The Basics of Financial Mathematics

Spring 2003

Richard F. Bass
Department of Mathematics
University of Connecticut

 If you find any errors, I would appreciate hearing from you: **bass@math.uconn.edu**

1. Introduction.

In this course we will study mathematical finance. Mathematical finance is **not** about predicting the price of a stock. What it is about is figuring out the price of options and derivatives.

The most familiar type of option is the option to buy a stock at a given price at a given time. For example, suppose Microsoft is currently selling today at $40 per share. A European call option is something I can buy that gives me the right to buy a share of Microsoft at some future date. To make up an example, suppose I have an option that allows me to buy a share of Microsoft for $50 in three months time, but does not compel me to do so. If Microsoft happens to be selling at $45 in three months time, the option is worthless. I would be silly to buy a share for $50 when I could call my broker and buy it for $45. So I would choose not to exercise the option. On the other hand, if Microsoft is selling for $60 three months from now, the option would be quite valuable. I could exercise the option and buy a share for $50. I could then turn around and sell the share on the open market for $60 and make a profit of $10 per share. Therefore this stock option I possess has some value. There is some chance it is worthless and some chance that it will lead me to a profit. The basic question is: how much is the option worth today?

The huge impetus in financial derivatives was the seminal paper of Black and Scholes in 1973. Although many researchers had studied this question, Black and Scholes gave a definitive answer, and a great deal of research has been done since. These are not just academic questions; today the market in financial derivatives is **larger** than the market in stock securities. In other words, more money is invested in options on stocks than in stocks themselves.

Options have been around for a long time. The earliest ones were used by manufacturers and food producers to hedge their risk. A farmer might agree to sell a bushel of wheat at a fixed price six months from now rather than take a chance on the vagaries of market prices. Similarly a steel refinery might want to lock in the price of iron ore at a fixed price.

The sections of these notes can be grouped into five categories. The first is elementary probability. Although someone who has had a course in undergraduate probability will be familiar with some of this, we will talk about a number of topics that are not usually covered in such a course: σ-fields, conditional expectations, martingales. The second category is the binomial asset pricing model. This is just about the simplest model of a stock that one can imagine, and this will provide a case where we can see most of the major ideas of mathematical finance, but in a very simple setting. Then we will turn to advanced probability, that is, ideas such as Brownian motion, stochastic integrals, stochastic differential equations, Girsanov transformation. Although to do this rigorously requires measure theory, we can still learn enough to understand and work with these concepts. We then return to finance and work with the continuous model. We will derive the Black-Scholes formula, see the Fundamental Theorem of Asset Pricing, work with equivalent martingale measures, and the like. The fifth main category is term structure models, which means models of interest rate behavior.

I found some unpublished notes of Steve Shreve extremely useful in preparing these notes. I hope that he has turned them into a book and that this book is now available. The stochastic calculus part of these notes is from my own book: *Probabilistic Techniques in Analysis*, Springer, New York, 1995.

2. Review of elementary probability.

Let's begin by recalling some of the definitions and basic concepts of elementary probability. We will only work with discrete models at first.

We start with an arbitrary set, called the probability space, which we will denote by Ω, the capital Greek letter "omega." We are given a class $\mathcal{F}$ of subsets of Ω. These are called events. We require $\mathcal{F}$ to be a σ-field.

Definition 2.1. *A collection $\mathcal{F}$ of subsets of Ω is called a σ-field if*

(1) $\emptyset \in \mathcal{F}$,
(2) $\Omega \in \mathcal{F}$,
(3) $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$, *and*
(4) $A_1, A_2, \ldots \in \mathcal{F}$ *implies both* $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$ *and* $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Here $A^c = \{\omega \in \Omega : \omega \notin A\}$ denotes the complement of A. $\emptyset$ denotes the empty set, that is, the set with no elements. We will use without special comment the usual notations of $\cup$ (union), $\cap$ (intersection), $\subset$ (contained in), $\in$ (is an element of).

Typically, in an elementary probability course, $\mathcal{F}$ will consist of all subsets of Ω, but we will later need to distinguish between various σ-fields. Here is an example. Suppose one tosses a coin two times and lets Ω denote all possible outcomes. So $\Omega = \{HH, HT, TH, TT\}$. A typical σ-field $\mathcal{F}$ would be the collection of all subsets of Ω. In this case it is trivial to show that $\mathcal{F}$ is a σ-field, since every subset is in $\mathcal{F}$. But if we let $\mathcal{G} = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}\}$, then $\mathcal{G}$ is also a σ-field. One has to check the definition, but to illustrate, the event $\{HH, HT\}$ is in $\mathcal{G}$, so we require the complement of that set to be in $\mathcal{G}$ as well. But the complement is $\{TH, TT\}$ and that event is indeed in $\mathcal{G}$.

One point of view which we will explore much more fully later on is that the σ-field tells you what events you "know." In this example, $\mathcal{F}$ is the σ-field where you "know" everything, while $\mathcal{G}$ is the σ-field where you "know" only the result of the first toss but not the second. We won't try to be precise here, but to try to add to the intuition, suppose one knows whether an event in $\mathcal{F}$ has happened or not for a particular outcome. We would then know which of the events $\{HH\}, \{HT\}, \{TH\}$, or $\{TT\}$ has happened and so would know what the two tosses of the coin showed. On the other hand, if we know which events in $\mathcal{G}$ happened, we would only know whether the event $\{HH, HT\}$ happened, which means we would know that the first toss was a heads, or we would know whether the event $\{TH, TT\}$ happened, in which case we would know that the first toss was a tails. But there is no way to tell what happened on the second toss from knowing which events in $\mathcal{G}$ happened. Much more on this later.

The third basic ingredient is a probability.

Definition 2.2. *A function $\mathbb{P}$ on $\mathcal{F}$ is a probability if it satisfies*

(1) if $A \in \mathcal{F}$, then $0 \le \mathbb{P}(A) \le 1$,
(2) $\mathbb{P}(\Omega) = 1$, and
(3) $\mathbb{P}(\emptyset) = 0$, and
(4) if $A_1, A_2, \ldots \in \mathcal{F}$ are pairwise disjoint, then $\mathbb{P}(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mathbb{P}(A_i)$.

A collection of sets A_i is pairwise disjoint if $A_i \cap A_j = \emptyset$ unless $i = j$.

There are a number of conclusions one can draw from this definition. As one example, if $A \subset B$, then $\mathbb{P}(A) \le \mathbb{P}(B)$ and $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. See Note 1 at the end of this section for a proof.

Someone who has had measure theory will realize that a σ-field is the same thing as a σ-algebra and a probability is a measure of total mass one.

A random variable (abbreviated r.v.) is a function X from Ω to $\mathbb{R}$, the reals. To be more precise, to be a r.v. X must also be measurable, which means that $\{\omega : X(\omega) \ge a\} \in \mathcal{F}$ for all reals a.

The notion of measurability has a simple definition but is a bit subtle. If we take the point of view that we know all the events in $\mathcal{G}$, then if Y is $\mathcal{G}$-measurable, then we know Y. Phrased another way, suppose we know whether or not the event has occurred for each event in $\mathcal{G}$. Then if Y is $\mathcal{G}$-measurable, we can compute the value of Y.

Here is an example. In the example above where we tossed a coin two times, let X be the number of heads in the two tosses. Then X is $\mathcal{F}$ measurable but not $\mathcal{G}$ measurable. To see this, let us consider $A_a = \{\omega \in \Omega : X(\omega) \ge a\}$. This event will equal

$$\begin{cases} \Omega & \text{if } a \le 0; \\ \{HH, HT, TH\} & \text{if } 0 < a \le 1; \\ \{HH\} & \text{if } 1 < a \le 2; \\ \emptyset & \text{if } 2 < a. \end{cases}$$

For example, if $a = \frac{3}{2}$, then the event where the number of heads is $\frac{3}{2}$ or greater is the event where we had two heads, namely, $\{HH\}$. Now observe that for each a the event A_a

is in $\mathcal{F}$ because $\mathcal{F}$ contains all subsets of Ω. Therefore X is measurable with respect to $\mathcal{F}$. However it is not true that A_a is in $\mathcal{G}$ for every value of a – take $a = \frac{3}{2}$ as just one example – the subset $\{HH\}$ is not in $\mathcal{G}$. So X is not measurable with respect to the σ-field $\mathcal{G}$.

A discrete r.v. is one where $\mathbb{P}(\omega : X(\omega) = a) = 0$ for all but countably many a's. In defining sets one usually omits the ω; thus $(X = x)$ means the same as $\{\omega : X(\omega) = x\}$.

In the discrete case, to check measurability with respect to a σ-field $\mathcal{F}$, it is enough that $(X = a) \in \mathcal{F}$ for all reals a. The reason for this is that if $x_1, x_2, \ldots$ are the values of x for which $\mathbb{P}(X = x) \neq 0$, then we can write $(X \geq a) = \cup_{x_i \geq a}(X = x_i)$ and we have a countable union. So if $(X = x_i) \in \mathcal{F}$, then $(X \geq a) \in \mathcal{F}$.

Given a discrete r.v. X, the expectation or mean is defined by

$$\mathbb{E}\, X = \sum_x x\mathbb{P}(X = x)$$

provided the sum converges. If X only takes finitely many values, then this is a finite sum and of course it will converge. This is the situation that we will consider for quite some time. However, if X can take an infinite number of values (but countable), convergence needs to be checked. For example, if $\mathbb{P}(X = 2^n) = 2^{-n}$ for $n = 1, 2, \ldots$, then $\mathbb{E}\, X = \sum_{n=1}^\infty 2^n \cdot 2^{-n} = \infty$.

There is an alternate definition of expectation which is equivalent in the discrete setting. Set

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

To see that this is the same, look at Note 2 at the end of the section. The advantage of the second definition is that some properties of expectation, such as $\mathbb{E}\,(X + Y) = \mathbb{E}\, X + \mathbb{E}\, Y$, are immediate, while with the first definition they require quite a bit of proof.

We say two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Two random variables X and Y are independent if $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(X \in B)$ for all A and B that are subsets of the reals. The comma in the expression $\mathbb{P}(X \in A, Y \in B)$ means "and." Thus

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

The extension of the definition of independence to the case of more than two events or random variables is what one would expect: $A_1, \ldots, A_n$ are independent if

$$\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i).$$

.

A common misconception is that an event is independent of itself. If A is an event that is independent of itself, then

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A) = (\mathbb{P}(A))^2.$$

The only finite solutions to the equation $x = x^2$ are $x = 0$ and $x = 1$, so an event is independent of itself only if it has probability 0 or 1.

Two σ-fields $\mathcal{F}$ and $\mathcal{G}$ are independent if A and B are independent whenever $A \in \mathcal{F}$ and $B \in \mathcal{G}$. A r.v. X and a σ-field $\mathcal{G}$ are independent if $\mathbb{P}((X \in A) \cap B) = \mathbb{P}(X \in A)\mathbb{P}(B)$ whenever A is a subset of the reals and $B \in \mathcal{G}$.

As an example, suppose we toss a coin two times and we define the σ-fields $\mathcal{G}_1 = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}\}$ and $\mathcal{G}_2 = \{\emptyset, \Omega, \{HH, TH\}, \{HT, TT\}\}$. Then $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent if $\mathbb{P}(HH) = \mathbb{P}(HT) = \mathbb{P}(TH) = \mathbb{P}(TT) = \frac{1}{4}$. (Here we are writing $\mathbb{P}(HH)$ when a more accurate way would be to write $\mathbb{P}(\{HH\})$.) An easy way to understand this is that if we look at an event in $\mathcal{G}_1$ that is not $\emptyset$ or Ω, then that is the event that the first toss is a heads or it is the event that the first toss is a tails. Similarly, a set other than $\emptyset$ or Ω in $\mathcal{G}_2$ will be the event that the second toss is a heads or that the second toss is a tails.

If two r.v.s X and Y are independent, we have the multiplication theorem, which says that $\mathbb{E}\,(XY) = (\mathbb{E}\,X)(\mathbb{E}\,Y)$ provided all the expectations are finite. See Note 3 for a proof.

Suppose $X_1, \ldots, X_n$ are n independent r.v.s, such that for each one $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = 0) = 1 - p$, where $p \in [0, 1]$. The random variable $S_n = \sum_{i=1}^n X_i$ is called a binomial r.v., and represents, for example, the number of successes in n trials, where the probability of a success is p. An important result in probability is that

$$\mathbb{P}(S_n = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

The variance of a random variable is

$$\operatorname{Var} X = \mathbb{E}\,[(X - \mathbb{E}\,X)^2].$$

This is also equal to

$$\mathbb{E}\,[X^2] - (\mathbb{E}\,X)^2.$$

It is an easy consequence of the multiplication theorem that if X and Y are independent,

$$\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y.$$

The expression $\mathbb{E}\,[X^2]$ is sometimes called the second moment of X.

We close this section with a definition of conditional probability. The probability of A given B, written $\mathbb{P}(A \mid B)$ is defined by

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

provided $\mathbb{P}(B) \neq 0$. The conditional expectation of X given B is defined to be

$$\frac{\mathbb{E}\,[X; B]}{\mathbb{P}(B)},$$

provided $\mathbb{P}(B) \neq 0$. The notation $\mathbb{E}\,[X; B]$ means $\mathbb{E}\,[X 1_B]$, where $1_B(\omega)$ is 1 if $\omega \in B$ and 0 otherwise. Another way of writing $\mathbb{E}\,[X; B]$ is

$$\mathbb{E}\,[X; B] = \sum_{\omega \in B} X(\omega)\mathbb{P}(\{\omega\}).$$

(We will use the notation $\mathbb{E}\,[X; B]$ frequently.)

Note 1. Suppose we have two disjoint sets C and D. Let $A_1 = C$, $A_2 = D$, and $A_i = \emptyset$ for $i \geq 3$. Then the A_i are pairwise disjoint and

$$\mathbb{P}(C \cup D) = \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \mathbb{P}(C) + \mathbb{P}(D) \tag{2.1}$$

by Definition 2.2(3) and (4). Therefore Definition 2.2(4) holds when there are only two sets instead of infinitely many, and a similar argument shows the same is true when there are an arbitrary (but finite) number of sets.

Now suppose $A \subset B$. Let $C = A$ and $D = B - A$, where $B - A$ is defined to be $B \cap A^c$ (this is frequently written $B \setminus A$ as well). Then C and D are disjoint, and by (2.1)

$$\mathbb{P}(B) = \mathbb{P}(C \cup D) = \mathbb{P}(C) + \mathbb{P}(D) \geq \mathbb{P}(C) = \mathbb{P}(A).$$

The other equality we mentioned is proved by letting $C = A$ and $D = A^c$. Then C and D are disjoint, and

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(C \cup D) = \mathbb{P}(C) + \mathbb{P}(D) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

Solving for $\mathbb{P}(A^c)$, we have

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

Note 2. Let us show the two definitions of expectation are the same (in the discrete case). Starting with the first definition we have

$$\begin{aligned}
\mathbb{E}\,X &= \sum_x x\mathbb{P}(X=x) \\
&= \sum_x x \sum_{\{\omega\in\Omega: X(\omega)=x\}} \mathbb{P}(\{\omega\}) \\
&= \sum_x \sum_{\{\omega\in\Omega: X(\omega)=x\}} X(\omega)\mathbb{P}(\{\omega\}) \\
&= \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}),
\end{aligned}$$

and we end up with the second definition.

Note 3. Suppose X can takes the values $x_1, x_2, \ldots$ and Y can take the values $y_1, y_2, \ldots$. Let $A_i = \{\omega : X(\omega) = x_i\}$ and $B_j = \{\omega : Y(\omega) = y_j\}$. Then

$$X = \sum_i x_i 1_{A_i}, \qquad Y = \sum_j y_j 1_{B_j},$$

and so

$$XY = \sum_i \sum_j x_i y_i 1_{A_i} 1_{B_j}.$$

Since $1_{A_i} 1_{B_j} = 1_{A_i \cap B_j}$, it follows that

$$\mathbb{E}\,[XY] = \sum_i \sum_j x_i y_j \mathbb{P}(A_i \cap B_j),$$

assuming the double sum converges. Since X and Y are independent, $A_i = (X = x_i)$ is independent of $B_j = (Y = y_j)$ and so

$$\begin{aligned}
\mathbb{E}\,[XY] &= \sum_i \sum_j x_i y_j \mathbb{P}(A_i)\mathbb{P}(B_j) \\
&= \sum_i x_i \mathbb{P}(A_i) \Big[\sum_j y_j \mathbb{P}(B_j)\Big] \\
&= \sum_i x_i \mathbb{P}(A_i) \mathbb{E}\,Y \\
&= (\mathbb{E}\,X)(\mathbb{E}\,Y).
\end{aligned}$$

3. Conditional expectation.

Suppose we have 200 men and 100 women, 70 of the men are smokers, and 50 of the women are smokers. If a person is chosen at random, then the conditional probability that the person is a smoker given that it is a man is 70 divided by 200, or 35%, while the conditional probability the person is a smoker given that it is a women is 50 divided by 100, or 50%. We will want to be able to encompass both facts in a single entity.

The way to do that is to make conditional probability a random variable rather than a number. To reiterate, we will make conditional probabilities random. Let M, W be man, woman, respectively, and S, S^c smoker and nonsmoker, respectively. We have

$$\mathbb{P}(S \mid M) = .35, \qquad \mathbb{P}(S \mid W) = .50.$$

We introduce the random variable

$$(.35)1_M + (.50)1_W$$

and use that for our conditional probability. So on the set M its value is .35 and on the set W its value is .50.

We need to give this random variable a name, so what we do is let $\mathcal{G}$ be the σ-field consisting of $\{\emptyset, \Omega, M, W\}$ and denote this random variable $\mathbb{P}(S \mid \mathcal{G})$. Thus we are going to talk about the conditional probability of an event given a σ-field.

What is the precise definition?

Definition 3.1. *Suppose there exist finitely (or countably) many sets $B_1, B_2, \ldots$, all having positive probability, such that they are pairwise disjoint, Ω is equal to their union, and $\mathcal{G}$ is the σ-field one obtains by taking all finite or countable unions of the B_i. Then the conditional probability of A given $\mathcal{G}$ is*

$$\mathbb{P}(A \mid \mathcal{G}) = \sum_i \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} 1_{B_i}(\omega).$$

In short, on the set B_i the conditional probability is equal to $\mathbb{P}(A \mid B_i)$.

Not every σ-field can be so represented, so this definition will need to be extended when we get to continuous models. σ-fields that can be represented as in Definition 3.B are called finitely (or countably) generated and are said to be generated by the sets $B_1, B_2, \ldots$.

Let's look at another example. Suppose Ω consists of the possible results when we toss a coin three times: HHH, HHT, etc. Let $\mathcal{F}_3$ denote all subsets of Ω. Let $\mathcal{F}_1$ consist of the sets $\emptyset, \Omega, \{HHH, HHT, HTH, HTT\}$, and $\{THH, THT, TTH, TTT\}$. So $\mathcal{F}_1$ consists of those events that can be determined by knowing the result of the first toss. We want to let $\mathcal{F}_2$ denote those events that can be determined by knowing the first two tosses. This will include the sets $\emptyset, \Omega, \{HHH, HHT\}, \{HTH, HTT\}, \{THH, THT\}, \{TTH, TTT\}$. This is not enough to make $\mathcal{F}_2$ a σ-field, so we add to $\mathcal{F}_2$ all sets that can be obtained by taking unions of these sets.

Suppose we tossed the coin independently and suppose that it was fair. Let us calculate $\mathbb{P}(A \mid \mathcal{F}_1), \mathbb{P}(A \mid \mathcal{F}_2)$, and $\mathbb{P}(A \mid \mathcal{F}_3)$ when A is the event $\{HHH\}$. First the conditional probability given $\mathcal{F}_1$. Let $C_1 = \{HHH, HHT, HTH, HTT\}$ and $C_2 = \{THH, THT, TTH, TTT\}$. On the set C_1 the conditional probability is $\mathbb{P}(A \cap C_1)/\mathbb{P}(C_1) = \mathbb{P}(HHH)/\mathbb{P}(C_1) = \frac{1}{8}/\frac{1}{2} = \frac{1}{4}$. On the set C_2 the conditional probability is $\mathbb{P}(A \cap C_2)/\mathbb{P}(C_2) = \mathbb{P}(\emptyset)/\mathbb{P}(C_2) = 0$. Therefore $\mathbb{P}(A \mid \mathcal{F}_1) = (.25)1_{C_1}$. This is plausible – the probability of getting three heads given the first toss is $\frac{1}{4}$ if the first toss is a heads and 0 otherwise.

Next let us calculate $\mathbb{P}(A \mid \mathcal{F}_2)$. Let $D_1 = \{HHH, HHT\}, D_2 = \{HTH, HTT\}, D_3 = \{THH, THT\}, D_4 = \{TTH, TTT\}$. So $\mathcal{F}_2$ is the σ-field consisting of all possible unions of some of the D_i's. $\mathbb{P}(A \mid D_1) = \mathbb{P}(HHH)/\mathbb{P}(D_1) = \frac{1}{8}/\frac{1}{4} = \frac{1}{2}$. Also, as above, $\mathbb{P}(A \mid D_i) = 0$ for $i = 2, 3, 4$. So $\mathbb{P}(A \mid \mathcal{F}_2) = (.50)1_{D_1}$. This is again plausible – the probability of getting three heads given the first two tosses is $\frac{1}{2}$ if the first two tosses were heads and 0 otherwise.

What about conditional expectation? Recall $\mathbb{E}[X; B_i] = \mathbb{E}[X1_{B_i}]$ and also that $\mathbb{E}[1_B] = 1 \cdot \mathbb{P}(1_B = 1) + 0 \cdot \mathbb{P}(1_B = 0) = \mathbb{P}(B)$. Given a random variable X, we define

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_i \frac{\mathbb{E}[X; B_i]}{\mathbb{P}(B_i)} 1_{B_i}.$$

This is the obvious definition, and it agrees with what we had before because $\mathbb{E}[1_A \mid \mathcal{G}]$ should be equal to $\mathbb{P}(A \mid \mathcal{G})$.

We now turn to some properties of conditional expectation. Some of the following propositions may seem a bit technical. In fact, they are! However, these properties are crucial to what follows and there is no choice but to master them.

Proposition 3.2. *$\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$ measurable, that is, if $Y = \mathbb{E}[X \mid \mathcal{G}]$, then $(Y > a)$ is a set in $\mathcal{G}$ for each real a.*

Proof. By the definition,

$$Y = \mathbb{E}[X \mid \mathcal{G}] = \sum_i \frac{\mathbb{E}[X; B_i]}{\mathbb{P}(B_i)} 1_{B_i} = \sum_i b_i 1_{B_i}$$

if we set $b_i = \mathbb{E}[X; B_i]/\mathbb{P}(B_i)$. The set $(Y \geq a)$ is a union of some of the B_i, namely, those B_i for which $b_i \geq a$. But the union of any collection of the B_i is in $\mathcal{G}$. $\square$

An example might help. Suppose

$$Y = 2 \cdot 1_{B_1} + 3 \cdot 1_{B_2} + 6 \cdot 1_{B_3} + 4 \cdot 1_{B_4}$$

and $a = 3.5$. Then $(Y \geq a) = B_3 \cup B_4$, which is in $\mathcal{G}$.

Proposition 3.3. *If $C \in \mathcal{G}$ and $Y = \mathbb{E}[X \mid \mathcal{G}]$, then $\mathbb{E}[Y; C] = \mathbb{E}[X; C]$.*

Proof. Since $Y = \sum \frac{\mathbb{E}[X;B_i]}{\mathbb{P}(B_i)} 1_{B_i}$ and the B_i are disjoint, then

$$\mathbb{E}[Y; B_j] = \frac{\mathbb{E}[X; B_j]}{\mathbb{P}(B_j)} \mathbb{E}\, 1_{B_j} = \mathbb{E}[X; B_j].$$

Now if $C = B_{j_1} \cup \cdots \cup B_{j_n} \cup \cdots$, summing the above over the j_k gives $\mathbb{E}[Y; C] = \mathbb{E}[X; C]$. $\square$

Let us look at the above example for this proposition, and let us do the case where $C = B_2$. Note $1_{B_2} 1_{B_2} = 1_{B_2}$ because the product is $1 \cdot 1 = 1$ if ω is in B_2 and 0 otherwise. On the other hand, it is not possible for an ω to be in more than one of the B_i, so $1_{B_2} 1_{B_i} = 0$ if $i \neq 2$. Multiplying Y in the above example by 1_{B_2}, we see that

$$\begin{aligned}\mathbb{E}[Y; C] &= \mathbb{E}[Y; B_2] = \mathbb{E}[Y 1_{B_2}] = \mathbb{E}[3 \cdot 1_{B_2}]\\ &= 3\mathbb{E}[1_{B_2}] = 3\mathbb{P}(B_2).\end{aligned}$$

However the number 3 is not just any number; it is $\mathbb{E}[X; B_2]/\mathbb{P}(B_2)$. So

$$3\mathbb{P}(B_2) = \frac{\mathbb{E}[X; B_2]}{\mathbb{P}(B_2)} \mathbb{P}(B_2) = \mathbb{E}[X; B_2] = \mathbb{E}[X; C],$$

just as we wanted. If $C = B_1 \cup B_4$, for example, we then write

$$\begin{aligned}\mathbb{E}[X; C] &= \mathbb{E}[X 1_C] = \mathbb{E}[X(1_{B_2} + 1_{B_4})]\\ &= \mathbb{E}[X 1_{B_2}] + \mathbb{E}[X 1_{B_4}] = \mathbb{E}[X; B_2] + \mathbb{E}[X; B_4].\end{aligned}$$

By the first part, this equals $\mathbb{E}[Y; B_2] + \mathbb{E}[Y; B_4]$, and we undo the above string of equalities but with Y instead of X to see that this is $\mathbb{E}[Y; C]$.

If a r.v. Y is $\mathcal{G}$ measurable, then for any a we have $(Y = a) \in \mathcal{G}$ which means that $(Y = a)$ is the union of one or more of the B_i. Since the B_i are disjoint, it follows that Y must be constant on each B_i.

Again let us look at an example. Suppose Z takes only the values $1, 3, 4, 7$. Let $D_1 = (Z = 1), D_2 = (Z = 3), D_3 = (Z = 4), D_4 = (Z = 7)$. Note that we can write

$$Z = 1 \cdot 1_{D_1} + 3 \cdot 1_{D_2} + 4 \cdot 1_{D_3} + 7 \cdot 1_{D_4}.$$

To see this, if $\omega \in D_2$, for example, the right hand side will be $0 + 3 \cdot 1 + 0 + 0$, which agrees with $Z(\omega)$. Now if Z is $\mathcal{G}$ measurable, then $(Z \geq a) \in \mathcal{G}$ for each a. Take $a = 7$, and we see $D_4 \in \mathcal{G}$. Take $a = 4$ and we see $D_3 \cup D_4 \in \mathcal{G}$. Taking $a = 3$ shows $D_2 \cup D_3 \cup D_4 \in \mathcal{G}$.

Now $D_3 = (D_3 \cup D_4) \cap D_4^c$, so since $\mathcal{G}$ is a σ-field, $D_3 \in \mathcal{G}$. Similarly $D_2, D_1 \in \mathcal{G}$. Because sets in $\mathcal{G}$ are unions of the B_i's, we must have Z constant on the B_i's. For example, if it so happened that $D_1 = B_1, D_2 = B_2 \cup B_4, D_3 = B_3 \cup B_6 \cup B_7$, and $D_4 = B_5$, then

$$Z = 1 \cdot 1_{B_1} + 3 \cdot 1_{B_2} + 4 \cdot 1_{B_3} + 3 \cdot 1_{B_4} + 7 \cdot 1_{B_5} + +4 \cdot 1_{B_6} + 4 \cdot 1_{B_7}.$$

We still restrict ourselves to the discrete case. In this context, the properties given in Propositions 3.2 and 3.3 uniquely determine $\mathbb{E}[X \mid \mathcal{G}]$.

Proposition 3.4. *Suppose Z is $\mathcal{G}$ measurable and $\mathbb{E}[Z; C] = \mathbb{E}[X; C]$ whenever $C \in \mathcal{G}$. Then $Z = \mathbb{E}[X \mid \mathcal{G}]$.*

Proof. Since Z is $\mathcal{G}$ measurable, then Z must be constant on each B_i. Let the value of Z on B_i be z_i. So $Z = \sum_i z_i 1_{B_i}$. Then

$$z_i \mathbb{P}(B_i) = \mathbb{E}[Z; B_i] = \mathbb{E}[X; B_i],$$

or $z_i = \mathbb{E}[X; B_i]/\mathbb{P}(B_i)$ as required. □

The following propositions contain the main facts about this new definition of conditional expectation that we will need.

Proposition 3.5. *(1) If $X_1 \geq X_2$, then $\mathbb{E}[X_1 \mid \mathcal{G}] \geq \mathbb{E}[X_2 \mid \mathcal{G}]$.*
(2) $\mathbb{E}[aX_1 + bX_2 \mid \mathcal{G}] = a\mathbb{E}[X_1 \mid \mathcal{G}] + b\mathbb{E}[X_2 \mid \mathcal{G}]$.
(3) If X is $\mathcal{G}$ measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$.
(4) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E} X$.
(5) If X is independent of $\mathcal{G}$, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E} X$.

We will prove Proposition 3.5 in Note 1 at the end of the section. At this point it is more fruitful to understand what the proposition says.

We will see in Proposition 3.8 below that we may think of $\mathbb{E}[X \mid \mathcal{G}]$ as the best prediction of X given $\mathcal{G}$. Accepting this for the moment, we can give an interpretation of (1)-(5). (1) says that if X_1 is larger than X_2, then the predicted value of X_1 should be larger than the predicted value of X_2. (2) says that the predicted value of $X_1 + X_2$ should be the sum of the predicted values. (3) says that if we know $\mathcal{G}$ and X is $\mathcal{G}$ measurable, then we know X and our best prediction of X is X itself. (4) says that the average of the predicted value of X should be the predicted value of X. (5) says that if knowing $\mathcal{G}$ gives us no additional information on X, then the best prediction for the value of X is just $\mathbb{E} X$.

Proposition 3.6. *If Z is $\mathcal{G}$ measurable, then $\mathbb{E}[XZ \mid \mathcal{G}] = Z\mathbb{E}[X \mid \mathcal{G}]$.*

We again defer the proof, this time to Note 2.

Proposition 3.6 says that as far as conditional expectations with respect to a σ-field $\mathcal{G}$ go, $\mathcal{G}$-measurable random variables act like constants: they can be taken inside or outside the conditional expectation at will.

Proposition 3.7. *If $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, then*

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{H}\right] \mid \mathcal{G}\right] = \mathbb{E}\left[X \mid \mathcal{H}\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right] \mid \mathcal{H}\right].$$

Proof. $\mathbb{E}\left[X \mid \mathcal{H}\right]$ is $\mathcal{H}$ measurable, hence $\mathcal{G}$ measurable, since $\mathcal{H} \subset \mathcal{G}$. The left hand equality now follows by Proposition 3.5(3). To get the right hand equality, let W be the right hand expression. It is $\mathcal{H}$ measurable, and if $C \in \mathcal{H} \subset \mathcal{G}$, then

$$\mathbb{E}\left[W; C\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right]; C\right] = \mathbb{E}\left[X; C\right]$$

as required. □

In words, if we are predicting a prediction of X given limited information, this is the same as a single prediction given the least amount of information.

