolio at time $k$, the **self-financing condition** can be expressed as:

$$\pi_{k+1} = \Delta_{k+1} S_{k+1} + y_{k+1} = \Delta_k S_{k+1} + y_k e^{r\Delta T}. \tag{3.3}$$

The interpretation is this: we hold $\Delta_k$ shares of stock and $y_k$ dollars at time $k$. At time $k + 1$, the stock price changes to $S_{k+1}$ and the money market grows to $y_k e^{r\Delta T}$. This is our portfolio value at time $k + 1$, $\pi_{k+1}$. Now we can rebalance our portfolio, if we want to. But all we at our disposal is $\pi_{k+1}$ dollars. Thus if we want to buy $\Delta_{k+1}$ shares of stock at price $S_{k+1}$, then the amount of cash we have in the bank, $y_{k+1}$ has to be such that

$$\pi_{k+1} = \Delta_{k+1} S_{k+1} + y_{k+1}.$$

Note that throughout this discussion, $y_k$ can be negative, with the interpretation that we borrow money from the bank, in stead of putting it in a saving account. Note that since

$$\pi_k = \Delta_k S_k + y_k,$$

it follows from (3.3) that

$$\pi_{k+1} = \Delta_k S_{k+1} + e^{r\Delta T}(\pi_k - \Delta_k S_k). \tag{3.4}$$

This equation can be looked as as a recurrence relation between $\pi_{k+1}$ and $\pi_k$. It has the advantage that the cash holding $y_k$ and $y_{k+1}$ do not show up in the equation, thus reducing the number of unknowns we have to deal with (only $\Delta_k$ needs to be found). observe that both (3.3) and (3.4) are equivalent, thus we'll refer to either one as the self-financing condition (or equivalently, the *self-financing equations*). More often equation (3.4) will be used. However, we should still use our judgement to decide which equation is best for which situation.

The self-financing equation (3.4) will be *the key* to solve for the replicating portfolio. The pricing by the replicating portfolio approach in the 1-period model is simply an application of the self-financing condition, as we'll show in the next section.

### 3.4.2 Self-financing equation in 1 period model

In the 1 period model, $N = 1$ thus the equation (3.4) reduces to

$$\pi_1 = \Delta_0 S_1 + e^{r\Delta T}(\pi_0 - \Delta_0 S_0).$$

How to use this equation? The portfolio is replicating, thus $pi_1 = V_1$. Keeping in mind that this equation has to hold for all outcomes, we have

$$\begin{aligned} V_u &= V_1(u) = \Delta_0 S_1(u) + e^{r\Delta T}(\pi_0 - \Delta_0 S_0) = \Delta_0 u S_0 + e^{r\Delta T}(\pi_0 - \Delta_0 S_0) \\ V_d &= V_1(d) = \Delta_0 S_1(d) + e^{r\Delta T}(\pi_0 - \Delta_0 S_0) = \Delta_0 d S_0 + e^{r\Delta T}(\pi_0 - \Delta_0 S_0). \end{aligned}$$

Thus we have $\Delta_0 = \frac{V_u - V_d}{S_0(u-d)}$, exactly as we had before.

### 3.4.3 The steps of solving the self-financing equation

### 3.4.4 Non-path dependent option

*A recursive approach* is used to find the replicating portfolio using the self-financing equation. We start at time $k = N - 1$ where equation (3.4) reads as

$$\pi_N = \Delta_{N-1} S_N + e^{r\Delta T}(\pi_{N-1} - \Delta_{N-1} S_{N-1}).$$

What are we solving for? **We are solving for $\Delta_{N-1}$ and $\pi_{N-1}$.** Once we know these then it is clear that we known how to construct a hedging portfolio at time $N-1$. However, observe that the above equation is an equation of *random variables*, so clearly we are NOT solving for explicit (numerical) value of $\Delta_{N-1}$ and $\pi_{N-1}$. Instead, we should ask we are solving for $\Delta_{N-1}$ and $\pi_{N-1}$ **in terms of what variable**? The answer is we are solving for $\Delta_{N-1}$ and $\pi_{N-1}$ in terms of $S_{N-1}$.

A crucial observation here is that since we are at time $k = N - 1, S_{N-1}$ *is a known value.* This has a real-life interpretation that once we arrive at time $N-1$, then by observing the stock market, we know what $S_{N-1}$ is. Since $\Delta_{N-1}$ and $\pi_{N-1}$ are expressed in terms of $S_{N-1}$, we can balance our portfolio accordingly.

What about $\pi_N$, how do we know its value? This is where *the replication property is used*. Since the portfolio is replicating, $\pi_N = V_N$ and if we are dealing with say a European-style derivative, then we can replace $V_N$ with $g(S_N)$ for some function $g$.

What about $S_N$? This is where *the binomial model is used.* Namely, for a given value of $S_{N-1}$ (recall we assumed it's known, since we're at time $N-1$), the only values $S_N$ can take are $uS_{N-1}$ or $dS_{N-1}$, thus essentially reducing us to the 1-period model's case. More details will be given in the example in class.

Now assume that we have solved this equation at time $N-1$. Proceed backwardly, the next equation at time $N-2$ is

$$\pi_{N-1} = \Delta_{N-2} S_{N-1} + e^{r\Delta T}(\pi_{N-2} - \Delta_{N-2} S_{N-2}).$$

But now we're back to the same procedure above, with $\pi_{N-1}$ being function of $S_{N-1}$, $S_{N-2}$ is known, and we're solving for $\Delta_{N-2}$ and $\pi_{N-2}$ in terms of $S_{N-2}$.

Thus we see that *provided, at each step k we can solve for $\Delta_k$ and $\pi_k$* then keep on going we will arrive at step $k = 0$, at which point we have completely obtained our replicating portfolio set up, as well as the price of the financial product $V_0$, given by $\pi_0$.

It is not hard to be convinced that we can solve for $\Delta_k$ and $\pi_k$ at each step $k$. An abstract proof can be given, but it can be messy. We'll demonstrate this by an example with $N = 3$ in class and we'll see how one can generally prove this fact.

### 3.4.5 Path dependent option

If the derivative is path-dependent, then the above system of equations needs to be replaced with

$$g(S_N, S_{N-1}, \cdots, S_0) = \Delta_{N-1} S_N + e^{r\Delta T}(\pi_{N-1} - \Delta_{N-1} S_{N-1}).$$

So now, we are solving for $\pi_{N-1}$ as a function of $S_{N-1}, S_{N-2}, \cdots, S_0$ instead of just $S_{N-1}$. This has to hold true for all $N+1$ outcomes of $S_N$. The question is, of course, is the above system solvable? The answer is yes.

The key again is for any outcomes of $s_{0,1}, \cdots, s_{N-1}$, $S_N$ can only take 2 values $us_{N-1}, ds_{N-1}$. Thus the above equation reads, for a particular outcome

$$\begin{aligned}
g(us_{n-1}, s_{N-1}, \cdots, s_0) &= \Delta_{N-1}(s_0, s_1, \cdots, s_{N-1}) us_{N-1} \\
&\quad + e^{r\Delta T}(\pi_{N-1}(s_0, s_1, \cdots, s_{N-1}) - \Delta_{N-1} s_{N-1}) \\
g(ds_{n-1}, s_{N-1}, \cdots, s_0) &= \Delta_{N-1}(s_0, s_1, \cdots, s_{N-1}) ds_{N-1} \\
&\quad + e^{r\Delta T}(\pi_{N-1}(s_0, s_1, \cdots, s_{N-1}) - \Delta_{N-1} s_{N-1})
\end{aligned}$$

This is a system of 2 equations with 2 unknowns, and we can verify that it is solvable.

## 3.5 Conditional expectation in the multi-period model

### 3.5.1 The value of a forward contract in the future

Suppose we're in the multi-period model with the present being $k = 0$. Consider the forward contract, which allows the holder to pay $K$ dollars for 1 share of the asset $S$ at time $N$. We've already discussed that its price $V_0$ should be $S_0 - Ke^{-rN\Delta T}$.

Now suppose at a time $n : 0 < n < N$ we want to sell this contract. How much should we charge it by? we should easily see that its price at time $n$ would be $V_n = S_n - Ke^{-r(N-n)\Delta T}$, by a replicating portfolio argument. But suppose we would like to apply the probabilistic approach in this case, how can we do it? Up to now, we used expectation under the risk neutral measure as a method for obtaining the no arbitrage price. But it's clear that taking expectation will not yield $V_n$ of the above forward contract; because taking expectation gives *a constant value*, while $V_n$ is clearly *a random variable.*

Of course the probabilistic approach can still be used, but instead of taking expectation we need to *take conditional expectation.* The intuition is that we are discussing a situation in the future, where *conditioning on the price of* $S_n$, we can decide the value of $V_n$. Indeed conditional expectation is fundamental in studying the multi-period model, as well as the continuous model later on. It is useful, for example, when we want to talk about not only the current price of a financial product, but its price evolution from time $0$ to the expiration time $N$. We'll give a few examples for the multi-period model in the next section.

### 3.5.2 The flow of information

We mentioned that at time $n$, the value of $S_n$ is known to us. This is correct. But to be more precise, at time $n$, all values $S_0, S_1, \cdots, S_n$ are known to us. Thus in deciding the price of a financial product at time $n$, we need to condition on information of $S_0, S_1, \cdots, S_n$ instead of just $S_n$. This would be clear, for example, when we deal with path-dependent or exotic option.

We will then look at expressions of the form

$$E(f(S_{n+k})|S_0, S_1, \cdots, S_n), k \geq 0.$$

We introduce a notation that represents the amount of information regarding $S_k, k = 1, 2, \cdots, n$ available at time $n$ :$\mathcal{F}_n^S$. When the asset in mind is clear (i.e. we're only discussing 1 asset $S$), we'll drop the super-script $S$ and just write $\mathcal{F}_n$. Thus

$$E(f(S_{n+k})|S_0, S_1, \cdots, S_n) = E(f(S_{n+k})|\mathcal{F}_n^S) = E(f(S_{n+k})|\mathcal{F}_n).$$

Now because the process $S_k$ is Markov, we have

$$E(f(S_{n+k})|\mathcal{F}_n) = E(f(S_{n+k})|S_n).$$

Thus most of the time, conditioning on $S_n$ is sufficient. There are exceptions, for example, when we deal with path-dependent option. It is clear that

$$E(S_1S_2|\mathcal{F}_2^S) = S_1S_2 \neq E(S_1S_2|S_2),$$

because

$$E(S_1S_2|S_2) = S_2E(S_1|S_2),$$

and generally $E(S_1|S_2) \neq S_1$.

### 3.5.3 Examples

When taking conditional expectation in the multi-period model, we should try to take advantage of the following:

1. The form of $S_n$ : $S_n = S_0X_1X_2\cdots X_n$.
2. The i.i.d property of $X_i, i = 1, \cdots n$.
3. The elementary properties of conditional expectation.
4. The form of $f$ in $E(f_{S_{n+k}}|S_k)$.

**Example 3.5.1.**

$$E(S_4|S_2) = E(S_2X_3X_4|S_2) = S_2E(X_3X_4|S_2) = S_2E(X_3)E(X_4) = S_2(pu + (1-p)d)^2.$$

**Example 3.5.2.**

$$E(S_3^2|S_2) = E((S_2X_3)^2|S_2) = S_2^E(X_3^2) = S_2(pu^2 + (1-p)d^2).$$

### 3.5.4 Conditional expectation revisited

When dealing with path-dependent options, we cannot rely on the Markovian property of $S$ as remarked above. So the following rule (the so-called tower property of conditional expectation) is important:

**If** $m \le n$ then for any random variable $\xi$:

$$E(E(\xi|\mathcal{F}_n^S)|\mathcal{F}_m^S) = E(E(\xi|\mathcal{F}_m^S)|\mathcal{F}_n^S) = E(\xi|\mathcal{F}_m^S).$$

In other words, when we condition on more information, and then condition on less information, (or the other way) the result is always the same as conditioning on less information.

*Proof.* We prove

$$E(E(\xi|\mathcal{F}_n^S)|\mathcal{F}_m^S) = E(\xi|\mathcal{F}_m^S)$$

and leave the other equality as exercise. First note that $E(E(\xi|\mathcal{F}_n^S)|\mathcal{F}_m^S)$ is a function of $S_0, S_1, \cdots, S_m$ by definition. Let's call it $g(S_0, S_1, \cdots, S_m)$. We need to check for any function $f(S_0, S_1, \cdots, S_m)$

$$E\Big[g(S_0, S_1, \cdots, S_m)f(S_0, S_1, \cdots, S_m)\Big] = E\Big[\xi f(S_0, S_1, \cdots, S_m)\Big].$$

But by definition,

$$E\Big[g(S_0, S_1, \cdots, S_m)f(S_0, S_1, \cdots, S_m)\Big] = E\Big[E(\xi|\mathcal{F}_n^S)f(S_0, S_1, \cdots, S_m)\Big].$$

observe that $f(S_0, S_1, \cdots, S_m)$ is also a function of $S_0, S_1, \cdots, S_n$ **since** $m \le n$. Therefore,

$$E\Big[E(\xi|\mathcal{F}_n^S)f(S_0, S_1, \cdots, S_m)\Big] = E\Big[\xi f(S_0, S_1, \cdots, S_m)\Big].$$

## 3.6 The risk neutral measure

### 3.6.1 Motivation

In the multi-period model, we do not have to limit ourselves to only consider expiration time $n = N$. Consider a forward contract on the asset $S$ with $0$ strike price that has expiration time $n \leq N$. What is the price for this contract at time $k$? Again, using the replicating portfolio apprach, we'll see that the price is $S_k$.

Recall how we define the risk neutral measure in the 1 period model as the measure $Q$ such that

$$E^Q(e^{-rT}S_T) = S_0.$$

The motivation for us there is exactly because the forward contract with 0 strike price expiration $T$ must be worth $S_0$ at time 0. Thus together with the above analysis, we can see that the the risk neutral measure $Q$ in the multi-period binomial model is such that for any $k \leq n$

$$E^Q(e^{-(n-k)\Delta T}S_n|S_k) = S_k. \tag{3.5}$$

### 3.6.2 The formula for the risk neutral measure

The equation (3.5) defines the risk neutral measure. But we want to find out concretely how to implement the risk neutral measure on the multi-period model, just as we did in the 1-period model. One important observation will help us here, that is *when limit to a 1 step period, such as from $n-1$ to $n$, the multi-period model looks exactly as a 1 period model. And the entire multi-period model can be re-produced by repeating so many such 1 step period movements.*

In terms of mathematics, what we're utilizing is the identical property of $X_i$. That is if we find out the distribution of $X_1$ under the risk neutral measure $Q$, then we've found out the distribution of all the $X_i$'s under $Q$ as well. And that completes the decription of risk neutral measure]

Concretely, the equation (3.5) for $n = 1$ and $k = 0$ reads

$$E^Q(e^{-\Delta T}S_1) = S_0.$$

But we have solved this equation before, of course. We conclude that $Q(X_1 = u) = q$ and $Q(X_1 = d) = 1 - q$ where

$$q = \frac{e^{r\Delta T} - d}{u - d}. \tag{3.6}$$

And thus under $Q$, $P(X_i = u) = q$ and $P(X_i = d) = 1 - q$ for all $i = 1, 2, \cdots, N$.

we may be suspicious. We derived this distribution from a 1 period analysis. Are we sure that the equation (3.5) holds for general $n$ and $k$?

To check, note this simple but also important observation:

$$E^Q(X_1) = \frac{e^{r\Delta T} - d}{u - d}u + \frac{u - e^{r\Delta T}}{u - d}d = e^{r\Delta T}.$$

Thus

$$\begin{aligned} E^Q(e^{-(n-k)\Delta T}S_n|S_k) &= E^Q(e^{-(n-k)\Delta T}S_kX_{k+1}X_{k+2}\cdots X_n|S_k) \\ &= e^{-(n-k)\Delta T}S_k[E(X_1)]^{n-k} = S_k, \end{aligned}$$

and equation (3.5) has been checked.

### 3.6.3 A model where volatility stays the same under physical and risk neutral measure

A measure of how volatile the asset is can be defined as followed:

$$(\sigma^P)^2\Delta T = Var^P(\frac{S_{k+1}-S_k}{S_k}).$$

Here the notation $P$ means that the Variance is taken under the probability $P$. We proceed to show that under the Binomial tree model, $(\sigma^P)^2$ is independent of $P$, and thus we just refer to it as $\sigma^2$. The volatility of the stock is $\sigma$.

Under the Binomial tree model, $\frac{S_{k+1}-S_k}{S_k} = X_{k+1}-1$. We suppose that under $P$, $S_k$ has growth rate $\mu$ and volatiltiy $\sigma^P$. That is

$$\begin{aligned} E^P(X_{k+1}) &= 1+\mu\Delta T \\ Var^P(X_{k+1}) &= (\sigma^P)^2\Delta T. \end{aligned}$$

Here we use discret compounding. The only requirement for the risk neutral probability $Q$ is that $S_k$ under it has growth rate $r$. That is

$$E^Q(X_{k+1}) = 1+r\Delta T$$

We can achieve the first set of conditions under $P$ as followed:

$$X_k = 1+\mu\Delta T+\sigma^P W_k,$$

where $W_k$ is a Bernoulli type RV with two states $u, d$ such that $E^P(W_k) = 0, Var^P(W_k) = \Delta T$. We can achieve the 2nd set of conditions under $Q$ by letting

$$X_k = 1+r\Delta T+\sigma^P\tilde{W}_k,$$

where

$$\tilde{W}_k = W_k + \frac{\mu-r}{\sigma^P}\Delta T.$$

We require Q to be such that $E^Q(\tilde{W}_k = 0$. That is

$$E^Q(W_k) = \frac{r-\mu}{\sigma^P}\Delta T.$$

### 3.6.4 Pricing by risk neutral measure

**Theorem 3.6.1.** *Suppose the asset $S_n$ follows the multi-period binomial model, where the probability $S_n$ goes up is given by equation (3.6). Then the no arbitrage price at time $k$ for any financial derivative with exercise time $N$ is*

$$V_k = E^Q(e^{-(N-k)\Delta T}V_N|\mathcal{F}_k^S). \tag{3.7}$$

*In particular, its value at 0 is*

$$V_0 = E^Q(e^{-N\Delta T}V_N).$$

Remark:

1. We will refer to equation (3.7) as the pricing formula (under risk neutral measure).

2. Note that in the pricing formula, the conditioning is on the *history of $S$, up to time $k$*. This formula becomes

$$V_k = E^Q(e^{-(N-k)\Delta T} V_N | S_k)$$

when we deal with European-style derivative for example. But in general, say, when dealing with exotic options, one cannot reduce conditioning on $\mathcal{F}_k^S$ down to $S_k$. Thus the pricing formula is a great theoeretical result for discussing the evolution of the derivative's price. Computing explicitly $V_k$ might take additional work.

3. The pricing formula also *only works for financial product with exercise time $N$*. In other words, it applies to European style and exotic derivatives, but NOT American option. We'll discuss why when we discusses the pricing of American options.

### 3.6.5 The fundamental theorems of asset pricing in multi-period model

we may also question the connection between the risk neutral measure, the existence of the replicating portfolio and the non-existence of arbitrage opportunity. Similar to the one period model, we also have two fundamental theorems that establish their connection here:

**Theorem 3.6.2.** *In the multi-period binomial model, the risk neutral measure exists if and only if there is no arbitrage opportunity.*

**Theorem 3.6.3.** *In the multi-period model, the risk neutral measure exists, and is unique, if and only if there is a replicating portfolio.*

Intuitively, these theorems are true because when we limit to any one step period, the multi-period model "looks like" the 1 period model. We have checked that for the one-period model, these theorems are true.

## 3.7 Remarks on using the binomial tree for pricing

It is common to use the "backward stepping" method to price a financial asset in the multi-period binomial model. This again makes use of the formula (3.7), where now we replace $N$ by $k+1$, by the property of conditional expectation:

$$V_k = E^Q(e^{-\Delta T} V_{k+1} | \mathcal{F}_k^S).$$

Even more explicitly, if we denote $\omega$ to be a vector of length $k$ consisting of $u$ and $d$ (so that $\omega$ denotes an outcome at time $k$) then the above formula becomes

$$V_k(\omega) = e^{-\Delta T}[qV_{k+1}(\omega u) + (1-q)V_{k+1}(\omega d)]. \tag{3.8}$$

This equation can be reduced further, in the case of Euro-style options, to finding the value of $V_k$ at "a certain node" on the binomial tree by the Markov property of $S$. Consider a time $k, 0 \le k \le N$ with the corresponding $k+1$ nodes. The price $V_k$ at a particular node $i, i = 1, \cdots, k+1$ can be computed as

$$V_k(i) = e^{-\Delta T}[qV_{k+1}(iu) + (1-q)V_{k+1}(id)].$$

For example, consider the following example on pricing a put option on a stock with the strike $K = 1.56$ and the expiration $N = 3$. The put price at each node is given in parentheses below the stock price. The risk neutral probability is $q = 0.4626$.

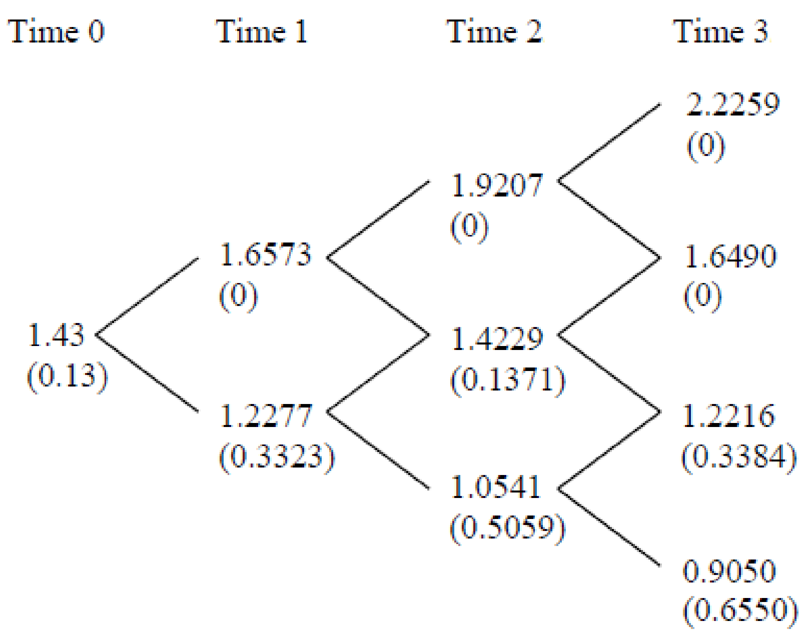



Figure 3.2: Binomial pricing of a put option

At the time $k = 2$ there are $k + 1 = 3$ nodes. At the bottom node $i = 3$, we have

$$\begin{aligned} V_2(3u) &= (1.56 - 1.2216)^+ = 0.3384 \\ V_2(3d) &= (1.56 - 0.9050)^+ = 0.6550 \\ V_2(3) &= 0.4626 \times 0.3384 + (1 - 0.4626) \times 0.6550 = 0.5059. \end{aligned}$$

Formula(3.8) implies that for any outcomes $\omega_1, \omega_2$ of length $k$ that consists of the same portion of $u$ and $d$ (for example $k = 3$ and $\omega_1 = ddu$ and $\omega_2 = dud$), we have

$$V_k(\omega_1) = V_k(\omega_2).$$

This is valid because the price of the financial derivative (in this case the European option) is *Markovian*. That is $V_k$ only depends on $S_k$ (and not $S_{k-1}, S_{k-2}, \cdots$), and $S_k(\omega_1) = S_k(\omega_2)$.

Indeed in the figure (3.2), we see that for the two outcomes $\omega_1 = ud, \omega_2 = du, V_2(\omega_1) = V_2(\omega_2) = 0.1371$. This is because $V_2$ only depends on the value of $S_2$. And in this example $S_2(du) = S_2(ud) = 1.4229$.

However, this finding value at a “certain node" will no longer be valid in a path dependent option, for example a down and out option. It is because now $V_k$ depends not only on $S_k$, but also on $S_{k-1}, S_{k-2}, \cdots$. It could happen that $S_k(\omega_1) = S_k(\omega_2)$ but $S_{k-1}(\omega_1) \neq S_{k-1}(\omega_2)$. For example in figure (3.2),

$$S_3(uud) = S_3(duu) = 1.6490$$

but

$$S_2(uud) = 1.9207 \neq S_2(duu) = 1.4229.$$

So one cannot conclude that $V_k(\omega_1) = V_k(\omega_2)$. For example, if we consider an look back option on the same stock as given in figure (3.2) then you can see that

$$V_3(uud) = \max(1.43, 1.6573, 1.9207, 1.6490) = 1.9207$$

while

$$V_3(duu) = \max(1.43, 1.2277, 1.4229, 1.6490) = 1.6490.$$

However we emphasize that the formula (3.8) is still valid. We just have to price the option via a “path by path" method.

## 3.8 The American options

### 3.8.1 Some preliminaries

*Definition*

An American call option with strike price $K$ and expiration $T$ on an underlying $S$ gives the holder the right to *choose a time* between $0$ and $T$ to *buy* 1 share of $S$ with price $K$.

An American put option with strike price $K$ and expiration $T$ on an underlying $S$ gives the holder the right to *choose a time* between $0$ and $T$ to *sell* 1 share of $S$ with price $K$.

Remark: The American option allows the possibility of buying and exercising the option *immediately*. Therefore, if we let $V_t^A$ be the price of an American call option with strike $K$ and expiration $T$, then $V_t^A \geq (S_t - K)^+$ for all $t$. It is because if $V_t^A < (S_t - K)^+$ then one can buy the option and exercise immediately for a positive profit, which is an arbitrage opportunity. Similarly, the price $V_t^A$ of an American put option also satisfies $V_t^A \geq (K - S_t)^+$ for all $t$.

*The optimal exercise time*

A holder of an American option will judiciously choose a time to exercise the option to maximize his expected profit. Such a time will be referred to as the optimal exercise time.

Also note that a prioi, the optimal exercise time is *a random time*. That is for different realizations of the paths of the underlying $S$, the holder might choose different times to exercise the option.

*Comparison with European option*

We have the following easy, but still important observation: Let $V_t^A$ be the (no arbitrage) price of an American call option and $V_t^C$ the (no arbitrage) price of the corresponding European call option purchased at time $t = 0$ with the same strike $K$ and expiry $T$. That is $V_T^C = (S_T - K)^+$ and the European option holder cannot exercise the option earlier than $T$. Then $V_t^A \geq V_t^C$. That is the price of an American option is always at least as expensive as its European counterpart. Note that this conclusion is model independent: we do not make any assumption on $S_t$.

Reason: Suppose $V_t^C > V_t^A$. Then we take a long position (that is we buy 1 share) on the American option and a short position (that is we sell 1 share) on the European option. Then we make a risk free positive profit equals $V_t^C - V_t^A$. At time $T$, we exercise the American option to close out our short position on the European option. Thus this is an arbitrage opportunity and therefore we must have $V_t^C \leq V_t^A$.

There is another rather surprising result. That is *in the case of a call option, the price of an American option on an asset* **that does not pay dividend** *is equal to the price of a European option for all time*: $V_t^A = V_t^C$, for all $t$. Thus the optimal exercise time of an American call option will always be the expiration time $T$. This result also is *model independent*. We present it in the next subsection.

*American call option*

Let $C(t)$ be the price of a European call with strike $K$ expiring at time $T$. Let $A(t)$ be the price of the corresponding American call. No particular model is put on the price process, except that we assume *no dividends are paid on the asset.* We will use a no-arbitrage argument to show that $C(t) > (S(t) - K)^+$ for all $t < T$, as long as $S(T)$ can fall to either side of $K$ with positive probability. Use this result to show that $A(t) = C(t)$ for all $t$ and that early exercise is never optimal. ( It is also helpful to keep this observation in mind: If $A(t) > (S_t - K)^+$ then it is not optimal to exercise the American option at time $t$ since trading the option itself gives higher pay off than exercising the option).

Ans: As long as $P(S(T) > K) > 0$, the price $C(t) > 0$, because the probability of a strictly positive payoff is greater than zero.

If $0 < C(t) \leq (S(t) - K)^+$ at some $t < T$, then $S(t) > K$ and $S(t) \geq C(t) + K$. This would create an arbitrage opportunity. Suppose you short one share of the underlying (that is you borrow $S(t)$ in cash from an agent and pay back one share of $S$ at time $T$ - Another way to think about it is you borrow 1 share of $S$ now and pay it back at time $T$) and buy the European call for $C(t)$, this leaves you with at least $S(t) - C(t) \geq K$ to invest at the risk-free rate. Since you owe a share of the underlying, your return from this position at $T$ is

$$\begin{aligned}(S(t)-C(t))e^{r(T-t)}+(S(T)-K)^+-S(T) &= (S(t)-C(t))e^{r(T-t)}-\min\{K,S(T)\}\\ &\geq Ke^{r(T-t)}-\min\{K,S(T)\}\end{aligned}$$

If $r > 0$, this is always positive, and so it yields an arbitrage. If $r = 0$, this payoff is non-negative and strictly positive on the event $S(t) < K$, which we assume happens with positive probability, so again we have an arbitrage. It follows that $C(t) > (S(t)-K)^+$, if $t \leq T$, in order that there is no arbitrage.

The price of the, $A(t)$ of the American call is always greater than or equal to $C(t)$. Thus $A(t) \geq C(t) > (S(t)-K)^+$ when $t < T$. Since the value of the American call is thus always strictly greater than the value of immediate exercise if $t < T$, it is optimal to exercise the American call only at $T$. It follows then that the American and European call have the same value: $A(t) = C(t)$, for all $t \leq T$.

### 3.8.2 Optimal stopping

Suppose you're the holder of an American option. Your goal is to choose a time to stop (to exercise the option) judiciously to maximize your return. What properties does this time have to satisfy? We have the following observations:

1. The exercise time would be random, rather tha deterministic : it is clear that for different realizations of the asset, you would want to choose different times to exercise the option.

2. Suppose we are currently at time $t$. The decision, whether or not, to exercise at time $t$ would be based on the past performance of the asset, up to time $t$. For example, if you hold an American put option, one possible stopping rule is to exercise when the asset price goes beyond a level $L$, where $L$ is a constant chosen at time $0$.

However, the exercise decision cannot be made based on the future information after time $t$. We say the random time is *a stopping time with respect to the filtration generated by the underlying asset* $S$. If we denote $\tau(\omega)$ to be the exercise time, then we write $\tau \in \mathcal{F}_t^S$.

3. The best (random) stopping strategy you can make can only be chosen among the stopping time strategies. This is a subtle point to appreciate, as besides stopping time decision, one can imagine another type of mixed strategy: if we have 2 strategies $\tau_1, \tau_2$ then we use $\tau_1$ with probability $p$ and $\tau_2$ with probability $1-p$, and this is done via an independent coin flip. For example, you can do the following: every day you flip a coin and exercise the option the first time the coin flip turns H. This will not be better than making your decision using purely stopping time. The reason is we can show there is an optimal stopping time $\tau^*$ that maximizes your return among other stopping times. Thus if you randomized your decision among stopping times, then you miss the "best optimal" stopping time sometimes. More conrete example will be showed below.

4. If we are in a discrete time model, the decision to stop must be made on discrete time points. In other words, the stopping times we referred to above are discrete stopping times. We study some preliminary properties of discrete stopping times in the next subsection.

### 3.8.3 Discrete stopping time

*Stopping time definition*

**Definition 3.8.1.** *Let $\tau$ be a random variable taking values $\{0, 1, ..., N\}$. We say $\tau$ is a stopping time with respect to $\mathcal{F}^S(n)$ if for all $n = 0, 1, ..., N$*

$$\{\tau \leq n\} \quad \in \quad \mathcal{F}^S(n)).$$

**Remark 3.8.2.** *Note that the notion of a stopping time is tied to a filtration (similar to the notion of a martingale). It could happen that $\tau$ is a stopping time w.r.t a filtration $\mathcal{F}^{S_1}(n)$ but not a stopping time w.r.t another filtration $\mathcal{F}^{S_2}(n)$.*

*Some propreties*

1. If $\tau$ is a $\mathcal{F}(n)$ stopping time then $\{\tau < n\} = \{\tau \leq n-1\} \in \mathcal{F}(n-1) \subseteq \mathcal{F}(n)$, we have

$$\{\tau \geq n\} \quad = \quad \{\tau < n\}^c \in \mathcal{F}(n)$$

Hence

$$\{\tau = n\} = \{\tau \leq n\} \cap \{\tau \geq n\} \in \mathcal{F}(n).$$

Conversely if $\{\tau = n\} \in \mathcal{F}(n)$ *for all* $n$ then $\{\tau \leq n\} = \cup_{i=0}^{n}\{\tau = i\} \in \mathcal{F}(n)$, for all $n$ as well. So we can use either conditions: $\{\tau = n\} \in \mathcal{F}(n)$ or $\{\tau \leq n\} \in \mathcal{F}(n)$ as definition for stopping time in discrete time.

