Martingales, Brownian Motion, and Stochastic Integrals

Michael Zazanis
Department of Statistics
Athens University of Economics and Business

Chapter 1

Probability Spaces, Measures, and Random Variables

1.1 Probability Spaces as Measurable Spaces

The standard formulation of probability theory starts with a sample space Ω. Events correspond to subsets of this space. Logic dictates that if a subset A of Ω corresponds to an event then its complement, A^c, should also correspond to an event, namely the non–occurrence of A. Similarly, if A and B are events then $A \cup B$ and $A \cap B$ should also correspond to events. Families of sets are usually called classes and from the above it should be clear that the class of all events events should be a *field* of sets.

Field: Let Ω be a set and $\mathscr{A}$ a class of subsets of Ω. $\mathscr{A}$ is a *field* if

F1. $\Omega \in \mathscr{A}$.

F2. $A, B \in \mathscr{A}$ implies that $A \cup B \in \Omega$.

F3. $A \in \mathscr{A}$ implies $A^c \in \mathscr{A}$.

Note that in view of the above definition, if $\mathscr{A}$ is a field then $\emptyset = \Omega^c \in \mathscr{A}$ and if A, B, both belong to Ω then $A \cap B = (A^c \cup B^c) \in \Omega$. Also, it follows by induction that if A_i, $i = 1, 2, \ldots, n$ belong in $\mathscr{A}$, then $\bigcup_{i=1}^n A_i$ and $\bigcap_{i=1}^n A_i$ also belong in $\mathscr{A}$ (i.e. a field is a class of subsets of Ω closed under finite unions and intersections). Note that the *set difference* of two sets in $\mathscr{A}$, defined as $A \setminus B := AB^c$ and the *symmetric difference* $A \triangle B := AB^c \cup A^cB$ also belong to $\mathscr{A}$.

The above framework is the adequate for the simplest situations that arise in probability theory, namely those that deal with finite sample spaces. Consider for instance the problem of casting a die. A natural choice of sample space in this case would be $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ where ω_1 corresponds to the case where the face that lands up is 1 etc. There are six "elementary events" in this case, $\{\omega_i\}$, $i = 1, 2, \ldots, 6$ and all conceivable events are unions of these[1]. There are $2^6 = 64$ possible events, including the empty set (impossible event) and the whole space (certain event). For instance, the event that the outcome is even is $\{\omega_2, \omega_4, \omega_6\}$, while the event that the outcome is greater than or equal to 5 is $\{\omega_5, \omega_6\}$.

If the sample space has a finite or a countably infinite number of elements $\{\omega_1, \omega_2, \omega_3, \ldots\}$ it is possible to think in terms of elementary events $\{\omega_i\}$. In the typical case however, when the sample space is uncountably infinite, one begins with a field of subsets of Ω. When Ω is not finite it is important to be able to extend the above considerations to *sequences of events.* In particular we wish to ensure that countable intersections and unions of events are again events and this leads us to extend the notion of the field to that of the $\sigma-$field.

σ–Field: Let S be a set and $\mathscr{F}$ a field of subsets of Ω. $\mathscr{F}$ is a *σ–field* if it also satisfies

F4. If A_i, $i = 1, 2, 3, \ldots$ belong to $\mathscr{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$.

Again we point out that $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c \in \mathscr{F}$. Thus the countable union property, together with closure under complementation and de Morgan's laws, imply closure under countable intersections as well and a σ–field is closed under countable set operations.

The following propositions are direct consequences of the definition.

Proposition 1: Let $\mathscr{F}_i$, $i \in I$ a family of σ–fields on S, where I is an index set. Then the class $\mathscr{F} := \bigcap_i \mathscr{F}_i$ is again a σ–field.

Proposition 2: The class $\mathscr{P}(S) := \{A : A \subset S\}$, i.e. the set of all subsets of Ω is a σ–field.

Let $\mathscr{C}$ a class of subsets of Ω. The σ–field it generates is the smallest σ–field that contains all its elements i.e. the *intersection* of all the σ–fields that contain $\mathscr{C}$. We know that the family of σ–fields that contain $\mathscr{C}$ is not empty since it contains at least $\mathscr{P}(S)$, the power set of Ω.

Definition 1. *Let Ω be a set and $\mathscr{F}$ a σ–field of subsets of Ω. A probability measure defined on (Ω,$\mathscr{F}$) is a set function $P : \mathscr{F} \to [0, 1]$ such that*

(i) $P(\Omega) = 1$,

(ii) $P(A^{\mathrm{c}}) = 1 - P(A)$ *for all* $A \in \mathscr{F}$,

[1]Note that the "elementary events" are not the *elements* ω_1, $\omega_2, \ldots$, but the *sets* $\{\omega_1\}$, $\{\omega_2\}, \ldots$. Events are always subsets of Ω.

(iii) *for all* $A_1, A_2, A_3, \ldots \in \mathscr{F}$ *with* $A_i \cap A_j = \emptyset$ *we have* $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

1.2 Sequences of Events

Let $\{A_n\}$ be a sequence of sets belonging to $\mathscr{F}$. We say that this sequence is *increasing* if $A_n \subseteq A_{n+1}$ for all n and *decreasing* if $A_n \supseteq A_{n+1}$ for all n. The limit of a monotone sequence of events is defined as $\lim_{n\to\infty} := \bigcup_{n=1}^{\infty} A_n$ for an increasing sequence $\{A_n\}$ and $\lim_{n\to\infty} := \bigcap_{n=1}^{\infty} A_n$ for a decreasing sequence.

If $\{A_n\}_{n=1,2,\ldots}$ is an increasing sequence of events, we can write $D_n = A_n \setminus A_{n-1}$, $n = 2, 3, \ldots$, $D_1 = A_1$. Note that $D_n \in \mathscr{F}$ and $D_n \bigcap D_m = \emptyset$ when $m \neq n$. Thus $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$ where the D_n's are disjoint and $P(D_n) = P(A_n) - P(A_{n-1})$, $n = 2, 3, \ldots$.

$$P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} D_n) = \sum_{n=1}^{\infty} P(D_n) = P(A_1) + \sum_{n=2}^{\infty} P(A_n) - P(A_{n-1})$$

However, the last series is telescopic and has the value $\lim_n P(A_n) - P(A_1)$. Thus

$$P(\lim_n A_n) = \lim_n P(A_n)$$

for increasing sequences. The same can be shown for decreasing sequences, hence the above equality holds for all monotonic sequences of events.

If $\{A_n\}$ is a sequence of events that is not monotonic, we define its *superior and inferior limits* as

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m, \qquad \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m.$$

The meaning of these two events can be understood as follows: $\omega \in \limsup A_n$ or $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$ means that $\omega \in \bigcup_{m=n}^{\infty} A_m$ *for all* n which in turn means that for every natural number n there exists another natural $n' > n$ such that $\omega \in A_{n'}$. In other words, $\omega \in \limsup_n A_n$ if there exists a subsequence (n_k) such that $\omega \in A_{n_k}$ for every k or equivalently if ω belongs to infinitely many A_n's. We also point out that the sets $B_n := \bigcup_{m=n}^{\infty} A_m$, $n = 1, 2, \ldots$, form a decreasing sequence.

Similarly, the sequence of sets $C_n = \bigcap_{m=n}^{\infty} A_m$ is an increasing sequence of sets hence $\liminf_n A_n = \bigcup_n C_n = \lim_n C_n$. Thus $\omega \in \liminf A_n$ or $\omega \in \bigcup_{n=1}^{\infty} C_n$ if there exists a natural number n such that $\omega \in C_n$, which in turn means that $\omega \in \bigcap_{m=n}^{\infty}$, i.e. that ω belongs to *all the* A_m*, for* $m \geq n$. Hence $\liminf_n A_n$ is the set of ω that belong to *all but a finite number of the* A_n*'s.*

Theorem 1. ***[Borel–Cantelli]*** *Let $\{A_n\}$ be a sequence of events such that*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \ . \tag{1.1}$$

Then, with probability 1, only a finite number of these events occurs.

Proof: Let Ω be the probability space and define

$$\mathbf{1}_{A_i}(\omega) = \begin{cases} 1 & \text{if } \omega \in A_i \\ 0 & \text{if } \omega \notin A_i \end{cases} \ .$$

Also $\{\omega : \text{a finite number of the } A_i\text{'s occur}\} = \{\omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_i}(\omega) < \infty\}$. Note however that

$$\sum_{n=1}^{\infty} P(A_n) = E \sum_{n=1}^{\infty} mathbf{1}_{A_n}(\omega)$$

and hence (1.1) implies that the rhs of the above equation is finite and hence that $\sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) < \infty$ w.p. 1. ♠

An alternative proof of the Borel–Cantelli lemma (as it is widely known) goes as follows. The probability that infinitely many of the events A_n occur is precisely $P(\limsup_n A_n)$ in view of the above discussion. But $\limsup_n A_n = \lim_{n\to\infty} \bigcup_{k\geq n} A_k$, hence

$$P(\limsup_n A_n) = P(\lim_{n\to\infty} \bigcup_{k\geq n} A_k) = \lim_{n\to\infty} P(\bigcup_{k\geq n} A_k) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0,$$

the last limit being zero since the series $\sum_{k=1}^{\infty} P(A_k)$ converges by assumption.

The Borel–Cantelli lemma has the following partial converse in the case where the events A_n are independent.

Theorem 2. ***[Second Borel-Cantelli Lemma]*** *If the events A_n, $n = 1, 2, \ldots$, are independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$ then the probability that infinitely many of the events A_n occur is 1, i.e. $P(\limsup_n A_n) = 1$.*

Proof: It suffices to show that $\lim_n P(\cup_{k\geq n} A_k) = 1$ or equivalently that $\lim_n P\left((\cup_{k\geq n} A_k)^c\right) = 0$. Using de Morgan's laws, $(\cup_{k\geq n} A_k)^c = \cap_{k\geq n} A_k^c$ hence $(\cup_{k\geq n} A_k)^c \subseteq \cap_{k=n}^m A_k^c$ for all $m \geq n$. Thus

$$P\left((\cup_{k\geq n} A_k)^c\right) \leq P\left(\cap_{k=n}^m A_k^c\right) = \prod_{k=n}^m P(A_k^c) = \prod_{k=n}^m \left(1 - P(A_k)\right),$$

where in the next to the last equality above we have used the independence of A_n. Using the inequality $1 - x \leq e^{-x}$ which is valid for all $x \in \mathbf{R}$ we have

$$P\left((\cup_{k\geq n} A_k)^c\right) \leq e^{-\sum_{k=n}^{m} P(A_k)}, \qquad \text{for all } m \geq n. \tag{1.2}$$

However, since the series $\sum_{k=1}^{\infty} P(A_k)$ diverges it follows that $\lim_{m\to\infty} \sum_{k=n}^{m} P(A_k) = \infty$ and hence, letting $m \to \infty$ in (1.2) we obtain

$$P\left((\cup_{k\geq n} A_k)^c\right) = 0,$$

or equivalently

$$P\left(\cup_{k\geq n} A_k\right) = 1,$$

whence

$$P(\liminf_n A_n) = P\left(\cap_{k=1}^{\infty} \cup_{k\geq n} A_k\right) = \lim_{n\to\infty} P\left(\cup_{k\geq n} A_k\right) = 1.$$

♠

1.3 Random Variables, Expectation

The Borel σ–field on the real line, denoted by $\mathscr{B}(\mathbb{R})$, or simply $\mathscr{B}$, is the smallest σ–field that contains all open sets or, as we say the smallest σ–field *generated by the open sets.* It is easy to see that such a σ–field must exist: Consider the family of *all σ-fields containing the open sets.* This family is of course not empty because it contains $\mathscr{P}(\mathbb{R})$ the set of all subsets of real numbers, which is a σ–field. The intersection of all these σ-fields is $\mathscr{B}$.

Proposition 1. $\mathscr{B}$ is also the σ–field generated by

1) all open intervals (a, b), $a, b \in \mathbb{R}$,

2) all closed intervals $[a, b]$, $a, b \in \mathbb{R}$,

3) all semi-infinite intervals of the form $(-\infty, a]$, $a \in \mathbb{R}$,

4) all semi-infinite intervals of the form $(-\infty, a)$, $a \in \mathbb{R}$.

If U, V, are two sets and $f : U \to V$ a function, then, for any $B \in V$, its inverse image under f is defined as $f^{-1}(B) := \{x \in U : f(x) \in B\}$.

Proposition 2. If I is a set of indices (not necessarily countable) and $\{B_i; i \in I\}$ a family of subsets of V then

1) $f^{-1}\left(\cup_{i\in I} B_i\right) = \cup_{i\in I} f^{-1}(B_i)$

2) $f^{-1}\left(\cap_{i\in I} B_i\right) = \cap_{i\in I} f^{-1}(B_i)$

2) If $f(U) = V$ then $f^{-1}(B^C) = \left(f^{-1}(B)\right)^C$.

Definition 2. A function $f : \mathbb{R} \to \mathbb{R}$ is Borel measurable (often we will simply call them measurable) if for all $B \in \mathscr{B}$, $f^{-1}(B) \in \mathscr{B}$.

Proposition 3. $f : \mathbb{R} \to \mathbb{R}$ is Borel measurable iff $f^{-1}(-\infty, x] \in \mathscr{B}$ for all $x \in \mathbb{R}$.

Definition 3. Suppose $(\Omega, \mathscr{F}, P)$ is a probability space. A function $X : \Omega \to \mathbb{R}$ is called a real random variable if, for all $B \in \mathscr{B}$, $X^{-1}(B) := \{\omega : X(\omega) \in B\}$.

Definition 4. A random variable $X : \Omega \to \mathbb{R}$ is called *simple* if there exists $n \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathbb{R}$ and $A_1, \ldots, A_n \in \mathscr{F}$ such that $X(\omega) = \sum_{i=1}^n a_i \mathbf{1}_{A_i}(\omega)$.

The expectation of a simple random variable $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ is defined as

$$EX = \sum_{i=1}^{n} a_i P(A_i)$$

The expectation of a *non-negative* random variable X is defined as

$$\sup\{EY : Y \text{ is a simple random variable and } Y \leq X\}.$$

σ–field generated by a random variable: If $X : \Omega \to \mathbb{R}$ is a real random variable, the collection of sets $\sigma - \{X\} := \{X^{-1}(B) : B \in \mathscr{B}\}$ is a σ–field contained in $\mathscr{F}$. (A sub-σ–field of $\mathscr{F}$.)

If $\mathscr{G}$ is a sigma field and $\subset \mathscr{F}$ then we say that $\mathscr{G}$ is a sub-sigma field of $\mathscr{F}$. If Y is a random variable such that $Y^{-1}(B) \in \mathscr{G}$ for any $B \in \mathscr{G}$, we say that Y is $\mathscr{G}$–measurable.

The conditional expectation of a random variable X with respect to the sub-sigma field $\mathscr{G}$ is defined to be a $\mathscr{G}$–measurable random variable Y for which

$$E[Y\mathbf{1}_G] \;=\; E[X\mathbf{1}_G] \quad \text{for all } G \in \mathscr{G}. \tag{1.3}$$

This conditional expectation is denoted by $Y = E[Y|\mathscr{G}]$.

Basic Properties of Conditional Expectation:

1. If $\mathscr{G} = \{\emptyset, \Omega\}$, i.e. if $\mathscr{G}$ is the trivial sigma field then $E[X|\mathscr{G}] = EX$.
2. $E[E[X|\mathscr{G}]] = EX$.

3. If $\mathscr{G}_1 \subset \mathscr{G}_2$ then $E[X|\mathscr{G}_1] = E[E[X|\mathscr{G}_2]|\mathscr{G}_1]$. (This is the so-called *tower property* of conditional expectation.

4. If $X \in \mathscr{G}$ then $X = E[X|\mathscr{G}]$.

5. If $Y \in \mathscr{G}$ then $E[XY|\mathscr{G}] = YE[X|\mathscr{G}]$.

We will sketch the proof of these properties. They all involve the use of the definition (1.3).