Let us verify that conditional expectation may be viewed as the best predictor of a random variable given a σ-field. If X is a r.v., a predictor Z is just another random variable, and the goodness of the prediction will be measured by $\mathbb{E}\left[(X-Z)^2\right]$, which is known as the mean square error.

Proposition 3.8. *If X is a r.v., the best predictor among the collection of $\mathcal{G}$-measurable random variables is $Y = \mathbb{E}\left[X \mid \mathcal{G}\right]$.*

Proof. Let Z be any $\mathcal{G}$-measurable random variable. We compute, using Proposition 3.5(3) and Proposition 3.6,

$$\begin{aligned}
\mathbb{E}\left[(X-Z)^2 \mid \mathcal{G}\right] &= \mathbb{E}\left[X^2 \mid \mathcal{G}\right] - 2\mathbb{E}\left[XZ \mid \mathcal{G}\right] + \mathbb{E}\left[Z^2 \mid \mathcal{G}\right] \\
&= \mathbb{E}\left[X^2 \mid \mathcal{G}\right] - 2Z\mathbb{E}\left[X \mid \mathcal{G}\right] + Z^2 \\
&= \mathbb{E}\left[X^2 \mid \mathcal{G}\right] - 2ZY + Z^2 \\
&= \mathbb{E}\left[X^2 \mid \mathcal{G}\right] - Y^2 + (Y-Z)^2 \\
&= \mathbb{E}\left[X^2 \mid \mathcal{G}\right] - 2Y\mathbb{E}\left[X \mid \mathcal{G}\right] + Y^2 + (Y-Z)^2 \\
&= \mathbb{E}\left[X^2 \mid \mathcal{G}\right] - 2\mathbb{E}\left[XY \mid \mathcal{G}\right] + \mathbb{E}\left[Y^2 \mid \mathcal{G}\right] + (Y-Z)^2 \\
&= \mathbb{E}\left[(X-Y)^2 \mid \mathcal{G}\right] + (Y-Z)^2.
\end{aligned}$$

We also used the fact that Y is $\mathcal{G}$ measurable. Taking expectations and using Proposition 3.5(4),

$$\mathbb{E}\left[(X-Z)^2\right] = \mathbb{E}\left[(X-Y)^2\right] + \mathbb{E}\left[(Y-Z)^2\right].$$

The right hand side is bigger than or equal to $\mathbb{E}\left[(X-Y)^2\right]$ because $(Y-Z)^2 \geq 0$. So the error in predicting X by Z is larger than the error in predicting X by Y, and will be equal if and only if $Z = Y$. So Y is the best predictor. □

There is one more interpretation of conditional expectation that may be useful. The collection of all random variables is a linear space, and the collection of all $\mathcal{G}$-measurable random variables is clearly a subspace. Given X, the conditional expectation $Y = \mathbb{E}[X \mid \mathcal{G}]$ is equal to the projection of X onto the subspace of $\mathcal{G}$-measurable random variables. To see this, we write $X = Y + (X - Y)$, and what we have to check is that the inner product of Y and $X - Y$ is 0, that is, Y and $X - Y$ are orthogonal. In this context, the inner product of X_1 and X_2 is defined to be $\mathbb{E}[X_1 X_2]$, so we must show $\mathbb{E}[Y(X - Y)] = 0$. Note

$$\mathbb{E}[Y(X-Y) \mid \mathcal{G}] = Y\mathbb{E}[X - Y \mid \mathcal{G}] = Y(\mathbb{E}[X \mid \mathcal{G}] - Y) = Y(Y - Y) = 0.$$

Taking expectations,

$$\mathbb{E}[Y(X-Y)] = \mathbb{E}[\mathbb{E}[Y(X-Y) \mid \mathcal{G}]] = 0,$$

just as we wished.

If Y is a discrete random variables, that is, it takes only countably many values $y_1, y_2, \ldots$, we let $B_i = (Y = y_i)$. These will be disjoint sets whose union is Ω. If $\sigma(Y)$ is the collection of all unions of the B_i, then $\sigma(Y)$ is a σ-field, and is called the σ-field generated by Y. It is easy to see that this is the smallest σ-field with respect to which Y is measurable. We write $\mathbb{E}[X \mid Y]$ for $\mathbb{E}[X \mid \sigma(Y)]$.

Note 1. We prove Proposition 3.5. (1) and (2) are immediate from the definition. To prove (3), note that if $Z = X$, then Z is $\mathcal{G}$ measurable and $\mathbb{E}[X; C] = \mathbb{E}[Z; C]$ for any $C \in \mathcal{G}$; this is trivial. By Proposition 3.4 it follows that $Z = \mathbb{E}[X \mid \mathcal{G}]$;this proves (3). To prove (4), if we let $C = \Omega$ and $Y = \mathbb{E}[X \mid \mathcal{G}]$, then $\mathbb{E}\,Y = \mathbb{E}[Y; C] = \mathbb{E}[X; C] = \mathbb{E}\,X$.

Last is (5). Let $Z = \mathbb{E}\,X$. Z is constant, so clearly $\mathcal{G}$ measurable. By the independence, if $C \in \mathcal{G}$, then $\mathbb{E}[X; C] = \mathbb{E}[X 1_C] = (\mathbb{E}\,X)(\mathbb{E}\,1_C) = (\mathbb{E}\,X)(\mathbb{P}(C))$. But $\mathbb{E}[Z; C] = (\mathbb{E}\,X)(\mathbb{P}(C))$ since Z is constant. By Proposition 3.4 we see $Z = \mathbb{E}[X \mid \mathcal{G}]$. $\square$

Note 2. We prove Proposition 3.6. Note that $Z\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$ measurable, so by Proposition 3.4 we need to show its expectation over sets C in $\mathcal{G}$ is the same as that of XZ. As in the proof of Proposition 3.3, it suffices to consider only the case when C is one of the B_i. Now Z is $\mathcal{G}$ measurable, hence it is constant on B_i; let its value be z_i. Then

$$\mathbb{E}[Z\mathbb{E}[X \mid \mathcal{G}]; B_i] = \mathbb{E}[z_i \mathbb{E}[X \mid \mathcal{G}]; B_i] = z_i \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]; B_i] = z_i \mathbb{E}[X; B_i] = \mathbb{E}[XZ; B_i]$$

as desired. $\square$

4. Martingales.

Suppose we have a sequence of σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \cdots$. An example would be repeatedly tossing a coin and letting $\mathcal{F}_k$ be the sets that can be determined by the first k tosses. Another example is to let $\mathcal{F}_k$ be the events that are determined by the values of a stock at times 1 through k. A third example is to let $X_1, X_2, \ldots$ be a sequence of random variables and let $\mathcal{F}_k$ be the σ-field generated by $X_1, \ldots, X_k$, the smallest σ-field with respect to which $X_1, \ldots, X_k$ are measurable.

Definition 4.1. *A r.v. X is integrable if $\mathbb{E}|X| < \infty$. Given an increasing sequence of σ-fields $\mathcal{F}_n$, a sequence of r.v.'s X_n is adapted if X_n is $\mathcal{F}_n$ measurable for each n.*

Definition 4.2. *A martingale M_n is a sequence of random variables such that*

(1) M_n is integrable for all n,
(2) M_n is adapted to $\mathcal{F}_n$, and
(3) for all n

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n. \tag{4.1}$$

Usually (1) and (2) are easy to check, and it is (3) that is the crucial property. If we have (1) and (2), but instead of (3) we have

(3′) for all n

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \geq M_n,$$

then we say M_n is a submartingale. If we have (1) and (2), but instead of (3) we have

(3″) for all n

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \leq M_n,$$

then we say M_n is a supermartingale.

Submartingales tends to increase and supermartingales tend to decrease. The nomenclature may seem like it goes the wrong way; Doob defined these terms by analogy with the notions of subharmonic and superharmonic functions in analysis. (Actually, it is more than an analogy: we won't explore this, but it turns out that the composition of a subharmonic function with Brownian motion yields a submartingale, and similarly for superharmonic functions.)

Note that the definition of martingale depends on the collection of σ-fields. When it is needed for clarity, one can say that $(M_n, \mathcal{F}_n)$ is a martingale. To define conditional expectation, one needs a probability, so a martingale depends on the probability as well. When we need to, we will say that M_n is a martingale with respect to the probability $\mathbb{P}$. This is an issue when there is more than one probability around.

We will see that martingales are ubiquitous in financial math. For example, security prices and one's wealth will turn out to be examples of martingales.

The word "martingale" is also used for the piece of a horse's bridle that runs from the horse's head to its chest. It keeps the horse from raising its head too high. It turns out that martingales in probability cannot get too large. The word also refers to a gambling system. I did some searching on the Internet, and there seems to be no consensus on the derivation of the term.

Here is an example of a martingale. Let $X_1, X_2, \dots$ be a sequence of independent r.v.'s with mean 0 that are independent. (Saying a r.v. X_i has mean 0 is the same as saying $\mathbb{E}\, X_i = 0$; this presupposes that $\mathbb{E}\,|X_1|$ is finite.) Set $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, the σ-field generated by $X_1, \dots, X_n$. Let $M_n = \sum_{i=1}^n X_i$. Definition 4.2(2) is easy to see. Since $\mathbb{E}\,|M_n| \le \sum_{i=1}^n \mathbb{E}\,|X_i|$, Definition 4.2(1) also holds. We now check

$$\mathbb{E}\,[M_{n+1} \mid \mathcal{F}_n] = X_1 + \cdots + X_n + \mathbb{E}\,[X_{n+1} \mid \mathcal{F}_n] = M_n + \mathbb{E}\, X_{n+1} = M_n,$$

where we used the independence.

Another example: suppose in the above that the X_k all have variance 1, and let $M_n = S_n^2 - n$, where $S_n = \sum_{i=1}^n X_i$. Again (1) and (2) of Definition 4.2 are easy to check. We compute

$$\mathbb{E}\,[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}\,[S_n^2 + 2X_{n+1}S_n + X_{n+1}^2 \mid \mathcal{F}_n] - (n+1).$$

We have $\mathbb{E}\,[S_n^2 \mid \mathcal{F}_n] = S_n^2$ since S_n is $\mathcal{F}_n$ measurable.

$$\mathbb{E}\,[2X_{n+1}S_n \mid \mathcal{F}_n] = 2S_n\mathbb{E}\,[X_{n+1} \mid \mathcal{F}_n] = 2S_n\mathbb{E}\, X_{n+1} = 0.$$

And $\mathbb{E}\,[X_{n+1}^2 \mid \mathcal{F}_n] = \mathbb{E}\, X_{n+1}^2 = 1$. Substituting, we obtain $\mathbb{E}\,[M_{n+1} \mid \mathcal{F}_n] = M_n$, or M_n is a martingale.

A third example: Suppose you start with a dollar and you are tossing a fair coin independently. If it turns up heads you double your fortune, tails you go broke. This is "double or nothing." Let M_n be your fortune at time n. To formalize this, let $X_1, X_2, \dots$ be independent r.v.'s that are equal to 2 with probability $\frac{1}{2}$ and 0 with probability $\frac{1}{2}$. Then $M_n = X_1 \cdots X_n$. Let $\mathcal{F}_n$ be the σ-field generated by $X_1, \dots, X_n$. Note $0 \le M_n \le 2^n$, and so Definition 4.2(1) is satisfied, while (2) is easy. To compute the conditional expectation, note $\mathbb{E}\, X_{n+1} = 1$. Then

$$\mathbb{E}\,[M_{n+1} \mid \mathcal{F}_n] = M_n\mathbb{E}\,[X_{n+1} \mid \mathcal{F}_n] = M_n\mathbb{E}\, X_{n+1} = M_n,$$

using the independence.

Before we give our fourth example, let us observe that

$$|\mathbb{E}\,[X \mid \mathcal{F}]| \le \mathbb{E}\,[|X| \mid \mathcal{F}]. \tag{4.2}$$

To see this, we have $-|X| \le X \le |X|$, so $-\mathbb{E}\,[|X| \mid \mathcal{F}] \le \mathbb{E}\,[X \mid \mathcal{F}] \le \mathbb{E}\,[|X| \mid \mathcal{F}]$. Since $\mathbb{E}\,[|X| \mid \mathcal{F}]$ is nonnegative, (4.2) follows.

Our fourth example will be used many times, so we state it as a proposition.

Proposition 4.3. *Let $\mathcal{F}_1, \mathcal{F}_2, \dots$ be given and let X be a fixed r.v. with $\mathbb{E}\,|X| < \infty$. Let $M_n = \mathbb{E}\,[X \mid \mathcal{F}_n]$. Then M_n is a martingale.*

Proof. Definition 4.2(2) is clear, while

$$\mathbb{E}\,|M_n| \le \mathbb{E}\,[\mathbb{E}\,[|X| \mid \mathcal{F}_n]] = \mathbb{E}\,|X| < \infty$$

by (4.2); this shows Definition 4.2(1). We have

$$\mathbb{E}\,[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}\,[\mathbb{E}\,[X \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = \mathbb{E}\,[X \mid \mathcal{F}_n] = M_n.$$

□

5. Properties of martingales.

When it comes to discussing American options, we will need the concept of stopping times. A mapping τ from Ω into the nonnegative integers is a stopping time if $(\tau = k) \in \mathcal{F}_k$ for each k.

An example is $\tau = \min\{k : S_k \ge A\}$. This is a stopping time because $(\tau = k) = (S_1, \dots, S_{k-1} < A, S_k \ge A) \in \mathcal{F}_k$. We can think of a stopping time as the first time something happens. $\sigma = \max\{k : S_k \ge A\}$, the last time, is not a stopping time.

Here is an intuitive description of a stopping time. If I tell you to drive to the city limits and then drive until you come to the second stop light after that, you know when you get there that you have arrived; you don't need to have been there before or to look ahead. But if I tell you to drive until you come to the second stop light **before** the city limits, either you must have been there before or else you have to go past where you are supposed to stop, continue on to the city limits, and then turn around and come back two stop lights. You don't know when you first get to the second stop light before the city limits that you get to stop there. The first set of instructions forms a stopping time, the second set does not.

Note $(\tau \le k) = \cup_{j=0}^{k} (\tau = j)$. Since $(\tau = j) \in \mathcal{F}_j \subset \mathcal{F}_k$, then the event $(\tau \le k) \in \mathcal{F}_k$ for all k. Conversely, if τ is a r.v. with $(\tau \le k) \in \mathcal{F}_k$ for all k, then

$$(\tau = k) = (\tau \le k) - (\tau \le k-1).$$

Since $(\tau \le k) \in \mathcal{F}_k$ and $(\tau \le k-1) \in \mathcal{F}_{k-1} \subset \mathcal{F}_k$, then $(\tau = k) \in \mathcal{F}_k$, and such a τ must be a stopping time.

Our first result is Jensen's inequality.

Proposition 5.1. *If g is convex, then*

$$g(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[g(X) \mid \mathcal{G}]$$

provided all the expectations exist.

For ordinary expectations rather than conditional expectations, this is still true. That is, if g is convex and the expectations exist, then

$$g(\mathbb{E} X) \leq \mathbb{E}[g(X)].$$

We already know some special cases of this: when $g(x) = |x|$, this says $|\mathbb{E} X| \leq \mathbb{E} |X|$; when $g(x) = x^2$, this says $(\mathbb{E} X)^2 \leq \mathbb{E} X^2$, which we know because $\mathbb{E} X^2 - (\mathbb{E} X)^2 = \mathbb{E} (X - \mathbb{E} X)^2 \geq 0$.

For Proposition 5.1 as well as many of the following propositions, the statement of the result is more important than the proof, and we relegate the proof to Note 1 below.

One reason we want Jensen's inequality is to show that a convex function applied to a martingale yields a submartingale.

Proposition 5.2. *If M_n is a martingale and g is convex, then $g(M_n)$ is a submartingale, provided all the expectations exist.*

Proof. By Jensen's inequality,

$$\mathbb{E}[g(M_{n+1}) \mid \mathcal{F}_n] \geq g(\mathbb{E}[M_{n+1} \mid \mathcal{F}_n]) = g(M_n).$$

$\square$

If M_n is a martingale, then $\mathbb{E} M_n = \mathbb{E}[\mathbb{E}[M_{n+1} \mid \mathcal{F}_n]] = \mathbb{E} M_{n+1}$. So $\mathbb{E} M_0 = \mathbb{E} M_1 = \cdots = \mathbb{E} M_n$. Doob's optional stopping theorem says the same thing holds when fixed times n are replaced by stopping times.

Theorem 5.3. *Suppose K is a positive integer, N is a stopping time such that $N \leq K$ a.s., and M_n is a martingale. Then*

$$\mathbb{E} M_N = \mathbb{E} M_K.$$

Here, to evaluate M_N, one first finds $N(\omega)$ and then evaluates $M_\cdot(\omega)$ for that value of N.

Proof. We have

$$\mathbb{E} M_N = \sum_{k=0}^{K} \mathbb{E}[M_N; N = k].$$

If we show that the k-th summand is $\mathbb{E}[M_n; N = k]$, then the sum will be

$$\sum_{k=0}^{K} \mathbb{E}[M_n; N = k] = \mathbb{E} M_n$$

as desired. We have

$$\mathbb{E}[M_N; N = k] = \mathbb{E}[M_k; N = k]$$

by the definition of M_N. Now $(N = k)$ is in $\mathcal{F}_k$, so by Proposition 2.2 and the fact that $M_k = \mathbb{E}[M_{k+1} \mid \mathcal{F}_k]$,

$$\mathbb{E}[M_k; N = k] = \mathbb{E}[M_{k+1}; N = k].$$

We have $(N = k) \in \mathcal{F}_k \subset \mathcal{F}_{k+1}$. Since $M_{k+1} = \mathbb{E}[M_{k+2} \mid \mathcal{F}_{k+1}]$, Proposition 2.2 tells us that

$$\mathbb{E}[M_{k+1}; N = k] = \mathbb{E}[M_{k+2}; N = k].$$

We continue, using $(N = k) \in \mathcal{F}_k \subset \mathcal{F}_{k+1} \subset \mathcal{F}_{k+2}$, and we obtain

$$\mathbb{E}[M_N; N = k] = \mathbb{E}[M_k; N = k] = \mathbb{E}[M_{k+1}; N = k] = \cdots = \mathbb{E}[M_n; N = k].$$

$\square$

If we change the equalities in the above to inequalities, the same result holds for submartingales.

As a corollary we have two of Doob's inequalities:

Theorem 5.4. *If M_n is a nonnegative submartingale,*

(a) $\mathbb{P}(\max_{k \le n} M_k \ge \lambda) \le \frac{1}{\lambda} \mathbb{E} M_n$.

(b) $\mathbb{E}(\max_{k \le n} M_k^2) \le 4 \mathbb{E} M_n^2$.

For the proof, see Note 2 below.

Note 1. We prove Proposition 5.1. If g is convex, then the graph of g lies above all the tangent lines. Even if g does not have a derivative at x_0, there is a line passing through x_0 which lies beneath the graph of g. So for each x_0 there exists $c(x_0)$ such that

$$g(x) \ge g(x_0) + c(x_0)(x - x_0).$$

Apply this with $x = X(\omega)$ and $x_0 = \mathbb{E}[X \mid \mathcal{G}](\omega)$. We then have

$$g(X) \ge g(\mathbb{E}[X \mid \mathcal{G}]) + c(\mathbb{E}[X \mid \mathcal{G}])(X - \mathbb{E}[X \mid \mathcal{G}]).$$

If g is differentiable, we let $c(x_0) = g'(x_0)$. In the case where g is not differentiable, then we choose c to be the left hand upper derivate, for example. (For those who are not familiar with derivates, this is essentially the left hand derivative.) One can check that if c is so chosen, then $c(\mathbb{E}[X \mid \mathcal{G}])$ is $\mathcal{G}$ measurable.

Now take the conditional expectation with respect to $\mathcal{G}$. The first term on the right is $\mathcal{G}$ measurable, so remains the same. The second term on the right is equal to

$$c(\mathbb{E}[X \mid \mathcal{G}])\mathbb{E}[X - \mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G}] = 0.$$

$\square$

Note 2. We prove Theorem 5.4. Set $M_{n+1} = M_n$. It is easy to see that the sequence $M_1, M_2, \ldots, M_{n+1}$ is also a martingale. Let $N = \min\{k : M_k \geq \lambda\} \wedge (n+1)$, the first time that M_k is greater than or equal to λ, where $a \wedge b = \min(a, b)$. Then

$$\mathbb{P}(\max_{k \leq n} M_k \geq \lambda) = \mathbb{P}(N \leq n)$$

and if $N \leq n$, then $M_N \geq \lambda$. Now

$$\begin{aligned} \mathbb{P}(\max_{k \leq n} M_k \geq \lambda) &= \mathbb{E}[1_{(N \leq n)}] \leq \mathbb{E}\Big[\frac{M_N}{\lambda}; N \leq n\Big] \\ &= \frac{1}{\lambda}\mathbb{E}[M_{N\wedge n}; N \leq n] \leq \frac{1}{\lambda}\mathbb{E}\, M_{N \wedge n}. \end{aligned} \tag{5.1}$$

Finally, since M_n is a submartingale, $\mathbb{E}\, M_{N\wedge n} \leq \mathbb{E}\, M_n$.

We now look at (b). Let us write M^* for $\max_{k \leq n} M_k$. We have

$$\mathbb{E}[M_{N\wedge n}; N \leq n] = \sum_{k=0}^{\infty} \mathbb{E}[M_{k\wedge n}; N = k].$$

Arguing as in the proof of Theorem 5.3,

$$\mathbb{E}[M_{k\wedge n}; N = k] \leq \mathbb{E}[M_n; N = k],$$

and so

$$\mathbb{E}[M_{N\wedge n}; N \leq n] \leq \sum_{k=0}^{\infty} \mathbb{E}[M_n; N = k] = \mathbb{E}[M_n; N \leq n].$$

The last expression is at most $\mathbb{E}[M_n; M^* \geq \lambda]$. If we multiply (5.1) by 2λ and integrate over λ from 0 to ∞, we obtain

$$\begin{aligned} \int_0^\infty 2\lambda \mathbb{P}(M^* \geq \lambda) d\lambda &\leq 2\int_0^\infty \mathbb{E}[M_n : M^* \geq \lambda] \\ &= 2\mathbb{E}\int_0^\infty M_n 1_{(M^* \geq \lambda)} d\lambda \\ &= 2\mathbb{E}\Big[M_n \int_0^{M^*} d\lambda\Big] \\ &= 2\mathbb{E}[M_n M^*]. \end{aligned}$$

Using Cauchy-Schwarz, this is bounded by

$$2(\mathbb{E}\,M_n^2)^{1/2}(\mathbb{E}\,(M^*)^2)^{1/2}.$$

On the other hand,

$$\begin{aligned}\int_0^\infty 2\lambda\mathbb{P}(M^*\geq\lambda)d\lambda &= \mathbb{E}\int_0^\infty 2\lambda 1_{(M^*\geq\lambda)}d\lambda\\ &= \mathbb{E}\int_0^{M^*} 2\lambda\,d\lambda = \mathbb{E}\,(M^*)^2.\end{aligned}$$

We therefore have

$$\mathbb{E}\,(M^*)^2 \leq 2(\mathbb{E}\,M_n^2)^{1/2}(\mathbb{E}\,(M^*)^2)^{1/2}.$$

Suppose $\mathbb{E}\,(M^*)^2<\infty$. We divide both sides by $(\mathbb{E}\,(M^*)^2)^{1/2}$, provided this is not infinite, square both sides, and obtain (a).

When $\mathbb{E}\,(M^*)^2$ is infinite, we need a way to circumvent the division by infinity. It goes as follows. The proof of (b) started with $\mathbb{P}(M_n^*\geq a)\leq \frac{1}{a}\mathbb{E}\,[|M_n|; M_n^*\geq a]$ and ended up with $\mathbb{E}\,[(M_n^*)^2]\leq 4\mathbb{E}\,|M_n|^2$. We will use the notation $a\wedge b=\min(a,b)$. Let $K>0$, and by considering the two cases $K>a$ and $K\leq a$, we see that this inequality implies

$$\mathbb{P}(M_n^*\wedge K\geq a)\leq \frac{1}{a}\mathbb{E}\,[|M_n|\wedge K; M_n^*\wedge K\geq a].$$

From the above argument, we obtain

$$\mathbb{E}\,(M_n^*\wedge K)^2\leq 4\mathbb{E}\,[(|M_n|\wedge K)^2].$$

Now let $K\to\infty$. The right hand side is bounded by $\mathbb{E}\,|M_n|^2$. Fatou's lemma from measure theory (Note 4) allows one to conclude that the limit of the left hand side is $\mathbb{E}\,(M_n^*)^2$. □

Note 3. We will show that bounded martingales converge. (The hypothesis of boundedness can be weakened; for example, $\mathbb{E}\,|M_n|\leq c<\infty$ for some c not depending on n suffices.)

Theorem 5.5. *Suppose M_n is a martingale bounded in absolute value by K. That is, $|M_n|\leq K$ for all n. Then $\lim_{n\to\infty}M_n$ exists a.s.*

Proof. Since M_n is bounded, it can't tend to $+\infty$ or $-\infty$. The only possibility is that it might oscillate. Let $a<b$ be two rationals. What might go wrong is that M_n might be larger than b infinitely often and less than a infinitely often. If we show the probability of this is 0, then taking the union over all pairs of rationals (a,b) shows that almost surely M_n cannot oscillate, and hence must converge.

Fix a, b and let $S_1 = \min\{k : M_k \le a\}$, $T_1 = \min\{k > S_1 : M_k \ge b\}$, $S_2 = \min\{k > T_1 : M_k \le a\}$, and so on. Let $U_n = \max\{k : T_k \le n\}$. U_n is called the number of upcrossings up to time n.

We can write

$$2K \ge M_n - M_0 = \sum_{k=1}^{n} (M_{S_{k+1}\wedge n} - M_{T_k \wedge n}) + \sum_{k=1}^{\infty} (M_{T_k\wedge n} - M_{S_k\wedge n}) + (M_{S_1\wedge n} - M_0).$$

Now take expectations. The expectation of the first sum on the right and the last term are zero by optional stopping. The middle term is larger than $(b-a)U_n$, so we conclude

$$(b-a)\mathbb{E}\, U_n \le 2K.$$

Let $n \to \infty$ to see that $\mathbb{E}\, \max_n U_n < \infty$, which implies $\max_n U_n < \infty$ a.s., which is what we needed. $\square$

Note 4. We will state Fatou's lemma in the following form.

If X_n is a sequence of nonnegative random variables converging to X a.s., then $\mathbb{E}\, X \le \sup_n \mathbb{E}\, X_n$.

This formulation is equivalent to the classical one and is better suited for our use.

6. The one step binomial asset pricing model.

Let us begin by giving the simplest possible model of a stock and see how a European call option should be valued in this context.

Suppose we have a single stock whose price is S_0. Let d and u be two numbers with $0 < d < 1 < u$. Here "d" is a mnemonic for "down" and "u" for "up." After one time unit the stock price will be either uS_0 with probability P or else dS_0 with probability Q, where $P + Q = 1$. Instead of purchasing shares in the stock, you can also put your money in the bank where one will earn interest at rate r. Alternatives to the bank are money market funds or bonds; the key point is that these are considered to be risk-free.

A European call option in this context is the option to buy one share of the stock at time 1 at price K. K is called the strike price. Let S_1 be the price of the stock at time 1. If S_1 is less than K, then the option is worthless at time 1. If S_1 is greater than K, you can use the option at time 1 to buy the stock at price K, immediately turn around and sell the stock for price S_1 and make a profit of $S_1 - K$. So the value of the option at time 1 is

$$V_1 = (S_1 - K)^+,$$

where x^+ is $\max(x, 0)$. The principal question to be answered is: what is the value V_0 of the option at time 0? In other words, how much should one pay for a European call option with strike price K?

It is possible to buy a negative number of shares of a stock. This is equivalent to selling shares of a stock you don't have and is called selling short. If you sell one share of stock short, then at time 1 you must buy one share at whatever the market price is at that time and turn it over to the person that you sold the stock short to. Similarly you can buy a negative number of options, that is, sell an option.

You can also deposit a negative amount of money in the bank, which is the same as borrowing. We assume that you can borrow at the same interest rate r, not exactly a totally realistic assumption. One way to make it seem more realistic is to assume you have a large amount of money on deposit, and when you borrow, you simply withdraw money from that account.

We are looking at the simplest possible model, so we are going to allow only one time step: one makes an investment, and looks at it again one day later.

Let's suppose the price of a European call option is V_0 and see what conditions one can put on V_0. Suppose you start out with V_0 dollars. One thing you could do is buy one option. The other thing you could do is use the money to buy Δ_0 shares of stock. If $V_0 > \Delta_0 S_0$, there will be some money left over and you put that in the bank. If $V_0 < \Delta_0 S_0$, you do not have enough money to buy the stock, and you make up the shortfall by borrowing money from the bank. In either case, at this point you have $V_0 - \Delta_0 S_0$ in the bank and Δ_0 shares of stock.

If the stock goes up, at time 1 you will have

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

and if it goes down,

$$\Delta_0 d S_0 + (1+r)(V_0 - \Delta_0 S_0).$$

We have not said what Δ_0 should be. Let us do that now. Let $V_1^u = (uS_0 - K)^+$ and $V_1^d = (dS_0 - K)^+$. Note these are deterministic quantities, i.e., not random. Let

$$\Delta_0 = \frac{V_1^u - V_1^d}{uS_0 - dS_0},$$

and we will also need

$$W_0 = \frac{1}{1+r}\Big[\frac{1+r-d}{u-d} V_1^u + \frac{u-(1+r)}{u-d} V_1^d\Big].$$

In a moment we will do some algebra and see that if the stock goes up and you had bought stock instead of the option you would now have

$$V_1^u + (1+r)(V_0 - W_0),$$

while if the stock went down, you would now have

$$V_1^d + (1+r)(V_0 - W_0).$$

Let's check the first of these, the second being similar. We need to show

$$\Delta_0 u S_0 + (1+r)(V_0 - \Delta_0 S_0) = V_1^u + (1+r)(V_0 - W_0). \tag{6.1}$$

The left hand side of (6.1) is equal to

$$\Delta_0 S_0 (u - (1+r)) + (1+r)V_0 = \frac{V_1^u - V_1^d}{u-d}(u - (1+r)) + (1+r)V_0. \tag{6.2}$$

The right hand side of (6.1) is equal to

$$V_1^u - \Big[\frac{1+r-d}{u-d}V_1^u + \frac{u-(1+r)}{u-d}V_1^d\Big] + (1+r)V_0. \tag{6.3}$$

Now check that the coefficients of V_0, of V_1^u, and of V_1^d agree in (6.2) and (6.3).