2. (*Important*) The event $\{\tau = 0\}$ has probability $0$ or $1$. The reason is $\{\tau = 0\} \in \mathcal{F}_0^S$, but $\mathcal{F}_0^S$ is the information of the asset up to today, which we assume know. Thus we must know whether or not $\{\tau = 0\}$ with probability 1 (simply put: we must know whether we exercise the option today or not).

3. Let $\tau_1, \tau_2$ be stopping times w.r.t. $\mathcal{F}^S(n)$. Then $\min(\tau_1, \tau_2)$ and $\max(\tau_1, \tau_2)$ are stopping times w.r.t $\mathcal{F}^S(n)$.

### 3.8.4 Pricing an American put option in 1 period model

*Mathematical definition*

For any time $k, 0 \leq k \leq N$, let $V_k$ denote the price of an American put option with strike $K$ and expiration $N$, *provided the option has not been exercised.* This is equivalent to say $V_k$ is the price of an American put option that allows you to choose an exercise time from $k$ to $N$. From our discussion above, we conclude

$$V_k = \max_{\tau:\ \tau \text{ stopping time taking values from } k \text{ to } N} E^Q\Big(e^{-(\tau-k)\Delta T}(K - S_\tau)^+ | \mathcal{F}_k^S\Big).$$

This will be **our mathematical definition** of the value of an American option at time $k$, *if it has not been exercised.*

We call the stopping time that achieves the maximum in the above definition *the optimal stopping time* for $V_k$, from here on denoted as $\tau_k^*$. We will show that it exists and characterize it in the later subsection.

*Pricing in 1 period*

1. We will now investigate how to solve for $V_k, k = 0, 1, \cdots, N$ in our discrete binomial model. First, suppose that $N = 0$, that is the option exires today, then it is clear that

$$V_0 = (K - S_0)^+.$$

3. Now suppose $N = 1$. Then $V_1 = (K - S_1)^+$. What about $V_0$? Note that if a stopping time $\tau \in \mathcal{F}_k, k = 0, 1$ then $P(\tau = 0) = 0$ or 1, as well as $P(\tau = 1) = 0$ or 1.

In other words, at time 0 we have 2 choices: exercise the option right away or exercise the option tomorrow. (Any strategy that says with 40% we exercise today and 50% we exercise tomorrow etc. will not be optimal, as we will show).

Let's denote $V_0^1 = E^Q(e^{-r\Delta T}(K - S_1)^+)$ to be the expected return of the option if we go with the strategy exercising tomorrow. Also $V_0^0 = (K - S_0)^+$ is the return we get if we go with the strategy exercising today. Then our decision is clear: exercise today if $V_0^0 \geq V_0^1$ and exercise tomorrow if $V_0^1 > V_0^0$. And the value (the price) of the American option today is

$$V_0 = \max(V_0^1, V_0^0).$$

4. Mathematically we can express the above as followed: define $\tau^*$ as

$$\tau^* := \mathbf{1}_{\{V_0^1 > V_0^0\}}.$$

Then (you can check) $\tau^*$ is a stopping time (actually it is deterministic) and

$$V_0 = E^Q(e^{-r(\Delta T)\tau^*}(K - S_{\tau^*})^+).$$

5. Why is the mixed strategy (40% today, 60% tomorrow) not optimal? Let $\tau$ be a random variable such that $P(\tau = 0) = .4, P(\tau = 1) = .6$ and $\tau$ is independent of $S$. Then the return of this strategy is

$$\begin{aligned} V_0^\tau &= E^Q(e^{-r(\Delta T)\tau}(K - S_\tau)^+) \\ &= E^Q\Big[E^Q(e^{-r(\Delta T)\tau}(K - S_\tau)^+ | \tau)\Big] \end{aligned}$$

Since $\tau$ is independent of $S$,

$$\begin{aligned} E^Q(e^{-r(\Delta T)\tau}(K - S_\tau)^+ | \tau) &= E^Q(e^{-r(\Delta T)}(K - S_1)^+)\mathbf{1}_{\tau=1} + (K - S_0^+)\mathbf{1}_{\tau=0} \\ &= V_0^1 \mathbf{1}_{\tau=1} + V_0^0 \mathbf{1}_{\tau=0}. \end{aligned}$$

Thus

$$V_0^\tau = V_0^1 P(\tau = 1) + V_0^0 P(\tau = 0) \leq \max(V_0^1, V_0^0)$$

and hence $\tau$ is not an optimal strategy.

### 3.8.5 Pricing an American put option in multi-period model

*An intuitive approach*

We want to find a formula for $V_k, 0 \leq k \leq N$. When $k = N$, the answer is easy: $V_N = (K - S_N)^+$. This is because at the expiration, exercising the option is always no worse than letting it expires worthless.

At $k = N-1$, the option holder has 2 choices: either exercise immediately, or wait to exercise at time $N$. Which way is better? If she exercises immediately, she'll get a pay off of $(K - S_{N-1})^+$. If she waits until time $N$ to exercise, the *risk neutral expected payoff* of this choice at time $N-1$ to her is

$$E^Q(e^{-r\Delta T}(K - S_N)^+|\mathcal{F}^S_{N-1}).$$

She can compare between these two values and decide her strategy depending on which one yields a better payoff. It also follows that at time $N-1$ the American option is worth

$$V_{N-1} = \max\Big\{(K - S_{N-1})^+, E^Q(e^{-r\Delta T}(K - S_N)^+|\mathcal{F}^S_{N-1})\Big\}.$$

Generally, at a time $k$, she has 2 choices: exercise immediately and receive $(K - S_k)^+$, or wait until time $k+1$ and get the pay off $V_{k+1}$. How? By following the optimal stopping strategy starting at time $k+1$ once she is at time $k+1$. Thus, at time $k$ the American option is worth

$$V_k = \max\Big\{(K - S_k)^+, E^Q(e^{-r\Delta T}V_{k+1}|\mathcal{F}^S_k)\Big\}.$$

This approach is intuitive, but needs justification on why it is correct. The reason is by definition, $V_k$ is the best risk neutral expected payoff among all stopping time strategies starting at time $k$. To reach this value, one may imagine the option holder searching among all strategies available to her at time $k$ and choose the optimal one among those. The assertion that

$$V_k = \max\Big\{(K - S_k)^+, E^Q(e^{-r\Delta T}V_{k+1}|\mathcal{F}^S_k)\Big\}$$

is equivalent to saying **searching optimally from time $k$ to $N$ gives the same payoff as searching optimally from time $k$ to $k+1$, and then search optimally from time $k+1$ to $N$.**

By searching optimally from $k$ to $k+1$, and then search optimally from $k+1$ to $N$ we mean choose a strategy that realizes the best payoff comparing if one stops at $k$ or stops at $k+1$, with the payoff at $k+1$ equals to the optimal payoff starting from $k+1$ to $N$.

But this assertion is not obvious. Stating that it is true amounts to proving *the dynamic programming principle* for $V_k$.

*The dynamic programming principle*

We now develop a formula for $V_k$. The claim is that

$$V_k = \max\Big\{(K - S_k)^+, E^Q(e^{-r\Delta T}V_{k+1}|\mathcal{F}^S_k)\Big\}. \tag{3.9}$$

For the binomial model, this translates to

$$V_k(\omega) = \max\Big\{(K - S_k)^+, e^{-rT}[qV_k(\omega u) + (1-q)V_k(\omega d)]\Big\}. \tag{3.10}$$

**Formula (3.9) is the basis for the intuitive approach, or the tree pricing method we described above.** Now we prove it.

*Proof of* (3.9)

Case 1: At $k = N-1$ Observe that, if $\tau$ stopping time taking values from $N-1$ to $N$, then by a similar argument from the previous subsection either $P(\tau = N-1|\mathcal{F}^S_{N-1}) = 1$ or $P(\tau = N-1|\mathcal{F}^S_{N-1}) = 1$.

Denote

$$V^N_{N-1} = E^Q(e^{-r\Delta T}(K - S_N)^+|\mathcal{F}^S_{N-1}).$$

Then by definition, we have

$$V_{N-1} = \max\Big\{(K - S_{N-1})^+, V^N_{N-1}\Big\},$$

which is (3.9).

Define

$$\tau^*_{N-1} := N\mathbf{1}_{\{V^N_{N-1} > (K-S_{N-1})^+\}} + (N-1)\mathbf{1}_{\{V^N_{N-1} \leq (K-S_{N-1})^+\}},$$

then we can check that $\tau^*_{N-1}$ is a stopping time and

$$V_{N-1} = E^Q(e^{-r(\Delta T)\tau^*_{N-1}}(K - S_{\tau^*_{N-1}})^+)|\mathcal{F}^S_{N-1}),$$

as in the 1 period model. More importantly,

$$\begin{aligned} V_{\tau^*_{N-1}} &= V_N\mathbf{1}_{\{V^N_{N-1} > (K-S_{N-1})^+\}} + V_{N-1}\mathbf{1}_{\{V^N_{N-1} \leq (K-S_{N-1})^+\}} \\ &= (K - S_N)^+\mathbf{1}_{\{\tau^*_{N-1}=N\}} + (K - S_{N-1})^+\mathbf{1}_{\{\tau^*_{N-1}=N-1\}} = (S_{\tau^*_{N-1}} - K)^+. \end{aligned}$$

**This is very important**: at the optimal exercise time, the value of the option is equal to the excercise value. We will show this is true for general $k$.

Case 2: At a general $k \leq N-1$

We proceed by induction. Suppose (3.9) is true at step $k+1$ and the optimal stopping time $\tau^*_{k+1}$ exists and the following relation holds:

$$V_{\tau^*_{k+1}} = (S_{\tau^*_{k+1}} - K)^+.$$

Recall that by definition

$$V_k = \max_{\tau:\ \tau \text{ stopping time taking values from } k \text{ to } N} E^Q\Big(e^{-r(\tau-k)\Delta T}(K - S_\tau)^+|\mathcal{F}^S_k\Big).$$

Again observe that, if $\tau$ is a stopping time taking values from $k$ to $N$, then either $P(\tau = k|\mathcal{F}_k^S) = 1$ or $P(\tau > k|\mathcal{F}_k^S) = 1$. The set of these two types of stoping times are mutually exclusive. Therefore, if $X$ is a function of $\tau$

$$\max_{\tau:\, \tau \text{ values from } k \text{ to } N} E(X(\tau)|\mathcal{F}_k^S) = \max\Big\{X_k, \max_{\tau:\, \tau \text{ values from } k+1 \text{ to } N} E(X(\tau)|\mathcal{F}_k^S)\Big\}.$$

Replacing $X(\tau)$ with $e^{-(\tau-k)\Delta T}(K - S_\tau)^+$ we have

$$E(\Big(e^{-(\tau-k)\Delta T}(K - S_\tau)^+|\mathcal{F}_k^S) = E\Big(e^{-r\Delta T}E(\Big(e^{-r(\tau-(k+1))\Delta T}(K - S_\tau)^+|\mathcal{F}_k^S)\Big)$$

and for any $\tau$ taking values from $k+1$ to $N$

$$\begin{aligned}&E\Big(e^{-r\Delta T}E(\Big(e^{-r(\tau-(k+1))\Delta T}(K - S_\tau)^+|\mathcal{F}_k^S)\Big) \leq \\ &E\Big(e^{-r\Delta T} \max_{\tau:\, \tau \text{ values from } k+1 \text{ to } N} E(\Big(e^{-r(\tau-(k+1))\Delta T}(K - S_\tau)^+|\mathcal{F}_k^S)\Big) = E\Big(e^{-r\Delta T}V_{k+1}\Big).\end{aligned}$$

It follows that

$$V_k \leq \max\Big\{(K - S_k)^+, E^Q(e^{-r\Delta T}V_{k+1}|\mathcal{F}_k^S)\Big\}.$$

On the other hand, it is clear that

$$V_k \geq (K - S_k)^+.$$

Moreover,

$$\begin{aligned}E^Q(e^{-r\Delta T}V_{k+1}|\mathcal{F}_k^S) &= E^Q\Big(e^{-r\Delta T}E^Q[e^{-r\Delta(\tau_{k+1}^*-(k+1))}(K - S_{\tau_{k+1}^*})^+|\mathcal{F}_{k+1}^S]|\mathcal{F}_k^S\Big) \\ &= E^Q\Big(e^{-r\Delta T(\tau_{k+1}^*-k)}(K - S_{\tau_{k+1}^*})^+|\mathcal{F}_k^S\Big) \leq V_k.\end{aligned}$$

Thus

$$V_k \geq \max\Big\{(K - S_k)^+, E^Q(e^{-r\Delta T}V_{k+1}|\mathcal{F}_k^S)\Big\}.$$

Finally denoting $V_k^{k+1} = E^Q(e^{-r\Delta T}V_{k+1}|\mathcal{F}_k^S)$ and define

$$\tau_k^* := \tau_{k+1}^* \mathbf{1}_{\{V_k^{k+1} > (K-S_k)^+\}} + k\mathbf{1}_{\{V_k^{k+1} \leq (K-S_k)^+\}},$$

we see that

$$\begin{aligned}V_{\tau_k^*} &= V_{\tau^*k+1}\mathbf{1}_{\{V_k^{k+1} > (K-S_k)^+\}} + V_k\mathbf{1}_{\{V_k^{k+1} \leq (K-S_k)^+\}} \\ &= (K - S_{\tau_{k+1}^*})^+\mathbf{1}_{\{\tau_k^* = \tau_{k+1}^*\}} + (K - S_k)^+\mathbf{1}_{\{\tau_k^* = k\}} = (S_{\tau_k^*} - K)^+,\end{aligned}$$

completing the induction step.

### 3.9 General approach for a discrete optimal stopping problem

In this section, we generalize the optimal stopping problem to other contexts than just the American options. In particular, consider the following two examples:

**Example 3.9.1.** *Consider a game where we toss a fair die. If it's 1 to 5 then we win 1 dollar and have the option to toss the die again. If it's 6 then we lose all the winning we have so far and the game stops (the winning is cumulative, that is if you toss 5 times and the die never shows a 6 then you win 5 dollars). What is the optimal strategy to play this game? What is the fair price to play this game?*

**Example 3.9.2.** *Consider a game where we toss a fair die three times. In the first two turns, if we roll a value $n$ other than 6 we receive $n$ dollars. We can then choose to continue rolling or stop the game. If we continue rolling we forego the $n$ dollar we receive from the turn. If we roll a 6 the game stops and we recieve nothing. The pay off on the last turn is similar to the first two, except that the game stops no matter what value we get from the roll. What is the optimal strategy to play this game? What is the fair price to play this game?*

Both of these problems are of the optimal stopping types. The first one is a perpetual problem (because it can potentially be played infinitely) and the second has a finite horizon. We give a general approach to both problems. Note that this approach assume a risk netral player. That is we assume that given a random payment $X$, the player is indifferent between receiving $X$ or $E(X)$.

*General approach for a finite horizon problem*

Given a finite horizon optimal stopping problem in a discrete setting there are two main considerations:

a. The continuation decision : Denote $V_n^{cont}$ to be the payoff the player would receive if they continue to play at time $n$ and $V_n^{stop}$ to be the payoff the player would receive if they stop at time $n$. Note that $V_n^{cont}$ is random from the perspective at time $n$ while $V_n^{stop}$ is known at time $n$. Lastly let $V_n$ to be the value of the game at time $n$. Then

$$V_n = \max(V_n^{stop}, E(V_n^{cont}|\mathcal{F}_n)).$$

The player would continue if $E(V_n^{cont}|\mathcal{F}_n) \geq V_n^{stop}$ and stop otherwise. An important observation here is that if the player indeed decides to continue at time $n$, then he would / should play optimally at time $n+1$ as well. That is $V_n^{cont} = V_{n+1}$.

b. The terminal condition : Start from the terminal condition where the game value is known and work backward using the continuation decision.

We use these two principles to solve example (3.9.2). The terminal condition is $V_3 = n$ if the die rolls $n \leq 5$ and $V_3 = 0$ if the die rolls 6. At $n = 2$, we see that $V_2^{cont} = V_3$ and $V_2^{stop} = n$ if the die rolls $n \leq 5$ and $V_2^{stop} = 0$ if the die rolls 6. $E(V_2^{cont}) = \frac{15}{6}$ and thus we stop if we roll a 4 or 5 or 6 on turn 2 and continue otherwise. We also see that $V_2 = \frac{15}{6}$

with probability 1/2 and $V_2 = 4$ or 5 with probability 1/6 each and $V_2 = 0$ with probability 1/6. At turn 1, $V_1^{cont} = V_2$ and $E(V_1^{cont}) = \frac{33}{12}$. Thus we would continue at time 1 if we roll 1 or 2 and stops if we roll a 3 or higher. Finally the value of the game is $E(V_1) = \frac{105}{36}$.

*General approach for a perpetual problem*

For a perpetual optimal stopping problem, we do not have a terminal condition to work with. Nevertheless, there are also three main considerations for a perpetual optimal stopping problem:

a. Expressing the value of the game as a function of the state of the underlying: Because there is no finite horizion, denoting the value of the game as a function of time (as we are used to so far) is not a good approach. In stead, we denote $V(n)$ (in contrast to $V_n$) as the game value when the underlying state has reached state $n$. For example, in (3.9.1) it is better to denote $V(n)$ as the value of the game when the player has winned $n$ dollars. The value of the game at the beginning is $V(0)$ as he starts the game with 0 winning.

b. The stopping condition : under what condition should the player stop playing the game? The general principle is again that he will stop if the current game value is higher than the expectation of the continuation game value. That is, we have

$$V(n) = \max(V^{stop}(n), E(V^{cont}(n)|\mathcal{F}(n)))$$

and the player would continue if $E(V^{cont}(n)|\mathcal{F}(n)) \geq V^{stop}(n)$ and stop otherwise. $\mathcal{F}(n)$ here denotes the condition of the system at state $n$. Note that $V^{cont}(n)$ is NOT necessarily $V(n+1)$ because the system does not necessarily move from state $n$ to state $n+1$ in the next step. For example, in (3.9.1) if $n$ denotes the winning of the player then in the next step he can move to states $n+1, n+2, \cdots, n+5$ and state 0 with probability 1/6 each. We use $V^{cont}(n)$ to express the value of the game at the possible next states from state $n$, whatever they are.

From this consideration, we can divide the states of the system into two sets: the continuation set and the stopped set. That is we can come up with criteria such that if the player reaches a certain state, he will know whether he should stop or continue.

c. Work backward from the stopped set: It usually is easy to figure out what $V(n)$ is for $n$ in the stopped set. For $n$ in the continuation set, see which states can be reached from $n$. Find those $n$ in the continuation set such that the immediate state from $n$ ends up in the stopped set. Call this the first layer. For such $n$ we can find $V(n) = E(V^{cont}(n)|\mathcal{F}(n)))$ because we know the value of the game at the next step. Next find those $n$ in the continuation set such that the immediate state from $n$ ends up in the stopped set or the first layer. Repeat this process until we find $V(n_0)$ where $n_0$ is the state at the beginning.

We use these three principles to solve example (3.9.2). Let $V(n)$ be the game value after the player as winned $n$ dollars. His stopping condition is when

$$V(n) > \sum_{k=1}^{5}(V(n)+k) = 5/6V(n) + 15/6,$$

that is if $V(n) > 15$. If the player stops, it is clear that $V(n) = n$. Thus he will stop if his winnning reaches 16 or above. Because after each roll he can only increases his fortune by $n = 1, 2, \cdots, 5$, and highest winning where he continues to roll is 15, the player will actually stop if his fortune reaches 16, 17, 18, 19, 20. That is the stopped set is $\{16, 17, 18, 19, 20\}$ and the continuation set is $\{0, 1, \cdots, 15\}$. We have $V(16) = 16, V(17) = 17, V(18) = 18, V(19) = 19, V(20) = 20$. The first layer in the continuation set is $n = 15$ since immediately after 15 the player will end up in the stopped set. We have

$$V(15) = V^{cont}(15) = \frac{1}{6}(V(16) + V(17) + \cdots + V(20) = \frac{1}{6}(16 + 17 + \cdots + 20).$$

Similary we can find $V(14)$

$$V(14) = V^{cont}(14) = \frac{1}{6}(V(15) + V(16) + \cdots + V(19) = \frac{1}{6}(V(15) + 16 + \cdots + 19),$$

where for $V(15)$ we plug in the value found above. Continue similarly we can find $V(0)$, the explicit value of which we'll skip.

### 3.10 Exercises

We consider the following formulation for the binominal model:

$$S_k = S_0 X_1 X_2 ... X_k, 1 \leq k \leq n$$

where for $0 < d \leq e^{r\tau} \leq u$, $X_i$ are i.i.d. with distribution

$$\begin{aligned} X_i &= u \text{ with probability } q \\ X_i &= d \text{ with probability } 1-q. \end{aligned}$$

We denote $Q$ as the risk neutral probability measure. We also take the length of one period, $\Delta T$ to be 1 and the risk free interest rate to be a constant $r$.

1. Suppose $r = 0$. Compute

$$E^Q\big((S_5 - S_3)^+ | S_3\big).$$

2. The Put-Call parity principle says: Holding a long position on a European Call Option and a short position on a European Put Option is the same as holding a long position on a Forward Contract (on the same stock $S$, with the same expiration date $n$ and strike price $K$). Suppose $S$ follows the multi-period Binomial model. h a) Express the Put-Call parity principle in terms of $V^{\text{put}}$, $V^{\text{call}}$ and $V^{\text{forward}}$.

b) Prove the Put-Call parity principle.

Consider the multiperiod model with $N = 2, S_0 = 8$. From any given period, the stock can either double its price or half its price in the next period (ask for a picture if this description is not clear). Also suppose the interest rate is 0.

3. Consider the multiperiod model with $N = 3, S_0 = 8$. From any given period, the stock can either double its price or half its price in the next period. Also suppose the interest rate is 0.

a. Find the price at time period 1, when the stock goes up, that is $V_1(u)$ of a down and out put option with barrier $L = 6$, strike $K = 40$ and expiration $N = 3$.

b. Find the price at time period 0 of the look back option on this stock with expiration $N = 3$. Explicit number is not necessary.

4. Consider a game where we toss a fair die. If it's 1 to 5 then we win 1 dollar and have the option to toss the die again. If it's 6 then we lose all the winning we have so far and the game stops (the winning is cumulative, that is if we toss 5 times and the die never shows a 6 then we win 5 dollars).

a. What is the optimal strategy to play this game?

b. What is the fair price to play this game?

5. Again consider the multiperiod model with $S_0 = 8$. From any given period, the stock can either double its price or half its price in the next period. Also suppose the interest rate is 0.

a. Use the replicating portfolio approach to find the price $V_0$ of a European call option with strike 10 and expiration $N = 2$ on this stock.

b. Use the replicating portfolio approach to find the price $V_0$ of a European put option with strike 20 and expiration $N = 2$ on this stock.

6. A financial math student, Mr. Solve-alot wants to try out as a rock artist. His first single, titled “I like Big Math" has just debuted. He plans to sell the single for 8 dollars during the first week. During the second week, he’ll either double or half the price of the first week, depending on the response. During the third week, he’ll either double or half the price of the second week, again depending on the response. Mathematically we model the price of his single as followed:

$$\begin{aligned} S_1 &= 8 \\ S_2 &= 8X_1 \\ S_3 &= 8X_1X_2, \end{aligned}$$

where $X_1, X_2$ are i.i.d with distribution $P(X_i = 2) = q, P(X_i = 1/2) = 1-q$. Assume that the interest rate $r = 0$.

a. Remembering his lessons, Mr. Solve-alot wants to look at his price movement in a risk neutral way. He also recalls that the risk neutral measure in this case is the $q$ so that $E^Q(X_1) = 1$. Determine this $q$.

b. What is the risk neutral probability that during the 3rd week, the single will sell for more than or equal to 8 dollars?

# CHAPTER 4 Discrete time martingales and fundamental theorems of asset pricing

## 4.1 Introduction

In this note we discuss the two fundamental theorems of asset pricing. The mathematical tool for discussion is martingale theory in discrete time. We will define the notion of martingales, and show that under the risk neutral measure, the discounted asset price is a martingale. By our pricing formula, the discounted value process of a non-American financial derivative is also a martingale under the risk neutral measure. This is used to prove the first fundamental theorem of asset pricing. Under the uniqueness of the risk neutral measure, we show the existence of the hedging portfolio for non-American financial derivatives. Finally, we show the existence of the hedging portfolio for American put option, by characterizing it as a super-martingale under the risk neutral measure.

## 4.2 Martingale in discrete time

A process $V_k$ is a martingale with respect to the filtration $\mathcal{F}_k^S$ under a probability measure $P$ if:

a. $V_k \in \mathcal{F}_k^S$ for all $k$.

b. For all $n \geq m$, $E(V_n|\mathcal{F}_m^S) = V_m$.

Remark:

1. Condition a means that each $V_k$ is a function of $S_0, S_1, \cdots, S_k$. This is consistent with our intuition that the value of the financial derivative should only depend on the historical price of the underlying asset. observe that $S_{k+1} \notin \mathcal{F}_k^S$ in our binomial model.

2. Condition b is the martingale condition. The expectation $E$ is taken under the probability $P$. This is essential: if we change the measure $P$, this condition may not hold. Thus $V_k$ can be that a process is a martingale under some measure but not under some other measure (think about the risk neutral meassure for example).

Similarly, $V_k$ is a sub (super)-martingale with respect to the filtration $\mathcal{F}_k^S$ under a probability measure $P$ if:

a. $V_k \in \mathcal{F}_k^S$ for all $k$.

b. For all $n \geq m$, $E(V_n|\mathcal{F}_m^S) \geq (\leq) V_m$.

3. Sometimes we just say $S_k$ is a martingale (under probability $P$). Then it is understood that the filtration is $S_k$'s own filtration ($\mathcal{F}_k^S$).

### 4.2.1 Some examples

The following are the most important examples we encountered so far:

1. The discounted stock price $e^{-rk\Delta_T} S_k$ is a martingale with respect to $\mathcal{F}_k^S$ under the risk neutral measure $Q$ (but not neccessarily under the physical measure $P$).

2. The discounted value of a European option $e^{-rk\Delta_T} V_k$ is a martingale with respect to $\mathcal{F}_k^S$ under the risk neutral measure $Q$.

3. The discounted value of an American option $e^{-rk\Delta_T} V_k$ is a super-martingale with respect to $\mathcal{F}_k^S$ under the risk neutral measure $Q$.

## 4.3 The first fundamental theorem of asset pricing

### 4.3.1 Betting against a martingale

A martingale is essentially a model for a fair game. First note that if $V_k$ is a martingale with respect to $\mathcal{F}_k^S$ under $P$ then $E(V_k) = E(V_{k-1}) = \cdots = E(V_0)$. Thus if we treat $V_k$ as our total earning when investing in $S$ then its expected earning is a constant in time if $V_k$ is a martingale. Actually something stronger is true: our expected earning based on $S$ at any time in the future, conditioned on the information up to the current time: $E(V_n|\mathcal{F}_m^S)$, is the same as our current earning based on $S$: $V_m$. This is what we think of as fair.

We can also look at the reverse direction. If $S_k$ is a martingale with respect to $\mathcal{F}_k$ then the total earning under any strategy we can form investing using $S_k$ is also a martingale, as long as our strategy at time $k$ only uses the information about $S$ up to time $k$ (this excludes insider trading for example). More specifically, let $\Delta_k$ be the number of shares we hold of $S$ at time $k$, then our net “winning" over the period $[k, k+1]$ is $\Delta_k(S_{k+1} - S_k)$ and our total winning up to a time $n$ is

$$\pi_n = \sum_{k=0}^{n-1} \Delta_k (S_{k+1} - S_k).$$

The notation $\pi$ is used specifically to represent the value of our portfolio of $S$ at time $n$. We have the following lemma:

**Lemma 4.3.1.** *If $\Delta_k \in \mathcal{F}_k^S$ and $S_k$ is also a martingale then $\pi_n$ is also a martingale with respect to $\mathcal{F}_n^S$.*

(From now on, we'll refer to any process $X_k$ that has the property $X_k \in \mathcal{F}_k^S$ as *being adapted to $\mathcal{F}_k^S$*.)

*Proof.* It suffices to show $E(\pi_{n+1}|\mathcal{F}_n^S) = \pi_n$ (Why?). It is equivalent to show

$$E(\pi_{n+1} - \pi_n|\mathcal{F}_n^S) = 0.$$

But note that

$$\pi_{n+1} - \pi_n = \Delta_n(S_{n+1} - S_n).$$

Thus

$$
\begin{aligned}
E(\pi_{n+1} - \pi_n | \mathcal{F}_n^S) &= E(\Delta_n (S_{n+1} - S_n) | \mathcal{F}_n^S) \\
&= \Delta_n E(S_{n+1} - S_n | \mathcal{F}_n^S) = 0.
\end{aligned}
$$

Similarly, we can show that if $S_k$ is a sub (super) martingale then $\pi_n$ is a sub (super) martingale under similar conditions. A sub (super) martingale represents a game that favors one particular side of the game, either the house or the player.

### 4.3.2 Self-financing portfolio as a martingale

**Remark 4.3.2.** *In our model, $S_k$ is NOT a martingale, but $e^{-rk\Delta_T} S_k$ is. But one* **does not invest in a discounted stock price process in reality**. *What one does is investing in (possibly multiple) financial assets and a saving account. The corresponding result is that if the portfolio is* **self financing** *then its discoutned value process is also a martingale under the risk neutral measure. The following lemma states the result more precisely.*

**Lemma 4.3.3.** *Suppose at any time $k$ we hold $\Delta_k$ shares of the asset $S$ and $y_k$ in the saving account. Suppose $e^{-rk\Delta_T} S_k$ is a martingale. If $\Delta_k$ is adapted to $\mathcal{F}_k^S$ and the self-financing condition holds:*

$$
\pi_{k+1} = \Delta_{k+1} S_{k+1} + y_{k+1} = \Delta_k S_{k+1} + y_k e^{r\Delta T}. \tag{4.1}
$$

*or equivalently*

$$
\pi_{k+1} = \Delta_k S_{k+1} + e^{r\Delta T}(\pi_k - \Delta_k S_k). \tag{4.2}
$$

*then $e^{-rk\Delta_T} \pi_k$ is also a martingale under $\mathcal{F}_k^S$.*

*Proof.* Suppose $e^{-rk\Delta_T} S_k$ is a martingale. It is enough to show

$$
E^Q \left[ e^{-r\Delta T} \pi_{k+1} | \mathcal{F}_k^S \right] = \pi_k.
$$

From the self-financing condition, we have

$$
\begin{aligned}
E^Q \left[ e^{-r\Delta T} \pi_{k+1} \middle| \mathcal{F}_k^S \right] &= E^Q \left[ e^{-r\Delta T} \Delta_k S_{k+1} + (\pi_k - \Delta_k S_k) \middle| \mathcal{F}_k^S \right] \\
&= \Delta_k E^Q \left[ e^{-r\Delta T} S_{k+1} - S_k | \mathcal{F}_k^S \right] + \pi_k = \pi_k.
\end{aligned}
$$

**Remark 4.3.4.** *We do NOT have a similar conclusion in the self-financing portfolio case when $e^{-rk\Delta_T} S_k$ is a super (sub) martingale, that is correspondingly $e^{-rk\Delta_T} \pi_k$ is also a super (sub) martingale. The reason is the sign of $\Delta_k$ matters in this case. If we short an asset that is a super martingale (that is it decreases on average), then we're likely to make money in the future (that is the portfolio will be a sub-martingale). But if we long an asset* *that is a super martingale, then we're likely to lose money (that is the portfolio remains a super-martingale). The following calculation makes it clear:*

*Suppose $e^{-k\Delta_T} S_k$ is a super martingale. Then*

$$E^Q\Big[e^{-r\Delta T} S_{k+1} | \mathcal{F}_k^S\Big] \leq S_k.$$

*From the self-financing condition, we have*

$$\begin{aligned} E^Q\Big[e^{-r\Delta T}\pi_{k+1}\Big|\mathcal{F}_k^S\Big] &= E^Q\Big[e^{-r\Delta T}\Delta_k S_{k+1} + (\pi_k - \Delta_k S_k)\Big|\mathcal{F}_k^S\Big] \\ &= \Delta_k E^Q\Big[e^{-r\Delta T} S_{k+1} - S_k | \mathcal{F}_k^S\Big] + \pi_k \\ &\leq \pi_k, \text{ if } \Delta_k \geq 0 \\ &\geq \pi_k, \text{ if } \Delta_k \leq 0. \end{aligned}$$

### 4.3.3 Market with more than 1 assets

The result about self-financing portfolio also holds in market with more than 1 asset $S^1, S^2, \cdots, S^m$. We just have to generalize the self-financing condition to:

$$\pi_{k+1} = \sum_i \Delta_k^i S_{k+1}^i + e^{r\Delta T}(\pi_k - \sum_i \Delta_k^i S_k^i).$$

Our result is

**Lemma 4.3.5.** *Suppose at any time $k$ we hold $\Delta_k^i$ shares of asset $S^i$ and $y_k$ in cash. Suppose $e^{-rk\Delta_T} S_k^i$ is a martingale for all $i$. If $\Delta_k^i$ is adapted to $\mathcal{F}_k^S$ and the self-financing condition holds:*

$$\pi_{k+1} = \sum_i \Delta_k^i S_{k+1}^i + e^{r\Delta T}(\pi_k - \sum_i \Delta_k^i S_k^i). \tag{4.3}$$

*then $e^{-rk\Delta_T}\pi_k$ is also a martingale under $\mathcal{F}_k^S$.*

*Proof.* Suppose $e^{-rk\Delta_T} S_k^i$ is a martingale for all $i$. It is enough to show

$$E^Q\Big[e^{-r\Delta T}\pi_{k+1} | \mathcal{F}_k^S\Big] = \pi_k.$$

From the self-financing condition, we have

$$\begin{aligned} E^Q\Big[e^{-r\Delta T}\pi_{k+1}\Big|\mathcal{F}_k^S\Big] &= E^Q\Big[e^{-r\Delta T}\Big\{\sum_i \Delta_k^i S_{k+1}^i\Big\} + (\pi_k - \sum_i \Delta_k^i S_k^i)\Big|\mathcal{F}_k^S\Big] \\ &= E^Q\Big[\sum_i \Delta_k^i\Big\{e^{-r\Delta T} S_{k+1}^i - S_k^i\Big\}\Big|\mathcal{F}_k^S\Big] + \pi_k \\ &= \sum_i \Delta_k^i E^Q\Big[e^{-r\Delta T} S_{k+1}^i - S_k^i | \mathcal{F}_k^S\Big] + \pi_k = \pi_k. \end{aligned}$$

### 4.3.4 The first fundamental theorem of asset pricing

We are in the position to prove the first version of the fundamental theorem of asset pricing for discrete time model.