1. Set $Y = E[X|\mathscr{G}]$. Since Y is measurable with respect to $\mathscr{G}$ which is the *trivial* $\sigma-$*field*. This necessarily means that Y is constant because, for any $x \in \mathbb{R}$, $\{\omega : Y(\omega) \leq x\} \in \{\emptyset, \Omega\}$, i.e. $Y(\omega) = c$ for some $c \in \mathbb{R}$ and all $\omega \in \Omega$. Thus, from the definition of conditional expectation, $E[X\mathbf{1}_\Omega] = E[Y\mathbf{1}_\Omega]$. Hence, since $E[X\mathbf{1}_\Omega] = EX$, and $E[Y\mathbf{1}_\Omega] = c$, the result follows.

2. Again, set $Y = E[X|\mathscr{G}]$. Apply (1.3) with $G = \Omega \in \mathscr{G}$. Then $E[Y\mathbf{1}_\Omega] = E[X\mathbf{1}_\Omega]$ or $EY = EX$.

3. Set $Y_1 := E[X|\mathscr{G}_1]$, $Y_2 := E[X|\mathscr{G}_2]$ and $Z := E[Y_2|\mathscr{G}_1]$. We need to show that $Z = Y_1$ a.s. and to this end it is sufficient to show that

$$E[Z\mathbf{1}_A] = E[X\mathbf{1}_A] \quad \text{for any } A \in \mathscr{G}_1. \tag{1.4}$$

However

$$E[Z\mathbf{1}_A] = E[Y_2\mathbf{1}_A] \quad \text{because } Z = E[Y_2|\mathscr{G}_1] \text{ and } A \in \mathscr{G}_1 \tag{1.5}$$

$$E[Y_2\mathbf{1}_A] = E[X\mathbf{1}_A] \quad \text{because } Y_2 = E[X|\mathscr{G}_2] \text{ and } A \in \mathscr{G}_2 \text{ since } A \in \mathscr{G}_1 \subset \mathscr{G}_2. \tag{1.6}$$

Hence (1.4) follows from (1.5) and (1.6).

4. Set $Y := E[X|\mathscr{G}]$. We then have $E[X\mathbf{1}_A] = E[Y\mathbf{1}_A]$ or (by the linearity of the expection) $E[(X-Y)\mathbf{1}_A] = 0$ for any $A \in \mathscr{G}$. Since $X \in \mathscr{G}$ and $Y \in \mathscr{G}$ by the definition of the conditional expectation, the set $A_k = \{X - Y > \frac{1}{k}\} \in \mathscr{G}$ for any $k \in \mathbb{N}$. We thus have

$$0 = E[(X-Y)\mathbf{1}_{A_k}] \geq \frac{1}{k}P(A_k)$$

whence it follows that $P(A_k) = 0$. This however means that

$$P(X - Y > 0) = P\left(\cup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k) = 0.$$

By a similar argument we show that $P(X - Y < 0) = 0$ and hence $X = Y$ a.s.

5. Set $Z := E[X|\mathscr{G}]$. We will first prove that $E[YX|\mathscr{G}] = YE[X|\mathscr{G}]$ when $Y = \mathbf{1}_B$ with $B \in \mathscr{G}$. To this end it suffices to show that

$$E[\mathbf{1}_A Y\, E[X|\mathscr{G}]\,] \;= = E[\mathbf{1}_A Y X], \quad \text{for any } A \in \mathscr{G}. \tag{1.7}$$

When $Y = \mathbf{1}_B$ the above equation becomes

$$E[\mathbf{1}_{A\cap B}\, E[X|\mathscr{G}]\,] \;=\; E[\mathbf{1}_{A\cap B} X]$$

which holds since $A \cap B \in \mathscr{G}$.

1.4 Convergence Concepts for Sequences of Random Variables

Let $\{X_n\}$ be a sequence of real random variables defined on a probability space $(\Omega, \mathscr{F}, P)$. Seeing that such a random variable is in fact a *measurable function* from Ω to $\mathbb{R}$ (we write $X : (\Omega, \mathscr{F}) \longrightarrow (\mathbb{R}, \mathscr{B})$, where $\mathscr{B}$ is the Borel σ–field in $\mathbb{R}$) we realize that the issue of convergence of a sequence of random variables is the same as that of a sequence of real functions defined on an measure space.

1.4.1 Convergence in Probability and Convergence with Probability 1

Definition 5. *The sequence* $\{X_n\}$ *converges in probability to the random variable* X *if* $\forall \epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0. \tag{1.8}$$

(Note that (1.8) is shorthand for the statement $\lim_{n\to\infty} P\left(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0$.) To signify that X_n converges to 0 in probability we often write $X_n \xrightarrow{\text{P}} 0$.

Equivalently we may say that, for every $\epsilon > 0$, $\delta > 0$, there exists n_0 such that $P(|X_n - X| > \epsilon) < \delta$ for all $n \geq n_0$.

Definition 6. *The sequence* $\{X_n\}$ *converges to the random variable* X with probability 1 *if there exists a set* Λ *such that* $P(\Lambda) = 0$ *and for all* $\omega \notin \Lambda$, $X_n(\omega) \to X(\omega)$.

The above is *pointwise convergence* for all ω not in Λ and is usually denoted as $X_n \to X$ w.p. 1 (with probability 1) or a.s. (almost surely). Equivalently we may write $P\left(\{\omega : X_n(\omega) \to X(\omega)\}\right) = 1$ or simply $P(X_n \to X) = 1$.

In order to understand the connection between the two modes of convergence we have discussed so far let us examine closely the definition of a.s. convergence. The set on which X_n converges pointwise, i.e. the set $\{\omega : X_n(\omega) \to X(\omega)\}$ can be written as

$$\{\omega : \forall \epsilon > 0 \; \exists n_0(\omega, \epsilon) \text{ such that } |X_m(\omega) - X(\omega)| < \epsilon \text{ for all } m \geq n_0(\epsilon)\}$$

or, equivalently,

$$\bigcap_{\epsilon>0} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| < \epsilon\}.$$

Let $A_\epsilon := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| < \epsilon\}$. If $\epsilon_1 < \epsilon_2$, then $A_{\epsilon_1} \subseteq A_{\epsilon_2}$ Also, nothing is lost if we let $\epsilon = 1/k$ where $k \in \mathbb{N}$ and we can thus say that the set on which X_n converges to X is the set

$$\lim_{k\to\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\}$$

or, equivalently, $\lim_{k\to\infty} \limsup_{m\to\infty} \{|X_m - X| < 1/k\}$. Convergence with probability 1 is equivalent to the condition

$$\lim_{n\to\infty} P\left(\bigcap_{m=n}^{\infty} \{|X_m - X| > \epsilon\}\right) = 0. \tag{1.9}$$

From the above discussion we see that X_n *converges in probability to* X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \tag{1.10}$$

whereas X_n converges to 0 *with probability 1* if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(\sup_{k\geq n} |X_k - X| > \epsilon) = 0. \tag{1.11}$$

It should be clear from the above that convergence with probability 1 is stronger: it implies convergence in probability, while convergence in probability does not imply convergence w.p.1. Similarly, X_n converges w.p.1 to X iff $\lim_{n\to\infty} P(\sup_{k\geq n} |X_k - X| > \epsilon) = 0$. Convergence with probability 1 is also referred to as almost sure (abbreviated a.s.) convergence.

Before we move further, let us consider the following examples:

Example 1: Suppose that $\{X_n; n \in \mathbb{N}\}$ is a sequence of independent Bernoulli random variables with $P(X_n = 0) = 1 - \frac{1}{n}$, $P(X_n = 1) = \frac{1}{n}$. It is easy to see that X_n converges to 0 in probability. Indeed, for any $\epsilon > 0$, $P(|X_n| > \epsilon) \leq \frac{1}{n} \to 0$, and hence (1.10) is satisfied. On the other hand we can see that X_n does not converge to 0 w.p.1. Indeed, if $\epsilon \in (0, 1)$, then

$$\{\sup_{k\geq n} |X_k| > \epsilon\} = \bigcup_{k=n}^{\infty} \{X_k = 1\}$$

and hence, by de Morgan's rule

$$
\begin{aligned}
P(\sup_{k\geq n}|X_k|>\epsilon) &= P\left(\bigcup_{k=n}^{\infty}\{X_k=1\}\right) \geq P\left(\bigcup_{k=n}^{n}\{X_k=1\}\right) = 1-P\left(\bigcap_{k=n}^{m}\{X_k=0\}\right) \\
&= 1-\frac{n-1}{n}\frac{n}{n+1}\cdots\frac{m-2}{m-1}\frac{m-1}{m} \;=\; 1-\frac{n-1}{m}.
\end{aligned}
$$

This being true for all $m >$ we may let $m \to \infty$ to obtain

$$P(\sup_{k\geq n}|X_k|>\epsilon)\geq 1$$

which, of course implies $P(\sup_{k\geq n}|X_k|>\epsilon)=1$ for all $n\in\mathbb{N}$ and, as a result, (1.11) is not satisfied. Let us appraise this situation: If we make n large enough we can make the probability $P(X_n=1)$ arbitrarily close to zero, i.e. we can, in the limit be sure that $X_n=0$. However, since $P\left(\bigcup_{k=n}^{\infty}\{X_k=1\}\right)=1$ we can also be sure that, no mater how large we take n to be, there will be another 1 in the sequence. Put differently, the total number of 1's in the sequence is infinite with probability 1.

Example 2: Suppose now that, in the previous example, $P(X_n=1)=\frac{1}{n^2}$. Again it is easy to see that X_n converges in probability to 0. This time we will also show that it converges to 0 w.p.1. The same arguments as above apply.

$$
\begin{aligned}
P(\sup_{k\geq n}|X_k|>\epsilon) &= P\left(\bigcup_{k=n}^{\infty}\{X_k=1\}\right) = \lim_{m\to\infty}P\left(\bigcup_{k=n}^{m}\{X_k=1\}\right) \\
&= \lim_{m\to\infty}\left(1-P\left(\bigcap_{k=n}^{m}\{X_k=0\}\right)\right)
\end{aligned}
$$

the second equality following from the continuity of probability measure and the third from de Morgan's laws.

$$
\begin{aligned}
P\left(\bigcap_{k=n}^{m}\{X_k=0\}\right) &= \prod_{k=n}^{m}\left(1-\frac{1}{k^2}\right) = \prod_{k=n}^{m}\left(\frac{k^2-1}{k^2}\right) \\
&= \frac{(n-1)(n+1)}{n^2}\frac{n(n+2)}{(n+1)^2}\cdots\frac{(m-2)m}{(m-1)^2}\frac{(m-1)(m+1)}{m^2} \\
&= \frac{(n-1)(m+1)}{nm}
\end{aligned}
$$

Hence

$$P(\sup_{k\geq n}|X_k|>\epsilon) = 1-\lim_{m\to\infty}\frac{(n-1)(m+1)}{nm} = 1-\frac{n-1}{n} = \frac{1}{n}$$

which (according to (1.11) establishes convergence w.p.1. Unlike example 1, here we see that the total number of 1's in the sequence is finite with probability 1.

There is however a case where convergence in probability implies convergence with probability 1. Suppose that $\{Y_n\}$ converges monotonically to Y in probability. To start with the simplest case, assume that $0 \leq Y_{n+1} \leq Y_n$ for all n and $Y_n \xrightarrow{P} 0$. Because of monotonicity

$$\sup_{m \geq n} Y_m = Y_n$$

hence

$$P(\sup_{m \geq n} Y_m > \epsilon) = P(Y_n > \epsilon) \to 0$$

which implies that $Y_n \to 0$ w.p. 1.

The above result generalizes immediately to the case where either $Y_{n+1} \leq Y_n$ for all n or $Y_{n+1} \geq Y_n$ for all n and $Y_n \xrightarrow{P} Y$ by considering the sequence $\tilde{Y}_n = |Y_n - Y|$. Note that in both cases $\tilde{Y}_n$ is decreasing and by definition converges to zero in probability. Hence

$$\lim_{n \to \infty} P(\sup_{m \geq n} |Y_m - Y| > \epsilon) = \lim_{n \to \infty} P(|Y_n - Y| > \epsilon) = 0.$$

1.4.2 Convergence in the L^p sense

Let $\{X_n\}$, $n = 1, 2, \ldots$, be a sequence of real random variables such that $E|X_n|^p < \infty$ where $p \geq 1$. We say that X_n converges to X in L^p (write $X_n \xrightarrow{L^p} X$) if

$$\lim_{n \to \infty} E|X_n - X|^p = 0.$$

The case $p = 2$ is of particular importance and L^2 convergence it is often referred to as *mean square* (m.s.) convergence.

It is easy to see that convergence in L^p implies convergence in probability. For this we shall need the following basic inequality (known as the Markov inequality).

Theorem 3. *Let $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ an increasing function and Y a real random variable such that $E\phi(|Y|) < \infty$. Then, for any $\alpha > 0$,*

$$P(|Y| > \alpha) \leq \frac{E\phi(|Y|)}{\phi(\alpha)}. \tag{1.12}$$

A particular choice of the function ϕ that is often useful is $\phi(x) = x^p$ with $p \geq 1$ which gives a bound on the tail of the distribution in terms of its moments.

Proof: It suffices to observe that

$$\begin{aligned} E\phi(|Y|) &= E[\phi(|Y|)\mathbf{1}(|Y| \leq \alpha)] + E[\phi(|Y|)\mathbf{1}(|Y| > \alpha)] \\ &\geq E[\phi(|Y|)\mathbf{1}(|Y| > \alpha)] \\ &\geq \phi(\alpha)E[\mathbf{1}(|Y| > \alpha)] = \phi(\alpha)P(|Y| > \alpha) \end{aligned}$$

where in the above inequalities we have used the fact that ϕ takes nonnegative values and that it is increasing. ♠

Hence, applying the above inequality with $\phi(x) = x^p$ we obtain

$$P(|X_n - X| > \epsilon) \leq \frac{E|X_n - X|^p}{\epsilon^p}. \tag{1.13}$$

If $X_n \xrightarrow{L^p} X$ then the numerator on the right hand side of (1.13) goes to 0 as $n \to \infty$, hence we have $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$ for any $\epsilon > 0$. Thus convergence in L^p implies convergence in probability.

The relationship between convergence in L^p and a.s. convergence is more complicated. Neither one implies the other, unless certain extra conditions are satisfied.

Finally, one important result which will be used in the sequel is the following.

Theorem 4. *If a sequence of random variables $\{X_n\}$ converges to X in probability then there exists a subsequence n_k such that $X_{n_k} \to X$ w.p. 1.*

Proof: Since X_n converges in probability to X, for every k there exists n_k such that

$$P(|X_{n_k} - X| > 2^{-k}) < 2^{-k}.$$

Call A_k the event $\{\omega : |X_{n_k}(\omega) - X(\omega)| > 2^{-k}\}$. Since $\sum_{k=1}^{\infty} P(A_k) < \sum_{k=1}^{\infty} 2^{-k} < \infty$ the Borel–Cantelli theorem assures us that, with probability one, only finitely many of the A_k's will occur, i.e. that with probability 1, $|X_{n_k} - X| < 2^{-k}$ for all $k \geq k_0(\omega)$. This insures that

$$\sum_{k=1}^{\infty} |X_{n_k} - X| < \infty \qquad w.p.1$$

since the tail of the series is dominated by the convergent series $\sum_k 2^{-k}$. Thus

$$\lim_{k\to\infty} \sup_{m\geq k} |X_{n_m} - X| \leq \lim_{k\to\infty} \sum_{m\geq k} |X_{n_m} - X| = 0$$

since the series converges. ♠

Chapter 2

Martingales in Discrete Time

2.1 Adapted and Predictable processes

On the probability space $(\Omega, \mathscr{F}, P)$ suppose that there has been defined an *increasing* sequence of σ–fields$\mathscr{F}_n$ such that $\mathscr{F}_n \subset \mathscr{F}_{n+1} \subset \mathscr{F}$ for all n. The family $\{\mathscr{F}_n\}$ is called a *filtration*. In practice $\mathscr{F}_n$ represents the information available at (discrete) time n.

