Shreve Stochastic Calculus Notes

Benjamin Marrow

February 13, 2019

1 General Probability Theory

1.1 Infinite Probability Spaces

Definition 1.1 (σ-Algebra)**.** Let Ω be a nonempty set, and let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a σ-algebra (called a σ-field by some authors) provided that:

- the empty set $\varnothing$ belongs to $\mathcal{F}$
- whenever a set A belongs to $\mathcal{F}$, its complement A^C also belongs to $\mathcal{F}$, and
- whenever a sequence of sets $A1, A2, \cdots$ belongs to $\mathcal{F}$, their union $\bigcup_{n=1}^{\infty} A_n$ also belongs to $\mathcal{F}$

For example, if the set of objects $\Omega = \{a, b, c, d\}$, some possible σ-algebras are:

- $\{\varnothing, \{a, b, c, d\}\}$
- $\{\varnothing, \{a, b, c, d\}, \{b\}, \{a, c, d\}\}$

The set of all possible subsets – known as the 'power set' is always a σ-algebra. Meanwhile, the following would not be considered σ-algebras:

- $\{\{a, b\}, \{c, d\}\}$ (doesn't contain empty set)
- $\{\varnothing, \Omega, \{a\}\}$ (doesn't contain a^c)
- $\{\varnothing, \Omega, \{a\}, \{b\}, \{c\}, \{d\}\}$ (doesn't contain unions e.g.).

Using the following property:

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c,$$

it follows that the intersection of finite or infinite collection of sets in the σ-algebra remains in the σ-algebra. The usefulness of defining σ-algebras is that we can perform operations of interest on the sets (unions, intersections) and remain within the σ-algebra.

Definition 1.2 (Probability Measure)**.** Let Ω be a nonempty set, and let $\mathcal{F}$ be a σ-algebra of subsets of Ω. A probability measure $\mathbb{P}$ is a function that, to every set $A \in \mathcal{F}$, assigns a number in $[0, 1]$, called the probability of A and written $\mathbb{P}(A)$. We requre:

- $\mathbb{P}(\Omega) = 1$, and
- (countable additivity) whenever $A_1, A_2, \ldots,$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \tag{1}$$

(In other words, for *disjoint* sets, the probability of the union is the sum of the individual probabilities).

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

From this definitions, several other properties naturally follow:

- First, since finite sequences can be made into countably infinite sequences through the additions of empty sets, the countable additivity guarantees finite additivity as well.
- Second, since $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\Omega) = 1$, we get that:

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

A probability measure must satisfy all these conditions.

Definition 1.3 (Borel-σ-algebra)**.** The σ-algebra obtained by beginning with closed intervals and adding everything else necessary in order to have a σ-algebra is called the *Borel σ-algebra* of subsets of $[0,1]$ and is denoted by $\mathcal{B}[0,1]$.

"To get the Borel subsets of $\mathbb{R}$, one begins with the closed intervals $[a,b] \in \mathbb{R}$ and adds all other set sthat are necessary in order to have a σ-algebra. This means that unions of sequences of closed intervals are Borel sets. In particular, every open interval is a Borel set, because it is the union of a sequence of open intervals."

"The distribution of a random variable is itself a probability measure, but it is a measure on subsets of $\mathbb{R}$ rather than subsets of Ω."

Definition 1.4 (Almost Surely)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If a set $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, we say that the event A occurs *almost surely.*

1.2 Random Variables and Distributions

Definition 1.5 (Random Variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is a real valued function X defined on Ω with the property that for every Borel subset B of $\mathbb{R}$, the subset of Ω given by:

$$\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\} \tag{2}$$

is in the σ-algebra $\mathcal{F}$.

In other words, a random variable is a function that maps a set of outcomes to a measurable space. Examples of random variables are (i) a function that maps the selection of a person (the experiment) to their height, or (ii) a function that maps a dice-roll tuple to a sum of the rolls.

Every open interval is a Borel set, because an open interval can be written as the union of a sequence of closed intervals. Every closed set is a Borel set because it is the complement of an open set.

Definition 1.6 (Distribution Measure)**.** Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution measure of X is the probability measure μ_X that assigns to each Borel subset B of $\mathbb{R}$ the mass $\mu_X(B) = \mathbb{P}\{X \in B\}$.

To see why distributions of random variables differ from random variables themselves, consider the case of $X(\omega) = \omega, Y(\omega) = 1 - \omega$. They have the same distributions, but $X \neq Y$. Importantly, different measures define different distributions.

We can also define the distribution measure of a random variable through its cumulative distribution function, $F(x)$, where the cumulative distribution is related to the distribution measure:

$$F(x) = \mu_X(\infty, x]$$

and, conversely:

$$\mu_X(x,y] = F(y) - F(x) \text{ for } x < y$$

We can define a density function:

$$\mu_X[a,b] = \mathbb{P}\{a \le X \le b\} = \int_a^b f(x)dx, -\infty < a \le b < \infty$$

1.3 Expectations

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We are interested in computing an average value. When Ω is finite,

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

Similarly, when Ω is countably infinite, we can define the expectation as an infinite sum:

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} X(\omega_k)\mathbb{P}(\omega_k)$$

Since sums across uncountable sets are not well defined, we must use the notion of an integral. However, Ω is often not a subset of $\mathbb{R}$, and in these cases, there is no natural way to partition the set Ω into tagged partitions. Since the measure μ_X can map abstract spaces to a $\mathbb{R}$, we can instead partition the y-axis, rather than the x-axis.

For each subinterval $[y_k, y_{k+1}]$, we set:

$$A_k = \{\omega \in \Omega; y_l \leq X(\omega) < y_{k+1}\},$$

and define the lower Lebesgue sum to be:

$$LS_{\Pi}^{-}(X) = \sum_{k=1}^{\infty} y_k \mathbb{P}(A_k)$$

As the norm of the partition converges, the lower sum converges to the Lebesgue integral. We now define the properties of the Lebesgue integral:

Theorem 1.1 (Lebesgue Integral Properties)**.** *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

i If X takes on finitely many values, $y_0, y_1, y_2, \ldots, y_n$, then

$$\int_{\Omega} X(\Omega)d\mathbb{P}(\Omega) = \sum_{k=0}^{n} y_k \mathbb{P}\{X = y_k\}$$

ii (Integrability) The random variable X is integrable IFF

$$\int_{\Omega} |X(\omega)| \, d\mathbb{P}(\omega) < \infty$$

Now let Y be another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

iii (Comparison) If $X \leq Y$ almost surely (i.e. $\mathbb{P}\{X \leq Y\} = 1$), and if $\int_{\Omega} X(\omega)d\mathbb{P}(\omega)$ and $\int_{\Omega} Y(\omega)d\mathbb{P}(\omega)$ are defined, then:

$$\int_{\Omega} X(\omega)d\mathbb{P}(\omega) \leq \int_{\Omega} Y(\omega)d\mathbb{P}(\omega)$$

In particular, if $X = Y$ almost surely and one of the integrals is defined, then they are both defined and

$$\int_{\Omega} X(\omega)d\mathbb{P}(\omega) = \int_{\Omega} Y(\omega)d\mathbb{P}(\omega)$$

iv (Linearity) If α and β are real constants and X and Y are integrable, or if α and β are nonnegative constants and X and Y are nonegative, then:

$$\int_{\Omega} (\alpha X(\omega) + \beta Y(\omega)) \, d\mathbb{P}(\omega) = \alpha \int_{\Omega} X(\omega)d\mathbb{P}(\omega) + \beta \int_{\Omega} Ym(\omega)d\mathbb{P}(\omega)$$

By way of linearity, it also follows that:

$$\int_{A \cup B} X(\omega)d\mathbb{P}(\omega) = \int_{A} X(\omega)d\mathbb{P}(\omega) + \int_{B} X(\omega)d\mathbb{P}(\omega)$$

We are now equipped to define the expected value.

Definition 1.7 (Expected Value)**.** Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The expectation of X is defined to be:

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$$

Theorem 1.2 (Properties of Expected Values)**.** *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

i If Ω is finite, then

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

ii (integrability) The random variable X is integrable IFF

$$\mathbb{E}|X| < \infty$$

Now let Y be another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

iii (Comparison) If $X \leq Y$ almost surely and X and Y are integrable or almost surely nonnegative, then:

$$\mathbb{E}[X] \leq \mathbb{E}[Y]$$

iv (Linearity) If α and β are real constants and X and Y are integrable, or if α and β are nonnegative constants and X and Y are nonnegative, then:

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$$

v (Jensen's inequality) If ϕ is a convex real-valued function defined on $\mathbb{R}$ and if $\mathbb{E}[X] < \infty$, then:

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

Definition 1.8 (Lebesgue Measure on $\mathbb{R}$)**.** Let $\mathcal{B}$ be the σ-algebra of Borel subsets of $\mathbb{R}$ (i.e. the smallest σ-algebra) containing all teh closed intervals $[a, b]$. The Lebesgue measure on $\mathbb{R}$, which we denote by $\mathcal{L}$, assigns to each est $B \in \mathcal{B}(\mathbb{R})$ a number in $[0, \infty)$ or the value ∞ so that:

i) $\mathcal{L}[a, b] = b - a$ whenever $a \leq b$, and

ii) if $B_1, B_2, B_3, \ldots$ is a sequence of disjoint sets in $\mathcal{B}(\mathbb{R})$, then we have the countable additivity property:

$$\mathcal{L}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathcal{L}(B_n)$$

A function f with the property that, for every Borel subset B or $\mathbb{R}$, the set $\{x; f(x) \in B\}$ is also a Borel subset of $\mathbb{R}$, is said to be Borel measurable. All continuous and piecewise continuous functions are Borel measurable.