**Theorem 4.3.6.** *Let a market have $m$ risky assets $S^1, S^2, \cdots, S^m$. Suppose an equivalent risk neutral measure $Q$ exists, that is $Q$ is equivalent to $P$ and*

$$S_k^i = E^Q(e^{-r\Delta T} S_{k+1}^i | \mathcal{F}_k^S), i = 1, \cdots, m.$$

*Suppose additionally that all derivatives that make payment $V_N$ at time $N$ satisfy*

$$V_k = E^Q(e^{-r\Delta T(N-k)} V_N | \mathcal{F}_k^S),$$

*then there is no self-financing portfolio consisting of $S^i, V$ and the saving account such that $\pi_0 = 0$ and $P(\pi_l \geq 0) = 1, P(\pi_l > 0) > 0$ for $0 < l \leq N$. That is the market is arbitrage free.*

*Proof.*
Suppose at time $0$ we hold $\Delta^i$ shares of asset $S^i$ and $y$ dollars in cash, as well as $\Delta$ shares of $V$ such that $\pi_0 = 0$. Then because $e^{-rk\Delta T} V_k$ is also a martingale under $Q$, we conclude $e^{-rk\Delta T}\pi_k$ is a martingale by Lemma (4.3.5). Thus

$$E^Q\left(e^{-rl\Delta T}\pi_l\right) = \pi_0 = 0.$$

Now since we're in a discrete space model, there are only finitely many outcomes $\omega_1, \omega_2, \cdots, \omega_n$ at time $l$. Let $P^Q(\omega_i) = q_i$ and note that by the equivalence condition, $q_i > 0, \forall i$. Then

$$q_1\pi_l(\omega_1) + q_2\pi_l(\omega_2) + \cdots + q_n\pi_l(\omega_n) = 0.$$

Thus it must follow that either $\pi_l(\omega_i) = 0, \forall i$ or there exists $i$ such that $\pi_l(\omega_i) < 0$.

**Theorem 4.3.7.** *Let a market have $m$ risky assets $S^1, S^2, \cdots, S^m$. Suppose that there is no arbitrage opportunity in the market. Then a risk neutral measure $Q$ exists, that is*

$$S_k^i = E^Q(e^{-r\Delta T} S_{k+1}^i | \mathcal{F}_k^S), i = 1, \cdots, m.$$

*Moreover, all derivatives that make payment $V_N$ at time $N$ must satisfy*

$$V_k = E^Q(e^{-r\Delta T(N-k)} V_N | \mathcal{F}_k^S),$$

**Remark 4.3.8.** *Theorem* (4.3.6) *and* (4.3.7) *together can be stated simply as a market is arbitrage free if and only if an equivalent risk neutral measure exists.*

*Proof.* Without loss of generality we prove the statement for $N = 1$. We restate the statements of Theorem (4.3.6) and (4.3.7) into the following:

There exists a vector $Q = [q_1, q_2, \cdots, q_n]^T > 0$ so that

$$\begin{bmatrix} e^{-r\Delta T}S_1^1(\omega_1) & e^{-r\Delta T}S_1^1(\omega_2) & \cdots & e^{-r\Delta T}S_1^1(\omega_n) \\ e^{-r\Delta T}S_1^2(\omega_1) & e^{-r\Delta T}S_1^2(\omega_2) & \cdots & e^{-r\Delta T}S_1^2(\omega_n) \\ \cdots & & & \\ e^{-r\Delta T}S_1^m(\omega_1) & e^{-r\Delta T}S_1^m(\omega_2) & \cdots & e^{-r\Delta T}S_1^m(\omega_n) \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \cdots \\ q_n \end{bmatrix} = \begin{bmatrix} S_0^1 \\ S_0^2 \\ \cdots \\ S_0^m \\ 1 \end{bmatrix},$$

if and only if we cannot find a vector $\Delta = [\Delta_1, \Delta_2, \cdots, \Delta_m, \Delta_{m+1}]^T$ so that

$$\begin{bmatrix} S_0^1 & S_0^2 & \cdots & S_0^m & 1 \end{bmatrix} \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \cdots \\ \Delta_m \\ \Delta_{m+1} \end{bmatrix} = 0 \tag{4.4}$$

and

$$\begin{bmatrix} S_1^1(\omega_1) & S_1^2(\omega_1) & \cdots & S_1^m(\omega_1) & e^{r\Delta T} \\ S_1^1(\omega_2) & S_1^2(\omega_2) & \cdots & S_1^m(\omega_2) & e^{r\Delta T} \\ \cdots & & & & \\ S_1^1(\omega_n) & S_1^2(\omega_n) & \cdots & S_1^m(\omega_n) & e^{r\Delta T} \end{bmatrix} \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \cdots \\ \Delta_m \\ \Delta_{m+1} \end{bmatrix} \gneqq 0, \tag{4.5}$$

where by $z \gneqq 0$ we mean $z \geq 0$ and $g \neq 0$.

Conditions (4.4) and (4.5) can be combined into one statement: we cannot find vector $\Delta = [\Delta_1, \Delta_2, \cdots, \Delta_m]^T$ so that

$$\begin{bmatrix} S_1^1(\omega_1) - e^{r\Delta T}S_0^1 & S_1^2(\omega_1) - e^{r\Delta T}S_0^2 & \cdots & S_1^m(\omega_1) - e^{r\Delta T}S_0^m \\ S_1^1(\omega_2) - e^{r\Delta T}S_0^1 & S_1^2(\omega_2) - e^{r\Delta T}S_0^2 & \cdots & S_1^m(\omega_2) - e^{r\Delta T}S_0^m \\ \cdots & & & \\ S_1^1(\omega_n) - e^{r\Delta T}S_0^1 & S_1^2(\omega_n) - e^{r\Delta T}S_0^2 & \cdots & S_1^m(\omega_n) - e^{r\Delta T}S_0^m \end{bmatrix} \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \cdots \\ \Delta_m \end{bmatrix} \gneqq 0.$$

Stated in this way, this is a well-known result in linear programming, known as Stiemke's Theorem. It is as followed: Let $A$ be a $m \times n$ matrix. Then exactly one of the following system has a solution:

a. $y^T A \gneqq 0$ for some $y \in \mathbb{R}^m$

or

b. $Ax = 0, x > 0, x \in \mathbb{R}^n$.

Applying Stiemke's Theorem to our situation with

$$A^T = \begin{bmatrix} S_1^1(\omega_1) - e^{r\Delta T}S_0^1 & S_1^2(\omega_1) - e^{r\Delta T}S_0^2 & \cdots & S_1^m(\omega_1) - e^{r\Delta T}S_0^m \\ S_1^1(\omega_2) - e^{r\Delta T}S_0^1 & S_1^2(\omega_2) - e^{r\Delta T}S_0^2 & \cdots & S_1^m(\omega_2) - e^{r\Delta T}S_0^m \\ \cdots & & & \\ S_1^1(\omega_n) - e^{r\Delta T}S_0^1 & S_1^2(\omega_n) - e^{r\Delta T}S_0^2 & \cdots & S_1^m(\omega_n) - e^{r\Delta T}S_0^m \end{bmatrix},$$

we see that if $y^T A \gneqq 0$ does not have a solution (the market is arbitrage free) then we must be able to find $x > 0 \in \mathbb{R}^n$ so that $Ax = 0$. we can easily see that this means we can find a probability vector $Q$ (by normalizing x) so that for all $j = 1, \cdots, m$

$$\sum_i q_i(S^j(\omega_i) - e^{r\Delta T} S_0^j) = 0.$$

But then $Q$ is exactly the risk neutral measure. Conversely, if a risk neutral measure exists, then the system $Ax = 0$ has a positive solution. Thus we cannot solve the system $y^T A \gneqq 0$. That is there is no arbitrage opportunity in the market.

**Remark 4.3.9.** *In Theorem (4.3.6), we did not include the American option in the portfolio. The reason is the discounted value of an American option is a super-martingale in general and the direction of the discounted portfolio value is unclear, as explained in Remark (4.3.4). However, one should also expect that the inclusion of American options should not affect the arbitrage property of the market. This is indeed the case, if the option holder acts in an optimal way. The description of this situation is slightly more complicated, so we reserandom variablee a separate section to discuss it.*

### 4.3.5 Hedging portfolio as a pricing tool

Theorem (4.3.6) already states the pricing we must follow for any financial derivative if we want our market to be arbitrage free, whether or not we can find a replicating portfolio for the derivatives. We learned in Lecture 2b that we can also price a financial derivative by the replicating portfolio, if it exists. These two methods should be consistent, that is they should give the same price. The following Lemma confirms this is the case.

**Lemma 4.3.10.** *Let a market have $m$ risky assets $S^1, S^2, \cdots, S^m$. If a risk neutral measure $Q$ exists, that is*

$$S_k^i = E^Q(e^{-r\Delta T} S_{k+1}^i | \mathcal{F}_k^S), i = 1, \cdots, m.$$

*Consider a financial derivative $V$, whose replicating portfolio exists. That is at any time $0 \le k \le N$, we can find $\Delta_k^i$ shares of asset $S^i$ and $y_k$ dollars in cash such that*

$$\pi_k = \sum_i \Delta_k^i S_k^i + y_k = V_k,$$

*and the self-financing condition is satisfied:*

$$\pi_{k+1} = \sum_i \Delta_k^i S_{k+1}^i + e^{r\Delta T}(\pi_k - \sum_i \Delta_k^i S_k^i).$$

*Then $V_k = E^Q(e^{-r\Delta T(N-k)} V_N | \mathcal{F}_k^S), \forall k.$*

*Proof.* By Lemma (4.3.5), $e^{-rk\Delta T}\pi_k$ is a martingale. Therefore

$$V_k = \pi_k = E^Q(e^{-r(N-k)\Delta T}\pi_N | \mathcal{F}_k^S) = E^Q(e^{-r(N-k)\Delta T} V_N | \mathcal{F}_k^S).$$

## 4.4 The second fundamental theorem of asset pricing

**Theorem 4.4.1.** *Let a market have $m$ risky assets $S^1, S^2, \cdots, S^m$. If a risk neutral measure $Q$ exists, that is*

$$S^i_k = E^Q(e^{-r\Delta T} S^i_{k+1} | \mathcal{F}^S_k), i = 1, \cdots, m.$$

*and it is* **unique***, then every financial derivative that pays $V_N$ at time $N$ can be replicated and the market is arbitrage-free.*

We will first prove this theorem for the case $N = 1$ and then for general $N$. Because we're in a discrete space, there are $n$ possible outcomes, $\omega_1, \omega_2, \cdots, \omega_n$ at time $N = 1$. The replicating condition requires that we are able to find $\Delta^i, i = 1, \cdots, m$ and $y$ such that

$$\sum_{i=1}^{m} \Delta^i S^i_1(\omega_j) + y e^{r\Delta T} = V_1(\omega_j), j = 1, \cdots, n.$$

*Note that the above is a system of $n$ equations in $m+1$ variables*. The unique martingale measure condition says that there exists a **unique** positive solution $q_1, q_2, \cdots, q_n$ to the system of equations

$$\begin{aligned} \sum_{i=1}^{n} q_i e^{-r\Delta T} S^j_1(\omega_i) &= S^j_0, j = 1, \cdots, m \\ \sum_i q_i = 1. \end{aligned}$$

*Note that the above is a system of $m + 1$ equations in $n$ variables* (the variables are the $q_i$). The left hand side matrix in the first system is

$$\begin{bmatrix} S^1_1(\omega_1) & S^2_1(\omega_1) & \cdots & S^m_1(\omega_1) & e^{rT} \\ S^1_1(\omega_2) & S^2_1(\omega_2) & \cdots & S^m_1(\omega_2) & e^{rT} \\ \cdots & & & & \\ S^1_1(\omega_n) & S^2_1(\omega_n) & \cdots & S^m_1(\omega_n) & e^{rT} \end{bmatrix}.$$

It is equivalent to

$$A = \begin{bmatrix} e^{-r\Delta T} S^1_1(\omega_1) & e^{-r\Delta T} S^2_1(\omega_1) & \cdots & e^{-r\Delta T} S^m_1(\omega_1) & 1 \\ e^{-r\Delta T} S^1_1(\omega_2) & e^{-r\Delta T} S^2_1(\omega_2) & \cdots & e^{-r\Delta T} S^m_1(\omega_2) & 1 \\ \cdots & & & & \\ e^{-r\Delta T} S^1_1(\omega_n) & e^{-r\Delta T} S^2_1(\omega_n) & \cdots & e^{-r\Delta T} S^m_1(\omega_n) & 1 \end{bmatrix},$$

as far as existence of solution is concerned.

The left hand side matrix in the second system is

$$B = \begin{bmatrix} e^{-r\Delta T} S^1_1(\omega_1) & e^{-r\Delta T} S^1_1(\omega_2) & \cdots & e^{-r\Delta T} S^1_1(\omega_n) \\ e^{-r\Delta T} S^2_1(\omega_1) & e^{-r\Delta T} S^2_1(\omega_2) & \cdots & e^{-r\Delta T} S^2_1(\omega_n) \\ \cdots & & & \\ e^{-r\Delta T} S^m_1(\omega_1) & e^{-r\Delta T} S^m_1(\omega_2) & \cdots & e^{-r\Delta T} S^m_1(\omega_n) \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Note that $A = B^T$. Thus what we need to prove is the following Lemma

**Lemma 4.4.2.** *Let $A$ be a $m \times n$ matrix. Suppose that* **there exists** *a vector $b \in \mathbb{R}^m$ such that the equation $Ax = b$ has a unique solution. Then for* **any** *vector $c \in \mathbb{R}^n$, the equation $A^T x = c$ has a solution.*

This is a well-known result in linear algebra. We provide the proof for completeness.

*Proof.* If the equation $Ax = b$ has a unique solution then the equation $Ax = 0$ has a unique solution (and vice versa). The equation $Ax = 0$ has a unique solution if and only if the columns of $A$ are linearly independent. But the columns of $A$ are the rows of $A^T$. Thus the matrix $A^T$ has full row rank. Thus for any vector $c$, the equation $A^T x = c$ has a solution.

**Remark 4.4.3.** *The fact that the system $Bq = S_0$ has a positive solution* **was not used** *in the Lemma. In fact it is not needed. We have seen this in the example of market that is complete but not arbitrage free. The condition for $q_i > 0$ is used to assert that the market is arbitrage free, and provide a link between pricing using expectation under the risk neutral measure and using the replicating portfolio, as described in Lemma (4.3.10).*

## CHAPTER 5 Continuous time stochastic processes

Discrete time process, such as Markov chain, has the advantage of giving us clear intuition and analysis. On the other hand, if we model a real life process using discrete time model, it has the dis-advantage of restricting the possible action on the discrete time points. For example, using the binomial tree model for a stock price $S_k$ has the implication that an investor can only balance his or her portfolio based on $S$ at $k = 0, 1, \cdots N$ and cannot do so in between $k$ and $k + 1$. To overcome this restriction and to obtain a model that captures the intuition that we can take action at any time, we need to use continuous time stochatic process $S_t, 0 \leq t \leq T$. A very common way to model $S_t$ is by decomposing it into two components: signal and noise. Intuitively, in an ideal environment where there is no noise, $S_t$ will follow an Ordinary Differential Equation:

$$\frac{dS_t}{dt} = f(t, S_t). \tag{5.1}$$

In reality, the movement of $S_t$ over time is subject to random perturbation. Therefore, we need to add a noise term to the equation (5.1) to obtain

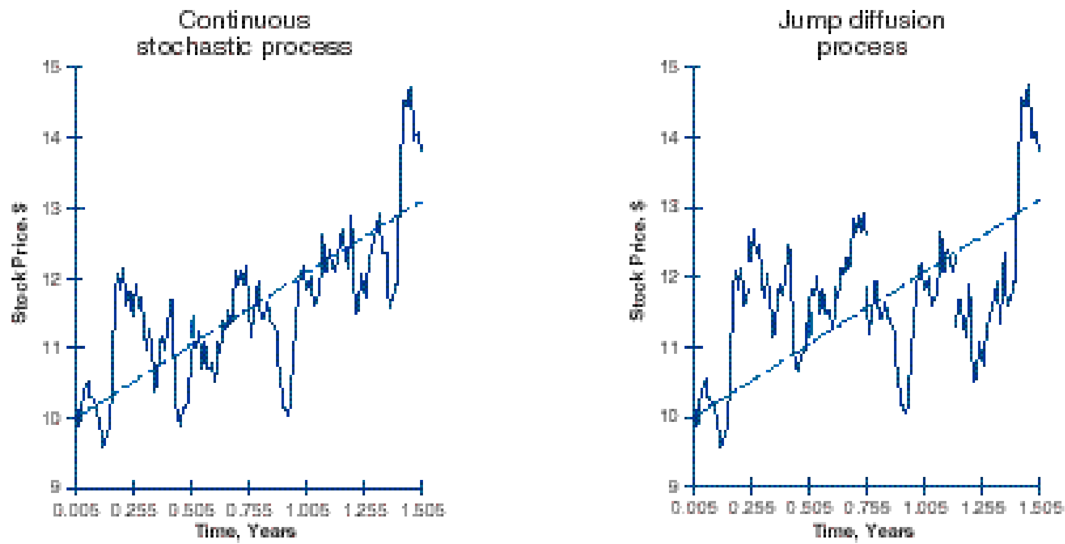
$$dS_t = f(t, S_t)dt + \text{“Noise"} . \tag{5.2}$$

There are typically two types of noise that are used when we model continuous time process: Brownian motion and Possion processes. These two types have specific interpretation when $S_t$ represents a stock price. Brownian motion represents the minute to minute fluctation that results from the continuous trading activities of $S$ from the investors. Poisson processes represent a shock (for example a market crash) that causes $S_t$ to jump in value. A typical model would incorporate both types of noise to capture various degree of change in the underlying process.

We note here that even though these two noises are “orthogonal" in the sense that the Brownian motion is continuous and the Poisson process is pure-jump, they belong to the same family known as the Levy processes. The Levy processes enjoy the properties of having independent and identically distributed increment. That is if $t_1 < t_2 < t_3 < t_4$ such that $t_2 - t_1 = t_4 - t_3$ and $S_t$ is a Levy process then

$$S_{t_4} - S_{t_3} \stackrel{d}{=} S_{t_2} - S_{t_1},$$

$S_{t_2} - S_{t_1}$ and $S_{t_4} - S_{t_3}$ are independent.




As we shall see, having the independent and identically distributed increment makes it nice and convenient to analyze the noise. Moreover, it also makes it easier for us to make sense out of the stochastic differential equation

$$dS_t = f(t, S_t)dt + \sigma(t, S_t) \text{ “Noise" } . \tag{5.3}$$

which is a generalization of the equation (5.2).

## 5.1 Discrete noise - Poisson process

### 5.1.1 Heuristics about Poisson process

We think of Poisson process as followed: suppose that we have an alarm clock that will ring after a *random* time $\tau$, where $\tau$ is exponentially distributed with some mean $\frac{1}{\lambda}$. We keep account of the value of the Poisson process at any time $t$ by the notation $N(t)$. At time 0, we set the alarm clock and set $N(0) = 0$. When the alarm rings, we increase the value of $N$ by 1, that is we set $N(\tau) = 1$ and repeat the whole process (i.e. we reset the alarm clock and increase the value of $N$ by 1 the next time the clock rings). The resulting process $N(t)$ is then a Poisson process with rate $\lambda$. We observe that the larger $\lambda$ is, the clock would be likely to ring sooner and the more jumps would likely happen in a given time interandom variableal [0,T]. It is also clear that $N(t)$ is constant in between the “ring" times.

### 5.1.2 Formal mathematical definition

a. $\tau$ (as a R.V.) is said to be exponentially distributed with rate $\lambda$ if it has the density

$$f(t) = \lambda e^{-\lambda t} \mathbf{1}_{(t \geq 0)}.$$

It follows that $E(\tau) = \frac{1}{\lambda}$ and $Var(\tau) = \frac{1}{\lambda^2}$. An important property of exponential random variable is the *memoryless property*:

$$\mathbb{P}(\tau > t + s | \tau > s) = \mathbb{P}(\tau > t).$$

b. Let $\tau_i, i = 1, 2, ...$ be a sequence of i.i.d. Exponential($\lambda$). Let $S_k := \sum_{i=1}^{k} \tau_i$. The Poisson process $N(t)$ with rate $\lambda$ is defined as:

$$N(t) = \sum_{i=1}^{\infty} \mathbf{1}_{(t \geq S_i)}.$$

$\tau_i$ is called the *inter-arrival time.* It is the wait time from the $(i-1)^{th}$ jump to the $i^{th}$ jump. $S_i$ is called the *arrival time.* It is the time of the $i^{th}$ jump.

### 5.1.3 Important basic properties

a. Distribution: $N(t)$ is has distribution Poisson($\lambda t$), that is

$$\mathbb{P}(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

*Proof.* Let $S_n = \sum_{i=1}^{n} \tau_i$ be the arrival time, then

$$\begin{aligned} \mathbb{P}(N(t) = k) &= \mathbb{P}(S_{k+1} > t, S_k \leq t) \\ &= \mathbb{P}(S_{k+1} > t) - \mathbb{P}(S_{k+1} > t, S_k > t) = \mathbb{P}(S_{k+1} > t) - \mathbb{P}(S_k > t). \end{aligned}$$

From Shreve's Lemma 11.2.1, $S_n$ has Gamma($\lambda, n$) distribution. That is, it has the density:

$$g_n(s) = \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda s}, s \geq 0.$$

It is a straight forward matter of integration now to verify that

$$\mathbb{P}(S_{k+1} > t) - \mathbb{P}(S_k > t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

The integration can be tedious, however. Another way to verify it is as followed: Denote $f(t) := \mathbb{P}(S_{k+1} > t) - \mathbb{P}(S_k > t)$ and note that $f(t)$ satisfies the following ODE:

$$\begin{aligned} f'(t) &= g_k(t) - g_{k+1}(t) = \frac{(\lambda t)^{k-1}}{(k-1)!} \lambda e^{-\lambda t} - \frac{(\lambda t)^k}{k!} \lambda e^{-\lambda t} \\ f(0) &= 0. \end{aligned}$$

It is clear that $f(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$ is the *unique* solution to the above ODE. The verification is complete. ■

b. $N(t)$ has independent increment. That is if we denote $\mathcal{F}_t$ to be the filtration generated by $N(s), 0 \le s \le t$ then for all $t \le t_1 < t_2$, $N(t_2) - N(t_1)$ is independent of $\mathcal{F}_t$.

Heuristic reason: Let $0 \le s < t$. Clearly $N(t) - N(s)$ counts the number of jumps starting from time $s$. Given all the information up to time $s$, what is the distribution of the first jump time after $s$? That is, we want to compute $\mathbb{P}(S_{N(s)+1} \ge t|\mathcal{F}_s)$, where $S_n$ is the arrival time as defined in Shreve (11.2.4). Note that since $N(s)$ represents the number of jumps up to time $s$, $S_{N(s)+1}$ is exactly the time of *the first jump after time* $s$.

But this is the same as computing $\mathbb{P}(\tau_{N(s)+1} \ge t - S_{N(s)}|\tau_{N(s)+1} \ge s - S_{N(s)})$. Note that $S_{N(s)}$ here represents the time of the *last jump before time s,* and $\tau_{N(s)+1}$ is *the wait time between the last jump before time* $s$ *and the first jump after time* $s$. So $\mathbb{P}(\tau_{N(s)+1} \ge t - S_{N(s)}|\tau_{N(s)+1} \ge s - S_{N(s)})$ asks for the probability that we have to wait until after time $t$ for the first jump after time $s$, given that we know we have waited up until time $s$ since the last jump before $s$, which has the same content as $\mathbb{P}(S_{N(s)+1} \ge t|\mathcal{F}_s)$.

Note also that $\mathbb{P}(\tau_{N(s)+1} \ge t - S_{N(s)}|\tau_{N(s)+1} \ge s - S_{N(s)}) = \mathbb{P}(\tau_{N(s)+1} \ge t - s + s - S_{N(s)}|\tau_{N(s)+1} \ge s - S_{N(s)})$. Since $\mathcal{F}_s$ is given, $N(s)$ should be looked at as a constant here. But from the memoryless property of $\tau_{N(s)+1}$, we get

$$\mathbb{P}(\tau_{N(s)+1} \ge t - s + s - \tau_{N(s)}|\tau_{N(s)+1} \ge s - \tau_{N(s)}) = \mathbb{P}(\tau_{N(s)+1} \ge t - s).$$

That is, the first jump time after $s$ can be looked at as an exponential clock starts at time $s$, hence independent of the past information. Using the independence of interarrival times, it is clear now that the increments of $N(t)$ after time $s$ is independent of the information up to time $s$. ■

c. $N(t)$ has stationary increment. More specifically, $N(t) - N(s)$ has distribution Possion($\lambda(t-s)$).

Heuristic reason: It follows from the same arguments of part b. ■

### 5.1.4 Compound Poisson process

The Poisson process we introduced has the satisfactory property that it jumps at *random* times. However, each of the jump is by definition of length 1, which is rather restrictive. It is desirable in terms of being realistic to have random jumps in our model. To that end, we proceed as followed.

Let $N(t)$ be a Poisson process with rate $\lambda$ and let $Y_0 = 0, Y_i, i = 1, 2, ...$ be i.i.d.(and also independent of N(t)) with $E(Y_i) = \mu$. Define

$$Q(t) = \sum_{i=0}^{N(t)} Y_i,$$

then $Q(t)$ is called a compound Poisson process. Similar to a Poisson process, $Q(t)$ also has the basic properties of *independent and stationary increments.* We *do not know* the specific distribution of $Q(t) - Q(s)$ (it depends on the distribution of $Y_i$'s , of course), but we do know that $E(Q(t) - Q(s)) = \mu\lambda(t - s)$.

## 5.2 Continuous noise - Brownian motion

### 5.2.1 Brownian motion as a source of noise from the limit of discrete time binomial model

We have seen that the discrete time binomial model provides some rich features to our modeling of a financial asset. In particular, it allows us to discuss the pricing path dependent derivatives and American options, which would be trivial in the 1 period model. However, the discrete time model is still limiting in a sense: we're only allowed to take action; and the underlying asset can only change value at certain discrete points in time. Between these points, no change and no action can happen.

Of course the situation can be improved by adding more discrete time points. The finer the decision making process (making decisions more often), or the richer the movement of the underlying becomes, the more points we will have to add to the model. Eventually we can see that our model approaches some kind of continuous time process. The question is how can we do it in a tractable way? Let's call our process $S_t$ and suppose it's defined for every $t$ in $[0, T]$ (by approximation). Tractable here means, for one thing, that our approximation should converge. (Not every refining procedure will converge, as we may have learned with a non-Riemann integrable function - the refining procedure there for the integral of such a function does not converge). For another, we want to be able to compute the distribution of $S_t$ at any time $t$, as well as the joint distribution of $S_{t_1}, S_{t_2}, \cdots, S_{t_n}$ for any time points $t_1, t_2, \cdots, t_n$. Beyond this we would like to know as much about $S_t$ as possible. The key tool for us is the Central Limit Theorem. As we shall see, the limiting process $S_t$ will have the form of the equation (5.3) where the noise term is replaced by the increment of a Brownian motion.

### 5.2.2 The limit of discrete time binomial model

As mentioned in the previous section, we need to do our approximation in the right way. For example, observe that if $S_N = S_0 X_1 X_2 \cdots X_N$ then

$$\log(S_n) = \log(S_0) + \sum_{k=1}^{N} \log(X_k).$$

Thus

$$\begin{aligned} E(\log(S_N)) &= \log(S_0) + NE(\log(X_1)); \\ Var(\log(S_N)) &= NVar(\log(X_1)). \end{aligned}$$

As we add more time points (increasing $N$) we would want to keep the $E(\log(S_N))$ and $Var(\log(S_N))$ constant. After all, it is the distribution of the asset at the terminal time $T = N\delta T$. Also, the spirit of Central Limit Theorem is that if we have a sequence of partial sum of i.i.d random variables, *scaled in such a way that ther expectation and* *variance remains constant*:

$$\begin{aligned} E\left(\frac{\sum_{i=1}^n X_i - \mu}{\sigma\sqrt{n}}\right) &= 0; \\ Var\left(\frac{\sum_{i=1}^n X_i - \mu}{\sigma\sqrt{n}}\right) &= 1. \end{aligned}$$

then the sequence of partial sum will converge.

This motivates us to model $X_i$ as followed: for a fixed $\mu, \sigma$

$$X_i = e^{\mu\Delta T + \sigma\sqrt{\Delta T}\xi_i},$$

where $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$, and the $\xi_i$ are independent.

Note that this is just another way to express our previous discrete time model, with

$$\begin{aligned} u = e^{\mu\Delta T + \sigma\sqrt{\Delta T}}; \\ d = e^{\mu\Delta T - \sigma\sqrt{\Delta T}}. \end{aligned}$$

Given $u, d$, we can solve for $\mu, \sigma$ such that the above system is satisfied. Note also that the no arbitrage condition requires:

$$\mu\Delta T - \sigma\sqrt{\Delta T} < r\Delta T < \mu\Delta T + \sigma\sqrt{\Delta T}.$$

Also note that the probability given above is *the real world probability, not the risk neutral probability*. The choice of 1/2 may seem a bit arbitrary, but keep in mind that the real world probability is irrelevant for pricing. We only need to build a model with rich enough structure to capture the up and down movement of the asset. The choice of 1/2 helps us to prove the convergence of the model in an easier way. Also note that we can even make $\mu$ and $\sigma$ be random variables, if we feel adventurous. For now we'll just keep them constants.

Lastly we want to verify that the expectation and variance of $\log(S_N)$ is constant:

$$\begin{aligned} E(\log(S_N)) &= \log(S_0) + NE(\log(X_1)) = \log(S_0) + \mu N\Delta T = \log(S_0) + \mu T; \\ Var(\log(S_N)) &= NVar(\log(X_1)) = \sigma^2 N\Delta T = \sigma^2 T. \end{aligned}$$

### 5.2.3 Converging to the continuous model

We will now denote $S_N$ as $S_T$ (since $T = N\Delta T$). From the previous section, we have

$$\begin{aligned} \log(S_T) &= \log(S_0) + \mu T + \sum_{i=1}^N \sigma\sqrt{\Delta T}\xi_i \\ &= \log(S_0) + \mu T + \sigma\sqrt{T}\frac{\sum_{i=1}^N \xi_i}{\sqrt{N}}. \end{aligned}$$

Let $N \to \infty$, by the Central Limit Theorem, we see that $S_T$ has distribution

$$\log(S_T) = \log(S_0) + \mu T + \sigma W_T,$$

where $W_T$ has distribution $N(0,T)$ (one can represent $W_T = \sqrt{T}N(0,1)$ as well).