Suppose that $V_0 > W_0$. What you want to do is come along with no money, sell one option for V_0 dollars, use the money to buy Δ_0 shares, and put the rest in the bank (or borrow if necessary). If the buyer of your option wants to exercise the option, you give him one share of stock and sell the rest. If he doesn't want to exercise the option, you sell your shares of stock and pocket the money. Remember it is possible to have a negative number of shares. You will have cleared $(1+r)(V_0 - W_0)$, whether the stock went up or down, with no risk.

If $V_0 < W_0$, you just do the opposite: sell Δ_0 shares of stock short, buy one option, and deposit or make up the shortfall from the bank. This time, you clear $(1+r)(W_0 - V_0)$, whether the stock goes up or down.

Now most people believe that you can't make a profit on the stock market without taking a risk. The name for this is "no free lunch," or "arbitrage opportunities do not exist." The only way to avoid this is if $V_0 = W_0$. In other words, we have shown that the only reasonable price for the European call option is W_0.

The "no arbitrage" condition is not just a reflection of the belief that one cannot get something for nothing. It also represents the belief that the market is freely competitive. The way it works is this: suppose $W_0 = \$3$. Suppose you could sell options at a price $V_0 = \$5$; this is larger than W_0 and you would earn $V_0 - W_0 = \$2$ per option without risk. Then someone else would observe this and decide to sell the same option at a price less than V_0 but larger than W_0, say \$4. This person would still make a profit, and customers would go to him and ignore you because they would be getting a better deal. But then a third person would decide to sell the option for less than your competition but more than W_0, say at \$3.50. This would continue as long as any one would try to sell an option above price W_0.

We will examine this problem of pricing options in more complicated contexts, and while doing so, it will become apparent where the formulas for Δ_0 and W_0 came from. At this point, we want to make a few observations.

Remark 6.1. First of all, if $1+r > u$, one would never buy stock, since one can always do better by putting money in the bank. So we may suppose $1+r < u$. We always have $1+r \geq 1 > d$. If we set

$$\overline{p} = \frac{1+r-d}{u-d}, \qquad \overline{q} = \frac{u-(1+r)}{u-d},$$

then $\overline{p}, \overline{q} \geq 0$ and $\overline{p}+\overline{q} = 1$. Thus $\overline{p}$ and $\overline{q}$ act like probabilities, but they have nothing to do with P and Q. Note also that the price $V_0 = W_0$ does not depend on P or Q. It does depend on $\overline{p}$ and $\overline{q}$, which seems to suggest that there is an underlying probability which controls the option price and is not the one that governs the stock price.

Remark 6.2. There is nothing special about European call options in our argument above. One could let V_1^u and V_d^1 be any two values of any option, which are paid out if the stock goes up or down, respectively. The above analysis shows we can exactly duplicate the result of buying any option V by instead buying some shares of stock. If in some model one can do this for any option, the market is called *complete* in this model.

Remark 6.3. If we let $\overline{\mathbb{P}}$ be the probability so that $S_1 = uS_0$ with probability $\overline{p}$ and $S_1 = dS_0$ with probability $\overline{q}$ and we let $\overline{\mathbb{E}}$ be the corresponding expectation, then some algebra shows that

$$V_0 = \frac{1}{1+r}\overline{\mathbb{E}}V_1.$$

This will be generalized later.

Remark 6.4. If one buys one share of stock at time 0, then one expects at time 1 to have $(Pu+Qd)S_0$. One then divides by $1+r$ to get the value of the stock in today's dollars. (r, the risk-free interest rate, can also be considered the rate of inflation. A dollar tomorrow is equivalent to $1/(1+r)$ dollars today.) Suppose instead of P and Q being the probabilities of going up and down, they were in fact $\overline{p}$ and $\overline{q}$. One would then expect to have $(\overline{p}u+\overline{q}d)S_0$ and then divide by $1+r$. Substituting the values for $\overline{p}$ and $\overline{q}$, this reduces to S_0. In other words, if $\overline{p}$ and $\overline{q}$ were the correct probabilities, one would expect to have the same amount of money one started with. When we get to the binomial asset pricing model with more than one step, we will see that the generalization of this fact is that the stock price at time n is a martingale, still with the assumption that $\overline{p}$ and $\overline{q}$ are the correct probabilities. This is a special case of the *fundamental theorem of finance*: there always exists some probability, not necessarily the one you observe, under which the stock price is a martingale.

Remark 6.5. Our model allows after one time step the possibility of the stock going up or going down, but only these two options. What if instead there are 3 (or more) possibilities. Suppose for example, that the stock goes up a factor u with probability P, down a factor d with probability Q, and remains constant with probability R, where $P + Q + R = 1$. The corresponding price of a European call option would be $(uS_0 - K)^+$, $(dS_0 - K)^+$, or $(S_0 - K)^+$. If one could replicate this outcome by buying and selling shares of the stock, then the "no arbitrage" rule would give the exact value of the call option in this model. But, except in very special circumstances, one cannot do this, and the theory falls apart. One has three equations one wants to satisfy, in terms of V_1^u, V_1^d, and V_1^c. (The "c" is a mnemonic for "constant.") There are however only two variables, Δ_0 and V_0 at your disposal, and most of the time three equations in two unknowns cannot be solved.

7. The multi-step binomial asset pricing model.

In this section we will obtain a formula for the pricing of options when there are n time steps, but each time the stock can only go up by a factor u or down by a factor d. The "Black-Scholes" formula we will obtain is already a nontrivial result that is useful.

We assume the following.

(1) Unlimited short selling of stock
(2) Unlimited borrowing
(3) No transaction costs
(4) Our buying and selling is on a small enough scale that it does not affect the market.

We need to set up the probability model. Ω will be all sequences of length n of H's and T's. S_0 will be a fixed number and we define $S_k(\omega) = u^j d^{k-j} S_0$ if the first k elements of a given $\omega \in \Omega$ has j occurrences of H and $k-j$ occurrences of T. (What we are doing is saying that if the j-th element of the sequence making up ω is an H, then the stock price goes up by a factor u; if T, then down by a factor d.) $\mathcal{F}_k$ will be the σ-field generated by $S_0, \ldots, S_k$.

Let
$$\overline{p} = \frac{(1+r) - d}{u - d}, \qquad \overline{q} = \frac{u - (1+r)}{u - d}$$
and define $\overline{\mathbb{P}}(\omega) = \overline{p}^j \overline{q}^{n-j}$ if ω has j appearances of H and $n-j$ appearances of T. We observe that under $\overline{\mathbb{P}}$ the random variables S_{k+1}/S_k are independent and equal to u with probability $\overline{p}$ and d with probability $\overline{q}$. To see this, let $Y_k = S_k/S_{k-1}$. Thus Y_k is the factor the stock price goes up or down at time k. Then $\mathbb{P}(Y_1 = y_1, \ldots, Y_n = y_n) = \overline{p}^j \overline{q}^{n-j}$,

where j is the number of the y_k that are equal to u. On the other hand, this is equal to $\mathbb{P}(Y_1 = y_1) \cdots \mathbb{P}(Y_n = y_n)$. Let $\overline{\mathbb{E}}$ denote the expectation corresponding to $\overline{\mathbb{P}}$.

The $\overline{\mathbb{P}}$ we construct may not be the true probabilities of going up or down. That doesn't matter - it will turn out that using the principle of "no arbitrage," it is $\overline{\mathbb{P}}$ that governs the price.

Our first result is the fundamental theorem of finance in the current context.

Proposition 7.1. *Under $\overline{\mathbb{P}}$ the discounted stock price $(1+r)^{-k} S_k$ is a martingale.*

Proof. Since the random variable S_{k+1}/S_k is independent of $\mathcal{F}_k$, we have

$$\overline{\mathbb{E}}\,[(1+r)^{-(k+1)} S_{k+1} \mid \mathcal{F}_k] = (1+r)^{-k} S_k (1+r)^{-1} \overline{\mathbb{E}}\,[S_{k+1}/S_k \mid \mathcal{F}_k].$$

Using the independence the conditional expectation on the right is equal to

$$\overline{\mathbb{E}}\,[S_{k+1}/S_k] = \overline{p} u + \overline{q} d = 1 + r.$$

Substituting yields the proposition. $\square$

Let Δ_k be the number of shares held between times k and $k+1$. We require Δ_k to be $\mathcal{F}_k$ measurable. $\Delta_0, \Delta_1, \ldots$ is called the portfolio process. Let W_0 be the amount of money you start with and let W_k be the amount of money you have at time k. W_k is the wealth process. If we have Δ_k shares between times k and $k+1$, then at time $k+1$ those shares will be worth $\Delta_k S_{k+1}$. The amount of cash we hold between time k and $k+1$ is W_k minus the amount held in stock, that is, $W_k - \Delta_k S_k$. At time $k+1$ this is worth $(1+r)[W_k - \Delta_k S_k]$. Therefore

$$W_{k+1} = \Delta_k S_{k+1} + (1+r)[W_k - \Delta_k S_k].$$

Note that in the case where $r = 0$ we have

$$W_{k+1} - W_k = \Delta_k (S_{k+1} - S_k),$$

or

$$W_{k+1} = \sum_{i=0}^{k} \Delta_i (S_{i+1} - S_i).$$

This is a discrete version of a stochastic integral. Since

$$\overline{\mathbb{E}}\,[W_{k+1} - W_k \mid \mathcal{F}_k] = \Delta_k \overline{\mathbb{E}}\,[S_{k+1} - S_k \mid \mathcal{F}_k] = 0,$$

it follows that in the case $r = 0$ that W_k is a martingale. More generally

Proposition 7.2. *Under $\overline{\mathbb{P}}$ the discounted wealth process $(1+r)^{-k}W_k$ is a martingale.*

Proof. We have

$$(1+r)^{-(k+1)}W_{k+1} = (1+r)^{-k}W_k + \Delta_k[(1+r)^{-(k+1)}S_{k+1} - (1+r)^{-k}S_k].$$

Observe that

$$\begin{aligned}\overline{\mathbb{E}}\,&[\Delta_k[(1+r)^{-(k+1)}S_{k+1} - (1+r)^{-k}S_k \mid \mathcal{F}_k] \\ &= \Delta_k\overline{\mathbb{E}}\,[(1+r)^{-(k+1)}S_{k+1} - (1+r)^{-k}S_k \mid \mathcal{F}_k] = 0.\end{aligned}$$

The result follows. □

Our next result is that the binomial model is complete. It is easy to lose the idea in the algebra, so first let us try to see why the theorem is true.

For simplicity let us first consider the case $r = 0$. Let $V_k = \mathbb{E}\,[V \mid \mathcal{F}_k]$; by Proposition 4.3 we see that V_k is a martingale. We want to construct a portfolio process, i.e., choose Δ_k's, so that $W_n = V$. We will do it inductively by arranging matters so that $W_k = V_k$ for all k. Recall that W_k is also a martingale.

Suppose we have $W_k = V_k$ at time k and we want to find Δ_k so that $W_{k+1} = V_{k+1}$. At the $(k+1)$-st step there are only two possible changes for the price of the stock and so since V_{k+1} is $\mathcal{F}_{k+1}$ measurable, only two possible values for V_{k+1}. We need to choose Δ_k so that $W_{k+1} = V_{k+1}$ for each of these two possibilities. We only have one parameter, Δ_k, to play with to match up two numbers, which may seem like an overconstrained system of equations. But both V and W are martingales, which is why the system can be solved.

Now let us turn to the details. In the following proof we allow $r \geq 0$.

Theorem 7.3. *The binomial asset pricing model is complete.*

The precise meaning of this is the following. If V is any random variable that is $\mathcal{F}_n$ measurable, there exists a constant W_0 and a portfolio process Δ_k so that the wealth process W_k satisfies $W_n = V$. In other words, starting with W_0 dollars, we can trade shares of stock to exactly duplicate the outcome of any option V.

Proof. Let

$$V_k = (1+r)^k\overline{\mathbb{E}}\,[(1+r)^{-n}V \mid \mathcal{F}_k].$$

By Proposition 4.3 $(1+r)^{-k}V_k$ is a martingale. If $\omega = (t_1, \ldots, t_n)$, where each t_i is an H or T, let

$$\Delta_k(\omega) = \frac{V_{k+1}(t_1,\ldots,t_k,H,t_{k+2},\ldots,t_n) - V_{k+1}(t_1,\ldots,t_k,T,t_{k+2},\ldots,t_n)}{S_{k+1}(t_1,\ldots,t_k,H,t_{k+2},\ldots,t_n) - S_{k+1}(t_1,\ldots,t_k,T,t_{k+2},\ldots,t_n)}.$$

Set $W_0 = V_0$, and we will show by induction that the wealth process at time k equals V_k.

The first thing to show is that Δ_k is $\mathcal{F}_k$ measurable. Neither S_{k+1} nor V_{k+1} depends on $t_{k+2}, \ldots, t_n$. So Δ_k depends only on the variables $t_1, \ldots, t_k$, hence is $\mathcal{F}_k$ measurable.

Now $t_{k+2}, \ldots, t_n$ play no role in the rest of the proof, and $t_1, \ldots, t_k$ will be fixed, so we drop the t's from the notation. If we write $V_{k+1}(H)$, this is an abbreviation for $V_{k+1}(t_1, \ldots, t_k, H, t_{k+2}, \ldots, t_n)$.

We know $(1+r)^{-k} V_k$ is a martingale under $\overline{\mathbb{P}}$ so that

$$\begin{aligned} V_k &= \overline{\mathbb{E}}\,[(1+r)^{-1} V_{k+1} \mid \mathcal{F}_k] \\ &= \frac{1}{1+r}[\overline{p} V_{k+1}(H) + \overline{q} V_{k+1}(T)]. \end{aligned} \tag{7.1}$$

(See Note 1.) We now suppose $W_k = V_k$ and want to show $W_{k+1}(H) = V_{k+1}(H)$ and $W_{k+1}(T) = V_{k+1}(T)$. Then using induction we have $W_n = V_n = V$ as required. We show the first equality, the second being similar.

$$\begin{aligned} W_{k+1}(H) &= \Delta_k S_{k+1}(H) + (1+r)[W_k - \Delta_k S_k] \\ &= \Delta_k [u S_k - (1+r) S_k] + (1+r) V_k \\ &= \frac{V_{k+1}(H) - V_{k+1}(T)}{(u-d) S_k} S_k [u - (1+r)] + \overline{p} V_{k+1}(H) + \overline{q} V_{k+1}(T) \\ &= V_{k+1}(H). \end{aligned}$$

We are done. □

Finally, we obtain the Black-Scholes formula in this context. Let V be any option that is $\mathcal{F}_n$-measurable. The one we have in mind is the European call, for which $V = (S_n - K)^+$, but the argument is the same for any option whatsoever.

Theorem 7.4. *The value of the option V at time 0 is $V_0 = (1+r)^{-n} \overline{\mathbb{E}}\, V$.*

Proof. We can construct a portfolio process Δ_k so that if we start with $W_0 = (1+r)^{-n} \overline{\mathbb{E}}\, V$, then the wealth at time n will equal V, no matter what the market does in between. If we could buy or sell the option V at a price other than W_0, we could obtain a riskless profit. That is, if the option V could be sold at a price c_0 larger than W_0, we would sell the option for c_0 dollars, use W_0 to buy and sell stock according to the portfolio process Δ_k, have a net worth of $V + (1+r)^n (c_0 - W_0)$ at time n, meet our obligation to the buyer of the option by using V dollars, and have a net profit, at no risk, of $(1+r)^n (c_0 - W_0)$. If c_0 were less than W_0, we would do the same except buy an option, hold $-\Delta_k$ shares at time k, and again make a riskless profit. By the "no arbitrage" rule, that can't happen, so the price of the option V must be W_0. □

Remark 7.5. Note that the proof of Theorem 7.4 tells you precisely what hedging strategy (i.e., what portfolio process to use).

In the binomial asset pricing model, there is no difficulty computing the price of a European call. We have

$$\overline{\mathbb{E}}\,(S_n - K)^+ = \sum_x (x - K)^+ \overline{\mathbb{P}}(S_n = x)$$

and

$$\mathbb{P}(S_n = x) = \binom{n}{k} \overline{p}^k \overline{q}^{n-k}$$

if $x = u^k d^{n-k} S_0$. Therefore the price of the European call is

$$\sum_{k=0}^{n} (u^k d^{n-k} S_0 - K)^+ \binom{n}{k} \overline{p}^k \overline{q}^{n-k}.$$

The formula in Theorem 7.4 holds for exotic options as well. Suppose

$$V = \max_{i=1,\ldots,n} S_i - \min_{j=1,\ldots,n} S_j.$$

In other words, you sell the stock for the maximum value it takes during the first n time steps and you buy at the minimum value the stock takes; you are allowed to wait until time n and look back to see what the maximum and minimum were. You can even do this if the maximum comes before the minimum. This V is still $\mathcal{F}_n$ measurable, so the theory applies. Naturally, such a "buy low, sell high" option is very desirable, and the price of such a V will be quite high. It is interesting that even without using options, you can duplicate the operation of buying low and selling high by holding an appropriate number of shares Δ_k at time k, where you do not look into the future to determine Δ_k.

Let us look at an example of a European call so that it is clear how to do the calculations. Consider the binomial asset pricing model with $n = 3$, $u = 2$, $d = \frac{1}{2}$, $r = 0.1$, $S_0 = 10$, and $K = 15$. If V is a European call with strike price K and exercise date n, let us compute explicitly the random variables V_1 and V_2 and calculate the value V_0. Let us also compute the hedging strategy Δ_0, Δ_1, and Δ_2.

Let

$$\overline{p} = \frac{(1+r) - d}{u - d} = .4, \qquad \overline{q} = \frac{u - (1+r)}{u - d} = .6.$$

The following table describes the values of the stock, the payoff V, and the probabilities for each possible outcome ω.

ω	S_1	S_2	S_3	V	Probability

HHH	$10u$	$10u^2$	$10u^3$	65	$\overline{p}^3$
HHT	$10u$	$10u^2$	$10u^2d$	5	$\overline{p}^2\overline{q}$
HTH	$10u$	$10ud$	$10u^2d$	5	$\overline{p}^2\overline{q}$
HTT	$10u$	$10ud$	$10ud^2$	0	$\overline{pq}^2$
THH	$10d$	$10ud$	$10u^2d$	5	$\overline{p}^2\overline{q}$
THT	$10d$	$10ud$	$10ud^2$	0	$\overline{pq}^2$
TTH	$10d$	$10d^2$	$10ud^2$	0	$\overline{pq}^2$
TTT	$10d$	$10d^2$	$10d^3$	0	$\overline{q}^3$

We then calculate

$$V_0 = (1+r)^{-3}\overline{\mathbb{E}}V = (1+r)^{-3}(65\overline{p}^3 + 15\overline{p}^2\overline{q}) = 4.2074.$$

$V_1 = (1+r)^{-2}\overline{\mathbb{E}}[V \mid \mathcal{F}_1]$, so we have

$$V_1(H) = (1+r)^{-2}(65\overline{p}^2 + 10\overline{pq}) = 10.5785, \qquad V_1(T) = (1+r)^{-2}5\overline{pq} = .9917.$$

$V_2 = (1+r)^{-1}\overline{\mathbb{E}}[V \mid \mathcal{F}_2]$, so we have

$$V_2(HH) = (1+r)^{-1}(65\overline{p} + 5\overline{q}) = 26.3636, \qquad V_2(HT) = (1+r)^{-1}5\overline{p} = 1.8182,$$
$$V_2(TH) = (1+r)^{-1}5\overline{p} = 1.8182, \qquad V_2(TT) = 0.$$

The formula for Δ_k is given by

$$\Delta_k = \frac{V_{k+1}(H) - V_{k+1}(T)}{S_{k+1}(H) - S_{k+1}(T)},$$

so

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)} = .6391,$$

where V_1 and S_1 are as above.

$$\Delta_1(H) = \frac{V_2(HH) - V_2(HT)}{S_2(HH) - S_2(HT)} = 1.2273, \qquad \Delta_1(T) = \frac{V_2(TH) - V_2(TT)}{S_2(TH) - S_2(TT)} = .1039.$$

$$\Delta_2(HH) = \frac{V_3(HHH) - V_3(HHT)}{S_3(HHH) - S_3(HHT)} = 1.0,$$
$$\Delta_2(HT) = \frac{V_3(HTH) - V_3(HTT)}{S_3(HTH) - S_3(HTT)} = .3333,$$
$$\Delta_2(TH) = \frac{V_3(THH) - V_3(THT)}{S_3(THH) - S_3(THT)} = .3333,$$
$$\Delta_2(TT) = \frac{V_3(TTH) - V_3(TTT)}{S_3(TTH) - S_3(TTT)} = 0.0.$$

Note 1. The second equality is (7.1) is not entirely obvious. Intuitively, it says that one has a heads with probability $\overline{p}$ and the value of V_{k+1} is $V_{k+1}(H)$ and one has tails with probability $\overline{q}$, and the value of V_{k+1} is $V_{k+1}(T)$.

Let us give a more rigorous proof of (7.1). The right hand side of (7.1) is $\mathcal{F}_k$ measurable, so we need to show that if $A \in \mathcal{F}_k$, then

$$\overline{\mathbb{E}}\,[V_{k+1}; A] = \overline{\mathbb{E}}\,[\overline{p}V_{k+1}(H) + \overline{q}V_{k+1}(T); A].$$

By linearity, it suffices to show this for $A = \{\omega = (t_1 t_2 \cdots t_n) : t_1 = s_1, \ldots, t_k = s_k\}$, where $s_1 s_2 \cdots s_k$ is any sequence of H's and T's. Now

$$\begin{aligned}\overline{\mathbb{E}}\,[V_{k+1}; s_1 \cdots s_k] &= \overline{\mathbb{E}}\,[V_{k+1}; s_1 \cdots s_k H] + \overline{\mathbb{E}}\,[V_{k+1}; s_1 \cdots s_k T]\\ &= V_{k+1}(s_1 \cdots s_k H)\overline{\mathbb{P}}(s_1 \cdots s_k H) + V_{k+1}(s_1 \cdots s_k T)\overline{\mathbb{P}}(s_1 \cdots s_k T).\end{aligned}$$

By independence this is

$$V_{k+1}(s_1 \cdots s_k H)\overline{\mathbb{P}}(s_1 \cdots s_k)\overline{p} + V_{k+1}(s_1 \cdots s_k T)\overline{\mathbb{P}}(s_1 \cdots s_k)\overline{q},$$

which is what we wanted.

8. American options.

An American option is one where you can exercise the option any time before some fixed time T. For example, on a European call, one can only use it to buy a share of stock at the expiration time T, while for an American call, at any time before time T, one can decide to pay K dollars and obtain a share of stock.

Let us give an informal argument on how to price an American call, giving a more rigorous argument in a moment. One can always wait until time T to exercise an American call, so the value must be at least as great as that of a European call. On the other hand, suppose you decide to exercise early. You pay K dollars, receive one share of stock, and your wealth is $S_t - K$. You hold onto the stock, and at time T you have one share of stock worth S_T, and for which you paid K dollars. So your wealth is $S_T - K \le (S_T - K)^+$. In fact, we have strict inequality, because you lost the interest on your K dollars that you would have received if you had waited to exercise until time T. Therefore an American call is worth no more than a European call, and hence its value must be the same as that of a European call.

This argument does not work for puts, because selling stock gives you some money on which you will receive interest, so it may be advantageous to exercise early. (A put is the option to sell a stock at a price K at time T.)

Here is the more rigorous argument. Suppose that if you exercise the option at time k, your payoff is $g(S_k)$. In present day dollars, that is, after correcting for inflation, you have $(1+r)^{-k}g(S_k)$. You have to make a decision on when to exercise the option, and that decision can only be based on what has already happened, not on what is going to happen in the future. In other words, we have to choose a stopping time τ, and we exercise the option at time $\tau(\omega)$. Thus our payoff is $(1+r)^{-\tau}g(S_\tau)$. This is a random quantity. What we want to do is find the stopping time that maximizes the expected value of this random variable. Of course, we work with $\overline{\mathbb{P}}$, and thus we are looking for the stopping time τ such that $\tau \le n$ and

$$\overline{\mathbb{E}}\,(1+r)^{-\tau}g(S_\tau)$$

is as large as possible. The problem of finding such a τ is called an optimal stopping problem.

Suppose $g(x)$ is convex with $g(0)=0$. Certainly $g(x)=(x-K)^+$ is such a function. We will show that $\tau \equiv n$ is the solution to the above optimal stopping problem: the best time to exercise is as late as possible.

We have

$$g(\lambda x) = g(\lambda x + (1-\lambda)\cdot 0) \le \lambda g(x) + (1-\lambda)g(0) = \lambda g(x). \tag{8.1}$$

By Jensen's inequality,

$$\begin{aligned}
\overline{\mathbb{E}}\,[(1+r)^{-(k+1)}g(S_{k+1}) \mid \mathcal{F}_k] &= (1+r)^{-k}\overline{\mathbb{E}}\Big[\frac{1}{1+r}g(S_{k+1}) \mid \mathcal{F}_k\Big]\\
&\ge (1+r)^{-k}\overline{\mathbb{E}}\Big[g\Big(\frac{1}{1+r}S_{k+1}\Big) \mid \mathcal{F}_k\Big]\\
&\ge (1+r)^{-k}g\Big(\overline{\mathbb{E}}\Big[\frac{1}{1+r}S_{k+1} \mid \mathcal{F}_k\Big]\Big)\\
&= (1+r)^{-k}g(S_k).
\end{aligned}$$

For the first inequality we used (8.1). So $(1+r)^{-k}g(S_k)$ is a submartingale. By optional stopping,

$$\overline{\mathbb{E}}\,[(1+r)^{-\tau}g(S_\tau)] \le \overline{\mathbb{E}}\,[(1+r)^{-n}g(S_n)],$$

so $\tau \equiv n$ always does best.

For puts, the payoff is $g(S_k)$, where $g(x)=(K-x)^+$. This is also convex function, but this time $g(0)\neq 0$, and the above argument fails.

Although good approximations are known, an exact solution to the problem of valuing an American put is unknown, and is one of the major unsolved problems in financial mathematics.

9. Continuous random variables.

We are now going to start working toward continuous times and stocks that can take any positive number as a value, so we need to prepare by extending some of our definitions.

Given any random variable $X \geq 0$, we can approximate it by r.v's X_n that are discrete. We let

$$X_n = \sum_{i=0}^{n2^n} \frac{i}{2^n} 1_{(i/2^n \leq X < (i+1)/2^n)}.$$

In words, if $X(\omega)$ lies between 0 and n, we let $X_n(\omega)$ be the closest value $i/2^n$ that is less than or equal to $X(\omega)$. For ω where $X(\omega) > n$ we set $X_n(\omega) = 0$. Clearly the X_n are discrete, and approximate X. In fact, on the set where $X \leq n$, we have that $|X(\omega) - X_n(\omega)| \leq 2^{-n}$.

For reasonable X we are going to define $\mathbb{E}\, X = \lim \mathbb{E}\, X_n$. Since the X_n increase with n, the limit must exist, although it could be $+\infty$. If X is not necessarily nonnegative, we define $\mathbb{E}\, X = \mathbb{E}\, X^+ - \mathbb{E}\, X^-$, provided at least one of $\mathbb{E}\, X^+$ and $\mathbb{E}\, X^-$ is finite. Here $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$.

There are some things one wants to prove, but all this has been worked out in measure theory and the theory of the Lebesgue integral; see Note 1. Let us confine ourselves here to showing this definition is the same as the usual one when X has a density.

Recall X has a density f_X if

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx$$

for all a and b. In this case

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

provided $\int_{-\infty}^{\infty} |x| f_X(x)dx < \infty$. With our definition of X_n we have

$$\mathbb{P}(X_n = i/2^n) = \mathbb{P}(X \in [i/2^n, (i+1)/2^n)) = \int_{i/2^n}^{(i+1)/2^n} f_X(x)dx.$$

Then

$$\mathbb{E}\, X_n = \sum_i \frac{i}{2^n} \mathbb{P}(X_n = i/2^n) = \sum_i \int_{i/2^n}^{(i+1)/2^n} \frac{i}{2^n} f_X(x)dx.$$

Since x differs from $i/2^n$ by at most $1/2^n$ when $x \in [i/2^n, (i+1)/2^n)$, this will tend to $\int x f_X(x)dx$, unless the contribution to the integral for $|x| \geq n$ does not go to 0 as $n \to \infty$. As long as $\int |x| f_X(x)dx < \infty$, one can show that this contribution does indeed go to 0.

We also need an extension of the definition of conditional probability. A r.v. is $\mathcal{G}$ measurable if $(X > a) \in \mathcal{G}$ for every a. How do we define $\mathbb{E}\,[Z \mid \mathcal{G}]$ when $\mathcal{G}$ is not generated by a countable collection of disjoint sets?

Again, there is a completely worked out theory that holds in all cases; see Note 2. Let us give a definition that is equivalent that works except for a very few cases. Let us suppose that for each n the σ-field $\mathcal{G}_n$ is finitely generated. This means that $\mathcal{G}_n$ is generated by finitely many disjoint sets $B_{n1}, \ldots, B_{nm_n}$. So for each n, the number of B_{ni} is finite but arbitrary, the B_{ni} are disjoint, and their union is Ω. Suppose also that $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \cdots$. Now $\cup_n \mathcal{G}_n$ will not in general be a σ-field, but suppose $\mathcal{G}$ is the smallest σ-field that contains all the $\mathcal{G}_n$. Finally, define $\mathbb{P}(A \mid \mathcal{G}) = \lim \mathbb{P}(A \mid \mathcal{G}_n)$.

This is a fairly general set-up. For example, let Ω be the real line and let $\mathcal{G}_n$ be generated by the sets $(-\infty, n), [n, \infty)$ and $[i/2^n, (i+1)/2^n)$. Then $\mathcal{G}$ will contain every interval that is closed on the left and open on the right, hence $\mathcal{G}$ must be the σ-field that one works with when one talks about Lebesgue measure on the line.

The question that one might ask is: how does one know the limit exists? Since the $\mathcal{G}_n$ increase, we know by Proposition 4.3 that $M_n = \mathbb{P}(A \mid \mathcal{G}_n)$ is a martingale with respect to the $\mathcal{G}_n$. It is certainly bounded above by 1 and bounded below by 0, so by the martingale convergence theorem, it must have a limit as $n \to \infty$.

Once one has a definition of conditional probability, one defines conditional expectation by what one expects. If X is discrete, one can write X as $\sum_j a_j 1_{A_j}$ and then one defines

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_j a_j \mathbb{P}(A_j \mid \mathcal{G}).$$

If the X is not discrete, one write $X = X^+ - X^-$, one approximates X^+ by discrete random variables, and takes a limit, and similarly for X^-. One has to worry about convergence, but everything does go through.

With this extended definition of conditional expectation, do all the properties of Section 2 hold? The answer is yes. See Note 2 again.

With continuous random variables, we need to be more cautious about what we mean when we say two random variables are equal. We say $X = Y$ almost surely, abbreviated "a.s.", if

$$\mathbb{P}(\{\omega : X(\omega) \neq Y(\omega)\}) = 0.$$

So $X = Y$ except for a set of probability 0. The a.s. terminology is used other places as well: $X_n \to Y$ a.s. means that except for a set of ω's of probability zero, $X_n(\omega) \to Y(\omega)$.