Theorem 1.3 (Comparison of Riemann and Lebesgue Integrals)**.** *Let f be a bounded function defined on $\mathbb{R}$, and let $a < b$ be numbers.*

(i) The Riemann integral $\int_a^v f(x)dx$ is defined IFF the set of points x in $[a, b]$ where $f(x)$ is not continuous has Lebesgue measure 0.

(ii) If the Riemann integral $\int_a^v f(x)dx$ is defined, then f is Borel measurable (so the Lebesgue integral $\int_{[a,b]} f(x) d\mathcal{L}(x)$ is also defined), and the Riemann and Lebesgue integrals agree.

Note that any finite set of points in $\mathbb{R}$ has Lebesgue measure 0.

Definition 1.9 (Property Holding Almost Everywhere)**.** If the set of numbers in $\mathbb{R}$ that fail to have some property is a set with Lebesgue measures zero, we say that the property holds almost everywhere.

1.4 Convergence of Integrals

Definition 1.10 (Convergence Almost Surely)**.** Let $X_1, X_2, X_3, \ldots$ be a sequence of random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be another random variable defined on this space. We say that $X_1, X_2, X_3, \ldots$ converges to X almost surely and write

$$\lim_{n\to\infty} X_n = X \text{ almost surely}$$

if the set of $\omega \in \Omega$ for which the sequence of numbers $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$ has limit $X(\omega)$ is a set with probability one. Equivalently, the set of $\omega \in \Omega$ for which the sequence of numbers $X_1(\omega), X_2(\omega), X_3(\omega), \ldots$ does not converge to $X(\omega)$ and is a set with probability zero.

For example, in the infinite coin-toss probability space, where we assign a 1 for heads and 0 for tails, the partial sum of outcomes divided by the number of trials converges to $1/2$ almost surely, even though there are an *uncountable* many number of sequences that do not converge to $1/2$.

Definition 1.11 (Convergence Almost Everywhere)**.** Let $f_1, f_2, f_3, \ldots$ be a sequence of real-valued, Borel-measurable functions defined on $\mathbb{R}$. Let f be another real-valued, Borel-measurable function defined on $\mathbb{R}$. We say that $f_1, f_2, f_3, \ldots$ converges to f almost everywhere and write:

$$\lim_{n\to\infty} X_n = X \text{ almost everywhere}$$

if the set of $x \in \mathbb{R}$ for which the sequence of numbers $f_1(x), f_2(x), f_3(x), \ldots$ does not have a limit $f(x)$ is a set with Lebesgue measure zero.

Theorem 1.4 (Monotone convergence)**.** *Let$X_1, X_2, X_3, \ldots$ be a sequence of random variables converging almost surely to another random variable X. If*

$$0 \le X_1 \le X_2 \le X_2 \le \ldots \text{ almost surely}$$

then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$$

Similarly, let $f_1, f_2, f_3, \ldots$ be a sequence of Borel-measurable functions on $\mathbb{R}$ converging almost everywhere to a function f. If

$$0 \le f_1 \le f_2 \le f_3 \ldots \text{ almost everywhere}$$

then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)dx = \int_{-\infty}^{\infty} f(x)dx$$

Theorem 1.5 (Dominated convergence)**.** *Let $X_1, X_2, \ldots$ be a sequence of random variables converging almost surely to a random variable X. If there is another random variable Y such that $\mathbb{E}[Y] < \infty$ and $|X_n| \le Y$ almost surely for every n, then*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}X$$

Let $f_1, f_2, \ldots$ be a sequence of Borel-measurable functions on $\mathbb{R}$ converging almost everywhere to a function f. If there is another function g such that $\int_{-\infty}^{\infty} g(x)dx < \infty$ and $|f_n| \le g$ almost everywhere for every n, then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)dx = \int_{-\infty}^{\infty} f(x)dx$$

1.5 Computation of Expectations

Theorem 1.6 (Relating Integrals over $\mathbb{R}$ to integrals over Ω)**.** *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let g be a Borel-measurable function on $\mathbb{R}$. Then:*

$$\mathbb{E}\,|g(X)| = \int_{\mathbb{R}} |g(x)|\, d\mu_X(c),$$

and if this quantity is finite, then

$$\mathbb{E}f(X) = \int_{\mathbb{R}} f(x) d\mu_X(x)$$

This theorem tells us that in order to compute the Lebesgue integral $\mathbb{E}X$ over the abstract space Ω, it suffices to compute the integral $\int_{\mathbb{R}} f(x) d\mu_X(x)$ over the set of reals. We are also able to use densities to compute expectations, in addition to measures according to a similar notion:

Theorem 1.7 (Expectations Using Densities)**.** *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let g be a Borel-measurable function on $\mathbb{R}$. Suppose that X has a desnity f. Then:*

$$\mathbb{E}|g(X)| = \int_{-\infty}^{\infty} |g(x)| f(x),$$

and if this quantity is finite, then

$$\mathbb{E}f(X) = \int_{-\infty}^{\infty} g(x) f(x)$$

1.6 Change of Measure

Theorem 1.8 (Change of Measure)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely nonnegative random variable with $\mathbb{E}Z = 1$. For $A \in \mathcal{F}$, define*

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega)$$

Then $\tilde{\mathbb{P}}$ is a probability measure. Furthermore, if X is a nonnegative random variable, then

$$\tilde{\mathbb{E}}X = \mathbb{E}[XZ]$$

If Z is almost surely strictly positive, we also have:

$$\mathbb{E}Y = \tilde{\mathbb{E}}\left[\frac{Y}{Z}\right]$$

for every nonnegative random variable Y

Definition 1.12 (Equivalence of Measures)**.** Let Ω be a nonempty set and $\mathcal{F}$ a σ-algebra of subsets of Ω. Two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ are said to be equivalent if they agree which sets in $\mathcal{F}$ have probability zero.

Definition 1.13 (Randon-Nikodym derivative)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\tilde{\mathbb{P}}$ be another probability measure on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$ and let Z be an almost surely positive random variable that relates $\mathbb{P}$ and $\tilde{\mathbb{P}}$ via 1.63. Then Z is called the Radon-Nikodym derivate of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$ and we write:

$$Z = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}$$

Definition 1.14 (Radon-Nikodym)**.** Let $\mathbb{P}$ and $(\Omega, \mathcal{F}, \mathbb{P})$ be equivalent probability measures defined on $(\Omega, \mathcal{F})$. Then there exists an almost surely positive random variable Z such that $\mathbb{E}Z = 1$ and:

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega) \text{ for every } A \in \mathcal{F}$$

1.7 Summary

- Probability begins with the notion of a probability space.
  - Ω is the set of all possible outcomes of a random experiemnt.
  - $\mathcal{F}$ is the collection of subsets of Ω whose probability is defined.
  - $\mathbb{P}$ is a function mapping $\mathcal{F}$ to $[0, 1]$.
- Two Axioms of probability spaces:
  1. $\mathbb{P}(\Omega) = 1$

2. For <u>countable</u>, <u>disjoint</u> sets, $\mathbb{P}(\bigcup X_i) = \sum_i \mathbb{P}(X_i)$

- $\mathcal{F}$ is a σ-algebra, which satisfies 3 conditions:
  1. It contains $\varnothing$
  2. It is closed under countable complements
  3. It is closed under countable unions.
- While σ-algebras can be defined on any field, a *Borel* σ-algebra, $\mathcal{B}(\mathbb{R})$ is the smallest σ-alebgra that contains all the closed interval $[a, b]$ in $\mathbb{R}$.
- A *random variable* X is a mapping from $\Omega \mapsto \mathbb{R}$.
  - Has property that, $\forall B \in \mathcal{B}(\mathbb{R}), \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$.
    * Intuitively, this is saying that the sets that X maps to in the codomain $\mathcal{B}(\mathbb{R})$ belong to sets in the domain $\mathcal{F}$. (X is an injective mapping from $\mathcal{F}$ to $\mathcal{B}(\mathbb{R})$).
- $\mathbb{P}$ is a probability measure on Ω, which, with X, determines a *distribution*.
  - Different r.v.'s can have the same distribution. The same r.v. can have different distributions.
- The distribution is described as a measure, μ_X that assigns to each Borel subset B of $\mathbb{R}$ the mass $\mu_X(B) = \mathbb{P}(x \in B)$.
  - if X has density $f(x)$, then $\mu_X(B) = \int_B f(x)dx$.
- The *expectation* of a r.v. X is $\mathbb{E}X = \int_\Omega X(\Omega)d\mathbb{P}(\Omega)$, where the RHS is a Lebesgue integral. Lebesgue integrals have the properties:
  1. (Compairison) if $X \leq Y$ almost surely, then $\mathbb{E}X \leq \mathbb{E}Y$
  2. (Linearity) $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}X + \beta\mathbb{E}Y$
  3. (Convexity Inequality) $\phi(\mathbb{E}X) \leq \mathbb{E}(\phi(X))$
- A probability density, $f(x)$, maps the Borel subsets to the relative likelihoods.
  - $\mathbb{E}X = \int_\infty^\infty xf(x)dx$. More generally, $\mathbb{E}g(X) = \int_\infty^\infty g(x)f(x)dx$
  - It is not always true that $\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}X_n$ when a sequence of random variables converges. But it *does* hold true when the sequence converges monotonically.
- We can only define a change of probability measure, $\mathbb{P} \to \tilde{\mathbb{P}}$. If $\mathbb{P}$ is a p. measure and Z is a nonnegative r.v. satisfying $\mathbb{E}Z = 1$, then $\tilde{\mathbb{P}}$ is also a p. measure if:

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)d\mathbb{P}(\omega) \ \forall \ A \in \mathcal{F}$$

  - When Z is strictly positive,, the measures are *equivalent*.

1.8 Problems

2 Chapter 2

2.1 Information and σ-algebras

Definition 2.1 (Resolved)**.** Sets that are *resolved* by information are sets that, conditional on the information provided, allow us to know whether or not the true ω is part of it.