By a similar argument, partitioning the interandom variableal $[t,T]$ into $N$ sub-interals, and let $N \to \infty$ we also have

$$\log(S_T) = \log(S_t) + \mu(T-t) + \sigma W_{T-t},$$

where $W_{T-t}$ has $N(0, T-t)$ distribution.

Moreover, since the increments $\xi_i$ are independent, it can be shown that $W_{T-t}$ is indendent of $W_t$ (where $W_t$ is the random variable we get for partitioning the interandom variableal $[0,t]$ and let $N \to \infty$). In fact, by the same argument, we can see that $W_{T-t}$ is indendent of $W_r, 0 \le r \le t$. This is the so called independent increment propety we'll discuss later.

Thus, in summary, one can say that our continuous model, derived from the limit of the discrete binomial model, satisfies the following properties in distribution:

a. For any $s < t$

$$S_t = S_s e^{\mu(t-s)+\sigma W_{t-s}},$$

$W_{t-s}$ has $N(0, t-s)$ distribution, $W_{t-s}$ is indendent of $W_r, 0 \le r \le s$.

b. In particular,

$$S_t = S_0 e^{\mu t + \sigma W_t}.$$

We say $S_t$ has log normal distribution (the $\log$ of $S_t$ has Normal distribution).

### 5.2.4 Brownian motion and Itô integral

*Definition*

A stochastic process $W_t$ is a Brownian motion (abbreviated BM) if

a. $W_0 = 0$

b. $W_t - W_s$ is independent of $W_r, 0 \le r \le s$.

c. $W_t - W_s$ has Normal(0, $t - s$) distribution.

Remark:

1. Property a can be changed to $W_0 = x$ which would be referred to as Brownian motion starting at $x$. Without specification, by default we refer to a Brownian motion starting at 0.

2. Property b and c together are referred to as independent and stationary increment. Omiting the normal distribution, we can write it as: for any $s_1 < t1 \le s_2 < t_2$ $W_{t_1} - W_{s_1}$ and $W_{t_2} - W_{s_2}$ are i.i.d. This can be generalized to $n$ pair of points.

**Example 5.2.1.** *Compute, for* $s < t$ $E(W_t)^2, E(W_sW_t), E(W_sW_t^2), E(W_t|W_s), E(W_t^2|W_S)$.

*Some terminologies*

1. We will say that an event $E$ happens *almost surely* (abbreviated a.s.) (under some reference probability $P$) if it happens with probability 1 under $P$: $P(E) = 1$.

2. A stochastic process can be viewed as a function of 2 variables: $t$ and $\omega$. Let $X_t$ be a stochastic process (the $\omega$ variable is usually suppressed when it's implicitly understood that $X_t$ also depends on $\omega$). By *a path of* $X_t$, we understand it as fixing an $\omega$ and viewing $X(\cdot, \omega)$ as a function of $t$.

*Some properties of Brownian motion*

a. Almost surely, Brownian motion path is continous . That is

$$P(W_t \text{ is continuous in } t \text{ on } [0, T]) = 1.$$

b. Almost surely, Brownian motion has nowhere differentiable path. This property is one of the reasons we need to use Itô Calculus to study integration with respect to Brownian motion.

c. Let $\Delta$ be a finite partition of $t, T]$, that is a collection of $t = t_0 < t_1 < t_2 < \cdots < t_n = T$. Then for all partition $\Delta$ of $[t, T]$ we have

$$E \sum_i (W_{t_{i+1}} - W_{t_i})^2 = T - t.$$

We say the quadratic variation of Brownian motion on the interandom variableal $[t, T]$ is $T - t$. This is also one important feature of Brownian motion that leads to Itô Calculus.

### 5.3 Exercises

1. Compute $\phi_X(t) := E\big(\exp(itX)\big)$, where $\exp(x) := e^x$, $i$ is the imaginary number: $i^2 = -1$ and $X$ has $N(\mu, \sigma^2)$ distribution. Recall that the density of $N(\mu, \sigma^2)$ is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

$\phi(t)$ is called the characteristic function of $X$.

2.
a) Suppose $\mu = 0$. Compute $\phi_X^{(4)}(t)$: the 4th derivative of $\phi_X$ with respect to $t$.
b) Use the following fact:

$$E[X^k] = (-i)^k \phi_X^{(k)}(0)$$

to compute $E[(W_t)^4]$ where $W$ is a Brownian motion.

3. Let $0 \le s \le t \le T$ and $W$ a Brownian motion. Compute the followings:
a) $E(W_s^2 W_t)$
b) $E(W_t^2 W_s)$
c) $E(\exp(\sigma W_t - \frac{1}{2}\sigma^2 t))$, where $\sigma$ is a constant.
d) $E(W_t^2 | W_s)$

# CHAPTER 6 An introduction to stochastic calculus

## 6.1 The dynamics of $S_t$

### 6.1.1 Evolution in discrete time

In the discrete time model, the evolution of $S_n$ is clear. At the next time point $n+1$, to get the value of $S_{n+1}$, we just have to multiply $S_n$ with $X_{n+1}$, where $X_{n+1} = e^{\mu \Delta T + \sigma \sqrt{\Delta T} \xi_{n+1}}$ as described above. In other words

$$S_{n+1} = S_n e^{\mu \Delta T + \sigma \sqrt{\Delta T} \xi_{n+1}}.$$

Equivalently, at the next time point $n+1$, to get the value of $\log(S_{n+1})$, we just have to add $\log(S_n)$ with $\mu \Delta T + \sigma \sqrt{\Delta T} \xi_{n+1}$. We say

$$\Delta \log(S_n) := \log(S_{n+1}) - \log(S_n) = \mu \Delta T + \sigma \sqrt{\Delta T} \xi_{n+1}.$$

This is a discrete recursion equation that specifies the dynamics of $\log(S_n)$. We would like to get an equation for the dynamics of $S_n$:

$$\Delta S_n := S_{n+1} - S_n =?$$

But in our discrete time model, this won't be anything nice. we'll see that the situation is different in the continuos time.

### 6.1.2 Evolution in continuous time

In the continuous time context, $\Delta T$ becomes $dt$ ($dt$ is not a real physical quantity, it's a differential and only makes sense within an integral sign. Nevertheless, it can be used to specify the dynamics of a process in time, as long as we are clear on what we mean by using $dt$).

Formally, we have

$$S_{t+dt} = S_t e^{\mu dt + \sigma \sqrt{dt} \xi_t},$$

where $\xi_t$ takes values $\pm 1$ with probability 1/2 each. Apply Taylor's expansion on the exponential term, we get

$$\begin{aligned} S_{t+dt} &= S_t(1 + \mu dt + \sigma \xi_t \sqrt{dt} + \frac{1}{2}\sigma^2 \xi_t^2 dt + \text{higher order terms}) \\ &= S_t(1 + [\mu + \frac{1}{2}\sigma^2]dt + \sigma \xi_t \sqrt{dt} + \text{higher order terms}) \\ &= S_t + S_t[\mu + \frac{1}{2}\sigma^2]dt + \sigma S_t \xi_t \sqrt{dt} + S_t \text{higher order terms}). \end{aligned}$$

since $\xi_t^2 = 1$. What we mean by the above is, provide we can make sense of the intergration, we have for $s < t$

$$S_t = S_s + \int_s^t [\mu + \frac{1}{2}\sigma^2] S_u du + \int_s^t \sigma S_u \xi_u \sqrt{du} + \int_s^t \text{higher order terms}.$$

By higher order terms, we mean terms of order higher than $dt^{3/2}$. As we will explain later, the intergral of the form

$$\int_s^t \text{higher order terms} = \int_s^t O(dt^{3/2}) = 0.$$

Also the integral needs to be explained, because of the term

$$\int_s^t \sigma S_u \xi_u \sqrt{du}.$$

As we'll also explain, $\int O(\sqrt{dt}) = \infty$. Thus the term $\int_s^t \sigma S_u \xi_u \sqrt{du}$ is undefined. However, $S_t$ is defined (we got it as a limit of convergence of the discrete model, and we have a distribution for $S_t$). Thus there must be a way to define $\int_s^t \sigma S_u \xi_u \sqrt{du}$. As we'll see, this will lead to the definition of Brownian motion and Itô Calculus.

The bottomline is, provided we can make sense of these technical details, we have arrived at the dynamics of $S_t$ as we wished for in the continuous time, using the Taylor's expansion:

$$dS_t := S_{t+dt} - S_t = S_t[\mu + \frac{1}{2}\sigma^2]dt + \sigma S_t \xi_t \sqrt{dt},$$

where we thow away the higher order terms since it disappears in the integral, which is the rigorous sense we want to give to the above dynamical equation anyway.

## 6.2 Prelude to Brownian motion and Itô Calculus

### 6.2.1 Brownian motion

As we can observe from the previous sections, the term that makes $S_t$ a random variable is $W_t$ in

$$S_t = S_0 e^{\mu t + \sigma W_t},$$

or $W_{t-s}$ is

$$S_t = S_s e^{\mu(t-s)+\sigma W_{t-s}},$$

or $\xi_u$ in

$$S_t = S_s + \int_s^t [\mu + \frac{1}{2}\sigma^2] S_u du + \int_s^t \sigma S_u \xi_u \sqrt{du}.$$

Indeed, there is reason to believe these are all different forms of one single process, let's call it $W_t$ where corresponding to the above we have

$$\begin{aligned} W_t &= W_t \\ W_{t-s} &= W_t - W_s \\ \sqrt{du}\xi_u &= dW_u. \end{aligned}$$

Let's see what we have learned about this process $W_t$ so far:
a. $W_t - W_s$ is independent of $W_r, 0 \le r \le s$.
b. $W_t - W_s$ has $N(0, t-s)$ distribution.
Surprisingly, these two characteristics are enough to specify a unique stochastic (random) process called the Brownian motion. As typical in mathematics, now that we have the intuition, we'll take the reverse approach and *define* Brownian motion as a process satisfying properties a and b. We then *build* a model for the underlying $S_t$ out of this Brownian motion $W_t$. The question is: how do we build $S_t$?

### 6.2.2 Itô Calculus

The short answer to the question how to build $S_t$ is to specify its dynamics:

$$\begin{aligned} dS_t &= S_t \mu dt + \sigma S_t dW_t, \\ S_0 &= x. \end{aligned}$$

Then we will find, if possible, a process $S_t$ that has the above dynamics. Yet this involves several significant difficulties. First of all, the above is a *Stochastic Differential Equation* (SDE). Equation because $S_t$ appears on both sides of the equality (recall Ordinary Differental Equation (ODE) and what makes it a differntial equation). Stochastic because the equation relates random quantities on both sides. Differential because it involves $dt$ and $dWt$.

Even ignoring the issue that we are dealing with a SDE, there is even a more fundamental issue: what do we mean by

$$\int_0^t f(u) dW_u,$$

for a (possibly random) process $f(u)$. The reason is we have seen in some sense

$$\int_0^t f(u)dW_u = \int_s^t f(u)\xi_u\sqrt{du},$$

and integrating with respect to the term $\sqrt{du}$ has to be interpreted in a special way. Being able to give a sense to integrating with $dWu$ is one fundamental result in the Itô Calculs: the so-called Itô's integral.

Similar to classical Calculus (but in some sense developing things in the reverse order), after we have a notion of the integral, we want to have notion of differentiation. That is, what is the "derivative" with respect to $t$ (to time) of a term like

$$S_t = \int_0^t \alpha(u)d_u + \int_0^t \sigma(u)dW_u?$$

In some sense, the answer has been given above, the "derivative" is given in terms of the differential:

$$dS_t = \alpha(t)dt + \sigma(t)dW_t.$$

The reason we don't have a proper derivative with respect to $t$ is because $W_t$ is NOT differentiable in $t$, another characteristic property of Brownian motion.

Then what about the "derivative" of a function os $S_t$, say $S_t^2$? This is to ask for the chain rule for the stochastic calculus. The chain rule in this case is referred to as Itô's formula.

Lastly, how can we solve a SDE? After having the chain rule, we can develop some basic techniques to find the explicit solution for some basic form of SDEs, in which the equation

$$\begin{aligned} dS_t &= S_t\mu dt + \sigma S_t dW_t, \\ S_0 &= x \end{aligned}$$

is included. $S_t$ in this case will be referred to as Geometric Brownian motion.

## 6.3 The Lebesgue-Stiltjes integral

### 6.3.1 The Riemann integral

Suppose we have a partition of the interandom variableal $[0, T]$: $0 = t_0 < t_1 < \cdots < t_n = T$ and a function $f(t)$ on $[0, T]$ that has nice properties. Define $\|\Delta\| = \max_i(t_{i+1} - t_i)$ as the mesh of this particular partition. Then

$$\sum_{i=1}^n f(t_i)(t_{i+1} - t_i) \to \int_0^t f(s)ds,$$

as $\|\Delta\| \to 0$.

Additionally, if we try the following:

$$\sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i)^{1+\varepsilon}$$

or

$$\sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i)^{1-\varepsilon}$$

for some $\varepsilon > 0$ we'll see that

$$\sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i)^{1+\varepsilon} \to 0$$

and

$$\sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i)^{1-\varepsilon} \to \infty$$

as $\|\Delta\| \to 0$. The intuitive reason is because

$$\sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i)^{1+\varepsilon} = \sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i)(t_{i+1} - t_i)^{\varepsilon}$$

and since

$$\sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i) \to \int_0^t f(s)ds,$$
$$(t_{i+1} - t_i)^{\varepsilon} \to 0,$$

the product converges to 0. Similarly,

$$\sum_{i=1}^{n} f(t_i)(t_{i+1} - t_i)^{1-\varepsilon} = \sum_{i=1}^{n} \frac{f(t_i)(t_{i+1} - t_i)}{(t_{i+1} - t_i)^{\varepsilon}}$$

which goes to infinity.

Formally we write

$$\begin{aligned} \int_0^t f(t)(dt)^{1+\varepsilon} &= 0 \\ \int_0^t f(t)(dt)^{1-\varepsilon} &= \infty, \end{aligned}$$

where

$$\begin{aligned}\int_0^t f(t)(dt)^{1+\varepsilon} &= \lim_{\|\Delta\|\to 0}\sum_{i=1}^{n} f(t_i)(t_{i+1}-t_i)^{1+\varepsilon}\\ \int_0^t f(t)(dt)^{1-\varepsilon} &= \lim_{\|\Delta\|\to 0}\sum_{i=1}^{n} f(t_i)(t_{i+1}-t_i)^{1-\varepsilon}\end{aligned}$$

In particular,

$$\begin{aligned}\int_0^t f(t)(dt)^2 &= 0\\ \int_0^t f(t)\sqrt{dt} &= \infty,\end{aligned}$$

### 6.3.2 Some examples

**Example 6.3.1.**

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{i}{n^2} = \int_0^1 tdt = \frac{t^2}{2}\Big|_0^1 = \frac{1}{2},$$

*where we see the sum on the left hand side as the function* $f(t) = t$ *evaluated at the grid point* $\frac{i}{n}$ *and* $t_{i+1} - t_i = \frac{i+1}{n} - \frac{i}{n} = \frac{1}{n}$.

Note that the sum above can be checked straightforwardly by using the identity

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

The point is this simple example allows us to verify that

$$\begin{aligned}\int_0^1 t\sqrt{dt} &= \infty\\ \int_0^1 t(dt)^2 &= 0\end{aligned}$$

since

$$\begin{aligned}\lim_{n\to\infty}\sum_{i=1}^{n}\frac{i}{n}\frac{1}{n^2} &= \lim_{n\to\infty}\frac{n(n+1)}{2n^3} = 0\\ \lim_{n\to\infty}\sum_{i=1}^{n}\frac{i}{n}\frac{1}{\sqrt{n}} &= \lim_{n\to\infty}\frac{n(n+1)}{2n\sqrt{n}} = \infty.\end{aligned}$$

### 6.3.3 The Lebesgue-Stiltjes integral

we may ask why do we even care about expression like $\int_0^t f(s)\sqrt{ds}$ or $\int_0^t f(s)(ds)^2$, since we never see them in calculus. This is true, because all we've dealt with there were Riemann integral. However, we can generalize the notion of integration in the following way: let $g(x)$ be a function defined on $[0, T]$ with nice property. Then we can define

$$\int_0^T f(s)dg(s) := \lim_{\|\Delta\|\to 0} \sum_{i=1}^{n} f(t_i)(g(t_{i+1}) - g(t_i)),$$

if the limit on the right hand side exists. This is called the Lebesgue-Stiltjes integral of $f$ against $g$.

And since now we deal with general function $g$, we can see that there are such functions that when $t_{i+1} - t_i$ is small

$$g(t_{i+1}) - g(t_i) \approx \sqrt{t_{i+1} - t_i}.$$

(Such function $g$ is not one of our classical calculus examples. At least if we look among the differentiable functions we won't find one. The reason is if $g$ is differentiable, then first order approximation tells us that

$$g(t_{i+1}) - g(t_i) \approx g'(t_i)(t_{i+1} - t_i),$$

so it's of order $O(dt)$, not $O(\sqrt{dt})$.)

Our discussion above shows us that for such function $g$ the Lebesgue-Stiltjes integral for $f$ against $g$ does not exist. As we will see, the Brownian motion paths give us such an example of a function $g$, which makes it necessary to define the Itô's integral.

### 6.3.4 An example

Actually if we have done u-substitution in Calculus, we have performed the Lebesgue-Stiltjes integral (possibly without realizing it). Consider the following example:

$$\int_0^1 2xe^{x^2}dx = \int_0^1 e^u du,$$

where we made the substitution

$$\begin{aligned} u &= x^2 \\ du &= 2xdx. \end{aligned}$$

More explicitly we consider $u$ as a function of $x$:

$$\begin{aligned} u(x) &= x^2 \\ du/dx &= 2x. \end{aligned}$$

But another way to write this is we're evaluating the integral

$$\int_0^1 e^{x^2} du(x) = \int_0^1 e^{x^2} u'(x) dx = \int_0^1 e^{x^2} 2x dx.$$

That is we integrate $e^{x^2}$ with respect to $u(x) = x^2$ over the interandom variableal $[0, 1]$.

### 6.3.5 The Lebesgue-Stiltjes integral as an example of portfolio value

we may ask when we need to use the Lebesgue-Stiltjes integral. Consider the following example.

Let $0 = t_0 < t_1 < ... < t_n = T$ be a partition of $[0, T]$

Consider an investor who invests in an underlying asset $S$ and the saving account such that the portfolio is self-financing. Let $\pi_k = \pi_{t_k}$ be the value of the portfolio and $\Delta_k$ be the number of shares of $S$ he holds at time $k$. Then

$$\pi_{k+1} = \Delta_k S_{k+1} + e^{r\Delta T}(\pi_k - \Delta_k S_k).$$

We will replace $e^{r\Delta T}$ with $1 + r\Delta T$, i.e. continuous compounding with discrete compouding. They should be very close, if $\Delta T$ is small. Then the self-financing equation reads

$$\pi_{k+1} = \Delta_k S_{k+1} + (1 + r\Delta T)(\pi_k - \Delta_k S_k),$$

or

$$\begin{aligned}
\pi_{k+1} &= \pi_k + \Delta_k(S_{k+1} - S_k) + r\Delta_T(\pi_k - \Delta_k S_k) \\
&= \pi_k + \Delta_k(S_{k+1} - S_k) + y_k r(t_{k+1} - t_k) \\
&= \sum_{i=1}^{k} \Delta_i(S_{i+1} - S_i) + y_i r(t_{i+1} - t_i).
\end{aligned}$$

where $y_k$ is the amount of cash we holds at time $k$. Thus we see that if we consider $\Delta(t), y(t)$ as a function of $t$, self-financing requiring that $\Delta(t) + y(t) = \pi(t)$, letting $\|\Delta\| \to 0$ we get

$$\pi_t = \int_0^t \Delta_u dS_u + \int_0^t y_u r du.$$

Thus the amount of money we get from investing in the stock in the continuous time is a Lebesge-Stiltjes integral.

Note: if we replace $y_u = \pi_u - rS_u$, then the above equation reads

$$\pi_t = \int_0^t \Delta_u(dS_u - rS_u du) + \int_0^t \pi_u r du,$$

which has the following interpretation:

When we invest in a risky asset ($S$) in such a way that our portfolio is self-financing, our gain can be decomposed in two components: the deterministic component, which is just the saving account: $\int_0^t \pi_u r du$. The other component is how the underlying asset performs versus the saving account: if it performs better: $dS_u > rS_u du$ then our portfolio will perform better than the traditional saving. If it performs worse: $dS_u < rS_u du$ then our portfolio will perform worse than the traditional saving.

### 6.3.6 Integrating with respect to Brownian motion

As we have seen in the previous chapter, the underlying asset in continuous time has the structure

$$S_t = \int_0^t \alpha(u) d_u + \int_0^t \sigma(u) dW_u.$$

Thus we need to make sense out of the term $\int_0^t \sigma(u) dW_u$. Naively, one can define it like this: let $\Delta$ be a partition of $[0, t]$. By $\|\Delta\|$ we mean the mesh of this partition: $\max_i |t_{i+1} - t_i|$. Then following the Riemann integration technique, we can define

$$\int_0^t \sigma(u) dW_u := \lim_{\|\Delta\| \to 0} \sum_i \sigma(t_i)(W_{t_{i+1}} - W_{t_i}).$$

There are several issues that need to be considered.

1. In what sense do we understand the convergence in the right hand side (if it's convergent at all)? The reason is both the left hand side and the right hand side are still functions of $\omega$. So the most straightforward (!) way is to require the right hand side to converge for all $\omega$ (or a.s.). This won't happen, because of the irregularity of BM. It turns out that we need to settle for some type of convergence on average (that is, when we average - take expectation of the square - in $\omega$, we see that the right hand side converges). More specifically, we say $X_n$ converges to $Y$ in mean square if as $n \to \infty$

$$E[(X_n - Y)^2] \to 0.$$

In this way, there exists a random variable, we call it $\int_0^T \sigma(u) dW_u$ so that $\sum_i \sigma(t_i)(W_{t_{i+1}} - W_{t_i})$ converges to $\int_0^T \sigma(u) dW_u$ as $\|\Delta\| \to 0$. This way of interpreting the convergence of the right hand side is the first feature of Itô's integral.

2. Notice that the integrand $\sigma$ is sampled at the left hand point of the partition: $t_i$. Recall that in Riemann integration, where we sample the integrand on the interandom variableal does not affect the value of the integral (we can use right hand, left hand or mid point). In integrating with respect to BM, it turns out that which sample point we use matters.

3. The choice of the left hand point is referred to as the Itô integral, which we use in math finance because of its connection to non-anticipating portfolio. The choice of the trapezoidal rule (using $\frac{1}{2}(\sigma(t_{i+1}) + \sigma(t_i))$) is referred to as the Stratonovich integral,

and more popular among the physicists and engineers because of its closer connection to classical calculus and a “more natural" interpretation of convergence. ( “More natural here means if we approximate the Brownian path with path of nicer poperty - say piecewise smooth - then the integrals against these paths converge to the Stratonovich integral.)

**Example 6.3.2.** *Let* $\Delta$ *be a partition of* $[0,t]$. *Then*

$$\sum_i 2W_{t_i}(W_{t_{i+1}} - W_{t_i})$$

*converges to* $W_t^2 - t$ *in mean square as* $\|\Delta\| \to 0$ *while*

$$2\sum_i \frac{1}{2}(W_{t_{i+1}} + W_{t_i})(W_{t_{i+1}} - W_{t_i})$$

*converges to* $W_t^2$ *in mean square as* $\|\Delta\| \to 0$.

Proof:
We show the first limit. The second is left as a homework exercise.
The key to proving the first limit is the identity

$$2W_{t_i}(W_{t_{i+1}} - W_{t_i}) = W_{t_{i+1}}^2 - W_{t_i}^2 - (W_{t_{i+1}} - W_{t_i})^2.$$

Thus

$$\begin{aligned}\sum_i 2W_{t_i}(W_{t_{i+1}} - W_{t_i}) &= \sum_i W_{t_{i+1}}^2 - W_{t_i}^2 - \sum_i (W_{t_{i+1}} - W_{t_i})^2 \\ &= W_t^2 - \sum_i (W_{t_{i+1}} - W_{t_i})^2.\end{aligned}$$

We will not show the details, but it can be showed without much difficulty that

$$\sum_i (W_{t_{i+1}} - W_{t_i})^2$$

converges to $t$ in mean square (The proof is left as a homework exercise). Just keep in mind that it is not the same as showing

$$E(\sum_i (W_{t_{i+1}} - W_{t_i})^2) \to t,$$

which is a simpler calculation and indeed true for every partition $\Delta$. What we need to show is

$$E\Big\{\Big[\sum_i (W_{t_{i+1}} - W_{t_i})^2 - t\Big]^2\Big\} \to 0,$$

as $\|\Delta\| \to 0$.

### 6.3.7 Some properties of the Itô integrals

1. With probability 1, $X_t = \int_0^t \alpha(u) dW_u$ is nowhere differentiable in $t$.
2.

$$E\Big\{ \int_0^t \alpha(u) dW_u \Big\} = 0.$$

$$E\Big\{ \big[ \int_0^t \alpha(u) dW_u \big]^2 \Big\} = \int_0^t E(\alpha(u)^2) du.$$

3. If $\alpha(u)$ is **deterministic**, then $\int_0^t \alpha_u du$ has N(0, $\int_0^t \alpha(u)^2 du$) distribution.

**Example 6.3.3.** *$\int_0^t u dW_u$ has N(0, $\frac{t^3}{3}$) distribution.*

**Example 6.3.4.** *$W_t^2 - t = \int_0^t 2W_u dW_u$ does NOT have a Normal distribution, because the integrand, $2W_u$, is NOT deterministic.*

*We can still compute*

$$E(W_t^2 - t) = t - t = 0,$$

*which is consistent with*

$$E(\int_0^t 2W_u dW_u) = 0.$$

*Computing $E((W_t^2 - t)^2) = E(W_t^4 - 2tW_t^2 + t^2)$ involves computing the fourth moment of a Normal distribution (we'll learn how to do that later with Itô's formula). There's an easier way to do it by computing the right hand side:*

$$E\Big\{ \big[ \int_0^t 2W_u dW_u \big]^2 \Big\} = \int_0^t 4E(W_u)^2 du = \int_0^t 4u du = 2t^2.$$

*Thus we can deduce that*

$$E(W_t^4) = 2t^2 - t^2 + 2tE(W_t^2) = 3t^2.$$

### 6.3.8 Itô integral versus Riemann integral

The convergence

$$\sum_i 2W_{t_i}(W_{t_{i+1}} - W_{t_i})$$

to $W_t^2 - t$ in mean square means that

$$\int_0^t 2W_u dW_u = W_t^2 - t.$$

Now suppose we have a process $x(t)$ that is differentiable in $t$ with $x(0) = 0$, so that we can write

$$dx(t) = \dot{x}(t)dt,$$

where $\dot{x}(t)$ is the derivative of $x(t)$ wrt to $t$. (The above is just another way to write $\frac{dx(t)}{dt} = \dot{x}(t)$. Then we have, by the chain rule,

$$\int_0^t 2x(u)dx(u) = \int_0^t 2x(u)\dot{x}(u)du = x^2(t) - x^2(0) = x^2(t).$$

Thus Itô integral has a “correction" term versus its Riemannian counter part, in this case the $-t$ in the Itô integral. This can be thought of also as a consequence of the non-differentiability of the BM, so Itô integral cannot be done in the Riemannian way to begin with (the $\dot{W}(t) = \frac{dW(t)}{dt}$ does not exist).

### 6.3.9 Itô process and differential form

Let $\alpha(t, \omega)$ and $\sigma(t, \omega)$ be nice enough process. The discussion above have given the meaning to the process $X_t$ defined as

$$X_t = x + \int_0^t \alpha(u)du + \int_0^t \sigma(u)dWu,$$

where the $dt$ term is understood in the Riemannian sense and the $dWt$ term is understood in the Itô sense. Any process $X_t$ with the above structure is referred to as an Itô process. Note that we also have $X_0 = x$ in the above equation.

It is the convention in stochastic calculus that we write the above equation for $X_t$ in differential form:

$$\begin{aligned} dX_t &= \alpha(t)dt + \sigma(t)dWt, \\ X_0 &= 0. \end{aligned}$$

Just keep in mind that the differential form has no more rigorous meaning than saying:

$$X_t = x + \int_0^t \alpha(u)du + \int_0^t \sigma(u)dWu.$$

So if this is our first time seeing the differential form, it is good to build the habit of automatically converting it to the integral form to give it meaning.

### 6.3.10 Explicit formula for Itô's integral

we may wonder if we have explicit formula for Itô integral calculation, like the example $\int_0^t W_u dW_u$. The answer is yes, for certain types of integrand, essentially like the classical calculus. Of course we do not want to go through the definition to learn what the answer is. Recall that we "compute" the integral in the classical case by guessing the anti-derivative, then prove that it is the true anti-derivative by differentiating this candidate.

The idea for the Itô integral is the same, except that we have the correction term, so it's not easy to guess precisely what the "anti-derivative" is - but we can get close. For example in $\int_0^t 2W_u dWu$ it is natural to guess it is $W_t^2$. The next step is to "differentiate" the candidate and find out what the correction term should be. However, we do not know what differentiation in the Itô's context means yet. It cannot be done like the classical derivative, because one can also show that the process $Z_t = \int_0^t \sigma(u) dWu$ is nowhere differentiable in $t$. The Itô's formula gives the us the meaning in "differentiating" in this case, as well as rules to "differentiate" correctly with the right correction term.

## 6.4 Itô formula

### 6.4.1 Another perspective of differentiation

Let $f(t)$ be a deterministic function of $t$. What do we mean by the derivative of $f$ wrt $t$: $f'(t)$ ? We can either use the classical definition, via the limit of the difference quotion, *or* we can define it as a function such that for all $s < t$

$$f(t) - f(s) = \int_s^t f'(u) du.$$

The only issue is such definition of $f'(t)$ needs to be unique (there might possibly be two different functions $f'(t)$ that would satisfy the above equation). As we may have expected from classical calculus, it turns out there is only one $f'(t)$ that satisfies the above equation.

In this way, one can develop classical calculus with the notion of the integral first, and define the derivative this way (which is the approach of the Lebesgue integral and the derivative associated with it).

As far as we're concerned, we've followed exactly this program with the Itô's integral. We developed what it means to have an Itô's integral, and now we ask what it means to differentiate in $t$. If we take the above definition as guidance, then we've already known the answer: the "derivative" of a stochastic process is $dX_t$. This statement contains nothing deep, it simply says:

$$X_t = \int_0^t dX_u,$$

which is true by partitioning the interandom variableal $[0, t]$ and check it the Riemannian way.

However, if $X_t$ is an Itô process then we have something more to say: if

$$X_t = x + \int_0^t \alpha(u) du + \int_0^t \sigma(u) dWu,$$

then

$$dX_t \quad = \quad \alpha(t)dt + \sigma(t)dWt.$$

Note that if we divide both sides of the differential form with $dt$ (a VERY formal operation, not to be taken in any rigorous sense) we have

$$\frac{dXt}{dt} = \alpha(t) + \sigma(t)\frac{dWt}{dt},$$

which is exactly the differentiation with respect to $t$ that we have in mind. The problem, of course is the term $\frac{dWt}{dt}$ does not exist, so we have to use the differential form, and interpret the two types of integrals in two different sense.

So it turns out that differentiation in Itô sense is not that exciting (there is not much we can say). Yet there are still questions worth asking: if

$$dX_t \quad = \quad \alpha(t)dt + \sigma(t)dWt,$$

what can we say about $dX_t^2$, or better, $df(X_t)$ for a smooth function of $f(X_t)$? Asking this question is equivalent to formulating the chain rule in the classical case. Since if we have a differentiable function $x(t)$ in $t$, we have

$$df(x(t)) = f'(x(t))\dot{x}(t)dt.$$

In the case that $X(t)$ is an Itô process, $X(t)$ is *not* differentiable in $t$. Thus the above formula does not work. The chain rule for a function of an Itô process, i.e. to figure out $df(X_t)$ when $X_t$ is an Itô proess, is referred to as the Itô's formula.