Note 1. The best way to define expected value is via the theory of the Lebesgue integral. A probability $\mathbb{P}$ is a measure that has total mass 1. So we define

$$\mathbb{E}\, X = \int X(\omega)\, \mathbb{P}(d\omega).$$

To recall how the definition goes, we say X is simple if $X(\omega) = \sum_{i=1}^m a_i 1_{A_i}(\omega)$ with each

$a_i \geq 0$, and for a simple X we define

$$\mathbb{E}\, X = \sum_{i=1}^{m} a_i \mathbb{P}(A_i).$$

If X is nonnegative, we define

$$\mathbb{E}\, X = \sup\{\mathbb{E}\, Y : Y \text{ simple },Y \leq X\}.$$

Finally, provided at least one of $\mathbb{E}\, X^+$ and $\mathbb{E}\, X^-$ is finite, we define

$$\mathbb{E}\, X = \mathbb{E}\, X^+ - \mathbb{E}\, X^-.$$

This is the same definition as described above.

Note 2. The Radon-Nikodym theorem from measure theory says that if $\mathbb{Q}$ and $\mathbb{P}$ are two finite measures on $(\Omega, \mathcal{G})$ and $\mathbb{Q}(A) = 0$ whenever $\mathbb{P}(A) = 0$ and $A \in \mathcal{G}$, then there exists an integrable function Y that is $\mathcal{G}$-measurable such that $\mathbb{Q}(A) = \int_A Y d\mathbb{P}$ for every measurable set A.

Let us apply the Radon-Nikodym theorem to the following situation. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $X \geq 0$ is integrable: $\mathbb{E}\, X < \infty$. Suppose $\mathcal{G} \subset \mathcal{F}$. Define two new probabilities on $\mathcal{G}$ as follows. Let $\mathbb{P}' = \mathbb{P}|_{\mathcal{G}}$, that is, $\mathbb{P}'(A) = \mathbb{P}(A)$ if $A \in \mathcal{G}$ and $\mathbb{P}'(A)$ is not defined if $A \in \mathcal{F} - \mathcal{G}$. Define $\mathbb{Q}$ by $\mathbb{Q}(A) = \int_A X d\mathbb{P} = \mathbb{E}\,[X; A]$ if $A \in \mathcal{G}$. One can show (using the monotone convergence theorem from measure theory) that $\mathbb{Q}$ is a finite measure on $\mathcal{G}$. (One can also use this definition to define $\mathbb{Q}(A)$ for $A \in \mathcal{F}$, but we only want to define $\mathbb{Q}$ on $\mathcal{G}$, as we will see in a moment.) So $\mathbb{Q}$ and $\mathbb{P}'$ are two finite measures on $(\Omega, \mathcal{G})$. If $A \in \mathcal{G}$ and $\mathbb{P}'(A) = 0$, then $\mathbb{P}(A) = 0$ and so it follows that $\mathbb{Q}(A) = 0$. By the Radon-Nikodym theorem there exists an integrable random variable Y such that Y is $\mathcal{G}$ measurable (this is why we worried about which σ-field we were working with) and

$$\mathbb{Q}(A) = \int_A Y d\mathbb{P}'$$

if $A \in \mathcal{G}$. Note

((a) Y is $\mathcal{G}$ measurable, and
(b) if $A \in \mathcal{G}$,

$$\mathbb{E}\,[Y; A] = \mathbb{E}\,[X; A]$$

because

$$\mathbb{E}\,[Y; A] = \mathbb{E}\,[Y 1_A] = \int_A Y d\mathbb{P} = \int_A Y d\mathbb{P}' = \mathbb{Q}(A) = \int_A X d\mathbb{P} = \mathbb{E}\,[X 1_A] = \mathbb{E}\,[X; A].$$

We define $\mathbb{E}[X \mid \mathcal{G}]$ to be the random variable Y. If X is integrable but not necessarily nonnegative, then X^+ and X^- will be integrable and we define

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X^+ \mid \mathcal{G}] - \mathbb{E}[X^- \mid \mathcal{G}].$$

We define

$$\mathbb{P}(B \mid \mathcal{G}) = \mathbb{E}[1_B \mid \mathcal{G}]$$

if $B \in \mathcal{F}$.

Let us show that there is only one r.v., up to almost sure equivalence, that satisfies (a) and (b) above. If Y and Z are $\mathcal{G}$ measurable, and $\mathbb{E}[Y; A] = \mathbb{E}[X; A] = \mathbb{E}[Z; A]$ for $A \in \mathcal{G}$, then the set $A_n = (Y > Z + \frac{1}{n})$ will be in $\mathcal{G}$, and so

$$\mathbb{E}[Z; A_n] + \tfrac{1}{n}\mathbb{P}(A_n) = \mathbb{E}[Z + \tfrac{1}{n}; A_n] \le \mathbb{E}[Y; A_n] = \mathbb{E}[Z; A_n].$$

Consequently $\mathbb{P}(A_n) = 0$. This is true for each positive integer n, so $\mathbb{P}(Y > Z) = 0$. By symmetry, $\mathbb{P}(Z > Y) = 0$, and therefore $\mathbb{P}(Y \ne Z) = 0$ as we wished.

If one checks the proofs of Propositions 2.3, 2.4, and 2.5, one sees that only properties (a) and (b) above were used. So the propositions hold for the new definition of conditional expectation as well.

In the case where $\mathcal{G}$ is finitely or countably generated, under both the new and old definitions (a) and (b) hold. By the uniqueness result, the new and old definitions agree.

10. Stochastic processes.

We will be talking about stochastic processes. Previously we discussed sequences $S_1, S_2, \ldots$ of r.v.'s. Now we want to talk about processes Y_t for $t \ge 0$. For example, we can think of S_t being the price of a stock at time t. Any nonnegative time t is allowed.

We typically let $\mathcal{F}_t$ be the smallest σ-field with respect to which Y_s is measurable for all $s \le t$. So $\mathcal{F}_t = \sigma(Y_s : s \le t)$. As you might imagine, there are a few technicalities one has to worry about. We will try to avoid thinking about them as much as possible, but see Note 1.

We call a collection of σ-fields $\mathcal{F}_t$ with $\mathcal{F}_s \subset \mathcal{F}_t$ if $s < t$ a filtration. We say the filtration satisfies the "usual conditions" if the $\mathcal{F}_t$ are right continuous and complete (see Note 1); all the filtrations we consider will satisfy the usual conditions.

We say a stochastic process has continuous paths if the following holds. For each ω, the map $t \to Y_t(\omega)$ defines a function from $[0, \infty)$ to $\mathbb{R}$. If this function is a continuous function for all ω's except for a set of probability zero, we say Y_t has continuous paths.

Definition 10.1. *A mapping $\tau : \Omega \to [0, \infty)$ is a stopping time if for each t we have*

$$(\tau \le t) \in \mathcal{F}_t.$$

Typically, τ will be a continuous random variable and $\mathbb{P}(\tau = t) = 0$ for each t, which is why we need a definition just a bit different from the discrete case.

Since $(\tau < t) = \cup_{n=1}^{\infty}(\tau \le t - \frac{1}{n})$ and $(\tau \le t - \frac{1}{n}) \in \mathcal{F}_{t-\frac{1}{n}} \subset \mathcal{F}_t$, then for a stopping time τ we have $(\tau < t) \in \mathcal{F}_t$ for all t.

Conversely, suppose τ is a nonnegative r.v. for which $(\tau < t) \in \mathcal{F}_t$ for all t. We claim τ is a stopping time. The proof is easy, but we need the right continuity of the $\mathcal{F}_t$ here, so we put the proof in Note 2.

A continuous time martingale (or submartingale) is what one expects: each M_t is integrable, each M_t is $\mathcal{F}_t$ measurable, and if $s < t$, then

$$\mathbb{E}\,[M_t \mid \mathcal{F}_s] = M_s.$$

(Here we are saying the left hand side and the right hand side are equal almost surely; we will usually not write the "a.s." since almost all of our equalities for random variables are only almost surely.)

The analogues of Doob's theorems go through. Note 3 has the proofs.

Note 1. For technical reasons, one typically defines $\mathcal{F}_t$ as follows. Let $\mathcal{F}_t^0 = \sigma(Y_s : s \le t)$. This is what we referred to as $\mathcal{F}_t$ above. Next add to $\mathcal{F}_t^0$ all sets N for which $\mathbb{P}(N) = 0$. Such sets are called null sets, and since they have probability 0, they don't affect anything. In fact, one wants to add all sets N that we think of being null sets, even though they might not be measurable. To be more precise, we say N is a null set if $\inf\{\mathbb{P}(A) : A \in \mathcal{F}, N \subset A\} = 0$. Recall we are starting with a σ-field $\mathcal{F}$ and all the $\mathcal{F}_t^0$'s are contained in $\mathcal{F}$. Let $\mathcal{F}_t^{00}$ be the σ-field generated by $\mathcal{F}_t^0$ and all null sets N, that is, the smallest σ-field containing $\mathcal{F}_t^0$ and every null set. In measure theory terminology, what we have done is to say $\mathcal{F}_t^{00}$ is the completion of $\mathcal{F}_t^0$.

Lastly, we want to make our σ-fields right continuous. We set $\mathcal{F}_t = \cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}^{00}$. Although the union of σ-fields is not necessarily a σ-field, the intersection of σ-fields is. $\mathcal{F}_t$ contains $\mathcal{F}_t^{00}$ but might possibly contain more besides. An example of an event that is in $\mathcal{F}_t$ but that may not be in $\mathcal{F}_t^{00}$ is

$$A = \{\omega : \lim_{n\uparrow 0} Y_{t+\frac{1}{n}}(\omega) \ge 0\}.$$

$A \in \mathcal{F}_{t+\frac{1}{m}}^{00}$ for each m, so it is in $\mathcal{F}_t$. There is no reason it needs to be in $\mathcal{F}_t^{00}$ if Y is not necessarily continuous at t. It is easy to see that $\cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon} = \mathcal{F}_t$, which is what we mean when we say $\mathcal{F}_t$ is right continuous.

When talking about a stochastic process Y_t, there are various types of measurability one can consider. Saying Y_t is adapted to $\mathcal{F}_t$ means Y_t is $\mathcal{F}_t$ measurable for each t. However, since Y_t is really a function of two variables, t and ω, there are other notions of measurability that come into play. We will be considering stochastic processes that have continuous paths or that are predictable (the definition will be given later), so these various types of measurability will not be an issue for us.

Note 2. Suppose $(\tau < t) \in \mathcal{F}_t$ for all t. Then for each positive integer n_0,

$$(\tau \le t) = \cap_{n=n_0}^{\infty} (\tau < t + \tfrac{1}{n}).$$

For $n \ge n_0$ we have $(\tau < t + \frac{1}{n}) \in \mathcal{F}_{t+\frac{1}{n}} \subset \mathcal{F}_{t+\frac{1}{n_0}}$. Therefore $(\tau \le t) \in \mathcal{F}_{t+\frac{1}{n_0}}$ for each n_0. Hence the set is in the intersection: $\cap_{n_0>1} \mathcal{F}_{t+\frac{1}{n_0}} \subset \cap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} = \mathcal{F}_t$.

Note 3. We want to prove the analogues of Theorems 5.3 and 5.4. The proof of Doob's inequalities are simpler. We only will need the analogue of Theorem 5.4(b).

Theorem 10.2. *Suppose M_t is a martingale with continuous paths and $\mathbb{E}\, M_t^2 < \infty$ for all t. Then for each t_0*

$$\mathbb{E}\,[(\sup_{s \le t_0} M_s)^2] \le 4\mathbb{E}\,[|M_{t_0}|^2].$$

Proof. By the definition of martingale in continuous time, N_k is a martingale in discrete time with respect to $\mathcal{G}_k$ when we set $N_k = M_{kt_0/2^n}$ and $\mathcal{G}_k = \mathcal{F}_{kt_0/2^n}$. By Theorem 5.4(b)

$$\mathbb{E}\,[\max_{0 \le k \le 2^n} M^2_{kt_0/2^n}] = \mathbb{E}\,[\max_{0 \le k \le 2^n} N_k^2] \le 4\mathbb{E}\, N^2_{2^n} = 4\mathbb{E}\, M^2_{t_0}.$$

(Recall $(\max_k a_k)^2 = \max a_k^2$ if all the $a_k \ge 0$.)

Now let $n \to \infty$. Since M_t has continuous paths, $\max_{0 \le k \le 2^n} M^2_{kt_0/2^n}$ increases up to $\sup_{s \le t_0} M_s^2$. Our result follows from the monotone convergence theorem from measure theory (see Note 4). $\square$

We now prove the analogue of Theorem 5.3. The proof is simpler if we assume that $\mathbb{E}\, M_t^2$ is finite; the result is still true without this assumption.

Theorem 10.3. *Suppose M_t is a martingale with continuous paths, $\mathbb{E}\, M_t^2 < \infty$ for all t, and τ is a stopping time bounded almost surely by t_0. Then $\mathbb{E}\, M_\tau = \mathbb{E}\, M_{t_0}$.*

Proof. We approximate τ by stopping times taking only finitely many values. For $n > 0$ define

$$\tau_n(\omega) = \inf\{kt_0/2^n : \tau(\omega) < kt_0/2^n\}.$$

τ_n takes only the values $kt_0/2^n$ for some $k \le 2^n$. The event $(\tau_n \le jt_0/2^n)$ is equal to $(\tau < jt_0/2^n)$, which is in $\mathcal{F}_{jt_0/2^n}$ since τ is a stopping time. So $(\tau_n \le s) \in \mathcal{F}_s$ if s is of the form $jt_0/2^n$ for some j. A moment's thought, using the fact that τ_n only takes values of the form $kt_0/2^n$, shows that τ_n is a stopping time.

It is clear that $\tau_n \downarrow \tau$ for every ω. Since M_t has continuous paths, $M_{\tau_n} \to M_\tau$ a.s.

Let N_k and $\mathcal{G}_k$ be as in the proof of Theorem 10.2. Let $\sigma_n = k$ if $\tau_n = kt_0/2^n$. By Theorem 5.3,

$$\mathbb{E}\, N_{\sigma_n} = \mathbb{E}\, N_{2^n},$$

which is the same as saying

$$\mathbb{E}\, M_{\tau_n} = \mathbb{E}\, M_{t_0}.$$

To complete the proof, we need to show $\mathbb{E}\, M_{\tau_n}$ converges to $\mathbb{E}\, M_\tau$. This is almost obvious, because we already observed that $M_{\tau_n} \to M_\tau$ a.s. Without the assumption that $\mathbb{E}\, M_t^2 < \infty$ for all t, this is actually quite a bit of work, but with the assumption it is not too bad.

Either $|M_{\tau_n} - M_\tau|$ is less than or equal to 1 or greater than 1. If it is greater than 1, it is less than $|M_{\tau_n} - M_\tau|^2$. So in either case,

$$|M_{\tau_n} - M_\tau| \le 1 + |M_{\tau_n} - M_\tau|^2. \tag{10.1}$$

Because both $|M_{\tau_n}|$ and $|M_\tau|$ are bounded by $\sup_{s \le t_0} |M_s|$, the right hand side of (10.1) is bounded by $1 + 4 \sup_{s \le t_0} |M_s|^2$, which is integrable by Theorem 10.2. $|M_{\tau_n} - M_\tau| \to 0$, and so by the dominated convergence theorem from measure theory (Note 4),

$$\mathbb{E}\, |M_{\tau_n} - M_\tau| \to 0.$$

Finally,

$$|\mathbb{E}\, M_{\tau_n} - \mathbb{E}\, M_\tau| = |\mathbb{E}\, (M_{\tau_n} - M_\tau)| \le \mathbb{E}\, |M_{\tau_n} - M_\tau| \to 0.$$

$\square$

Note 4. The dominated convergence theorem says that if $X_n \to X$ a.s. and $|X_n| \le Y$ a.s. for each n, where $\mathbb{E}\, Y < \infty$, then $\mathbb{E}\, X_n \to \mathbb{E}\, X$.

The monotone convergence theorem says that if $X_n \ge 0$ for each n, $X_n \le X_{n+1}$ for each n, and $X_n \to X$, then $\mathbb{E}\, X_n \to \mathbb{E}\, X$.

11. Brownian motion.

First, let us review a few facts about normal random variables. We say X is a normal random variable with mean a and variance b^2 if

$$\mathbb{P}(c \le X \le d) = \int_c^d \frac{1}{\sqrt{2\pi b^2}} e^{-(y-a)^2/2b^2}\, dy$$

and we will abbreviate this by saying X is $\mathcal{N}(a, b^2)$. If X is $\mathcal{N}(a, b^2)$, then $\mathbb{E}\, X = a$, $\operatorname{Var} X = b^2$, and $\mathbb{E}\, |X|^p < \infty$ is finite for every positive integer p. Moreover

$$\mathbb{E}\, e^{tX} = e^{at} e^{t^2 b^2/2}.$$

Let S_n be a simple symmetric random walk. This means that $Y_k = S_k - S_{k-1}$ equals $+1$ with probability $\frac{1}{2}$, equals -1 with probability $\frac{1}{2}$, and is independent of Y_j for $j < k$. We notice that $\mathbb{E}\, S_n = 0$ while $\mathbb{E}\, S_n^2 = \sum_{i=1}^n \mathbb{E}\, Y_i^2 + \sum_{i \neq j} \mathbb{E}\, Y_i Y_j = n$ using the fact that $\mathbb{E}\,[Y_i Y_j] = (\mathbb{E}\, Y_i)(\mathbb{E}\, Y_j) = 0$.

Define $X_t^n = S_{nt}/\sqrt{n}$ if nt is an integer and by linear interpolation for other t. If nt is an integer, $\mathbb{E}\, X_t^n = 0$ and $\mathbb{E}\,(X_t^n)^2 = t$. It turns out X_t^n does not converge for any ω.

However there is another kind of convergence, called weak convergence, that takes place. There exists a process Z_t such that for each k, each $t_1 < t_2 < \cdots < t_k$, and each $a_1 < b_1, a_2 < b_2, \ldots, a_k < b_k$, we have

(1) The paths of Z_t are continuous as a function of t.

(2) $\mathbb{P}(X_{t_1}^n \in [a_1, b_1], \ldots, X_{t_k}^n \in [a_k, b_k]) \to \mathbb{P}(Z_{t_1} \in [a_1, b_1], \ldots, Z_{t_k} \in [a_k, b_k])$.

See Note 1 for more discussion of weak convergence.

The limit Z_t is called a Brownian motion starting at 0. It has the following properties.

(1) $\mathbb{E}\, Z_t = 0$.

(2) $\mathbb{E}\, Z_t^2 = t$.

(3) $Z_t - Z_s$ is independent of $\mathcal{F}_s = \sigma(Z_r, r \leq s)$.

(4) $Z_t - Z_s$ has the distribution of a normal random variable with mean 0 and variance $t - s$. This means

$$\mathbb{P}(Z_t - Z_s \in [a, b]) = \int_a^b \frac{1}{\sqrt{2\pi(t-s)}} e^{-y^2/2(t-s)} dy.$$

(This result follows from the central limit theorem.)

(5) The map $t \to Z_t(\omega)$ is continuous for almost all ω.

See Note 2 for a few remarks on this definition.

It is common to use B_t ("B" for Brownian) or W_t ("W" for Wiener, who was the first person to prove rigorously that Brownian motion exists). We will most often use W_t.

We will use Brownian motion extensively and develop some of its properties. As one might imagine for a limit of a simple random walk, the paths of Brownian motion have a huge number of oscillations. It turns out that the function $t \to W_t(\omega)$ is continuous, but it is not differentiable; in fact one cannot define a derivative at any value of t. Another bizarre property: if one looks at the set of times at which $W_t(\omega)$ is equal to 0, this is a set which is uncountable, but contains no intervals. There is nothing special about 0 – the same is true for the set of times at which $W_t(\omega)$ is equal to a for any level a.

In what follows, one of the crucial properties of a Brownian motion is that it is a martingale with continuous paths. Let us prove this.

Proposition 11.1. *W_t is a martingale with respect to $\mathcal{F}_t$ and W_t has continuous paths.*

Proof. As part of the definition of a Brownian motion, W_t has continuous paths. W_t is $\mathcal{F}_t$ measurable by the definition of $\mathcal{F}_t$. Since the distribution of W_t is that of a normal random variable with mean 0 and variance t, then $\mathbb{E}\,|W_t| < \infty$ for all t. (In fact, $\mathbb{E}\,|W_t|^n < \infty$ for all n.)

The key property is to show $\mathbb{E}\,[W_t \mid \mathcal{F}_s] = W_s$.

$$\mathbb{E}\,[W_t \mid \mathcal{F}_s] = \mathbb{E}\,[W_t - W_s \mid \mathcal{F}_s] + \mathbb{E}\,[W_s \mid \mathcal{F}_s] = \mathbb{E}\,[W_t - W_s] + W_s = W_s.$$

We used here the facts that $W_t - W_s$ is independent of $\mathcal{F}_s$ and that $\mathbb{E}\,[W_t - W_s] = 0$ because W_t and W_s have mean 0. $\square$

We will also need

Proposition 11.2. *$W_t^2 - t$ is a martingale with continuous paths with respect to $\mathcal{F}_t$.*

Proof. That $W_t^2 - t$ is integrable and is $\mathcal{F}_t$ measurable is as in the above proof. We calculate

$$\begin{aligned}
\mathbb{E}\,[W_t^2 - t \mid \mathcal{F}_s] &= \mathbb{E}\,[((W_t - W_s) + W_s)^2 \mid \mathcal{F}_s] - t \\
&= \mathbb{E}\,[(W_t - W_s)^2 \mid \mathcal{F}_s] + 2\mathbb{E}\,[(W_t - W_s)W_s \mid \mathcal{F}_s] + \mathbb{E}\,[W_s^2 \mid \mathcal{F}_s] - t \\
&= \mathbb{E}\,[(W_t - W_s)^2] + 2W_s\mathbb{E}\,[W_t - W_s \mid \mathcal{F}_s] + W_s^2 - t.
\end{aligned}$$

We used the facts that W_s is $\mathcal{F}_s$ measurable and that $(W_t - W_s)^2$ is independent of $\mathcal{F}_s$ because $W_t - W_s$ is. The second term on the last line is equal to $W_s\mathbb{E}\,[W_t - W_s] = 0$. The first term, because $W_t - W_s$ is normal with mean 0 and variance $t - s$, is equal to $t - s$. Substituting, the last line is equal to

$$(t - s) + 0 + W_s^2 - t = W_s^2 - s$$

as required. $\square$

Note 1. A sequence of random variables X_n converges weakly to X if $\mathbb{P}(a < X_n < b) \to \mathbb{P}(a < X < b)$ for all $a, b \in [-\infty, \infty]$ such that $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 0$. a and b can be infinite. If X_n converges to a normal random variable, then $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 0$ for all a and b. This is the type of convergence that takes place in the central limit theorem. It will not be true in general that X_n converges to X almost surely.

For a sequence of random vectors $(X_1^n, \ldots, X_k^m)$ to converge to a random vector $(X_1, \ldots, X_k)$, one can give an analogous definition. But saying that the normalized random walks $X_n(t)$ above converge weakly to Z_t actually says more than (2). A result from probability theory says that X_n converges to X weakly if and only if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ whenever f is a bounded continuous function on $\mathbb{R}$. We use this to define weak convergence for stochastic processes. Let $C([0,\infty)$ be the collection of all continuous functions from $[0,\infty)$ to the reals. This is a metric space, so the notion of function from $C([0,\infty))$ to $\mathbb{R}$ being continuous makes sense. We say that the processes X_n converge weakly to the process Z, and mean by this that $\mathbb{E}[F(X_n)] \to \mathbb{E}[F(Z)]$ whenever F is a bounded continuous function on $C([0,\infty))$. One example of such a function F would be $F(f) = \sup_{0 \le t < \infty} |f(t)|$ if $f \in C([0,\infty))$; Another would be $F(f) = \int_0^1 f(t)dt$.

The reason one wants to show that X_n converges weakly to Z instead of just showing (2) is that weak convergence can be shown to imply that Z has continuous paths.

Note 2. First of all, there is some redundancy is the definition: one can show that parts of the definition are implied by the remaining parts, but we won't worry about this. Second, we actually want to let $\mathcal{F}_t$ to be the completion of $\sigma(Z_s : s \le t)$, that is, we throw in all the null sets into each $\mathcal{F}_t$. One can prove that the resulting $\mathcal{F}_t$ are right continuous, and hence the filtration $\mathcal{F}_t$ satisfies the "usual" conditions. Finally, the "almost all" in (5) means that $t \to Z_t(\omega)$ is continuous for all ω, except for a set of ω of probability zero.

12. Stochastic integrals.

If one wants to consider the (deterministic) integral $\int_0^t f(s)\,dg(s)$, where f and g are continuous and g is continuously differentiable, we can define it analogously to the usual Riemann integral as the limit of Riemann sums $\sum_{i=1}^n f(s_i)[g(s_i) - g(s_{i-1})]$, where $s_1 < s_2 < \cdots < s_n$ is a partition of $[0,t]$. This is known as the Riemann-Stieltjes integral. One can show (using the mean value theorem, for example) that

$$\int_0^t f(s)\,dg(s) = \int_0^t f(s)g'(s)\,ds.$$

If we were to take $f(s) = 1_{[0,a]}(s)$ (which is not continuous, but that is a minor matter here), one would expect the following:

$$\int_0^t 1_{[0,a]}(s)\,dg(s) = \int_0^t 1_{[0,a]}(s)g'(s)\,ds = \int_0^a g'(s)\,ds = g(a) - g(0).$$

Note that although we use the fact that g is differentiable in the intermediate stages, the first and last terms make sense for any g.

We now want to replace g by a Brownian path and f by a random integrand. The expression $\int f(s)\,dW(s)$ does not make sense as a Riemann-Stieltjes integral because it is a fact that $W(s)$ is not differentiable as a function of t. We need to define the expression by some other means. We will show that it can be defined as the limit in L^2 of Riemann sums. The resulting integral is called a stochastic integral.

Let us consider a very special case first. Suppose f is continuous and deterministic (i.e., does not depend on ω). Suppose we take a Riemann sum approximation

$$I_n = \sum_{i=0}^{2^n-1} f(\tfrac{i}{2^n})[W(\tfrac{i+1}{2^n}) - W(\tfrac{i}{2^n})].$$

Since W_t has zero expectation for each t, $\mathbb{E}\, I_n = 0$. Let us calculate the second moment:

$$\begin{aligned}\mathbb{E}\, I_n^2 &= \mathbb{E}\Big[\Big(\sum_i f(\tfrac{i}{2^n})[W(\tfrac{i+1}{2^n}) - W(\tfrac{i}{2^n})]\Big)^2\Big] \qquad (12.1)\\ &= \mathbb{E} \sum_{i=0}^{2^n-1} f(\tfrac{i}{2^n})^2[W(\tfrac{i+1}{2^n}) - W(\tfrac{i}{2^n})]^2\\ &\quad + \mathbb{E}\sum_{i\neq j} f(\tfrac{i}{2^n})f(\tfrac{j}{2^n})[W(\tfrac{i+1}{2^n}) - W(\tfrac{i}{2^n})]\,[W(\tfrac{j+1}{2^n}) - W(\tfrac{j}{2^n})].\end{aligned}$$

The first sum is bounded by

$$\sum_i f(\tfrac{i}{2^n})^2 \frac{1}{2^n} \approx \int_0^1 f(t)^2 dt,$$

since the second moment of $W(\frac{i+1}{2^n}) - W(\frac{i}{2^n})$ is $1/2^n$. Using the independence and the fact that W_t has mean zero,

$$\mathbb{E}\big[[W(\tfrac{i+1}{2^n} - W(\tfrac{i}{2^n})]\,[W(\tfrac{j+1}{2^n} - W(\tfrac{j}{2^n})]\big] = \mathbb{E}\,[W(\tfrac{i+1}{2^n} - W(\tfrac{i}{2^n})]\mathbb{E}\,[W(\tfrac{j+1}{2^n} - W(\tfrac{j}{2^n})] = 0,$$

and so the second sum on the right hand side of (12.1) is zero. This calculation is the key to the stochastic integral.

We now turn to the construction. Let W_t be a Brownian motion. We will only consider integrands H_s such that H_s is $\mathcal{F}_s$ measurable for each s (see Note 1). We will construct $\int_0^t H_s\, dW_s$ for all H with

$$\mathbb{E}\int_0^t H_s^2\, ds < \infty. \qquad (12.2)$$

Before we proceed we will need to define the quadratic variation of a continuous martingale. We will use the following theorem without proof because in our applications we can construct the desired increasing process directly. We often say a process is a continuous process if its paths are continuous, and similarly a continuous martingale is a martingale with continuous paths.

Theorem 12.1. *Suppose M_t is a continuous martingale such that $\mathbb{E}\, M_t^2 < \infty$ for all t. There exists one and only one increasing process A_t that is adapted to $\mathcal{F}_t$, has continuous paths, and $A_0 = 0$ such that $M_t^2 - A_t$ is a martingale.*

The simplest example of such a martingale is Brownian motion. If W_t is a Brownian motion, we saw in Proposition 11.2 that $W_t^2 - t$ is a martingale. So in this case $A_t = t$ almost surely, for all t. Hence $\langle W \rangle_t = t$.

We use the notation $\langle M \rangle_t$ for the increasing process given in Theorem 12.1 and call it the quadratic variation process of M. We will see later that in the case of stochastic integrals, where

$$N_t = \int_0^t H_s dW_s,$$

it turns out that $\langle N \rangle_t = \int_0^t H_s^2 ds$.

We will use the following frequently, and in fact, these are the only two properties of Brownian motion that play a significant role in the construction.

Lemma 12.1. *(a)* $\mathbb{E}\,[W_b - W_a \mid \mathcal{F}_a] = 0$.
(b) $\mathbb{E}\,[W_b^2 - W_a^2 \mid \mathcal{F}_a] = \mathbb{E}\,[(W_b - W_a)^2 \mid \mathcal{F}_a] = b - a$.

Proof. (a) This is $\mathbb{E}\,[W_b - W_a] = 0$ by the independence of $W_b - W_a$ from $\mathcal{F}_a$ and the fact that W_b and W_a have mean zero.

(b) $(W_b - W_a)^2$ is independent of $\mathcal{F}_a$, so the conditional expectation is the same as $\mathbb{E}\,[(W_b - W_a)^2]$. Since $W_b - W_a$ is a $\mathcal{N}(0, b-a)$, the second equality in (b) follows.