As time moves forward, the resolution increases. In other words, the set of resolved sets after three coin tosses *includes* all the sets resolved after two coin tosses (we are only increasing the information). Specifically, if $n < m$, $\mathcal{F}_m$ contains every set in $\mathcal{F}_n$ and more. Hence the collection of increasing σ-algebras, $\mathcal{F}_1, \mathcal{F}_2, \cdots$ is called a *filtration*.

Definition 2.2 (filtration)**.** Let Ω be a nonempty set. Let T be a fixed positive number, and assume that for each $t \in [0, T]$ there is a σ-algebra $\mathcal{F}(t)$. Assume further that if $s < t$, then every set in $\mathcal{F}(s)$ is also in $\mathcal{F}(t)$. Then we call the collection of σ-algebras $\mathcal{F}(t), 0 \le t \le T$ a filtration.

Definition 2.3 (σ-algebra generated by X)**.** Let X be a random variable defined on a nonempty sample space Ω. The σ-algebra generated by X, denoted $\sigma(X)$, is the collection of all subsets of Ω of the form $\{X \in B\}$ where B ranges over the Borel subsets of $\mathbb{R}$

Definition 2.4 (Measurable)**.** Let X be a random variable defined on a nonempty sample space n. Let g be a σ-algebra of subsets of n. If every set in $\sigma(X)$ is also in $\mathcal{G}$, we say that X is $\mathcal{G}$-measurable.

Definition 2.5 (Adapted Stochastic Process)**.** Let Ω be a nonempty sample space equipped with a filtration $\mathcal{F}(t), 0 \le t \le T$. Let $X(t)$ be a collection of random variables indexed by $t \in [0, T]$. We say this collection of random variables is an *adapted stochastic process* if, for each t, the random variable $X(t)$ is $\mathcal{F}(t)-$*measurable*.

2.2 Independence

Independence can be affected by changes in the probability measure; measurability is not.

Two sets are independent if:

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

Intuitively, independence of sets A and B means that knowing that the outcome ω of a random experiment is in A does not change our estimation of the probability that it is in B.

Definition 2.6 (Independence)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$ (i.e., the sets in $\mathcal{G}$ and the sets in $\mathcal{H}$ are also in $\mathcal{F}$). We say these two σ-algebras are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \;\forall\; A \in \mathcal{G}, B \in \mathcal{H}$$

Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say these two random variables are independent if the σ-algebras they generate, $\sigma(X)$ and $\sigma(Y)$, are independent. We say that the random variable X is independent of the σ-algebra $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.

Theorem 2.1 (Independence of Functions of RVs)**.** *Let X and Y be independent random variables, and let f and g be Borel-measurable functions on $\mathbb{R}$. Then $f(X)$ and $g(Y)$ are independent random variables.*

Definition 2.7 (Joint Distribution)**.** Let X and Y be random variables. The pair of random variables (X, Y) takes values in the plane $\mathbb{R}^2$, and the joint distribution measure of (X, Y) is given by:

$$\mu_{X,Y}(C) = \mathbb{P}\{(X, Y) \in C\} \text{ for all Borel sets } C \subset \mathbb{R}^2$$

This is a probability measure (i.e., a way of assigning measure between 0 and 1 to subsets of $\mathbb{R}^2$ so that $\mu_{X,Y}(\mathbb{R}^2) = 1$ and the countable additivity property is satisfied. The joint cumulative distribution function of (X, Y) is:

$$F_{X,Y}(a, b) = \mu_{X,Y}((-\infty, a] \times (-\infty, b]) = \mathbb{P}\{X \le a, Y \le b\}, a \in \mathbb{R}, b \in \mathbb{R}$$

We say that a nonnegative, Borel-measurable function $f_{X,Y}(x, y)$ is a joint density for the pair of random variables (X, Y) if

$$\mu_{X,Y}(C) = \int_{\infty}^{\infty}\int_{\infty}^{\infty} \mathbb{I}_C(x, y) f_{X,Y}(x, y) dy dx \text{ for all Borel sets} C \subset \mathbb{R}^2$$

Theorem 2.2 (Independence & Joint Distributions)**.** *Let X and Y be random variables. The following conditions are equivalent.*

(i) X and Y are independent.

(ii) The joint distribution measure factors:

$$\mu_{X,Y}(A \times B) = \mu_X(A) \cdot \mu_Y(B) \;\forall\; \text{ Borel subsets } A \subset \mathbb{R}, B \subset \mathbb{R}$$

(iii) *The joint cumulative distribution measure factors:*

$$F_{X,Y}(a,b) = F_X(a) \cdot F_Y(b) \; \forall \;\; \textit{Borel subsets } a \in \mathbb{R}, b \in \mathbb{R}$$

(iv) *The joint moment-generating function factors*

$$\mathbb{E}e^{uX+vY} = \mathbb{E}e^{uX} + \mathbb{E}e^{vY}$$

for all $u, v \in \mathbb{R}$ *for which the expectations are finite.*

If there is a joint density, each of the conditions above is equivalent to the following.

(v) *The joint density factors:*

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \text{ for almost every } x, y \in \mathbb{R}$$

The conditions above imply but are not equivalent to the following.

(vi)

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

provided $\mathbb{E}[XY] < \infty$

Definition 2.8 (Variance, Covariance, St. Dev)**.** Let X be a random variable whose expected value is defined. The variance of X, denoted $\mathrm{Var}(X)$ is

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

Because $(X - \mathbb{E}X)^2$ is nonnegative, $\mathrm{Var}(X)$ is always defined, although it may be infinite. The standard deviation of X is $\sqrt{\mathrm{Var}(X)}$. The linearity of expectations show that:

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

Let Y be another random variable and assume that $\mathbb{E}X, \mathrm{Var}X, \mathbb{E}Y,$ and $\mathrm{Var}(Y)$ are all finite. The covariance of X and Y is:

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

The linearity of expectations shows that:

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y$$

In particular,, $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$ IFF $\mathrm{Cov}(X,Y) = 0$. Assume, in addition to the finiteness of expectations and variances, that $\mathrm{Var}(X) > 0$ and $\mathrm{Var}(Y) > 0$.The correlation coefficient of X and Y is:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

If $\rho(X,Y) = 0$ (or, equivalently, $\mathrm{Cov}(X,Y) = 0$), we say that X and Y are uncorrelated.

Linear combinations of jointly normal random variables (i.e., sums of constants time the random variables) are jointly normal.

2.3 General Conditional Expectations

Definition 2.9 (Conditional Expectations)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ and let X be a random variable that is either nonnegative or integrable. The conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}[X|\mathcal{G}]$ is any random variable that satisfies:

(i) (Measurability) $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$ measurable, and

(ii) (Partial Averaging)

$$\int_A \mathbb{E}[X|\mathcal{G}](\omega)d\mathbb{P}(\omega) = \int_A X(\omega)d\mathbb{P}(\omega) \ \forall \ A \in \mathcal{G}$$

If $\mathcal{G}$ is the σ−algebra generated by some other random variable W, i.e., $\mathcal{G} = \sigma(W)$, we generally write $\mathbb{E}[X|W]$ rather than $\mathbb{E}[X|\sigma(W)]$.

Theorem 2.3 (Properties of Conditional Expectations)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.*

(i) (Linearity of conditional expectations) If X and Y are integrable random variables, and $c_1 + c_2$ are constnats, then:

$$\mathbb{E}[c_1X + c_2Y|\mathcal{G}] = c_1\mathbb{E}[X|\mathcal{G}] + c_2\mathbb{E}[Y|\mathcal{G}]$$

(This equation also holds if we assume that X and Y are nonnegative (rather than itnegrable) and c_1 and c_2 are positive, although both sides may be $+\infty$).

(ii) (Taking out what is known) If X and Y are integrable random variables, Y and XY are integrable, and X is $\mathcal{G}$−measurable then:

$$\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|G]$$

This equation also holds if we assume that X is positive and Y is nonnegative (rather than integrable), althouh both sides may be $+\infty$.

(iii) (Iterated conditioning) If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$ ($\mathcal{H}$ contains less information than $\mathcal{G}$) and X is an integrable random variable, then:

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$$

This equation also holds if we assume that X is positive and Y is nonnegative (rather than integrable), althouh both sides may be $+\infty$.

(iv) (Independence) If X is integrable and independent of $\mathcal{G}$ then:

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$$

This equation also holds if we assume that X is positive (rather than integrable), althouh both sides may be $+\infty$.

(v) (Conditional Jensen's Inequality) If $\varphi(x)$ is a convex function of a dummy variable x and X is integrable, then:

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbb{E}[X|\mathcal{G}])$$

Definition 2.10 (Martingale)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t), 0 \leq t \leq T$ be a filtration of sub-σ-algebras of $\mathcal{F}$. Consider an adapted stochastic process $M(t), 0 \leq t \leq T$.

i If

$$\mathbb{E}[M(t)|\mathcal{F}(s)] = M(s) \ \forall - \leq s \leq t \leq T$$

we say this process is a martingale. It has no tendency to rise or fall.

ii If

$$\mathbb{E}[M(t)|\mathcal{F}(s)] \geq M(s) \ \forall - \leq s \leq t \leq T$$

we say this process is a submartingale. It has no tendency to fall; it may have a tendency to rise.

iii If

$$\mathbb{E}[M(t)|\mathcal{F}(s)] \leq M(s) \ \forall - \leq s \leq t \leq T$$

we say this process is a supermartingale. It has no tendency to rise; it may have a tendency to fall.

Definition 2.11 (Markov Process)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let T be a fixed positive number, and let $\mathcal{F}(t), 0 \le t \le T$ be a filtration of sub-σ-algebras of $\mathcal{F}$. Consider an adapted stochastic process $X(t), 0 \le t \le T$. Assume that for al $0 \le s \le t \le T$ and for every nonnegative, Borel measurable function f there is another Borel measurable function g such that:

$$\mathbb{E}[f(X(t))|\mathcal{F}(s)] = g(X(s))$$

Then we say that X is a Markov process.

