MARTINGALES, DIFFUSIONS AND FINANCIAL MATHEMATICS

A.W. van der Vaart

Preliminary Notes with (too many) mistakes.

CONTENTS

LITERATURE

There are very many books on the topics of the course. The list below is a small selection.

Discrete martingales are discussed in most advanced introductions to general probability theory. The book by David Williams is particularly close to our presentation.

For an introduction to stochastic integration we prefer the book by Chung and Williams (Ruth Williams this time). It has introductions to most of the important topics and is very well written. The two volumes by Rogers and Williams (David again) are a classic, but they are not easy and perhaps even a bit messy at times. The book by Karatzas and Shreve is more accessible, and good if you like the details. The book by Revuz and Yor has a wider scope on stochastic processes. Unlike Chung and Williams or Rogers and Williams the latter two books are restricted to martingales with continuous sample paths, which obscures some interesting aspects, but also makes some things easier.

The theory of stochastic integration and much of the theory of abstract stochastic processes was originally developed by the "french school", with Meyer as the most famous proponent. Few people can appreciate the fairly abstract and detailed original books (Look for Dellacherie and Meyer, volumes 1, 2, 3, 4). The book by Elliott is in this tradition, but somewhat more readable. The first chapter of Jacod and Shiryaev is an excellent summary and reference, but is not meant for introductory reading.

The book by Øksendal is a popular introduction. It does not belong to my personal favourites.

The book by Stroock and Varadhan is a classic on stochastic differential equations and particularly important as a source on the "martingale problem".

There are also many books on financial calculus. Some of them are written from the perspective of differential equations. Then Brownian motion is reduced to a process such that $(dB_t)^2 = dt$. The books mentioned below are of course written from a probabilistic point of view. Baxter and Rennie have written their book for a wide audience. It is interesting how they formulate "theorems" very imprecisely, but never wrong. It is good to read to get a feel for the subject. Karatzas and Shreve, and Kopp and Elliott have written rigorous mathematical books that give you less feel, but more theorems.

[1] Baxter, M. and Rennie, A., (1996). *Financial calculus*. Cambridge University Press, Cambridge.

[2] Chung, K.L. and Williams, R.J., (1990). *Introduction to stochastic integration, second edition*. Birkhäuser, London.

[3] Elliott, R.J., (1982). *Stochastic calculus and applications*. Springer Verlag, New York.

[4] Jacod, J. and Shiryaev, A.N., (1987). *Limit theorems for stochastic processes*. Springer-Verlag, Berlin.

[5] Kopp, P.E. and Elliott, R.J., (1999). *Mathematics and financial markets*. Springer-Verlag, New York.

[6] Karatzas, I. and Shreve, S.E., (1988). *Brownian motion and stochastic calculus*. Springer-Verlag, Berlin.

[7] Karatzas, I. and Shreve, S.E., (1998). *Methods of mathematical finance*. Springer-Verlag, Berlin.

[8] Øksendal, B., (1998). *Stochastic differential equations, 5th edition*. Springer, New York.

[9] Revuz, D. and Yor, M., (1994). *Continuous martingales and Brownian motion*. Springer, New York.

[10] Rogers, L.C.G. and Williams, D., (2000). *Diffusions, Markov Processes and Martingales, volumes 1 and 2*. Cambridge University Press, Cambridge.

[11] Stroock, D.W. and Varadhan, S.R.S., (1979). *Multidimensional Diffusion Processes*. Springer-Verlag, Berlin.

[12] van der Vaart, A.W. and Wellner, J.A., (1996). *Weak Convergence and Empirical Processes*. Springer Verlag, New York.

[13] Williams, D., (1991). *Probability with Martingales*. Cambridge University Press, Cambridge.

EXAM

The written exam will consist of problems as in these notes, questions to work out examples as in the notes or variations thereof, and will require to give precise definitions and statements of theorems plus a numbers of proofs.

The requirements for the oral exam are the same. For a very high mark it is, of course, necessary to know everything.

Very important is to be able to give a good overview of the main points of the course and their connections.

Starred sections or lemmas in the lecture notes can be skipped completely. Starred exercises may be harder than other exercises.

Proofs to learn by heart:

2.13, 2.43, 2.44 for $p = 2$.

4.21, 4.22, 4.26, 4.28.

5.22, 5.25(i)-(iii), 5.43, 5.46 case that M is continuous, 5.53, 5.58, 5.82, 5.93.

6.1, 6.9(ii),

7.8 case that $\mathrm{E}\xi^2 < \infty$ and (7.6) holds for every x, y, 7.15.

1
Measure Theory

In this chapter we review or introduce a number of results from measure theory that are especially important in the following.

1.1 Conditional Expectation

Let X be an integrable random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In other words $X: \Omega \to \mathbb{R}$ is a measurable map (relative to $\mathcal{F}$ and the Borel sets on $\mathbb{R}$) with $\mathrm{E}|X| < \infty$.

1.1 Definition. *Given a sub σ-field $\mathcal{F}_0 \subset \mathcal{F}$ the conditional expectation of X relative to $\mathcal{F}_0$ is a $\mathcal{F}_0$-measurable map $X': \Omega \to \mathbb{R}$ such that*

$$\mathrm{E}X1_F = \mathrm{E}X'1_F, \qquad \text{for every } F \in \mathcal{F}_0, \tag{1.2}$$

The random variable X' is denoted by $\mathrm{E}(X|\,\mathcal{F}_0)$.

It is clear from this definition that any other $\mathcal{F}_0$-measurable map $X'': \Omega \to \mathbb{R}$ such that $X' = X''$ almost surely is also a conditional expectation. In the following theorem it is shown that conditional expectations exist and are unique, apart from this indeterminacy on null sets.

1.3 Theorem. *Let X be a random variable with $\mathrm{E}|X| < \infty$ and $\mathcal{F}_0 \subset \mathcal{F}$ a σ-field. Then there exists an $\mathcal{F}_0$-measurable map $X': \Omega \to \mathbb{R}$ such that (1.2) holds. Furthermore, any two such maps X' agree almost surely.*

Proof. If $X \geq 0$, then on the σ-field $\mathcal{F}_0$ we can define a measure $\mu(F) = \int_F X\,d\mathbb{P}$. Clearly this measure is finite and absolutely continuous relative to the restriction of $\mathbb{P}$ to $\mathcal{F}_0$. By the Radon-Nikodym theorem there exists an $\mathcal{F}_0$-measurable function X', unique up to null sets, such that $\mu(F) = \int_F X'\,d\mathbb{P}$ for every $F \in \mathcal{F}_0$. This is the desired map X'. For a general X we apply this argument separately to X^+ and X^- and take differences.

Suppose that $\mathrm{E}(X' - X'')1_F = 0$ for every F in a σ-field for which $X' - X''$ is measurable. Then we may choose $F = \{X' > X\}$ to see that the probability of this set is zero, because the integral of a strictly positive variable over a set of positive measure must be positive. Similarly we see that the set $F = \{X' < X''\}$ must be a null set. Thus $X' = X''$ almost surely. ■

The definition of a conditional expectation is not terribly insightful, even though the name suggests an easy interpretation as an expected value. A number of examples will make the definition clearer.

A measurable map $Y{:}\,\Omega \to (\mathbb{D}, \mathcal{D})$ generates a σ-field $\sigma(Y)$. We use the notation $\mathrm{E}(X|\,Y)$ as an abbreviation of $\mathrm{E}(X|\,\sigma(Y))$.

1.4 Example (Ordinary expectation). The expectation $\mathrm{E}X$ of a random variable X is a number, and as such can of course be viewed as a degenerate random variable. Actually, it is also the conditional expectation relative to the trivial σ-field $\mathcal{F}_0 = \{\emptyset, \Omega\}$. More generally, we have that $\mathrm{E}(X|\,\mathcal{F}_0) = \mathrm{E}X$ if X and $\mathcal{F}_0$ are independent. In this case $\mathcal{F}_0$ gives "no information" about X and hence the expectation given $\mathcal{F}_0$ is the "unconditional" expectation.