To prove the first equality in (b), we write

$$\begin{aligned}
\mathbb{E}\,[W_b^2 - W_a^2 \mid \mathcal{F}_a] &= \mathbb{E}\,[((W_b - W_a) + W_a)^2 \mid \mathcal{F}_a] - \mathbb{E}\,[W_a^2 \mid \mathcal{F}_a] \\
&= \mathbb{E}\,[(W_b - W_a)^2 \mid \mathcal{F}_a] + 2\mathbb{E}\,[W_a(W_b - W_a) \mid \mathcal{F}_a] + \mathbb{E}\,[W_a^2 \mid \mathcal{F}_a] \\
&\qquad - \mathbb{E}\,[W_a^2 \mid \mathcal{F}_a] \\
&= \mathbb{E}\,[(W_b - W_a)^2 \mid \mathcal{F}_a] + 2W_a\mathbb{E}\,[W_b - W_a \mid \mathcal{F}_a],
\end{aligned}$$

and the first equality follows by applying (a). $\square$

We construct the stochastic integral in three steps. We say an integrand $H_s = H_s(\omega)$ is elementary if

$$H_s(\omega) = G(\omega) 1_{(a,b]}(s)$$

where $0 \le a < b$ and G is bounded and $\mathcal{F}_a$ measurable. We say H is simple if it is a finite linear combination of elementary processes, that is,

$$H_s(\omega) = \sum_{i=1}^n G_i(\omega) 1_{(a_i, b_i]}(s). \tag{12.3}$$

We first construct the stochastic integral for H elementary; the work here is showing the stochastic integral is a martingale. We next construct the integral for H simple and here the difficulty is calculating the second moment. Finally we consider the case of general H.

First step. If G is bounded and $\mathcal{F}_a$ measurable, let $H_s(\omega) = G(\omega)1_{(a,b]}(s)$, and define the stochastic integral to be the process N_t, where $N_t = G(W_{t\wedge b} - W_{t\wedge a})$. Compare this to the first paragraph of this section, where we considered Riemann-Stieltjes integrals.

Proposition 12.2. *N_t is a continuous martingale, $\mathbb{E}\, N_\infty^2 = \mathbb{E}\,[G^2(b-a)]$ and*

$$\langle N\rangle_t = \int_0^t G^2 1_{[a,b]}(s)\, ds.$$

Proof. The continuity is clear. Let us look at $\mathbb{E}\,[N_t \mid \mathcal{F}_s]$. In the case $a < s < t < b$, this is equal to

$$\mathbb{E}\,[G(W_t - W_a) \mid \mathcal{F}_s] = G\mathbb{E}\,[(W_t - W_a) \mid \mathcal{F}_s] = G(W_s - W_a) = N_s.$$

In the case $s < a < t < b$, $\mathbb{E}\,[N_t \mid \mathcal{F}_s]$ is equal to

$$\mathbb{E}\,[G(W_t - W_a) \mid \mathcal{F}_s] = \mathbb{E}\,[G\mathbb{E}\,[W_t - W_a \mid \mathcal{F}_a] \mid \mathcal{F}_s] = 0 = N_s.$$

The other possibilities are $s < t < a < b$, $a < b < s < t$, $as < a < b < t$, and $a < s < b < t$; these are done similarly.

For $\mathbb{E}\, N_\infty^2$, we have using Lemma 12.1(b)

$$\mathbb{E}\, N_\infty^2 = \mathbb{E}\,[G^2\mathbb{E}\,[(W_b - W_a)^2 \mid \mathcal{F}_a]] = \mathbb{E}\,[G^2\mathbb{E}\,[W_b^2 - W_a^2 \mid \mathcal{F}_a]] = \mathbb{E}\,[G^2(b-a)].$$

For $\langle N\rangle_t$, we need to show

$$\begin{aligned}\mathbb{E}\,[G^2(W_{t\wedge b} - W_{t\wedge a})^2 - G^2(t\wedge b - t\wedge a) \mid \mathcal{F}_s]& \\ = G^2(W_{s\wedge b} - W_{s\wedge a})^2 - G^2(s\wedge b - s\wedge a).&\end{aligned}$$

We do this by checking all six cases for the relative locations of $a, b, s,$ and t; we do one of the cases in Note 2. □

Second step. Next suppose H_s is simple as in (12.3). In this case define the stochastic integral

$$N_t = \int_0^t H_s\, dW_s = \sum_{i=1}^n G_i(W_{b_i\wedge t} - W_{a_i\wedge t}).$$

Proposition 12.3. *N_t is a continuous martingale, $\mathbb{E}\, N_\infty^2 = \mathbb{E} \int_0^\infty H_s^2\, ds$, and $\langle N \rangle_t = \int_0^t H_s^2\, ds$.*

Proof. We may rewrite H so that the intervals $(a_i, b_i]$ satisfy $a_1 \le b_1 \le a_2 \le b_2 \le \cdots \le b_n$. For example, if we had $a_1 < a_2 < b_1 < b_2$, we could write

$$H_s = G_1 1_{(a_1,a_2]} + (G_1 + G_2) 1_{(a_2,b_1]} + G_2 1_{(b_1,b_2]},$$

and then if we set $G_1' = G_1, G_2' = G_1 + G_2, G_3' = G_2$ and $a_1' = a_1, b_1' = a_2, a_2' = a_2, b_2' = b_1, a_3' = b_1, b_3' = b_2$, we have written H as

$$\sum_{i=1}^3 G_i' 1_{(a_i', b_i']}.$$

So now we have H simple but with the intervals $(a_i', b_i']$ non-overlapping.

Since the sum of martingales is clearly a martingale, N_t is a martingale. The sum of continuous processes will be continuous, so N_t has continuous paths.

We have

$$\mathbb{E}\, N_\infty^2 = \mathbb{E}\Big[\sum G_i^2 (W_{b_i} - W_{a_i})^2\Big] + 2\mathbb{E}\Big[\sum_{i<j} G_i G_j (W_{b_i} - W_{a_i})(W_{b_j} - W_{a_j})\Big].$$

The terms in the second sum vanish, because when we condition on $\mathcal{F}_{a_j}$, we have

$$\mathbb{E}\,[G_i G_j (W_{b_i} - W_{a_i}) \mathbb{E}\,[(W_{b_j} - W_{a_j}) \mid \mathcal{F}_{a_j}] = 0.$$

Taking expectations,

$$\mathbb{E}\,[G_i G_j (W_{b_i} - W_{a_i}) \mathbb{E}\,[(W_{b_j} - W_{a_j})] = 0.$$

For the terms in the first sum, by Lemma 12.1

$$\mathbb{E}\,[G_i^2 (W_{b_i} - W_{a_i})^2] = \mathbb{E}\,[G_i^2 \mathbb{E}\,[(W_{b_i} - W_{a_i})^2 \mid \mathcal{F}_{a_i}]] = \mathbb{E}\,[G_i^2([b_i - a_i)].$$

So

$$\mathbb{E}\, N_\infty^2 = \sum_{i=1}^n E[G_i^2([b_i - a_i)],$$

and this is the same as $\mathbb{E} \int_0^\infty H_s^2 ds$. □

Third step. Now suppose H_s is adapted and $\mathbb{E} \int_0^\infty H_s^2\, ds < \infty$. Using some results from measure theory (Note 3), we can choose H_s^n simple such that $\mathbb{E} \int_0^\infty (H_s^n - H_s)^2\, ds \to 0$. The triangle inequality then implies (see Note 3 again)

$$\mathbb{E} \int_0^\infty (H_s^n - H_s^m)^2\, ds \to 0.$$

Define $N^n_t = \int_0^t H^n_s dW_s$ using Step 2. By Doob's inequality (Theorem 10.3) we have

$$\begin{aligned}\mathbb{E}\left[\sup_t (N^n_t - N^m_t)^2\right] &= \mathbb{E}\left[\sup_t \Big(\int_0^t (H^n_s - H^m_s)\, dW_s\Big)^2\right]\\ &\le 4\mathbb{E}\Big(\int_0^\infty (H^n_s - H^m_s)\, dW_s\Big)^2\\ &= 4\mathbb{E}\int_0^\infty (H^n_s - H^m_s)^2\, ds \to 0.\end{aligned}$$

This should look reminiscent of the definition of Cauchy sequences, and in fact that is what is going on here; Note 3 has details. In the present context Cauchy sequences converge, and one can show (Note 3) that there exists a process N_t such that

$$\mathbb{E}\left[\Big(\sup_t \Big|\int_0^t H^n_s\, dW_s - N_t\Big|\Big)^2\right] \to 0.$$

If H^n_s and $H^{n\prime}_s$ are two sequences converging to H, then $\mathbb{E}\,(\int_0^t (H^n_s - H^{n\prime}_s)\, dW_s)^2 = \mathbb{E}\,\int_0^t (H^n_s - H^{n\prime}_s)^2\, ds \to 0$, or the limit is independent of which sequence H^n we choose. See Note 4 for the proof that N_t is a martingale, $\mathbb{E}\, N^2_t = \mathbb{E}\,\int_0^t H^2_s\, ds$, and $\langle N\rangle_t = \int_0^t H^2_s\, ds$. Because $\sup_t[\int_0^t H^n_s\, dW_s - N_t] \to 0$, one can show there exists a subsequence such that the convergence takes place almost surely, and with probability one, N_t has continuous paths (Note 5).

We write $N_t = \int_0^t H_s\, dW_s$ and call N_t the stochastic integral of H with respect to W.

We discuss some extensions of the definition. First of all, if we replace W_t by a continuous martingale M_t and H_s is adapted with $\mathbb{E}\,\int_0^t H^2_s d\langle M\rangle_s < \infty$, we can duplicate everything we just did (see Note 6) with ds replaced by $d\langle M\rangle_s$ and get a stochastic integral. In particular, if $d\langle M\rangle_s = K^2_s ds$, we replace ds by $K^2_s ds$.

There are some other extensions of the definition that are not hard. If the random variable $\int_0^\infty H^2_s d\langle M\rangle_s$ is finite but without its expectation being finite, we can define the stochastic integral by defining it for $t \le T_N$ for suitable stopping times T_N and then letting $T_N \to \infty$; look at Note 7.

A process A_t is of bounded variation if the paths of A_t have bounded variation. This means that one can write $A_t = A^+_t - A^-_t$, where A^+_t and A^-_t have paths that are increasing. $|A|_t$ is then defined to be $A^+_t + A^-_t$. A semimartingale is the sum of a martingale and a process of bounded variation. If $\int_0^\infty H^2_s\, d\langle M\rangle_s + \int_0^\infty |H_s|\, |dA_s| < \infty$ and $X_t = M_t + A_t$, we define

$$\int_0^t H_s\, dX_s = \int_0^t H_s\, dM_s + \int_0^t H_s\, dA_s,$$

where the first integral on the right is a stochastic integral and the second is a Riemann-Stieltjes or Lebesgue-Stieltjes integral. For a semimartingale, we define $\langle X \rangle_t = \langle M_t \rangle$. Note 7 has more on this.

Given two semimartingales X and Y we define $\langle X, Y \rangle_t$ by what is known as polarization:

$$\langle X, Y \rangle_t = \tfrac{1}{2}[\langle X + Y \rangle_t - \langle X \rangle_t - \langle Y \rangle_t].$$

As an example, if $X_t = \int_0^t H_s dW_s$ and $Y_t = \int_0^t K_s dW_s$, then $(X+Y)_t = \int_0^t (H_s + K_s) dW_s$, so

$$\langle X + Y \rangle_t = \int_0^t (H_s + K_s)^2 ds = \int_0^t H_s^2 ds + \int_0^t 2H_s K_s ds + \int_0^t K_s^2 ds.$$

Since $\langle X \rangle_t = \int_0^t H_s^2 ds$ with a similar formula for $\langle Y \rangle_t$, we conclude

$$\langle X, Y \rangle_t = \int_0^t H_s K_s ds.$$

The following holds, which is what one would expect.

Proposition 12.4. *Suppose K_s is adapted to $\mathcal{F}_s$ and $\mathbb{E} \int_0^\infty K_s^2 ds < \infty$. Let $N_t = \int_0^t K_s dW_s$. Suppose H_s is adapted and $\mathbb{E} \int_0^\infty H_s^2 d\langle N \rangle_s < \infty$. Then $\mathbb{E} \int_0^\infty H_s^2 K_s^2 ds < \infty$ and*

$$\int_0^t H_s dN_s = \int_0^t H_s K_s dW_s.$$

The argument for the proof is given in Note 8.

What does a stochastic integral mean? If one thinks of the derivative of Z_t as being a white noise, then $\int_0^t H_s dZ_s$ is like a filter that increases or decreases the volume by a factor H_s.

For us, an interpretation is that Z_t represents a stock price. Then $\int_0^t H_s dZ_s$ represents our profit (or loss) if we hold H_s shares at time s. This can be seen most easily if $H_s = G1_{[a,b]}$. So we buy $G(\omega)$ shares at time a and sell them at time b. The stochastic integral represents our profit or loss.

Since we are in continuous time, we are allowed to buy and sell continuously and instantaneously. What we are not allowed to do is look into the future to make our decisions, which is where the H_s adapted condition comes in.

Note 1. Let us be more precise concerning the measurability of H that is needed. H is a stochastic process, so can be viewed as a map from $[0, \infty) \times \Omega$ to $\mathbb{R}$ by $H : (s, \omega) \to H_s(\omega)$. We define a σ-field $\mathcal{P}$ on $[0, \infty) \times \Omega$ as follows. Consider the collection of processes of the form $G(\omega) 1_{(a,b])}(s)$ where G is bounded and $\mathcal{F}_a$ measurable for some $a < b$. Define $\mathcal{P}$ to be the smallest σ-field with respect to which every process of this form is measurable. $\mathcal{P}$ is called the predictable or previsible σ-field, and if a process H is measurable with respect to $\mathcal{P}$, then the process is called predictable. What we require for our integrands H is that they be predictable processes.

If H_s has continuous paths, then approximating continuous functions by step functions shows that such an H can be approximated by linear combinations of processes of the form $G(\omega)1_{(a,b])}(s)$. So continuous processes are predictable. The majority of the integrands we will consider will be continuous.

If one is slightly more careful, one sees that processes whose paths are functions which are continuous from the left at each time point are also predictable. This gives an indication of where the name comes from. If H_s has paths which are left continuous, then $H_t = \lim_{n\to\infty} H_{t-\frac{1}{n}}$, and we can "predict" the value of H_t from the values at times that come before t. If H_t is only right continuous and a path has a jump at time t, this is not possible.

Note 2. Let us consider the case $a < s < t < b$; again similar arguments take care of the other five cases. We need to show

$$\mathbb{E}\,[G^2(W_t - W_a)^2 - G^2(t-a) \mid \mathcal{F}_s] = G^2(W_s - W_a)^2 - G^2(s-a). \tag{12.4}$$

The left hand side is equal to $G^2\mathbb{E}\,[(W_t - W_a)^2 - (t-a) \mid \mathcal{F}_s]$. We write this as

$$\begin{aligned}
&G^2\mathbb{E}\,[((W_t - W_s) + (W_s - W_a))^2 - (t-a) \mid \mathcal{F}_s]\\
&\quad = G^2\Big\{\mathbb{E}\,[(W_t - W_s)^2 \mid \mathcal{F}_s] + 2\mathbb{E}\,[(W_t - W_s)(W_s - W_a) \mid \mathcal{F}_s]\\
&\qquad\quad + \mathbb{E}\,[(W_s - W_a)^2 \mid \mathcal{F}_s] - (t-a)\Big\}\\
&\quad = G^2\Big\{\mathbb{E}\,[(W_t - W_s)^2] + 2(W_s - W_a)\mathbb{E}\,[W_t - W_s \mid \mathcal{F}_s] + (W_s - W_a)^2 - (t-a)\Big\}\\
&\quad = G^2\Big\{(t-s) + 0 + (W_s - W_a)^2 - (t-a)\Big\}.
\end{aligned}$$

The last expression is equal to the right hand side of (12.4).

Note 3. A definition from measure theory says that if μ is a measure, $\|f\|_2$, the L^2 norm of f with respect to the measure μ, is defined as $\Big(\int f(x)^2\mu(dx)\Big)^{1/2}$. The space L^2 is defined to be the set of functions f for which $\|f\|_2 < \infty$. (A technical proviso: one has to identify as equal functions which differ only on a set of measure 0.) If one defines a distance between two functions f and g by $d(f,g) = \|f-g\|_2$, this is a metric on the space L^2, and a theorem from measure theory says that L^2 is complete with respect to this metric. Another theorem from measure theory says that the collection of simple functions (functions of the form $\sum_{i=1}^n c_i 1_{A_i}$) is dense in L^2 with respect to the metric.

Let us define a norm on stochastic processes; this is essentially an L^2 norm. Define

$$\|N\| = \Big(\mathbb{E} \sup_{0\le t<\infty} N_t^2\Big)^{1/2}.$$

One can show that this is a norm, and hence that the triangle inequality holds. Moreover, the space of processes N such that $\|N\| < \infty$ is complete with respect to this norm. This means that if N^n is a Cauchy sequence, i.e., if given ε there exists n_0 such that $\|N^n - N^m\| < \varepsilon$ whenever $n, m \geq n_0$, then the Cauchy sequence converges, that is, there exists N with $\|N\| < \infty$ such that $\|N^n - N\| \to 0$.

We can define another norm on stochastic processes. Define

$$\|H\|_2 = \Big(\mathbb{E} \int_0^\infty H_s^2 ds\Big)^{1/2}.$$

This can be viewed as a standard L^2 norm, namely, the L^2 norm with respect to the measure μ defined on $\mathcal{P}$ by

$$\mu(A) = \mathbb{E} \int_0^\infty 1_A(s, \omega) ds.$$

Since the set of simple functions with respect to μ is dense in L^2, this says that if H is measurable with respect to $\mathcal{P}$, then there exist simple processes H_s^n that are also measurable with respect to $\mathcal{P}$ such that $\|H^n - H\|_2 \to 0$.

Note 4. We have $\|N^n - N\| \to 0$, where the norm here is the one described in Note 3. Each N^n is a stochastic integral of the type described in Step 2 of the construction, hence each N_t^n is a martingale. Let $s < t$ and $A \in \mathcal{F}_s$. Since $\mathbb{E}\,[N_t^n \mid \mathcal{F}_s] = N_s^n$, then

$$\mathbb{E}\,[N_t^n; A] = \mathbb{E}\,[N_s^n; A]. \tag{12.5}$$

By Cauchy-Schwarz,

$$\begin{aligned} |\mathbb{E}\,[N_t^n; A] - \mathbb{E}\,[N_t; A]| &\leq \mathbb{E}\,[\,|N_t^n - N_t|; A] \leq \Big(\mathbb{E}\,[(N_t^n - N_t)^2]\Big)^{1/2} \Big(\mathbb{E}\,[1_A^2]\Big)^{1/2} \\ &\leq \|N^n - N\| \to 0. \end{aligned} \tag{12.6}$$

We have a similar limit when t is replaced by s, so taking the limit in (12.5) yields

$$\mathbb{E}\,[N_t; A] = \mathbb{E}\,[N_s; A].$$

Since N_s is $\mathcal{F}_s$ measurable and has the same expectation over sets $A \in \mathcal{F}_s$ as N_t does, then by Proposition 4.3 $\mathbb{E}\,[N_t \mid \mathcal{F}_s] = N_s$, or N_t is a martingale.

Suppose $\|N^n - N\| \to 0$. Given $\varepsilon > 0$ there exists n_0 such that $\|N^n - N\| < \varepsilon$ if $n \geq n_0$. Take $\varepsilon = 1$ and choose n_0. By the triangle inequality,

$$\|N\| \leq \|N^n\| + \|N^n - N\| \leq \|N^n\| + 1 < \infty$$

since $\|N^n\|$ is finite for each n.

That $N_t^2 - \langle N \rangle_t$ is a martingale is similar to the proof that N_t is a martingale, but slightly more delicate. We leave the proof to the reader, but note that in place of (SEC.402) one writes

$$|\mathbb{E}\,[(N_t^n)^2; A] - \mathbb{E}\,[(N_t)^2; A]| \le \mathbb{E}\,[\,|(N_t^n)^2 - (N_t)^2|] \le \mathbb{E}\,[\,|N_t^n - N_t|\,|N_t^n + N_t|]. \qquad (12.7)$$

By Cauchy-Schwarz this is less than

$$\|N_t^n - N_t\|\ \|N_t^n + N_t\|.$$

since $\|N_t^n + N_t\| \le \|N_t^n\| + \|N_t\|$ is bounded independently of n, we see that the left hand side of (12.7) tends to 0.

Note 5. We have $\|N^n - N\| \to 0$, where the norm is described in Note 3. This means that

$$\mathbb{E}\,[\sup_t |N_t^n - N_t|^2] \to 0.$$

A result from measure theory implies that there exists a subsequence n_k such that

$$\sup_t |N_t^{n_k} - N_t|^2 \to 0, \qquad \text{a.s.}$$

So except for a set of ω's of probability 0, $N_t^{n_k}(\omega)$ converges to $N_t(\omega)$ uniformly. Each $N_t^{n_k}(\omega)$ is continuous by Step 2, and the uniform limit of continuous functions is continuous, therefore $N_t(\omega)$ is a continuous function of t. Incidentally, this is the primary reason we considered Doob's inequalities.

Note 6. If M_t is a continuous martingale, $\mathbb{E}\,[M_b - M_a \mid \mathcal{F}_a] = \mathbb{E}\,[M_b \mid \mathcal{F}_a] - M_a = M_a - M_a = 0$. This is the analogue of Lemma 12.1(a). To show the analogue of Lemma 12.1(b),

$$\begin{aligned}
\mathbb{E}\,[(M_b - M_a)^2 \mid \mathcal{F}_a] &= \mathbb{E}\,[M_b^2 \mid \mathcal{F}_a] - 2\mathbb{E}\,[M_b M_a \mid \mathcal{F}_a] + \mathbb{E}\,[M_a^2 \mid \mathcal{F}_a] \\
&= \mathbb{E}\,[M_b^2 \mid \mathcal{F}_a] - 2M_a\mathbb{E}\,[M_b \mid \mathcal{F}_a] + M_a^2 \\
&= \mathbb{E}\,[M_b^2 - \langle M \rangle_b \mid \mathcal{F}_a] + \mathbb{E}\,[\langle M \rangle_b - \langle M \rangle_a \mid \mathcal{F}_a] + M_a^2 - \langle M \rangle_a \\
&= \mathbb{E}\,[\langle M \rangle_b - \langle M \rangle_a \mid \mathcal{F}_a],
\end{aligned}$$

since $M_t^2 - \langle M \rangle_t$ is a martingale. That

$$\mathbb{E}\,[M_b^2 - M_a^2 \mid \mathcal{F}_a] = \mathbb{E}\,[\langle M \rangle_b - \langle M \rangle_a \mid \mathcal{F}_a]$$

is just a rewriting of

$$\mathbb{E}\,[M_b^2 - \langle M \rangle_b \mid \mathcal{F}_a] = M_a^2 - \langle M \rangle_a = \mathbb{E}\,[M_a^2 - \langle M \rangle_a \mid \mathcal{F}_a].$$

With these two properties in place of Lemma 12.1, replacing W_s by M_s and ds by $d\langle M\rangle_s$, the construction of the stochastic integral $\int_0^t H_s dM_s$ goes through exactly as above.

Note 7. If we let $T_K = \inf\{t > 0 : \int_0^t H_s^2 d\langle M\rangle_s \geq K\}$, the first time the integral is larger than or equal to K, and we let $H_s^K = H_s 1_{(s\leq T_K)}$, then $\int_0^\infty H_s^K d\langle M\rangle_s \leq K$ and there is no difficulty defining $N_t^K = \int_0^t H_s^K dM_s$ for every t. One can show that if $t \leq T_{K_1}$ and T_{K_2}, then $N_t^{K_1} = N_t^{K_2}$ a.s. If $\int_0^t H_s d\langle M\rangle_s$ is finite for every t, then $T_K \to \infty$ as $K \to \infty$. If we call the common value N_t, this allows one to define the stochastic integral N_t for each t in the case where the integral $\int_0^t H_s^2 d\langle M\rangle_s$ is finite for every t, even if the expectation of the integral is not.

We can do something similar is M_t is a martingale but where we do not have $\mathbb{E}\,\langle M\rangle_\infty < \infty$. Let $S_K = \inf\{t : |M_t| \geq K\}$, the first time $|M_t|$ is larger than or equal to K. If we let $M_t^K = M_{t\wedge S_K}$, where $t \wedge S_k = \min(t, S_K)$, then one can show M^K is a martingale bounded in absolute value by K. So we can define $J_t^K = \int_0^t H_s dM_t^K$ for every t, using the paragraph above to handle the wider class of H's, if necessary. Again, one can show that if $t \leq S_{K_1}$ and $t \leq S_{K_2}$, then the value of the stochastic integral will be the same no matter whether we use M^{K_1} or M^{K_2} as our martingale. We use the common value as a definition of the stochastic integral J_t. We have $S_K \to \infty$ as $K \to \infty$, so we have a definition of J_t for each t.

Note 8. We only outline how the proof goes. To show

$$\int_0^t H_s dN_s = \int_0^t H_s K_s dW_s, \tag{12.8}$$

one shows that (SEC.801) holds for H_s simple and then takes limits. To show this, it suffices to look at H_s elementary and use linearity. To show (12.8) for H_s elementary, first prove this in the case when K_s is elementary, use linearity to extend it to the case when K is simple, and then take limits to obtain it for arbitrary K. Thus one reduces the proof to showing (12.8) when both H and K are elementary. In this situation, one can explicitly write out both sides of the equation and see that they are equal.

13. Ito's formula.

Suppose W_t is a Brownian motion and $f : \mathbb{R} \to \mathbb{R}$ is a C^2 function, that is, f and its first two derivatives are continuous. Ito's formula, which is sometime known as the change of variables formula, says that

$$f(W_t) - f(W_0) = \int_0^t f'(W_s) dW_s + \tfrac{1}{2}\int_0^t f''(W_s) ds.$$

Compare this with the fundamental theorem of calculus:

$$f(t) - f(0) = \int_0^t f'(s) ds.$$

In Ito's formula we have a second order term to carry along.

The idea behind the proof is quite simple. By Taylor's theorem.

$$\begin{aligned} f(W_t) - f(W_0) &= \sum_{i=0}^{n-1} [f(W_{(i+1)t/n}) - f(W_{it/n})] \\ &\approx \sum_{i=1}^{n-1} f'(W_{it/n})(W_{(i+1)t/n} - W_{it/n}) \\ &\qquad + \tfrac{1}{2} \sum_{i=0}^{n-1} f''(W_{it/n})(W_{(i+1)t/n} - W_{it/n})^2. \end{aligned}$$

The first sum on the right is approximately the stochastic integral and the second is approximately the quadratic variation.

For a more general semimartingale $X_t = M_t + A_t$, Ito's formula reads

Theorem 13.1. *If $f \in C^2$, then*

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s + \tfrac{1}{2} \int_0^t f''(X_s) d\langle M \rangle_s.$$

Let us look at an example. Let W_t be Brownian motion, $X_t = \sigma W_t - \sigma^2 t/2$, and $f(x) = e^x$. Then $\langle X \rangle_t = \langle \sigma W \rangle_t = \sigma^2 t$, $f'(x) = f"(x) = e^x$, and

$$\begin{aligned} e^{\sigma W_t - \sigma^2 t/2} &= 1 + \int_0^t e^{X_s} \sigma dW_s - \tfrac{1}{2} \int_0^t e^{X_s} \tfrac{1}{2} \sigma^2 ds \\ &\qquad + \tfrac{1}{2} \int_0^t e^{X_s} \sigma^2 ds \\ &= 1 + \int_0^t e^{X_s} \sigma dW_s. \end{aligned} \tag{13.1}$$

This example will be revisited many times later on.

Let us give another example of the use of Ito's formula. Let $X_t = W_t$ and let $f(x) = x^k$. Then $f'(x) = kx^{k-1}$ and $f''(x) = k(k-1)x^{k-2}$. We then have

$$\begin{aligned} W_t^k &= W_0^k + \int_0^t k W_s^{k-1} dW_s + \tfrac{1}{2} \int_0^t k(k-1) W_s^{k-2} d\langle W \rangle_s \\ &= \int_0^t k W_s^{k-1} dW_s + \frac{k(k-1)}{2} \int_0^t W_s^{k-2} ds. \end{aligned}$$

When $k = 3$, this says $W_t^3 - 3 \int_0^t W_s \, ds$ is a stochastic integral with respect to a Brownian motion, and hence a martingale.

For a semimartingale $X_t = M_t + A_t$ we set $\langle X \rangle_t = \langle M \rangle_t$. Given two semimartingales X, Y, we define

$$\langle X, Y \rangle_t = \tfrac{1}{2}[\langle X + Y \rangle_t - \langle X \rangle_t - \langle Y \rangle_t].$$

The following is known as Ito's product formula. It may also be viewed as an integration by parts formula.

Proposition 13.2. *If X_t and Y_t are semimartingales,*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y \rangle_t.$$

Proof. Applying Ito's formula with $f(x) = x^2$ to $X_t + Y_t$, we obtain

$$(X_t + Y_t)^2 = (X_0 + Y_0)^2 + 2 \int_0^t (X_s + Y_s)(dX_s + dY_s) + \langle X + Y \rangle_t.$$

Applying Ito's formula with $f(x) = x^2$ to X and to Y, then

$$X_t^2 = X_0^2 + 2 \int_0^t X_s dX_s + \langle X \rangle_t$$

and

$$Y_t^2 = Y_0^2 + 2 \int_0^t Y_s dY_s + \langle Y \rangle_t.$$

Then some algebra and the fact that

$$X_t Y_t = \tfrac{1}{2}[(X_t + Y_t)^2 - X_t^2 - Y_t^2]$$

yields the formula. □

There is a multidimensional version of Ito's formula: if $X_t = (X_t^1, \ldots, X_t^d)$ is a vector, each component of which is a semimartingale, and $f \in C^2$, then

$$f(X_t) - f(X_0) = \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dX_s^i + \tfrac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x_i^2}(X_s) d\langle X^i, X^j \rangle_s.$$

The following application of Ito's formula, known as Lévy's theorem, is important.

Theorem 13.3. *Suppose M_t is a continuous martingale with $\langle M\rangle_t = t$. Then M_t is a Brownian motion.*

Before proving this, recall from undergraduate probability that the moment generating function of a r.v. X is defined by $m_X(a) = \mathbb{E}\, e^{aX}$ and that if two random variables have the same moment generating function, they have the same law. This is also true if we replace a by iu. In this case we have $\varphi_X(u) = \mathbb{E}\, e^{iuX}$ and φ_X is called the characteristic function of X. The reason for looking at the characteristic function is that φ_X always exists, whereas $m_X(a)$ might be infinite. The one special case we will need is that if X is a normal r.v. with mean 0 and variance t, then $\varphi_X(u) = e^{-u^2t/2}$. This follows from the formula for $m_X(a)$ with a replaced by iu (this can be justified rigorously).

Proof. We will prove that M_t is a $\mathcal{N}(0,t)$; for the remainder of the proof see Note 1.