### 6.4.2 The Itô's formula

**Theorem 6.4.1.** *Let $X_t$ be a Itô process:*

$$dX_t \quad = \quad \alpha(t)dt + \sigma(t)dWt.$$

*Let $f(t,x)$ be once continously differentiable in $t$, twice continuously differentiable in $x$ (abbreviated as $C^{1,2}$). Then*

$$\begin{aligned} df(t,X_t) &= f_t(t,X_t)dt + f_x(t,X_t)dX_t + \frac{1}{2}f_{xx}(t,X_t)\sigma^2(t)dt \\ &= [f_t(t,X_t) + f_x(t,X_t)\alpha(t) + \frac{1}{2}\sigma^2(t)f_{xx}(t,X_t)]dt + f_x(t,X_t)\sigma(t)dWt. \end{aligned}$$

Remark: If we understood that the formula should be evaluated at $(t,X_t)$ we can neglect the arguments and write the Itô's formula in a somewhat cleaner form:

$$\begin{aligned} df(t,X_t) &= f_t dt + f_x dX_t + \frac{1}{2}f_{xx}\sigma_t^2 dt \\ &= [f_t + f_x\alpha_t + \frac{1}{2}\sigma_t f_{xx}]dt + f_x\sigma_t dWt. \end{aligned}$$

**Example 6.4.2.** *Apply Itô's formula to $W_t^2$:*

$$dW_t^2 = 2W_t dW_t + dt$$

*That is*

$$W_t^2 - W_0^2 = \int_0^t 2W_u dW_u + t.$$

*Or*

$$\int_0^t 2W_u dW_u = W_t^2 - t,$$

*recovering the formula we developed above.*

**Example 6.4.3.** *Apply Itô's formula to $tW_t^2$:*

$$d(tW_t^2) = W_t^2 dt + 2tW_t dW_t + tdt.$$

*That is*

$$tW_t^2 = \int_0^t W_u^2 du + \int_0^t 2uW_u dW_u + \int_0^t udu.$$

*Or*

$$tW_t^2 - \frac{t^2}{2} = \int_0^t W_u^2 du + \int_0^t 2uW_u dW_u.$$

This is an example of the integration by parts formula in the Itô's context:

$$\begin{aligned}\int_0^t W_u^2 du &= tW_t^2 - \int_0^t ud(W_u^2) \\ &= tW_t^2 - \int_0^t 2udW_u - \int_0^t udu.\end{aligned}$$

### 6.4.3 Intuition of the Itô's formula

We won't give a rigorous proof of the Itô's formula. We'll only list here the steps to help we understand why such formula is intuitively correct. We'll do this by considering different forms of $f(t, X_t)$. Note that intuitively

$$df(t, X_t) \approx \Delta f(t, X_t) = f(t + h, X_{t+h}) - f(t, X_t),$$

for very small $h$. The key to Itô's formula is the second order Taylor expansion. Indeed Itô's formula can be looked at as Taylor's expansion with the $dW_t$ term.

$f(t, X_t) = f(W_t)$

Apply Taylor's expansion

$$f(W_{t+h}) - f(W_t) = f'(W_t)(W_{t+h} - W_t) + \frac{1}{2}f''(W_{t+\varepsilon})(W_{t+h} - W_t)^2,$$

for some $0 \leq \varepsilon \leq h$.

Therefore, if $\Delta$ is a partition of $[0, t]$ such that $\|\Delta\| \leq h$, we have

$$\begin{aligned} f(W_t) - f(W_0) &= \sum_i f(W_{t_{i+1}}) - f(W_{t_i}) \\ &= \sum_i f'(W_{t_i})(W_{t_i+h} - W_{t_i}) + \frac{1}{2}f''(W_{t_i+\varepsilon})(W_{t_i+h} - W_{t_i})^2 \end{aligned}$$

Let $\|\Delta\| \to 0$, we'll see that

$$\sum_i f'(W_{t_i})(W_{t_i+h} - W_{t_i}) \to \int_0^t f(W_u)dW_u,$$

in mean square.

We have discussed how $\sum_i (W_{t_i+h} - W_{t_i})^2 \to t$ in mean square as $\|\Delta\| \to 0$. In a similar manner, we can believe that

$$\sum_i f''(W_{t_i+\varepsilon})(W_{t_i+h} - W_{t_i})^2 \to \int_0^t f''(W_t)dt,$$

in mean square. That is, we have showed

$$f(W_t) - f(W_0) = \int_0^t f'(W_u)dW_u + \frac{1}{2}\int_0^t f''(W_u)du.$$

$f(t, X_t) = f(t, W_t)$

This case is very similar to the above, except the Taylor's expansion becomes

$$\begin{aligned} f(t+h, W_{t+h}) - f(t, W_t) &= f_t(t, W_t)(t + h - h) + f_x(t, W_t)(W_{t+h} - W_t) \\ &+ \frac{1}{2}f_{xx}(t, W_t)(W_{t+h} - W_t)^2 + f_{tx}(t, W_t)h(W_{t+h} - W_t) \\ &+ \text{higher order term} \end{aligned}$$

where by higher order term we mean terms of order $h^k(W_{t+h} - W_t)^j$ such that $j + k \geq 3$.

Apply the partition of $[0, t]$ as before, and note that

$$\begin{aligned}
\sum_i f_t(t_i, W_{t_i})h &\to \int_0^t f(u, W_u)du, \\
\sum_i f_x(t_i, W_{t_i})(W_{t_i+h} - W_{t_i}) &\to \int_0^t f_x(W_u)dW_u, \\
\sum_i f_{xx}(t_i, W_{t_i})(W_{t_i+h} - W_{t_i})^2 &\to \int_0^t f_{xx}(W_t)dt, \\
\sum_i \text{higher order term} &\to 0
\end{aligned}$$

in mean square as before. The only point to note is that the term

$$\sum_i f_{tx}(t_i, W_{t_i})(W_{t_i+h} - W_{t_i})h \to 0$$

as $h \to 0$.

The heuristic reason is this: since $\sum_i (W_{t_i+h} - W_{t_i})^2 \to t$ in mean square, we can believe that $W_{t_i+h} - W_{t_i}$ " $\to$ " $\sqrt{h}$ for small $h$. (This is not to be taken in any rigorous way). Thus, if $f_{tx}$ is bounded by a constant $C$:

$$\Big| \sum_i f_{tx}(t_i, W_{t_i})(W_{t_i+h} - W_{t_i})h \Big| \le C \sum_i h^{3/2} \le C\frac{T}{h}h^{3/2} \to 0,$$

as $h \to 0$.

The above gives the following informal rule about product of differentials when we deal with Taylor expansion involving $dWt$.

*Informal rule for product of differentials*

We have

$$\begin{aligned}
(dt)^2 &= 0; \\
dtdW_t &= 0; \\
(dW_t)^2 &= dt.
\end{aligned}$$

More rigorously, the above means, for a partition $\Delta$ of $[0, T]$ and bounded $f$

$$\begin{aligned}
\sum_i f(t_i)(t_{i+1} - t_i)^2 &\to 0; \\
\sum_i f(t_i)(t_{i+1} - t_i)(W_{t_{i+1}} - W_{t_i}) &\to 0; \\
\sum_i f(t_i)(W_{t_{i+1}} - W_{t_i})^2 &\to \int_0^t f(s)ds
\end{aligned}$$

in mean square as $\|\Delta\| \to 0$.

With these informal rule, we can give an intuition about the Itô's formula for the general case that $X_t$ is an Itô process.

*$f(t, X_t), X_t$ an Itô process*

By Taylor's expansion

$$\begin{aligned} f(t+h, X_{t+h}) - f(t, X_t) &= f_t(t, X_t)(t+h-t) + f_x(t, W_t)(X_{t+h} - X_t) \\ &+ \frac{1}{2} f_{xx}(t, X_t)(X_{t+h} - X_t)^2 + f_{tx}(t, X_t)h(X_{t+h} - X_t) \\ &+ \text{ higher order term} \end{aligned}$$

Note that

$$X_{t+h} - X_t \approx \alpha(t)h + \sigma(t)(W_{t+h} - W_t).$$

Rewriting the Taylor's expansion, replacing the difference term with differential term we have

$$\begin{aligned} df(t, X_t) &= f_t dt + f_x dX_t + \frac{1}{2} f_{xx}(\alpha_t^2 (dt)^2 \\ &+ \sigma_t^2 (dWt)^2 + 2\alpha_t \sigma_t dt dW_t) + f_{tx} dt(\alpha_t dt + \sigma_t dW_t). \end{aligned}$$

Applying the informal rule of product of differentials, we get

$$df(t, X_t) = f_t dt + f_x dX_t + \frac{1}{2} f_{xx} \sigma_t^2 dt,$$

which is exactly the Itô's formula.

### 6.4.4 An application of Itô's formula

Itô's formula is fundamental in stochastic analysis as it provides the equivalence of the chain rule in this context. It also provides a nice tool for some computations, for example, the moments of a Normal random variable as followed.

**Example 6.4.4.** *Let $X$ have Normal$(0, \sigma^2)$ distribution. Compute $E(X^4)$.*

Ans: If we were to apply the usual formula

$$E(X^4) = \int_{-\infty}^{\infty} x^4 f_X(x) dx,$$

where $f_X(x)$ is the density of the Normal $(0, \sigma^2)$ it would be a tedious calculation.

Instead, let's compute $E(W_t^4)$ where $W_t$ is a BM. If we can do this, then

$$E(X^4) = \sigma^4 E(Z^4) = \frac{\sigma^4}{t^2} E(W_t^4),$$

where $Z$ is a standard Normal.

Now apply Itô's formula to $W_t^4$ we have

$$W_t^4 = W_0^4 + \int_0^t 4W_s^3 dW_s + \int_0^t 6W_s^2 ds.$$

The key is the expectation of $\int_0^t 4W_s^3 dW_s$ is 0. Thus taking expectation on both sides, we get

$$E(W_t^4) = 6\int_0^t E(W_s)^2 ds = 6\int_0^t s^2 ds = 3t^2.$$

Therefore $E(X^4) = 3\sigma^4$.

Remark: It is clear that $E(X^{2k+1})$, $k$ an integer is always 0. So we're only interested in computing the even moments of $X$. And for that purpose, we can just focus on computing the even moment of $W_t$ using the conversion trick described above.

**Example 6.4.5.** *Compute* $E(W_t^{2k})$.

Ans: Apply Itô's formula to $W_t^{2k}$:

$$W_t^{2k} = W_0^{2k} + \int_0^t 2kW_s^{2k-1} dW_s + \int_0^t k(2k-1)W_s^{2k-2} ds.$$

Taking expectations on both sides give

$$E(W_t^{2k}) = \int_0^t k(2k-1)E(W_s^{2k-2}) ds.$$

This provides a recursion formula to compute the even moments of $W_t$. For example, for $k = 3$

$$\begin{aligned} E(W_t^6) &= \int_0^t 3 \cdot 5 \cdot 3s^2 ds \\ &= 5 \cdot 3t^3. \end{aligned}$$

With this, we can compute for the case $k = 4$ for $E(W_t^8)$. In fact, it is not hard to guess that

$$E(W_t^8) = 7 \cdot 5 \cdot 3t^4,$$

and in general

$$E(W_t^{2k}) = (2k-1)!!t^k,$$

where

$$(2k-1)!! = (2k-1)(2k-3)\cdots 3 \cdot 1.$$

### 6.5 Geometric Brownian motion

#### 6.5.1 Definition

As motivated in Lecture notes 5a, a natural model for us to use for the underlying process $S_t$, as the limit of the multiperiod binomial model is

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

where $\mu, \sigma$ are constants, $W_t$ is a BM. We'll use the convention that $\sigma > 0$ (even though $S_t$ would have the same distribution if $\sigma < 0$).

**Definition 6.5.1.** *A process $S_t$ following the dynamics*

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

*is referred to as a Geometric BM.*

The word geometric comes from the intuition that each increment in $S_t$ is a product of $S_t$ with another increment:

$$dS_t = S_t(\mu dt + \sigma dW_t).$$

#### 6.5.2 Explicit formula

Now we derive an explicit formula for $S_t$ a Geometric BM. Formally dividing both sides of (6.1) by $S_t$ we have

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t.$$

Note that the right hand side is free of $S_t$. Therefore, if we can write the left hand side as $df(S_t)$ then we'll be almost done, since then

$$f(S_t) - f(S_0) = \int_0^t df(S_u)du = \int_0^t \mu du + \int_0^t \sigma dW_u,$$

and we can solve for $S_t$ by taking the inverse of $f$, if possible.

Recalling from classical calculus, the function with a derivative of the form $1/x$ is $\log(x)$. Thus we can try applying Ito's formula to $\log(S_t)$. We have

$$\begin{aligned} d\log(S_t) &= \frac{1}{S_t}dS_t - \frac{1}{2S_t^2}\sigma^2 S_t^2 dt \\ &= \mu dt + \sigma dW_t - \frac{1}{2}\sigma^2 dt. \end{aligned}$$

Therefore,

$$\log(S_t) - \log(S_0) = \int_0^t (\mu - \frac{1}{2}\sigma^2)du + \int_0^t \sigma dW_u$$
$$= (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t.$$

Therefore,

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}.$$

Excerise: Verify that

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}$$

satisfies

$$dS_t = \mu S_t dt + \sigma S_t dW_t.$$

## 6.6 Risk neutral measure

### 6.6.1 Definition

So far, the dynamics of $S_t$ we gave were under the physical measure $P$. That is properly we should write

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

under $P$ and

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}$$

under $P$.

How would we define the risk neutral measure in continuous time? Recalling the intutions that

a. The risk neutral measure must be equivalent to the physical measure $P$.

b. The risk neutral measure should satisfy

$$V_t = E^Q(e^{-r(T-t)} V_T | \mathcal{F}_t^S),$$

for any financial derivative that pays $V_T$ at time $t$.

c. The risk neutral price at time $t$ for a financial product that pays $S_T$ at time $T$ (a forward contract on $S$ with zero strike) is $S_t$.

Thus we see that the proper definition of a risk neutral measure $Q$ is a measure that is equivalent to $P$ such that for any $t \leq T$

$$E^Q(e^{-r(T-t)} S_T | S_t) = S_t.$$

### 6.6.2 Conditional expectation for Geometric BM

It is clear that to check whether or not a measure $Q$ is risk neutral, we need to perform a conditional expectation computation of the type

$$E^Q(f(S_T)|S_t),$$

for some function $f$.

It is reasonable to believe that under the risk neutral measure, $S_t$ is still a geometric Brownian motion, with possibly different $\mu, \sigma, W_t$. That is it has the following dynamics under $Q$:

$$dS_t = \mu^Q S_t dt + \sigma^Q S_t dW_t^Q,$$

where $\mu^Q, \sigma^Q$ are constants and $W_t^Q$ a Q-Brownian motion. For our computational purpose, we might as well just perform the expectation under $P$ to have an idea of what we're getting.

Since we're conditioning on $S_t$, we want to rewrite $S_T$ in terms of $S_t$, that is

$$E\Big\{f(S_T)\Big|S_t\Big\} = E\Big\{f\Big(S_t e^{(\mu-\frac{1}{2}\sigma^2)(T-t)+\sigma(W_T-W_t)}\Big)\Big|S_t\Big\}.$$

Note that we have a very familiar situation as what we dealt with in the discrete time: $W_T - W_t$ is independent of $S_t$. (To see this, note that $S_t$ is a function of $W_t$ by its explicit formula. The independence follows from the independent increment of BM). Therefore, we can apply the Independence Lemma, in the following form:

**Theorem 6.6.1.** *Independence Lemma Let $X$ be a random variable independent of $S_t$. Then*

$$E\Big\{f(X, S_t)|S_t\Big\} = E\Big\{f(X, x)\Big\}|_{x=S_t}.$$

Thus by the Independence Lemma, we see that

$$E\Big\{f(S_T)|S_t\Big\} = E\Big\{f\Big(xe^{(\mu-\frac{1}{2}\sigma^2)(T-t)+\sigma(W_T-W_t)}\Big)\Big\}|_{x=S_t}.$$

The expectation is taken over the random variable $W_T - W_t$, which has Normal(0,T-t) distribution. In particular, we have

$$E(S_T|S_t) = E\Big\{xe^{(\mu-\frac{1}{2}\sigma^2)(T-t)+\sigma(W_T-W_t)}\Big\}|_{x=S_t} = xe^{\mu(T-t)}|_{x=S_t} = S_t e^{\mu(T-t)},$$

where we have used the fact that if $X$ has Normal(0, $\sigma^2$) distribution

$$E(e^X) = e^{\frac{1}{2}\sigma^2}.$$

### 6.6.3 The dynamics of $S_t$ under the risk neutral measure

From the previous section, we see that if under $Q$ $S_t$ has the following dynamics :

$$dS_t = \mu^Q S_t dt + \sigma^Q S_t dW_t^Q,$$

then

$$E^Q(e^{-r(T-t)} S_T | S_t) = S_t e^{(\mu^Q - r)(T-t)}.$$

Thus if $\mu^Q = r$ under $Q$ then $S_t$ satisfies the risk neutral pricing condition under $Q$. What about $\sigma^Q$ ? What can we say about it? It turns out that,$\sigma^Q$ has to be equal to $\sigma$ if $Q$ is equivalent to $P$. The reason will be explained in section (7.2) for continuity of exposition.

It turns out that, there is a theorem called the Girsanov theorem, that says there exists such an equivalent measure $Q$. That is there exists a $Q$ equivalent to $P$ such that under $Q$

$$dS_t = r S_t dt + \sigma S_t dW_t^Q,$$

where $W^Q$ is a BM under $Q$. We will take this as given in our further discussion without further dicussing of the proof of Girsanov theorem. Moreover, we will slightly abuse notation and write $W_t$ instead of $W_t^Q$ even when we discuss the distribution of $S_t$ under $Q$.

## 6.7 Exercises

1. Use Ito formula to compute the following

a) $d\sin(W_t)$

b) $d\exp(W_t)$

2. Let $0 = t_0 < t_1 < t_2 < ... < t_n = T$. Show that

$$E\Big[(\sum_{i=0}^{n-1}\big(W_{t_{i+1}} - W_{t_i})^2 - T\big)^2\Big] = 2\sum_{i=0}^{n-1}(t_{i+1} - t_i)^2.$$

3. Suppose $S_t$ satisfies

$$dS_t = \sin(S_t)t^2 dt + \exp(\sqrt{St} - t)dW_t.$$

Compute

a) $d\log(S_t)$

b) $d\exp(S_t^2)$

c) $d\sqrt{S_t}$

Please note that this exercise is only to symbolically practice Ito's formula. There may not exist a process $S_t$ that satisfies the above dynamics, or even if it does exist, to make sense in the expression a,b and c.

4. $E(\exp(\int_0^t \sin(s)dW_s))$.

# CHAPTER 7 The Black-Scholes formula

### 7.0.1 The first fundamental theorem of asset pricing in continuous time

**Theorem 7.0.1.** *Let $S_t$ be a non-negative continuous stochastic process under $P$. Then the market consisting of $S_t$ and a saving account is arbitrage free if and only if there exists an equivalent risk neutral measure $Q$. Moreover, for any financial derivative based on $S$ that makes a payment $V_T$ at time $T$, the no-arbitrage price of $V$ at time $t$ is*

$$V_t = E^Q(e^{-r(T-t)}V_T|\mathcal{F}_t^S).$$

### 7.0.2 Pricing of European style derivative

Now suppose $V$ is a European style derivative: $V_T = \phi(S_T)$ (that is it is not path-dependent). Then

$$\begin{aligned} V_t &= E^Q(e^{-r(T-t)}\phi(S_T)|\mathcal{F}_t^S) \\ &= E^Q\Big\{e^{-r(T-t)}\phi\Big(S_t e^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(W_T-W_t)}\Big)\Big|\mathcal{F}_t^S\Big\}. \end{aligned}$$

Note that $S_t$ is constant given $\mathcal{F}_t^S$. $W_T - W_t$ is independent of $\mathcal{F}_t^S$. That is, we can use the Independence Lemma in the following form:

$$E^Q(f(S_t, W_T - W_t)|\mathcal{F}_t^S) = E^Q(f(x, W_T - W_t))|_{x=S_t}.$$

## 7.1 Pricing of the European-call option - The Black-Scholes formula

### 7.1.1 The Black-Scholes formula

**Theorem 7.1.1.** *Suppose that under $Q$,*

$$dS_t = rS_t dt + \sigma S_t dW_t.$$

*Then the no-arbitrage price at time $0$, $V_0$ of a European call with strike $K$ and expiration time $T$ satisfies*

$$V_0 = S_0 N(d_1) - Ke^{-rT}N(d_2),$$

*where $N(d_1) = P(Z \leq d_1)$, $Z$ standard Normal and*

$$\begin{aligned} d_1 &= \frac{(r + \frac{1}{2}\sigma^2)T - \log(\frac{K}{S_0})}{\sigma\sqrt{T}} \\ d_2 &= \frac{(r - \frac{1}{2}\sigma^2)T - \log(\frac{K}{S_0})}{\sigma\sqrt{T}} \end{aligned}$$

*The formula for $V_0$ is referred to as the Black-Scholes formula.*

### 7.1.2 Outline of the proof of Black-Scholes formula

At the heart of it, the Black-Scholes formula is just computing the expectation of $(X-K)^+$, where $X$ has log normal distribution. However, because of the parameters and the log normal density involved, the computation may seem intimidating for the first time. To help we keep track of what's going on, just keep in mind the big steps that we have to perform in this computation.

**1. Computing $E(X-K)^+$, for a continuous random variable $X$ with some density $\phi_X(x)$.**

The key for this step is to break the function $(x-K)+$ into two parts: $x-K$ for $x > K$ and $0$ for $x < K$. That is

$$\begin{aligned} \int_{-\infty}^{\infty} (x-K)^+ \phi_X(x) &= \int_K^{\infty} (x-K)\phi_X(x) \\ &= \int_K^{\infty} x\phi_X(x) - \int_K^{\infty} \phi_X(x). \end{aligned}$$

**2. Recognize that $\int_K^{\infty} \phi_X(x)$ is just $P(X > K)$.**

**3. Computing $P(X > K)$ for $X$ having log normal distribution.**

The key to this step is just to write $X$ as what it is in distribution: $X = e^Y$, where $Y$ have Normal distribution. Then

$$P(X > K) = P(e^Y > K) = P(Y > \log(K)),$$

and we can use the $Z$-transform to turn $P(Y > \log(K))$ into a standard Normal calculation.

**4. Recognizing that $\int_K^{\infty} x\phi_X(x)$ is $E(X\mathbf{1}_{X \geq K})$ where $X$ has log normal distribution (recall the indicator function definition as $\mathbf{1}_E = 1$ if $E$ happens and 0 otherwise).**

**5. Computing $E(X\mathbf{1}_{X \geq K})$ where $X$ has log normal distribution.**

The key to this step is also to write $X$ as what it is in distribution: $X = e^Y$, where $Y$ have Normal distribution. Then

$$E(e^Y \mathbf{1}_{e^Y \geq K}) = E(e^Y \mathbf{1}_{Y \geq \log(K)}) = \int_{\log(K)}^{\infty} e^y \phi_Y(y) dy,$$

where $\phi_Y(y)$ is a Normal density. What we accomplished here is transforming a log-normal expectation computation to a normal expectation computation (because it's complicated to figure out the density of a log-normal distribution).

**6. Computing expression of the type $E(e^Y)$, where $Y$ has Normal distribution.**
The first key to this step is to recognize that $e^y\phi_Y(y)$, where $\phi_Y(y)$ is a Normal density will be an exponential function, since it will have the form

$$e^y\phi_Y(y) = e^y e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

The second key is to **complete the square** for the exponent of the exponential. That is we want to write

$$e^y e^{-\frac{(y-\mu)^2}{2\sigma^2}} = e - \frac{(y-\tilde{\mu})^2}{2\tilde{\sigma}^2} + c,$$

for some constants $\tilde{\mu}, \tilde{\sigma}, c$. The point is

$$e^{-\frac{(y-\tilde{\mu})^2}{2\tilde{\sigma}^2}}$$

is (modulo a constant) the density of a Normal($\tilde{\mu}, \tilde{\sigma}$) so it will integrate to one. The constant $e^c$ will just factor out of the integration.

### 7.1.3 Proof of Black-Scholes formula

1. By the risk neutral pricing formula:

$$\begin{aligned} V_0 &= E^Q(e^{-rT}V_T) = E^Q(e^{-rT}(S_T-K)^+) = E^Q\Big[e^{-rT}\big(S_0e^{(r-\frac{1}{2}\sigma^2)T+\sigma W_T}-K\big)^+\Big] \\ &= E^Q\Big[e^{-rT}\big(S_0e^X-K)^+\Big], \end{aligned}$$

where $X$ has distribution $N((r-\frac{1}{2}\sigma^2)T, \sigma^2T)$. Note that $S_0e^X - K \geq 0$ iff $X \geq \log(\frac{K}{S_0})$. Therefore

$$\begin{aligned} V_0 &= \int_{\log(\frac{K}{S_0})}^{\infty} e^{-rT}(S_0e^x-K)\phi_X(x)dx \\ &= \int_{\log(\frac{K}{S_0})}^{\infty} e^{-rT}S_0e^x\phi_X(x)dx - \int_{\log(\frac{K}{S_0})}^{\infty} e^{-rT}K\phi_X(x)dx \\ &:= A - B, \end{aligned}$$

where $\phi_X(x) := \frac{1}{\sqrt{2\pi\sigma^2T}}\exp\big(-\frac{(x-\mu)^2}{2\sigma^2T}\big), \mu := (r-\frac{1}{2}\sigma^2)T$ is the density of $X$. We will compute $A$ and $B$ separately.

2.

$$\begin{aligned} B &= \int_{\log(\frac{K}{S_0})}^{\infty} e^{-rT} K \phi_X(x) dx = e^{-rT} K P(X > \log(\frac{K}{S_0})) \\ &= e^{-rT} K P(Z > \frac{\log(\frac{K}{S_0}) - (r - \frac{1}{2}\sigma^2)T}{\sigma T}) \\ &= e^{-rT} K N(d_2), \end{aligned}$$

where

$$d_2 := \frac{(r - \frac{1}{2}\sigma^2)T - \log(\frac{K}{S_0})}{\sigma\sqrt{T}}.$$

3.

$$\begin{aligned} A &= \int_{\log(\frac{K}{S_0})}^{\infty} e^{-rT} S_0 e^x \phi_X(x) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2 T}} \int_{\log(\frac{K}{S_0})}^{\infty} e^{-rT} S_0 \exp\big(x - \frac{(x-\mu)^2}{2\sigma^2 T}\big) dx. \end{aligned}$$

Clearly,

$$x - \frac{(x-\mu)^2}{2\sigma^2 T} = \frac{2\sigma^2 T x - (x-\mu)^2}{2\sigma^2 T}.$$

We complete the square in the numerator of the above fraction:

$$\begin{aligned} 2\sigma^2 T x - (x-\mu)^2 &= -x^2 + 2(\sigma^2 T + \mu)x - \mu^2 = -\big(x - (\sigma^2 T + \mu)\big)^2 + (\sigma^2 T + \mu)^2 - \mu^2 \\ &= -(x - (\sigma^2 T + \mu))^2 + \sigma^4 T^2 + 2\mu\sigma^2 T. \end{aligned}$$

Thus

$$\begin{aligned} x - \frac{(x-\mu)^2}{2\sigma^2 T} &= \frac{-(x - (\sigma^2 T + \mu))^2 + \sigma^4 T^2 + 2\mu\sigma^2 T}{2\sigma^2 T} \\ &= -\frac{(x - (\sigma^2 T + \mu))^2}{2\sigma^2 T} + \frac{1}{2}\sigma^2 T + \mu. \end{aligned}$$

Also note that

$$\frac{1}{2}\sigma^2 T + \mu = \frac{1}{2}\sigma^2 T + (r - \frac{1}{2}\sigma^2)T = rT.$$

and

$$\sigma^2 T + \mu = \sigma^2 T + (r - \frac{1}{2}\sigma^2)T = (r + \frac{1}{2}\sigma^2)T.$$

Therefore

$$
\begin{aligned}
A &= \frac{1}{\sqrt{2\pi\sigma^2 T}} \int_{\log(\frac{K}{S_0})}^{\infty} e^{-rT} S_0 \exp\big(x - \frac{(x-\mu)^2}{2\sigma^2 T}\big) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2 T}} \int_{\log(\frac{K}{S_0})}^{\infty} S_0 \exp\big(- \frac{(x-(r+\frac{1}{2}\sigma^2)T)^2}{2\sigma^2 T}\big) dx \\
&= S_0 P\big(\tilde{X} \geq \log(\frac{K}{S_0})\big),
\end{aligned}
$$

where $\tilde{X}$ has distribution $N((r+\frac{1}{2}\sigma^2)T, \sigma^2 T)$. Thus

$$
\begin{aligned}
A &= S_0 P(Z \geq \frac{\log(\frac{K}{S_0}) - (r+\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}) \\
&= S_0 N(d_1),
\end{aligned}
$$

where $d_1 = \frac{(r+\frac{1}{2}\sigma^2)T - \log(\frac{K}{S_0})}{\sigma\sqrt{T}}$.

This finishes the derivation of Black-Scholes formula.

**Remark 7.1.2.** *As we mentioned the heart of Black-Scholes formula is just computing the expectation of a log-normal random variable. So if we understand the technique, we can handle much more general European style derivative than just the European Call option. For example, we can price a European style derivative that pays $(S_T^2 - K)^+$ at time $T$. The details are given in the Section (7.2).*

### 7.1.4 Pricing a European call option at time $t$

**Theorem 7.1.3.** *Suppose that under $Q$,*

$$dS_t = rS_t dt + \sigma S_t dW_t.$$

*Then the no-arbitrage price at time $t$, $V_t$ of a European call with strike $K$ and expiration time $T$ satisfies*

$$V_t = S_t N(d_1(t)) - Ke^{-r(T-t)} N(d_2(t)),$$

*where*

$$
\begin{aligned}
d_1(t) &= \frac{(r+\frac{1}{2}\sigma^2)(T-t) - \log(\frac{K}{S_t})}{\sigma\sqrt{T-t}} \\
d_2(t) &= \frac{(r-\frac{1}{2}\sigma^2)(T-t) - \log(\frac{K}{S_t})}{\sigma\sqrt{T-t}}
\end{aligned}
$$

*Proof.* By the risk neutral pricing formula and the Independence Lemma:

$$\begin{aligned} V_t &= E^Q(e^{-r(T-t)}V_T|S_t) = E^Q(e^{-r(T-t)}(S_T - K)^+|S_t) \\ &= E^Q\Big[e^{-r(T-t)}\big(xe^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(W_T-W_t)} - K\big)^+\Big]\Big|_{x=S_t} \\ &= E^Q\Big[e^{-rT}\big(xe^X - K\big)^+\Big]\Big|_{x=S_t}, \end{aligned}$$

where $X$ has distribution $N((r - \frac{1}{2}\sigma^2)(T-t), \sigma^2(T-t))$.

Thus we see that we just repeat exactly the same calculation as the original Black-Scholes formula at time $t = 0$, replacing $S_0$ with $S_t$ and $T$ with $T - t$. The conclusion now follows.

## 7.2 Further discussion and examples

### 7.2.1 $\sigma^Q = \sigma$ between the risk neutral and physical measures

Why does $\sigma^Q$ have to be equal to $\sigma$ if $Q$ is equivalent to $P$? It is because of the following law of iterated logarithm: if $W_t$ is a BM under $P$ then

$$P\Big\{\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log(t)}} = 1\Big\} = 1.$$

Thus if $Q$ is to be equivalent to $P$, the Brownian motion $W_t$ cannot be a scaled version of $W_t^Q$. If $\sigma^Q \neq \sigma$ then essentially we would have $W_t^Q$ as a scaled version of $W_t$. Then the law of iterated logarithm would require

$$P^Q\Big\{\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log(t)}} = 1\Big\} = 1,$$

since $Q$ and $P$ are equivalent. But this cannot happen, since we also have

$$P^Q\Big\{\limsup_{t\to\infty} \frac{W_t^Q}{\sqrt{2t\log\log(t)}} = 1\Big\} = P^Q\Big\{\limsup_{t\to\infty} \frac{cW_t}{\sqrt{2t\log\log(t)}} = 1\Big\} = 1,$$

where $c$ is the scaling factor.

Since the two sets

$$\Big\{\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log(t)}} = 1\Big\}, \Big\{\limsup_{t\to\infty} \frac{W_t^Q}{\sqrt{2t\log\log(t)}} = 1\Big\}$$

are disjoint, it cannot happen that $P^Q$ of these two sets are 1.