To see this note that $\mathrm{E}(\mathrm{E}X)1_F = \mathrm{E}X\mathrm{E}1_F = \mathrm{E}X1_F$ for every F such that X and F are independent. □

1.5 Example. At the other extreme we have that $\mathrm{E}(X|\,\mathcal{F}_0) = X$ if X itself is $\mathcal{F}_0$-measurable. This is immediate from the definition. "Given $\mathcal{F}_0$ we then know X exactly." □

1.6 Example. Let $(X, Y){:}\,\Omega \to \mathbb{R} \times \mathbb{R}^k$ be measurable and possess a density $f(x, y)$ relative to a σ-finite product measure $\mu \times \nu$ on $\mathbb{R} \times \mathbb{R}^k$ (for instance, the Lebesgue measure on $\mathbb{R}^{k+1}$). Then it is customary to define a *conditional density* of X given $Y = y$ by

$$f(x|\,y) = \frac{f(x, y)}{\int f(x, y)\,d\mu(x)}.$$

This is well defined for every y for which the denominator is positive, i.e. for all y in a set of measure one under the distribution of Y.

We now have that the conditional expection is given by the "usual formula"

$$\mathrm{E}(X|\,Y) = \int x f(x|\,Y)\,d\mu(x),$$

where we may define the right hand zero as zero if the expression is not well defined.

That this formula is the conditional expectation according to the abstract definition follows by a number of applications of Fubini's theorem. Note that, to begin with, it is a part of the statement of Fubini's theorem that the function on the right is a measurable function of Y. □

1.7 Example (Partitioned Ω). If $\mathcal{F}_0 = \sigma(F_1, \ldots, F_k)$ for a partition $\Omega = \cup_{i=1}^k F_i$, then

$$\mathrm{E}(X|\,\mathcal{F}_0) = \sum_{i=1}^k \mathrm{E}(X|\,F_i)1_{F_i},$$

where $\mathrm{E}(X|\,F_i)$ is defined as $\mathrm{E}X1_{F_i}/\mathbb{P}(F_i)$ if $\mathbb{P}(F_i) > 0$ and arbitrary otherwise. Thus the conditional expectation is constant on every of the partitioning sets F_i (as it needs to be to be $\mathcal{F}_0$-measurable) and the constant values are equal to the average values of X over these sets.

The validity of (1.2) is easy to verify for $F = F_j$ and every j. And then also for every $F \in \mathcal{F}_0$ by taking sums, since every $F \in \mathcal{F}_0$ is a union of a number of F_j's.

This example extends to σ-fields generated by a countable partition of Ω. In particular, $\mathrm{E}(X|\,Y)$ is exactly what we would think it should be if Y is a discrete random variable. □

A different perspective on an expectation is to view it as a best prediction if "best" is defined through minimizing a second moment. For instance, the ordinary expectation $\mathrm{E}X$ minimizes $\mu \mapsto \mathrm{E}(X-\mu)^2$ over $\mu \in \mathbb{R}$. A conditional expectation is a best prediction by an $\mathcal{F}_0$-measurable variable.

1.8 Lemma (L_2-projection). *If* $\mathrm{E}X^2 < \infty$, *then* $\mathrm{E}(X|\,\mathcal{F}_0)$ *minimizes* $\mathrm{E}(X-Y)^2$ *over all* $\mathcal{F}_0$*-measurable random variables* Y.

Proof. We first show that $X' = \mathrm{E}(X|\,\mathcal{F}_0)$ satisfies $\mathrm{E}X'Z = \mathrm{E}XZ$ for every $\mathcal{F}_0$-measurable Z with $\mathrm{E}Z^2 < \infty$.

By linearity of the conditional expectation we have that $\mathrm{E}X'Z = \mathrm{E}XZ$ for every $\mathcal{F}_0$-simple variable Z. If Z is $\mathcal{F}_0$-measurable with $\mathrm{E}Z^2 < \infty$, then there exists a sequence Z_n of $\mathcal{F}_0$-simple variables with $\mathrm{E}(Z_n - Z)^2 \to 0$. Then $\mathrm{E}X'Z_n \to \mathrm{E}X'Z$ and similarly with X instead of X' and hence $\mathrm{E}X'Z = \mathrm{E}XZ$.

Now we decompose, for arbitrary square-integrable Y,

$$\mathrm{E}(X-Y)^2 = \mathrm{E}(X-X')^2 + 2\mathrm{E}(X-X')(X'-Y) + \mathrm{E}(X'-Y)^2.$$

The middle term vanishes, because $Z = X' - Y$ is $\mathcal{F}_0$-measurable and square-integrable. The third term on the right is clearly minimal for $X' = Y$. ■

1.9 Lemma (Properties).
(i) $\mathrm{E}\mathrm{E}(X|\,\mathcal{F}_0) = \mathrm{E}X$.
(ii) *If Z is $\mathcal{F}_0$-measurable, then* $\mathrm{E}(ZX|\,\mathcal{F}_0) = Z\mathrm{E}(X|\,\mathcal{F}_0)$ a.s.. *(Here require that $X \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ and $Z \in L_q(\Omega, \mathcal{F}, \mathbb{P})$ for $1 \le p \le \infty$ and $p^{-1} + q^{-1} = 1$.)*
(iii) *(linearity)* $\mathrm{E}(\alpha X + \beta Y|\,\mathcal{F}_0) = \alpha\mathrm{E}(X|\,\mathcal{F}_0) + \beta\mathrm{E}(Y|\,\mathcal{F}_0)$ a.s..
(iv) *(positivity) If* $X \ge 0$ a.s., *then* $\mathrm{E}(X|\,\mathcal{F}_0) \ge 0$ a.s..
(v) *(towering property) If* $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}$, *then* $\mathrm{E}\big(\mathrm{E}(X|\,\mathcal{F}_1)|\,\mathcal{F}_0\big) = \mathrm{E}(X|\,\mathcal{F}_0)$ a.s..
(vi) *(Jensen) If $\phi: \mathbb{R} \to \mathbb{R}$ is convex, then* $\mathrm{E}(\phi(X)|\,\mathcal{F}_0) \ge \phi\big(\mathrm{E}(X|\,\mathcal{F}_0)\big)$ a.s.. *(Here require that $\phi(X)$ is integrable.)*
(vii) $\|\mathrm{E}(X|\,\mathcal{F}_0)\|_p \le \|X\|_p \quad (p \ge 1)$.

* **1.10 Lemma (Convergence theorems).**
(i) *If* $0 \le X_n \uparrow X$ a.s., *then* $0 \le \mathrm{E}(X_n|\,\mathcal{F}_0) \uparrow \mathrm{E}(X|\,\mathcal{F}_0)$ a.s..
(ii) *If* $X_n \ge 0$ a.s. *for every n, then* $\mathrm{E}(\liminf X_n|\,\mathcal{F}_0) \le \liminf \mathrm{E}(X_n|\,\mathcal{F}_0)$ a.s..
(iii) *If $|X_n| \le Y$ for every n and and integrable variable Y, and $X_n \overset{\text{as}}{\to} X$, then* $\mathrm{E}(X_n|\,\mathcal{F}_0) \overset{\text{as}}{\to} \mathrm{E}(X|\,\mathcal{F}_0)$ a.s..

The conditional expectation $\mathrm{E}(X|\,Y)$ given a random vector Y is by definition a $\sigma(Y)$-measurable function. For most Y, this means that it is a measurable function $g(Y)$ of Y. (See the following lemma.) The value $g(y)$ is often denoted by $\mathrm{E}(X|\,Y = y)$.

Warning. Unless $\mathbb{P}(Y = y) > 0$ it is not right to give a meaning to $\mathrm{E}(X|\,Y = y)$ for a fixed, single y, even though the interpretation as an expectation given "that we know that $Y = y$" often makes this tempting. We may only think of a conditional expectation as a function $y \mapsto \mathrm{E}(X|\,Y = y)$ and this is only determined up to null sets.

1.11 Lemma. *Let $\{Y_\alpha: \alpha \in A\}$ be random variables on Ω and let X be a $\sigma(Y_\alpha: \alpha \in A)$-measurable random variable.*
(i) *If $A = \{1, 2, \ldots, k\}$, then there exists a measurable map $g: \mathbb{R}^k \to \mathbb{R}$ such that $X = g(Y_1, \ldots, Y_k)$.*
(ii) *If $|A| = \infty$, then there exists a countable subset $\{\alpha_n\}_{n=1}^\infty \subset A$ and a measurable map $g: \mathbb{R}^\infty \to \mathbb{R}$ such that $X = g(Y_{\alpha_1}, Y_{\alpha_2}, \ldots)$.*

1.2 Uniform Integrability

In many courses on measure theory the dominated convergence theorem is one of the best results. Actually, domination is not the right concept, uniform integrability is.