We apply Ito's formula with $f(x) = e^{iux}$. Then

$$e^{iuM_t} = 1 + \int_0^t iue^{iuM_s}\,dM_s + \tfrac{1}{2}\int_0^t (-u^2)e^{iuM_s}\,d\langle M\rangle_s.$$

Taking expectations and using $\langle M\rangle_s = s$ and the fact that a stochastic integral is a martingale, hence has 0 expectation, we have

$$\mathbb{E}\, e^{iuM_t} = 1 - \frac{u^2}{2}\int_0^t e^{iuM_s}\,ds.$$

Let $J(t) = \mathbb{E}\, e^{iuM_t}$. The equation can be rewritten

$$J(t) = 1 - \frac{u^2}{2}\int_0^t J(s)ds.$$

So $J'(t) = -\frac{1}{2}u^2J(t)$ with $J(0) = 1$. The solution to this elementary ODE is $J(t) = e^{-u^2t/2}$. Since

$$\mathbb{E}\, e^{iuM_t} = e^{-u^2t/2},$$

then by our remarks above the law of M_t must be that of a $\mathcal{N}(0,t)$, which shows that M_t is a mean 0 variance t normal r.v. □

Note 1. If $A \in \mathcal{F}_s$ and we do the same argument with M_t replaced by $M_{s+t} - M_s$, we have

$$e^{iu(M_{s+t}-M_s)} = 1 + \int_0^t iue^{iu(M_{s+r}-M_s)}\,dM_r + \tfrac{1}{2}\int_0^t (-u^2)e^{iu(M_{s+r}-M_s)}\,d\langle M\rangle_r.$$

Multiply this by 1_A and take expectations. Since a stochastic integral is a martingale, the stochastic integral term again has expectation 0. If we let $K(t) = \mathbb{E}\,[e^{iu(M_{t+s}-M_t)}; A]$, we now arrive at $K'(t) = -\frac{1}{2}u^2K(t)$ with $K(0) = \mathbb{P}(A)$, so

$$K(t) = \mathbb{P}(A)e^{-u^2t/2}.$$

Therefore

$$\mathbb{E}\left[e^{iu(M_{t+s}-M_s)};A\right] = \mathbb{E}\,e^{iu(M_{t+s}-M_s)}\mathbb{P}(A). \tag{13.2}$$

If f is a nice function and $\widehat{f}$ is its Fourier transform, replace u in the above by $-u$, multiply by $\widehat{f}(u)$, and integrate over u. (To do the integral, we approximate the integral by a Riemann sum and then take limits.) We then have

$$\mathbb{E}\,[f(M_{s+t}-M_s);A] = \mathbb{E}\,[f((M_{s+t}-M_s)]\mathbb{P}(A).$$

By taking limits we have this for $f = 1_B$, so

$$\mathbb{P}(M_{s+t}-M_s \in B, A) = \mathbb{P}(M_{s+t}-M_s \in B)\mathbb{P}(A).$$

This implies that $M_{s+t}-M_s$ is independent of $\mathcal{F}_s$.

Note $\mathrm{Var}\,(M_t - M_s) = t - s$; take $A = \Omega$ in (13.2).

14. The Girsanov theorem.

Suppose $\mathbb{P}$ is a probability and

$$dX_t = dW_t + \mu(X_t)dt,$$

where W_t is a Brownian motion. This is short hand for

$$X_t = X_0 + W_t + \int_0^t \mu(X_s)ds. \tag{14.1}$$

Let

$$M_t = \exp\Big(-\int_0^t \mu(X_s)dW_s - \int_0^t \mu(X_s)^2 ds/2\Big). \tag{14.2}$$

Then as we have seen before, by Ito's formula, M_t is a martingale. This calculation is reviewed in Note 1. We also observe that $M_0 = 1$.

Now let us define a new probability by setting

$$\mathbb{Q}(A) = \mathbb{E}\,[M_t;A] \tag{14.3}$$

if $A \in \mathcal{F}_t$. We had better be sure this $\mathbb{Q}$ is well defined. If $A \in \mathcal{F}_s \subset \mathcal{F}_t$, then $\mathbb{E}\,[M_t;A] = \mathbb{E}\,[M_s;A]$ because M_t is a martingale. We also check that $\mathbb{Q}(\Omega) = \mathbb{E}\,[M_t;\Omega] = \mathbb{E}\,M_t$. This is equal to $\mathbb{E}\,M_0 = 1$, since M_t is a martingale.

What the Girsanov theorem says is

Theorem 14.1. *Under $\mathbb{Q}$, X_t is a Brownian motion.*

Under $\mathbb{P}$, W_t is a Brownian motion and X_t is not. Under $\mathbb{Q}$, the process W_t is no longer a Brownian motion.

In order for a process X_t to be a Brownian motion, we need at a minimum that X_t is mean zero and variance t. To define mean and variance, we need a probability. Therefore a process might be a Brownian motion with respect to one probability and not another. Most of the other parts of the definition of being a Brownian motion also depend on the probability.

Similarly, to be a martingale, we need conditional expectations, and the conditional expectation of a random variable depends on what probability is being used.

There is a more general version of the Girsanov theorem.

Theorem 14.2. *If X_t is a martingale under $\mathbb{P}$, then under $\mathbb{Q}$ the process $X_t - D_t$ is a martingale where*

$$D_t = \int_0^t \frac{1}{M_s} d\langle X, M\rangle_s.$$

$\langle X\rangle_t$ is the same under both $\mathbb{P}$ and $\mathbb{Q}$.

Let us see how Theorem 14.1 can be used. Let S_t be the stock price, and suppose

$$dS_t = \sigma S_t dW_t + mS_t dt.$$

(So in the above formulation, $\mu(x) = m$ for all x.) Define

$$M_t = e^{(-m/\sigma)(W_t) - (m^2/2\sigma^2)t}.$$

Then from (13.1) M_t is a martingale and

$$M_t = 1 + \int_0^t \Big(-\frac{m}{\sigma}\Big) M_s dW_s.$$

Let $X_t = W_t$. Then

$$\langle X, M\rangle_t = \int_0^t \Big(-\frac{m}{\sigma}\Big) M_s ds = -\int_0^t M_s \frac{m}{\sigma} ds.$$

Therefore

$$\int_0^t \frac{1}{M_s} d\langle X, M\rangle_s = -\int_0^t \frac{m}{\sigma} ds = -(m/\sigma)t.$$

Define $\mathbb{Q}$ by (14.3). By Theorem 14.2, under $\mathbb{Q}$ the process $\widetilde{W}_t = W_t + (m/\sigma)t$ is a martingale. Hence

$$dS_t = \sigma S_t(dW_t + (m/\sigma)dt) = \sigma S_t d\widetilde{W}_t,$$

or

$$S_t = S_0 + \int_0^t \sigma S_s d\widetilde{W}_s$$

is a martingale. So we have found a probability under which the asset price is a martingale. This means that $\mathbb{Q}$ is the risk-neutral probability, which we have been calling $\overline{\mathbb{P}}$.

Let us give another example of the use of the Girsanov theorem. Suppose $X_t = W_t + \mu t$, where μ is a constant. We want to compute the probability that X_t exceeds the level a by time t_0.

We first need the probability that a Brownian motion crosses a level a by time t_0. If $A_t = \sup_{s \le t} W_t$, (note we are not looking at $|W_t|$), we have

$$\mathbb{P}(A_t > a, c \le W_t \le d) = \int_c^d \varphi(t, a, x), \tag{14.4}$$

where

$$\varphi(t, a, x) = \begin{cases} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} & x \ge a \\ \frac{1}{\sqrt{2\pi t}} e^{-(2a-x)^2/2t} & x < a. \end{cases}$$

This is called the reflection principle, and the name is due to the derivation, given in Note 2. Sometimes one says

$$\mathbb{P}(W_t = x, A_t > a) = \mathbb{P}(W_t = 2a - x), \qquad x < a,$$

but this is not precise because W_t is a continuous random variable and both sides of the above equation are zero; (14.4) is the rigorous version of the reflection principle.

Now let W_t be a Brownian motion under $\mathbb{P}$. Let $d\mathbb{Q}/d\mathbb{P} = M_t = e^{\mu W_t - \mu^2 t/2}$. Let $Y_t = W_t - \mu t$. Theorem 14.1 says that under $\mathbb{Q}$, Y_t is a Brownian motion. We have $W_t = Y_t + \mu t$.

Let $A = (\sup_{s \le t_0} W_s \ge a)$. We want to calculate

$$\mathbb{P}(\sup_{s \le t_0} (W_s + \mu s) \ge a).$$

W_t is a Brownian motion under $\mathbb{P}$ while Y_t is a Brownian motion under $\mathbb{Q}$. So this probability is equal to

$$\mathbb{Q}(\sup_{s \le t_0} (Y_s + \mu s) \ge a).$$

This in turn is equal to

$$\mathbb{Q}(\sup_{s \le t_0} W_s \ge a) = \mathbb{Q}(A).$$

Now we use the expression for M_t:

$$\begin{aligned}\mathbb{Q}(A) &= \mathbb{E}_{\mathbb{P}}[e^{\mu W_{t_0}-\mu^2 t_0/2}; A]\\ &= \int_{-\infty}^{\infty} e^{\mu x-\mu^2 t_0/2}\mathbb{P}(\sup_{s\le t_0} W_s \ge a, W_{t_0}=x)dx\\ &= e^{-\mu^2 t_0/2}\Big[\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi t_0}} e^{\mu x} e^{-(2a-x)^2/2t_0}\,dx + \int_a^{\infty} \frac{1}{\sqrt{2\pi t_0}} e^{\mu x} e^{-x^2/2t_0}\,dx.\Big]\end{aligned}$$

Proof of Theorem 14.1. Using Ito's formula with $f(x)=e^x$,

$$M_t = 1 - \int_0^t \mu(X_r) M_r dW_r.$$

So

$$\langle W, M\rangle_t = -\int_0^t \mu(X_r) M_r\, dr.$$

Since $\mathbb{Q}(A)=\mathbb{E}_{\mathbb{P}}[M_t; A]$, it is not hard to see that

$$\mathbb{E}_{\mathbb{Q}}[W_t; A] = \mathbb{E}_{\mathbb{P}}[M_t W_t; A].$$

By Ito's product formula this is

$$\mathbb{E}_{\mathbb{P}}\Big[\int_0^t M_r dW_r; A\Big] + \mathbb{E}_{\mathbb{P}}\Big[\int_0^t W_r dM_r; A\Big] + \mathbb{E}_{\mathbb{P}}\Big[\langle W, M\rangle_t; A\Big].$$

Since $\int_0^t M_r dW_r$ and $\int_0^t W_r dM_r$ are stochastic integrals with respect to martingales, they are themselves martingales. Thus the above is equal to

$$\mathbb{E}_{\mathbb{P}}\Big[\int_0^s M_r dW_r; A\Big] + \mathbb{E}_{\mathbb{P}}\Big[\int_0^s W_r dM_r; A\Big] + \mathbb{E}_{\mathbb{P}}\Big[\langle W, M\rangle_t; A\Big].$$

Using the product formula again, this is

$$\mathbb{E}_{\mathbb{P}}[M_s W_s; A] + \mathbb{E}_{\mathbb{P}}[\langle W, M\rangle_t - \langle W, M\rangle_s; A] = \mathbb{E}_{\mathbb{Q}}[W_s; A] + \mathbb{E}_{\mathbb{P}}[\langle W, M\rangle_t - \langle W, M\rangle_s; A].$$

The last term on the right is equal to

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\Big[\int_s^t d\langle W, M\rangle_r; A\Big] &= \mathbb{E}_{\mathbb{P}}\Big[-\int_s^t M_r \mu(X_r) dr; A\Big] = \mathbb{E}_{\mathbb{P}}\Big[-\int_s^t \mathbb{E}_{\mathbb{P}}[M_t \mid \mathcal{F}_r]\mu(X_r) dr; A\Big]\\ &= \mathbb{E}_{\mathbb{P}}\Big[-\int_s^t M_t \mu(X_r) dr; A\Big] = \mathbb{E}_{\mathbb{Q}}\Big[-\int_s^t \mu(X_r) dr; A\Big]\\ &= -\mathbb{E}_{\mathbb{Q}}\Big[\int_0^t \mu(X_r)\, dr; A\Big] + \mathbb{E}_{\mathbb{Q}}\Big[\int_0^s \mu(X_r)\, dr; A\Big].\end{aligned}$$

Therefore

$$\mathbb{E}_{\mathbb{Q}}\Big[W_t + \int_0^t \mu(X_r)dr; A\Big] = \mathbb{E}_{\mathbb{Q}}\Big[W_s + \int_0^s \mu(X_r)dr; A\Big],$$

which shows X_t is a martingale with respect to $\mathbb{Q}$.

Similarly, $X_t^2 - t$ is a martingale with respect to $\mathbb{Q}$. By Lévy's theorem, X_t is a Brownian motion. □

In Note 3 we give a proof of Theorem 14.2 and in Note 4 we show how Theorem 14.1 is really a special case of Theorem 14.2.

Note 1. Let

$$Y_t = -\int_0^t \mu(X_s)dW_s - \tfrac{1}{2}\int_0^t [\mu(X_s)]^2 ds.$$

We apply Ito's formula with the function $f(x) = e^x$. Note the martingale part of Y_t is the stochastic integral term and the quadratic variation of Y is the quadratic variation of the martingale part, so

$$\langle Y\rangle_t = \int_0^t [-\mu(X_s)]^2 ds.$$

Then $f'(x) = e^x, f''(x) = e^x$, and hence

$$\begin{aligned} M_t = e^{Y_t} &= e^{Y_0} + \int_0^t e^{Y_s} dY_s + \tfrac{1}{2}\int_0^t e^{Y_s} d\langle Y\rangle_s \\ &= 1 + \int_0^t M_s(-\mu(X_s)dW_s - \tfrac{1}{2}\int_0^t [\mu(X_s)]^2 ds \\ &\quad + \tfrac{1}{2}\int_0^t M_s[-\mu(X_s)]^2 ds \\ &= 1 - \int_0^t M_s\mu(X_s)dW_s. \end{aligned}$$

Since stochastic integrals with respect to a Brownian motion are martingales, this completes the argument that M_t is a martingale.

Note 2. Let S_n be a simple random walk. This means that $X_1, X_2, \ldots,$ are independent and identically distributed random variables with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$; let $S_n = \sum_{i=1}^n X_i$. If you are playing a game where you toss a fair coin and win \$1 if it comes up heads and lose \$1 if it comes up tails, then S_n will be your fortune at time n. Let $A_n = \max_{0\le k\le n} S_k$. We will show the analogue of (14.4) for S_n, which is

$$\mathbb{P}(S_n = x, A_n \ge a) = \begin{cases} \mathbb{P}(S_n = x) & x \ge a \\ \mathbb{P}(S_n = 2a - x & x < a. \end{cases} \tag{14.5}$$

(14.4) can be derived from this using a weak convergence argument.

To establish (14.5), note that if $x \geq a$ and $S_n = x$, then automatically $A_n \geq a$, so the only case to consider if when $x < a$. Any path that crosses a but is at level x at time n has a corresponding path determined by reflecting across level a at the first time the Brownian motion hits a; the reflected path will end up at $a + (a - x) = 2a - x$. The probability on the left hand side of (14.5) is the number of paths that hit a and end up at x divided by the total number of paths. Since the number of paths that hit a and end up at x is equal to the number of paths that end up at $2a - x$, then the probability on the left is equal to the number of paths that end up at $2a - x$ divided by the total number of paths; this is $\mathbb{P}(S_n = 2a - x)$, which is the right hand side.

Note 3. To prove Theorem 14.2, we proceed as follows. Assume without loss of generality that $X_0 = 0$. Then if $A \in \mathcal{F}_s$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[X_t; A] &= \mathbb{E}_{\mathbb{P}}[M_t X_t; A] \\
&= \mathbb{E}_{\mathbb{P}}\Big[\int_0^t M_r \, dX_r; A\Big] + \mathbb{E}_{\mathbb{P}}\Big[\int_0^t X_r \, dM_r; A\Big] + \mathbb{E}_{\mathbb{P}}[\langle X, M\rangle_t; A] \\
&= \mathbb{E}_{\mathbb{P}}\Big[\int_0^s M_r \, dX_r; A\Big] + \mathbb{E}_{\mathbb{P}}\Big[\int_0^s X_r \, dM_r; A\Big] + \mathbb{E}_{\mathbb{P}}[\langle X, M\rangle_t; A] \\
&= \mathbb{E}_{\mathbb{Q}}[X_s; A] + \mathbb{E}_{\mathbb{Q}}[\langle X, M\rangle_t - \langle X, M\rangle_s; A].
\end{aligned}$$

Here we used the fact that stochastic integrals with respect to the martingales X and M are again martingales.

On the other hand,

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}[\langle X, M\rangle_t - \langle X, M\rangle_s; A] &= \mathbb{E}_{\mathbb{P}}\Big[\int_s^t d\langle X, M\rangle_r; A\Big] \\
&= \mathbb{E}_{\mathbb{P}}\Big[\int_s^t M_r \, dD_r; A\Big] \\
&= \mathbb{E}_{\mathbb{P}}\Big[\int_s^t \mathbb{E}_{\mathbb{P}}[M_t \mid \mathcal{F}_r] \, dD_r; A\Big] \\
&= \mathbb{E}_{\mathbb{P}}\Big[\int_s^t M_t \, dD_r; A\Big] \\
&= \mathbb{E}_{\mathbb{P}}[(D_t - D_s) M_t; A] \\
&= \mathbb{E}_{\mathbb{Q}}[D_t - D_s; A].
\end{aligned}$$

The proof of the quadratic variation assertion is similar. □

Note 4. Here is an argument showing how Theorem 14.1 can also be derived from Theorem 14.2.

From our formula for M we have $dM_t = -M_t\mu(X_t)dW_t$, and therefore $d\langle X, M\rangle_t = -M_t\mu(X_t)dt$. Hence by Theorem 14.2 we see that under $\mathbb{Q}$, X_t is a continuous martingale with $\langle X\rangle_t = t$. By Lévy's theorem, this means that X is a Brownian motion under $\mathbb{Q}$. $\square$

15. Stochastic differential equations.

Let W_t be a Brownian motion. We are interested in the existence and uniqueness for stochastic differential equations (SDEs) of the form

$$dX_t = \sigma(X_t)\,dW_t + b(X_t)\,dt, \qquad X_0 = 0. \tag{15.1}$$

This means X_t satisfies

$$X_t = x_0 + \int_0^t \sigma(X_s)\,dW_s + \int_0^t b(X_s)\,ds. \tag{15.2}$$

Here W_t is a Brownian motion, and (15.2) holds for almost every ω.

We have to make some assumptions on σ and b. We assume they are Lipschitz, which means:

$$|\sigma(x) - \sigma(y)| \le c|x-y|, \qquad |b(x) - b(y)| \le c|x-y|$$

for some constant c. We also suppose that σ and b grow at most linearly, which means:

$$|\sigma(x)| \le c(1+|x|), \qquad |b(x)| \le c(1+|x|).$$

Theorem 15.1. *There exists one and only one solution to (15.2).*

The idea of the proof is Picard iteration, which is how existence and uniqueness for ordinary differential equations is proved; see Note 1.

The intuition behind (15.1) is that X_t behaves locally like a multiple of Brownian motion plus a constant drift: locally $X_{t+h} - X_t \approx \sigma(W_{t+h} - W_t) + \mu((t+h)-t)$. However the constants σ and μ depend on the current value of X_t. When X_t is at different points, the coefficients vary, which is why they are written $\sigma(X_t)$ and $\mu(X_t)$. σ is sometimes called the diffusion coefficient and μ is sometimes called the drift coefficient.

The above theorem also works in higher dimensions. We want to solve

$$dX_t^i = \sum_{j=1}^d \sigma_{ij}(X_s)dW_s^j + b_i(X_s)ds, \qquad i = 1, \ldots, d.$$

This is an abbreviation for the equation

$$X_t^i = x_0^i + \int_0^t \sum_{j=1}^d \sigma_{ij}(X_s)dW_s^j + \int_0^t b_i(X_s)ds.$$

Here the initial value is $x_0 = (x_0^1, \dots, x_0^d)$, the solution process is $X_t = (X_t^1, \dots, X_t^d)$, and $W_t^1, \dots, W_t^d$ are d independent Brownian motions. If all of the σ_{ij} and b_i are Lipschitz and grow at most linearly, we have existence and uniqueness for the solution.

Suppose one wants to solve

$$dZ_t = aZ_t\, dW_t + bZ_t\, dt.$$

Note that this equation is linear in Z_t, and it turns out that linear equations are almost the only ones that have an explicit solution. In this case we can write down the explicit solution and then verify that it satisfies the SDE. The uniqueness result above (Theorem 15.1) shows that we have in fact found the solution.

Let

$$Z_t = Z_0 e^{aW_t - a^2 t/2 + bt}.$$

We will verify that this is correct by using Ito's formula. Let $X_t = aW_t - a^2t/2 + bt$. Then X_t is a semimartingale with martingale part aW_t and $\langle X\rangle_t = a^2 t$. $Z_t = e^{X_t}$. By Ito's formula with $f(x) = e^x$,

$$\begin{aligned} Z_t &= Z_0 + \int_0^t e^{X_s} dX_s + \tfrac{1}{2} \int_0^t e^{X_s} a^2 ds \\ &= Z_0 + \int_0^t aZ_s dW_s - \int_0^t \frac{a^2}{2} Z_s ds + \int_0^t b\, ds \\ &\quad + \tfrac{1}{2} \int_0^t a^2 Z_s ds \\ &= \int_0^t aZ_s dW_s + \int_0^t bZ_s ds. \end{aligned}$$

This is the integrated form of the equation we wanted to solve.

There is a connection between SDEs and partial differential equations. Let f be a C^2 function. If we apply Ito's formula,

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \tfrac{1}{2} \int_0^t f''(X_s) d\langle X\rangle_s.$$

From (15.2) we know $\langle X\rangle_t = \int_0^t \sigma(X_s)^2 ds$. If we substitute for dX_s and $d\langle X\rangle_s$, we obtain

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_s) dW_s + \int_0^t \mu(X_s) ds \\ &\quad + \tfrac{1}{2} \int_0^t f''(X_s)\sigma(X_s)^2 ds \\ &= f(X_0) + \int_0^t f'(X_s) dW_s + \int_0^t \mathcal{L}f(X_s) ds, \end{aligned}$$

where we write

$$\mathcal{L}f(x) = \tfrac{1}{2}\sigma(x)^2 f''(x) + \mu(x)f'(x).$$

$\mathcal{L}$ is an example of a differential operator. Since the stochastic integral with respect to a Brownian motion is a martingale, we see from the above that

$$f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)ds$$

is a martingale. This fact can be exploited to derive results about PDEs from SDEs and vice versa.

Note 1. Let us illustrate the uniqueness part, and for simplicity, assume b is identically 0.

Proof of uniqueness. If X and Y are two solutions,

$$X_t - Y_t = \int_0^t [\sigma(X_s) - \sigma(Y_s)]dW_s.$$

So

$$\mathbb{E}\,|X_t - Y_t|^2 = \mathbb{E}\int_0^t |\sigma(X_s) - \sigma(Y_s)|^2 ds \le c\int_0^t \mathbb{E}\,|X_s - Y_s|^2 ds,$$

using the Lipschitz hypothesis on σ. If we let $g(t) = \mathbb{E}\,|X_t - Y_t|^2$, we have

$$g(t) \le c\int_0^t g(s)\,ds.$$

Then

$$g(t) \le c\int_0^t \left[c\int_0^s g(r)\,dr\right] ds.$$

g is easily seen to be bounded on finite intervals, and iteration implies

$$g(t) \le At^n/n!$$

for each n, which implies g must be 0. □

16. Continuous time financial models.

The most common model by far in finance is one where the security price is based on a Brownian motion. One does not want to say the price is some multiple of Brownian motion for two reasons. First, of all, a Brownian motion can become negative, which doesn't make sense for stock prices. Second, if one invests \$1,000 in a stock selling for \$1 and it goes up to \$2, one has the same profit, namely, \$1,000, as if one invests \$1,000 in a stock selling for \$100 and it goes up to \$200. It is the proportional increase one wants.

Therefore one sets $\Delta S_t/S_t$ to be the quantity related to a Brownian motion. Different stocks have different volatilities σ (consider a high-tech stock versus a pharmaceutical). In addition, one expects a mean rate of return μ on ones investment that is positive (otherwise, why not just put the money in the bank?). In fact, one expects the mean rate of return to be higher than the risk-free interest rate r because one expects something in return for undertaking risk.

So the model that is used is to let the stock price be modeled by the SDE

$$dS_t/S_t = \sigma dW_t + \mu dt,$$

or what looks better,

$$dS_t = \sigma S_t dW_t + \mu S_t dt. \tag{16.1}$$

Fortunately this SDE is one of those that can be solved explicitly, and in fact we gave the solution in Section 15.

Proposition 16.1. *The solution to (16.1) is given by*

$$S_t = S_0 e^{\sigma W_t + (\mu - (\sigma^2/2)t)}. \tag{16.2}$$

Proof. Using Theorem 15.1 there will only be one solution, so we need to verify that S_t as given in (16.2) satisfies (16.1). We already did this, but it is important enough that we will do it again. Let us first assume $S_0 = 1$. Let $X_t = \sigma W_t + (\mu - (\sigma^2/2)t$, let $f(x) = e^x$, and apply Ito's formula. We obtain

$$\begin{aligned} S_t = e^{X_t} &= e^{X_0} + \int_0^t e^{X_s} dX_s + \tfrac{1}{2}\int_0^t e^{X_s} d\langle X\rangle_s \\ &= 1 + \int_0^t S_s \sigma dW_s + \int_0^t S_s(\mu - \tfrac{1}{2}\sigma^2) ds \\ &\qquad + \tfrac{1}{2}\int_0^t S_s \sigma^2 ds \\ &= 1 + \int_0^t S_s \sigma dW_s + \int_0^t S_s \mu ds, \end{aligned}$$

which is (16.1). If $S_0 \neq 0$, just multiply both sides by S_0. □

If one purchases Δ_0 shares (possibly a negative number) at time t_0, then changes the investment to Δ_1 shares at time t_1, then changes the investment to Δ_2 at time t_2, etc., then ones wealth at time t will be

$$X_{t_0} + \Delta_0(S_{t_1} - S_{t_0}) + \Delta_1(S_{t_2} - S_{t_1}) + \cdots + \Delta_i(S_{t_{i+1}} - S_{t_i}). \tag{16.3}$$

To see this, at time t_0 one has the original wealth X_{t_0}. One buys Δ_0 shares and the cost is $\Delta_0 S_{t_0}$. At time t_1 one sells the Δ_0 shares for the price of S_{t_1} per share, and so ones wealth is now $X_{t_0} + \Delta_0(S_{t_1} - S_{t_0})$. One now pays $\Delta_1 S_{t_1}$ for Δ_1 shares at time t_1 and continues. The right hand side of (16.3) is the same as

$$X_{t_0} + \int_0^t \Delta(s) dS_s,$$

where we have $t \geq t_{i+1}$ and $\Delta(s) = \Delta_i$ for $t_i \leq s < t_{i+1}$. In other words, our wealth is given by a stochastic integral with respect to the stock price. The requirement that the integrand of a stochastic integral be adapted is very natural: we cannot base the number of shares we own at time s on information that will not be available until the future.

The continuous time model of finance is then that the security price is given by (16.1) (often called geometric Brownian motion), that there are no transaction costs, but one can trade as many shares as one wants and vary the amount held in a continuous fashion. This clearly is not the way the market actually works, for example, stock prices are discrete, but this model has proved to be a very good one.

17. Markov properties of Brownian motion.

Let W_t be a Brownian motion. Because $W_{t+r} - W_t$ is independent of $\sigma(W_s : s \leq t)$, then knowing the path of W up to time s gives no help in predicting $W_{t+r} - W_t$. In particular, if we want to predict W_{t+r} and we know W_t, then knowing the path up to time t gives no additional advantage in predicting W_{t+r}. Phrased another way, this says that to predict the future, we only need to know where we are and not how we got there.

Let's try to give a more precise description of this property, which is known as the Markov property.

Fix r and let $Z_t = W_{t+r} - W_r$. Clearly the map $t \to Z_t$ is continuous since the same is true for W. Since $Z_t - Z_s = W_{t+r} - W_{s+r}$, then the distribution of $Z_t - Z_s$ is normal with mean zero and variance $(t+r) - (s+r)$. One can also check the other parts of the definition to show that Z_t is also a Brownian motion.

Recall that a stopping time in the continuous framework is a r.v. T taking values in $[0, \infty)$ such that $(T \leq t) \in \mathcal{F}_t$ for all t. To make a satisfactory theory, we need that the $\mathcal{F}_t$ be right continuous (see Section 10), but this is fairly technical and we will ignore it.

If T is a stopping time, $\mathcal{F}_T$ is the collection of events A such that $A \cap (T > t) \in \mathcal{F}_t$ for all t.

Let us try to provide some motivation for this definition of $\mathcal{F}_T$. It will be simpler to consider the discrete time case. The analogue of $\mathcal{F}_T$ in the discrete case is the following: if N is a stopping time, let

$$\mathcal{F}_N = \{A : A \cap (N \leq k) \in \mathcal{F}_k \text{ for all } k\}.$$

If X_k is a sequence that is adapted to the σ-fields $\mathcal{F}_k$, that is, X_k is $\mathcal{F}_k$ measurable when $k = 0, 1, 2, \ldots$, then knowing which events in $\mathcal{F}_k$ have occurred allows us to calculate X_k for each k. So a reasonable definition of $\mathcal{F}_N$ should allow us to calculate X_N whenever we know which events in $\mathcal{F}_N$ have occurred or not. Or phrased another way, we want X_N to be $\mathcal{F}_N$ measurable. Where did the sequence X_k come from? It could be any adapted sequence. Therefore one definition of the σ-field of events occurring before time N might be:

> Consider the collection of random variables X_N where X_k is a sequence adapted to $\mathcal{F}_k$. Let $\mathcal{G}_N$ be the smallest σ-field with respect to which each of these random variables X_N is measurable.

In other words, we want $\mathcal{G}_N$ to be the σ-field generated by the collection of random variables X_N for all sequences X_k that are adapted to $\mathcal{F}_k$.

We show in Note 1 that $\mathcal{F}_N = \mathcal{G}_N$. The σ-field $\mathcal{F}_N$ is just a bit easier to work with.

Now we proceed to the strong Markov property for Brownian motion, the proof of which is given in Note 2.

Proposition 17.1. *If X_t is a Brownian motion and T is a bounded stopping time, then $X_{T+t} - X_T$ is a mean 0 variance t random variable and is independent of $\mathcal{F}_T$.*

This proposition says: if you want to predict X_{T+t}, you could do it knowing all of $\mathcal{F}_T$ or just knowing X_T. Since $X_{T+t} - X_T$ is independent of $\mathcal{F}_T$, the extra information given in $\mathcal{F}_T$ does you no good at all.

We need a way of expressing the Markov and strong Markov properties that will generalize to other processes.

Let W_t be a Brownian motion. Consider the process $W_t^x = x + W_t$, which is known as Brownian motion started at x. Define Ω' to be set of continuous functions on $[0, \infty)$, let $X_t(\omega) = \omega(t)$, and let the σ-field be the one generated by the X_t. Define $\mathbb{P}^x$ on $(\Omega', \mathcal{F}')$ by

$$\mathbb{P}^x(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n) = \mathbb{P}(W_{t_1}^x \in A_1, \ldots, W_{t_n}^x \in A_n).$$

What we have done is gone from one probability space Ω with many processes W_t^x to one process X_t with many probability measures $\mathbb{P}^x$.