### 7.2.2 Pricing a European style derivative with payment $V_T = (S_T^2 - K)^+$

**Lemma 7.2.1.** *Suppose we have a European-style derivative that pays $(S_T^2 - K)^+$ at time $T$ where $S_t$ is a geometric BM:*

$$dS_t = rS_t dt + \sigma S_t dW_t$$

*Then the no arbitrage at time 0 of this derivative is*

$$V_0 = e^{(r+\sigma^2)T}[S_0^2 N(\bar{d}_1) - e^{-rT}\bar{K}N(\bar{d}_2)].$$

*where*

$$\begin{aligned} \bar{d}_1 &= \frac{(r+2\sigma^2)T - \log(\frac{\bar{K}}{S_0^2})}{2\sigma\sqrt{T}} \\ \bar{d}_2 &= \frac{(r-2\sigma^2)T - \log(\frac{\bar{K}}{S_0^2})}{2\sigma\sqrt{T}} \\ \bar{K} &= \frac{K}{e^{(r+\sigma^2)T}} \end{aligned}$$

*Proof.* From the pricing formula:

$$V_0 = E(e^{-rT}(S_T^2 - K)^+).$$

Note that

$$S_T^2 = S_0^2 \exp((2r - \sigma^2)T + 2\sigma W_T).$$

We need to utilize the Black-Scholes formula, so we want to compare $S_t^2$ with a process with volatility $2\sigma$. So we consider the process $\bar{S}_t$ where

$$\begin{aligned} d\bar{S}_t &= r\bar{S}_t dt + 2\sigma\bar{S}_t dW_t \\ \bar{S}_0 &= S_0^2. \end{aligned}$$

That is

$$\bar{S}_t = S_0^2 \exp((r - 2\sigma^2)t + 2\sigma W_t). \tag{7.1}$$

The Black-Scholes formula gives

$$E(e^{-rT}(\bar{S}_T - K)^+) = \bar{S}_0 N(\bar{d}_1) - Ke^{-rT}N(\bar{d}_2),$$

where

$$\begin{aligned} \bar{d}_1 = \frac{(r+2\sigma^2)T - \log(\frac{K}{S_0^2})}{2\sigma\sqrt{T}} \\ \bar{d}_2 = \frac{(r-2\sigma^2)T - \log(\frac{K}{S_0^2})}{2\sigma\sqrt{T}} \end{aligned}$$

So in the original computation:

$$
\begin{aligned}
V_0 &= E(e^{-rT}(S_T^2 - K)^+) \\
&= E\left(e^{-rT}(S_0^2 \exp((2r-\sigma^2)T + 2\sigma W_T) - K)^+\right) \\
&= e^{(r+\sigma^2)T} E\left(e^{-rT}(S_0^2 \exp((r-2\sigma^2)T + 2\sigma W_T) - \bar{K})^+\right) \\
&= e^{(r+\sigma^2)T} E\left(e^{-rT}(\bar{S}_T - \bar{K})^+\right),
\end{aligned}
$$

where $\bar{K} = \frac{K}{e^{(r+\sigma^2)T}}$. Now we have rewritten the formula in the form similar to (7.1) with $\bar{S}_0 = S_0^2$. Therefore, the conclusion is

$$
V_0 = e^{(r+\sigma^2)T}[S_0^2 N(\bar{d}_1) - e^{-rT}\bar{K}N(\bar{d}_2)].
$$

■

*Proof.* Alternative derivation

We can rephrase Black-Scholes formula this way:

$$
e^{-rT}E^Q\Big((\bar{S}_0 e^{\bar{\mu}T+\bar{\sigma}W_T} - K)^+\Big) = e^{-rT}\Big(e^{(\bar{\mu}+\frac{1}{2}\bar{\sigma}^2)T}\bar{S}_0 N(d_1) - KN(d_2)\Big),
$$

where

$$
\begin{aligned}
d_1 &= \frac{(\bar{\mu}+\bar{\sigma}^2)T - \log\left(\frac{K}{S_0}\right)}{\bar{\sigma}\sqrt{T}} \\
d_2 &= \frac{\bar{\mu}T - \log\left(\frac{K}{S_0}\right)}{\bar{\sigma}\sqrt{T}}.
\end{aligned}
$$

Thus in computing

$$
V_0 = E(e^{-rT}(S_T^2 - K)^+) = E\Big(e^{-rT}(S_0^2 e^{(2r-\sigma^2)T+2\sigma W_T})^+\Big),
$$

we only have to plug in the above with

$$
\begin{aligned}
\bar{S}_0 &= S_0^2 \\
\bar{\mu} &= 2r - \sigma^2 \\
\bar{\sigma} &= 2\sigma.
\end{aligned}
$$

In this case, we have

$$
\begin{aligned}
V_0 &= e^{-rT}\Big(e^{(2r+2\sigma^2)T}S_0^2 N(d_1) - KN(d_2)\Big) \\
&= e^{(r+\sigma^2)T}S_0^2 N(d_1) - Ke^{-rT}N(d_2),
\end{aligned}
$$

where

$$d_1 = \frac{(2r+3\sigma^2)T - \log\left(\frac{K}{S_0^2}\right)}{2\sigma\sqrt{T}}$$

$$d_2 = \frac{(2r-\sigma^2)T - \log\left(\frac{K}{S_0^2}\right)}{2\sigma\sqrt{T}}.$$

we can check that this is exactly what we got in the first derivation. ■

### 7.2.3 Some further Black-Scholes computation suggestion

we can practice manipulating Black-Scholes formula by considering pricing, that is, find $V_t$, for the following European style derivative, all with expiration $T$, assuming the Black-Scholes model:

a. $V_T = \log(S_T)$;
b. $V_T = S_T^\beta, \beta$ a constant;
c. $V_T = \mathbf{1}_{S_T \geq K}$, the so-called Binary or cash or nothing option;
d. $V_T = \mathbf{1}_{K_1 \leq S_T \leq K_2}$, a generalization of the Binary option;
e. $V_{T_2} = \frac{S_{T_1}}{S_{T_2}}$, $T_1 < T_2$ are two fixed times, the option expires at $T_2$.

## 7.3 Merton's structural model of credit risk

In this section we discuss a nice example of Black-Scholes formula used in the context of modeling a company's credit risk. Roughly speaking, credit risk concerns the possibility of financial losses due to changes in the credit quality of the company. The most notable change in credit quality is a default event. Usually, default is triggered by a failure of the firm to meet its debt servicing obligations, which usually quickly leads to bankruptcy proceedings. Under structural models, a default event is deemed to occur for a firm when its assets reach a sufficiently low level compared to its liabilities. These models require strong assumptions on the dynamics of the firm's asset, its debt and how its capital is structured. The main advantage of structural models is that they provide an intuitive picture, as well as an endogenous explanation for default. We will discuss other advantages and some of their disadvantages in what follows.

The Merton model assumes that the total value $A_t$ of a firmâĂŹs assets follows a geometric Brownian motion under the physical probability:

$$dA_t = \mu A_t dt + \sigma A_t dW_t.$$

The company is funded by shares (equity) and bonds (debt). The Merton model assumes that debt consists of a single outstanding bond with face value $D$ and maturity T. At maturity, if the total value of the assets is greater than the debt, the latter is paid in full and the remainder is distributed among shareholders. However, if $A_T < D$ then default occurs: the bondholders exercise the right to liquidate the firm and receive the liquidation value (equal to the total firm value since there are no bankruptcy costs) in lieu of the debt. Shareholders receive nothing in this case, but by the principle of limited liability are not required to inject any additional funds to pay for the debt.

From these simple observations, we see that shareholders have a cash flow at T equal to

$$(A_T - D)^+,$$

and so equity can be viewed as a European call option on the firm's assets. That is

$$E_T = (A_T - D)^+.$$

The equity value $E_t$ of the firm at time $t$ can be found using Black-Scholes formula:

$$E_t = E^Q(e^{-r(T-t)}(A_T - D)^+ | A_t) = A_t N(d+) - De^{-r(T-t)} N(d-),$$

where

$$d\pm = \frac{(r \pm \frac{1}{2}\sigma^2)(T-t) - \log(\frac{D}{A_t})}{\sigma\sqrt{T-t}.}$$

Bond holders, on the other hand, receive

$$\begin{aligned} \min(D, A_T) &= A_T - (A_T - D)^+ \\ &= D - (D - A_T)^+. \end{aligned}$$

We can imagine this debt as a single defaultable bond with face value $D$ and recovered value $D - (D - A_T)^+$. The physical default probability is

$$P(B(T,T) < D) = P(A_T < D) = N(-(d-)).$$

The value $B(t,T)$ for this bond at earlier times $t < T$ can be obtained as the value of a zero-coupon bond minus a European put option. In the assumption of constant interest rate:

$$\begin{aligned} B(t,T) &= De^{-r(T-t)} - [De^{-r(T-t)} N(-(d-)) - A_t N(-d+)] \\ &= De^{-r(T-t)} N(d-) + A_t N(-d+). \end{aligned}$$

Even though it is not obvious in this form, it is true that

$$B(t,T) \leq e^{-r(T-t)} D.$$

One way to see this is from the risk neutral pricing formula of $B(t,T)$. This reflects the fact that the company bond holder faces greater risk (namely the default risk) than a non-defaultable bond holder. Thus they require a higher compensation, reflecting in the lower bond price than that of the zero-coupon non-defaultable bond.

The yield to maturity $Y(t,T)$ of a particular bond with face value $D$ is defined by

$$B(t,T) = e^{-Y(t,T)(T-t)}D.$$

That is

$$Y(t,T) = -\frac{1}{T-t}\log(\frac{B(t,T)}{D}).$$

Thus a zero-coupon non-defaultable bond has yield to maturity exactly equal to $r$. For the defaultable bond in the Merton model it is

$$Y(t,T) = -\frac{1}{T-t}\log\left(e^{-r(T-t)}N(d-) + \frac{A_t}{D}N(-d+)\right).$$

The difference $S(t,T)$ between the yield to maturity of a defaultable bond and the short rate is referred to as the yield-spread or the credit-spread. In this case it is

$$S(t,T) = -\frac{1}{T-t}\log\left(N(d-) + e^{r(T-t)}\frac{A_t}{D}N(-d+)\right).$$

$S(t,T)$ can be used as a measure of how “risky" the firm is. In practice, since $B(t,T)$ and $r$ are observable in the market, $S(t,T)$ can be directly calculated from market data.

The qualitative behaviour of the credit spread term structure is that credit spreads start at zero for $T = 0$, increase sharply to a maximum, and then decrease either to zero at large times if $r - \sigma^2/2 \leq 0$ or a positive value if $r - \sigma^2/2 > 0$. s. It is at odds with empirical observations in two respects: (i) observed spreads remain positive even for small time horizons and (ii) tend to increase as the time horizon increases.

*Estimating a firm’s asset and volatility using Merton’s model*

Structural models like MertonâĂŹs model depend on the unobserved variable $A_t$. On the other hand, for publicly traded companies, the share price (and hence the total equity) is closely observed in the market. The usual âĂIJad-hocâĂİ approach to obtaining an estimate for the firmâĂŹs asset values $A_t$ and volatility $\sigma$ in Merton’s model uses the Black-Scholes formula for a call option, that is,

$$E_t = BSCall(A_t, D, r, \sigma, T-t), \tag{7.2}$$

where $D$ and $T$ are determined by the firmâĂŹs debt structure. One combines this with a second equation by equating the equity volatility to the coefficient of the Brownian term obtained by applying Ito’s formula to 7.2, namely,

$$\sigma A_t = \frac{\partial BSCall}{\partial A} = \sigma^E E_t.$$

A consistent method is to use Duan’s maximum likelihood to estimate $\sigma$ and $\mu$ directly from the equity time series $E_i$. Once an estimate for $\sigma$ is obtained in this way, it can be inserted back into the pricing formula 7.2 in order to produce estimates for the firm values $A_i$ .

### 7.4 Binomial approximation to Black-Scholes price

In this section, we compare the price obtained by Black-Scholes formula with the Binomial tree model, where $S_t$ is a Geometric BM:

$$dS_t = rS_t dt + \sigma S_t dW_t \tag{7.3}$$

We will show through an example that the two prices are close to each other. This is not a surprising result, as we showed that the Black-Scholes are indeed obtained from the continuous limit of the Binomial model. we can believe that as the number of periods in the Binomial model grow (so that the length of the period decreases) the price obtained by the Binomial model will converge to the price given by the Black-Scholes model.

#### 7.4.1 The Black-Scholes price:

For simplicty, we let $r = 0, \sigma = 0.1, T = 1, S_0 = 1000$ and $K = 1000$. Then the Black-Scholes formula for European-Call is

$$V_0 = S_0 N(d_1) - KN(d_2)$$

where

$$\begin{aligned} d_1 &= \frac{\frac{1}{2}\sigma^2 T - \log(\frac{K}{S_0})}{\sigma\sqrt{T}} \\ &= 0.05 \end{aligned}$$

Similarly $d_2 = -0.05$. Thus $V_0 = 1000(0.52 - 0.48) = 40$.

### 7.5 The approximation:

We now divide $[0, 1]$ into $n = 5$ intervals. The discrete approximation to (7.3) is

$$S_{t_{k+1}} - S_{t_k} = rS_k(t_{k+1} - t_k) + \sigma S_k(B_{t_{k+1}} - B_{t_k})$$

where $k = 0, 1, ..., 5$ and $t_0 = 0, t_1 = 0.2, ..., t_4 = 0.8, t_5 = 1$.

$B_{t_{k+1}} - B_{t_k}$ has distribution $N(0, t_{k+1} - t_k)$. We approximate this by $\sqrt{t_{k+1} - t_k} Y_k$ where

$$\begin{aligned} Y_k &= 1 \text{ with probability } \frac{1}{2} \\ &= -1 \text{ with probability } \frac{1}{2}. \end{aligned}$$

For short-hand, we will write $S_k$ for $S_{t_k}$. The evolution equation for $S_k$ becomes

$$S_{k+1} = S_k(1 + \sigma\sqrt{t_{k+1} - t_k} Y_k).$$

Note that this is exactly the Binomial model we have studied before with $X_k = 1 + \sigma\sqrt{t_{k+1} - t_k} Y_k$. Plug in , we have

$$\begin{aligned} X_k &= 1.044 \text{ with probability } \frac{1}{2} \\ &= .956 \text{ with probability } \frac{1}{2}. \end{aligned}$$

Draw out the binomial tree, we see that the price for Euro Call on $S_k$ with strike 1000 and expiration time $n = 5$ is

$$V_0^b = \big(240 + 5 \times 135 + 10 \times 39.9\big)\frac{1}{2^5} = 41.06$$

This is not a very precise approximation to the Black-Scholes price of course (which gives 40 as in Section 2) but considering we only used 5 steps it is not terrible. The point of this computation is to convince you again that indeed the Geometric Brownian motion can be viewed as the limit of the Binomial tree as the time step gets closer to 0.

### 7.6 Exercises

In all of the following questions, suppose $S_t$ follow the Black-Scholes model under a risk neutral measure $Q$:

$$dS_t = rS_t dt + \sigma S_t dW_t.$$

1. Derive the Black-Scholes formula. Try to go as far as we can before consulting the notes or textbook.

2. Compute $V_0$ for the following European-style derivatives:

a. $V_T = (S_T^\beta - K)^+, \beta$ a constant.

b. $V_T = 1$ is $S_T < K$, $V_T = 0$ otherwise.

c. $V_T = S_T^\beta, \beta$ a constant.

d. $V_T = \log(S_T)$.

e. $V_T = 1$ if $K_1 < S_T < K_2$, $V_T = 0$ otherwise.

f. $V_T = (K - S_T^\beta)^+, \beta$ a constant.

3. Let $r = 0, \sigma = 0.1, T = 1, S_0 = 1000$ and $K = 1000$. Consider a European Call option on $S$ with strike $K$ and expiration $T$.

a. Compute the Black-Scholes price for $V_0$.

b. Compute the Binomial approximation to the Black-Scholes price with the number of steps $n = 5$.

c. Repeat part a with $r = 0.05$.

d. Repeat part b with $r = 0.05$.

# CHAPTER 8 The Black-Scholes PDE

## 8.1 Introduction

Consider the price $V_t$ at time of a European call with strike $K$ and expiration $T$:

$$V_t = E^Q(e^{-r(T-t)}(S_T - K)^+ | \mathcal{F}_t).$$

From either the Markov property of $S_t$ or the Indendence Lemma, we see that there is a function $v(t, x)$ (deterministic!) such that for all $t$,

$$V_t = v(t, S_t).$$

The question is can we derive an equation for $v(t, x)$? The answer is yes, and the equation is a Partial Differential Equation (PDE): an equation connecting the partial derivatives of $v$ in $t$ and $x$, hence the name.

This equation is of interest because if we can solve it, then to decide $V_t$ we only need to plug in $S_t$ for $x$. Of course we can decide $V_t$ by taking Expectation via the Independence Lemma, which leads to the Black-Scholes formula. Numerically, this would lead to the pricing by simulation method: we simulate the paths of $S_t$ and summing over the paths as way to approximate the expectation. The pricing of $V_t$ by by figuring out $v(t, x)$ would like to the numerical solution of PDE approach. This provides us with an alternative (and sometimes possibly more powerful) approach to the simulation method described above.

## 8.2 Two approaches to derive the Black-Scholes PDE

There are two approaches to derive the Black-Scholes PDE, by constructing either the replicating or the game theory portfolio. We describe both approaches here. The common key to both approaches is the dynamics of a self-financing portfolio: Let $\pi_t$ be the value of a self financing portfolio consisting of $n$ assets: $S^1, S^2, \cdots, S^n$. (They can be anthing from the underlying assets, the saving account to the financial derivative based on the underlyings). Let $\Delta_t^i, i = 1, \cdots, n$ be the number of shares of $S^i$ in the portfolio at time $i$. Then

$$d\pi_t = \Delta_t^1 dS_t^1 + \Delta_t^2 dS_t^2 + \cdots + \Delta_t^n dS_t^n.$$

The intuition is since the portfolio is self-financing, the only change in the portfolio value (from one period to another) is from the change in the asset price. You can make this more rigorous by writing from the period $[t_i, t_{i+1}]$:

$$\begin{aligned}
\pi_{t_i} &= \Delta^1_{t_i} S^1_{t_i} + \Delta^2_{t_i} dS^2_{t_i} + \cdots + \Delta^n_{t_i} dS^n_{t_i} \\
\pi_{t_{i+1}} &= \Delta^1_{t_i} S^1_{t_{i+1}} + \Delta^2_{t_i} dS^2_{t_{i+1}} + \cdots + \Delta^n_{t_i} dS^n_{t_{i+1}}.
\end{aligned}$$

Subtracting the two equations, letting $t_i \to t_{i+1}$ we have the dynamics of $\pi_t$ as described.

## 8.3 The game theory approach

### 8.3.1 The idea

Recall the game theory portfolio is a portfolio consisting of the underlying $S$ and the derivative $V$ such $\pi_T$ is a constant in $\omega$. By the no arbitrage condition this forces $e^{rT}\pi_0 = \pi_T$.

This argument can be repeat in a small time interval $[t, t+h]$ to the conclusion that

$$e^{rh}\pi_t = \pi_{t+h}$$

Using Taylor's approximation $e^{rh} \approx 1 + rh$ we have

$$r\pi_t h \approx \pi_{t+h} - \pi_t.$$

Letting $h$ goes to 0, we get

$$d\pi_t = r\pi_t dt.$$

Thus the first key is that **if the portfolio is a game theory portfolio, it has to satisfy**

$$d\pi_t = r\pi_t dt.$$

This is not surprising actually, since it says if a portfolio has deterministic growth, then the growth rate has to be the interest.

The next key is to use the self-financing condition: suppose we hold $\Delta_t$ shares of the underlying $S$ and 1 share of $V$ for our game theory portfolio then

$$d\pi_t = \Delta_t dS_t + dV_t.$$

Equating these two equations give

$$\Delta_t dS_t + dV_t = r(\Delta_t S_t + V_t)dt,$$

since $\pi_t = \Delta_t S_t + V_t$. The term $dV_t$ can be expanded by a Taylor expansion to terms involving the partials of the function $v(t,x)$ (recall that $V_t = v(t, S_t)$) with respect to $t$ and $x$ (by applying the Ito's formula). This will lead to a PDE for $v(t,x)$.

### 8.3.2 Game theory portfolio in continuous time

The above approach did not address a subtle but crucial point: how can we build a game theory portfolio in continuous time? I.e, how can we choose $\Delta_t$? The answer is not simple. In the 1 period model, we could build a portfolio because by the nature of the model, the stock price only changes at time $T$ and it only has 2 outcomes. In the continuous time, the stock price changes on the interval $[0, T]$ and it takes on a continuum of values. So solving for the number of shares $\Delta_t$ of the underlying $S_t$ at every time $t$ cannot be done directly. Instead, we introduce the following idea.

Let $S_t$ be a financial asset in continuous time with the following dynamics:

$$dS_t = \mu_t dt + \sigma_t dW_t.$$

We introduce the following terminology: we call $S_t$ a **risky asset** if $\sigma_t \neq 0$ and we call it a **risk-free** asset if $si_t = 0, \forall t$. That is a risk free asset must have its dynamics as:

$$dS_t = \mu_t dt.$$

Note that a risk-free asset does NOT have to be deterministic. The only requirement is its Brownian motion component is 0. In this way, the bond (or a saving account with variable (in time) and random interest rate) is a risk-free asset.

Back to our game theory portfolio, by applying Ito's formula to find $dV_t = dv(t, S_t)$, we see that $d\pi_t$ consists of a $dt$ and a $dW_t$ term. That is, $\pi_t$ is an Ito process. For simplicity let us write

$$d\pi_t = \mu_t^\pi dt + \sigma_t^\pi dW_t,$$

where $\mu_t^\pi, \sigma_t^\pi$ are some stochastic processes. Observe that by choosing $\Delta_t$ carefully, we can control $\sigma_t^\pi$ or $\mu_t^\pi$ (for example, possibly making $\sigma_t^\pi$ to be 0). If we can do this, then *we can turn $\pi_t$ into a risk-free asset.*

Why is this important? It is because the ONLY risk-free asset in an arbitrage-free market is the bond, or the saving account. More precisely, we have the following result:

**Lemma 8.3.1.** *Let $S_t$ be a risk-free asset. That is suppose*

$$dS_t = \mu_t dt.$$

*In addition, suppose $\mu_t$ is continuous in $t$. If the market is arbitrage free, then*

$$\mu_t = rS_t, \forall t.$$

*Proof.* Suppose not, then WLOG we assume $\mu_s > rS_s$ for some $s$. Since $\mu_s$ is continuous, we can find $t > s$ such that $\mu_u > rS_u, u \in [s, t]$. Then

$$S_t = S_s + \int_s^t \mu_u du > S_s + \int_s^t rS_u du.$$

That is

$$S_t > S_s e^{r(t-s)}.$$

Then it is clear that by borrowing money from the bank to invest in $S$ at time $s$, we will have an arbitrage opportunity. ■

To conclude, *to build a game-theory portfolio in continuous time is equivalent to build a risk-free portfolio.* This will be our approach in deriving the Black-Scholes PDE.

### 8.3.3 Derivation of the Black-Scholes PDE

*Goal:*

To show that under the model

$$dS_t = rS_t dt + \sigma S_t dB_t,$$

the price $v(t, S_t)$ of a European-style derivative that pays $\phi(S_t)$ at time $T$ satisfies

$$\begin{aligned} \frac{\partial}{\partial t}v(t,x) + \frac{\partial}{\partial x}v(t,x)rx + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x)\sigma^2 x^2 - rv &= 0 \\ v(t,x) &= \phi(x). \end{aligned}$$

*Ingredients*

1. Ito's formula
2. Game-theory portfolio: We hold $\Delta_t$ shares of stock at time $t$ and 1 share of $V$. We choose $\Delta_t$ such that the return of the portfolio is "deterministic".
3. No arbitrage principle: If a portfolio $\pi_t$ satisfies

$$d\pi_t = \mu(t)\pi_t d_t, \tag{8.1}$$

then we must have $\mu(t) = r$, for all $t$.

*Derivation of Black-Scholes PDE*

1. Apply Ito's formula:

$$dV_t = \frac{\partial}{\partial t}v(t,S_t) + \frac{\partial}{\partial x}v(t,S_t)dS_t + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,S_t)\sigma^2 S_t^2 dt.$$

Since

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

grouping the $d_t$ and $dB_t$ terms together we have

$$dV_t = \Big[\frac{\partial}{\partial t}v(t,S_t) + \frac{\partial}{\partial x}v(t,S_t)rS_t + \frac{1}{2}\frac{\partial^2}{\partial x^2}V(t,S_t)\sigma^2 S_t^2\Big]dt + \Big[\frac{\partial}{\partial x}v(t,S_t)\sigma S_t\Big]dW_t. \tag{8.2}$$

2.

a. By definition: $\pi_t = \Delta_t S_t + V_t$. By self-financing requirement:

$$\begin{aligned} d\pi_t &= \Delta_t dS_t + dV_t \\ &= \Delta_t\big(rS_t dt + \sigma S_t dW_t\big) + dV_t. \end{aligned}$$

Replace $dV_t$ by (8.2) and group $d_t, dW_t$ terms again we have

$$\begin{aligned} d\pi_t = \big[\Delta_t r S_t + \frac{\partial}{\partial t}v(t,S_t) + \frac{\partial}{\partial x}v(t,S_t) r S_t + \frac{1}{2}\frac{\partial^2}{\partial x^2}V(t,S_t)\sigma^2 S_t^2\big]dt \\ +\big[\Delta_t \sigma S_t + \frac{\partial}{\partial x}v(t,S_t)\sigma S_t\big]dW_t. \end{aligned}$$

b. Since $\pi_t$ is a game theory portfolio, it has the dynamics

$$d\pi_t = r\pi_t dt.$$

Comparing with the above equation, this forces the $dW_t$ term to be 0 or

$$\Delta_t = -\frac{\partial}{\partial x}v(t,S_t).$$

Then

$$\begin{aligned} d\pi_t &= \big[\frac{\partial}{\partial t}v(t,S_t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,S_t)\sigma^2 S_t^2\big]dt \\ &= \frac{\big[\frac{\partial}{\partial t}v(t,S_t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,S_t)\sigma^2 S_t^2\big]}{\pi_t}\pi_t d_t. \end{aligned}$$

This is in the form of (8.1) with

$$\mu(t) = \frac{\big[\frac{\partial}{\partial t}v(t,S_t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,S_t)\sigma^2 S_t^2\big]}{\pi_t}.$$

Therefore, we conclude that

$$\frac{\big[\frac{\partial}{\partial t}v(t,S_t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,S_t)\sigma^2 S_t^2\big]}{\pi_t} = r.$$

But $\pi_t = \Delta_t S_t + v(t,S_t) = -\frac{\partial}{\partial x}v(t,S_t)S_t + v(t,S_t)$. So we have

$$\frac{\partial}{\partial t}v(t,S_t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,S_t)\sigma^2 S_t^2 = r(-\frac{\partial}{\partial x}v(t,S_t)S_t + v(t,S_t)).$$

In other words

$$\frac{\partial}{\partial t}v(t,S_t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,S_t)\sigma^2 S_t^2 + r\frac{\partial}{\partial x}v(t,S_t)S_t - rv(t,S_t) = 0.$$

Lastly, since this is true for any value of $S_t$, replacing $S_t$ by $x$ we have

$$\frac{\partial}{\partial t}v(t,x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x)\sigma^2 x^2 + r\frac{\partial}{\partial x}v(t,x)x - rv(t,x) = 0.$$

This is the Black-Scholes PDE.

### 8.4 The replicating portfolio approach

#### 8.4.1 The idea

Recall the replicating portfolio is a portfolio consisting of the underlying $S$ and the saving account $y$ such $\pi_t = V_t$ for any time $t$. But this forces the dynamics of $\pi_t$ and $V_t$ to be the same:

$$dV_t = d\pi_t.$$

If we hold $\Delta_t$ in $S$ and $y_t$ in the saving account at time $t$ then (recalling that $dy_t = ry_t dt$)

$$dV_t = d\pi_t = \Delta_t dS_t + ry_t dt.$$

The $dV_t$ term can be expanded by Ito's formula as before. The key now is the above equation will become

$$\text{(something1) } dt + \text{ (something2) } dW_t = 0.$$

The second key is we have the **freedom** to choose $\Delta_t$ to make the $dW_t$ term to be 0. This forces the $dt$ term to be 0 as well, from which we can derive the PDE. The details are as followed.

#### 8.4.2 The derivation

1. Apply Ito's formula:

$$\begin{aligned} dV_t &= \frac{\partial}{\partial t}v(t,S_t) + \frac{\partial}{\partial x}v(t,S_t)dS_t + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,S_t)\sigma^2 S_t^2 dt \\ &= [\frac{\partial}{\partial t}v + rS_t\frac{\partial}{\partial x}v + \frac{1}{2}\frac{\partial^2}{\partial x^2}v\sigma^2 S_t^2]dt + \frac{\partial}{\partial x}v\sigma S_t dW_t. \end{aligned}$$

2. Since

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

we have

$$\begin{aligned} d\pi_t &= \Delta_t(rS_t dt + \sigma S_t dW_t) + ry_t dt \\ &= \Delta_t(rS_t dt + \sigma S_t dW_t) + r(\pi_t - \Delta_t S_t)dt. \end{aligned}$$

Since $V_t = \pi_t$, we conclude $dV_t = dS_t$ and

$$\begin{aligned} [\frac{\partial}{\partial t}v + rS_t\frac{\partial}{\partial x}v + \frac{1}{2}\frac{\partial^2}{\partial x^2}v\sigma^2 S_t^2]dt + \frac{\partial}{\partial x}v\sigma S_t dW_t &= \Delta_t(rS_t dt + \sigma S_t dW_t) + r(\pi_t - \Delta_t S_t)dt \\ &= \Delta_t(rS_t dt + \sigma S_t dW_t) + r(V_t - \Delta_t S_t)dt \\ &= rvdt + \Delta_t\sigma S_t dW_t. \end{aligned}$$

The above equation becomes:

$$[\frac{\partial}{\partial t}v + rS_t\frac{\partial}{\partial x}v + \frac{1}{2}\frac{\partial^2}{\partial x^2}v\sigma^2 S_t^2 - rv]dt + (\frac{\partial}{\partial x}v - \Delta_t)\sigma S_t dW_t = 0.$$

Choosing $\Delta_t = \frac{\partial}{\partial x}v(t, S_t)$ the $dW_t$ term is 0. (Note how this term *is exactly the negative of the choice for $\Delta_t$ in the game theory portfolio.* This is related to the Delta hedging concept). Then the $dt$ term has to be 0 as well, leaving us with

$$\frac{\partial}{\partial t}v + rS_t\frac{\partial}{\partial x}v + \frac{1}{2}\frac{\partial^2}{\partial x^2}v\sigma^2 S_t^2 - rv = 0.$$

This is exactly the same Black-Scholes equation we derived before.

### 8.5 Some remarks on Delta hedging

As we see from the derivation, in the game theory portfolio, we have to use $\Delta_t = -\frac{\partial}{\partial x}v(t, S_t)$ and in the replicating portfolio, we have to use $\Delta_t = \frac{\partial}{\partial x}v(t, S_t)$. This is an example of one of the Greeks in Math Finance, which we'll cover later.

The partial derivative of a financial derivative price (or a portfolio price) with respect to the underlying asset price is referred to as the Delta of the financial derivative at time $t$. It measures the sensitivity of the derivative price with respect to the underlying asset. (Since there are 2 "Deltas" floating around, I will use the Greek symbol $\Delta_t$ for the number of assets in the portfolio, and the English word Delta for the concept of partial derivative with respect to the underlying price).

We have seen that it may be desirable for a portfolio to have "stable" return over time (as in the game-theory portfolio). The intuitive idea is to make the Delta of our portfolio to be 0 at all time, so that the portfolio is protected against small change in the underlying asset price in the short run. In particular if we hold $\Delta_t$ share of $S_t$ and 1 share of the financial derivative in our portfolio, then

$$\pi_t = \Delta_t S_t + v(t, S_t).$$

Thus (assuming $\Delta_t$ is constant)

$$\begin{aligned}\frac{\partial}{\partial S_t}\pi_t &= \Delta_t\frac{\partial}{\partial S_t}S_t + \frac{\partial}{\partial S_t}v(t, S_t)\\ &= \Delta_t + \frac{\partial}{\partial S_t}v(t, S_t).\end{aligned}$$

It follows that the choice $\Delta_t = -\frac{\partial}{\partial x}v(t, S_t)$ will make $\frac{\partial}{\partial S_t}\pi_t = 0$. This choice of $\Delta_t$ is referred to as Delta hedging: it makes the Delta of the portfolio value to be 0 in the short run.

Note that the above derivation is NOT rigorous: we assume $\Delta_t$ to be constant (or at least independent of $S_t$) to pass the partial derivative w.r.t $S_t$ through it, only to conclude that it equals $-\frac{\partial}{\partial x}v(t, S_t)$ hence depends on $S_t$. But the calculation can be heuristically justified as in the short run, $\Delta_t$ can be thought of as approximately constant, and all of the above derivation should be looked at only in the approximate sense.

Thus, in practice, at any time moment $t$ one can choose the number of shares $\Delta_t$ so that the Delta of the portfolio is approximately 0 for a short time. But then the approximation will no longer be valid after a while (maybe a minute, half an hour etc, say at time $t + \varepsilon$). Then one will need to rebalance the portfolio at that moment to keep the Delta approximately 0 again. One cannot hope to choose $\Delta$ for all time $t$ (the buy and hold strategy) while also keep the Delta of the portfolio to be approximately 0 at all time.