1.12 Definition. *A collection* $\{X_\alpha: \alpha \in A\}$ *of random variables is uniformly integrable if*

$$\lim_{M\to\infty} \sup_{\alpha\in A} \mathrm{E}|X_\alpha| 1_{|X_\alpha|>M} = 0.$$

1.13 Example. A finite collection of integrable random variables is uniformly integrable.

This follows because $\mathrm{E}|X| 1_{|X|>M} \to 0$ as $M \to \infty$ for any integrable variable X, by the dominated convergence theorem. □

1.14 Example. A dominated collection of random variables is uniformly integrable: if $|X_\alpha| \leq Y$ and $\mathrm{E}Y < \infty$, then $\{X_\alpha: \alpha \in A\}$ is uniformly integrable.

To see this note that $|X_\alpha| 1_{|X_\alpha|>M} \leq Y 1_{Y>M}$. □

1.15 Example. If the collection of random variables $\{X_\alpha: \alpha \in A\}$ is bounded in L_2, then it is is uniformly integrable.

This follows from the inequality $\mathrm{E}|X| 1_{|X|>M} \leq M^{-1}\mathrm{E}X^2$, which is valid for any random variable X.

Similarly, it suffices for uniform integrability that $\sup_\alpha \mathrm{E}|X_\alpha|^p < \infty$ for some $p > 1$. □

1.16 EXERCISE. Show that a uniformly integrable collection of random variables is bounded in $L_1(\Omega, \mathcal{F}, \mathbb{P})$.

1.17 EXERCISE. Show that any converging sequence X_n in $L_1(\Omega, \mathcal{F}, \mathbb{P})$ is uniformly integrable.

1.18 Theorem. *Suppose that* $\{X_n: n \in \mathbb{N}\} \subset L_1(\Omega, \mathcal{F}, \mathbb{P})$. *Then* $\mathrm{E}|X_n - X| \to 0$ *for some* $X \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ *if and only if* $X_n \xrightarrow{\mathrm{P}} X$ *and* $\{X_n: n \in \mathbb{N}\}$ *is uniformly integrable.*

Proof. We only give the proof of "if". (The main part of the proof in the other direction is the preceding exercise.)

If $X_n \xrightarrow{\mathrm{P}} X$, then there is a subsequence X_{n_j} that converges almost surely to X. By Fatou's lemma $\mathrm{E}|X| \leq \liminf \mathrm{E}|X_{n_j}|$. If X_n is uniformly integrable, then the right side is finite and hence $X \in L_1(\Omega, \mathcal{F}, \mathbb{P})$.

For any random variables X and Y and positive numbers M and N,

$$\begin{aligned} \mathrm{E}|X| 1_{|Y|>M} &\leq \mathrm{E}|X| 1_{|X|>N} 1_{|Y|>M} + N\mathbb{P}\big(|Y| > M\big) \\ &\leq \mathrm{E}|X| 1_{|X|>N} + \frac{N}{M}\mathrm{E}|Y| 1_{|Y|>M}. \end{aligned} \tag{1.19}$$

Applying this with $M = N$ and (X, Y) equal to the four pairs that can be formed of X_n and X we find, for any $M > 0$,

$$\mathrm{E}|X_n - X|(1_{|X_n|>M} + 1_{|X|>M}) \leq 2\mathrm{E}|X_n| 1_{|X_n|>M} + 2\mathrm{E}|X| 1_{|X|>M}.$$

We can make this arbitrarily small by making M sufficiently large. Next, for any $\varepsilon > 0$,

$$\mathrm{E}|X_n - X|1_{|X_n|\le M, |X|\le M} \le \varepsilon + 2M\mathbb{P}\bigl(|X_n - X| > \varepsilon\bigr).$$

As $n \to \infty$ the second term on the right converges to zero for every fixed $\varepsilon > 0$ and M. ∎

1.20 EXERCISE. If $\{|X_n|^p : n \in \mathbb{N}\}$ is uniformly integrable ($p \ge 1$) and $X_n \xrightarrow{\mathrm{P}} X$, then $\mathrm{E}|X_n - X|^p \to 0$. Show this.

1.21 Lemma. *If $X \in L_1(\Omega, \mathcal{F}, \mathbb{P})$, then the collection of all conditional expectations $\mathrm{E}(X|\,\mathcal{F}_0)$ with $\mathcal{F}_0$ ranging over all sub σ-fields of $\mathcal{F}$ is uniformly integrable.*

Proof. By Jensen's inequality $|\mathrm{E}(X|\,\mathcal{F}_0)| \le \mathrm{E}(|X|\,|\,\mathcal{F}_0)$ almost surely. It therefore suffices to show that the conditional expectations $\mathrm{E}(|X|\,|\,\mathcal{F}_0)$ are uniformly integrable. For simplicity of notation suppose that $X \ge 0$.

With $X' = \mathrm{E}(X|\,\mathcal{F}_0)$ and arguing as in (1.19) we see that

$$\mathrm{E}X'1_{X'>M} = \mathrm{E}X1_{X'>M} \le \mathrm{E}X1_{X>N} + \frac{N}{M}\mathrm{E}X'.$$

We can make the right side arbitrarily small by first choosing N and next M sufficiently large. ∎

We conclude with a lemma that is sometimes useful.

1.22 Lemma. *Suppose that X_n and X are random variables such that $X_n \xrightarrow{\mathrm{P}} X$ and $\limsup \mathrm{E}|X_n|^p \le \mathrm{E}|X|^p < \infty$ for some $p \ge 1$. Then $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable and $\mathrm{E}|X_n - X|^p \to 0$.*

1.3 Monotone Class Theorem

Many arguments in measure theory are carried out first for simple types of functions and then extended to general functions by taking limits. A monotone class theorem is meant to codify this procedure. This purpose of standardizing proofs is only partly successful, as there are many monotone class theorems in the literature, each tailored to a particular purpose. The following theorem will be of use to us.

We say that a class $\mathcal{H}$ of functions $h : \Omega \to \mathbb{R}$ is *closed under monotone limits* if for each sequence $\{h_n\} \subset \mathcal{H}$ such that $0 \le h_n \uparrow h$ for some function h, the limit h is contained in $\mathcal{H}$. We say that it is closed under bounded monotone limits if this is true for every such sequence h_n with a (uniformly) bounded limit. A class of sets is *intersection-stable* if it contains the intersection of every pair of its elements (i.e. is a π-system).

1.23 Theorem. *Let $\mathcal{H}$ be a vector space of functions $h: \Omega \to \mathbb{R}$ on a measurable space $(\Omega, \mathcal{F})$ that contains the constant functions and the indicator of every set in a collection $\mathcal{F}_0 \subset \mathcal{F}$, and is closed under (bounded) monotone limits. If $\mathcal{F}_0$ is intersection-stable, then $\mathcal{H}$ contains all (bounded) $\sigma(\mathcal{F}_0)$-measurable functions.*

Proof. See e.g. Williams, A3.1 on p205. ∎

2
Discrete Time Martingales

A *stochastic process* X in discrete time is a sequence $X_0, X_1, X_2, \dots$ of random variables defined on some common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The index n of X_n is referred to as "time" and a map $n \mapsto X_n(\omega)$, for a fixed $\omega \in \Omega$, is a *sample path.* (Later we replace n by a continuous parameter $t \in [0, \infty)$ and use the same terminology.) Usually the discrete time set is $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$. Sometimes we delete 0 or add ∞ to get $\mathbb{N}$ or $\bar{\mathbb{Z}}_+ = \mathbb{N} \cup \{0, \infty\}$, and add or delete a corresponding random variable X_∞ or X_0 to form the stochastic process.