- The process $\{X_n\}_{n\geq 0}$ is **adapted** to $\{\mathscr{F}_n\}$ if for every n X_n is measurable with respect to $\mathscr{F}_n$. We will write $X_n \in \mathscr{F}_n$.
- The process $\{X_n\}_{n\geq 0}$ is **predictable** with respect to $\{\mathscr{F}_n\}$ if for every n X_n is measurable with respect to $\mathscr{F}_{n-1}$ or symbolically $X_n \in \mathscr{F}_{n-1}$.

2.2 Stopping Times

Let T be a nonnegative, integer valued, random variable. T is a stopping time w.r.t. the filtration $\{\mathscr{F}_n\}$ iff the sequence of random variables $\mathbf{1}(T = n)$, $n = 0, 1, 2, \ldots$, is adapted to $\{\mathscr{F}_n\}$. In particular, note that if T is a stopping time then $\{\mathbf{1}(T \leq n)\}$ is also an adapted sequence, while $\{\mathbf{1}(T > n)\}$ is a *predictable* sequence. To see this, write

$$\mathbf{1}(T \leq n) = \sum_{k=0}^{k=n} \mathbf{1}(T = k)$$

and observe that $\mathbf{1}(T = k) \in \mathscr{F}_k \subset \mathscr{F}_n$ for $k \leq n$. This establishes that $\mathbf{1}(T \leq n) \in \mathscr{F}_n$. On the other hand $\mathbf{1}(T > n) = 1 - \mathbf{1}(T \leq n-1)$ which, in view of the above is a *predictable* sequence.

Proposition 4. *If S, T, are $\mathscr{F}_n$–stopping times then $S+T$, $S \vee T$, $S \wedge T$ are also $\mathscr{F}_n$–stopping times.*

Proof: To prove the first statement, note that

$$\mathbf{1}(S+T=n) = \sum_{k=0}^{n} \mathbf{1}(S=k)\mathbf{1}(T=n-k) \in \mathscr{F}_n.$$

The second follows from $\mathbf{1}(S \vee T \leq n) = \mathbf{1}(S \leq n)\mathbf{1}(T \leq n)$, and the fact that both $\mathbf{1}(S \leq n)$ and $\mathbf{1}(S \leq n)$ are in $\mathscr{F}_n$ since T and S are stopping times. Finally $\mathbf{1}(S \wedge T > n) = \mathbf{1}(S > n)\mathbf{1}(T > n)$. ♠

2.3 Martingales in Discrete Time

Theorem 5. *A process $\{X_n\}$ is a martingale w.r.t. the filtration $\{\mathscr{F}_n\}$ if*

- *X_n is an adapted process, i.e. $X_n \in \mathscr{F}_n$,*
- *$E|X_n| < \infty \; \forall n$,*
- *$E[X_{n+1}|\mathscr{F}_n] = X_n \; \forall n$.*

Example 1: Let $\{Y_i\}$ be independent random variables with $E|Y_i| < \infty$ for all i and consider the filtration $\mathscr{F}_n = \sigma - \{Y_1, Y_2, \ldots, Y_n\}$. (This is sometimes called the natural filtration of the process.) Let $EX_i = \mu_i$. The process $X_n = \sum_{i=1}^{n} Y_i - \mu_i$ is an $\mathscr{F}_n$–martingale.

Example 2: Using the setup of the previous example suppose that, for all i, $\sigma_i^2 = \mathrm{Var}(Y_i) < \infty$. The process $X_n = \left(\sum_{i=1}^{n} Y_i - \mu_i\right)^2 - \sum_{i=1}^{n} \sigma_i^2$ is an $\mathscr{F}_n$–martingale.

Example 3: Using again the same setup we assume that Y_i has distribution F_i and $\tilde{F}_i(s) := \int_{-\infty}^{\infty} e^{-sx} dF_i(x)$ is finite for s in a neighborhood of 0. Then

$$X_n := \frac{e^{-s\sum_{i=1}^{n} Y_i}}{\prod_{i=1}^{n} \tilde{F}_i(s)} \,,$$

is an $\mathscr{F}_n$–martingale.

Example 4: Let $\{Y_n\}$ be a Discrete Time Markov Chain with state space $\mathscr{S}$ and transition probability matrix $P(i,j)$. Also suppose that $f : \mathscr{S} \to \mathbb{R}$ be a real function. Then

$$X_n := \sum_{k=1}^{n} \left(f(Y_k) - \sum_{j \in \mathscr{S}} P(Y_{k-1}, j) f(j) \right)$$

is an $\mathscr{F}_n$–martingale.

Example 5: [Right Regular Sequences and Induced Martingales for Markov Chains] Let $\{Y_n\}$ be a Discrete Time Markov Chain with state space $\mathscr{S}$ and transition probability matrix $P(i,j)$. Let $f : \mathscr{S} \to \mathbb{R}$ be bounded and satisfy

$$f(i) = \sum_{j \in \mathscr{S}} P(i,j) f(j) \ , \quad \forall i \in \mathscr{S} \ .$$

Such sequences (right eigenvectors corresponding to eigenvalue 1) are called right regular sequences. Then

$$X_n = f(Y_n)$$

is a martingale.

Example 6: The above example is a special case of the following more general class of martingales. Let f be a right eigenvector corresponding to an eigenvalue λ of P, i.e.

$$\lambda f(i) = \sum_{j \in \mathscr{S}} P(i,j) f(j) \ , \quad \forall i \in \mathscr{S} \ .$$

Assuming that $E|f(Y_n)| < \infty$,

$$X_n = \lambda^{-n} f(Y_n)$$

is a martingale.

Example 7: [Likelihood Ratios] Let $\{Y_n\}$ be an i.i.d. sequence with density g. Let f be another density function. Then the process

$$X_n = \frac{f(Y_0) f(Y_1) \cdots f(Y_n)}{g(Y_0) g(Y_1) \cdots g(Y_n)}$$

is a martingale.

2.4 Submartingales, Supermartingales, and Martingale Transforms

Theorem 6. *$\{X_n\}$ is a **submartingale** w.r.t. $\{\mathscr{F}_n\}$ iff*

a) $X_n \in \mathscr{F}_n$

b) $E|X_n| < \infty \ \forall n$

c) $E[X_{n+1} | \mathscr{F}_n] \geq X_n, \ \forall n.$

*$\{X_n\}$ is a **supermartingale** w.r.t. $\{\mathscr{F}_n\}$ iff it satisfies a) and b) above and*

c') $E[X_{n+1}|\mathscr{F}_n] \leq X_n$, $\forall n$.

2.4.1 Convexity and Jensen's Inequality

A function $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ is convex iff, for every $\lambda \in (0,1)$ and every $x_1, x_2 \in \mathbb{R}$

$$\phi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\phi(x_1) + (1-\lambda)\phi(x_2)$$

It can be shown that, ϕ is convex iff for every $x_0 \in \mathbb{R}$ there exists $\beta \in \mathbb{R}$ such that

$$\phi(x_0) + \beta(x - x_0) \leq \phi(x) \qquad \text{for all } x \in \mathbb{R}.$$

(This result, which is in fact true for convex functions in $\mathbb{R}^n$ is known as the *supporting hyperplane theorem.*) We are now ready to state the central result about convex functions which we shall need here:

Jensen's Inequality Let ϕ be a convex function and X a random variable with $EX < \infty$. Then

$$g(EX) \leq Eg(X) \; . \tag{2.1}$$

Proof: Apply the supporting hyperplane theorem with $x_0 = EX$ to obtain

$$\phi(EX) + \beta(x - EX) \leq \phi(x) \qquad \text{for all } x \in \mathbb{R}.$$

Hence

$$\phi(EX) + \beta(X - EX) \leq \phi(X)$$

and taking expectations in the above equation establishes (2.1) since $E(X - EX) = 0$. ♠

Theorem 7. *Let $\{X_n\}$ be a martingale and g a convex function. Then $\{g(X_n)\}$ is a submartingale, provided that $E|g(X_n)| < \infty$.*

Examples: Suppose $\{X_n\}$ is a martingale. Then $\{X_n^2\}$ and $\{(X_n - a)^+\}$ are submartingales.

2.4.2 Martingale Transforms

Let $\{M_n\}$ be an $\mathscr{F}_n$–martingale and $\{C_n\}$ an $\mathscr{F}_n$–predictable process. Set $\Delta M_n = M_n - M_{n-1}$,

$$X_n = C_0 M_0 + \sum_{k=1}^{n} C_k \Delta M_k \; .$$

$\{X_n\}$ is a *Martingale Transform.* It is easy to see that martingale transforms are martingales:

Proposition 5. *Suppose $|C_n| \le K\ \forall n$, where K is a positive real number. Then $\{X_n\}$ is a martingale.*

Proof: We first show that $E|X_n| < \infty$:

$$\begin{aligned} E|X_n| &\le E|C_0|E|M_0| + \sum_{k=1}^{n} E|C_k|E|\Delta M_k| \\ &\le K\left(E|M_0| + \sum_{k=1}^{n} E|M_k| + E|M_{k-1}|\right) < \infty \end{aligned}$$

($E|M_k| < \infty$ for all k since $\{M_k\}$ is a martingale.)

Next, check that $E[X_{n+1}|\mathscr{F}_n] = X_n$. Indeed,

$$X_{n+1} = X_n + C_{n+1}\left(M_{n+1} - M_n\right)$$

and, taking expectations,

$$E[X_{n+1}|\mathscr{F}_n] = E[X_n + C_{n+1}\left(M_{n+1} - M_n\right)|\mathscr{F}_n] = X_n + C_{n+1}E[M_{n+1} - M_n|\mathscr{F}_n] = 0$$

The second equality following from the fact that $X_n, C_{n+1} \in \mathscr{F}_n$, and the last from the fact that M_n is a martingale. ♠

2.5 Square–integrable martingales and orthogonality of increments

Theorem 8. *Let $\{X_n\}$ be an $\mathscr{F}_n$–martingale with $EX_n^2 < \infty$. Then, for all integers $i \le j \le k \le l$,*

$$E(X_l - X_k)(X_j - X_i) = 0\ . \tag{2.2}$$

Furthermore

$$EX_n^2 = EX_0^2 + \sum_{k=1}^{n} E(X_k - X_{k-1})^2\ . \tag{2.3}$$

Proof: To establish (2.2) note that

$$E[(X_l - X_k)(X_j - X_i)|\mathscr{F}_k] = (X_j - X_i)E[X_l - X_k|\mathscr{F}_k] = 0\ .$$

To show (2.3) write $X_n = X_0 + \sum_{k=1}^{n}(X_k - X_{k-1})$. Then

$$X_n^2 = X_0^2 + 2X_0\sum_{k=1}^{n}(X_k - X_{k-1}) + \sum_{k=1}^{n}(X_k - X_{k-1})^2 + 2\sum_{k=2}^{n}\sum_{j=1}^{k-1}(X_k - X_{k-1})(X_j - X_{j-1}).$$

Taking expectations and using (2.2) yields (2.3). ♠

2.6 The Doob–Meyer Decomposition

Theorem 9. *Let $\{X_n\}$ be a process adapted to $\mathscr{F}_n$. Then there exists an $\mathscr{F}_n$–martingale $\{M_n\}$ and an $\mathscr{F}_n$–predictable process $\{A_n\}$ with $M_0 = 0$, $A_0 = 0$ such that*

$$X_n = X_0 + M_n + A_n \ . \tag{2.4}$$

This decomposition is essentially unique in that if $X_n = X_0 + \tilde{M}_n + \tilde{A}_n$ $\forall n$, $M_n = \tilde{M}_n$, $A_n = \tilde{A}_n$ $\forall n$ (with probability 1).

If $\{X_n\}$ is a submartingale then $\{A_n\}$ is a nondecreasing process, i.e. $A_{n+1} \geq A_n$ $\forall n$ (w.p. 1).

Proof: If (2.4) is true then

$$X_{n+1} - X_n = M_{n+1} - M_n + A_{n+1} - A_n$$

hence

$$E[X_{n+1} - X_n | \mathscr{F}_n] = E[M_{n+1} - M_n | \mathscr{F}_n] + E[A_{n+1} - A_n | \mathscr{F}_n] \ .$$

Since $\{M_n\}$ is a martingale, $E[M_{n+1} - M_n | \mathscr{F}_n] = 0$. Since $\{A_n\}$ is predictable, $E[A_{n+1} - A_n | \mathscr{F}_n] = A_{n+1} - A_n$. Hence,

$$A_{n+1} = A_n + E[X_{n+1} | \mathscr{F}_n] - X_n \ .$$

Set

$$A_n = \sum_{k=1}^{n} E[X_k | \mathscr{F}_{k-1}] - X_{k-1}, \tag{2.5}$$

$$M_n = \sum_{k=1}^{n} X_k - E[X_k | \mathscr{F}_{k-1}]. \tag{2.6}$$

From (2.5) you can verify that $\{A_n\}$ is $\mathscr{F}_n$–predictable, from (2.6) that $\{M_n\}$ is an $\mathscr{F}_n$–martingale, and adding (2.5)+ (2.6) gives

$$M_n + A_n = X_n + X_0.$$

Note that if $\{X_n\}$ is a submartingale then it is $\mathscr{F}_n$–adapted and therefore the Doob-Meyer decomposition holds with A_n, M_n given by (2.5), (2.6). From (2.5) it follows that

$$A_{n+1} - A_n = E[X_{n+1} | \mathscr{F}_n] - X_n \geq 0,$$

since $\{X_n\}$ is a submartingale.

To show uniqueness, suppose we also have $X_n = X_0 + \tilde{M}_n + \tilde{A}_n$. Then $M_n + A_n = \tilde{M}_n + \tilde{A}_n$ or

$$M_n - \tilde{M}_n = \tilde{A}_n - A_n \ . \tag{2.7}$$

Taking conditional expectations we get

$$E[M_n|\mathscr{F}_{n-1}] - E[\tilde{M}_n|\mathscr{F}_{n-1}] = E[\tilde{A}_n|\mathscr{F}_{n-1}] - E[A_n|\mathscr{F}_{n-1}]$$

However

$$E[M_n|\mathscr{F}_{n-1}] = M_{n-1}\text{(martingale)}$$
$$E[A_n|\mathscr{F}_{n-1}] = M_{n-1}\text{(predictable)}$$

(The same relations hold for $\tilde{M}_n$ and $\tilde{A}_n$.) Therefore

$$M_{n-1} - \tilde{M}_{n-1} = \tilde{A}_n - A_n \ . \tag{2.8}$$

From (2.7) and (2.8) we get

$$M_{n-1} - \tilde{M}_{n-1} = M_n - \tilde{M}_n \ . \tag{2.9}$$

(2.9) holds for all n and, by induction,

$$M_n - \tilde{M}_n = M_0 - \tilde{M}_0 = 0 \ .$$

From (2.7) it follows that

$$A_n - \tilde{A}_n = 0 \ .$$

♠

Example: An application of the Doob–Meyer decomposition Let $\{X_n\}$ be an $\mathscr{F}_n$–martingale. Then $\{X_n^2\}$ is a submartingale and

$$X_n^2 = X_0^2 + A_n + M_n$$

where M_n is a martingale and A_n is predictable and are given by the expressions

$$A_n = \sum_{k=1}^{n} E[\Delta X_k^2|\mathscr{F}_{k-1}], \tag{2.10}$$

$$M_n = \sum_{k=1}^{n} X_k^2 - E[X_k^2|\mathscr{F}_{k-1}]. \tag{2.11}$$

2.6.1 Quadratic Variation of a Martingale

Let $\{X_n\}$ be an $\mathscr{F}_n$–martingale. Then $\{X_n^2\}$ is a submartingale which we can decompose into a martingale and an increasing process. This increasing process is called the *quadratic variation* of X, $\langle X\rangle$. We write

$$X_n^2 = M_n + \langle X\rangle_n \ .$$

From the Doob–Meyer decomposition we have

$$\langle X\rangle_n = EX_0^2 + \sum_{k=1}^{n} E[(X_k - X_{k-1})^2|\mathscr{F}_{k-1}] \ .$$