An example in the Markov chain setting might help. No knowledge of Markov chains is necessary to understand this. Suppose we have a Markov chain with 3 states, A, B, and C. Suppose we have a probability $\mathbb{P}$ and three different Markov chains. The first, called X_n^A, represents the position at time n for the chain started at A. So $X_0^A = A$, and X_1^A can be one of A, B, C, X_2^A can be one of A, B, C, and so on. Similarly we have X_n^B, the chain started at B, and X_n^C. Define $\Omega' = \{(AAA), (AAB), (ABA), \ldots, (BAA), (BAB), \ldots\}$.

So Ω' denotes the possible sequence of states for time $n = 0, 1, 2$. If $\omega = ABA$, set $Y_0(\omega) = A, Y_1(\omega) = B, Y_2(\omega) = A$, and similarly for all the other 26 values of ω. Define $\mathbb{P}^A(AAA) = \mathbb{P}(X_0^A = A, X_1^A = A, X_2^A = A)$. Similarly define $\mathbb{P}^A(AAB), \ldots$. Define $\mathbb{P}^B(AAA) = \mathbb{P}(X_0^B = A, X_1^B = A, X_2^B = A)$ (this will be 0 because we know $X_0^B = B$), and similarly for the other values of ω. We also define $\mathbb{P}^C$. So we now have one process, Y_n, and three probabilities $\mathbb{P}^A, \mathbb{P}^B, \mathbb{P}^C$. As you can see, there really isn't all that much going on here.

Here is another formulation of the Markov property.

Proposition 17.2. *If $s < t$ and f is bounded or nonnegative, then*

$$\mathbb{E}^x[f(X_t) \mid \mathcal{F}_s] = \mathbb{E}^{X_s}[f(X_{t-s})], \qquad \text{a.s.}$$

The right hand side is to be interpreted as follows. Define $\varphi(x) = \mathbb{E}^x f(X_{t-s})$. Then $\mathbb{E}^{X_s} f(X_{t-s})$ means $\varphi(X_s(\omega))$. One often writes $P_t f(x)$ for $\mathbb{E}^x f(X_t)$. We prove this in Note 3.

This formula generalizes: If $s < t < u$, then

$$\mathbb{E}^x[f(X_t)g(X_u) \mid \mathcal{F}_s] = \mathbb{E}^{X_s}[f(X_{t-s})g(X_{u-s})],$$

and so on for functions of X at more times.

Using Proposition 17.1, the statement and proof of Proposition 17.2 can be extended to stopping times.

Proposition 17.3. *If T is a bounded stopping time, then*

$$\mathbb{E}^x[f(X_{T+t}) \mid \mathcal{F}_T] = \mathbb{E}^{X_T}[f(X_t)].$$

We can also establish the Markov property and strong Markov property in the context of solutions of stochastic differential equations. If we let X_t^x denote the solution to

$$X_t^x = x + \int_0^t \sigma(X_s^x) dW_s + \int_0^t b(X_s^x) ds,$$

so that X_t^x is the solution of the SDE started at x, we can define new probabilities by

$$\mathbb{P}^x(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n) = \mathbb{P}(X_{t_1}^x \in A_1, \ldots, X_{t_n}^x \in A_n).$$

This is similar to what we did in defining $\mathbb{P}^x$ for Brownian motion, but here we do not have translation invariance. One can show that when there is uniqueness for the solution to the SDE, the family $(\mathbb{P}^x, X_t)$ satisfies the Markov and strong Markov property. The statement is precisely the same as the statement of Proposition 17.3.

Note 1. We want to show $\mathcal{G}_N = \mathcal{F}_N$. Since $\mathcal{G}_N$ is the smallest σ-field with respect to which X_N is measurable for all adapted sequences X_k and it is easy to see that $\mathcal{F}_N$ is a σ-field, to show $\mathcal{G}_N \subset \mathcal{F}_N$, it suffices to show that X_N is measurable with respect to $\mathcal{F}_N$ whenever X_k is adapted. Therefore we need to show that for such a sequence X_k and any real number a, the event $(X_N > a) \in \mathcal{F}_N$.

Now $(X_N > a) \cap (N = j) = (X_j > a) \cap (N = j)$. The event $(X_j > a) \in \mathcal{F}_j$ since X is an adapted sequence. Since N is a stopping time, then $(N \le j) \in \mathcal{F}_j$ and $(N \le j-1)^c \in \mathcal{F}_{j-1} \subset \mathcal{F}_j$, and so the event $(N = j) = (N \le j) \cap (N \le j-1)^c$ is in $\mathcal{F}_j$. If $j \le k$, then $(N = j) \in \mathcal{F}_j \subset \mathcal{F}_k$. Therefore

$$(X_N > a) \cap (N \le k) = \cup_{j=0}^{k}((X_N > a) \cap (N = j)) \in \mathcal{F}_k,$$

which proves that $(X_N > a) \in \mathcal{F}_N$.

To show $\mathcal{F}_N \subset \mathcal{G}_N$, we suppose that $A \in \mathcal{F}_N$. Let $X_k = 1_{A\cap(N\le k)}$. Since $A \in \mathcal{F}_N$, then $A \cap (N \le k) \in \mathcal{F}_k$, so X_k is $\mathcal{F}_k$ measurable. But $X_N = 1_{A\cap(N\le N)} = 1_A$, so $A = (X_N > 0) \in \mathcal{G}_N$. We have thus shown that $\mathcal{F}_N \subset \mathcal{G}_N$, and combining with the previous paragraph, we conclude $\mathcal{F}_N = \mathcal{G}_N$. $\square$

Note 2. Let T_n be defined by $T_n(\omega) = (k+1)/2^n$ if $T(\omega) \in [k/2^n, (k+1)/2^n)$. It is easy to check that T_n is a stopping time. Let f be continuous and $A \in \mathcal{F}_T$. Then $A \in \mathcal{F}_{T_n}$ as well. We have

$$\begin{aligned}\mathbb{E}\,[f(X_{T_n+t} - X_{T_n}); A] &= \sum \mathbb{E}\,[f(X_{\frac{k}{2^n}+t} - X_{\frac{k}{2^n}}); A \cap T_n = k/2^n] \\ &= \sum \mathbb{E}\,[f(X_{\frac{k}{2^n}+t} - X_{\frac{k}{2^n}})]\mathbb{P}(A \cap T_n = k/2^n) \\ &= \mathbb{E}\, f(X_t)\mathbb{P}(A).\end{aligned}$$

Let $n \to \infty$, so

$$\mathbb{E}\,[f(X_{T+t} - X_T); A] = \mathbb{E}\, f(X_t)\mathbb{P}(A).$$

Taking limits this equation holds for all bounded f.

If we take $A = \Omega$ and $f = 1_B$, we see that $X_{T+t} - X_T$ has the same distribution as X_t, which is that of a mean 0 variance t normal random variable. If we let $A \in \mathcal{F}_T$ be arbitrary and $f = 1_B$, we see that

$$\mathbb{P}(X_{T+t} - X_T \in B, A) = \mathbb{P}(X_t \in B)\mathbb{P}(A) = \mathbb{P}(X_{T+t} - X_T \in B)\mathbb{P}(A),$$

which implies that $X_{T+t} - X_T$ is independent of $\mathcal{F}_T$. $\square$

Note 3. Before proving Proposition 17.2, recall from undergraduate analysis that every bounded function is the limit of linear combinations of functions e^{iux}, $u \in \mathbb{R}$. This follows from using the inversion formula for Fourier transforms. There are various slightly different formulas for the Fourier transform. We use $\widehat{f}(u) = \int e^{iux} f(x)\,dx$. If f is smooth enough and has compact support, then one can recover f by the formula

$$f(x) = \frac{1}{2\pi} \int e^{-iux} \widehat{f}(u)\,du.$$

We can first approximate this improper integral by

$$\frac{1}{2\pi} \int_{-N}^{N} e^{-iux} \widehat{f}(u)\,du$$

by taking N larger and larger. For each N we can approximate $\frac{1}{2\pi}\int_{-N}^{N} e^{-iux}\widehat{f}(u)\,du$ by using Riemann sums. Thus we can approximate $f(x)$ by a linear combination of terms of the form $e^{iu_j x}$. Finally, bounded functions can be approximated by smooth functions with compact support.

Proof. Let $f(x) = e^{iux}$. Then

$$\begin{aligned}\mathbb{E}^{x}[e^{iuX_t} \mid \mathcal{F}_s] &= e^{iuX_s}\mathbb{E}^{x}[e^{iu(X_t - X_s)} \mid \mathcal{F}_s] \\ &= e^{iuX_s} e^{-u^2(t-s)/2}.\end{aligned}$$

On the other hand,

$$\varphi(y) = \mathbb{E}^{y}[f(X_{t-s})] = \mathbb{E}\,[e^{iu(W_{t-s}+y)}] = e^{iuy} e^{-u^2(t-s)/2}.$$

So $\varphi(X_s) = \mathbb{E}^{x}[e^{iuX_t} \mid \mathcal{F}_s]$. Using linearity and taking limits, we have the lemma for all f. $\square$

18. Martingale representation theorem.

In this section we want to show that every random variable that is $\mathcal{F}_t$ measurable can be written as a stochastic integral of Brownian motion. In the next section we use this to show that under the model of geometric Brownian motion the market is complete. This means that no matter what option one comes up with, one can exactly replicate the result (no matter what the market does) by buying and selling shares of stock.

In mathematical terms, we let $\mathcal{F}_t$ be the σ-field generated by $W_s, s \le t$. From (16.2) we see that $\mathcal{F}_t$ is also the same as the σ-field generated by $S_s, s \le t$, so it doesn't matter which one we work with. We want to show that if V is $\mathcal{F}_t$ measurable, then there exists H_s adapted such that

$$V = V_0 + \int H_s dW_s, \tag{18.1}$$

where V_0 is a constant.

Our goal is to prove

Theorem 18.1. *If V is $\mathcal{F}_t$ measurable and $\mathbb{E}\,V^2 < \infty$, then there exists a constant c and an adapted integrand H_s with $\mathbb{E}\int_0^t H_s^2 ds < \infty$ such that*

$$V = c + \int_0^t H_s dW_s.$$

Before we prove this, let us explain why this is called a martingale representation theorem. Suppose M_s is a martingale adapted to $\mathcal{F}_s$, where the $\mathcal{F}_s$ are the σ-field generated by a Brownian motion. Suppose also that $\mathbb{E}\,M_t^2 < \infty$. Set $V = M_t$. By Theorem 18.1, we can write

$$M_t = V = c + \int_0^t H_s dW_s.$$

The stochastic integral is a martingale, so for $r \le t$,

$$M_r = \mathbb{E}\,[M_t \mid \mathcal{F}_r] = c + \mathbb{E}\Big[\int_0^t H_s dW_s \mid \mathcal{F}_r\Big] = c + \int_0^r H_s dW_s.$$

We already knew that stochastic integrals were martingales; what this says is the converse: every martingale can be represented as a stochastic integral. Don't forget that we need $\mathbb{E}\,M_t^2 < \infty$ and M_s adapted to the σ-fields of a Brownian motion.

In Note 1 we show that if every martingale can be represented as a stochastic integral, then every random variable V that is $\mathcal{F}_t$ measurable can, too, provided $\mathbb{E}\,V^2 < \infty$.

There are several proofs of Theorem 18.1. Unfortunately, they are all technical. We outline one proof here, giving details in the notes. We start with the following, proved in Note 2.

Proposition 18.2. *Suppose*

$$V^n = c_n + \int_0^t H_s^n dW_s,$$

$c_n \to c$,

$$\mathbb{E}\,|V^n - V|^2 \to 0,$$

and for each n the process H^n is adapted with $\mathbb{E}\int_0^t (H_s^n)^2 ds < \infty$. Then there exist a constant c and an adapted H_s with $\mathbb{E}\int_0^t H_s^2 ds < \infty$ so that

$$V_t = c + \int_0^t H_s dW_s.$$

What this proposition says is that if we can represent a sequence of random variables V_n and $V_n \to V$, then we can represent V.

Let $\mathcal{R}$ be the collection of random variables that can be represented as stochastic integrals. By this we mean

$$\mathcal{R} = \{V : \mathbb{E}\, V^2 < \infty, V \text{ is } \mathcal{F}_t \text{ measurable}, V = c + \int_0^t H_s dW_s$$
$$\text{for some adapted } H \text{ with } \mathbb{E} \int_0^t H_s^2 ds < \infty\}.$$

Next we show $\mathcal{R}$ contains a particular collection of random variables. (The proof is in Note 3.)

Proposition 18.3. *If g is bounded, the random variable $g(W_t)$ is in $\mathcal{R}$.*

An almost identical proof shows that if f is bounded, then

$$f(W_t - W_s) = c + \int_s^t H_r dW_r$$

for some c and H_r.

Proposition 18.4. *If $t_0 \le t_1 \le \cdots \le t_n \le t$ and $f_1, \ldots, f_n$ are bounded functions, then $f_1(W_{t_1} - W_{t_0}) f_2(W_{t_2} - W_{t_1}) \cdots f_n(W_{t_n} - W_{t_{n-1}})$ is in $\mathcal{R}$.*

See Note 4 for the proof.

We now finish the proof of Theorem 18.1. We have shown that a large class of random variables is contained in $\mathcal{R}$.

Proof of Theorem 18.1. We have shown that random variables of the form

$$f_1(W_{t_1} - W_{t_0}) f_2(W_{t_2} - W_{t_1}) \cdots f_n(W_{t_n} - W_{t_{n-1}}) \tag{18.2}$$

are in $\mathcal{R}$. Clearly if $V_i \in \mathcal{R}$ for $i = 1, \ldots, m$, and a_i are constants, then $a_1 V_1 + \cdots a_m V_m$ is also in $\mathcal{R}$. Finally, from measure theory we know that if $\mathbb{E}\, V^2 < \infty$ and V is $\mathcal{F}_t$ measurable, we can find a sequence V_k such that $\mathbb{E}\, |V_k - V|^2 \to 0$ and each V_k is a linear combination of random variables of the form given in (18.2). Now apply Proposition 18.2. □

Note 1. Suppose we know that every martingale M_s adapted to $\mathcal{F}_s$ with $\mathbb{E}\, M_t^2$ can be represented as $M_r = c + \int_0^r H_s dW_s$ for some suitable H. If V is $\mathcal{F}_t$ measurable with $\mathbb{E}\, V^2 < \infty$, let $M_r = \mathbb{E}\,[V \mid \mathcal{F}_r]$. We know this is a martingale, so

$$M_r = c + \int_0^r H_s dW_s$$

for suitable H. Applying this with $r = t$,

$$V = \mathbb{E}\,[V \mid \mathcal{F}_t] = M_t = c + \int_0^t H_s dW_s.$$

Note 2. We prove Proposition 18.2. By our assumptions,

$$\mathbb{E}\,|(V^n - c_n) - (V^m - c_m)|^2 \to 0$$

as $n, m \to \infty$. So

$$\mathbb{E}\,\Big|\int_0^t (H_s^n - H_s^m)dW_s\Big|^2 \to 0.$$

From our formulas for stochastic integrals, this means

$$\mathbb{E}\int_0^t |H_s^n - H_s^m|^2 ds \to 0.$$

This says that H_s^n is a Cauchy sequence in the space L^2 (with respect to the norm $\|\cdot\|_2$ given by $\|Y\|_2 = \Big(\mathbb{E}\int_0^t Y_s^2 ds\Big)^{1/2}$). Measure theory tells us that L^2 is a complete metric space, so there exists H_s such that

$$\mathbb{E}\int_0^t |H_s^n - H_s|^2 ds \to 0.$$

In particular $H_s^n \to H_s$, and this implies H_s is adapted. Another consequence, due to Fatou's lemma, is that $\mathbb{E}\int_0^t H_s^2 ds$.

Let $U_t = \int_0^t H_s dW_s$. Then as above,

$$\mathbb{E}\,|(V^n - c_n) - U_t|^2 = \mathbb{E}\int_0^t (H_s^n - H_s)^2 ds \to 0.$$

Therefore $U_t = V - c$, and U has the desired form. $\square$

Note 3. Here is the proof of Proposition 18.3. By Ito's formula with $X_s = -iuW_s + u^2 s/2$ and $f(x) = e^x$,

$$\begin{aligned} e^{X_t} &= 1 + \int_0^t e^{X_s}(-iu)dW_s + \int_0^t e^{X_s}(u^2/2)ds \\ &\qquad + \tfrac{1}{2}\int_0^t e^{X_s}(-iu)^2 ds \\ &= 1 - iu\int_0^t e^{X_s}dW_s. \end{aligned}$$

If we multiply both sides by $e^{-u^2 t/2}$, which is a constant and hence adapted, we obtain

$$e^{-iuW_t} = c_u + \int_0^t H_s^u dW_s \tag{18.3}$$

for an appropriate constant c_u and integrand H^u.

If f is a smooth function (e.g., C^∞ with compact support), then its Fourier transform $\widehat{f}$ will also be very nice. So if we multiply (18.3) by $\widehat{f}(u)$ and integrate over u from $-\infty$ to ∞, we obtain

$$f(W_t) = c + \int_0^t H_s dW_s$$

for some constant c and some adapted integrand H. (We implicitly used Proposition 18.2, because we approximate our integral by Riemann sums, and then take a limit.) Now using Proposition 18.2 we take limits and obtain the proposition. $\square$

Note 4. The argument is by induction; let us do the case $n = 2$ for clarity. So we suppose

$$V = f(W_t)g(W_u - W_t).$$

From Proposition 18.3 we now have that

$$f(W_t) = c + \int_0^t H_s dW_s, \qquad g(W_u - W_t) = d + \int_t^u K_s dW_s.$$

Set $\overline{H}_r = H_r$ if $0 \le s < t$ and 0 otherwise. Set $\overline{K}_r = K_r$ if $s \le r < t$ and 0 otherwise. Let $X_s = c + \int_0^s \overline{H}_r dW_r$ and $Y_s = d + \int_0^s \overline{K}_r dW_r$. Then

$$\langle X, Y \rangle_s = \int_0^s \overline{H}_r \overline{K}_r dr = 0.$$

Then by the Ito product formula,

$$\begin{aligned} X_s Y_s &= X_0 Y_0 + \int_0^s X_r dY_r + \int_0^s Y_r dX_r \\ &\quad + \langle X, Y \rangle_s \\ &= cd + \int_0^s [X_r \overline{K}_r + Y_r \overline{H}_r] dW_r. \end{aligned}$$

If we now take $s = u$, that is exactly what we wanted. Note that $X_r \overline{K}_r + Y_r \overline{H}_r$ is 0 if $r > u$; this is needed to do the general induction step. $\square$

19. Completeness.

Now let S_t be a geometric Brownian motion. As we mentioned in the last section, if $S_t = S_0 \exp(\sigma W_t + (\mu - \sigma^2/2)t)$, then given S_t we can determine W_t and vice versa, so the σ fields generated by S_t and W_t are the same. Recall S_t satisfies

$$dS_t = \sigma S_t dW_t + \mu S_t dt.$$

Define a new probability $\overline{\mathbb{P}}$ by

$$\frac{d\overline{\mathbb{P}}}{d\mathbb{P}} = M_t = \exp(aW_t - a^2t/2).$$

By the Girsanov theorem,

$$\widetilde{W}_t = W_t - at$$

is a Brownian motion under $\overline{\mathbb{P}}$. So

$$dS_t = \sigma S_t d\widetilde{W}_t + \sigma S_t a dt + \mu S_t dt.$$

If we choose $a = -\mu/\sigma$, we then have

$$dS_t = \sigma S_t d\widetilde{W}_t. \tag{19.1}$$

Since $\widetilde{W}_t$ is a Brownian motion under $\overline{\mathbb{P}}$, then S_t must be a martingale, since it is a stochastic integral of a Brownian motion. We can rewrite (19.1) as

$$d\widetilde{W}_t = \sigma^{-1} S_t^{-1} dS_t. \tag{19.2}$$

Given a $\mathcal{F}_t$ measurable variable V, we know by Theorem 18.1 that there exist a constant and an adapted process H_s such that $\mathbb{E} \int_0^t H_s^2 ds < \infty$ and

$$V = c + \int_0^t H_s d\widetilde{W}_s.$$

But then using (19.2) we have

$$V = c + \int_0^t H_s \sigma^{-1} S_s^{-1} dS_s.$$

We have therefore proved

Theorem 19.1. *If S_t is a geometric Brownian motion and V is $\mathcal{F}_t$ measurable and integrable, then there exist a constant c and an adapted process K_s such that*

$$V = c + \int_0^t K_s dS_s.$$

Moreover, there is a probability $\overline{\mathbb{P}}$ under which S_t is a martingale.

The probability $\overline{\mathbb{P}}$ is called the risk-neutral measure. Under $\overline{\mathbb{P}}$ the stock price is a martingale.

20. Black-Scholes formula, I.

We can now derive the formula for the price of any option. Let $T \geq 0$ be a fixed real. If V is $\mathcal{F}_T$ measurable, we have by Theorem 19.1 that

$$V = c + \int_0^T K_s dS_s, \tag{20.1}$$

and under $\overline{\mathbb{P}}$, the process S_s is a martingale.

Theorem 20.1. *The price of V must be $\overline{\mathbb{E}}V$.*

Proof. This is the "no arbitrage" principle again. Suppose the price of the option V at time 0 is W. Starting with 0 dollars, we can sell the option V for W dollars, and use the W dollars to buy and trade shares of the stock. In fact, if we use c of those dollars, and invest according to the strategy of holding K_s shares at time s, then at time T we will have

$$e^{rT}(W_0 - c) + V$$

dollars. At time T the buyer of our option exercises it and we use V dollars to meet that obligation. That leaves us a profit of $e^{rT}(W_0 - c)$ if $W_0 > c$, without any risk. Therefore W_0 must be less than or equal to c. If $W_0 < c$, we just reverse things: we buy the option instead of sell it, and hold $-K_s$ shares of stock at time s. By the same argument, since we can't get a riskless profit, we must have $W_0 \geq c$, or $W_0 = c$.

Finally, under $\overline{\mathbb{P}}$ the process S_t is a martingale. So taking expectations in (20.1), we obtain

$$\overline{\mathbb{E}}V = c.$$

□

Note that there is a slight difference from the approach we used in the binomial asset model. There we showed that $(1+r)^{-k}S_k$ was a martingale under $\overline{\mathbb{P}}$. We could do the analogue to that here, i.e., find a probability under which $e^{-rt}S_t$ was a martingale, but it is slightly simpler to make S_t a martingale under $\overline{\mathbb{P}}$ and to incorporate the interest rate into the definition of V. So, for example, in the case of pricing a European call, we let

$$V = e^{-rT}(S_T - K)^+.$$

We can think of this as saying the value of V at time T is $(S_T - K)^+$ in terms of the value of the dollar at time T. In terms of present day dollars the value of V is $e^{-rT}(S_T - K)^+$.

The formula in the statement of Theorem 20.1. is amenable to calculation. Suppose we have the standard European option, where $V = e^{-rt}(S_t - K)^+$. Recall that under $\overline{\mathbb{P}}$ the stock price satisfies

$$dS_t = \sigma S_t d\widetilde{W}_t,$$

where $\widetilde{W}_t$ is a Brownian motion under $\overline{\mathbb{P}}$. So then

$$S_t = S_0 e^{\sigma \widetilde{W}_t - \sigma^2 t/2}.$$

Hence

$$\begin{aligned} \overline{\mathbb{E}}V &= \overline{\mathbb{E}}[e^{-rT}(S_T - K)^+] \\ &= \overline{\mathbb{E}}\Big[e^{-rT}[S_0 e^{\sigma \widetilde{W}_T - (\sigma^2/2)T} - K]^+\Big]. \end{aligned} \tag{20.2}$$

We know the density of $\widetilde{W}_T$ is just $(2\pi T)^{-1/2}e^{-y^2/(2T)}$, so we can do the calculations and end up with the famous Black-Scholes formula:

$$W_0 = x\Phi(g(x,T)) - Ke^{-rT}\Phi(h(x,T)),$$

where $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-y^2/2}\,dy$, $x = S_0$,

$$g(x,T) = \frac{\log(x/K) + (r+\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$h(x,T) = g(x,T) - \sigma\sqrt{T}.$$

It is of considerable interest that the final formula depends on σ but is completely independent of μ. The reason for that can be explained as follows. Under $\overline{\mathbb{P}}$ the process S_t satisfies $dS_t = \sigma S_t d\widetilde{W}_t$, where $\widetilde{W}_t$ is a Brownian motion. Therefore, similarly to formulas we have already done,

$$S_t = S_0 e^{\sigma\widetilde{W}_t - \sigma^2 t/2},$$

and there is no μ present here. (We used the Girsanov formula to get rid of the μ.) The price of the option V is

$$\overline{\mathbb{E}}\,e^{-rT}[S_T - K]^+, \tag{20.3}$$

which is independent of μ since S_t is. Therefore we can use this instead of (20.2), i.e., we can assume μ is zero, and the calculations become much simpler.

21. Hedging strategies.

The previous section allows us to compute the value of any option, but we would also like to know what the hedging strategy is. This means, if we know $V = \overline{\mathbb{E}}\,V + \int_0^T H_s dS_s$, what should H_s be? This might be important to know if we wanted to duplicate an option that was not available in the marketplace, or if we worked for a bank and wanted to provide an option for sale.

It is not always possible to compute H, but in many cases of interest it is possible. We illustrate one technique with two examples.

First, suppose we want to hedge the standard European call $V = e^{-rT}(S_T - K)^+$. We are working here with the risk-neutral probability only. It turns out it makes no difference: the definition of $\int_0^t H_s dX_s$ for a semimartingale X does not depend on the probability $\mathbb{P}$, other than worrying about some integrability conditions.

We can rewrite V as

$$V = \overline{\mathbb{E}}\,V + g(\widetilde{W}_T),$$

where

$$g(x) = e^{-rT}(e^{\sigma x - \sigma^2 T/2} - K)^+ - \overline{\mathbb{E}}\,V.$$

Therefore the expectation of $g(\widetilde{W}_T)$ is 0. Recall that under $\overline{\mathbb{P}}$, $\widetilde{W}$ is a Brownian motion. If we write $g(\widetilde{W}_T)$ as

$$\int_0^T H_s d\widetilde{W}_s, \tag{21.1}$$

then since $dS_t = \sigma S_t d\widetilde{W}_t$, we have

$$g(\widetilde{W}_T) = c + \int_0^T \frac{1}{\sigma S_s} H_s dS_s. \tag{21.2}$$

Therefore it suffices to find the representation of the form (21.1).

Recall from the section on the Markov property that

$$P_t f(x) = \overline{\mathbb{E}}^x f(\widetilde{W}_t) = \overline{\mathbb{E}} f(x + \widetilde{W}_t) = \int \frac{1}{\sqrt{2\pi t}} e^{-(y)^2/2t} f(x+y) dy.$$

Let $M_t = \overline{\mathbb{E}}[g(\widetilde{W}_T) \mid \mathcal{F}_t]$. By Proposition 4.3, we know that M_t is a martingale. By the Markov property Proposition 17.2, we see that

$$M_t = \overline{\mathbb{E}}^{\widetilde{W}_t}[g(\widetilde{W}_{T-t}] = P_{T-t} g(\widetilde{W}_t). \tag{21.3}$$

Now let us apply Ito's formula with the function $f(x_1, x_2) = P_{x_2} g(x_1)$ to the process $X_t = (X_t^1, X_t^2) = (\widetilde{W}_t, T-t)$. So we need to use the multidimensional version of Ito's formula. We have $dX_t^1 = d\widetilde{W}_t$ and $dX_t^2 = -dt$. Since X_t^2 is a decreasing process and has no martingale part, then $d\langle X^2 \rangle_t = 0$ and $d\langle X^1, X^2 \rangle_t = 0$, while $d\langle X^1 \rangle_t = dt$. Ito's formula says that

$$\begin{aligned} f(X_t^1, X_t^2) &= f(X_0^1, X_0^2) + \int_0^t \sum_{i=1}^2 \frac{\partial f}{\partial x_i}(X_t) dX_t^i \\ &\quad + \tfrac{1}{2} \int_0^t \sum_{i,j=1}^2 \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t) d\langle X^i, X^j \rangle_t \\ &= c + \int_0^t \frac{\partial f}{\partial x_1}(X_t) d\widetilde{W}_t + \text{ some terms with } dt. \end{aligned}$$

But we know that $f(X_t) = P_{T-t} g(\widetilde{W}_t) = M_t$ is a martingale, so the sum of the terms involving dt must be zero; if not, $f(X_t)$ would have a bounded variation part. We conclude

$$M_t = \int_0^t \frac{\partial}{\partial x} P_{T-s} g(\widetilde{W}_s) d\widetilde{W}_s.$$

If we take $t = T$, we then have

$$g(\widetilde{W}_T) = M_T = \int_0^T \frac{\partial}{\partial x} P_{T-s} g(\widetilde{W}_s) d\widetilde{W}_s,$$

and we have our representation.

For a second example, let's look at the sell-high option. Here the payoff is $\sup_{s \le T} S_s$, the largest the stock price ever is up to time T. This is $\mathcal{F}_T$ measurable, so we can compute its value. How can one get the equivalent outcome without looking into the future?

For simplicity, let us suppose the interest rate r is 0. Let $N_t = \sup_{s \le t} S_s$, the maximum up to time t. It is not the case that N_t is a Markov process. Intuitively, the reasoning goes like this: suppose the maximum up to time 1 is \$100, and we want to predict the maximum up to time 2. If the stock price at time 1 is close to \$100, then we have one prediction, while if the stock price at time 1 is close to \$2, we would definitely have another prediction. So the prediction for N_2 does not depend just on N_1, but also the stock price at time 1. This same intuitive reasoning does suggest, however, that the triple $Z_t = (S_t, N_t, t)$ is a Markov process, and this turns out to be correct. Adding in the information about the current stock price gives a certain amount of evidence to predict the future values of N_t; adding in the history of the stock prices up to time t gives no additional information.

Once we believe this, the rest of the argument is very similar to the first example. Let $P_u f(z) = \overline{\mathbb{E}}^z f(Z_u)$, where $z = (s, n, t)$. Let $g(Z_t) = N_t - \overline{\mathbb{E}} N_T$. Then

$$M_t = \overline{\mathbb{E}}[g(Z_T) \mid \mathcal{F}_t] = \overline{\mathbb{E}}^{Z_t}[g(Z_{T-t})] = P_{T-t}g(Z_t).$$

We then let $f(s,n,t) = P_{T-t}g(s,n,t)$ and apply Ito's formula. The process N_t is always increasing, so has no martingale part, and hence $\langle N \rangle_t = 0$. When we apply Ito's formula, we get a dS_t term, which is the martingale term, we get some terms involving dt, which are of bounded variation, and we get a term involving dN_t, which is also of bounded variation. But M_t is a martingale, so all the dt and dN_t terms must cancel. Therefore we should be left with the martingale term, which is

$$\int_0^t \frac{\partial}{\partial s} P_{T-s} g(S_s, N_s, s) dS_s,$$

where again $g(s,n,t) = n$. This gives us our hedging strategy for the sell-high option, and it can be explicitly calculated.