2.1 Martingales

A *filtration* $\{\mathcal{F}_n\}$ (in discrete time) on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a nested sequence of σ-fields

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

The σ-field $\mathcal{F}_n$ is interpreted as the events F of which it is known at "time" n whether F has occurred or not. A stochastic process X is said to be *adapted* if X_n is $\mathcal{F}_n$-measurable for every $n \geq 0$. The quadruple $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ is called a "filtered probability space" or "stochastic basis".

A typical example of a filtration is the *natural filtration* generated by a stochastic process X, defined as

$$\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n).$$

Then $F \in \mathcal{F}_n$ if and only if $F = \{(X_0, \dots, X_n) \in B\}$ for some Borel set B. Once $X_0, \dots, X_n$ are realized we know whether F has occurred or not. The natural filtration is the smallest filtration to which X is adapted.

2.1 Definition. *An adapted, integrable stochastic process X on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ is a*
(i) martingale if $\mathrm{E}(X_n|\,\mathcal{F}_m) = X_m$ a.s. *for all $m \leq n$.*
(ii) submartingale if $\mathrm{E}(X_n|\,\mathcal{F}_m) \geq X_m$ a.s. *for all $m \leq n$.*
(ii) supermartingale if $\mathrm{E}(X_n|\,\mathcal{F}_m) \leq X_m$ a.s. *for all $m \leq n$.*

A different way of writing the martingale property is

$$\mathrm{E}(X_n - X_m|\,\mathcal{F}_m) = 0, \qquad m \leq n.$$

Thus given all information at time m the expected increment $X_n - X_m$ in the future time interval $(m, n]$ is zero, for every initial time m. This shows that a martingale X_n can be interpreted as the total gain up to time n in a fair game: at every time m we expect to make a zero gain in the future (but may have gained in the past and we expect to keep this). In particular, the expectation $\mathrm{E}X_n$ of a martingale is constant in n.

Submartingales and supermartingales can be interpreted similarly as total gains in favourable and unfavourable games. If you are not able to remember which inequalities correspond to “sub” and “super”, that is probably normal. It helps a bit to try and remember that a submartingale is increasing in mean: $\mathrm{E}X_m \leq \mathrm{E}X_n$ if $m \leq n$.

2.2 EXERCISE. If $\mathrm{E}(X_{n+1}|\,\mathcal{F}_n) = X_n$ for every $n \geq 0$, then automatically $\mathrm{E}(X_n|\,\mathcal{F}_m) = X_m$ for every $m \leq n$ and hence X is a martingale. Similarly for sub/super. Show this.

2.3 Example. Let $Y_1, Y_2, \ldots$ be a sequence of independent random variables with mean zero. Then the sequence of partial sums $X_n = Y_1 + \cdots + Y_n$ is a martingale relative to the filtration $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. Set $X_0 = 0$.

This follows upon noting that for $m \leq n$ the increment $X_n - X_m = \sum_{m<i\leq n} Y_i$ is independent of $\mathcal{F}_m$ and hence $\mathrm{E}(X_n - X_m|\,\mathcal{F}_m) = \mathrm{E}(X_n - X_m) = 0$. □

2.4 EXERCISE. In the preceding example show that $\sigma(Y_1, \ldots, Y_n) = \sigma(X_1, \ldots, X_n)$.

2.5 EXERCISE. If $\{N(t)\colon t \geq 0\}$ is a standard Poisson process and $0 \leq t_0 < t_1 < \cdots$ is a fixed sequence of numbers, then $X_n = N(t_n) - t_n$ is a martingale relative to the filtration $\mathcal{F}_n = \sigma(N(t)\colon t \leq t_n)$. Show this, using the fact that the Poisson process has independent increments.

2.6 Example. Let ξ be a fixed, integrable random variable and $\mathcal{F}_n$ an arbitrary filtration. Then $X_n = \mathrm{E}(\xi|\,\mathcal{F}_n)$ is a martingale.

This is an immediate consequence of the towering property of conditional expectations, which gives that $\mathrm{E}(X_n|\,\mathcal{F}_m) = \mathrm{E}\big(\mathrm{E}(\xi|\,\mathcal{F}_n)|\,\mathcal{F}_m\big) = \mathrm{E}(\xi|\,\mathcal{F}_m)$ for every $m \leq n$.

By Theorem 1.18 this martingale X is uniformly integrable. Later we shall see that any uniformly integrable martingale takes this form. Moreover, we can choose ξ such that $X_n \stackrel{\text{as}}{\to} \xi$ as $n \to \infty$. □

It is part of the definition of a martingale X that every of the random variables X_n is integrable. If $\sup_n \mathrm{E}|X_n| < \infty$, then we call the martingale L_1-*bounded*. If $\mathrm{E}|X_n|^p < \infty$ for ally n and some p, then we call X an L_p-*martingale* and if $\sup_n \mathrm{E}|X_n|^p < \infty$, then we call X L_p-*bounded*.

Warning. Some authors use the phrase "L_p-martingale" for a martingale that is bounded in $L_p(\Omega, \mathcal{F}, \mathbb{P})$. To avoid this confusion, it is perhaps better to use the more complete phrases "martingale in L_p" and "martingale that is bounded in L_p".

2.7 Lemma. *If $\phi: \mathbb{R} \to \mathbb{R}$ is convex and X a martingale, then $\{\phi(X_n)\}$ is a submartingale relative to the same filtration, provided that $\phi(X_n)$ is integrable for every n.*

Proof. Because a convex function is automatically measurable, the variable $\phi(X_n)$ is adapted for every n. By Jensen's inequality $\mathrm{E}\big(\phi(X_n)|\,\mathcal{F}_m\big) \geq \phi\big(\mathrm{E}(X_n|\,\mathcal{F}_m)\big)$ almost surely. The right side is $\phi(X_m)$ almost surely if $m \leq n$, by the martingale property. ■

2.8 EXERCISE. If $\phi: \mathbb{R} \to \mathbb{R}$ is convex and nondecreasing and X is a submartingale, then $\{\phi(X_n)\}$ is a submartingale relative to the same filtration, provided that $\phi(X_n)$ is integrable for every n. Show this.

2.2 Stopped Martingales

If X_n is interpreted as the total gain at time n, then a natural question is if we can maximize profit by quitting the game at a suitable time. If X_n is a martingale with $\mathrm{E}X_0 = 0$ and we quit at a fixed time T, then our expected profit is $\mathrm{E}X_T = \mathrm{E}X_0 = 0$ and hence quitting the game does not help. However, this does not exclude the possibility that stopping at a random time might help. This is the gambler's dream.

If we could let our choice to stop depend on the future, then it is easy to win. For instance, if we were allowed to stop just before we incurred a big loss. This we prohibit by considering only "stopping times" as in the following definition.

2.9 Definition. *A random variable $T: \Omega \to \bar{\mathbb{Z}}_+$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ is a stopping time if $\{T \leq n\} \in \mathcal{F}_n$ for every $n \geq 0$.*

Warning. A stopping time is permitted to take the value ∞.

2.10 EXERCISE. Let X be an adapted stochastic process and let $B \subset \mathbb{R}$ be measurable. Show that $T = \inf\{n: X_n \in B\}$ defines a stopping time. (Set $\inf \emptyset = \infty$.)

2.11 EXERCISE. Show that T is a stopping time if and only if $\{T = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N} \cup \{0\}$.

The restriction to stopping times is natural. If we are to stop playing at time T, then for every time $n = 0, 1, 2 \ldots$ we must know if $T = n$ at time n. If the filtration is generated by the process X, then the event $\{T = n\}$ must, for every n, depend on the history $X_0, \ldots, X_n$ of the process up to time n only, if T is a stopping time. So we are allowed to base our decision to stop on the past history of gains or losses, but not on future times.

The question now is if we can find a stopping time T such that $\mathrm{E} X_T > 0$. We shall see that this is usually not the case. Here the random variable X_T is defined as

$$(X_T)(\omega) = X_{T(\omega)}(\omega). \tag{2.12}$$

If T can take the value ∞, this requires that X_∞ is defined.

A first step towards answering this question is to note that the *stopped process* X^T defined by

$$(X^T)_n(\omega) = X_{T(\omega)\wedge n}(\omega),$$

is a martingale whenever X is one.