2.7 The Optional Sampling Theorem

2.7.1 Optional Sampling Theorem for Submartingales

Let $\{X_n\}$ be a submartingale w.r.t. $\{\mathscr{F}_n\}$ and S, T, be stopping times such that $0 \leq S \leq T \leq m$ (where m is a given integer). Then

$$EX_S \leq EX_T \ . \tag{2.12}$$

Proof: Write $X_T = X_0 + (X_1 - X_0) + \cdots + (X_T - X_{T-1})$, or

$$X_T = X_0 + \sum_{k=1}^{m} (X_k - X_{k-1}\mathbf{1}(T \geq k) \ .$$

Similarly,

$$X_S = X_0 + \sum_{k=1}^{m} (X_k - X_{k-1}\mathbf{1}(S \geq k) \ .$$

Taking expectations we can write

$$EX_T = EX_0 + \sum_{k=1}^{m} E[E[(X_k - X_{k-1}\mathbf{1}(T \geq k)|\mathscr{F}_{k-1}]] \ .$$

Note that $\mathbf{1}(T \geq k) = 1 - \sum_{i=0}^{k-1} \mathbf{1}(T = i) \in \mathscr{F}_{k-1}$ and hence

$$E[(X_k - X_{k-1}\mathbf{1}(T \geq k)|\mathscr{F}_{k-1}] = \mathbf{1}(T \geq k)E[(X_k - X_{k-1}|\mathscr{F}_{k-1}]$$

Since $T \geq S$, $\mathbf{1}(T \geq k) \geq \mathbf{1}(S \geq k)$. Also $E[X_k - X_{k-1}|\mathscr{F}_{k-1}] \geq 0$ ($\{X_n\}$ is a submartingale). Hence

$$\mathbf{1}(T \geq k)E[(X_k - X_{k-1}|\mathscr{F}_{k-1}] \geq \mathbf{1}(S \geq k)E[(X_k - X_{k-1}|\mathscr{F}_{k-1}] \ .$$

From the above it follows that

$$X_0 + \sum_{k=1}^{m} E[(X_k - X_{k-1}\mathbf{1}(T \geq k)|\mathscr{F}_{k-1}] \geq X_0 + \sum_{k=1}^{m} E[(X_k - X_{k-1}\mathbf{1}(S \geq k)|\mathscr{F}_{k-1}] \ .$$

Taking expectations:

$$E\left[X_0 + \sum_{k=1}^{m} (X_k - X_{k-1}\mathbf{1}(T \geq k)\right] \geq E\left[X_0 + \sum_{k=1}^{m} E[(X_k - X_{k-1}\mathbf{1}(S \geq k)\right] \ ,$$

or

$$EX_T \geq EX_S \ .$$

♠

2.7.2 Doob's Maximal Inequality

Let $\{X_n\}$ be a nonnegative submartingale (i.e. $X_n \geq 0 \; \forall n$). Then, $\forall \lambda > 0$, $\forall n$,

$$\lambda P\{\max_{0 \leq k \leq n} X_k > \lambda\} \leq EX_n \; . \tag{2.13}$$

Proof: Define the stopping time T as

$$T = \begin{cases} \min\{k : X_k > \lambda\} & \text{if} \quad \max_{0 \leq k \leq n} X_k > \lambda \\ n & \text{if} \quad \max_{0 \leq k \leq n} X_k \leq \lambda \end{cases} .$$

Notice that $\{X_T > \lambda\} = \{\max_{0 \leq k \leq n} X_k > \lambda\}$ and therefore

$$P\{X_T > \lambda\} = P\{\max_{0 \leq k \leq n} X_k > \lambda\} \; .$$

However, from Markov's inequality

$$\lambda P\{X_T > \lambda\} \leq EX_T \; ,$$

while from the Optional Sampling Theorem,

$$EX_T \leq EX_n \; .$$

The conclusion of the theorem follows from the above. ♠

2.7.3 The Optional Sampling Theorem for Martingales

Let $\{X_n\}$ be a martingale w.r.t. $\{\mathscr{F}_n\}$. We know that $EX_n = EX_0$. If T is a stopping time, under what conditions is $EX_T = EX_0$? We start with

Lemma 1. *Let $\{X_n\}$ be a martingale and T a stopping time w.r.t. $\{\mathscr{F}_n\}$. Then, for all $n \geq k$,*

$$E[X_n \mathbf{1}(T = k)] = E[X_k \mathbf{1}(T = k)] \; .$$

Proof: Indeed

$$E[X_n \mathbf{1}(T = k)] = E[E[X_n \mathbf{1}(T = k) | \mathscr{F}_n]] = E[\mathbf{1}(T = k) E[X_k | \mathscr{F}_n]] = E[\mathbf{1}(T = k) X_n]$$

♠

Lemma 2. *With the assumptions of the previous lemma*

$$E[X_{T \wedge n}] = EX_0 \; .$$

Proof: We can write $X_{T\wedge n} = \sum_{k=0}^{n-1} X_k \mathbf{1}(T=k) + X_n \mathbf{1}(T \geq n)$ and taking expectations,

$$\begin{aligned}
E[X_{T\wedge n}] &= \sum_{k=0}^{n-1} E[X_k \mathbf{1}(T=k)] + E[X_n \mathbf{1}(T \geq n)] \\
&= \sum_{k=0}^{n-1} E[X_n \mathbf{1}(T=k)] + E[X_n \mathbf{1}(T \geq n)] \\
&= E\left[X_n \left(\sum_{k=0}^{n-1} \mathbf{1}(T=k) + \mathbf{1}(T \geq n)\right)\right] \\
&= EX_n = EX_0 \ .
\end{aligned}$$

Theorem 10. *Let $\{X_n\}$ be a martingale and T a stopping time w.r.t. $\{\mathscr{F}_n\}$. Suppose that $P(T<\infty)=1$ and $E[\sup_k |X_{T\wedge k}|] < \infty$. Then $EX_T = EX_0$.*

Proof: From the previous lemma we have $EX_{T\wedge n} = EX_0 \ \forall n$. Since $P(T<\infty)=1$, $\lim_{n\to\infty} X_{T\wedge n} = X_T$. Finally, $X_{T\wedge n} \geq \sup_k |X_{T\wedge k}|$. Use the Dominated Convergence Theorem to conclude that

$$\lim_{n\to\infty} E[X_{T\wedge n}] = E[\lim_{n\to\infty} X_{T\wedge n}] = EX_T \ .$$

♠

2.7.4 The Kolmogorov–Doob Inequality

Theorem 11. *Let X_n be a square–integrable martingale (i.e. $EX_n^2 < \infty$ for all n). Then*

$$P(\max_{0\leq i\leq n} |X_i| \geq \epsilon) \leq \frac{EX_n^2}{\epsilon^2} \ .$$

Proof: Define the sets $A_k = \{|X_i| < \epsilon, \ i \leq k\}$, $B_k = A_{k-1} \cap \{|X_k| \geq \epsilon\}$. Then $\Omega = A_n \bigcup \bigcup_{k=0}^{n} B_k$ and

$$EX_n^2 = \sum_{k=0}^{n} E[X_n^2 \mathbf{1}(B_k)] \ + \ E[X_n^2 \mathbf{1}(A_n)] \geq \sum_{k=0}^{n} E[X_n^2 \mathbf{1}(B_k)]$$

We have however

$$\begin{aligned}
E[X_n^2 \mathbf{1}(B_k)] &= E[(X_n - X_k + X_k)^2 \mathbf{1}(B_k)] \\
&= E[(X_n - X_k)^2 \mathbf{1}(B_k)] + 2E[(X_n - X_k)\mathbf{1}(B_k)] + E[X_k^2 \mathbf{1}(B_k)] \\
&\geq E[X_k^2 \mathbf{1}(B_k)] \ .
\end{aligned}$$

Hence

$$EX_n^2 \geq \sum_{k=0}^{n} E[X_k^2 \mathbf{1}(B_k)] \geq \epsilon^2 \sum_{k=0}^{n} E\mathbf{1}(B_k) = \epsilon^2 E[\sum_{k=0}^{n} \mathbf{1}(B_k)] = \epsilon^2 P(\bigcup_{k=0}^{n} B_k) \ ,$$

from which the conclusion of the theorem follows immediately. ♠

Chapter 3

Brownian Motion

3.1 Brownian Motion

A stochastic process $\{W_t, t \geq 0\}$ is called Standard Brownian Motion if it satisfies the following three postulates

i) $P(W_0 = 0) = 1$, i.e. the process starts with probability 1 from 0 at time 0.

ii) $\{W_t, t \geq 0\}$ has continuous paths with probability 1.

iii) The increments are independent i.e. if $0 \leq t_i < t_2 < \cdots < t_k$ then $P(W_{t_i} - W_{t_{i-1}} \in H_i; i = 1, 2, \ldots, k) = \prod_{i=1}^{k} P(W_{t_i} - W_{t_{i-1}} \in H_i)$ for any (Borel) subsets H_i of $\mathbb{R}$.

iv) For $0 \leq s < t$, $W_t - W_s$ is normally distributed with mean 0 and variance $t - s$:

$$P(W_t - W_s \in H) = \frac{1}{\sqrt{2\pi(t-s)}} \int_H e^{-x^2/2(t-s)} dx$$

From the above postulates it follows that the finite dimensional distributions of the process W_t are given by

$$P\left(W_{t_1} \in (x_1, x_1 + dx_1), \ldots, W_{t_n} \in (x_n, x_n + dx_n)\right) = f(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) dx_1 \cdots dx_n$$

with

$$\begin{aligned} f(x_1, \ldots, x_n; t_1, t_2, \ldots, t_n) \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{t_1(t_2 - t_1) \cdots (t_n - t_{n-1})}} e^{-\frac{1}{2}\left\{\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \cdots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}}\right\}} \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{|\Sigma|}} e^{-\frac{1}{2} x^T \Sigma^{-1} x} \end{aligned}$$

where x^T denotes the transpose of $x = (x_1, \ldots, x_n)$ and

$$\Sigma = E \begin{pmatrix} W_{t_1} \\ \vdots \\ W_{t_n} \end{pmatrix} (W_{t_1}, \ldots, W_{t_n}) = \begin{bmatrix} EW_{t_1}W_{t_1} & \cdots & EW_{t_1}W_{t_n} \\ \cdots & EW_{t_i}W_{t_j} & \cdots \\ EW_{t_n}W_{t_1} & \cdots & EW_{t_n}W_{t_n} \end{bmatrix} = [t_i \wedge t_j]_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}$$

is the corresponding covariance matrix, i.e. the finite dimensional distributions of brownian motion are normal. This means that brownian motion is a Gaussian process.

3.1.1 Properties of Standard Brownian Motion

1. Markov Property. Brownian motion is a Markov process with stationary transition probabilities

$$\begin{aligned} P_t(x, A) &= P(W_{t+s} \in A | W_s = x) = P(W_{t+s} - W_s \in A - x | W_s = x) \\ &= P(W_t \in A - x) = \int_{A-x} \phi(u) du \end{aligned}$$

   where $A - x$ is the set $\{y - x : y \in A\}$ and $\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$.

2. Scaling Property. $\forall c > 0$ $\{\sqrt{c} W_{t/c}; t \geq 0\} \stackrel{\mathrm{d}}{=} \{W_t; t \geq 0\}$. Indeed, $\sqrt{c} W_{t/c}$ has continuous paths, stationary and independent increments, and the correct distribution.

3. Symmetry. $\{-W_t; t \geq 0\} \stackrel{\mathrm{d}}{=} \{W_t; t \geq 0\}$.

4. Time reversal. $\{tW_{1/t}; t \geq 0\} \stackrel{\mathrm{d}}{=} \{W_t; t \geq 0\}$.

3.2 Martingales associated with Brownian Motion

It is easy to see that standard brownian motion is a martingale. If we denote by $\mathscr{F}_s := \sigma\{W_u; 0 \leq u \leq s\}$ the history of the process up to time s then

$$E[W_t | \mathscr{F}_s] = W_s + E[W_t - W_s | \mathscr{F}_s] = W_s$$

the second term in the above equation vanishing as a result of the independent increments property.

This property, together with the *optional stopping theorem* allows us to compute probabilities of reaching boundaries. Suppose that $W_0 = x$ and let $a < x < b$. Set $\tau = \inf\{t \geq 0 : W_t = a \text{ or } b\}$. Then, by the optional stopping theorem we have

$$EW_\tau = EW_0 = x.$$

However $W_\tau = a\mathbf{1}(W_\tau = a) + b\mathbf{1}(W_\tau = b$, and if we denote by $p_a = P(W_\tau = a)$ (and similarly for p_b) we have $ap_a + bp_b = x$ which gives (since $p_a + p_b = 1$)

$$p_a = \frac{b-x}{b-a}.$$

Similarly, one can easily show that the process $S_t = W_t^2 - t$ is also a martingale. Indeed,

$$\begin{aligned} E[W_t^2 - t|\mathscr{F}_s] &= E[(W_t - W_s)^2 + 2W_s(W_t - W_s) + W_s^2 - t|\mathscr{F}_s] \\ &= E[(W_t - W_s)^2|\mathscr{F}_s] + 2W_s E[W_t - W_s|\mathscr{F}_s] + W_s^2 - t \\ &= (t-s) + 0 + W_s^2 - t = W_s^2 - s \end{aligned}$$

With τ defined as before let us use the optional sampling theorem again. This time we obtain

$$EW_\tau^2 - E\tau = x^2$$

which gives

$$p_a a^2 + p_b b^2 - E\tau = x^2$$

or

$$\frac{(b-x)a^2 + (x-a)b^2}{b-a} - x^2 = E\tau$$

from which we obtain

$$E\tau = ab.$$

An important martingale associated with brownian motion is the exponential martingale. Suppose here that W_t is $BM(\mu, \sigma^2)$. Then, if θ is any real number

$$M_t := e^{\theta W_t - q(\theta)t}, \qquad \text{with } q(\theta) = \mu\theta + \frac{1}{2}\theta^2\sigma^2$$

is a martingale. Indeed,

$$E[M_t|\mathscr{F}_s] = E[e^{\theta(W_t - W_s) - q(\theta)(t-s)}|\mathscr{F}_s]M_s = M_s$$

the last equality following from the fact that $Ee^{\theta(W_t - W_s)} = e^{\mu\theta(t-s) + \frac{1}{2}\theta^2\sigma^2(t-s)}$.