2.4 Summary

- Thinking about information
  - A random experiment is performed, and an outcome ω is determined, but the value of ω is not revealed. Instead, for each set in the σ-algebra $\mathcal{G}$, we are told whether ω is in the set. The more sets there are on $\mathcal{G}$, the more information this provides. If $\mathcal{G}$ is the trivial σ-algebra containing only $\varnothing$ and Ω, this provides no information.
- $\mathcal{G}$–measurable
  - An r.v. X is $\mathcal{G}$–measurable IFF $\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\}$ is in $\mathcal{G}$ for every Borel subset in $\mathbb{R}$.
  - In this case, the information in $\mathcal{G}$ is enough to determine the value of the random variable $X(\omega)$, even though it may not be enough to determine the value ω of the outcome of the random experiment.
- Independence
  - This occurs when the information in the σ–algebra is irrelevant to X.
  - X and $\mathcal{G}$ are indepdendent IFF $\forall\ A \in \mathcal{G}, B \in \mathbb{R}$, we have:

$$\mathbb{P}\{\omega \in \Omega; \omega \in A \text{ and } X(\omega) \in B\} = \mathbb{P}(A) \cdot \mathbb{P}\{\omega \in \Omega; X(\omega) \in B\}$$

  - Independence implies uncorrelatedness, but uncorrelated random variables do not need to be independent.
  - Jointly normally distributed random variables are uncorrelated IFF they are indepdendent, but normally distributed random variables do not need to be jointly normal.
- Conditional Expectation
  - This occurs when $\mathcal{G}$ tells us some information about X, but not enough to estimate it. $\mathbb{E}[X|\mathcal{G}]$ is a random variable.
  - Conditional expectations behave in many ways like expectations, except that expectations do not depend on ω and conditional expectations do.
- Information in Continuous Time:
  - We usually have $(\Omega, \mathcal{F}, \mathbb{P})$ where we interpret $\mathcal{F}(t)$ as the information available at time t. Within this context, we have an *adapted stochastic process*: a collection of rvs $\{X(t); 0 \le t \le T\}$ such that for every t, $X(t)$ is $\mathcal{F}(t)$ measurable.
- Martingale
  - Informally, a martingale is a process such that its future expectation is equal to its current value.
- Markov Process
  - Informally, a Markov Process is a process such that our estimate of $f(X(t))$ made at time s depends only on process value $X(s)$ at time s, and not on the path of the series before time s.

3 Brownian Motion

3.1 Introduction

The most importanat properties of borwnian are that (i) it is a martingale and (ii) that it accumulates quadratic variation at rate one per unit time.

3.2 Scaled Random Walks

3.2.1 Symmetric Random Walk

If we define the sequence of successive outcomes of tosses by $\omega = \omega_1\omega_2\omega_3\ldots$, with

$$X_j = \begin{cases} 1 \text{ if } \omega_j - H, \\ -1 \text{ if } \omega_j = T \end{cases}$$

and $M_9 = 0$,

$$M_k = \sum_{j=1}^{k} X_j, k = 1, 2, \cdots$$

The process $M_k, k = 0, 1, 2, \ldots$ is a *symmetric random walk* (symmetric because on each toss, you are equally likely to move up or down). In other words, the random walk is the partial sum of the discrete innovations.

3.2.2 Increments of the Symmetric Random Walk

A random walk has *independent increments*. Each of the random variables,

$$M_{k_{i+1}} - M_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} X_j$$

is an *increment* of the random walk. Note that $\mathbb{E}[M_{k_{i+1}} - M_{k_i}] = 0, \mathrm{Var}[M_{k_{i+1}} - M_{k_i}] = k_{i+1} - k_i$. To see why the variance is this, consider that $\mathrm{Var}(X_j) = \mathbb{E}[X_j^2] - \mathbb{E}[X]^2$. But $\mathbb{E}[X] = 0$, and $X^2 = 1 \;\forall\; \omega \implies \mathbb{E}[X^2] = 1$. So $\mathrm{Var}(X_j) = 1$. Also recall that for independent variables, the variance of a sum is the sum of the variances. The implication of this is that the variance of the increment over any time interval k to l for nonnegative integers, $k < l$ is $l - k$.

3.2.3 Martingale Property for the Symmetric Random Walk

We show that symmetric random walk is martingale:

$$\begin{aligned}
\mathbb{E}[M_l|\mathcal{F}_k] &= \mathbb{E}[(M_l - M_k) + M_k|F_k] \\
&= \mathbb{E}[(M_l - M_k)|\mathcal{F}_k] + \mathbb{E}[M_k|\mathcal{F}_k] \\
&= \mathbb{E}[(M_l - M_k)|\mathcal{F}_k] + M_k \\
&= \mathbb{E}[(M_l - M_k)] + M_k = M_k
\end{aligned}$$

The key step is the move from lines (3) to (4), which uses independence of the increment from $\mathcal{F}_k$.

3.2.4 Quadratic Variation of the Symmetric Random Walk

The *quadratic variation* of the symmetric random walk up to time k is defined to be:

$$[M, M]_k = \sum_{j=1}^{k} (M_j - M_{j-1})^2 = k$$

In other words, the quadratic variation is the partial sum of the increments (it is path-dependent). However, since in our case, the increment is ± 1, the increment squared is 1, and so the quadratic variation is simply $\sum_{j=1}^{k} 1 = k$. Even though in this case above, the quadratic variation evaluates to the same as the variance, they are different: $\mathrm{Var}(M_k)$ is computed by taking an average over all paths, taking probabilities into account. $[M, M]_k$ is computed along a single path; probabilities do not matter (realizations do!). Variances are computed theoretically; quadratic variations are path dependent.

3.2.5 Scaled Symmetric Random Walk

To speed up time / scale down step of a symmetric random walk, we define a *scaled symmetric walk*:

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt}$$

for a fixed positive integer n. The Brownian motion is found in the limit as $n \to \infty$. The scaled random walk also has independent increments. As a result, the same martingale property obtains for the scaled random walk:

$$\begin{aligned}
W^{(n)}(t) &= \left(W^{(n)}(t) - W^{(n)}(s)\right) + W^{(n)}(s) \\
\mathbb{E}[W^{(n)}(t)|\mathcal{F}(s)] &= \mathbb{E}[\left(W^{(n)}(t) - W^{(n)}(s)\right)|\mathcal{F}(s)] + \mathbb{E}[W^{(n)}(s)|\mathcal{F}(s)] \\
&= \mathbb{E}[W^{(n)}(t) - W^{(n)}(s)] + W^{(n)}(s) \\
&= 0 + W^{(n)}(s) = W^{(n)}(s)
\end{aligned}$$

We consider the quadratic variation of the *scaled symmetric random walk*:

$$\begin{aligned}
\left[W^{(n)}, W^{(n)}\right](t) &= \sum_{j=1}^{nt} \left[W^{(n)}\left(\frac{j}{n}\right) - W^{(n)}\left(\frac{j-1}{n}\right)\right]^2 \\
&= \sum_{j=1}^{nt} \left[\frac{1}{\sqrt{n}} X_j\right]^2 = \sum_{j=1}^{nt} \frac{1}{n} = t
\end{aligned}$$

Again, over the path from time 0 to time t, we obtain a quadratic variation equal to t.

3.2.6 Limiting Distribution of the Scaled Random Walk

Theorem 3.1 (Central Limit Theorem)**.** *Fix $t \geq 0$. As $n \to \infty$, the distribution of the scaled random walk $W^{(n)}(t)$ evaluated at time t converges to the normal distribution with mean zero and variance t.*

3.2.7 Log-Normal Distribution as the Limit of the Binomial Model

We use the CLT to show that the limit of a properly scaled binomial asset-pricing model leads to a stock price with a log-normal distribution. In other words, the binomial model is a discrete=time version of the geometric Brownian motion model.

Theorem 3.2. *As $n \to \infty$, the distribution of $S_n(t)$ converges to the distribution of*

$$S(t) = S(0) \exp\left\{\sigma W(t) - \frac{1}{2}\sigma^2 t\right\}$$

This distribution is called log-normal.

3.3 Brownian Motion

3.3.1 Definition of Brownian Motion

Definition 3.1 (Brownian Motion)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose there is a continuous function $W(t)$ of $t \geq 0$ that satisfies $W(0) = 0$ and that depends on ω. Then $W(t), t \geq 0$, is a Brownian motion if $\forall\ 0 = t_0 < t_1 < \cdots < t_m$ the increments

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), \ldots, W(t_m) - W(t_{m-1})$$

are independent and each of these increments is normally distributed with:

$$\mathbb{E}[W(t_{i+1}) - W(t_i)] = 0,$$
$$\text{Var}[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i$$

Not only is $W(t) = W(t) - W(0)$ normally distributed for each t, but the increments $W(t) - W(s)$ are normally distributed for $0 \leq s < t$. One can think of ω in the experiment as the Brownian motion path itself.