## 8.6 An example

### 8.6.1 Goal:

To show that the price for a cash or nothing derivative: $V_T = \mathbf{1}_{\{S_T \geq K\}}$ at time $t$, which is

$$\begin{aligned} v(t, S_t) &= e^{-r(T-t)}N(d_2(t, S_t)), \\ d_2(t, S_t) &= \frac{(r - \frac{1}{2}\sigma^2)(T-t) - \log(\frac{K}{S_t})}{\sigma\sqrt{T-t}} \end{aligned}$$

satisfies the Black-Scholes PDE:

$$\begin{aligned} \frac{\partial}{\partial t}V + \frac{\partial}{\partial x}Vrx + \frac{\partial^2}{\partial x^2}V\sigma^2x^2 - rV &= 0 \\ v(t, x) = \mathbf{1}_{\{x \geq K\}}. \end{aligned}$$

### 8.6.2 Check the terminal condition:

We want to show that

$$\begin{aligned} v(t, x) &= 1 \text{ if } x \geq K \\ &= 0 \text{ if } x < K. \end{aligned}$$

Indeed if $x > K$ then $\frac{K}{x} < 1$ and $\log(\frac{K}{x}) < 0$. Therefore $d_2(T, x) = \infty$ and $N(d_2(T, x)) = 1$.

Similarly $x \leq K$ then $\frac{K}{x} \geq 1$ and $\log(\frac{K}{x}) \geq 0$. Therefore $d_2(T, x) = -\infty$ and $N(d_2(T, x)) = 0$.

### 8.6.3 Calculations:

1. Derivatives of $d_2(t,x)$:

$$\begin{aligned}
\frac{\partial}{\partial t}d_2(t,x) &= -\frac{r-\frac{1}{2}\sigma^2}{2\sigma\sqrt{T-t}} + \frac{\log(\frac{x}{K})}{2\sigma\sqrt{T-t}^3} \\
&= \frac{1}{2(T-t)}\left[d_2 - \frac{2}{\sigma}(r-\frac{1}{2}\sigma^2)\sqrt{T-t}\right] \\
\frac{\partial}{\partial x}d_2(t,x) &= \frac{1}{x\sigma\sqrt{T-t}} \\
\frac{\partial^2}{\partial x^2}d_2(t,x) &= -\frac{1}{x^2\sigma\sqrt{T-t}}.
\end{aligned}$$

2. Derivatives of $V$:

$$\begin{aligned}
\frac{\partial}{\partial t}V &= re^{-r(T-t)}N(d_2(t,x)) + e^{-r(T-t)}\phi_z(d_2(t,x))\frac{\partial}{\partial t}d_2(t,x) \\
&= rV + e^{-r(T-t)}\phi_z(d_2(t,x))\frac{\partial}{\partial t}d_2(t,x) \\
\frac{\partial}{\partial x}V &= e^{-r(T-t)}\phi_z(d_2(t,x))\frac{\partial}{\partial x}d_2(t,x) \\
\frac{\partial^2}{\partial x^2}V &= -e^{-r(T-t)}d_2(t,x)\phi_z(d_2(t,x))(\frac{\partial}{\partial x}d_2(t,x))^2 + e^{-r(T-t)}\phi_z(d_2(t,x))\frac{\partial^2}{\partial x^2}d_2(t,x).
\end{aligned}$$

3. Check the cancellations:

a.

$$\begin{aligned}
\frac{\partial}{\partial t}V - rV &= e^{-r(T-t)}\phi_z(d_2)\frac{1}{2(T-t)}\left[d_2 - \frac{2}{\sigma}(r-\frac{1}{2}\sigma^2)\sqrt{T-t}\right] \\
&= e^{-r(T-t)}\phi_z(d_2)\frac{1}{2(T-t)}d_2 - e^{-r(T-t)}\phi_z(d_2)\frac{1}{\sigma\sqrt{T-t}}(r-\frac{1}{2}\sigma^2).
\end{aligned}$$

b.

$$\begin{aligned}
\frac{\partial}{\partial x}Vrx &= \frac{re^{-r(T-t)}\phi_z(d_2)}{\sigma\sqrt{T-t}} \\
\frac{1}{2}\frac{\partial^2}{\partial x^2}V\sigma^2x^2 &= \frac{1}{2}\left[-\frac{e^{-r(t-t)}\phi_z(d_2)d_2}{T-t} - \frac{e^{-r(T-t)}\phi_z(d_2)\sigma}{\sqrt{T-t}}\right].
\end{aligned}$$

It is easy to see that everything cancels out now.

## 8.7 Exercises

In all of the following questions, suppose $S_t$ follow the Black-Scholes model under a risk neutral measure $Q$:

$$dS_t = rS_tdt + \sigma S_tdW_t.$$

1. Derive the Black-Scholes PDE using either the game theory or replicating portfolio. Try to go as far as you can before consulting the notes or textbook.

2. Compute $V_t$ for the following Euro-style derivatives:
a. $V_T = 1$ is $S_T < K$, $V_T = 0$ otherwise.
b. $V_T = S_T^{\beta}$, $\beta$ a constant.
c. $V_T = \log(S_T)$.

3. Verify that each of the $V_t$ you find in Question 2 satisfies the Black-Scholes PDE.

## CHAPTER 9 The Greeks

### 9.1 Motivation:

Suppose an option seller sells a Euro-style derivative that pays $V_T = \phi(S_T)$ at time $T$. We already learned that he should charge $V_0 = E^(e^{-rT}\phi(S_T))$ for the option at time 0.

Now the question is what should the option seller should do with $V_0$? He is obligated to pay out $\phi(S_T)$ (For example, $\phi(S_T) = (S_T - K)^+$ if the derivative is a Euro Call option) at time $T$. Certainly he cannot just invest $V_0$ in the bank and hope that he will have enough money to cover the random amount $\phi(S_T)$ that needs to be paid out at time $T$. Clearly he needs to invest $V_0$ in a portfolio that is a combination of the stock $S$ and the money market.

But how much should he hold in stocks? Recall from the binomial tree model, we learned that to hedge a Euro-style derivative, at any time $k$ the option seller should hold $\Delta_k := \frac{V_{k+1}-V_k}{S_{k+1}-S_k}$ shares of stock and put the rest of his money into the money market. Then at the expiration time $n$, the value of his portfolio will be exactly equal to $V_n$, the amount that needs to be paid out. We will apply this idea in continuous time as well. This is the idea of Delta hedging.

### 9.2 Delta hedging:

The idea: We divide the interval $[0, T]$ into $n$ subintervals, each with length $\delta$ ($\delta$ small). We denote each grid point of these subintervals by $t_k, 0 = t_0 < t_1 < ... < t_n = T$.

We construct a self-financing portfolio that consists of the underlying stock and the money market as followed: At each time $t_k$, we will hold $\Delta_k := \frac{\partial V}{\partial S}(t_k)$ shares of stock. We claim that in this way, the value of the portfolio at time $T$ will approximately be equal to the value of the derivative $V_T = \phi(S_T)$.

Reason: By Ito's formula

$$V_{t_{k+1}} - V_{t_k} \approx \big(\frac{\partial V}{\partial t}(t_k) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(t_k)\sigma^2 S_{t_k}^2\big)\delta + \frac{\partial V}{\partial S}(t_k)(S_{t_{k+1}} - S_{t_k}).$$

Since $V_t$ satisfies the Black-Scholes PDE, we have

$$\frac{\partial V}{\partial t}(t_k) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(t_k)\sigma^2 S_{t_k}^2 = -\frac{\partial V}{\partial S} rS(t_k) + rV(t_k).$$

Plug this in the above:

$$\begin{aligned} V_{t_{k+1}} - V_{t_k} &\approx \big(-\frac{\partial V}{\partial S} r S(t_k) + rV(t_k)\big)\delta + \frac{\partial V}{\partial S}(t_k)(S_{t_{k+1}} - S_{t_k}) \\ &= \big(V(t_k) - \frac{\partial V}{\partial S} S(t_k)\big) r\delta + \frac{\partial V}{\partial S}(t_k)(S_{t_{k+1}} - S_{t_k}). \end{aligned}$$

Now suppose at time $t_k$ we have a portfolio $\pi$ that satisfies $\pi(t_k) \approx V(t_k)$. We purchase $\frac{\partial V}{\partial S}(t_k)$ shares of stock, which leaves us with $\pi(t_k) - \frac{\partial V}{\partial S} S(t_k)$ to put into the bank. At time $t_{k+1}$ the value of our portfolio is (because of self-financing)

$$\pi(t_{k+1}) = \pi(t_k) + \big(\pi(t_k) - \frac{\partial V}{\partial S} S(t_k)\big) r\delta + \frac{\partial V}{\partial S}(t_k)(S_{t_{k+1}} - S_{t_k})$$

Note that we're in discrete time so the growth in 1 period of time of the money market portion is the interest rate times the length of that period, which is $\delta$.

But since $\pi(t_k) \approx V(t_k)$ we have

$$\begin{aligned} \pi(t_{k+1}) &\approx V(t_k) + \big(V(t_k) - \frac{\partial V}{\partial S} S(t_k)\big) r\delta + \frac{\partial V}{\partial S}(t_k)(S_{t_{k+1}} - S_{t_k}) \\ &\approx V(t_{k+1}). \end{aligned}$$

So the approximation extends to the next period. The quantity $\frac{\partial V}{\partial S}$, the first partial derivative of $V$ with respect to $S$, is thus seen to be very important in hedging, and it's called the Delta, in symbol $\Delta$, the first Greek we encounter in this section.

## 9.3 Computing $\frac{\partial V}{\partial S}$:

The above derivation is valid for any Euro-style derivative. However, the relevant question is: how much is exactly $\frac{\partial V}{\partial S}$? Or how to compute the Delta of a certain Euro derivative? This is difficult in general and usually one needs to use numerical techniques. However, when we specialize to certain cases of $V_T = \phi(S_T)$, for example $\phi(S_T) = S_T^k$ for some integer $k$ then explicit computation of the Delta is posssible. In this section we show how to compute the Delta of the most important derivative we encounter in this class: the Euro-Call option.

Recall that the Black-Scholes formula gives for a Euro call that pays $(S_T - K)^+$ at time $T$:

$$\begin{aligned} V(t, S_t) &= S_t N(d_1(t, S_t)) - Ke^{-r(T-t)} N(d_2(t, S_t)), \\ d_1(t, S_t) &= \frac{(r + \frac{1}{2}\sigma^2)(T-t) - \log(\frac{K}{S_t})}{\sigma\sqrt{T-t}} \\ d_2(t, S_t) &= \frac{(r - \frac{1}{2}\sigma^2)(T-t) - \log(\frac{K}{S_t})}{\sigma\sqrt{T-t}} \end{aligned}$$

It is also easy to see that

$$\frac{\partial}{\partial S} d_2(t, S_t) = \frac{\partial}{\partial S} d_1(t, S_t) = \frac{1}{S_t \sigma \sqrt{T-t}}.$$

Therefore,

$$\frac{\partial V}{\partial S}(t) = N(d_1(t,S_t)) + S_t\phi_Z(d_1(t,S_t))\frac{1}{S_t\sigma\sqrt{T-t}} - Ke^{-r(T-t)}\phi_Z(d_2(t,S_t))\frac{1}{S_t\sigma\sqrt{T-t}}.$$

We claim that

$$\phi_Z(d_1(t,S_t)) = Ke^{-r(T-t)}\phi_Z(d_2(t,S_t))\frac{1}{S_t}.$$

To see this, note that $d_1(t,S_t) = d_2(t,S_t) + \sigma\sqrt{T-t}$. Therefore,

$$\begin{aligned}\phi_Z(d_1) &= \phi_Z\big(d_2 + \sigma\sqrt{T-t}\big) \\ &= \frac{1}{\sqrt{2\pi}}\exp\big(-\frac{(d_2+\sigma\sqrt{T-t})^2}{2}\big) \\ &= \phi_Z(d_2)\exp\big(\frac{-2d_2\sigma\sqrt{T-t}-\sigma^2(T-t)}{2}\big).\end{aligned}$$

One can check that

$$2d_2(t,S_t)\sigma\sqrt{T-t} + \sigma^2(T-t) = 2\big(r(T-t) - \log(K) + \log(S_t)\big).$$

Plug this into the above expression, the claim is checked. Thus we see a surprisingly simple result: $\frac{\partial V}{\partial S}(t) = N(d_1(t,S_t))$.

### 9.4 Predicting the future price of Euro Call option - Theta, Delta and Gamma

So we see that the partial derivative of $V$ with respect to $S$ plays an important role in hedging. Indeed the Greeks are just various partial derivatives of $V$ with respect to different parameters in the Black-Scholes model: $t, r, \sigma, T, S$. Some of them show up more ofen than others. In particular, two more Greeks that are important for our purpose are the ones that appear in Ito's formula:

$$\begin{aligned}\Theta(t) &:= \frac{\partial V}{\partial t}(t) \\ \Gamma(t) &:= \frac{\partial^2 V}{\partial S^2}(t),\end{aligned}$$

and of course previously we have

$$\Delta(t) := \frac{\partial V}{\partial S}(t).$$

Note that in this way the Greeks are random processes. They are functions of $t$ and $S_t$. Their use is to measure the sensitivity of the option price with respect to the change of other parameters in the model. Again in general it may be difficult to compute the $\Theta, \Gamma$ of a general derivative. But if we specialize to certain form of $\phi(S_T)$ then the computation can be doable. In particular, for the Euro-Call option:

$$\begin{aligned}\Gamma(t) &= \frac{\phi_Z(d_1(t,S_t))}{\sigma S_t\sqrt{T-t}}\\ \Theta(t) &= rKe^{-r(T-t)}N(d_2(t,S_t)) + \frac{\sigma S_t\phi_Z(d_1(t,S_t))}{2\sqrt{T-t}}.\end{aligned}$$

The formulas are complicated, but they are explicit and one can compute these quantities provided $S_t, \sigma, r, T$ are given. Also at time $t$, using Black-Scholes formula we also know $V_t$. Therefore, Ito's formula gives for a small change in time $t+\delta$

$$V_{t+\delta} \approx V_t + \big(\Theta(t) + \frac{1}{2}\Gamma(t)\sigma^2 S^2(t)\big)\delta + \Delta(t)(S_{t+\delta} - S_t).$$

Note: The book used $V_{new}$ for $V_{t+\delta}$ and only consider the case $t = 0$. So their formula is simpler than ours and our formula is slightly more general.

## 9.5 Comparing option price - Vega and Rho

### 9.5.1 Vega - $\nu(t, S_t)$

We're interested in the following question: Suppose

$$\begin{aligned}dS_t^i &= \mu^i S^i t + \sigma S_t^i dW_t,\\ S_0^i &= S_0,\end{aligned}$$

*under the physical measure* $P$. That is the two stocks have different average return and same volitility, with the same initial price. Let $V_t^i$ be the price of Euro Call option with expiration $T$ and strike $K$ on $S^i$. Suppose that $\mu^1 > \mu^2$. Can we conclude that $V_t^1 > V_t^2$?

This is actually a trick question. The answer is NO, $V_t^1 = V_t^2$. The reason is we need to price these options under the risk neutral measure. And under the risk neutral measure, the dynamics of $S^i$ are

$$\begin{aligned}dS_t^i &= rS^i t + \sigma S_t^i dW_t,\\ S_0^i &= S_0.\end{aligned}$$

That is they have the SAME dynamics with the same initial condition. So the corresponding call options on them have the same price.

A more interesting question would be what if they have different volitility? Suppose that

$$\begin{aligned}dS_t^i &= rS^i t + \sigma^i S_t^i dW_t,\\ S_0^i &= S_0,\end{aligned}$$

where $\sigma^1 > \sigma^2$ under $Q$. What can we say about $V_t^1$ versus $V_t^2$? Note that appealing to the Black-Scholes formula is not straightforward because

$$\begin{aligned} V(t,S_t) &= S_t N(d_1(t,S_t)) - Ke^{-r(T-t)}N(d_2(t,S_t)), \\ d_1(t,S_t) &= \frac{(r+\frac{1}{2}\sigma^2)(T-t) - \log(\frac{K}{S_t})}{\sigma\sqrt{T-t}} \\ d_2(t,S_t) &= \frac{(r-\frac{1}{2}\sigma^2)(T-t) - \log(\frac{K}{S_t})}{\sigma\sqrt{T-t}}, \end{aligned}$$

and we have the presence of $\sigma$ at BOTH the numerator and the denominator of the fractions in $d_1, d_2$. So one can't say straightfowardly that if $si$ increases then $V(t, S_t)$ increases.

The way to answer this question is to differentiate $V(t, S_t)$ with respect to $\sigma$ and look at the sign of the derivative. Doing this is equivalent to find the Greek $\nu(t, S_t) = \frac{\partial}{\partial\sigma}V(t, S_t)$ of the Euro Call option. The computation is as followed: note that

$$\begin{aligned} \frac{\partial}{\partial\sigma}d_1(t,S_t) &= \frac{-r+\log(\frac{K}{S_t})}{\sigma^2\sqrt{T-t}} + \frac{1}{2}\sqrt{T-t} \\ \frac{\partial}{\partial\sigma}d_2(t,S_t) &= \frac{-r+\log(\frac{K}{S_t})}{\sigma^2\sqrt{T-t}} - \frac{1}{2}\sqrt{T-t}. \end{aligned}$$

Therefore

$$\begin{aligned} \nu(t,S_t) &= S_t\phi_Z(d_1(t,S_t))\left[\frac{-r+\log(\frac{K}{S_t})}{\sigma^2\sqrt{T-t}} + \frac{1}{2}\sqrt{T-t}\right] \\ &- Ke^{-r(T-t)}\phi_Z(d_2(t,S_t))\left[\frac{-r+\log(\frac{K}{S_t})}{\sigma^2\sqrt{T-t}} - \frac{1}{2}\sqrt{T-t}\right]. \end{aligned}$$

This, coupled with the fact discusses above that

$$\phi_Z(d_1(t,S_t)) = Ke^{-r(T-t)}\phi_Z(d_2(t,S_t))\frac{1}{S_t}$$

gives a rather simple formula for $\nu$:

$$\nu(t,S_t) = S_t\phi_Z(d_1(t,S_t))\sqrt{T-t}.$$

We can now answer our original question. If $\sigma_1 > \sigma_2$ then $V_t^1 > V_T^2$, since $\nu(t, S_t) \geq 0$.

Remark: The above result can also be justified with the following intuitive argument: since a call option is like an insurance, and since the larger the volitility, the riskier the stock is, the call option on the stock with the higher volitility should have a higher price. Note that this argument also applies to the put option. So we predict that the put option on the stock with the higher volitility should have a higher price. We do NOT have to compute the vega of the Put option to justify this. It simply follows from the Put-Call parity: $V_t^c - V_t^p = S_t - K$. Thus if $V_t^c$ increases, $V_t^p$ also have to increase.

### 9.5.2 Rho - $\rho(t, S_t)$

We consider the following situation

$$\begin{aligned} dS_t^i &= r^i S^i t + \sigma S_t^i dW_t, \\ S_0^i &= S_0, \end{aligned}$$

and $V_t^i$ corresponding price of Euro Call option on $S^i$ with strike $K$ and expiration $T$. This can be interpreted as evaluating the price of the call option on the same stock in different periods of the economy that has different interest rate. Can we compare $V^1$ and $V^2$, say if $r^1 > r^2$?

From the discussion in the previous section, you see that we need to compute $\rho(t, S_t) = \frac{\partial}{\partial r} V(t, S_t)$. You can verify that

$$\rho(t, S_t) = K(T-t)e^{-r(T-t)} N(d_2(t, S_t)).$$

Thus, since $\rho(t, S_t) \geq 0$, we also have $V_t^1 > V_t^2$.

## 9.6 Exercises

1. Let $\rho = \frac{\partial V_0}{\partial r}$ where $V_T = (S_T - K)^+$. Prove that

$$\rho = KTe^{-rT}N(d_2),$$

where

$$d_2 = \frac{(r - \frac{1}{2}\sigma^2)T - \log(K/S_0)}{\sigma\sqrt{T}}$$

# CHAPTER 10 Discrete time interest rate model

## 10.1 Discrete bond model

Consider the binomial multi-period model $0, 1, \cdots, N$. At each time $k$ there are $k+1$ possible outcomes. Consider a bond with maturity $N$ and face value $F$. That is at time $N$ the bond holder will receive $F$ dollars (fixed at time 0). We want to model interest rate $r_k$ and the bond price $B_k$ at time $k$. The length of each period will be denoted as $\Delta T$.

However, what we learned from the previous binomial model cannot be directly used. First we cannot model the bond price following the stock's approach: given $B_0$, define

$$\begin{aligned} B_{k+1} &= uB_k \text{ with probability } p \\ &= dB_k \text{ with probability } 1-p. \end{aligned}$$

The reason is that the bond is financial product that we want to price, so we cannot model it directly. A zero-coupon bond with maturity $N$ is a financial contract that pays the holder 1 dollar at time $N$. Observe that if we follow the stock's binominal modelling approach then there is no way we can get $B_N = 1$. We need to work backward from $B_N = 1$ to figure out $B_k, k = 0, \cdots, n-1$.

The bond is not a financial derivative (it does not derive its value from other underlying). However, the bond price is influenced by interest rate. This suggests that we model the interest rate directly and simply use the pricing formula:

$$B_k = E^Q\left(\frac{1}{1+r_k\Delta T}B_{k+1}|\mathcal{F}_k\right). \tag{10.1}$$

We discuss the notation. $\mathcal{F}_k$ represents the state of the world at time $k$, which is known to us at time $k$. We use disrete compounding in this section, hence the discounting factor $\frac{1}{1+r_k\Delta T}$. Interest rate is allowed to be random and varying with time, that is why we need to put $\frac{1}{1+r_k\Delta T}$ inside the expectation.

In general, the pricing formula for a bond $B$ with face value $F$ (note: $F$ is a constant) and maturity $N$ is

$$B_k = E^Q\left(\prod_{i=k}^{N-1}\frac{1}{1+r_i\Delta T}F|\mathcal{F}_k\right). \tag{10.2}$$

Note that $r_k$ is known at time $k$ but $r_i, k+1 \le i \le N-1$ is not known until time $i$.

$r_k$ is referred to as the short rate. In the context of the discrete model, we understand $r_k$ as the interest rate available for depositing with the money market account (or borrowing) in the period $[k, k+1]$.

If we want to deposit for a longer period (say from time $k$ to $k+2$) there are two options. Either we deposit from time $[k, k+1]$ with interest rate $r_k$ and then roll over to time $[k+1, k+2]$ with interest rate $r_{k+1}$; or we can deposit from time $[k, k+2]$ with the spot rate $R(k, k+2)$ (compounding twice at $t = k, k+1$). The spot rate $R(k, N)$ is simply the interest that we would earn for depositing during time period $[k, N]$ (compounding $N-k$ times at $t = k, k+1, \cdots, N-1$). That is if we deposit 1 dollar at time $k$ and do not withdraw until time $N$ then at time $N$ we would receive $(1 + R(k, N)\Delta T)^{N-k}$ dollars. There must be a consistency between these rates; otherwise there would be an arbitrage opportunity. In particular we have

$$B(k, N) = \frac{1}{(1 + R(k, N)\Delta T)^{N-k}} = E^Q(\prod_{i=k}^{N-1} \frac{1}{1 + r_i \Delta T} F|\mathcal{F}_k).$$

The difference is as followed: $R(k, N)$ is a random variable that is known at time $k$. $r_k$ is also known at time $k$ but $r_i, k+1 \le i \le N-1$ is not known until time $i$. Note that in this sense, the spot rate $R(k, N)$ coincides with the yield to maturity of the zero-coupon bond $B(k, N)$ for the time period $[k, N]$. See also section (10.6).

Observe also that we assume here that $R(k, N)$ is a compound interest rate that is compounded at the end of each period. If $R(k, N)$ is a simple interest rate the above equation would read

$$B(k, N) = \frac{1}{(1 + R(k, N)(N-k)\Delta T)} = E^Q(\prod_{i=k}^{N-1} \frac{1}{1 + r_i \Delta T} F|\mathcal{F}_k).$$

That is, the interest is only compouned once at time $t = k$. In this case, $R(k, N)$ is referred to as the LIBOR rate available for the period $[k, N]$. See also the section on the forward rate agreement below for further discussion.

## 10.2 Stochastic short rate model

Now we turn to modelling $r_k, k = 0, \cdots, N$. One important observation here is that we need to model $r_k$ directly under the risk neutral measure $Q$. That is we need to take the risk neutral measure $Q$ as given exogenously (maybe from working with other financial dirvatives) and model $r_k$ under this measure. The reason is if our market only consists of bond then there is no reference underlying, such as a stock, for us to define $Q$ upon. In the lack of such reference product, we take our default risk neutral measure as $q = 1-q = 1/2$ for example.

We need to be careful on how to model $r_k$. We can use the binomial approach

$$\begin{aligned} r_{k+1} &= ur_k \text{ with probability } q \\ &= dr_k \text{ with probability } 1-q; \end{aligned}$$

but it may not be realistic as this allows for the interest rate to grow quite large in some particular event (in contrast to the real world observation that interest rate tends to stay around some fixed level, say around $1\%$, for a period of time). If we want to avoid specifying a formula, we can just simply 'fill in the nodes of the binomial tree for the intrest rate value at that node. For example if $N = 3$ we can specify

$$r_3(uuu) = 0.03, r_3(udu) = 0.025, r_d(ddu) = 0.0125, r_d(ddd) = 0.025,$$

and so on for $r_2, r_1, r_0$. Once we have the interest rate tree, we can use formula (10.1) to figure out the bond price with all maturities up to $N$.

### 10.2.1 A mean reverting model

Instead of filling out the nodes of a binomial short rate tree, we can specify a mean-reverting formulation for the short rate as followed:

$$r_{k+1} = r_k + \theta(\mu - r_k)\Delta T + \sigma X_k.$$

Here $X_k, k = 0, 2, \cdots, N-1$ are iid Bernoulli-type random variables such that

$$E^Q(X_k) = qu + (1-q)d = 0.$$

$\mu$ is the average level that the short rate "reverts" to in the long run and $\theta$ is the mean-reverting rate. In terms of modeling, here we get the freedom to choose $u, d$ once $q$ is given. See also section (10.4) for more information on how $q$ might be determined. This is a discretization version of the continuous time Hull-White model. The idea is on average, the movement of $r_k$ looks like

$$r_{k+1} = r_k + \theta(\mu - r_k)\Delta T.$$

Thus if $r_k > \mu$' the term $\theta(\mu - r_k)\Delta T < 0$ which pulls $r_k$ closer to $\mu$. Similarly if $r_k < \mu$ the term $\theta(\mu - r_k)\Delta T > 0$ which again pulls $r_k$ closer to $\mu$.

## 10.3 The money market account

If we deposit 1 dollar at time $k$, we will earn $1 + r_k\Delta T$ dollars at time $k+1$. The money market account $M_k$ is the value at time $k$ of the account that deposits 1 dollar at time $0$. That is

$$M_k = \prod_{i=0}^{k-1}(1 + r_i\Delta T).$$

Observe that even though $M_k$ is random, $M_{k+1}$ is known at time $k$ (since $r_k$ is determined at time $k$). In this way we say the money market account process $M_k$ is predictable.

The money market account can be used in the pricing formula as followed: for $0 \leq i < j \leq N$

$$B_i = E^Q(\prod_{k=i}^{j-1} \frac{1}{1 + r_k \Delta T} B_j | \mathcal{F}_i).$$

Observe that $\prod_{k=i}^{j-1} \frac{1}{1+r_k\Delta T} = \frac{M_j}{M_i}$ and $M_i$ is known at time $i$ we can write the pricing formula succintly as

$$M_i B_i = E^Q(M_j B_j | \mathcal{F}_i).$$

## 10.4 Option pricing in stochastic interest rate model

Consider a binomial tree model where we have the underlying $S$ that follows the model

$$\begin{aligned} S_{k+1} &= uS_k \text{ with probability } p \\ &= dS_k \text{ with probability } 1 - p. \end{aligned}$$

Suppose we have a call option on $S$ with strike $K$ and expiration $N$. That is $V_N = (S_N - K)^+$. We want to find $V_0$, given a stochastic interest rate model. To this end, suppose that a stochastic interest tree has already be given as in section (10.2). Note though that we cannot specify the risk neutral measure $q = 1/2$ by default as in that section. The reason is because we have a reference underlying here. So we need to use the definition:

$$S_k = E^Q(\frac{1}{1 + r_k \Delta T} S_{k+1} | S_k).$$

This implies

$$(1 + r_k \Delta T) S_k = q_k u S_k + (1 - q_k) d S_k,$$

or

$$(1 + r_k \Delta T) = q_k u + (1 - q_k) d.$$

One can solve for $q_k$ from there. Note something subtle here. Not only that $q$ depends on $k$ as $r$ depends on $k$ but it also depends on the random event that happens at time $k$. That is the $q_k$ in the above equation is a conditional probability. It is the risk neutral probability that given an even $\omega_k$ at time $k$, the next event twill be $\omega_k u$:

$$P^Q(\omega_k u | \omega_k) = q_k(\omega_k).$$

Once we can fill out all the risk neutral probabilities this way we can proceed to compute the option price via the backward pricing method as we discussed before.

## 10.5 Coupon paying bond and future cash flow

So far we have discussed zero-coupon bond. That is bond with only a face value, here denominated as 1 dollar, which is paid at the maturity time $N$. More generally, one can consider a coupon paying bond with coupon rate $C$ and face value $F$. That is at each period $k = 0, 1, \cdots, N$ (determined by the bond, usually annually or semi-annually) the bond holder receives $C\%$ of the face value and receive the face value $F$ dollars at the maturity time $N$. The pricing of such a coupon paying bond, given an interest rate tree, is similar to the method discuss above. Indeed, we can look at the zero-coupon bond as a future income stream that pays us $C_k = \frac{C}{100}F$ at time $k$. We can compute the present-value $C_k^0$ of a future payment $C_k$ at time $k$ by the risk neutral pricing formula

$$C_k^0 = E^Q(\frac{1}{M_k}C_k).$$

Then the present value of the zero-coupon bond is the sum of the present values in the future income stream:

$$C_0 = \sum_{k=0}^{N} C_k^0.$$

In the case of a coupon bond with a fixed coupon rate (say $5\%$) there is even a more straightforward method to compute its present value without the need for modeling the interest rate. Indeed, the present value of $1$ dollar paid at time $k$ is just $B(0,k)$, the price of a zero coupon bond that paids $1$ dollar at time $k$. Thus in the above formula, $C_k^0 = B(0,k)C_k = B(0,k)\frac{C}{100}F$.

## 10.6 Yield to maturity

Precisely speaking, bond price is a function of two time variables: the current time $k$ and the maturity time $N$. Thus we write $B(k,N)$ for the price at time $k$ of a bond with maturity $N$. For a present value of time, say $k = 0$, we can discuss the price of the bond that matures at time $N = 1, 2, \cdots$. Term structure refers to the concept of fixing a present time and varies the maturity of a fixed income security, typically a bond. The term structure reflects the public opinion on how the interest rate will behave in the future.

A particular quantity of interest is the bond's yield to maturity. For a bond with maturity at time $N$, the yield to maturity $\lambda(0,N)$ is defined such that

$$B(0,N) = \frac{1}{(1+\lambda(0,N)\Delta T)^N}. \tag{10.3}$$

If we plot $B(0,N)$ as a function of $N$ it will be a decreasing graph. However, if we plot $\lambda(0,N)$ as a function of $N$ it will usually be an increasing graph. This reflects the fact that people usually require a higher compensation for a longer duration of the loan.

Furthermore, for a coupon paying bond with coupon rate $C$, face value $F$ and maturity $N$, its yield to maturity $\lambda(0,N)$ is defined such that

$$B(0,N) = \sum_{k=0}^{N} \frac{\frac{C}{100}F}{(1+\lambda(0,N)\Delta T)^k} + \frac{F}{(1+\lambda(0,N)\Delta T)^N}. \tag{10.4}$$

### 10.7 Forward rate agreement

A forward rate agreement (FRA) is a contract that allows the holder to pay a fixed rate $K(0,T_1,T_2)$ and receive a floating rate $L(T_1,T_2)$ for the loan period $[T_1,T_2]$. Suppose all interests are simple. The contract holder will receive (for a notional amount of 1 dollar)

$$(L(T_1,T_2) - K(0,T_1,T_2))(T_2 - T_1),$$

at time $T_2$. Here $L(T_1,T_2)$ is simply understood as the simple interest available for depositing or borrowing 1 dollar in the period $[T_1,T_2]$. A typical quote for $L(T_1,T_2)$ follows the Libor (London Interbank Offered Rate). Libor rates are calculated for 5 currencies and 7 borrowing periods ranging from overnight to one year and are published each business day by Thomson Reuters. For convenience, we can assume $L(T_1,T_2)$ falls into these 7 borrowing periods. If not, it can be figured out using certain interpolation methods from the given Libor rates. We emphasize that $L(T_1,T_2)$ is a random variable that is only known at time $T_1$ while $K(0,T_1,T_2)$ is a constant that is known at time $t=0$.