2.13 Theorem. *If T is a stopping time and X is a martingale, then X^T is a martingale.*

Proof. We can write (with an empty sum denoting zero)

$$X_n^T = X_0 + \sum_{i=1}^{n} 1_{i \le T}(X_i - X_{i-1}).$$

Hence $X_{n+1}^T - X_n^T = 1_{n+1 \le T}(X_{n+1} - X_n)$. The variable $1_{n+1 \le T} = 1 - 1_{T \le n}$ is $\mathcal{F}_n$-measurable. Taking the conditional expectation relative to $\mathcal{F}_n$ we find that

$$\mathrm{E}(X_{n+1}^T - X_n^T \,|\, \mathcal{F}_n) = 1_{n+1 \le T} \mathrm{E}(X_{n+1} - X_n \,|\, \mathcal{F}_n) = 0, \qquad \text{a.s.}$$

because X is a martingale. (To be complete, also note that $|X_n^T| \le \max_{1 \le i \le n} |X_i|$ is integrable for every fixed n and verify that X^T is a stochastic process.) ■

2.14 EXERCISE. Show that the sub- and supermartingale properties are also retained under stopping.

If the stopped process X^T is a martingale, then $\mathrm{E}X_n^T = \mathrm{E}X_{T\wedge n}$ is constant in n. If T is bounded and $\mathrm{E}X_0 = 0$, then we can immediately conclude that $\mathrm{E}X_T = 0$ and hence stopping does not help. For general T we would like to take the limit as $n \to \infty$ in the relation $\mathrm{E}X_{T\wedge n} = 0$ and obtain the same conclusion that $\mathrm{E}X_T = 0$. Here we must be careful. If $T < \infty$ we always have that $X_{T\wedge n} \xrightarrow{\text{as}} X_T$ as $n \to \infty$, but we need some integrability to be able to conclude that the expectations converge as well. Domination of X suffices. Later we shall see that uniform integrability is also sufficient, and then we can also allow the stopping time T to take the value ∞ (after defining X_∞ appropriately).

2.15 EXERCISE. Suppose that X is a martingale with uniformly bounded increments: $|X_{n+1} - X_n| \le M$ for every n and some constant M. Show that $\mathrm{E}X_T = 0$ for every stopping time T with $\mathrm{E}T < \infty$.

2.3 Martingale Transforms

Another way to try and beat the system would be to change stakes. If $X_n - X_{n-1}$ is the standard pay-off at time n, we could devise a new game in which our pay-off is $C_n(X_n - X_{n-1})$ at time n. Then our total capital at time n is

$$(C\cdot X)_n := \sum_{i=1}^n C_i(X_i - X_{i-1}), \qquad (C\cdot X)_0 = 0. \tag{2.16}$$

If C_n were allowed to depend on $X_n - X_{n-1}$, then it would be easy to make a profit. We exclude this by requiring that C_n may depend on knowledge of the past only.

2.17 Definition. *A stochastic process C on $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ is predictable if C_n is $\mathcal{F}_{n-1}$ measurable for every $n \ge 1$.*

The process $C\cdot X$ in (2.16) is called a *martingale transform* of X (if X is a martingale). It is the discrete time version of the stochastic integral that we shall be concerned with later. Again we cannot beat the system: the martingale transform is a martingale.

2.18 Theorem. *Suppose that $C_n \in L_p(\Omega, \mathcal{F}, \mathbb{P})$ and $X_n \in L_q(\Omega, \mathcal{F}, \mathbb{P})$ for all n and some $p^{-1} + q^{-1} = 1$.*
(i) If C is predictable and X a martingale, then $C \cdot X$ is a martingale.
(ii) If C is predictable and nonnegative and X is a supermartingale, then $C \cdot X$ is a supermartingale.

Proof. If $Y = C \cdot X$, then $Y_{n+1} - Y_n = C_n(X_{n+1} - X_n)$. Because C_n is $\mathcal{F}_n$-measurable, $\mathrm{E}(Y_{n+1} - Y_n \mid \mathcal{F}_n) = C_n \mathrm{E}(X_{n+1} - X_n \mid \mathcal{F}_n)$ almost surely. Both (i) and (ii) are now immediate. ■

2.4 Doob's Upcrossing Inequality

Let $a < b$ be given numbers. The number of *upcrossings* of the interval $[a, b]$ by the process X in the time interval $\{0, 1, \ldots, n\}$ is defined as the largest integer k for which we can find

$$0 \le s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \le n,$$

with

$$X_{s_i} < a, \quad X_{t_i} > b, \qquad i = 1, 2, \ldots, k.$$

The number of upcrossings is denoted by $U_n[a, b]$. The definition is meant to be "ω"-wise and hence $U_n[a, b]$ is a function on Ω. Because the description involves only finitely many steps, $U_n[a, b]$ is a random variable.

A high number of upcrossings of $[a, b]$ indicates that X is "variable" around the level $[a, b]$. The upcrossing numbers $U_n[a, b]$ are therefore an important tool to study convergence properties of processes. For supermartingales Doob's lemma gives a surprisingly simple bound on the size of the upcrossings, just in terms of the last variable.

2.19 Lemma. *If X is a supermartingale, then*

$$(b - a)\mathrm{E}U_n[a, b] \le \mathrm{E}(X_n - a)^-.$$

Proof. We define a process $C_1, C_2, \ldots$ taking values "0" and "1" only as follows. If $X_0 \ge a$, then $C_n = 0$ until and including the first time n that $X_n < a$, then $C_n = 1$ until and including the first time that $X_n > b$, next $C_n = 0$ until and including the first time that $X_n < a$, etcetera. If $X_0 < a$, then $C_n = 1$ until and including the first time that $X_n > b$, then $C_n = 0$ etcetera. Thus the process is switched "on" and "off" each time the process X crosses the levels a or b. It is "on" during each crossing of the interval $[a, b]$.

We claim that

$$(b-a)U_n[a,b] \le (C\cdot X)_n + (X_n - a)^-, \tag{2.20}$$

where $C\cdot X$ is the martingale transform of the preceding section. To see this note that $(C\cdot X)_n$ is the sum of all increments $X_i - X_{i-1}$ for which $C_i = 1$. A given realization of the process C is a sequence of n zeros and ones. Every consecutive series of ones (a “run”) corresponds to a crossing of $[a,b]$ by X, except possibly the final run (if this ends at position n). The final run (as every run) starts when X is below a and ends at X_n, which could be anywhere. Thus the final run contributes positively to $(C\cdot X)_n$ if $X_n > a$ and can contribute negatively only if $X_n < a$. In the last case it can contribute in absolute value never more than $|X_n - a|$. Thus if we add $(X_n-a)^-$ to $(C\cdot X)_n$, then we obtain at least the sum of the increments over all completed crossings.

It follows from the description, that C_n depends on $C_1,\ldots,C_{n-1}$ and X_{n-1} only. Hence, by induction, the process C is predictable. By Theorem 2.18 the martingale transform $C\cdot X$ is a supermartingale and has nonincreasing mean $\mathrm{E}(C\cdot X)_n \le \mathrm{E}(C\cdot X)_0 = 0$. Taking means across (2.20) concludes the proof. ∎

2.5 Martingale Convergence

In this section we give conditions under which a (sub/super) martingale converges to a limit X_∞, almost surely or in pth mean. Furthermore, we investigate if we can add X_∞ to the end of the sequence $X_0, X_1, \ldots$ and obtain a (sub/super) martingale $X_0, X_1, \ldots, X_\infty$ (with the definition extended to include the time ∞ in the obvious way).

2.21 Theorem. *If X_n is a (sub/super) martingale with $\sup_n \mathrm{E}|X_n| < \infty$, then there exists an integrable random variable X_∞ with $X_n \to X_\infty$ almost surely.*

Proof. If we can show that X_n converges almost surely to a limit X_∞ in $[-\infty,\infty]$, then X_∞ is automatically integrable, because by Fatou's lemma $\mathrm{E}|X_\infty| \le \liminf \mathrm{E}|X_n| < \infty$.