3.3.2 Distribution of Brownian Motion

Since the individual innovations of the Brownian motion are independent and normally distributed, they are jointly normally distributed. Hence we can compute the moments of the joint distribution. We begin with the covariance of the innovations. (Recall that $\mathbb{E}[W(s)] = \mathbb{E}[W(t)] = 0$

$$\begin{aligned}\mathbb{E}[W(s)W(t)] &= \mathbb{E}[W(s)(W(t) - W(s)) + W^2(s)] \\ &= \mathbb{E}[W(s)] \cdot \mathbb{E}[W(t) - W(s)] + \mathbb{E}[W^2(s)] \\ &= 0 + \text{Var}[W(s)] = s\end{aligned}$$

Interestingly, this leads to the following variance-covariance matrix for Brownian motion:

$$\begin{bmatrix} \mathbb{E}[W^2(t_1)] & \mathbb{E}[W(t_1)W(t_2)] & \cdots & \mathbb{E}[W(t_1)W(t_m)] \\ \mathbb{E}[W(t_2)W(t_1)] & \mathbb{E}[W^2(t_2)] & \cdots & \mathbb{E}[W(t_2)W(t_m)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[W(t_m)W(t_1)] & \mathbb{E}[W(t_m)W(t_2)] & \cdots & \mathbb{E}[W^2(t_m)] \end{bmatrix} = \begin{bmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \tag{3}$$

Theorem 3.3 (Alternative characterizations of Brownian motion)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose thre is a continuous function $W(t)$ of $t \geq 0$ that satisfies $W(0) = 0$ and that depends on ω. The following three properties are equivalent.*

(i) For all $0 = t_0 < t_1 < \ldots < t_m$, the increments

$$W(t_1) = W(t_1) - W(t_0), W(t_2) - W(t_1), \ldots, W(t_m) - W(t_{m-1})$$

are independent and each of these increments is normally distribtued with mean 0 and variance given by $\text{Var}[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i$.

(ii) For all $0 = t_0 < t_1 < \ldots < t_m$, the random variables $W(t_1), W(t_2), \ldots, W(t_m)$ are jointly normally distributed with means equal to zero and covariance matrix given by equation 3.

(iii) For all $0 = t_0 < t_1 < \ldots < t_m$, the random variables $W(t_1), W(t_2), \ldots, W(t_m)$ have a joint moment generating function given by (see textbook).

3.3.3 Filtration for Brownian Motion

Definition 3.2 (Filtration for Brownian Motion)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is defined a Brownian motion $W(t), t \geq 0.$. A filtration for the Brownian motion is a collection of σ-algebras, $\mathcal{F}(t), t \geq 0$ satisfying:

(i) (Information accumulates) For $0 \leq s < t$, every set in $\mathcal{F}(s)$ is also in $\mathcal{F}(t)$. In other words, there is at least as much information available at the later time $\mathcal{F}(t)$ as there is at the earlier time $\mathcal{F}(s)$.

(ii) (Adaptivity) For each $t \geq 0$, the Brownian motion W(t) at time t is $\mathcal{F}(t)$ measurable. In other words, the information at time t is sufficient to evaluate the Brownian motion $W(t)$ at that time.

(iii) (Independence of Future Increments) For $0 \leq t < u$, the increment $W(u) - W(t)$ is independent of $\mathcal{F}(t)$. In other words, any increment of the Brownian motion after time t is independent of the information available at time t.

Let $\Delta(t), t \geq 0$ be a stochastic process. We say that $\Delta(t)$ is adapted to the filtration $\mathcal{F}(t)$ if for each $t \geq 0$ the random variable $\Delta(t)$ is $\mathcal{F}(t)$-measurable.

Note that equation (3) gives rise to efficient markets hypothesis, as it suggests that the information available at time t is of no use predicting the future movements of the Brownian motion.

3.3.4 Martingale Property for Brownian Motion

Theorem 3.4. *Brownian motion is a martingale.*

Proof. Let $0 \leq s < t$ be given.

$$\begin{aligned}\mathbb{E}[W(t)|\mathcal{F}(s)] &= \mathbb{E}[(W(t) - W(s)) + W(s)|\mathcal{F}(t)] \\ &= \mathbb{E}[W(t) - W(s)|\mathcal{F}(s)] + \mathbb{E}[W(s)|\mathcal{F}(s)] \\ &= \mathbb{E}[W(t) - W(s)] + \mathbb{E}[W(s)] \\ &= 0 + \mathbb{E}[W(s)] = W(s)\end{aligned}$$

□

3.4 Quadratic Variation

The paths of Brownian motion are unusual in that their quadratic variation is not zero. This is what makes stochastic calculus different from ordinary calculus!

3.4.1 First-Order Variation

To compute the first-order variation of a function, we are interested in the total amount of up and down oscillation undergone by a function between time 0 and T. We effectively compute:

$$FV_T(f) = \int_0^T |f'(t)|\, dt$$

3.4.2 Quadratic Variation

Definition 3.3 (Quadratic Variation)**.** Let $f(t)$ be a function defined for $0 \leq t \leq T$. The quadratic variation of f up to time T is:

$$[f, f](T) = \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2,$$

where $\Pi = \{t_0, t_1, \ldots, t_n\}$ and $0 = t_0 < t_1 < \ldots < t_n = T$

If the function has a continuous derivative, then the quadratic variation evaluates to 0.

Theorem 3.5. *Let W be a Brownian motion. Then $[W, W](T) = T$ for all $T \geq 0$ almost surely.*

In other words, the set of paths for which $[W, W](T) \neq T$ has zero probability. We conclude that Brownian motion accumulates quadratic variation at rate one per unit of time:

$$dW(t)dW(t) = dW^2(t) = dW(t)$$

Similarly,

$$dW(t)dt = 0, dt^2 = 0$$

3.4.3 Volatility of Geometric Brownian Motion

Let α and $\sigma > 0$ be constants, and define the geometric Brownian motion,

$$S(t) = S(0) \exp\left\{\sigma W(t) + \left(\alpha - \frac{1}{2}\sigma^2\right) t\right\}$$

3.5 Markov Property

Theorem 3.6 (Brownian Motion as Markov Process). *Let $W(t), t \geq 0$ be a Brownian motion and let $\mathcal{F}(t), t \geq 0$ be a filtration for this Brownian motion. Then $W(t), t \geq 0$ is a Markov process.*

Recall that, by definition, a Markov process is defined by the property that $\exists\, g$ s.t. :

$$\mathbb{E}[f(W(t))|\mathcal{F}(s)] = g(W(s))$$

Specifically we can write:

$$\mathbb{E}[f(W(t))|\mathcal{F}(s)] = \int_{-\infty}^{\infty} f(y)p(\tau, W(s), y)dy \tag{4}$$

where $p(\cdot)$ give the *transition density* defined by:

$$\frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(y-x)^2}{2\tau}}$$

We interpret equation 4 as follows: conditioned on information in $\mathcal{F}(s)$, the conditional density of $W(t)$ is given by $p(t, W(s), y)$. The density is normal with mean $W(s)$ and variance $\tau = t - s$. But the only information from $\mathcal{F}(s)$ that is relevant – the essence of the Markov property.

3.6 First Passage Time Distribution

We begin with a martingale containing Brownian motion in the exponential function. Fixing a constant σ, the *exponential martingale* corresponding to σ is given by:

$$Z(t) = \exp\left\{\sigma W(t) - \frac{1}{2}\sigma^2 t\right\} \tag{5}$$

Theorem 3.7 (Exponential Martingale). *Let $W(t), t \geq 0$, be a Brownian motion with a filtration $\mathcal{F}(t), t \geq 0$, and let σ be a constant. The process $Z(t), t \geq 0$, of equation (5) is a martingale.*

Proof. We have:

$$\begin{aligned}
\mathbb{E}[Z(t)|\mathcal{F}(s)] &= \mathbb{E}\left[\exp\left\{\sigma W(t) - \frac{1}{2}\sigma^2 t\right\}|\mathcal{F}(s)\right] \\
&= \mathbb{E}\left[\exp\left\{\sigma(W(t) - W(s))\right\} \cdot \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 t\right\}|\mathcal{F}(s)\right] \\
&= \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 t\right\} \cdot \mathbb{E}\left[\exp\left\{\sigma(W(t) - W(s))\right\}|\mathcal{F}(s)\right] \\
&= \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 t\right\} \cdot \mathbb{E}\left[\exp\left\{\sigma(W(t) - W(s))\right\}\right]
\end{aligned}$$

Since $W(t) - W(s)$ is normally distributed with mean 0 and variance $t - s$, this expectation is equal to is $\exp\left\{\frac{1}{2}\sigma^2(t-s)\right\}$. Substituting in, we get:

$$\mathbb{E}[Z(t)|\mathcal{F}(s)] = \exp\left\{\sigma W(s) - \frac{1}{2}\sigma^2 s\right\}$$

$\square$

We define the *first passage time* to level m as the first time the Brownian motion W reaches the level m, i.e. $\tau_m = \min\{t \geq 0; W(t) = m\}$

Theorem 3.8 (First Passage Time). *For $m \in \mathbb{R}$, the first passage time of Brownian motion to level m is finite almost surely, and the Laplace transform of its distribution is given by:*

$$\mathbb{E}e^{-\alpha\tau_m} = e^{-|m|\sqrt{2\alpha}} \;\forall\; \alpha > 0$$

3.7 Reflection Principle

3.7.1 Reflection Equality

We fix a positive level m and a positive time t. We wish to "count" the Brownian motion paths that reach level m at or before time t. For paths that reach m before t, but by time t are below, we consider the reflection equality:

$$\mathbb{P}\{\tau_m \leq t, W(t) \leq w\} = \mathbb{P}\{W(t) \geq 2m - w\}, w \leq m, m > 0$$

3.7.2 First Passage Time Distribution

Theorem 3.9. *For all $m \neq 0$, the random variable τ_m has cumulative distribution function:*

$$\mathbb{P}\{\tau_m \leq t\} = \frac{2}{\sqrt{2\pi}} \int_{\frac{|m|}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} dy, t \geq 0$$

and density:

$$f_{\tau_m}(t) = \frac{d}{dt}\mathbb{P}\{\tau_m \leq t\} = \frac{|m|}{\sqrt{t}} e^{-\frac{m^2}{2t}}, t \geq 0$$

3.7.3 Distribution of Brownian Motion and Its Maximum

We define the *maximum to date* for Brownian motion to be:

$$M(t) = \max_{-\leq s \leq t} W(s)$$

For positive m, we have $M(t) \geq mIFF\tau_m \leq t$. Hence we can rewrite the reflection equation as:

$$\mathbb{P}\{M(t) \geq M, W(t) \leq w\} = \mathbb{P}\{W(t) \geq 2m - w\}, w \leq m, m > 0$$