There is another way to calculate hedging strategies, using what is known as the Clark-Haussmann-Ocone formula. This is a more complicated procedure, and most cases can be done as well by an appropriate use of the Markov property.

22. Solving PDE.

Without going into the theory of PDE, let us look at how to solve some simple PDE using probability. Let us consider

$$f_t(x,t) = a(x) f_{xx}(x,t) + b(x) f_x(x,t), \qquad f(x,0) = g(x). \tag{22.1}$$

1

Here f is a function of x and t, a and b are given functions of x and g is also given. The above equation is known as the Cauchy problem.

Proposition 22.1. *Let $A = \sqrt{2a}$ and let X_t be the solution to*

$$dX_t = A(X_t)dW_t + b(X_t)dt. \tag{22.2}$$

The solution to the above equation is given by

$$f(x,t) = \mathbb{E}^x g(X_t).$$

Proof. Fix t_0 and let $M_t = f(X_t, t_0 - t)$. We first show M_t is a martingale. By Ito's formula,

$$\begin{aligned} dM_s &= f_x(X_s, t_0 - s)dX_s - f_t(X_s, t_0 - s)ds + \tfrac{1}{2} f_{xx}(X_s, t_0 - s)A^2(X_s)ds \qquad (22.3)\\ &= f_x(X_s, t_0 - s)A(X_s)dW_s + f_x(X_s, t_0 - s)b(X_s)ds + \tfrac{1}{2} f_{xx}(X_s, t_0 - s)A^2(X_s)ds \\ &\qquad - f_t(X_s, t_0 - s)ds. \end{aligned}$$

Since f solves (22.1), then

$$dM_t = f_x(X_s, t_0 - s)A(X_s)dW_s,$$

which is a stochastic integral of a Brownian motion, hence a martingale.

Now $\mathbb{E}^x M_0 = f(x, t_0)$ and $\mathbb{E}^x M_{t_0} = \mathbb{E}^x f(X_{t_0}, 0) = \mathbb{E}^x g(X_{t_0})$. Since martingales have constant expectation,

$$f(x, t_0) = \mathbb{E}^x g(X_{t_0}).$$

Since t_0 is arbitrary, the proposition is proved. □

To solve the equation

$$f_t(x,t) = a(x)f_{xx}(x,t) + b(x)f_x(x,t) + c(x)f(x,t), \qquad f(x,0) = g(x),$$

we will use similar methods to show that the solution is given by

$$f(x,t) = \mathbb{E}^x \Big[g(X_t) e^{\int_0^t c(X_s)ds} \Big],$$

where X_t is the solution to (22.2). (This is known as the Feynman-Kac formula.) To see this, if we let

$$N_t = M_t e^{\int_0^t c(X_s)ds},$$

where $M_t = f(X_t, t_0 - t)$, then the Ito product formula yields

$$dN_t = M_t e^{\int_0^t c(X_s)ds} c(X_t)dt + e^{\int_0^t c(X_s)ds} dM_t.$$

Using (22.3) and the fact that $af_{xx} + bf_x + cf = 0$, we see that that the dt term is 0 and N_t is a martingale. Using $\mathbb{E}\, N_0 = \mathbb{E}\, N_{t_0}$ leads to the desired representation of the solution.

Let us look at an example:

$$f_t = \tfrac{1}{2}\sigma^2 x^2 f_{xx} + rxf_x - rf,$$

which we will see is the PDE that arises in Black-Scholes. Here $a = \frac{1}{2}\sigma^2 x^2$ so that $A = \sigma x$, $b = rx$, and $c = -r$. The SDE to be solved, then, is

$$dX_t = \sigma X_t dW_t + rX_t dt, \qquad X_0 = x.$$

We know the solution to this is

$$X_t = xe^{\sigma W_t - \sigma^2 t/2 + rt}.$$

Hence

$$f(x,t) = \mathbb{E}^{\,x}\Big[g(X_t)e^{-rt}\Big] = \mathbb{E}\,\Big[e^{-rt} g\Big(xe^{\sigma W_t - \sigma^2 t/2 + rt}\Big)\Big].$$

Since we know the density of W_t, we can get an explicit expression (as an integral) for $f(x,t)$, namely,

$$f(x,t) = \frac{e^{-rt}}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} g\Big(xe^{\sigma y - \sigma^2 t/2 + rt}\Big) e^{-y^2/2t} dy.$$

23. Black-Scholes formula, II.

Here is a second approach to the Black-Scholes formula. This approach works for European calls and several other options, but does not work in the generality that the first approach does. On the other hand, it allows one to compute more easily what the equivalent strategy of buying or selling stock should be to duplicate the outcome of the given option.

Let V_t be the value of the portfolio and assume $V_t = f(S_t, T - t)$ for all t, where f is some function that is sufficiently smooth. We also want $V_T = (S_T - K)^+$.

Recall Ito's formula. The multivariate version is

$$f(X_t) = f(X_0) + \int_0^t \sum_{i=1}^d f_{x_i}(X_s)\, dX_s^i + \frac{1}{2}\int_0^t \sum_{i,j=1}^d f_{x_i x_j}(X_s)\, d\langle X^i, X^j\rangle_s.$$

Here $X_t = (X_t^1, \ldots, X_t^d)$ and f_{x_i} denotes the partial derivative of f in the x_i direction, and similarly for the second partial derivatives.

We apply this with $d = 2$ and $X_t = (S_t, T - t)$. From the SDE that S_t solves, $d\langle X^1 \rangle_t = \sigma^2 S_t^2 \, dt$, $\langle X^2 \rangle_t = 0$ (since $T - t$ is of bounded variation and hence has no martingale part), and $\langle X^1, X^2 \rangle_t = 0$. Also, $dX_t^2 = -dt$. Then

$$
\begin{aligned}
V_t - V_0 &= f(S_t, T-t) - f(S_0, T) \qquad (23.1)\\
&= \int_0^t f_x(S_u, T-u) \, dS_u - \int_0^t f_s(S_u, T-u) \, du \\
&\quad + \tfrac{1}{2} \int_0^t \sigma^2 S_u^2 f_{xx}(S_u, T-u) \, du.
\end{aligned}
$$

On the other hand, if a_u and b_u are the number of shares of stock and bonds, respectively, held at time u,

$$V_t - V_0 = \int_0^t a_u \, dS_u + \int_0^t b_u \, d\beta_u. \qquad (23.2)$$

This formula says that the increase in net worth is given by the profit we obtain by holding a_u shares of stock and b_u bonds at time u. Since the value of the portfolio at time t is

$$V_t = a_t S_t + b_t \beta_t,$$

we must have

$$b_t = (V_t - a_t S_t)/\beta_t. \qquad (23.3)$$

Also, recall

$$\beta_t = \beta_0 e^{rt}. \qquad (23.4)$$

To match up (23.2) with (23.1), we must therefore have

$$a_t = f_x(S_t, T-t) \qquad (23.5)$$

and

$$r[f(S_t, T-t) - S_t f_x(S_t, T-t)] = -f_s(S_t, T-t) + \frac{1}{2}\sigma^2 S_t^2 f_{xx}(S_t, T-t) \qquad (23.6)$$

for all t and all S_t. (23.6) leads to the parabolic PDE

$$f_s = \tfrac{1}{2}\sigma^2 x^2 f_{xx} + rx f_x - rf, \qquad (x, s) \in (0, \infty) \times [0, T), \qquad (23.7)$$

and

$$f(x, 0) = (x - K)^+. \qquad (23.8)$$

Solving this equation for f, $f(x,T)$ is what V_0 should be, i.e., the cost of setting up the equivalent portfolio. Equation (23.5) shows what the trading strategy should be. In the previous section we showed how to solve this PDE.

24. The fundamental theorem of finance.

In Section 19, we showed there was a probability measure under which S_t was a martingale. This is true very generally. Let S_t be the price of a security. We will suppose S_t is a continuous semimartingale, and can be written $S_t = M_t + A_t$.

Arbitrage means that there is a trading strategy H_s such that there is no chance that we lose anything and there is a positive profit with positive probability. Mathematically, arbitrage exists if there exists H_s that is adapted and satisfies a suitable integrability condition with

$$\int_0^T H_s dS_s \geq 0, \qquad \text{a.s.}$$

and

$$\mathbb{P}\Big(\int_0^T H_s dS_s > b\Big) > \varepsilon$$

for some $b, \varepsilon > 0$. It turns out that to get a necessary and sufficient condition for S_t to be a martingale, we need a slightly weaker condition.

The NFLVR condition ("no free lunch with vanishing risk") is that there do not exist a fixed time T, $\varepsilon, b > 0$, and H_n (that are adapted and satisfy the appropriate integrability conditions) such that

$$\int_0^T H_n(s)\, dS_s > -\frac{1}{n}, \qquad \text{a.s.}$$

for all t and

$$\mathbb{P}\Big(\int_0^T H_n(s)\, dS_s > b\Big) > \varepsilon.$$

Here T, b, ε do not depend on n. The condition says that one can with positive probability ε make a profit of b and with a loss no larger than $1/n$.

Two probabilities $\mathbb{P}$ and $\mathbb{Q}$ are equivalent if $\mathbb{P}(A) = 0$ if and only $\mathbb{Q}(A) = 0$, i.e., the two probabilities have the same collection of sets of probability zero. $\mathbb{Q}$ is an equivalent martingale measure if $\mathbb{Q}$ is a probability measure, $\mathbb{Q}$ is equivalent to $\mathbb{P}$, and S_t is a martingale under $\mathbb{Q}$.

Theorem 24.1. *If S_t is a continuous semimartingale and the NFLVR conditions holds, then there exists an equivalent martingale measure $\mathbb{Q}$.*

The proof is rather technical and involves some heavy-duty measure theory, so we will only point examine a part of it. Suppose that we happened to have $S_t = W_t + f(t)$, where $f(t)$ is a deterministic increasing continuous function. To obtain the equivalent martingale measure, we would want to let

$$M_t = e^{-\int_0^t f'(s)dW_s - \frac{1}{2}\int_0^t (f'(s))^2 ds}.$$

In order for M_t to make sense, we need f to be differentiable. A result from measure theory says that if f is not differentiable, then we can find a subset A of $[0, \infty)$ such that $\int_0^t 1_A(s)ds = 0$ but the amount of increase of f over the set A is positive. This last statement is phrased mathematically by saying

$$\int_0^t 1_A(s)df(s) > 0,$$

where the integral is a Riemann-Stieltjes (or better, a Lebesgue-Stieltjes) integral. Then if we hold $H_s = 1_A(s)$ shares at time s, our net profit is

$$\int_0^t H_s dS_s = \int_0^t 1_A(s)dW_s + \int_0^t 1_A(s)\, df(s).$$

The second term would be positive since this is the amount of increase of f over the set A. The first term is 0, since $\mathbb{E}\,(\int_0^t 1_A(s)dW_s)^2 = \int_0^t 1_A(s)^2 ds = 0$. So our net profit is nonrandom and positive, or in other words, we have made a net gain without risk. This contradicts "no arbitrage." See Note 1 for more on this.

Sometime Theorem 24.1 is called the first fundamental theorem of asset pricing. The second fundamental theorem is the following.

Theorem 24.2. *The equivalent martingale measure is unique if and only if the market is complete.*

We will not prove this.

Note 1. We will not prove Theorem 24.1, but let us give a few more indications of what is going on. First of all, recall the Cantor set. This is where $E_1 = [0, 1]$, E_2 is the set obtained from E_1 by removing the open interval $(\frac{1}{3}, \frac{2}{3})$, E_3 is the set obtained from E_2 by removing the middle third from each of the two intervals making up E_2, and so on. The intersection, $E = \cap_{n=1}^{\infty} E_n$, is the Cantor set, and is closed, nonempty, in fact uncountable, yet it contains no intervals. Also, the Lebesgue measure of A is 0. We set $A = E$. Let f be the Cantor-Lebesgue function. This is the function that is equal to 0 on $(-\infty, 0]$, 1 on $[1, \infty)$, equal to $\frac{1}{2}$ on the interval $[\frac{1}{3}, \frac{2}{3}]$, equal to $\frac{1}{4}$ on $[\frac{1}{9}, \frac{2}{9}]$, equal to $\frac{3}{4}$ on $[\frac{7}{9}, \frac{8}{9}]$, and is defined similarly on each interval making up the complement of A. It turns out we can define f on A so that it is continuous, and one can show $\int_0^1 1_A(s)df(s) = 1$. So this A and f provide a concrete example of what we were discussing.

25. American puts.

The proper valuation of American puts is one of the important unsolved problems in mathematical finance. Recall that a European put pays out $(K - S_T)^+$ at time T, while an American put allows one to exercise early. If one exercises an American put at time $t < T$, one receives $(K - S_t)^+$. Then during the period $[t, T]$ one receives interest, and the amount one has is $(K - S_t)^+ e^{r(T-t)}$. In today's dollars that is the equivalent of $(K - S_t)^+ e^{-rt}$. One wants to find a rule, known as the exercise policy, for when to exercise the put, and then one wants to see what the value is for that policy. Since one cannot look into the future, one is in fact looking for a stopping time τ that maximizes

$$\overline{\mathbb{E}}\, e^{-r\tau}(K - S_\tau)^+.$$

There is no good theoretical solution to finding the stopping time τ, although good approximations exist. We will, however, discuss just a bit of the theory of optimal stopping, which reworks the problem into another form.

Let G_t denote the amount you will receive at time t. For American puts, we set

$$G_t = e^{-rt}(K - S_t)^+.$$

Our problem is to maximize $\overline{\mathbb{E}}\, G_\tau$ over all stopping times τ.

We first need

Proposition 25.1. *If S and T are bounded stopping times with $S \leq T$ and M is a martingale, then*

$$\mathbb{E}\,[M_T \mid \mathcal{F}_S] = M_S.$$

Proof. Let $A \in \mathcal{F}_S$. Define U by

$$U(\omega) = \begin{cases} S(\omega) & \text{if } \omega \in A, \\ T(\omega) & \text{if } \omega \notin A. \end{cases}$$

It is easy to see that U is a stopping time, so by Doob's optional stopping theorem,

$$\mathbb{E}\, M_0 = \mathbb{E}\, M_U = \mathbb{E}\,[M_S; A] + \mathbb{E}\,[M_T; A^c].$$

Also,

$$\mathbb{E}\, M_0 = \mathbb{E}\, M_T = \mathbb{E}\,[M_T; A] + \mathbb{E}\,[M_T; A^c].$$

Taking the difference, $\mathbb{E}\,[M_T; A] = \mathbb{E}\,[M_s; A]$, which is what we needed to show. $\square$

Given two supermartingales X_t and Y_t, it is routine to check that $X_t \wedge Y_t$ is also a supermartingale. Also, if X_t^n are supermartingales with $X_t^n \downarrow X_t$, one can check that X_t

is again a supermartingale. With these facts, one can show that given a process such as G_t, there is a least supermartingale larger than G_t.

So we define W_t to be a supermartingale (with respect to $\overline{\mathbb{P}}$, of course) such that $W_t \geq G_t$ a.s for each t and if Y_t is another supermartingale with $Y_t \geq G_t$ for all t, then $W_t \leq Y_t$ for all t. We set $\overline{\tau} = \inf\{t : W_t = G_t\}$. We will show that $\overline{\tau}$ is the solution to the problem of finding the optimal stopping time. Of course, computing W_t and $\overline{\tau}$ is another problem entirely.

Let

$$\mathcal{T}_t = \{\tau : \tau \text{ is a stopping time}, t \leq \tau \leq T\}.$$

Let

$$V_t = \sup_{\tau \in \mathcal{T}_t} \overline{\mathbb{E}}\,[G_\tau \mid \mathcal{F}_t].$$

Proposition 25.2. *V_t is a supermartingale and $V_t \geq G_t$ for all t.*

Proof. The fixed time t is a stopping time in $\mathcal{T}_t$, so $V_t \geq \overline{\mathbb{E}}\,[G_t \mid \mathcal{F}_t] = G_t$, or $V_t \geq G_t$. so we only need to show that V_t is a supermartingale.

Suppose $s < t$. Let π be the stopping time in $\mathcal{T}_t$ for which $V_t = \overline{\mathbb{E}}\,[G_\pi \mid \mathcal{F}_t]$. $\pi \in \mathcal{T}_t \subset \mathcal{T}_s$. Then

$$\overline{\mathbb{E}}\,[V_t \mid \mathcal{F}_s] = \overline{\mathbb{E}}\,[G_\pi \mid \mathcal{F}_s] \leq \sup_{\tau \in \mathcal{T}_s} \mathbb{E}\,[G_\tau \mid \mathcal{F}_s] = V_s.$$

$\square$

Proposition 25.3. *If Y_t is a supermartingale with $Y_t \geq G_t$ for all t, then $Y_t \geq V_t$.*

Proof. If $\tau \in \mathcal{T}_t$, then since Y_t is a supermartingale, we have

$$\overline{\mathbb{E}}\,[Y_\tau \mid \mathcal{F}_t] \leq Y_t.$$

So

$$V_t = \sup_{\tau \in \mathcal{T}_t} \overline{\mathbb{E}}\,[G_\tau \mid \mathcal{F}_t] \leq \sup_{\tau \in \mathcal{T}_t} \overline{\mathbb{E}}\,[Y_\tau \mid \mathcal{F}_t] \leq Y_t.$$

$\square$

What we have shown is that W_t is equal to V_t. It remains to show that $\overline{\tau}$ is optimal. There may in fact be more than one optimal time, but in any case $\overline{\tau}$ is one of them. Recall we have $\mathcal{F}_0$ is the σ-field generated by S_0, and hence consists of only $\emptyset$ and Ω.

Proposition 25.4. *$\overline{\tau}$ is an optimal stopping time.*

Proof. Since $\mathcal{F}_0$ is trivial, $V_0 = \sup_{\tau \in \mathcal{T}_0} \overline{\mathbb{E}}[G_\tau \mid \mathcal{F}_0] = \sup_\tau \overline{\mathbb{E}}[G_\tau]$. Let σ be a stopping time where the supremum is attained. Then

$$V_0 \geq \overline{\mathbb{E}}[V_\sigma \mid \mathcal{F}_0] = \overline{\mathbb{E}}[V_\sigma] \geq \overline{\mathbb{E}}[G_\sigma] = V_0.$$

Therefore all the inequalities must be equalities. Since $V_\sigma \geq G_\sigma$, we must have $V_\sigma = G_\sigma$. Since $\overline{\tau}$ was the first time that W_t equals G_t and $W_t = V_t$, we see that $\overline{\tau} \leq \sigma$. Then

$$\overline{\mathbb{E}}[G_{\overline{\tau}}] = \overline{\mathbb{E}}[V_{\overline{\tau}}] \geq \overline{E} V_\sigma = \overline{\mathbb{E}} G_\sigma.$$

Therefore the expected value of $G_{\overline{\tau}}$ is as least as large as the expected value of G_σ, and hence $\overline{\tau}$ is also an optimal stopping time. $\square$

The above representation of the optimal stopping problem may seem rather bizarre. However, this procedure gives good usable results for some optimal stopping problems. An example is where G_t is a function of just W_t.

26. Term structure.

We now want to consider the case where the interest rate is nondeterministic, that is, it has a random component. To do so, we take another look at option pricing.

Accumulation factor. Let $r(t)$ be the (random) interest rate at time t. Let

$$\beta(t) = e^{\int_0^t r(u)du}$$

be the accumulation factor. One dollar at time T will be worth $1/\beta(T)$ in today's dollars.

Let $V = (S_T - K)^+$ be the payoff on the standard European call option at time T with strike price K, where S_t is the stock price. In today's dollars it is worth, as we have seen, $V/\beta(T)$. Therefore the price of the option should be

$$\overline{\mathbb{E}}\Big[\frac{V}{\beta(T)}\Big].$$

We can also get an expression for the value of the option at time t. The payoff, in terms of dollars at time t, should be the payoff at time T discounted by the interest or inflation rate, and so should be

$$e^{-\int_t^T r(u)du}(S_T - K)^+.$$

Therefore the value at time t is

$$\overline{\mathbb{E}}\Big[e^{-\int_t^T r(u)du}(S_T - K)^+ \mid \mathcal{F}_t\Big] = \overline{\mathbb{E}}\Big[\frac{\beta(t)}{\beta(T)} V \mid \mathcal{F}_t\Big] = \beta(t)\overline{\mathbb{E}}\Big[\frac{V}{\beta(T)} \mid \mathcal{F}_t\Big].$$

From now on we assume we have already changed to the risk-neutral measure and we write $\mathbb{P}$ instead of $\overline{\mathbb{P}}$.

Zero coupon. A zero coupon bond with maturity date T pays \$1 at time T and nothing before. This is equivalent to an option with payoff value $V = 1$. So its price at time t, as above, should be

$$B(t,T) = \beta(t)\mathbb{E}\left[\frac{1}{\beta(T)} \mid \mathcal{F}_t\right] = \mathbb{E}\left[e^{-\int_t^T r(u)du} \mid \mathcal{F}_t\right].$$

Let's derive the SDE satisfied by $B(t,T)$. Let $N_t = \mathbb{E}\left[1/\beta(T) \mid \mathcal{F}_t\right]$. This is a martingale. By the martingale representation theorem,

$$N_t = \mathbb{E}\left[1/\beta(T)\right] + \int_0^t H_s dW_s$$

for some adapted integrand H_s. So $B(t,T) = \beta(t)N_t$. Here T is fixed. By Ito's product formula,

$$\begin{aligned}
dB(t,T) &= \beta(t)dN_t + N_t d\beta(t)\\
&= \beta(t)H_t dW_t + N_t r(t)\beta(t)dt\\
&= \beta(t)H_t dW_t + B(t,T)r(t)dt,
\end{aligned}$$

and we thus have

$$dB(t,T) = \beta(t)H_t dW_t + B(t,T)r(t)dt. \tag{26.1}$$

Forward rates. We now discuss forward rates. If one holds T fixed and graphs $B(t,T)$ as a function of t, the graph will not clearly show the behavior of r. One sometimes specifies interest rates by what are known as forward rates.

Suppose we want to borrow \$1 at time T and repay it with interest at time $T+\varepsilon$. At the present time we are at time $t \le T$. Let us try to accomplish this by buying a zero coupon bond with maturity date T and shorting (i.e., selling) N zero coupon bonds with maturity date $T+\varepsilon$. Our outlay of money at time t is

$$B(t,T) - NB(t,T+\varepsilon) = 0.$$

If we set

$$N = B(t,T)/B(t,T+\varepsilon),$$

our outlay at time t is 0. At time T we receive \$1. At time $T+\varepsilon$ we pay $B(t,T)/B(t,T+\varepsilon)$. The effective rate of interest R over the time period T to $T+\varepsilon$ is

$$e^{\varepsilon R} = \frac{B(t,T)}{B(t,T+\varepsilon)}.$$

Solving for R, we have

$$R = \frac{\log B(t,T) - \log B(t,T+\varepsilon)}{\varepsilon}.$$

We now let $\varepsilon \to 0$. We define the forward rate by

$$f(t,T) = -\frac{\partial}{\partial T} \log B(t,T). \tag{26.2}$$

Sometimes interest rates are specified by giving $f(t,T)$ instead of $B(t,T)$ or $r(t)$.

Recovering B from f. Let us see how to recover $B(t,T)$ from $f(t,T)$. Integrating, we have

$$\begin{aligned}\int_t^T f(t,u)du &= -\int_t^T \frac{\partial}{\partial u} \log B(t,u)du = -\log B(t,u) \mid_{u=t}^{u=T} \\ &= -\log B(t,T) + \log B(t,t).\end{aligned}$$

Since $B(t,t)$ is the value of a zero coupon bond at time t which expires at time t, it is equal to 1, and its log is 0. Solving for $B(t,T)$, we have

$$B(t,T) = e^{-\int_t^T f(t,u)du}. \tag{26.3}$$

Recovering r from f. Next, let us show how to recover $r(t)$ from the forward rates. We have

$$B(t,T) = \mathbb{E}\left[e^{-\int_t^T r(u)du} \mid \mathcal{F}_t\right].$$

Differentiating,

$$\frac{\partial}{\partial T} B(t,T) = \mathbb{E}\left[-r(T)e^{-\int_t^T r(u)du} \mid \mathcal{F}_t\right].$$

Evaluating this when $T = t$, we obtain

$$\mathbb{E}\left[-r(t) \mid \mathcal{F}_t\right] = -r(t). \tag{26.4}$$

On the other hand, from (26.3) we have

$$\frac{\partial}{\partial T} B(t,T) = -f(t,T)e^{-\int_t^T f(t,u)du}.$$

Setting $T = t$ we obtain $-f(t,t)$. Comparing with (26.4) yields

$$r(t) = f(t,t). \tag{26.5}$$

27. Some interest rate models.

Heath-Jarrow-Morton model

Instead of specifying r, the Heath-Jarrow-Morton model (HJM) specifies the forward rates:

$$df(t,T) = \sigma(t,T)dW_t + \alpha(t,T)dt. \tag{27.1}$$

Let us derive the SDE that $B(t,T)$ satisfies. Let

$$\alpha^*(t,T) = \int_t^T \alpha(t,u)du, \qquad \sigma^*(t,T) = \int_t^T \sigma(t,u)du.$$

Since $B(t,T) = \exp(-\int_t^T f(t,u)du)$, we derive the SDE for B by using Ito's formula with the function e^x and $X_t = -\int_t^T f(t,u)du$. We have

$$\begin{aligned}
dX_t &= f(t,t)dt - \int_t^T df(t,u)du \\
&= r(t)dt - \int_t^T [\alpha(t,u)dt + \sigma(t,u)dW_t]\, du \\
&= r(t)dt - \Big[\int_t^T \alpha(t,u)du\Big]dt - \Big[\int_t^T \sigma(t,u)du\Big]dW_t \\
&= r(t)dt - \alpha^*(t,T)dt - \sigma^*(t,T)dW_t.
\end{aligned}$$

Therefore, using Ito's formula,

$$\begin{aligned}
dB(t,T) &= B(t,T)dX_t + \tfrac{1}{2}B(t,T)(\sigma^*(t,T))^2dt \\
&= B(t,T)\Big[r(t) - \alpha^* + \tfrac{1}{2}(\sigma^*)^2\Big]dt - \sigma^*B(t,T)dW_t.
\end{aligned}$$

From (26.1) we know the dt term must be $B(t,T)r(t)dt$, hence

$$dB(t,T) = B(t,T)r(t)dt - \sigma^*B(t,T)dW_t.$$

Comparing with (27.1), we see that if $\mathbb{P}$ is the risk-neutral measure, we have $\alpha^* = \frac{1}{2}(\sigma^*)^2$. See Note 1 for more on this.

Hull and White model

In this model, the interest rate r is specified as the solution to the SDE

$$dr(t) = \sigma(t)dW_t + (a(t) - b(t)r(t))dt. \tag{27.2}$$

Here σ, a, b are deterministic functions. The stochastic integral term introduces randomness, while the $a - br$ term causes a drift toward $a(t)/b(t)$. (Note that if $\sigma(t) = \sigma, a(t) = a, b(t) = b$ are constants and $\sigma = 0$, then the solution to (27.2) becomes $r(t) = a/b$.)

(27.2) is one of those SDE's that can be solved explicitly. Let $K(t) = \int_0^t b(u)du$. Then

$$d\Big[e^{K(t)}r(t)\Big] = e^{K(t)}r(t)b(t)dt + e^{K(t)}\Big[a(t) - b(t)r(t)\Big]dt + e^{K(t)}[\sigma(t)dW_t]$$
$$= e^{K(t)}a(t)dt + e^{K(t)}[\sigma(t)dW_t].$$

Integrating both sides,

$$e^{K(t)}r(t) = r(0) + \int_0^t e^{K(u)}a(u)du + \int_0^t e^{K(u)}\sigma(u)dW_u.$$

Multiplying both sides by $e^{-K(t)}$, we have the explicit solution

$$r(t) = e^{-K(t)}\Big[r(0) + \int_0^t e^{K(u)}a(u)du + \int_0^t e^{K(u)}\sigma(u)dW_u\Big].$$

If $F(u)$ is deterministic, then

$$\int_0^t F(u)dW_u = \lim \sum F(u_i)(W_{u_{i+1}} - W_{u_i}).$$

From undergraduate probability, linear combinations of Gaussian r.v.'s (Gaussian = normal) are Gaussian, and also limits of Gaussian r.v.'s are Gaussian, so we conclude that the r.v. $\int_0^t F(u)dW_u$ is Gaussian. We see that the mean at time t is

$$\mathbb{E}\, r(t) = e^{-K(t)}\Big[r(0) + \int_0^t e^{K(u)}a(u)du\Big].$$

We know how to calculate the second moment of a stochastic integral, so

$$\operatorname{Var} r(t) = e^{-2K(t)}\int_0^t e^{2K(u)}\sigma(u)^2du.$$

(One can similarly calculate the covariance of $r(s)$ and $r(t)$.) Limits of linear combinations of Gaussians are Gaussian, so we can calculate the mean and variance of $\int_0^T r(t)dt$ and get an explicit expression for

$$B(0,T) = \mathbb{E}\, e^{-\int_0^T r(u)du}.$$

Cox-Ingersoll-Ross model

One drawback of the Hull and White model is that since $r(t)$ is Gaussian, it can take negative values with positive probability, which doesn't make sense. The Cox-Ingersoll-Ross model avoids this by modeling r by the SDE

$$dr(t) = (a - br(t))dt + \sigma\sqrt{r(t)}dW_t.$$

The difference from the Hull and White model is the square root of r in the stochastic integral term. This square root term implies that when $r(t)$ is small, the fluctuations in $r(t)$ are larger than they are in the Hull and White model. Provided $a \geq \frac{1}{2}\sigma^2$, it can be shown that $r(t)$ will never hit 0 and will always be positive. Although one cannot solve for r explicitly, one can calculate the distribution of r. It turns out to be related to the square of what are known in probability theory as Bessel processes. (The density of $r(t)$, for example, will be given in terms of Bessel functions.)

Note 1. If $\mathbb{P}$ is not the risk-neutral measure, it is still possible that one exists. Let $\theta(t)$ be a function of t, let $M_t = \exp(-\int_0^t \theta(u)dW_u - \frac{1}{2}\int_0^t \theta(u)^2 du)$ and define $\overline{\mathbb{P}}(A) = \mathbb{E}\left[M_T; A\right]$ for $A \in \mathcal{F}_T$. By the Girsanov theorem,

$$dB(t,T) = B(t,T)\Big[r(t) - \alpha^* + \tfrac{1}{2}(\sigma^*)^2 + \sigma^*\theta\Big]dt - \sigma^* B(t,T)d\widetilde{W}_t,$$

where $\widetilde{W}_t$ is a Brownian motion under $\overline{\mathbb{P}}$. Again, comparing this with (26.1) we must have

$$\alpha^* = \tfrac{1}{2}(\sigma^*)^2 + \sigma^*\theta.$$

Differentiating with respect to T, we obtain

$$\alpha(t,T) = \sigma(t,T)\sigma^*(t,T) + \sigma(t,T)\theta(t).$$

If we try to solve this equation for θ, there is no reason off-hand that θ depends only on t and not T. However, if θ does not depend on T, $\overline{\mathbb{P}}$ will be the risk-neutral measure.