We can assume without loss of generality that X_n is a supermartingale. For a fixed pair of numbers $a<b$, let

$$F_{a,b} = \Bigl\{\omega\in\Omega\colon \liminf_{n\to\infty} X_n(\omega) < a \le b < \limsup_{n\to\infty} X_n(\omega)\Bigr\}.$$

If $\lim_{n\to\infty} X_n(\omega)$ does not exist in $[-\infty,\infty]$, then we can find $a<b$ such that $\omega\in F_{a,b}$. Because the rational numbers are dense in $\mathbb{R}$, we can even

find such $a < b$ among the rational numbers. The theorem is proved if we can show that $\mathbb{P}(F_{a,b}) = 0$ for every of the countably many pairs $(a, b) \in \mathbb{Q}^2$.

Fix $a < b$ and let $U_n[a, b]$ be the number of upcrossings of $[a, b]$ on $\{0, \ldots, n\}$ by X. If $\omega \in F_{a,b}$, then $U_n[a, b] \uparrow \infty$ as $n \to \infty$ and hence by monotone convergence $\mathrm{E}U_n[a, b] \uparrow \infty$ if $\mathbb{P}(F_{a,b}) > 0$. However, by Doob's upcrossing's inequality

$$(b - a)\mathrm{E}U_n[a, b] \le \mathrm{E}(X_n - a)^- \le \mathrm{E}|X_n - a| \le \sup_n \mathrm{E}|X_n| + |a|.$$

The right side is finite by assumption and hence the left side cannot increase to ∞. We conclude that $\mathbb{P}(F_{a,b}) = 0$. ∎

2.22 EXERCISE. Let X_n be a nonnegative supermartingale. Show that $\sup_n \mathrm{E}|X_n| < \infty$ and hence X_n converges almost surely to some limit.

If we define X_∞ as $\lim X_n$ if this limit exists and as 0 otherwise, then, if X is adapted, X_∞ is measurable relative to the σ-field

$$\mathcal{F}_\infty = \sigma(\mathcal{F}_1, \mathcal{F}_2, \ldots).$$

Then the stochastic process $X_0, X_1, \ldots, X_\infty$ is adapted to the filtration $\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_\infty$. We may ask whether the martingale property $\mathrm{E}(X_n|\,\mathcal{F}_m) = X_m$ (for $n \ge m$) extends to the case $n = \infty$. The martingale is then called *closed*. From Example 2.6 we know that the martingale $X_m = \mathrm{E}(X_\infty|\,\mathcal{F}_m)$ is uniformly integrable. This condition is also sufficient.

2.23 Theorem. *If X is a uniformly integrable (sub/super) martingale, then there exists a random variable X_∞ such that $X_n \to X_\infty$ almost surely and in L_1. Moreoever,*
(i) If X is a martingale, then $X_n = \mathrm{E}(X_\infty|\,\mathcal{F}_n)$ almost surely for every $n \ge 0$.
(ii) If X is a submartingale, then $X_n \le \mathrm{E}(X_\infty|\,\mathcal{F}_n)$ almost surely for every $n \ge 0$.

Proof. The first assertion is a corollary of the preceding theorem and the fact that a uniformly integrable sequence of random variables that converges almost surely converges in L_1 as well.

Statement (i) follows by taking the L_1-limit as $n \to \infty$ in the equality $X_m = \mathrm{E}(X_n|\,\mathcal{F}_m)$, where we use that $\|\mathrm{E}(X_n|\,\mathcal{F}_m) - \mathrm{E}(X_\infty|\,\mathcal{F}_m)\|_1 \le \|X_n - X_\infty\|_1 \to 0$, so that the right side converges to $\mathrm{E}(X_\infty|\,\mathcal{F}_m)$.

Statement (ii) follows similarly (where we must note that L_1-convergence retains ordering almost surely), or by the following argument. By the submartingale property, for every $m \le n$, $\mathrm{E}X_m 1_F \le \mathrm{E}X_n 1_F$. By uniformly integrability of the process $X1_F$ we can take the limit as $n \to \infty$ in this and obtain that $\mathrm{E}X_m 1_F \le \mathrm{E}\mathrm{E}(X_n|\,\mathcal{F}_m)1_F = \mathrm{E}X_\infty 1_F$ for every $F \in \mathcal{F}_m$. The right side equals $\mathrm{E}X'_m 1_F$ for $X'_m = \mathrm{E}(X_\infty|\,\mathcal{F}_m)$ and hence

$\mathrm{E}(X_m - X_m')1_F \leq 0$ for every $F \in \mathcal{F}_m$. This implies that $X_m - X_m' \leq 0$ almost surely. ∎

2.24 Corollary. *If ξ is an integrable random variable and $X_n = \mathrm{E}(\xi|\,\mathcal{F}_n)$ for a filtration $\{\mathcal{F}_n\}$, then $X_n \to \mathrm{E}(\xi|\,\mathcal{F}_\infty)$ almost surely and in L_1.*

Proof. Because X is a uniformly integrable martingale, the preceding theorem gives that $X_n \to X_\infty$ almost surely and in L_1 for some integrable random variable X_∞, and $X_n = \mathrm{E}(X_\infty|\,\mathcal{F}_n)$ for every n. The variable X_∞ can be chosen $\mathcal{F}_\infty$ measurable (a matter of null sets). It follows that $\mathrm{E}(\xi|\,\mathcal{F}_n) = X_n = \mathrm{E}(X_\infty|\,\mathcal{F}_n)$ almost surely for every n and hence $\mathrm{E}\xi 1_F = \mathrm{E}X_\infty 1_F$ for every $F \in \cup_n \mathcal{F}_n$. But the set of F for which this holds is a σ-field and hence $\mathrm{E}\xi 1_F = \mathrm{E}X_\infty 1_F$ for every $F \in \mathcal{F}_\infty$. This shows that $X_\infty = \mathrm{E}(\xi|\,\mathcal{F}_\infty)$. ∎

The preceding theorem applies in particular to L_p-bounded martingales (for $p > 1$). But then more is true.

2.25 Theorem. *If X is an L_p-bounded martingale ($p > 1$), then there exists a random variable X_∞ such that $X_n \to X_\infty$ almost surely and in L_p.*

Proof. By the preceding theorem $X_n \to X_\infty$ almost surely and in L_1 and moreover $\mathrm{E}(X_\infty|\,\mathcal{F}_n) = X_n$ almost surely for every n. By Jensen's inequality $|X_n|^p = \bigl|\mathrm{E}(X_\infty|\,\mathcal{F}_n)\bigr|^p \leq \mathrm{E}\bigl(|X_\infty|^p|\,\mathcal{F}_n\bigr)$ and hence $\mathrm{E}|X_n|^p \leq \mathrm{E}|X_\infty|^p$ for every n. The theorem follows from Lemma 1.22. ∎

2.26 EXERCISE. Show that the theorem remains true if X is a nonnegative submartingale.

Warning. A stochastic process that is bounded in L_p and converges almost surely to a limit does not necessarily converge in L_p. For this $|X|^p$ must be uniformly integrable. The preceding theorem makes essential use of the martingale property of X. Also see Section 2.9.

2.6 Reverse Martingale Convergence

Thus far we have considered filtrations that are increasing. In this section, and in this section only, we consider a *reverse filtration*

$$\mathcal{F} \supset \mathcal{F}_0 \supset \mathcal{F}_1 \supset \cdots \supset \mathcal{F}_\infty = \cap_n \mathcal{F}_n.$$

2.27 Definition. *An adapted, integrable stochastic process X on the reverse filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$ is a*
(i) reverse martingale if $\mathrm{E}(X_m|\, \mathcal{F}_n) = X_n$ a.s. *for all $m \leq n$.*
(ii) reverse submartingale if $\mathrm{E}(X_m|\, \mathcal{F}_n) \geq X_n$ a.s. *for all $m \leq n$.*
(ii) reverse supermartingale if $\mathrm{E}(X_m|\, \mathcal{F}_n) \leq X_n$ a.s. *for all $m \leq n$.*

It is more insightful to say that a reverse (sub/super) martingale is a process $X = (X_0, X_1, \ldots)$ such that the sequence $\ldots, X_2, X_1, X_0$ is a (sub/super) martingale as defined before, relative to the filtration $\cdots \subset \mathcal{F}_2 \subset \mathcal{F}_1 \subset \mathcal{F}_0$. In deviation from the definition of (sub/super) martingales, the time index $\ldots, 2, 1, 0$ then runs against the natural order and there is a "final time" 0. Thus the (sub/super) martingales obtained by reversing a reverse (sub/super) martingale are automatically closed (by the "final element" X_0).