This allows us to describe the joint density:

Theorem 3.10 (Joint density of $M(t)$ and $W(t)$)**.** *For $t > 0$, the joint density of $(M(t), W(t))$ is*

$$f_{M(t),W(t)}(m, w) = \frac{2(2m - w)}{t\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}}, w \leq m, m > 0$$

Theorem 3.11 (Distribution of $M(t)|W(t)$)**.** *The conditional distribution of $M(t)$ given $W(t)$ is:*

$$f_{M(t)|W(t)}(m|w) = \frac{2(2m - w)}{t} e^{-\frac{2m(m-w)}{t}}$$

3.8 Summary

- Brownian Motion
  - continuous stochastic process, $W(t), t \geq 0$, with independent, normally distributed increments.
  - $\mathbb{E}[W(t_{i+1}) - W(t_i)] = 0, \mathrm{Var}[W(t_{i+1}) - W(t_i)] = t_{i+1} - t_i$
  - We associate Brownian motion with filtration $\mathcal{F}(t)$ s.t. $\forall\ t \geq 0, u \geq t, W(t)$ is $\mathcal{F}(t)$-measurable and $W(u) - W(t)$ is independent of $\mathcal{F}(t)$.
  - Both martingale and Markov process.
- Quadratic Variation Property
  - Brownian motion accumulates quadratic variation at a rate one per unit time. This is true *regardless of the path along which we do the computation*, i.e. $dW(t)dW(t) = dt$.
  - $dW(t)dt = dtdt = 0$

- First passage time of Brownian motion
  - Defined as the first time the Brownian motion reaches level m.
  - The r.v. τ_m has density
$$f_{\tau_m}(t) = \frac{|m|}{\sqrt{2\pi t}}$$
- Joint Density of $M(t), W(t)$
  -
$$f_{M(t),W(t)}(m,w) = \frac{2(2m-w)}{t\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}}, w \le m, m > 0$$

4 Stochastic Calculus

4.1 Introduction

Stochastic Calculus differs from ordinary calculus due to the fact that Brownian motion has a nonzero quadratic variation.

4.2 Itô's Integral for Simple Integrands

We fix a positive number T and seek to make sense of

$$\int_0^T \Delta(t) dW(t)$$

Here, $\Delta(t)$ is an adapted stochastic process. As a consequence, we require $\Delta(t)$ to be $\mathcal{F}(t)$−measurable for each $t \ge 0$. Since increments of W after time t are independent of $\mathcal{F}(t)$, it follows that $\Delta(t)$ is independent of future Brownian increments.

In ordinary calculus, with a differentiable function $g(t)$, we can define

$$\int_0^T \Delta(t) dg(t) = \int_0^t \Delta(t) g'(t) dt$$

But $g(t)$ is not differentiable in the case of Brownian motion.

4.2.1 Construction of the Integral

Let $\Pi = \{t_0, t_1, \ldots, t_n\}$ be a partition of $[0, T]$; assume that $\Delta(t)$ is constant in t on each subinterval $[t_j, t_{j+1})$. We call $\Delta(t)$ a simple process. We fix $f(t_j)$ throughout the entire interval. In general, for $t_k \le t \le t_{k+1}$ we can define

$$\begin{aligned} I(t) &= \sum_{j=0}^{k-1} \Delta(t_j)[W(t_{j+1}) - W(t_j)] + \Delta(t_k)[W(t) - W(t_k)] \\ &= \int_0^t \Delta(u) dW(u) \end{aligned}$$

4.2.2 Properties of the Integral

The Itô integral is defined as the gain from trading in the martingale $W(t)$. Since $W(t)$ is a martingale, it follows that $I(t)$ should have no tendency to rise or fall.

Theorem 4.1 (Itô Martingale)**.** *The Itô integral is a martingale.*

Because $I(t)$ is a martingale and $I(0) = 0$ we have that $\mathbb{E}I(t) = 0 \; \forall \; t \ge 0$, hence $\mathrm{Var} I(t) = \mathbb{E}I^2(t)$.

Theorem 4.2 (Itô Isometry)**.** *The Itô integral defined above satisfies*

$$\mathbb{E}I^2(t) = \mathbb{E}\int_0^t \Delta^2(u)du$$

Theorem 4.3 (QV Itô)**.** *The quadratic variation accumulated up to time t by the Itô integral is*

$$[I,I](t) = \int_0^t \Delta^2(u)du$$

We also note that the Itô integral accumulates quadratic variation at a rate $\Delta^2(t)$ per unit time.

$$dI(t)dI(t) = \Delta^2(t)dW(t)dW(t) = \Delta^2(t)dt$$

Also,

$$dI(t) = \Delta(t)dW(t)$$

Intuitively, this equation says that when we move forward a little bit in time from time t, the change in the Itô integral is $\Delta(t)$ times the change in the Brownian motion W.

4.3 Itô's Integral for General Integrands

Here we no longer assume that $\Delta(t)$ is a simple process. We do, however, assume the filtration process and the square integrability condition that:

$$\mathbb{E}\int_0^T \Delta^2(t)dt < \infty$$

We then construct the integral as the limit of simple processes:

$$\int_0^t \Delta(u)dW(u) = \lim_{n\to\infty}\int_0^t \Delta_n(u)dW(u), 0 \le t \le T$$

Theorem 4.4 (Itô Integral Properties)**.** *Let T be a positive constant and let $\Delta(t), 0 \le t \le T$ be an adapted stochastic process that satisfies the square integrability condition. Then $I(t) = \int_0^t \Delta(u)dW(u)$ has the following properties:*

(i) (Continuity) As a function of the upper limit of integration t, the paths of $I(t)$ are continuous

(ii) (Adaptivity) For each t, $I(t)$ is $\mathcal{F}(t)$−measurable.

(iii) (Linearity) If $I(t) = \int_0^t \Delta(u)dW(u)$ and $J(t) = \int_0^t \Gamma(u)dW(u)$, then

$$I(t) \pm J(t) = \int_0^t (\Delta(u) \pm \Gamma(u))dW(u);$$

furthermore for every constant c, $cI(t) = \int_0^t c\Delta(u)dW(u)$

(iv) $I(t)$ is a martingale.

(v) (Itô isometry) $\mathbb{E}I^2(t) = \mathbb{E}\int_0^t \Delta^2(u)du$.

(vi) (Quadratic Variation) $[I,I](t) = \int_0^t \Delta^2(u)du$

One way we see the effect of quadratic drift in action:

$$\int_0^t W(u)dW(u) = \frac{1}{2}W^2(t) - \frac{1}{2}t, t \ge 0$$

Since $\mathbb{E}W^2(t) = t$, if the term $-\frac{1}{2}t$ term were not present, we would not have a martingale.

4.4 Itô-Doeblin Formula

4.4.1 Formula for Brownian Motion

In finance, we are interested in differentiating functions of Brownian motion, of the form $f(W(t))$. If $W(t)$ were also differentiable, we could apply chain rule:

$$df(W(t)) = f'(W(t))W'(t)dt = f'(W(t))dW(t)$$

Because of QV, we instead have:

$$df(W(t)) = f'(W(t))dW(t) + \frac{1}{2}f''(W(t))dt$$

If we integrate, we obtain:

$$f(W(t)) - f(W(0)) = \int_0^t f'(W(u))dW(u) + \frac{1}{2}\int_0^t f''(W(u))du$$

Note that the second term on the right is an ordinary (Lebesgue) integral w.r.t the time variable.

Theorem 4.5 (Itô-Doeblin formula for Brownian motion)**.** *Let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x).f_X(T,x), andf_{xx}(t,x)$ are defined and continuous, and let $W(t)$ be a Brownian motion. The, for every $T \geq 0$,*

$$\begin{aligned} f(T, W(T)) = f(0, W(0) + \int_0^T f_t(t, W(t)) \\ + \int_0^T f_x(t, W(t))dW(t) + \frac{1}{2}\int_0^T f_{xx}(t, W(t))dt \end{aligned}$$

4.4.2 Formula for Itô Processes

Here we extend the Itô-Doeblin to stochastic processes more general than Brownian motions

Definition 4.1 (Ito Processes)**.** Let $W(t), t \geq 0$ be a Brownian motion, and let $\mathcal{F}(t), t \geq 0$, be an associated filtration. An Itô process is a stochastic process of the form:

$$X(t) = X(0) + \int_0^t \Delta(u)dW(u) + \int_0^t \Theta(u)du$$

where $X(0)$ is nonrandom and $\Delta(u)$ and $\Theta(u)$ are adapted stochastic processes.