We have thus seen that M_t is a martingale for any choice of θ. If we set $\theta = \theta_0 = -\frac{2\mu}{\sigma^2}$ we see that $q(\theta_0) = 0$ and thus the exponential martingale becomes $e^{\theta_0 W_t}$. We can use this to compute p_a and p_b (defined as before) when $\mu \neq 0$. Indeed, in this case, from the optional sampling theorem we have

$$E[e^{\theta_0 W_\tau}] = e^{\theta_0 x}$$

or

$$p_a e^{\theta_0 a} + p_b e^{\theta_0 b} = e^{\theta_0 x}$$

which gives

$$p_a = \frac{e^{\frac{2\mu}{\sigma^2}(b-x)} - 1}{e^{\frac{2\mu}{\sigma^2}(b-a)} - 1}.$$

The optional sampling theorem can also be used to obtain the Laplace transform of the time until we hit the boundary. Here we will assume that $\mu = 0$, $\sigma = 1$ which corresponds to $q(\theta) = \frac{1}{2}\theta^2$, in order to simplify the algebra. We start with

$$E[e^{\theta W_\tau - \tau q(\theta)}] = e^{\theta x}.$$

or

$$\begin{aligned} e^{\theta x} &= p_a E[e^{\theta W_\tau - \tau q(\theta)} | W_\tau = a] + p_b E[e^{\theta W_\tau - \tau q(\theta)} | W_\tau = b] \\ &= p_a e^{\theta a} E[e^{-q(\theta)\tau} | W_\tau = a] + p_b e^{\theta b} E[e^{-q(\theta)\tau} | W_\tau = b]. \end{aligned}$$

We seem to have the problem that this is one equation and we have two unknowns, $E[e^{-q(\theta)\tau}|W_\tau = a]$ and $E[e^{-q(\theta)\tau}|W_\tau = b]$ but in fact we can get around this problem by setting

$$s = q(\theta) = \frac{1}{2}\theta^2.$$

There are two solutions to this equation,

$$\theta_1 = \sqrt{2s}, \qquad \text{and} \qquad \theta_2 = -\sqrt{2s}.$$

Thus, if we set $f_a(s) = E[e^{-s\tau}; W_\tau = a]$ and $f_b(s) = E[e^{-s\tau}; W_\tau = b]$, we have

$$\begin{aligned} e^{x\sqrt{2s}} &= e^{a\sqrt{2s}} f_a(s) + e^{b\sqrt{2s}} f_b(s) \\ e^{-x\sqrt{2s}} &= e^{-a\sqrt{2s}} f_a(s) + e^{-b\sqrt{2s}} f_b(s). \end{aligned}$$

From this system we can compute $f_a(s)$, $f_b(s)$ separately, and hence also $Ee^{-s\tau} = f_a(s) + f_b(s)$. In fact, adding and subtracting the above equations we get

$$\begin{aligned} \cosh(x\sqrt{2s}) &= \cosh(a\sqrt{2s}) f_a(s) + \cosh(b\sqrt{2s}) \\ \sinh(x\sqrt{2s}) &= \sinh(a\sqrt{2s}) f_a(s) + \sinh(b\sqrt{2s}) \end{aligned}$$

or, using the fact that $\sinh(\alpha - \beta) = \sinh\alpha \cosh\beta - \cosh\alpha \sinh\beta$, we obtain

$$f_a(s) \sinh(b-a)\sqrt{2s} = \sinh(b-x)\sqrt{2s}$$

$$f_b(s) \sinh(b-a)\sqrt{2s} = \sinh(x-a)\sqrt{2s}$$

We thus have

$$f(s) = f_a(s) + f_b(s) = \frac{\sinh\left((x-a)\sqrt{2s}\right) + \sinh\left((b-x)\sqrt{2s}\right)}{\sinh\left((b-a)\sqrt{2s}\right)}$$

and using the formulas $\sinh 2\alpha = 2\sinh\alpha\cosh\alpha$, $\sinh\alpha+\sinh\beta = 2\cosh\left(\frac{\beta-\alpha}{2}\right)\sinh\left(\frac{\alpha+\beta}{2}\right)$ we obtain

$$f(s) = \frac{\cosh\left(\left(\frac{b+a}{2} - x\right)\sqrt{2s}\right)}{\cosh\left(\frac{b-a}{2}\sqrt{2s}\right)}$$

Since we can take $x = 0$ without loss of generality, this formula simplifies as follows

$$f(s) = \frac{\cosh\left(\frac{b+a}{2}\sqrt{2s}\right)}{\cosh\left(\frac{b-a}{2}\sqrt{2s}\right)}.$$

In particular, when $b = \ell > 0$, $a = -\ell$, then

$$f(s) = \frac{1}{\cosh\left(\ell\sqrt{2s}\right)}$$

3.3 Total and Quadratic Variation

Let f be a real function. The *total variation* of f over an interval $[a, b]$ is defined by the limit

$$Vf(a,b) = \lim_{n\to\infty}\sum_{k=0}^{n-1}|f(t_{k+1}) - f(t_k)|$$

$t_k = a + \frac{k}{n}(b-a)$, $k = 0, 1, \ldots, n-1$.

Remark: If f is monotonic, $Vf(a,b) = |f(b) - f(a)|$. To give another example, suppose f is right continuous and there exist points (countably many at the most) T_i, $i = 1, 2, \ldots$, such that f is absolutely continuous on (T_i, T_{i+1}) and has jumps of size J_i at T_i. In that case

$$f(t) = f(a) + \sum_{a<T_i\le t} J_i + \int_a^t f'(u)du \tag{3.1}$$

and

$$Vf(a,b) = \int_a^b |f'(u)|du + \sum_{a<T_i\le b}|J_i|\ .$$

The *quadratic variation* of f is defined as

$$Qf(a,b) = \lim_{n\to\infty}\sum_{k=0}^{n-1}|f(t_{k+1}) - f(t_k)|^2\ .$$

Suppose f is absolutely continuous, i.e. $f(t) = f(a) + \int_0^t f'(u)du$. Then, from the mean value theorem,

$$Qf(a,b) = \lim_{n\to\infty}\sum_{k=0}^{n-1}|f'(t_k+\xi_k)|^2\frac{1}{n^2}$$

where $\xi_k \in (0, \frac{1}{n})$. If $|f'(t)| \leq B \ \forall t \in [a,b]$, $0 \leq Qf(a,b) \leq \lim_{n\to\infty} B^2 \sum_{k=0}^{n-1} \frac{1}{n^2} = 0$. If f is as in (3.1 then

$$Qf(a,b) = \sum_{a<T_i\leq t} J_i^2 \ .$$

3.3.1 Quadratic Variation of Brownian Sample Paths

Let $W(t)$ be a standard Brownian motion. For every fixed $t > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left[W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right]^2 = t \qquad \text{w.p.1} \tag{3.2}$$

i.e. the quadratic variation of brownian paths is $QW(0,t) = t$. (The equality holds both with probability 1 and in the mean square sense.) One implication of this is that the total variation of the paths is infinite

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left|W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right| = \infty \qquad \text{w.p.1.}$$

This is a consequence of the inequality

$$\sum_{k=1}^{2^n} \left|W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right| \geq \frac{\sum_{k=1}^{2^n} \left[W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right]^2}{\max_{j=1,\ldots,2^n} \left|W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right|} \ .$$

The numerator converges to t w.p.1 as $n \to \infty$ while the denominator converges to zero since $W(t)$ is continuous (and hence uniformly continuous on bounded intervals) w.p. 1.

To show (3.2) (with convergence in the mean square sense) consider the sum

$$\sum_{k=1}^{2^n} \left[W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right]^2 \ - \ t \ = \ \sum_{k=1}^{2^n} \delta_{k,n}$$

where

$$\delta_{k,n} := \left(W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right)^2 - \frac{t}{2^n} \ .$$

Note that $E\delta_{k,n} = t2^{-n} - t2^{-n} = 0$ and $E\delta_{k,n}^2 = 3 \cdot 2^{-2n}t^2$ (fourth moment of a zero mean normal random variable). It suffices to show that $\sum_{k=1}^{2^n} \delta_{k,n} \overset{\text{m.s.}}{\to} 0$. The independence of brownian motion increments implies that

$$E\left(\sum_{k=1}^{2^n} \delta_{k,n}\right)^2 = \sum_{k=1}^{2^n} E[\delta_{k,n}^2] = 2^n \cdot 3 \cdot 2^{-2n}t^2 = 3 \cdot 2^{-n}t^2 \to 0 \quad \text{as } n \to \infty.$$

Thus we have shown that (3.2) holds in the L^2 sense. We next show that it holds w.p. 1 as well. Fix $\epsilon > 0$ and let

$$A_n = \{\omega : |\sum_{k=1}^{2^n} \delta_{k,n}| > \epsilon\}.$$

Then, from Chebychev's inequality

$$P(A_n) \leq \frac{E(\sum_{k=1}^{2^n} \delta_{k,n})^2}{\epsilon^2} = \frac{3t^2}{\epsilon^2 2^n}$$

and thus

$$\sum_{n=1}^{\infty} P(A_n) = \frac{3t^2}{\epsilon^2} \sum_{n=1}^{\infty} 2^{-n} < \infty.$$

Hence, for any given ϵ, the Borel–Cantelli theorem implies that only finitely many of the A_n occur, i.e. that, for any ϵ there exists $n_0(\epsilon)$ such that $n > n_0(\epsilon)$ implies $\sum_{k=1}^{2^n} \delta_{k,n} < \epsilon$. This establishes convergence w.p. 1 in (3.2).

3.4 Gaussian Processes

A stochastic process $\{X_t; t \in \mathbb{R}\}$ is called a Gaussian process if, for every k and every $t_1 < t_2 < \cdots < t_k$, the distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is multidimensional Gauss. It is clear that to define the joint distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ it suffices to determine the vector $(m(t_1), m(t_2), \ldots, m(t_k))$ and the covariance matrix

$$\begin{bmatrix} R(t_1,t_1) & R(t_1,t_2) & \cdots & R(t_1,t_k) \\ R(t_2,t_1) & R(t_2,t_2) & \cdots & R(t_2,t_k) \\ \vdots & \vdots & & \vdots \\ R(t_k,t_1) & R(t_k,t_2) & \cdots & R(t_k,t_k) \end{bmatrix}. \quad (3.3)$$

A Gaussian process with $m(t) = 0$ is called a *centered* Gaussian process. The standard Brownian motion is a centered Gaussian process with covariance function given by $R(t_i, t_j) = t_i \wedge t_j$. A Gaussian process for which $m(t) = \mu$ for all t and $R(s,t) = r(|t-s|)$ is called stationary since, in that case, $(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+s}, \ldots, X_{t_n+s})$ for all n, $t_1, \ldots, t_n$, s. The standard Brownian motion is not stationary. An example of such a process is the stationary Ornstein-Uhlenbeck process with covariance function $R(s,t) = \sigma^2 e^{-\alpha|t-s|}$ where $\alpha > 0$.

Chapter 4

Stochastic Integrals

4.1 L^2 spaces of random variables

4.1.1 A brief overview of linear spaces of random variables

Consider the family, L^2, of all random variables on the probability space $(\Omega, \mathscr{F}, P)$ that have zero mean and finite second moment, i.e. for every $X \in L^2$, $EX = 0$, $EX^2 < \infty$. It is easy to see that this family is *a linear space over* $\mathbb{R}$ i.e. that it satisfies the axioms of a real linear space. In fact, the only property that we need to check is that, if X, Y, belongs to L^2 then $X + Y$ belongs to L^2 as well. This however is a consequence of the inequality $\|X+Y\| \leq \|X\| + \|Y\|$. Thus, the finiteness of the second moment of X and Y implies that of their sum, $X+Y$. A norm is a function $\|\cdot\| : L^2 \to \mathbb{R}_0^+$ from the elements of L^2 to the nonnegative reals that has the following properties

N1. (Nonnegativity) For all $x \in L^2$, $\|x\| \geq 0$,

N2. $\|x\| = 0$ iff $x = 0$,

N3. $\|\alpha x\| = |\alpha| \, \|x\|$ for all $x \in L^2$, $\alpha \in \mathbb{R}$.

N4. (Triangular inequality) $\|x + y\| \leq \|x\| + \|y\|$

An inner product is a function $(\cdot, \cdot) : L^2 \times L^2 \to \mathbb{R}$ such that

IP1. $\langle aX, Y \rangle = \langle X, aY \rangle = a\langle X, Y \rangle$

IP2. $\langle X + Y, Z \rangle = \langle X, Z \rangle + \langle Y, Z \rangle$

We define an *inner product* on L^2 via the relationship

$$\langle X, Y \rangle := EXY \tag{4.1}$$

A linear space on which an inner product has been defined is an *inner product space.* The inner product induces a norm via the definition $\|x\| = \sqrt{\langle x, x \rangle}$.

The elements $x_1, x_2, \dots, x_n$ of L^2 are *linearly independent* iff

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0$$

implies $\alpha_i = 0$, $i = 1, 2, \dots, n$.

At this point it is customary to define the *dimension* of the linear space as the maximum number of linearly independent elements of the space. In the ordinary Euclidean space $\mathbb{R}^n$ this dimension is of course n. However, since we are willing to assume that our probability space $(\Omega, \mathscr{F}, P)$ is rich enough to support sequences of independent random variables X_i, we have to dispense with the requirement that our space has finite dimension.

Note that two random variables that differ only on a set of measure 0 have to be identified here: Indeed, if $P(X = Y) = 1$ then certainly $E(X-Y)^2 = 0$, hence $\|X-Y\| = 0$ which implies that $X - Y = 0$ according to (N2). Thus when we deal with random variables in L^2 we have to think of them rather as equivalence classes.

A sequence of elements of L^2, $\{X_n\}$, is said to converge to an element of $X \in L^2$ if $\|X_n - X\| \to 0$ as $n \to \infty$. Note that this is precisely L^2 convergence for the sequence of random variables.

A sequence $\{X_n\}$ is Cauchy, if

$$\|X_n - X_m\| \to 0 \qquad \text{as } n, m \to \infty. \tag{4.2}$$

Clearly every convergent sequence is Cauchy since, if $X_n \to X$ then, using the triangular inequality (N4) we have

$$\|X_n - X_m\| \le \|X_n - X\| + \|X_m - X\|$$

and each of the two terms on the right side go to 0 as n and m go to infinity. On the other hand, a Cauchy sequence is not necessarily convergent. While (4.2) guarantees that the elements of the sequence approach each other more and more as m and n grow large, there is no guarantee that the limit this sequence is approaching is actually an element of L^2.

All Cauchy sequences are bounded, i.e. if $\{X_n\}$ is a Cauchy sequence then $\sup_n \|X_n\| < \infty$ which means that there exists $M > 0$ such that $EX_n^2 \le M$ for all $n \in \mathbb{N}$.

The space L^2 is *complete* if every Cauchy sequence of elements of L^2 converges to an element of L^2.

Theorem 12. *L^2 is complete.*

Proof. We start with a Cauchy sequence X_n of elements of L^2. According to the definition we have to show that there exists a random variable X with $EX^2 < \infty$ such that $\|X_n - X\| \to 0$ as $n \to \infty$. Choose a subsequence n_k such that $\|X_m - X_n\| < 2^{-3k/2}$ when m and n are greater than or equal to n_k. This means that $E(X_{n_{k+1}} - X_{n_k})^2 < 2^{-3k}$. Using Chebychev's inequality we have

$$P\left(\left|X_{n_{k+1}} - X_{n_k}\right| > 2^{-k}\right) \leq \frac{E(X_{n_{k+1}} - X_{n_k})^2}{2^{-2k}} < 2^{-k}.$$

Since

$$\sum_{k=1}^{\infty} P\left(\left|X_{n_{k+1}} - X_{n_k}\right| > 2^{-k}\right) < \sum_{k=1}^{\infty} 2^{-k} < \infty,$$

the Borel–Cantelli lemma insures that, with probability 1, only finitely many of the inequalities $\left|X_{n_{k+1}} - X_{n_k}\right| > 2^{-k}$ are true. This is equivalent to saying that there exists some k_0 (which may depend on ω) such that, for all $k \geq k_0$, $\left|X_{n_{k+1}} - X_{n_k}\right| \leq 2^{-k}$. Hence, w.p. 1 the series $\sum_{k=1}^{\infty} |X_{n_{k=1}} - X_{n_k}|$ converges (since it is dominated by a convergent series). This in turn implies that the telescopic series $\sum_{k=1}^{\infty}(X_{n_{k+1}} - X_{n_k})$ converges absolutely and thus that $\lim_{k\to\infty} X_{n_k} =: X$ exists.

We next show that $EX^2 < \infty$ and hence that X is an element of L^2. Fix $k \in \mathbb{N}$. Since $E\left(X_{n_m} - X_{n_k}\right)^2 < 2^{-3k}$ for all $m > k$, from Fatou's lemma it follows that

$$E[\liminf_{m\to\infty} \left(X_{n_m} - X_{n_k}\right)^2] \leq \liminf_{m\to\infty} E[(X_{n_m} - X_{n_k})^2] \leq 2^{-3k}$$

and since $\lim_{m\to\infty} X_{n_m} = X$ w.p.1 this can also be written as $\|X - X_{n_k}\|^2 \leq 2^{-2k}$ or $\|X - X_{n_k}\| \leq 2^{-3k/2}$. From the triangle inequality we then have

$$\|X\| \leq \|X_{n_k}\| + 2^{-3k/2} < \infty$$

since $X_{n_k} \in L^2$. This shows that X is also an element of L^2.