2.28 Example. If ξ is an integrable random variable and $\{\mathcal{F}_n\}$ an arbitrary reverse filtration, then $X_n = \mathrm{E}(\xi|\, \mathcal{F}_n)$ defines a reverse martingale. We can include $n = \infty$ in this definition.

Because every reverse martingale satisfies $X_n = \mathrm{E}(X_0|\, \mathcal{F}_n)$, this is actually the only type of reverse martingale. □

2.29 Example. If $\{N(t){:}\, t > 0\}$ is a standard Poisson process, and $t_1 > t_2 > \cdots \geq 0$ a decreasing sequence of numbers, then $X_n = N(t_n) - t_n$ is a reverse martingale relative to the reverse filtration $\mathcal{F}_n = \sigma(N(t){:}\, t \leq t_n)$.

The verification of this is exactly the same as the for the corresponding martingale property of this process for an increasing sequence of times. □

That a reverse martingale becomes an ordinary martingale if we turn it around may be true, but it is not very helpful for the convergence results that we are interested in. The results on (sub/super) martingales do not imply those for reverse (sub/super) martingales, because the "infiniteness" is on the other end of the sequence. Fortunately, the same techniques apply.

2.30 Theorem. *If X is a uniformly integrable reverse (sub/super) martingale, then there exists a random variable X_∞ such that $X_n \to X_\infty$ almost surely and in mean as $n \to \infty$. Moreover,*
(i) If X is a reverse martingale, then $\mathrm{E}(X_m|\, \mathcal{F}_\infty) = X_\infty$ a.s. *for every m.*
(ii) If X is a reverse submartingale, then $\mathrm{E}(X_m|\, \mathcal{F}_\infty) \geq X_\infty$ a.s. *for every m.*

Proof. Doob's upcrossings inequality is applicable to bound the number of upcrossings of $X_0, \ldots, X_n$, because $X_n, X_{n-1}, \ldots, X_0$ is a supermartingale if X is a reverse supermartingale. Thus we can mimic the proof of Theorem 2.21 to prove the existence of an almost sure limit X_∞. By uniform integrability this is then also a limit in L_1.

The submartingale property implies that $\mathrm{E}X_m 1_F \geq \mathrm{E}X_n 1_F$ for every $F \in \mathcal{F}_n$ and $n \geq m$. In particular, this is true for every $F \in \mathcal{F}_\infty$. Upon taking the limit as $n \to \infty$, we see that $\mathrm{E}X_m 1_F \geq \mathrm{E}X_\infty 1_F$ for every $F \in \mathcal{F}_\infty$. This proves the relationship in (ii). The proof of (i) is easier. ∎

2.31 EXERCISE. Let $\{\mathcal{F}_n\}$ be a reverse filtration and ξ integrable. Show that $\mathrm{E}(\xi|\,\mathcal{F}_n) \to \mathrm{E}(\xi|\,\mathcal{F}_\infty)$ in L_1 and in mean for $\mathcal{F}_\infty = \cap_n \mathcal{F}_n$. What if $X_1, X_2, \ldots$ are i.i.d.?

* **2.32 Example (Strong law of large numbers).** A stochastic process $X = (X_1, X_2, \ldots)$ is called *exchangeable* if for every n the distribution of $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ is the same for every permutation $(\sigma(1), \ldots, \sigma(n))$ of $(1, \ldots, n)$. If $\mathrm{E}|X_1| < \infty$, then the sequence of averages $\bar{X}_n$ converges almost surely and in mean to a limit (which may be stochastic).

To prove this consider the reverse filtration $\mathcal{F}_n = \sigma(\overline{X}_n, \overline{X}_{n+1}, \ldots)$. The σ-field $\mathcal{F}_n$ "depends" on $X_1, \ldots, X_n$ only through $X_1 + \cdots + X_n$ and hence by symmetry and exchangeability $\mathrm{E}(X_i|\,\mathcal{F}_n)$ is the same for $i = 1, \ldots, n$. Then

$$\overline{X}_n = \mathrm{E}(\overline{X}_n|\,\mathcal{F}_n) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i|\,\mathcal{F}_n) = \mathrm{E}(X_1|\,\mathcal{F}_n), \qquad \text{a.s..}$$

The right side converges almost surely and in mean by the preceding theorem. □

2.33 EXERCISE. Identify the limit in the preceding example as $\mathrm{E}(X_1|\,\mathcal{F}_\infty)$ for $\mathcal{F}_\infty = \cap_n \mathcal{F}_n$.

Because, by definition, a reverse martingale satisfies $X_n = \mathrm{E}(X_0|\,\mathcal{F}_n)$, a reverse martingale is automatically uniformly integrable. Consequently the preceding theorem applies to any reverse martingale. A reverse (sub/super) martingale is uniformly integrable as soon as it is bounded in L_1. In fact, it suffices to verify that $\mathrm{E}X_n$ is bounded below/above.

2.34 Lemma. *A reverse supermartingale X is uniformly integrable if and only if $\mathrm{E}X_n$ is bounded above (in which case it increases to a finite limit as $n \to \infty$).*

Proof. The expectations $\mathrm{E}X_n$ of any uniformly integrable process X are bounded. Therefore, the "only if" part of the lemma is clear and the "if" part is the nontrivial part of the lemma. Suppose that X is a reverse supermartingale.

The sequence of expectations $\mathrm{E}X_n$ is nondecreasing in n by the reverse supermartingale property. Because it is bounded above it converges to a finite limit. Furthermore, $X_n \geq \mathrm{E}(X_0|\,\mathcal{F}_n)$ for every n and hence X^- is uniformly integrable, since $\mathrm{E}(X_0|\,\mathcal{F}_n)$ is. It suffices to show that X^+ is uniformly integrable, or equivalently that $\mathrm{E}X_n 1_{X_n>M} \to 0$ as $M \to \infty$, uniformly in n.

By the supermartingale property and because $\{X_n \le M\} \in \mathcal{F}_n$, for every $M, N > 0$ and every $m \le n$,

$$\begin{aligned}\mathrm{E}X_n 1_{X_n>M} &= \mathrm{E}X_n - \mathrm{E}X_n 1_{X_n\le M} \le \mathrm{E}X_n - \mathrm{E}X_m 1_{X_n\le M}\\ &= \mathrm{E}X_n - \mathrm{E}X_m + \mathrm{E}X_m 1_{X_n>M}\\ &\le \mathrm{E}X_n - \mathrm{E}X_m + \mathrm{E}X_m^+ 1_{X_m>N} + \frac{N}{M}\mathrm{E}X_n^+.\end{aligned}$$

We can make the right side arbitrarily small, uniformly in $n \ge m$, by first choosing m sufficiently large (so that $\mathrm{E}X_n - \mathrm{E}X_m$ is small), next choosing N sufficiently large and finally choosing M large. For the given m we can increase M, if necessary, to ensure that $\mathrm{E}X_n 1_{X_n>M}$ is also small for every $0 \le n \le m$. ∎

* 2.7 Doob Decomposition

If a martingale is a model for a fair game, then non-martingale processes should correspond to unfair games. This can be made precise by the Doob decomposition of an adapted process as a sum of a martingale and a predictable process. The Doob decomposition is the discrete time version of the celebrated (and much more complicated) Doob-Meyer decomposition of a “semi-martingale” in continuous time. We need it here to extend some results on martingales to (sub/super) martingales.

2.35 Theorem. *For any adapted process X there exists a martingale M and a predictable process A, unique up to null sets, both 0 at 0, such that $X_n = X_0 + M_n + A_n$, for every $n \ge 0$,*

Proof. If we set $A_0 = 0$ and $A_n - A_{n-1} = \mathrm{E}(X_n - X_{n-1}|\,\mathcal{F}_{n-1})$, then A is predictable. In order to satisfy the equation, we must set

$$M_0 = 0, \qquad M_n - M_{n-1} = X_n - X_{n-1} - \mathrm{E}(X_n - X_{n-1}|\,\mathcal{F}_{n-1}).$$

This clearly defines a martingale M.