Lemma 1 (QV of Itô Processes)**.** *The quadratic variation of the Itô Process is:*

$$[X, X](t) = \int_0^t \Delta^2(u)du$$

Definition 4.2 (Integral w.r.t. Itô Process)**.** Let $X(t), t \geq 0$, be an Itô process, and let $\Gamma(t), t \geq 0$ be an adapted process. We define the integral with respect to an Itô process

$$\int_0^t \Gamma(u)dX(u) = \int_0^t \Gamma(u)\Delta(u)dW(u) + \int_0^t \Gamma(u)\Theta(u)du$$

Theorem 4.6 (Itô-Doeblin formula for an Itô process)**.** *. Let $X(t), t \geq 0$ be an Itô process, and let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x)$, $f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous. Then, for every $T \geq 0$,*

$$\begin{aligned} f(T, X(T)) &= f(0, X(0)) + \int_0^T f_t(t, X(t))dt + \int_0^T f_x(t, X(t))dX(t) \\ &\quad + \frac{1}{2}\int_0^T f_{xx}(t, X(t))d[X, X](t) \\ &= f(0, X(0)) + \int_0^T f_t(t, X(t))dt + \int_0^T f_x(t, X(t))\Delta(t)dW(t) \\ &\quad + \int_0^T f_x(t, X(t))\Theta(t)dt + \frac{1}{2}\int_0^T f_{xx}(t, X(t))\Delta^2(t)dt \end{aligned}$$

We may rewrite this in differential notation as:

$$df(t,X(t)) = f_t(t,X(t))dt + f_x(t,X(t))dX(t) + \frac{1}{2}f_{xx}(t,X(t))dX(t)dX(t)$$

Intuitively, we take the Taylor series expansion of f w.r.t all its arguments, with first order expansion for all arguments with 0 quadratic variation (i.e. t) and second order for every argument with nonzero quadratic variation (i.e. $X(t)$). In fact, we can further simplify to:

$$df(t,X(t)) = f_t(t,X(t))dt + f_x(t,X(t))\Delta(t)dW(t) + f_x(t,X(t))\Theta(t)dt + \frac{1}{2}f_{xx}(t,X(t))\Delta^2(t)dt$$

4.4.3 Examples

Example 4.1 (Generalized geometric Brownian motion)**.** . Let $W(t), t \geq 0$ be a Brownian motion, let $F(t), t \geq 0$ be an associated filtration, and let $\alpha(t)$ and $\sigma(t)$ be adapted processes. Define the Itô process:

$$X(t) = \int_0^t \sigma(s)dW(s) + \int_0^t (\alpha(s) - \frac{1}{2}\sigma^2(s))ds$$

Then

$$dX(t) = \sigma(t)dW(t) + (\alpha(t) - \frac{1}{2}\sigma^2(t))dt$$

and

$$(dX(t))^2 = \sigma^2(t)dW(t)dW(t) = \sigma^2(t)dt$$

Consider an asset price process given by:

$$S(t) = S(0)e^{X(t)} = S(0)\exp\left\{\int_0^t \sigma(s)dW(s) + \int_0^t (\alpha(s) - \frac{1}{2}\sigma^2(s))ds\right\}$$

We may write $S(t) = f(X(t))$. Then

$$\begin{aligned}
dS(t) &= df(X(t)) \\
&= f'(X(t))dX(t) + \frac{1}{2}f''(X(t))dX(t)dX(t) \\
&= s(0)e^{X(t)}dX(t) + \frac{1}{2}S(0)e^{X(t)}dX(t)dX(t) \\
&= S(t)dX(t) + \frac{1}{2}S(t)dX(t)dX(t) \\
&= \alpha(t)S(t)dt + \sigma(t)S(t)dW(t)m
\end{aligned}$$

Theorem 4.7 (Itô integral of a deterministic integrand)**.** . *Let $W(s), s \geq 0$, be a Brownian motion, and let $\Delta(s)$ be a nonrandom function of time. Define $I(t) = \int_0^t \Delta(s)dW(s)$. For each $t \geq 0$, the random variable $I(t)$ is normally distributed with expected value zero and variance $\int_0^t \Delta^2(s)ds$.*

4.5 Black-Scholes-Merton Equation

In this section, we derive the Black-Scholes-Merton (BSM) partial differential equation for the price of an option on an asset modeled as a GBM.

4.5.1 Evolution of Portfolio Value

Consider an agent who at eacht ime t has a portfolio valued at $X(t)$. This portfolio invests in a money market account paying a constant rate of interest r and in a stock modeled by the GBM:

$$dS(t) = \alpha S(t)dt + \sigma S(t)dW(t)$$

Suppose at each time t, the investor holds $\Delta(t)$ shares of stock. The position $\Delta(t)$ can be random but must be adapted to the filtration associated with the Brownian motion $W(t), t \geq 0$. The remainder of the portfolio value, $X(t) - \Delta(t)S(t)$ is invested in the money market account. The differential $dX(t)$ for the investor's portfolio value at each time t is due to two factors, the capital gain $\Delta(t)dS(t)$ on the stock position and the interest earnings $r(X(t) - \Delta(t)S(t))dt$ on the cash position. Hence:

$$\begin{aligned} dX(t) &= \Delta(t)dS(t) + r(X(t) - \Delta(t)S(t))dt \\ &= \Delta(t)\left(\alpha S(t)dt + \sigma S(t)dW(t)\right) + r(X(t) - \Delta(t)S(t))dt \\ &= rX(t)dt + \Delta(t)(\alpha - r)S(t)dt + \Delta(t)\sigma S(t)dW(t) \end{aligned}$$

The first term on the RHS gives the underlying rate of return r on the portfolio. The second gives the risk premium for investing in the stock. The final term gives the volatility term proportional to the size of the stock investment.

Note that we want to consider the *discounted* stock price $e^{-rt}S(t)$. We have that

$$\begin{aligned} d(e^{-rt}S(t)) &= (\alpha - r)d^{-rt}S(t)dt + \sigma e^{-rt}S(t)dW(t) \\ d(e^{-rt}X(t)) &= \Delta(t)d(e^{-rt}S(t)) \end{aligned}$$

4.5.2 Evolution of Optional Value

Consider a European call option that pays $(S(T) - K)^+$ at time T. The strike price K is some nonnegative constant.

4.6 Multivariate Stochastic Calculus

4.6.1 Multiple Brownian Motions

Definition 4.3 (d-dimensional Brownian Motion)**.** A d-dimensional Brownian motion is a process

$$W(t) = \big(W_1(t), \dots, W_d(t)\big)$$

with the following properties

(i) Each $W_i(t)$ is a one-dimensional Brownian motion.

(ii) If $i \neq j$, then the processes $W_i(t)$ and $W_j(t)$ are independent.

Associated with a d-dimensional Brownian motion, we have a filtration $\mathcal{F}(t), t \geq 0$, such that the following holds.

(iii) (Information accumulates) For $0 \leq s < t$, every set in $\mathcal{F}(s)$ is also in $\mathcal{F}(t)$.

(iv) (Adaptivity) For each $t \geq 0$, the random vector $W(t)$ is $\mathcal{F}(t)$-measurable.

(v) (Independence of future increments) For $0 \leq t < u$, the vector of increments $W(u) - W(t)$ is independent of $\mathcal{F}(t)$.

Note that it is still possible to build correlated Brownian motions from this. Within each Brownian motion, the quadratic variation obviously inheres, i.e. $dW_i(t)dW_i(t) = dt$; however with independent Brownian motions, we have that

$$dW_i(t)dW_j(t) = 0, \quad i \neq j$$

4.6.2 Itô-Doeblin Formula for Multiple Processes

Let $X(t)$ and $Y(t)$ be Itô processes, which means they are processes of the form:

$$\begin{aligned} X(t) &= X(0) + \int_0^t \Theta_1(u)du + \int_0^t \sigma_{11}(u)dW_1(u) + \int_0^t \sigma_{12}(u)dW_2(u), \\ Y(t) &= Y(0) + \int_0^t \Theta_2(u)du + \int_0^t \sigma_{21}(u)dW_1(u) + \int_0^t \sigma_{22}(u)dW_2(u) \end{aligned}$$

In differential notation, we write

$$dX(t) = \Theta_1(t)dt + \sigma_{11}(t)dW_1(t) + \sigma_{12}(t)dW_2(t), \tag{4.6.1}$$

$$dY(t) = \Theta_2(t)dt + \sigma_{21}(t)dW_1(t) + \sigma_{22}(t)dW_2(t) \tag{4.6.2}$$

We have that $X(t)$ accumulates quadratic variation at a rate $\sigma_{11}^2(t) + \sigma_{12}^2(t)$ per unit time. In differential form,

$$dX(t)dX(t) = (\sigma_{11}^2(t) + \sigma_{12}^2(t))dt$$

In similar manner, we may derive the differential formulas

$$dY(t)dY(t) = (\sigma_{21}^2(t) + \sigma_{22}^2(t))dt$$
$$dX(t)dY(t) = \big(\sigma_{11}(t)\sigma_{21}(t) + \sigma_{12}(t)\sigma_{22}(t)\big)dt$$

Theorem 4.8 (Two-dimensional Itô-Doeblin formula). *Let $f(t,x,y)$ be a function whose partial derivatives, f_t, f_x, f_y, f_{xx}, f_{xy}, f_{yx}, and f_{yy} are defined and are continuous. Let $X(t)$ and $Y(t)$ be Itô processes as discussed above. The two-dimensional Itô-Doeblin formula in differential form is:*

$$\begin{aligned} df(t,X(t),Y(t)) &= f_t(t,X(t),Y(t))dt + f_x(t,X(t),Y(t))dX(t) + f_y(t,X(t),Y(t))dY(t) \\ &+ \frac{1}{2}f_{xx}(t,X(t)Y(t))dX(t)dX(t) + f_{xy}(t,X(t)Y(t))dX(t)dY(t) \\ &+ \frac{1}{2}f_{yy}(t,X(t)Y(t))dY(t)dY(t) \end{aligned}$$

Compactly,

$$df(t,X,Y) = f_t dt + f_x dX + f_y dY + \frac{1}{2}f_{xx}dXdX + f_{xy}dXdY + \frac{1}{2}f_{yy}dYdY$$

Since $dtdt$, $dtdX$, and $dtdY$ are zero, we don't include these in the Taylor expansion.