Finally we have to show that $\|X_n - X\| \to 0$ when $n \to \infty$. Indeed, for any given ϵ, choose n_k such that $\|X_{n_k} - X\| < \epsilon/2$ and N such that $\|X_n - X_m\| < \epsilon/2$ whenever $m \geq N$, $n \geq N$. Then, from the triangle inequality,

$$\|X_n - X\| \leq \|X_n - X_{n_k}\| + \|X_{n_k} - X\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

∎

4.2 Integration with respect to functions of bounded variation

To understand the challenges involved in defining the stochastic integral we have to recall first the definition of the ordinary integral. Historically, arriving at a satisfactory definition was by no means a simple task and it has only been completed in the first two decades of the twentieth century. Suppose that F is a function of bounded variation, f a continuous function defined on a closed interval $[a, b]$, and let $t_i^n := a + \frac{i}{n}(b - a)$, $i = 0, 1, 2, \ldots, n$. Then the so–called *Riemann–Stieltjes* integral can be defined as

$$\int_a^b f(x)dF(x) := \lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i^n) \left[F(t_{i+1}^n) - F(t_i^n)\right].$$

There is of course nothing special about the equally spaced partition we have used above and in fact, one can show that any partition, $a = t_0^n < t_1^n < \cdots < t_i^n < \cdots < t_n^n = b$ of the interval $[a, b]$ will yield the same limit as $n \to \infty$, provided of course that $\max_{0\le i\le n-1}(t_{i+1}^n - t_i^n) \to 0$ as $n \to \infty$.

Why do we require F to be a function of bounded variation? The reason is that functions of bounded variation correspond to *signed measures.* Any increasing function F defines a measure on the real line via the relationship $\mu(a, b] = F(b) - F(a)$. Think of the measure $\mu(a, b]$ of the interval $(a, b]$ as the total mass of the interval. Since F is increasing, the mass of any interval is nonnegative and if F is absolutely continuous, i.e. if $F(x) - F(0) = \int_0^x F'(u)du$, the derivative of F correspond to the mass density.

Similarly, a function of bounded variation can be written as the difference of two increasing functions G^+ and G^-: $F(x) = G^+(x) - G^-(x)$. This representation is unique (up to an arbitrary initial value, say $G^-(-\infty) = G^+(-\infty) = 0$). Thus, if we can think of increasing functions as mass distributions on the real line, we can think of bounded variation as electrical charge distributions that can be positive in some places and negative in others. In this case the signed measure μ of the interval $(a, b]$ is the total charge of the interval (positive − negative) i.e. $\mu(a, b] = F(b) - F(a) = (G^+(b) - G^+(a)) - (G^-(b) - G^-(a))$.

The real problem that presents itself when we try to define

$$\int_0^t f_s dW_s$$

is that, since W_t has paths of infinite total variation, they do not define a (signed) measure the way a bounded variation function does, so it is not at all clear how to define the integral and what its precise meaning would be.

4.3 Definition of the Ito Integral

Denote by H^2 the set of all adapted processes on the filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, P)$ satisfying

$$E\int_0^t X_s^2 ds < \infty \qquad \forall t \geq 0.$$

($\mathbb{F} = \{\mathscr{F}_t\}$ is the filtration and an adapted process is one for which $X_t \in \mathscr{F}_t$ for all $t \geq 0$. A process X is *simple* if there exists a real sequence $\{t_k\}$, $t_k \to \infty$ as $k \to \infty$ and a sequence of random variables $\{F_k\}$ such that $F_k \in \mathscr{F}_{t_k}$ for all k and

$$X(t,\omega) = \sum_{k=1}^{\infty} F_k(\omega)\mathbf{1}_{[t_k,t_{k+1}]}(t).$$

Also define

S: The set of all simple adapted processes,

S^2: The set of all simple adapted processes in H^2,

L^2: The set of all random variables ξ in $(\Omega, \mathscr{F}, P)$ such that $E(\xi^2)^{1/2} < \infty$.

Define a norm in H^2 by means of

$$\|X\| = \left(E\int_0^t X_s^2 ds\right)^{1/2}.$$

Theorem 13. *S^2 is dense in H^2 i.e. for all $X \in H^2$ there exist simple processes $\{X_n\}$ such that*

$$\|X_n - X\| \to 0, \qquad n \to \infty$$

We will denote the stochastic integral which we are about to define by $I_t(X) := \int_0^t X_s dW_s$, $t \geq 0$. For simple processes this task is easy. We set

$$I(X) = \sum_{k=0}^{n-1} X_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right) . \tag{4.3}$$

The stochastic integral defined above has the following two important properties.

Proposition 6. *For $X \in S^2$, $EI(X) = 0$ and $\|I(X)\| = \|X\|$.*

Proof:

$$EI(X) = \sum_{k=0}^{n-1} E\left[\, E\left[X_{t_k}\left(W_{t_{k+1}} - W_{t_k}\right)\middle|\, \mathscr{F}_{t_k}\right]\,\right]$$

However

$$E\left[X_{t_k}\left(W_{t_{k+1}}-W_{t_k}\right)\middle|\mathscr{F}_{t_k}\right] = X_{t_k}E\left[W_{t_{k+1}}-W_{t_k}\middle|\mathscr{F}_{t_k}\right]=0$$

whence we obtain $EI(X)=0$. To show that the *isometry* $\|I(X)\|=\|X\|$ holds as well, i.e. that $EI(X)^2=E\int_0^t X_s^2ds$ we first note that

$$I^2(X)=\sum_{k=0}^{n-1}X_{t_k}^2\left(W_{t_{k+1}}-W_{t_k}\right)^2 + 2\sum_{j=0}^{n-2}\sum_{k=j+1}^{n-1}X_{t_j}X_{t_k}\left(W_{t_{j+1}}-W_{t_j}\right)\left(W_{t_{k+1}}-W_{t_k}\right).$$

Taking expectations on both sides of the above equation we will have to deal with two types of terms.

$$E[X_{t_k}^2\left(W_{t_{k+1}}-W_{t_k}\right)^2] \tag{4.4}$$

and

$$E[X_{t_j}X_{t_k}\left(W_{t_{j+1}}-W_{t_j}\right)\left(W_{t_{k+1}}-W_{t_k}\right)] \tag{4.5}$$

Both expectations can be computed by conditioning appropriately. (4.4) becomes

$$\begin{aligned} E\left[\,E\left[X_{t_k}^2\left(W_{t_{k+1}}-W_{t_k}\right)^2\middle|\mathscr{F}_{t_k}\right]\,\right] &= E\left[\,X_{t_k}^2E\left[\left(W_{t_{k+1}}-W_{t_k}\right)^2\middle|\mathscr{F}_{t_k}\right]\,\right] \\ &= E\left[\,X_{t_k}^2(t_{k+1}-t_k)\right]. \end{aligned}$$

To compute (4.5) note that $j<k$ hence $j+1\le k$ and $t_{j+1}\le t_k$, which implies that $\mathscr{F}_{t_{j+1}}\subseteq\mathscr{F}_{t_k}$. Hence the expectation in (4.5) becomes

$$\begin{aligned} &E\left[\,E\left[X_{t_j}X_{t_k}\left(W_{t_{j+1}}-W_{t_j}\right)\left(W_{t_{k+1}}-W_{t_k}\right)\middle|\mathscr{F}_{t_k}\right]\,\right] \\ &\qquad = E\left[\,X_{t_j}X_{t_k}\left(W_{t_{j+1}}-W_{t_j}\right)E\left[\left(W_{t_{k+1}}-W_{t_k}\right)\middle|\mathscr{F}_{t_k}\right]\,\right] \;=0, \end{aligned}$$

the last equation following since $E\left[\left(W_{t_{k+1}}-W_{t_k}\right)\middle|\mathscr{F}_{t_k}\right]=0$. Thus we have

$$EI(X)^2=\sum_{k=0}^{n-1}E\left[X_{t_k}^2(t_{k+1}-t_k)\right]=E\left[\sum_{k=0}^{n-1}X_{t_k}^2(t_{k+1}-t_k)\right]=E\int_0^t X_s^2ds \tag{4.6}$$

since X_t is a simple process. ♠

Proposition 7. *Suppose $X\in H^2$. There exists a random variable $I(X)\in L^2$, unique up to a null set, such that $I(X_n)\to I(X)$.*

Proof: Let $\{X_n\}$ be a sequence in S^2 such that $X_n\to X$. Then $\|X_n-X_m\|\to 0$ (Cauchy sequence in H^2). From the previous proposition

$$\|I(X_m)-I(X_n)\|=\|I(X_m-X_n)\|=\|X_n-X_m\|\to 0$$

Hence $\{I(X_n)\}$ is a Cauchy sequence in L^2 and, in view of the completeness of L^2, there exists a random variable $I(X)\in L^2$ such that $I(X_n)\to I(X)$. We have thus been able to define the stochastic integral $I(X)$ for arbitrary integrants $X_t\in H^2$ (not necessarily simple processes) by means of approximating them by sequences of simple processes. This definition however would not be satisfactory if the resulting limit $I(X)$ depended on the approximating sequence. In other words it is essential to establish uniqueness, i.e. to show that any other sequence of simple processes would lead to the same result. Suppose that $\{X'_n\}$ is another S^2 sequence such that $X'_n \to X$. Then

$$\|X_n - X'_n\| \le \|X_n - X\| + \|X - X'_n\| \to 0$$

where we have used the triangle inequality and the fact that both X_n and X'_n converge in H^2 to X. However, using the linearity of the integral $I(X)$ for simple functions and the isometry $\|I(Y)\| = \|Y\|$ we have established for any simple process we have

$$\|I(X_n) - I(X'_n)\| = \|I(X_n - X'_n)\| = \|X_n - X'_n\| \to 0$$

Thus,

$$\|I(X'_n) - I(X)\| \le \|I(X'_n) - I(X_n)\| + \|I(X_n) - I(X)\| \to 0 \,.$$

This establishes the uniqueness of the stochastic integral $I(X)$ since it shows that $I(X'_n) \to I(X)$ in L^2. Finally, in order to show that $EI(X) = 0$ and $\|I(X)\| = \|X\|$ note that, for any sequence of random variables $\xi_n \to \xi$ in L^2, $E\xi_n \to E\xi$ and $\|\xi_n\| \to \|\xi\|$ as a consequence of the Dominated Convergence Theorem. ♠

Note that, up to this point, we have defined the Ito integral of a process X in H^2 for each t. We have not however defined $I_t(X)$ as a function of t, i.e. as a function of the upper limit of integration. This will be done presently. Let us see first an example.

We will compute explicitly $I_t(W) = \int_0^t W_s dW_s$. Since $W_t \in \mathscr{F}_t$ by assumption the integrand is an adapted process. Also $E\int_0^t W_s^2 ds = \int_0^t EW_s^2 = \int_0^t s ds = t^2/2 < \infty$, thus $W \in H^2$. Fix $t > 0$ and consider the simple functions $\{X_n\}$ defined by

$$X_n(s) = W\left(tk2^{-n}\right) \qquad \text{for } s \in \left[\tfrac{kt}{2^n}, \tfrac{(k+1)t}{2^n}\right), \qquad k = 0, 1, 2, \ldots, 2^n - 1.$$

It is easy to see that $\{X_n\}$ is a sequence of adapted processes in S^2. Also

$$\begin{aligned}
\|W - X_n\| &= E\int_0^t (W_s - X_n(s))^2 ds = \int_0^t E(W_s - X_n(s))^2 ds \\
&= \sum_{k=0}^{2^n-1} \int_0^{t/2^n} E\left(W\left(tk2^{-n} + s\right) - W\left(tk2^{-n}\right)\right)^2 ds = \sum_{k=0}^{2^n-1} \int_0^{t/2^n} s ds \\
&= 2^n \cdot \frac{1}{2}\left(\frac{t}{2^n}\right)^2 = \frac{t^2}{2^{n+1}}
\end{aligned}$$

Thus

$$\|W - X_n\| = \frac{t}{2^{\frac{n+1}{2}}} \to 0$$

which implies $I_t(X_n) \to I_t(W)$.

Write for simplicity $t_k = \frac{kt}{2^n}$.

$$\begin{aligned}
I_t(X_n) &= \sum_{k=0}^{2^n-1} W(t_k)\,[W(t_{k+1}) - W(t_k)] \\
&= \frac{1}{2}\sum_{k=0}^{2^n-1} \left[W^2(t_{k+1}) - W^2(t_k)\right] - \left[W^2(t_{k+1}) + W^2(t_k) - 2W(t_k)W(t_{k+1})\right] \\
&= \frac{1}{2}\sum_{k=0}^{2^n-1} \left[W^2(t_{k+1}) - W^2(t_k)\right] - \frac{1}{2}\sum_{k=0}^{2^n-1} [W(t_{k+1}) - W(t_k)]^2 \\
&= \frac{1}{2}W^2(t) - \frac{1}{2}\sum_{k=0}^{2^n-1} [W(t_{k+1}) - W(t_k)]^2
\end{aligned}$$

However the last term in the above string of equations is the quadratic variation of the brownian motion and it converges (in L^2) to t:

$$\sum_{k=0}^{2^n-1} [W(t_{k+1}) - W(t_k)]^2 \ \stackrel{L^2}{\longrightarrow}\ t$$

Consequently

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t$$

The above explicit computation can be repeated for other type of integrands. It is akin to the evaluation of integrals in ordinary calculus via approximating sequences. In practice stochastic integrals are most often evaluated via the Ito formula (which is essentially the stochastic counterpart of the "change–of–variables" formula of ordinary calculus).

4.4 The Ito Formula

Suppose that X_s, Y_s are adapted processes in H^2 and Z_t is an *Ito process*, i.e. a process expressed as

$$Z_t = Z_0 + \int_0^t X_s dW_s + \int_0^t Y_s ds. \tag{4.7}$$

The above is also often expressed in shorthand differential form (even though it only makes symbolic sense) as

$$dZ_t = X_t dW_t + Y_t dt. \tag{4.8}$$

Theorem 14. ***[Ito formula]*** *Suppose $f : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable and Z is given by (4.7). Then*

$$f(Z_t) = f(Z_0) + \int_0^t f'(Z_s) X_s dW_s + \int_0^t f'(Z_s) Y_s ds + \frac{1}{2}\int_0^t f''(Z_s) X_s^2 ds$$

In order to evaluate stochastic integrals of the form $\int_0^t f(W_s)dW_s$ we can apply the above formula with $X_s = 1$, $Y_s = 0$ which gives $Z_t = W_t$, and $F(t) = F(0) + \int_0^t f(s)ds$. Of course, f must be continuously differentiable in order for F'' to exist and be continuous. Then the Ito formula gives

$$F(W_t) = F(0) + \int_0^t f(W_s)dW_s + \frac{1}{2}\int_0^t f'(W_s)ds \tag{4.9}$$

For instance, suppose we wanted to evaluate $\int_0^t W_s dW_s$. Take $f(x) = \frac{1}{2}x^2$ and apply the above formula to obtain

$$\frac{1}{2}W_t^2 = \int_0^t W_s dW_s + \frac{1}{2}\int_0^t 1ds$$

which gives the result we had obtained in the previous section.

Similarly, to compute $\int_0^t e^{W_s}dW_s$ take $f(x) = e^x$ to obtain

$$e^{W_t} = 1 + \int_0^t e^{W_s}dW_s + \frac{1}{2}\int_0^t e^{W_s}ds$$

whence we obtain

$$\int_0^t e^{W_s}dW_s = e^{W_t} - 1 - \frac{1}{2}\int_0^t e^{W_s}ds.$$

Note that the integral appearing on the right hand side of the above equation is an ordinary Riemann integral, in view of the continuity of the paths of Brownian motion with probability 1.

We now proceed to give the proof of (4.9).