Assume $T_1, T_2$ are two points on our discrete grid. The Libor rate must be consistent with the short rate to avoid arbitrage. In fact, the simple Libor rate is related to the compouding short rate as followed

$$1 + L(T_1,T_2)(T_2 - T_1) = \prod_{i=1}^{n}(1 + r_i(t_{i+1} - t_i)), \tag{10.5}$$

where we assume $t_1 = T_1$ and $t_n = T_2$. There are more implications about the relations between LIBOR rate and short rate in terms of forward rate and interest rate swap which we will see more below.

The forward rate $f(0,T_1,T_2)$ is the fixed rate (the strike) $K(0,T_1,T_2)$ so that the FRA costs nothing to enter at the present time $t=0$. It turns out that we can calculate $f(0,T_1,T_2)$ as followed:

$$f(0,T_1,T_2) = \frac{1}{T_2 - T_1}\left(\frac{B(0,T_1)}{B(0,T_2)} - 1\right).$$

The reasoning is as followed. Consider two investing schemes. The first one is to enter a FRA at time 0, invest 1 dollar at time $T_1$ which gives us $(1+f(0,T_1,T_2))(T_2-T_1)$ at time $T_2$. The present value of this scheme is $B(0,T_2)(1+f(0,T_1,T_2))(T_2-T_1)$. The second one is to invest 1 dollar at time $T_1$ which gives us $1 + L(T_1,T_2)(T_2-T_1)$ at time $T_2$. The present value of this scheme is $B(0, T_1)$. This may be a bit surprising. But it really comes from the fact that $1 + L(T_1, T_2)(T_2 - T_1)$ translates to 1 dollar at time $T_1$ which translates to $B(0, T_1)$ at time 0. These two schemes should have the same present value at time 0 to avoid arbitrage. That is

$$B(0, T_1) = B(0, T_2)(1 + f(0, T_1, T_2))(T_2 - T_1).$$

From this equation we can solve for $f(0, T_1, T_2)$ as indicated.

This equation also suggests a replication scheme for the FRA as followed: at time 0 we sell 1 FRA, short 1 share of zero coupon bond with maturity $T_1$ and long $(1 + f(0, T_1, T_2))(T_2 - T_1)$ shares of zero coupon bond with maturity $T_2$. At time $T_1$ we receive 1 dollar from the FRA buyer and we use it to close out our short position in the bond $B(0, T_1)$. At time $T_2$ we receive $(1 + f(0, T_1, T_2))(T_2 - T_1)$ from our long position with the bond $B(0, T_2)$ and use it to close out our short position in the FRA contract.

### 10.8 Interest rate futures

An interest rate future is a contract that costs nothing to enter, and allows the holder to pay a fixed rate $F(0, T_1, T_2)$ and receive a floating rate $R(T_1, T_2)$ for the loan period $[T_1, T_2]$. The contract holder is required, however, to post a margin account to reflect the fluctuation in the contract price as time moves from $t = 0$ to $t = T_1$.

In practice, the interest rate future is executed as followed:

1. The contract is settled at time $T_1$ and the contract holder does not (typically) enter into the loan period.

2. The contract holder is required to post a margin account. If the future rate $F(t, T_1, T_2)$ increases, he has to deposit more money into the account. If the future rate $F(t, T_1, T_2)$ decreases he receives money into the account.

3. The future contract is quoted with the price $100 - F(t, T_1, T_2)$ for $0 \leq t \leq T_1$. Its price at time $T_1$ is $100 - R(T_1, T_2)$.

If the interest rate is constant then the futures rate and forward rate are equal. However, in practice the forward rate is less than the future rate, due to the margining requirement. This phenomenon is called the negative convexity of the futures rate. Intuitively, it can be explained as followed. The margining always works against the futures contract holder (and works for the futures contract seller). When they receive margin, this will coincide with a drop in the rate at which the balance in the margin account is rolled. When they have to deposit margin, this will coincide with an increase in the borrowing rate. Thus if the forward rate is equal to the futures rate, one can buy a forward rate agreement and sell a futures contract to profit from the margining effect.

### 10.9 Interest rate swaps

An interest rate swap, set over a period of times $t_0, t_1, \cdots, t_n$ is like a sequence of FRA where at each time $t_i$ one party (referred to as the payer) pays an agreed upon (fixed) rate in exchange for a floating rate, typically the Libor during the period $[t_i, t_{i+1}]$. The party who receives the fixed rate (and pays the floating rate) is referred to as the receiver. The contract is entered at time $t = 0$ and the first swap payment happens at $t = t_1$.

If we denote the fixed rate as $K$ then at each time $t_i$ the paying leg receives

$$(L(t_{i-1}, t_i) - K)(t_i - t_{i-1}).$$

We compute the present value of the total cash flow of the fixed leg and the floating leg separately. The present value of the total cash flow of the fixed leg is

$$\sum_{i=1}^{n} K(t_i - t_{i-1})B(0, t_i).$$

To compute the present value of the floating rate payment, we need to compute the present value $P_i$ of a payment of the type $L(t_{i-1}, t_i)(t_i - t_{i-1})$ received at time $t_i$. This is because $L(t_{i-1}, t_i)$ is not known until time $t_{i-1}$ and we need to use risk neutral valuation to compute its present value. Following the method we discuss aboved, we have

$$P_i = E^Q\Big(\prod_{k=1}^{i} \frac{1}{1 + r_k(t_k - t_{k-1})} L(t_{i-1}, t_i)(t_i - t_{i-1})\Big).$$

It turns out that we can express $P_i$ in a more informative way using our analysis with the FRA. Recall that $P_i$ is the price such that one is indifferent between receiving $P_i$ right now or receiving $L(t_{i-1}, t_i)(t_i - t_{i-1})$ at time $t_i$. If one to enter in to a FRA in the period $[t_{i-1}, t_i]$ with the forward rate $f(0, t_{i-1}, t_i)$, one would also be indifferent between receiving a fixed payment $f(0, t_{i-1}, t_i)(t_i - t_{i-1})$ or a floating payment $L(t_{i-1}, t_i)(t_i - t_{i-1})$. We know the present value of the fixed payment $f(0, t_{i-1}, t_i)(t_i - t_{i-1})$ is

$$\begin{aligned} B(0, t_i)f(0, t_{i-1}, t_i)(t_i - t_{i-1}) &= B(0, t_i)\frac{1}{t_i - t_{i-1}}\Big(\frac{B(0, t_{i-1})}{B(0, t_i)} - 1\Big)(t_i - t_{i-1}) \\ &= B(0, t_{i-1}) - B(0, t_i). \end{aligned}$$

This implies that

$$P_i = B(0, t_{i-1}) - B(0, t_i),$$

as well. Indeed, we can derive this result using the relation (10.5).

Thus the present value of the total cash flow for the floating payment leg is just

$$\sum_{i=1}^{n} P_i = 1 - B(0, t_n).$$

Similar to the forward rate, a swap rate $S_0$ is the fixed rate $H$ such that the swap contract costs nothing to enter. In other words, it is the fixed rate such that the present value of the fixed leg payment equals the present value of the floating leg payment. Using our above results, this means

$$S_0 \sum_{i=1}^{n} (t_i - t_{i-1}) B(0, t_i) = 1 - B(0, t_n).$$

Thus the swap rate is

$$S_0 = \frac{1 - B(0, t_n)}{\sum_{i=1}^{n} (t_i - t_{i-1}) B(0, t_i)}.$$

Remark 1: A swap contract has zero value at time 0 but most definitely will not at time $t_i, i > 0$. Indeed at time $t = 1$, a swap contract entered at time 0 will be similar to a $n - 1$ paying period contract entered at time 1. The difference is the rate has already been set for the contract entered at time 0 (namely $S_0$). On the other hand, if one enters a new contract at time $t = 1$, the swap rate will be

$$S_1 = \frac{1 - B(t_1, t_n)}{\sum_{i=2}^{n} (t_i - t_{i-1}) B(t_1, t_i)}.$$

It is easy to see that unless $S_1 = S_0$, the contract entered at time 0 will not have zero value at time 1. In particular, if $S_0 > S_1$ then the contract has positive value for the receiver side of the contract and negative value for the payer side of the contract.

Remark 2: An interst rate swap does not have to be entered in at time 0. One can easily construct an interest rate swap to be commenced at time $t_i$ and to expire at time $t_j$ (where the last payment will be swapped) and the payment is made at every interval of specified length $\Delta T = t_{i+1} - t_i$. In this case we use the notation $S(t_i, t_, \delta t)$ for the swap rate of suh contract and we can easily see that

$$S(t_i, t_, \delta t) = \frac{B(0, t_i) - B(0, t_j)}{\sum_{n=i+1}^{j} B(t_i, t_n) \Delta T}.$$

### 10.10 Remark on swap in general

In general, we can view a swap as an exchange of two products in the future so that the parties involved are indifferent about such exchange at the current moment. One of the products would have value that is fixed, i.e. known at the current moment, such as the forward rate or the forward price. The other would have value that is floating, such as a risky asset or a floating LIBOR rate. The main question is how to choose this fixed price. The principle to decide it is by comparing the present values of the two legs : fixed and floating. The present values must be equal so that the two parties are indifferent about the exchange. So far we have seen three examples of swap in this course:

a. Swapping a fixed cash value $K$ with a share of a risky asset $S$ at a future time $T$. This can also be viewed as swapping a zero coupon bond with face value $K$ and maturity

$T$ with the risky asset $S$. Since the present value of $S$ is $S_0$ and the present value of a zero coupon bond with face value 1 and maturity $T$ is $B(0,T)$ we see that $K = \frac{S_0}{B(0,T)}$.

b. Swapping a fixed rate with a floating rate for the borrowing / lending period $[t_i, t_{i+1}]$. This can be viewed as swapping a zero coupon bond with maturity $t_i$, face value 1 with a zero coupon bond with maturity $t_j$ face value $1 + f(0, t_i, t_{i+1})(t_{i+1} - t_i)$.

c. Swapping a fixed rate with a floating rate for the borrowing / lending period $[t_i, t_j]$ where the payment is made every interval of length $\Delta T$. The fixed leg can be viewed as swapping a series of zero-coupon bonds with face value $1 + s(0, t_i, t_j)\Delta T$ with maturities $t_{i+1}, t_{i+2}, \cdots, t_j$. The floating leg can be viewed as swapping a series of zero-coupon bonds with face value 1 with maturities $t_i, t_{i+1}, \cdots, t_{j-1}$.

## 10.11 Duration and convexity of a bond

Observe that it is equivalent to quote either the bond price or the yield to maturity; and more often it is more important to look at the yield to maturity as an indication for the bond's "value". The yield to maturity is affected by other variables, such as the short rate (which may change by the FED's decision, for example). Thus we want to know how sensitive the bond price is with respect to a change in the yield to maturity. The concept that captures this is called the bond's duration:

$$\begin{aligned} D &= -\frac{1}{B(0,T)}\frac{dB(0,T)}{d\lambda(0,T)} \\ &= -\frac{d}{d\lambda(0,T)}\log(B(0,T)). \end{aligned}$$

The unit of the yield to maturity $\lambda(0,T)$ used in this formula is in percentage (for example yield going from 8 % per year ( $\lambda(0,T) = 0.08$ ) to 9 % per year ($\lambda(0,T) = 0.09$) ) . Thus the unit of duration $D$ is *percent change in price* (since it is $\frac{dB(0,T)}{B(0,T)}$ ) per one percentage point change in yield per year. The duration $D$ captures sensitivity of bond price to interest rate.

Remark: The definition we gave above technically is called the modified duration. There is another concept called the Macaulay duration which measures the weighted average time until cash flows are received. Let $C_i$ be a sequence of future cash flow that is received at time $t_i$, whose present value at time $t = 0$ is $P_i$. The present value of this cash flow is

$$V = \sum_i P_i.$$

The Macaulay duration is defined as

$$D_{Macaulay} = \sum_i \frac{P_i t_i}{V}.$$

It turns out that the Macaulay duration and the modified duration are pretty similar in practice. Thus we have the following important observation: the duration of a zero coupon bond is equal to its maturity (under continuous compounding) and the duration of a coupon paying bond is always less than its maturity. Generally speaking, the longer the duration of a bond, the riskier it is and the higher its price volatility is.

A related concept of duration is called the dollar duration. It is defined as

$$DV_{01} = -\frac{dB(0,T)}{10000 d\lambda(0,T)}.$$

Here $\lambda(0,T)$ is measured in the unit of basis point. One basis point equals .01 %, hence the 10000 at the denominator. Dollar duration or DV01 is the change in price in dollars, not in percentage. It gives the dollar variation in a bond's value per unit change (in basis point) in the yield.

A bond's convexity is defined as

$$C = \frac{1}{B(0,T)} \frac{d^2 dB(0,T)}{\lambda(0,T)^2}.$$

Thus one can approximate the change in the bond price with respect to the change in yield if one knows its (modified) duration and convexity as followed:

$$\frac{dB(0,T)}{B(0,T)} \approx -D_{mod} d\lambda(0,T) + \frac{1}{2} C (d\lambda(0,T))^2.$$

Given a portfolio consisting of $n$ bonds with present value $P_i, i = 1, \cdots, n$; modified durations $D_i, i = 1, \cdots, n$ and convexities $C_i, i = 1, \cdots, n$, we can calculate the duration and convexity of the portfolio as the weighted average of the individual convexity:

$$D_{portfolio} = \sum_{i=1}^{n} \frac{D_i P_i}{P},$$

where $P = \sum_{i=1}^{n} P_i$ and

$$C_{portfolio} = \sum_{i=1}^{n} \frac{C_i P_i}{P}.$$

# CHAPTER 11 Continuous time interest rate model

## 11.1 Introduction

### 11.1.1 Various rates connected with bond

Since the payment of the bond with maturity $T$ at time $T$ is fixed, one can use the bonds (with various maturities, if necessary) to “lock-in" certain interest rates. Thus, zero-coupon bond prices are used as the standard for calculating interest rates. Throughout, it is assumed we deal with the market for risk free bonds and loans. The price at time $t \leq T$ of a zero-coupon bond that pays \$1 at time $T$ shall always be denoted by $B(t,T)$. Notice that $B(T,T) = 1$. There are various interest rates associated with $B$.

(i) Continuous compounding:

In this discussion all interest rates are quoted assuming continuous compounding. Consider an account which at time $S$ has \$$L_S$ and at time $T > S$ has \$$L_T$, where $S$ and $T$ are measured in years. Then the interest rate $r$, per annum, continuously compounded, earned over $[S,T]$ is determined by the equation $L_S e^{r(T-S)} = L_T$, or

$$r = \frac{1}{T-S} \ln\left(\frac{L_T}{L_S}\right) \tag{11.1}$$

(ii) The spot rate (yield to maturity)

The *zero rate* for the period $[t,T]$, also called the *spot rate*, or, more precisely, the *continuously compounded spot rate for the period* $[t,T]$ is the function which gives the interest rate of a zero coupon bond over the interval $[t,T]$. That is, if we denote this rate by $R(t,T)$ then

$$1 = B(t,T)\exp\Big(R(t,T)(T-t)\Big).$$

From which it follows that

$$R(t,T) = \frac{1}{T-t}\ln\left(\frac{1}{B(t,T)}\right) = -\frac{\ln B(t,T)}{T-t} \tag{11.2}$$

Thus

$$B(t,T) = e^{-(T-t)R(t,T)}. \tag{11.3}$$

For a fixed $t$, a plot of $R(t,T)$ as a function of $T$ is called a spot rate curve. It gives the (continuously compounded) interest rates available for risk free zero coupon bonds for all maturities starting from $t$. The notable fact about the spot rate curve is that it is not constant—normally it tends to be upward sloping. This phenomenon is called the term structure of interest rates. $R(t,T)$ is also referred to as the yield to maturity of the zero coupon bond at time $t$ for the *time to maturity* $\tau = T - t$.

(iii) The forward rate

Consider times $t < S < T$. Suppose at time $t$ we would want to lock in certain spot rate for the time interval $[S,T]$. Let's call this rate $F(t,S,T)$. Then clearly this rate must be related with the bond price $B(t,S)$ and $B(t,T)$. So we should determine what $F(t,S,T)$ is and further inquire into whether we can indeed lock in this rate at time $t$ by trading certain shares of the bonds with maturities at $S$ and $T$.

To answer the first question, clearly what we want is if we invest 1 dollar at time $S$ then we should receive $\exp\Big(F(t,S,T)(T-S\Big)$ at time $T$. Note that, 1 dollar at time $S$ is equivalent to $B(t,S)$ at time $t$ and $\exp\Big(F(t,S,T)(T-S)\Big)$ dollars at time $T$ is equivalent to $B(t,T)\exp\Big(F(t,S,T)(T-S)\Big)$ at time $t$. These clearly should be equal if we want to lock in the rate $F(t,S,T)$ at time $t$. Therefore

$$B(t,S) = B(t,T)\exp\Big(F(t,S,T)(T-S)\Big),$$

or

$$F(t;S,T) = \frac{1}{T-S}\ln\left(\frac{B(t,S)}{B(t,T)}\right) = -\frac{\ln B(t,T) - \ln B(t,S)}{T-S}.$$

To answer the second question, first plug

$$F(t,S,T) = -\frac{\ln B(t,T) - \ln B(t,S)}{T-S}$$

into $B(t,T)\exp\Big(F(t,S,T)(T-S)\Big)$ and observe that

$$B(t,T)\exp\Big(F(t,S,T)(T-S)\Big) = B(t,T)\frac{B(t,S)}{B(t,T)}.$$

This suggests that we should hold $\frac{B(t,S)}{B(t,T)}$ shares of bond with maturity $T$ at time $t$. To finance this position, we should sell 1 share of bond with maturity $S$ at time $t$ (since we expect to invest 1 dollar at time $S$). This turns out to be the right scheme because this costs nothing at time $t$: If at $t$ we sell one zero-coupon bond maturing at $S$ for $B(t,S)$ and with this money buy $B(t,S)/B(t,T)$ zero-coupon bonds maturing at $T$, the net value of this transaction for us is $0$. At time $S$ we pay out a dollar and at time $T$ receive $B(t,S)/B(t,T)$. This is indeed equivalent to earning, at $T$, the amount $B(t,S)/B(t,T) = \exp\Big(F(t,S,T)(T-S)\Big)$ from a deposit of \$ 1 at $S$. The rate of interest earned by this transaction, $F(t,S,T)$ is called the *forward rate for* $[S,T]$ *contracted at* $t$.

Remark: $F(t,S,T)$ is known at time $t$ by observing $B(t,S)$ and $B(t,T)$, that is $F(t,S,T) \in \mathcal{F}_t$, where $\mathcal{F}_t$ is the filtration generated by $B(t,S)$ and $B(t,T)$.

(iv) The instantaneous forward rate

The forward rate $F(t,S,T)$ has the formula

$$F(t;S,T) = -\frac{\ln B(t,T) - \ln B(t,S)}{T-S}.$$

If we let $T$ goes to $S$, then the right hand side should go to $-\frac{\partial}{\partial T}\ln[B(t,S)]$, if the derivative exists. Indeed, if we assume $B(t,T)$ is differentible in $T$, then this is the case. This is not an unreasonable assumption since for a fixed $t$, one can believe that the bond price is a smooth function of different maturities. (On the other hand, for a fix maturity $T$, the bond price should *not* be a smooth function of $t$. It should be very irregular, indeed, in $t$, similar to behavior of the graph of a Brownian motion in $t$).

So we define the *instantaneous forward rate at* $t$ *for investing at time* $T$ as

$$f(t,T) = -\frac{\partial}{\partial T}\ln[B(t,T)]. \tag{11.4}$$

By integrating in $T$, it follows that

$$B(t,T) = \exp\{-\int_t^T f(t,u)\,du\} \tag{11.5}$$

For brevity, we refer to $f(t,T)$ as the forward rate function.

Remark: Again, note that here $f(t,T)$ is known at time $t$, that is $f(t,T) \in \mathcal{F}_t$ where $\mathcal{F}_t$ is the filtration generated by $B(t,T)$.

(v) The short rate

The *short rate* is the rate available at time $t$ for the shortest period loans. Formally it is defined as

$$R(t) = f(t,t) \tag{11.6}$$

*Remark:* It seems reasonable to define $R(t) = f(t,t)$ (just from the understanding of what the forward rate is). First, it is reasonable to believe the spot rate $R(t,T)$ should converge to the short rate $R(t)$ when $T \to t$. Recall

$$R(t,T) = -\frac{\ln B(t,T)}{T-t} = \frac{\int_t^T f(t,u)du}{T-t}.$$

The RHS converges to $f(t,t)$ (by Lesbegue differentation theorem) as $T \to t$. So if we expect $R(t,T)$ to converge to $R(t)$, then it is reasonable to set $R(t) = f(t,t)$.

Second, from comparing the risk neutral pricing formula

$$B(t,T) = \tilde{E}\big(e^{-\int_t^T R(u)du}\big)$$

with the definition of the forward rate:

$$B(t,T) = e^{-\int_t^T f(t,u)du},$$

we should expect $R(t) = f(t,t)$ as well. Indeed, suppose $R(t) > f(t,t)$. Then if we suppose $R(u)$ and $f(t,u)$ are continuous functions of $u$ (which is reasonable) then there must exist some $T > t$ so that $R(u) > f(t,u)$ for $u \in [t,T]$. But then we have

$$B(t,T) = e^{-\int_t^T f(t,u)du} > \tilde{E}\Big(e^{-\int_t^T R(u)du}\big|\mathcal{F}(t)\Big) = B(t,T),$$

which is a contradiction. So this cannot happen.

## 11.2 One factor Hull-White model

### 11.2.1 Explicit formula for $R(t)$

The one-factor Hull-White model is defined by a short rate which solves a linear differential equation

$$dR(t) = k(\mu - R(t))\,dt + \sigma\,d\tilde{W}(t),$$

where $\tilde{W}$ is a Brownian motion under the risk-neutral measure. To find the explicit solution for $R(t)$, note that

$$d(e^{kt}R(t)) = e^{kt}(kR(t)dt + dR(t)).$$

Thus we get

$$e^{kt}R(t) = R(0) + \int_0^t k\mu e^{ks}ds + \int_0^t \sigma e^{ks}d\tilde{W}_s.$$

And

$$R(t) = e^{-kt}R(0) + \mu(1 - e^{-kt}) + e^{-kt}\int_0^t \sigma e^{ks}d\tilde{W}_s.$$

From this formula we can see that as $t$ approaches infinity, $R(t)$ is centered around $\mu$ with variance $\frac{\sigma^2}{2k}$. The formula also explains why we refer to this model as mean-reverting and $k$ as the rate of mean reversion.

### 11.2.2 Explicit formula for $B(t,T)$

From the risk neutral pricing formula

$$B(t,T) = \tilde{E}[e^{-\int_t^T R(u)du}|\mathcal{F}_t].$$

Note that for $u \geq t$

$$R(u) = e^{-k(u-t)}R(t) + \mu(1 - e^{-k(u-t)}) + e^{-k(u-t)} \int_t^u \sigma e^{k(s-t)} d\tilde{W}_s.$$

We need to integrate

$$\begin{aligned}\int_t^T e^{-k(u-t)} \int_t^u \sigma e^{k(s-t)} d\tilde{W}_s du &= \int_t^T \int_t^u e^{-k(u-t)} \sigma e^{k(s-t)} d\tilde{W}_s du \\ &= \sigma \int_t^T \int_t^u e^{-k(u-s)} d\tilde{W}_s du\end{aligned}$$

Accepting the fact that change of order of integration applies to this double integral of $d\tilde{W}_s$ and $du$ we have

$$\sigma \int_t^T \int_t^u e^{-k(u-s)} d\tilde{W}_s du = \sigma \int_t^T \int_s^T e^{-k(u-s)} du d\tilde{W}_s.$$

The above stochastic integral is independent of $\mathcal{F}_t$ and has normal distribution with mean 0 and variance

$$\begin{aligned}\sigma^2(T-t) &= \sigma^2 \int_t^T (\int_s^T e^{-k(u-s)} du)^2 ds \\ &= \frac{\sigma^2}{k^2}\Big[T - t + \frac{2}{k}(1 - e^{-k(T-t)}) + \frac{1}{2k}(1 - e^{-2k(T-t)})\Big].\end{aligned}$$

Thus the conditional distribution of the integral $\int_t^T R_u du$ on $\mathcal{F}_t$ has normal distribution with mean

$$\mu(T-t) = -\Big[R(t)\frac{1}{k}(1 - e^{-k(T-t)}) + \mu(T-t) - \frac{\mu}{k}(1 - e^{-k(T-t)})\Big]$$

and variance $A(T-t)$.

Thus by the Independence Lemma and the moment genrating function of a Normal random variable we have

$$B(t,T) = e^{\mu(T-t) + \frac{1}{2}\sigma^2(T-t)}.$$

### 11.2.3 Model calibration

By model calibration we mean making the choices for $\mu, k, \sigma$ in the Hull-White model so that the resulting bond prices $B(t,T)$ most closely match the market data.

It should be remarked that we possibly do not observe the short rate $R(t)$ directly in the market; if by $R(t)$ we mean the process such that

$$B(t,T) = \tilde{E}(e^{-\int_t^T R(u)du}|\mathcal{F}_t).$$

There are some proxys for the risk free short rate, such as the OIS (overnight interest rate swap) but strictly speaking this is not the rate $R(t)$ we have in mind when plugging in the pricing formula for the treasury bond $B(t,T)$. This is why even though we have the dynamics of $R(t)$ directly dependent on $\mu, k, \sigma$ we cannot use that dynamics to calibrate these parameters. The yield to maturity $R(t,T)$ on the other hand, is readily observable as a function of the bond price $B(t,T)$.

Observe that the previous formula for $B(t,T)$ can be written as

$$B(t,T) = e^{-C(T-t)R(t)-D(T-t)}$$

where

$$\begin{aligned} C(T-t) &= \frac{1}{k}(1-e^{-k(T-t)}) \\ D(T-t) &= \mu(T-t) - \frac{\mu}{k}(1-e^{-k(T-t)}) - \frac{1}{2}A(T-t). \end{aligned}$$

Recall that in terms of the yield to maturity $R(t,T)$

$$B(t,T) = e^{-R(t,T)(T-t)}.$$

Thus we conclude that

$$R(t,T) = \frac{C(T-t)}{T-t}R(t) + \frac{D(T-t)}{T-t}.$$

Now we start to look at $R(t,T)$ as dependent instead on $t$ and the time to maturity $\tau = T - t$ and we keep $\tau$ *fixed* for varying $t$. This is subtle conceptually because for $s \neq t, R(s,\tau)$ and $R(t,\tau)$ refers to the yields to maturity of *different* bonds with the same time to maturity $\tau$. Nevertheless these quantities are readily observable in the market and we have

$$dR(t,\tau) = \frac{C(\tau)}{\tau}dR(t) + \frac{D(\tau)}{\tau}$$

Thus

$$dR(t,\tau) = k(\tau)(\mu(\tau) - R(t,\tau))dt + \sigma(\tau)d\tilde{W}_t,$$

where the coefficients $k(\tau), \mu(\tau), \sigma(\tau)$ depend only on $\tau, k, \mu, \sigma$. Observe that $R(t, \tau)$ is again a mean-reverting process like $R(t)$. Since we can observe $R(t, \tau)$ at different $t$, there are readily availble methods to estimate the coefficients $k(\tau), \mu(\tau), \sigma(\tau)$ (see e.g. calibration for Ornstein-Uhlenbeck process using linear regression technique) from which we can get estimates of $k, \mu, \sigma$.

Remark: The dynamics for $R(t, \tau)$ given above holds for any $\tau$. In practice for each time to maturity $\tau$ we have a different set of data $R(t, \tau), 0 \leq t \leq T$ for example. Each set of data potentially gives rise to a different estimate of $k, \mu, \sigma$. Thus we may have the issue of having too much data for too few parameters. Or put it in another way: the Hull-White model may not be flexible enough to fit the market term structure. A solution for this is of course to increase the number of parameters, by going to the multi-factor setting. However, it's good to keep in mind that since the bond market is huge, with many different time to maturity it probably is not easy to come up with a model that captures the term structure perfectly across the spectrum, from the short end to the long end.

### 11.2.4 PDE representation for Hull-White Model

The Hull-White model is part of a affine-yield family of models. That is, the short rate is modelled in such a way that the zero-coupon bond price $B(t, T)$ can be expressed as

$$B(t, T) = e^{-A(t,T)R_t - B(t,T)}.$$

We demonstrate how to find the equations that $A(t, T), B(t, T)$ satisfy. This gives an alternative approach to finding the bond price when explicit computation of $r_t$ is not possible (see below for the case of CIR model).

First, we accept as a fact that the bond can be expressed as a function $u(t, R_t)$ of time and the short rate. The PDE that $u(t, x)$ satisfies is

$$\begin{aligned} -xu + u_t + u_x k(\mu - x) + \frac{1}{2} u_{xx} \sigma^2 &= 0 \\ u(T, x) &= 1. \end{aligned}$$

This is derived in a similar manner to the Black-Scholes PDE, via the martingale approach. That is we observe

$$e^{-\int_0^t R_s ds} B(t, T) = e^{-\int_0^t R_s ds} u(t, R_t)$$

is a martingale. Applying the Ito's formula to $X_t = e^{-\int_0^t R_s ds} u(t, R_t)$ we have

$$dX_t = e^{-\int_0^t R_s ds} \Big\{ (-R_t u + u_t + u_x k(\mu - R_t) + \frac{1}{2} u_{xx} \sigma^2) dt + u_x \sigma d\tilde{W}_t \Big\}.$$

Since $X_t$ is a martingale, its "dt" part must be 0 and we obtain the PDE. On the other hand, the affine yield representation suggests that

$$u(t, x) = e^{-A(t,T)x - B(t,T)}.$$

Plug this into the PDE we obtain

$$u(-x-(A_t x+B_t)-Ak(\mu-x)+\frac{1}{2}A^2\sigma^2)=0.$$

Since $u\neq 0$ we have

$$x+(A_t x+B_t)+Ak(\mu-x)-\frac{1}{2}A^2\sigma^2=0.$$

Factoring the terms involving $x$ we have

$$(-kA+A_t+1)x+k\mu A+\frac{1}{2}A^2\sigma^2+B_t=0.$$

Note that since $B(T,T)=1$, we necessarily have $A(T,T)=B(T,T)=0$. Thus the approach is first to solve the ODE with terminal condition

$$\begin{aligned}-kA+A_t+1 &= 0\\ A(T,T) &= 0.\end{aligned}$$

Once we obtain $A$ we can plug that in and solve for the ODE for $B$:

$$\begin{aligned}k\mu A+\frac{1}{2}A^2\sigma^2+B_t &= 0\\ B(T,T) &= 0.\end{aligned}$$

### 11.3 One factor CIR model

Suppose the short rate follows the dynamics

$$dR_t=k(\mu-R_t)dt+\sigma\sqrt{R_t}d\tilde{W}_t.$$

This is a mean reveversion equation but it does not have an explicit solution. By Markov property, there exists a function $u(t,R_t)$ such that

$$u(t,R_t)=B(t,T)=\tilde{E}(e^{-\int_t^T R_u du}|\mathcal{F}_t).$$

The PDE that $u(t,x)$ satisfies is

$$\begin{aligned}-xu+u_t+u_x k(\mu-x)+\frac{1}{2}u_{xx}\sigma^2 x &= 0\\ u(T,x)=1.\end{aligned}$$

On the other hand, to be affine-yield, $u(t,x)$ has the representation

$$u(t,x)=e^{-A(t,T)x-B(t,T)}.$$

Plug this form into the PDE, factoring the terms that involves $x$ gives an ODE system in $t$

$$\begin{aligned} B_t + k\mu A &= 0 \\ A_t - kA - \frac{1}{2}\sigma^2 A^2 + 1 = 0. \end{aligned}$$

The terminal conditions are $A(T,T) = B(T,T) = 0$. This is as far as we can go to describe the form of the explicit solution for $B(t,T)$. It is not our interest here to find the explicit solution to the above system of ODEs. The interested readers can find the solution in Shreve section 6.6.

## References


[1] Hurd, T. (2010) *Structural Modes of Credit Risk* - *Course Notes*

[2] Shreve, S. (2012). *Stochastic calculus for finance I: the binomial asset pricing model*. Springer Science & Business Media.

[3] Stampfli, J., & Goodman, V. (2001). *The mathematics of finance: modeling and hedging.*