Corollary 4.8.1 (Itô product rule). *Let $X(t)$ and $Y(t)$ be Itô processes. Then*

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t) + dX(t)dY(t)$$

4.6.3 Recognizing a Brownian Motion

Theorem 4.9 (Levy, one dimension). *Let $M(t), t \geq 0$ be a martingale relative to a filtration $\mathcal{F}(t), t \geq 0$. Assume that $M(0) = 0, M(t)$ has continuous paths, and $[M,M](t) = t$ for all $t \geq 0$. Then $M(t)$ is a Brownian motion.*

Even though this theorem does not say anything about normality, implicit in the conclusion is that $M(t)$ is normally distributed. Sine the properties mentioned in Levy's theorem are the only ones we use in our construction of the Itô-Doeblin formula, we apply it to M under a function $f(t,x)$:

$$df(t,M(t)) = f_t(t,M(t))dt + f_x(t,M(t))dM(t) + \frac{1}{2}f_{xx}(t,M(t))dt$$

The last term uses the fact that $dM(t)dM(t) = dt$. Note since $M(t)$ is a martingale, $\int_0^t f_x(s,M(s))dM(s)$ is also, hence its expectation is 0.

Theorem 4.10 (Levy, two dimensions). *Let $M_1(t)$ and $M_2(t), t \geq 0$, be martingales relative to filtration $\mathcal{F}(t), t \geq 0$. Assume that for $i = 1,2$ we have $M_i(0) = 0$, $M_i(t)$ has continuous paths, and $[M_i, M_i](t) = t \;\forall\; t \geq 0$. If, in addition, $[M_1, M_2](t) = 0 \;\forall\; t \geq 0$, then $M_1(t)$ and $M_2(t)$ are independent Brownian motions.*

Example 4.2 (Correlated Stock Prices)**.** Suppose

$$\frac{dS_1(t)}{S_1(t)} = \alpha_1 dt + \sigma_1 dW_1(t)$$
$$\frac{dS_2(t)}{S_2(t)} = \alpha_2 dt + \sigma_2 \left[\rho dW_1(t) + \sqrt{1-\rho^2} dW_2(t)\right]$$

To analyze second stock price process, define

$$W_3(t) = \rho W_1(t) + \sqrt{1-\rho^2} W_2(t)$$

Then $W_2(t)$ is a continuous martingale with $W_3(0) = 0$ and

$$\begin{aligned} dW_3(t)dW_3(t) &= \rho^2 dW_1(t)dW_1(t) + 2\rho\sqrt{1-\rho^2} dW_1(t)dW_2(t) + (1-\rho^2)dW_2(t)dW_2(t) \\ &= \rho^2 dt + (1-\rho^2)dt = dt \end{aligned}$$

By Levy theorem, W_3 is Brownian. Hence we can write

$$\frac{dS_2(t)}{S_2(t)} = \alpha_2 dt + \sigma_2 dW_3(t)$$

where we note that Brownian motions $W_1(t)$ and $W_3(t)$ are now correlated. In particular,

$$\begin{aligned} d(W_1(t)W_3(t)) &= W_1 dW_3(t) + W_3(t)dW_1(t) + dW_1(t)dW_3(t) \\ &= W_1 dW_3(t) + W_3(t)dW_1(t) + \rho dt \end{aligned}$$

Integrating,

$$W_1(t)W_3(t) = \int_0^t W_1(s)dW_3(s) + \int_0^t W_3(s)dW_1(s) + \rho t$$

Hence $\mathbb{E}[W_1(t)W_2(t)] = \rho t$

5 Connections with Partial Differential Equations

5.1 Introduction

This chapter shows how to connect the risk neutral pricing problem to partial differential equations. Solutions to stochastic differential equation have the Markov property.

5.2 Stochastic Differential Equations

A *stochastic differential equation* is an equation of the form

$$dX(u) = \beta(u, X(u))du + \gamma(u, X(u))dW(u)$$

where $\beta(u,x)$ and $\gamma(u,x)$ are given functions, called the *drift* and *diffusion*. We also specify an initial condition of the form $X(t) = x, t \geq 0, x \in \mathbb{R}$. We aim to find a stochastic process $X(T)$ defined for $T \geq t$ such that

$$\begin{aligned} X(t) &= x, \\ X(T) &= X(t) + \int_t^T \beta(u, X(u))du + \int_t^T \gamma(u, X(u))dW(u) \end{aligned}$$

Under some conditions on the functions $\beta(\cdot), \gamma(\cdot)$, there exists a unique solution. The solution $X(T)$ will be $\mathcal{F}(T)$–measurable. *One dimensional linear stochastic differential equations* can be solved explicitly. This is an SDE of the form:

$$dX(u) = (a(u) + b(u)X(u))\, du + (\gamma(u) + \sigma(u)X(u))\, dW(u)$$

where $a(u), b(u), \sigma(u)$, and $\gamma(u)$ are nonrandom functions of time.

Example 5.1 (Geometric Brownian Motion)**.** We have:

$$dS(u) = \alpha S(u)du + \sigma S(u)dW(u)$$

implying that $\beta(u,x) = \alpha x$ and $\gamma(u,x) = \sigma(x)$

$$\begin{aligned} S(t) &= S(0)\exp\left\{\sigma W(t) + \left(\alpha - \frac{1}{2}\sigma^2\right)t\right\} \\ S(T) &= S(0)\exp\left\{\sigma W(T) + \left(\alpha - \frac{1}{2}\sigma^2\right)T\right\} \\ \implies \frac{S(T)}{S(t)} &= \exp\left\{\sigma(W(T) - W(t)) + \left(\alpha - \frac{1}{2}\sigma^2\right)(T-t)\right\} \end{aligned}$$

5.3 The Markov Property

Consider the standard SDE, let $0 \leq t \leq T$ be given, and let $h(y)$ be a Borel-measurable function. Denote the expectation

$$g(t,x) = \mathbb{E}^{t,x}h(X(T))$$

To numerically compute the expectation numerically, we could use the Euler method by choosing a small positive step with size δ and then set:

$$X(t+\delta) - x + \beta(t,x)\delta + \gamma(t,x)\sqrt{\delta}\epsilon_1,$$

Then set

$$X(t+2\delta) = X(t+\delta) = \beta(t+\delta, X(t+\delta))\delta + \gamma(t+\delta, X(t+\delta))\sqrt{\delta}\epsilon_2$$

Theorem 5.1 (Expectation of SDE)**.** *Let $X(u), u \geq 0$, be a solution to the SDE with inital condition given at time* 0*. Then, for $0 \leq t \leq T$,*

$$\mathbb{E}[h(X(T))|\mathcal{F}(t)] = g(t, X(t))$$

Intuitively, the only relevant piece of information when computing $\mathbb{E}[h(X(T))|\mathcal{F}(t)]$ is $X(t)$.

Corollary 5.1.1. *Solutions to stochastic differential equations are Markov processes.*

5.4 Partial Differential Equations

The Feynman-Kac Theorem relates SDE's and PDE's.

Theorem 5.2 (Feynman-Kac)**.** *Consider the SDE*

$$dX(u) = \beta(u, X(u))du + \gamma(u, X(u))dW(u)$$

Let $h(y)$ be a Borel-measurable function. Fix $T > 0$, and let $t \in [0,T]$ be given. Define the function

$$g(t,x) = \mathbb{E}^{t,x}h(X(T))$$

(We assume that $\mathbb{E}^{t,x}|h(X(T))| < \infty$ for all t and x.) Then $g(t,x)$ satisfies the PDE

$$g_t(t,x) + \beta(t,x)g_x(t,x) + \frac{1}{2}\gamma^2(t,x)g_{xx}(t,x) = 0$$

and the terminal condition

$$g(T,x) = h(x) \ \forall \ x$$

The proof of the Feynman-Kac Theorem depends on the following lemma.

Lemma 2. *Let $X(u)$ be a solution to the SDE with initial condition given at time* 0*. Let $h(y)$ be a Borel-measurable function, fix $T > 0$, and let $g(t,x)$ be given by 6.3.1. Then the stochastic process*

$$f(t, X(t)), 0 \leq t \leq T$$

is a martingale.

The general principle behind the proof of the Feynman-Kac theorem is

1. find the martingale
2. take the differential, and
3. set the dt term equal to 0.

Theorem 5.3 (Discounted Feynman-Kac)**.** *Consider the SDE*

$$dX(u) = \beta(u, X(u))du + \gamma(u, X(u))dW(u)$$

Let $h(y)$ be a Borel-measurable function. Fix $T > 0$, and let $t \in [0, T]$ be given. Define the function

$$f(t, x) = \mathbb{E}^{t,x}[e^{-r(T-t)}h(X(T))]$$

(We assume that $\mathbb{E}^{t,x}|h(X(T))| < \infty$ for all t and x.) Then $f(t, x)$ satisfies the PDE

$$f_t(t, x) + \beta(t, x)f_x(t, x) + \frac{1}{2}\gamma^2(t, x)f_{xx}(t, x) = rf(t, x)$$

and the terminal condition

$$f(T, x) = h(x) \ \forall \ x$$

6 External Measure Theory Notes

For an interval defined on $\mathbb{R}$, we are able to assign a quantity, known as a measure. For example, given an interval $(a, b])$

$$\lambda((a,b]) = b - a$$

Our goal is to extend this function to a function which is defined on *all* subsets of $\mathbb{R}$, not simply intervals. That is, we want a $\lambda : \mathfrak{P}(\mathbb{R}) \mapsto \mathbb{R}_+ \bigcup \{+\infty\}$, which extends the idea of length to all subsets ($\mathfrak{P}$), such as the set of all rational numbers between 0 and 1. We call this function the measure. We develop some properties we would like this function to have (note that we already have two properties: the mapping of the function, and the measure of an interval). We also want:

- Scalar Shifts don't change the measure: $A \subseteq \mathbb{R}, A + x = \{x + y, y \in A\}, \forall A \subseteq \mathbb{R}, \forall x \in \mathbb{R}, \lambda(A) = \lambda(A + x)$
- σ-additive: measure of union of disjoint sets equals the sum of the measures: $A = \bigcup_{j\geq 1} A_j, A_j \bigcap A_i = \varnothing, \lambda(A) = \sum_{j=1}^{n} \lambda(A_j)$

It turns out that there is no such function that applies to *all* subsets. So the condition we drop is that it applies to all subsets, while we retain the others.