Proof of (4.9) We shall establish this special case of the Ito formula under the additional assumption that $\int_0^t E[f(W_s)]ds < \infty$. The integral $\int_0^t f(W_s)dWs$ makes sense as an Ito integral since the process $f(W_s)$ is adapted to $\mathscr{F}_s$ and $E\int_0^t f(W_s)dW_s = \int_0^t E[f(W_s)]ds < \infty$. Then if we set $t_k^{(n)} := \frac{kt}{2^n}$ for $k = 0, 1, 2, \ldots, 2^n - 1$ and we define an approximating sequence of simple processes via

$$X_n(s) = f(W(t_k^{(n)})) \qquad \text{when } s \in \left[t_k^{(n)}, t_{k+1}^{(n)}\right), \qquad k = 1, 2, \ldots, 2^n - 1.$$

In the sequel we shall suppress the dependence of $t_k^{(n)}$ on n and write simply t_k. The simple processes X_n are obviously adapted and belong to S^2, hence we can define their Ito integrals as

$$I(X_n) = \sum_{k=0}^{2^n-1} f(W(t_k))\left(W(t_{k+1}) - W(t_k)\right).$$

Let us use now Taylor's theorem for the function F (remember that $F' = f$) to obtain

$$F(W(t_{k+1})) - F(W(t_k)) = f(W(t_k))(W(t_{k+1} - W(t_k)) + \frac{1}{2}f'(\xi_k)(W(t_{k+1} - W(t_k))^2$$

where ξ_k is between $W(t_k)$ and $W(t_{k+1})$. Thus

$$\begin{aligned} I(X_n) &= \sum_{k=0}^{2^n-1} F(W(t_{k+1})) - F(W(t_k)) \quad - \quad \frac{1}{2}\sum_{k=0}^{2^n-1} f'(\xi_k)(W(t_{k+1} - W(t_k))^2 \\ &= F(W(t)) - F(0) \quad -\frac{1}{2}\sum_{k=0}^{2^n-1} f'(\xi_k)(W(t_{k+1}) - W(t_k))^2 \end{aligned}$$

since the first sum is telescopic. It remains to examine the limit of the second sum as $n \to \infty$. In fact we will show that it converges in L^2 to $\frac{1}{2}\int_0^t f'(W_s)ds$. To simplify the notation set

$$\begin{aligned} \Phi_n &= \sum_{k=0}^{2^n-1} f'(\xi_k)(W(t_{k+1}) - W(t_k))^2, \\ \Phi &= \int_0^t f'(W_s)ds, \\ \Psi_n &= \sum_{k=0}^{2^n-1} f'(\xi_k)(t_{k+1} - t_k). \end{aligned}$$

To show that $\Phi_n \overset{L^2}{\longrightarrow} \Phi$ we must establish that $\|\Phi_n - \Phi\| \to 0$. Using the triangle inequality

$$\|\Phi_n - \Phi\| \le \|\Phi_n - \Psi_n\| + \|\Psi_n - \Phi\|.$$

Now we have

$$\|\Phi_n - \Psi_n\|^2 = E\left(\sum_{k=0}^{2^n-1} f'(\xi_k)\left[(W(t_{k+1}) - W(t_k))^2 - (t_{k+1} - t_k)\right]\right)^2$$

To ease the notation define

$$\delta_{k,n} := \left[(W(t_{k+1}) - W(t_k))^2 - (t_{k+1} - t_k)\right] \tag{4.10}$$

We have used the same quantities before, namely when we were trying to compute the quadratic variation of Brownian motion. There we had seen that

$$E\delta_{k,n} = t2^{-n} \qquad \text{and} \qquad E\delta_{k,n}^2 = 3t^2 2^{-2n}. \tag{4.11}$$

Thus

$$\|\Phi_n - \Psi_n\|^2 = E\left(\sum_{k=0}^{2^n-1} f'(\xi_k)^2\delta_{k,n}^2\right) + 2E\left(\sum_{k=1}^{2^n-1}\sum_{l=0}^{k-1} f'(\xi_k)f'(\xi_l)\delta_{k,n}\delta_{l,n}\right)$$

At this point we note that the second expectation on the right hand side of the above equation vanishes. Also, since f' is continuous on $[0, t]$, and therefore bounded on this interval, say by M, we obtain the inequality

$$\|\Phi_n - \Psi_n\|^2 \le \sum_{k=0}^{2^n-1} ME\delta_{k,n}^2 \le 3t^2 2^{-2n} M 2^n = 3t^2 M 2^{-n} \to 0,$$

the second inequality following from (4.11). This completes the proof. ♠

4.5 A more general Ito formula

Often in applications the following, more general Ito formula is useful. Suppose that we are given a function $F : \mathbf{R}^2 \to \mathbf{R}$ continuously differentiable with respect of its first argument, t, and twice continuously differentiable with respect to its second argument, z. If, as before, Z_t is an Ito process, i.e.

$$dZ_t = X_t dW_t + Y_t dt,$$

then the following change of variables holds

$$\begin{aligned} dF(t, Z_t) &= \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial z} dZ_t + \frac{1}{2}\frac{\partial^2 F}{\partial z^2}(dZ_t)^2 \\ &= \left(\frac{\partial F}{\partial t} + \frac{\partial F}{\partial z} Y_t + \frac{1}{2}\frac{\partial^2 F}{\partial z^2} X_t^2\right) dt + \frac{\partial F}{\partial z} X_t dW_t. \end{aligned} \tag{4.12}$$

In integral form this can be written as

$$F(t, Z_t) - F(0, Z_0) = \int_0^t \left(\frac{\partial F}{\partial s} + \frac{\partial F}{\partial z} Y_s + \frac{1}{2}\frac{\partial^2 F}{\partial z^2} X_s^2\right) ds + \int_0^t \frac{\partial F}{\partial z} X_s dW_s.$$

4.6 Multidimensional version of Ito' s formula

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function and X_t and Ito process described by the equation

$$dX_t = u_t dt + v_t dW_t$$

where

$$u_t := \begin{pmatrix} u_t^1 \\ u_t^2 \\ \vdots \\ u_t^n \end{pmatrix}, \qquad v_t := \begin{bmatrix} v_t^{11} & v_t^{12} & \cdots & v_t^{1m} \\ v_t^{21} & v_t^{22} & \cdots & v_t^{2m} \\ \vdots & \vdots & & \vdots \\ v_t^{n1} & v_t^{n2} & \cdots & v_t^{nm} \end{bmatrix}, \qquad W_t := \begin{pmatrix} W_t^1 \\ W_t^2 \\ \vdots \\ W_t^m \end{pmatrix}$$

It is assumed that the processes u_t^i, v_t^{ij} are adapted and that W_t^i, $i = 1, 2, \ldots, m$ are independent standard brownian motions. The Ito formula is written symbolically as

$$df(X_t) = \nabla f dX_t + \frac{1}{2} dX_t^T H dX_t \tag{4.13}$$

where

$$H := \begin{bmatrix} D_{11}f(X_t) & D_{12}f(X_t) & \cdots & D_{1n}f(X_t) \\ D_{21}f(X_t) & D_{22}f(X_t) & \cdots & D_{2n}f(X_t) \\ \vdots & \vdots & & \vdots \\ D_{n1}f(X_t) & D_{n2}f(X_t) & \cdots & D_{nn}f(X_t) \end{bmatrix}.$$

We thus have

$$
\begin{aligned}
dX_t^T H dX_t &= dW_t^T V^T H V dW_t = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{m} dW_t^i V_{ij}^T H_{jk} V_{kl} dW_t^l \\
&= dt \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} V_{ij}^T H_{jk} V_{ki} = dt \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} V_{ji} H_{jk} V_{ki}
\end{aligned}
$$

where in the above string of equalities we have taken into account that

$$
dW_t^i dW_t^j = \delta_{ij} dt.
$$

(δ_{ij}, called Kronecker's delta, is defined to be equal to 1 if $i = j$ and zero otherwise.) We can thus write (4.13) in more detailed form as

$$
\begin{aligned}
df(X_t) &= \sum_{i=1}^{n} D_i f(X_t) u_t^i dt + \sum_{i=1}^{n} D_i f(X_t) \sum_{j=1}^{m} v_t^{ij} dW_t^j \\
&\quad + \frac{1}{2}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} v_t^{ji} D_{jk} f(X_t) v_t^{ki} \right) \qquad (4.14)
\end{aligned}
$$

Chapter 5

Hermite Polynomials, Brownian Motion, and Gaussian Spaces

Let $\{W_t; t \geq 0\}$ be standard Brownian motion. For $\theta \in \mathbb{R}$ define the process

$$U_t(\theta) = e^{\theta W_t - \frac{1}{2}\theta^2 t}, \qquad t \geq 0. \tag{5.1}$$

Show that $\{U_t(\theta); t \geq 0\}$ is a martingale.

The nth derivative of $U_t(\theta)$ with respect to θ evaluated at $\theta = 0$ is also a martingale. (Provide a proof or accept it as given and continue.) What are the martingales obtained in this fashion for $n = 1, 2, 3, 4$?

The *Hermite polynomials* are defined as follows:

$$h_n(x) := (-1)^n e^{\frac{1}{2}x^2} \frac{d^n}{dx^n} e^{-\frac{1}{2}x^2}, \qquad n = 0, 1, 2, 3, \ldots. \tag{5.2}$$

(There are slightly different definitions in other areas, notably in Physics. This is the standard definition in Probability.) From this definition we see that the first few Hermite polynomials are

$$\begin{aligned} h_0(x) &= 1 \\ h_1(x) &= x \\ h_2(x) &= x^2 - 1 \\ h_3(x) &= x^3 - 3x \\ h_4(x) &= x^4 - 6x^2 + 3. \end{aligned}$$

Consider the expression $e^{\theta x - \frac{1}{2}\theta^2}$ as a function of $\theta \in \mathbb{R}$.

Consider the analytic function $f(x) := e^{-\frac{1}{2}x^2}$. The following Taylor expansion is valid:

$$f(x-\theta) = \sum_{n=0}^{\infty} \frac{(-\theta)^n}{n!} f^{(n)}(x)$$

or, equivalently

$$e^{-\frac{1}{2}(x-\theta)^2} = e^{-\frac{1}{2}x^2+\theta x-\frac{1}{2}\theta^2} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!}(-1)^n \frac{d^n}{dx^n} e^{-\frac{1}{2}x^2}$$

which gives

$$e^{\theta x-\frac{1}{2}\theta^2} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!}(-1)^n e^{\frac{1}{2}x^2} \frac{d^n}{dx^n} e^{-\frac{1}{2}x^2}$$

or, equivalently,

$$e^{\theta x-\frac{1}{2}\theta^2} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} h_n(x). \tag{5.3}$$

This last expression provides the generating function for the Hermite polynomials.

Consider now the related expansion

$$e^{\theta x-\frac{1}{2}t\theta^2} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} H_n(x,t). \tag{5.4}$$

Here of course $H_n(x,t) := \frac{d^n}{d\theta^n} e^{\theta x-\frac{1}{2}t\theta^2}\Big|_{\theta=0}$. Setting $\lambda = \theta\sqrt{t}$, $y = x/\sqrt{t}$ we have $x\theta - \frac{1}{2}\theta^2 t = y\lambda - \frac{1}{2}\lambda^2$ and thus

$$e^{\theta x-\frac{1}{2}t\theta^2} = e^{\lambda y-\frac{1}{2}\lambda^2} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} h_n(y) = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} t^{n/2} h_n(x/\sqrt{t}).$$

Comparing with (5.4) we obtain

$$H_n(x,t) = t^{n/2} h_n(x/\sqrt{t}). \tag{5.5}$$

Now, suppose that X, Y are jointly normal random variables with mean zero, variance 1, and covariance ρ. Their moment generating function is given by

$$\mathbb{E}e^{\theta X+\eta Y} = e^{\frac{1}{2}\theta^2+\rho\eta\theta+\frac{1}{2}\eta^2}. \tag{5.6}$$

The above is equivalent to writing

$$\mathbb{E}[e^{\theta X-\frac{1}{2}\theta^2} e^{\eta Y-\frac{1}{2}\eta^2}] = e^{\rho\eta\theta}.$$

$$\mathbb{E}\left[\sum_{n=0}^{\infty} \frac{\theta^n}{n!} h_n(X) \sum_{n=0}^{\infty} \frac{\eta^n}{n!} h_n(Y)\right] = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{\theta^n \eta^m}{n!m!} \mathbb{E}[h_n(X)h_m(Y)] = e^{\rho\eta\theta} = \sum_{n=0}^{\infty} \frac{\theta^n \eta^n}{n!} \rho^n.$$

Since the above must hold for all θ and η we conclude that

$$\mathbb{E}[h_n(X)h_m(Y)] = \begin{cases} n!\rho^n & \text{if} \quad n = m \\ 0 & \text{if} \quad n \neq m \end{cases}. \tag{5.7}$$

In particular, when $Y = X$, $\rho = 1$ and the above relationship becomes

$$\mathbb{E}[h_n(X)h_m(X)] = \int_{\mathbb{R}} h_n(x)h_m(x)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx = n!1(n = m). \tag{5.8}$$

The *normalized* Hermite polynomials are defined as $\tilde{h}_n(x) := \frac{1}{\sqrt{n!}}h_n(x)$ and form an orthonormal sequence.

Referring back to (5.1) we have

$$U_t(\theta) = \sum_{n=0}^{\infty} \frac{\theta^n}{n!}H_n(W_t, t) = \sum_{n=0}^{\infty} \frac{\theta^n}{n!}t^{n/2}h_n(W_t/\sqrt{t})$$

For instance,

$$H_4(W_t, t) = t^2\left(\frac{W_t^4}{t^2} - 6\frac{W_t^2}{t} + 3\right) = W_t^4 - 6W_t^2 t + 3t.$$

We have seen that, according to the Itô rule, for any twice continuously differentiable function f,

$$f(W_t, t) = f(0, 0) + \int_0^t \frac{\partial}{\partial x}f(W_s, s)dW_s + \int_0^t \frac{\partial}{\partial t}f(W_s, s)ds + \frac{1}{2}\int_0^t \frac{\partial^2}{\partial x^2}f(W_s, s)ds.$$

Applying the above formula to the function $f(x, t) = e^{\theta x - \frac{1}{2}\theta^2 t}$, where θ is a given real parameter, we obtain

$$\begin{aligned} e^{\theta W_t - \frac{1}{2}\theta^2 t} &= 1 + \int_0^t \theta e^{\theta W_s - \frac{1}{2}\theta^2 s}dW_s + \int_0^t (-\frac{1}{2}\theta^2)e^{\theta W_s - \frac{1}{2}\theta^2 s}ds + \frac{1}{2}\int_0^t \theta^2 e^{\theta W_s - \frac{1}{2}\theta^2 s}ds \\ &= 1 + \theta\int_0^t e^{\theta W_s - \frac{1}{2}\theta^2 s}dW_s. \end{aligned}$$

Chapter 6

Stochastic Differential Equations

6.1 Introduction

The evolution of many physical systems can be described by means of equations that relate the rate of change of a quantity to the quantity itself, and perhaps other variables. For instance, if $N(t)$ denotes the size of a given population at time t and if we assume that the rate of change (births minus deaths) of the population depends only on the size of the population and time, and that the functional dependence is described by a known function, $f(N(t), t)$, then the evolution of size of the population in question may be described by means of the differential equation

$$\frac{dN(t)}{dt} = f(N(t), t), \quad N(0) \text{ known initial value.} \tag{6.1}$$

Let us consider a concrete example, known as the Verhulst model of population growth. This model assumes that $f(N) := rN(C - N)$ where $r > 0$ is a measure of the innate ability of the population to grow, while $C > 0$ is the environments *carrying capacity*. When the size of a population exceeds this limit the population rate of growth becomes negative. Thus we have the differential equation

$$\frac{dN(t)}{dt} = rN(C - N), \quad N(0) = N_0 \tag{6.2}$$

(where N_0 is the known initial value of the population size). This differential equation is of the separable variables type, i.e. we may write

$$\frac{dN}{N(C - N)} = rdt \qquad \text{or} \qquad \frac{dN}{N} - \frac{dN}{C - N} = rCdt \tag{6.3}$$

(where we have expanded in partial fractions). Integrating, we obtain

$$\log N - \log(C - N) = K_0 + rCt$$

where K_0 is an integration constant. This gives

$$\frac{N}{C-N} = e^{K_0} e^{rCt}$$

and setting $K = e^{K_0} > 0$ we obtain

$$N(t) = \frac{KC}{K + e^{-rCt}}.$$

The value of K is determined by the requirement that $N(0) = \frac{KC}{K+1} = N_0$. Substituting, we obtain the expression for the evolution of the population size as a function of time.

$$N(t) = \frac{N_0 C}{N_0 + (C - N_0)e^{-Crt}}, \quad t \geq 0. \tag{6.4}$$

Let us now consider a first, stochastic version of (6.2). Suppose that the growth parameter r is no longer constant, but instead, a stochastic process $r(t,\omega)$ taking non-negative values with probability 1. Furthermore, assume that the initial population size is a random variable $N_0(\omega)$ with known distribution. (We may assume N_0 and $\{r(t); t \geq 0\}$ to be independent but this is not required by the arguments we will use.) Thus (6.2) becomes

$$\frac{dN(t)}{dt} = r(t,\omega)N(C-N), \quad N(0) = N_0(\omega) \tag{6.5}$$

(We include the usually omitted argument ω to emphasize the stochastic nature of this differential equation.) Provided that the process $\{r(t); t \geq 0\}$ has with probability 1 integrable sample paths we can integrate (6.3) and obtain

$$N(t,\omega) = \frac{N_0(\omega)C}{N_0(\omega) + (C - N_0(\omega))e^{-C\int_0^t r(s,\omega)ds}}, \quad t \geq 0. \tag{6.6}$$

The statistics of the process $\{N(t); t \geq 0\}$ can in principle be computed from those of $\{r(t)\}$, N_0 and (6.5). Such equations as (6.5) are usually called *random differential equations*. Their solution is obtained pathwise (for each ω) by applying the usual techniques of the theory of ordinary differential equations. No new concepts are necessary for their study.

A different type of differential equation subject to random disturbances is an equation of the form

$$dX_t = f(X_t)dt + dW_t \tag{6.7}$$

where $\{W_t; t \geq 0\}$ is standard brownian motion. The above is to be understood as the continuous version of the discrete recurrence equation

$$X_{t_{i+1}} = X_{t_i} + f(X_{t_i})h + \sqrt{h}\xi_i, \quad i = 0, 1, 2, \ldots,$$

where $h > 0$ is a (small) discrete step, $t_i = ih$, and $\{\xi_i\}$ an i.i.d. sequence of standard normal random variables. Equation (6.7) is to be understood in its equivalent integral form as

$$X_t = X_0 + \int_0^t f(X_s)ds + W_t.$$

More generally we are interested in considering Stochastic Differential Equations of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \qquad \text{with initial condition } X_0 \tag{6.8}$$

where b and σ are appropriately smooth functions $\mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$. and $\{W_t; t \geq 0\}$ is standard Brownian motion. Because of the fact that the Wiener process has sample paths that are everywhere non-differentiable with probability 1, the above equation is a shorthand for the integral equation

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{6.9}$$

Here the first integral on the right hand side is an ordinary Riemann (or Lebesgue) integral for each ω, whereas the second integral is to be interpreted as an Itô integral.

It can be shown (see for instance Øksendal, 2003) than if the functions b and σ satisfy the following two conditions then (6.9) has a unique solution. Suppose that $T > 0$ and that

1. *Lipschiz Condition.* There exists $K > 0$ such that $|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y|$ for all $x, y \in \mathbb{R}$ and $t \in [0, T]$.
2. *Linear Growth Condition.* There exists $G > 0$ such that $|b(t, x)| + |\sigma(t, x)| \leq G(1 + |x|)$ for all $x, y \in \mathbb{R}$ and $t \in [0, T]$.

Then (6.9) has a unique solution with sample paths continuous with probability 1 in the interval $[0, T]$.

The proof is based on the idea of Picard iteration, familiar from ordinary differential equations, whose idea is to construct a sequence of processes $\{X_t^n; t \in [0, T]\}$ as follows

$$\begin{aligned}
X_t^0 &= X_0, \quad t \in [0, T] \\
X_t^1 &= X_0 + \int_0^t b(s, X_0)ds + \int_0^t \sigma(s, X_0)dW_s, \quad t \in [0, T] \\
&\vdots \\
X_t^{n+1} &= X_0 + \int_0^t b(s, X_s^n)ds + \int_0^t \sigma(s, X_s^n)dW_s, \quad t \in [0, T], \quad n = 1, 2, \dots,
\end{aligned}$$

It can then be shown that the sequence of processes $\{X_t^n\}$ converges to a process $\{X_t\}$ which is the unique solution of (6.9).

In the same spirit we would like to study the differential equation

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)\frac{dW_t}{dt} \tag{6.10}$$

where X_0 is given and W_t is standard brownian motion. The problem we are faced with here is that W_t has paths that are nondifferentiable, hence the way this equation is written makes no sense. This can be circumvented if we pose the problem in its integral form

It is customary to write the above equation in differential form as follows:

Next we consider a couple of simple stochastic differential equations important in applications.

6.1.1 Geometric Brownian Motion

Consider the stochastic differential equation

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \qquad X_0 = x_0.$$

In order to solve it we may apply the Itô rule with $f(x) = \log x$ to obtain

$$f(X_t) - f(X_0) = \int_0^t \left(f'(X_s)\mu X_s + \frac{1}{2}f''(X_s)\sigma_s^2 X_s^2\right) ds + \int_0^t f'(X_s)\sigma X_s dW_s$$

$$\log(X_t/X_0) = \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) ds + \int_0^t \sigma dW_s = \left(\mu + \frac{1}{2}\sigma^2\right) t + \sigma W_t$$

whence we obtain

$$X_t = x_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t} . \tag{6.11}$$

It is easy to determine the distribution of X_t. Noting that $W_t \sim N(0, t)$ (i.e. normal with mean 0 and variance equal to t) and setting $a := \mu - \frac{1}{2}\sigma^2$ we see that

$$P(X_t \le x) = P(x_0 e^{at+\sigma W_t} \le x) = P(at+\sigma W_t \le \log(x/x_0)) = P\left(W_t \le \frac{1}{\sigma}(\log(x/x_0) - at)\right).$$

Denoting by $\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-\frac{1}{2}u^2} du$ the distribution function of the standard normal distribution and taking into account that $W_t \overset{d}{=} \sqrt{t}Z$, where Z is a standard normal random variable, from the above equation we have

$$P(X_t \le x) = P\left(Z \le \frac{1}{\sigma\sqrt{t}}(\log(x/x_0) - at)\right) = \Phi\left(\frac{1}{\sigma\sqrt{t}}(\log(x/x_0) - at)\right). \tag{6.12}$$

If we set $F_t(x) := P(X_t \le x)$ and $f_t(x) := \frac{d}{dx}F_t(x)$, and similarly $\varphi(x) := \frac{d}{dx}\Phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ we see from (6.12) that

$$\begin{aligned} f_t(x) &= \frac{d}{dx}\Phi\left(\frac{1}{\sigma\sqrt{t}}(\log(x/x_0)-at)\right) = \varphi\left(\frac{1}{\sigma\sqrt{t}}(\log(x/x_0)-at)\right)\frac{1}{x\sigma\sqrt{t}} \\ &= \frac{1}{x\sigma\sqrt{2\pi t}}e^{-\frac{1}{2\sigma^2 t}(\log(x/x_0)-at)^2} \end{aligned}$$

or

$$f_t(x) = \frac{1}{x\sigma\sqrt{2\pi t}}e^{-\frac{1}{2\sigma^2 t}(\log(x/x_0)-(\mu-\frac{1}{2}\sigma^2)t)^2}, \qquad x > 0. \tag{6.13}$$

The above is a lognormal density. The moments are easily determined from (6.11) by taking into account the moment generating function of the standard normal distribution, i.e. $Ee^{\theta Z} = e^{\frac{1}{2}\theta^2}$. Thus

$$EX_t = E\left[x_0 e^{\left(\mu-\frac{1}{2}\sigma^2\right)t+\sigma W_t}\right] = x_0 e^{\left(\mu-\frac{1}{2}\sigma^2\right)t}E\left[e^{\sigma W_t}\right] = x_0 e^{\left(\mu-\frac{1}{2}\sigma^2\right)t}e^{\frac{1}{2}\sigma^2 t} = x_0 e^{\mu t}$$

and

$$EX_t^2 = E\left[x_0^2 e^{2\left(\mu-\frac{1}{2}\sigma^2\right)t+2\sigma W_t}\right] = x_0^2 e^{\left(2\mu-\sigma^2\right)t}E\left[e^{2\sigma W_t}\right] = x_0^2 e^{\left(2\mu-\sigma^2\right)t}e^{2\sigma^2 t} = x_0^2 e^{(2\mu+\sigma^2)t}.$$

Finally,

$$\mathrm{Var}(X_t) = EX_t^2 - (EX_t)^2 = x_0^2 e^{(2\mu+\sigma^2)t} - x_0^2 e^{2\mu t} = x_0^2 e^{2\mu t}\left(e^{\sigma^2 t}-1\right).$$

As a special case consider $\sigma = 1$, $\mu = \frac{1}{2}$, $x_0 = 1$. In this case we have the stochastic differential equation

$$dX_t = \frac{1}{2}X_t dt + X_t dW_t.$$

which has the solution

$$X_t = e^{W_t}\ .$$

Then $P(X_t \le x) = P(W_t \le \log x)$ and

$$f_t(x) = \frac{1}{\sqrt{2\pi t}}x^{-1}e^{-\frac{1}{2t}(\log x)^2}\ , \qquad x \ge 0.$$

The mean and variance in this case are $EX_t = e^{\frac{1}{2}t}$ and $\mathsf{Var}(X_t) = e^{2t} - e^t$.

6.2 Ornstein–Uhlenbeck Equation

This is the equation

$$dX_t = \alpha X_t dt + \beta dW_t, \qquad X_0 = x_0. \tag{6.14}$$

or, equivalently, in integral form

$$X_t = x_0 + \int_0^t \alpha X_s ds + \beta W_t.$$

The solution is easy if we use Ito's formula for the function $f(x,t) = e^{-at}x$. Note that $\frac{\partial f}{\partial t} = -ae^{-at}x$, $\frac{\partial f}{\partial x} = e^{-at}$, and $\frac{\partial^2 f}{\partial x^2} = 0$. Ito's formula yields

$$\begin{aligned} f(X_t,t) - f(X_0,0) &= \int_0^t \left(\frac{\partial f}{\partial t}(X_s,s) + \frac{\partial f}{\partial x}(X_s,s) a X_s + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\beta^2 \right) ds + \int_0^t \frac{\partial f}{\partial x}(X_s,s)\beta dW_s \\ &= \int_0^t \left(-ae^{-as}X_s + e^{-at} a X_s \right) ds + \int_0^t e^{-as}\beta dW_s \end{aligned}$$

whence we obtain

$$e^{-at}X_t - x_0 = \int_0^t e^{-as}\beta dW_s$$

or equivalently

$$X_t = x_0 e^{\alpha t} + \int_0^t \beta e^{\alpha(t-s)} dW_s. \tag{6.15}$$

Note that the above is a gaussian process with mean

$$EX_t = x_0 e^{\alpha t}$$

$$\mathsf{Var}(X_t) = \int_0^t \beta^2 e^{2\alpha(t-s)} ds = \beta^2 e^{2\alpha t} \int_0^t e^{-2\alpha s} ds = \frac{\beta^2}{2\alpha} e^{2\alpha t}\left(1 - e^{-2\alpha t}\right) = \frac{\beta^2}{2\alpha}\left(e^{2\alpha t} - 1\right).$$

Also, assuming $s < t$,

$$\begin{aligned} \mathrm{Cov}\left(X_s, X_t\right) &= \mathrm{Cov}\left(\int_0^s \beta e^{\alpha(s-u)} dW_u, \int_0^t \beta e^{\alpha(t-u)} dW_u \right) \\ &= \mathrm{Cov}\left(\int_0^s \beta e^{\alpha(s-u)} dW_u, \int_0^s \beta e^{\alpha(t-u)} dW_u + \int_s^t \beta e^{\alpha(t-u)} dW_u \right) \\ &= \mathrm{Cov}\left(\int_0^s \beta e^{\alpha(s-u)} dW_u, \int_0^s \beta e^{\alpha(t-u)} dW_u \right) = \int_0^s \beta^2 e^{\alpha(s-u)} e^{\alpha(t-u)} du \\ &= \beta^2 e^{\alpha(s+t)} \int_0^s e^{-2\alpha u} du = \frac{\beta^2}{2\alpha} e^{\alpha(s+t)}\left(1 - e^{-2\alpha s}\right) = \frac{\beta^2}{2\alpha}\left(e^{\alpha(s+t)} - e^{\alpha(t-s)}\right). \end{aligned}$$

In general setting $s \wedge t = \min(s,t)$, $s \vee t = \max(s,t)$,

$$\mathrm{Cov}\left(X_s, X_t\right) = \frac{\beta^2}{2\alpha} e^{a(s\vee t)}\left(e^{\alpha(s\wedge t)} - e^{-\alpha(s\wedge t)}\right).$$

6.3 The Brownian Bridge Process

Suppose that W_t, $t \geq 0$, is standard brownian motion. The brownian bridge process X_t, $t \in [0, 1]$, is defined (at least informally for the time being) as the process B_t, *conditional on the event* $W_1 = 0$. Based on this informal idea we can easily compute the joint distribution of the process X_t as follows. Thus, for $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$ we have that

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \stackrel{\mathrm{d}}{=} (W_{t_1}, W_{t_2}, \ldots, W_{t_n}) | \, W_1 = 0.$$

The right hand side of the above equation is the conditional distribution of a Gaussian vector given one of its components, hence it is again Gaussian.

Two useful characterizations of the brownian bridge process are the following.

$$X_t = W_t - tW_1, \qquad 0 \leq t \leq 1 \tag{6.16}$$

and

$$dX_t = -\frac{1}{1-t} X_t dt + dW_t, \qquad 0 \leq t < 1, \qquad X_0 = 0. \tag{6.17}$$

It is easy to check that the process described in (6.16) is a Gaussian process which has the right mean and covariance function and therefore that it is indeed the standard brownian bridge. Regarding the SDE of (6.17) note that it can be solved through the use of an integrating factor: Indeed, $\int \frac{dt}{1-t} = -\log(1-t)$ and $e^{\int \frac{dt}{1-t}} = \frac{1}{1-t}$. Thus, multiplying the equation (6.17) by $\frac{1}{1-t}$ we obtain

$$\frac{1}{1-t} dX_t + \frac{1}{(1-t)^2} X_t dt = \frac{dW_t}{1-t}$$

or

$$d\left(\frac{1}{1-t} X_t\right) = \frac{dW_t}{1-t}$$

where we have used Ito's rule, hence

$$X_t = (1-t) \int_0^t \frac{dW_s}{1-s}.$$

In the above derivation we also took into account the initial condition $X_0 = 0$. If the process starts at time $s \in (0, 1)$ at the point X_s, then the solution would have been

$$X_t = (1-t) \left(\frac{X_s}{1-s} + \int_s^t \frac{dW_u}{1-u} \right) \tag{6.18}$$

To see that this is again the standard brownian bridge it is enough to note that it is a Gaussian process with zero mean and covariance function given, when $s < t$, by (6.18) we have

$$EX_s X_t = E\left[X_s^2 \frac{1-t}{1-s} \right] + E\left[X_s (1-t) \int_s^t \frac{dW_u}{1-u} \right].$$

The second integral on the right hand side of the above equation is zero by the martingale property of stochastic integrals. Thus, in view of the fact that

$$\begin{aligned} E\left[X_s^2\right] &= (1-s)^2 E\left(\int_0^s \frac{dW_u}{1-u}\right)^2 = (1-s)^2 \int_0^s \frac{du}{(1-u)^2} \\ &= (1-s)^2\left(\frac{1}{1-s}-1\right) = s(1-s), \end{aligned}$$

gives

$$EX_sX_t = s(1-t)$$

which is the correct expression for the covariance. (Note that in the above derivation we have also made use of the Ito isometry.)

Bibliography

[1] Bernt Øksendal. *Stochastic Differential Equations. An Introduction with Applications.* 6th Edition, Springer Verlag, 2003.

[2] Fima Klebaner. Introduction to Stochastic Calculus with Applications. 2nd Edition, Imperial College Press, 2004.