Conversely, if the decomposition holds as stated, then $\mathrm{E}(X_n - X_{n-1}|\,\mathcal{F}_{n-1}) = \mathrm{E}(A_n - A_{n-1}|\,\mathcal{F}_{n-1})$, because M is a martingale. The right side is equal to $A_n - A_{n-1}$ because A is predictable. ∎

If $X_n - X_{n-1} = (M_n - M_{n-1}) + (A_n - A_{n-1})$ were our gain in the nth game, then our strategy could be to play if $A_n - A_{n-1} > 0$ and not to play if this is negative. Because A is predictable, we "know" this before time n and hence this would be a valid strategy. The martingale part M corresponds to a fair game and would give us expected gain zero. Relative to the predictable part we would avoid all losses and make all gains. Thus our expected profit would certainly be positive (unless we never play). We conclude that only martingales correspond to fair games.

From the fact that $A_n - A_{n-1} = \mathrm{E}(X_n - X_{n-1} \,|\, \mathcal{F}_{n-1})$ it is clear that (sub/super) martingales X correspond precisely to the cases that the sample paths of A are increasing or decreasing. These are the case where we would always or never play.

2.8 Optional Stopping

Let T be a stopping time relative to the filtration $\mathcal{F}_n$. Just as $\mathcal{F}_n$ are the events "known at time n", we like to introduce a σ-field $\mathcal{F}_T$ of "events known at time T". This is to be an ordinary σ-field. Plugging T into $\mathcal{F}_n$ would not do, as this would give something random.

2.36 Definition. *The σ-field $\mathcal{F}_T$ is defined as the collection of all $F \subset \Omega$ such that $F \cap \{T \le n\} \in \mathcal{F}_n$ for all $n \in \bar{\mathbb{Z}}_+$. (This includes $n = \infty$, where $\mathcal{F}_\infty = \sigma(\mathcal{F}_0, \mathcal{F}_1, \ldots)$.)*

2.37 EXERCISE. Show that $\mathcal{F}_T$ is indeed a σ-field.

2.38 EXERCISE. Show that $\mathcal{F}_T$ can be equivalently described as the collection of all $F \subset \Omega$ such that $F \cap \{T = n\} \in \mathcal{F}_n$ for all $n \in \bar{\mathbb{Z}}_+$.

2.39 EXERCISE. Show that $\mathcal{F}_T = \mathcal{F}_n$ if $T \equiv n$.

2.40 EXERCISE. Show that X_T is $\mathcal{F}_T$-measurable if $\{X_n : n \in \bar{\mathbb{Z}}_+\}$ is adapted.

2.41 Lemma. *Let S and T be stopping times. Then*
(i) if $S \le T$, then $\mathcal{F}_S \subset \mathcal{F}_T$.
(ii) $\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S \wedge T}$.

Proof. (i). If $S \le T$, then $F \cap \{T \le n\} = \big(F \cap \{S \le n\}\big) \cap \{T \le n\}$. If $F \in \mathcal{F}_S$, then $F \cap \{S \le n\} \in \mathcal{F}_n$ and hence, because always $\{T \le n\} \in \mathcal{F}_n$, the right side is in $\mathcal{F}_n$. Thus $F \in \mathcal{F}_T$.

(ii). By (i) we have $\mathcal{F}_{S \wedge T} \subset \mathcal{F}_S \cap \mathcal{F}_T$. Conversely, if $F \in \mathcal{F}_S \cap \mathcal{F}_T$, then $F \cap \{S \wedge T \le n\} = (F \cap \{S \le n\}) \cup (F \cap \{T \le n\}) \in \mathcal{F}_n$ for every n and hence $F \in \mathcal{F}_{S \wedge T}$. ∎

If the (sub/super) martingale X is uniformly integrable, then there exists an integrable random variable X_∞ such that $X_n \to X_\infty$ almost surely and in mean, by Theorem 2.23. Then we can define X_T as in (2.12), also if T assumes the value ∞. The optional stopping theorem shows that in this case we may replace the fixed times $m \le n$ in the defining martingale relationship $\mathrm{E}(X_n|\,\mathcal{F}_m) = X_m$ by stopping times $S \le T$.

2.42 Theorem (Optional stopping). *If X is a uniformly integrable supermartingale, then X_T is integrable for any stopping time T. Furthermore,*
(i) If T is a stopping time, then $\mathrm{E}(X_\infty|\,\mathcal{F}_T) \le X_T$ a.s..
(ii) If $S \le T$ are stopping times, then $\mathrm{E}(X_T|\,\mathcal{F}_S) \le X_S$ a.s..

Proof. First we note that X_T is $\mathcal{F}_T$-measurable (see Exercise 2.40). For (i) we wish to prove that $\mathrm{E}X_\infty 1_F \le \mathrm{E}X_T 1_F$ for all $F \in \mathcal{F}_T$. Now

$$\mathrm{E}X_\infty 1_F = \mathrm{E}\sum_{n=0}^{\infty+} X_\infty 1_F 1_{T=n} = \sum_{n=0}^{\infty+} \mathrm{E}X_\infty 1_F 1_{T=n},$$

by the dominated convergence theorem. (The "+" in the upper limit $\infty+$ of the sums indicates that the sums also include a term $n = \infty$.) Because $F \cap \{T = n\} \in \mathcal{F}_n$ and $\mathrm{E}(X_\infty|\,\mathcal{F}_n) \le X_n$ for every n, the supermartingale property gives that the right side is bounded above by

$$\sum_{n=0}^{\infty+} \mathrm{E}X_n 1_F 1_{T=n} = \mathrm{E}X_T 1_F,$$

if X_T is integrable, by the dominated convergence theorem. This gives the desired inequality and concludes the proof of (i) for any stopping time T for which X_T is integrable.

If T is bounded, then $|X_T| \le \max_{m\le n} |X_m|$ for n an upper bound on T and hence X_T is integrable. Thus we can apply the preceding paragraph to see that $\mathrm{E}(X_\infty|\,\mathcal{F}_{T\wedge n}) \le X_{T\wedge n}$ almost surely for every n. If X is a martingale, then this inequality is valid for both X and $-X$ and hence, for every n,

$$X_{T\wedge n} = \mathrm{E}(X_\infty|\,\mathcal{F}_{T\wedge n}), \qquad \text{a.s..}$$

for every n. If $n \to \infty$ the left side converges to X_T. The right side is a uniformly integrable martingale that converges to an integrable limit in L_1 by Theorem 2.23. Because the limits must agree, X_T is integrable.

Combining the preceding we see that $X_T = \mathrm{E}(X_\infty|\,\mathcal{F}_T)$ for every stopping time T if X is a uniformly integrable martingale. Then for stopping times $S \le T$ the towering property of conditional expectations gives $\mathrm{E}(X|\,\mathcal{F}_S) = \mathrm{E}\big(\mathrm{E}(X_\infty|\,\mathcal{F}_T)|\,\mathcal{F}_S\big) = \mathrm{E}(X_\infty|\,\mathcal{F}_S)$, because $\mathcal{F}_S \subset \mathcal{F}_T$. Applying (i) again we see that the right side is equal to X_S. This proves (ii) in the case that X is a martingale.

To extend the proof to supermartingales X, we employ the Doob decomposition $X_n = X_0 + M_n - A_n$, where M is a martingale with $M_0 = 0$ and A is a nondecreasing (predictable) process with $A_0 = 0$. Then $\mathrm{E}A_n = \mathrm{E}X_0 - \mathrm{E}X_n$ is bounded if X is uniformly integrable. Hence $A_\infty = \lim A_n$ is integrable, and A is dominated (by A_∞) and hence uniformly integrable. Then M must be uniformly integrable as well, whence, by the preceding, M_T is integrable and $\mathrm{E}(M_T|\,\mathcal{F}_S) = M_T$. It follows that $X_T = X_0 + M_T - A_T$ is integrable. Furthermore, by linearity of the conditional expectation, for $S \leq T$,

$$\begin{aligned}\mathrm{E}(X_T|\,\mathcal{F}_S) &= X_0 + \mathrm{E}(M_T|\,\mathcal{F}_S) - \mathrm{E}(A_T|\,\mathcal{F}_S)\\ &\leq X_0 + M_S - A_S = X_S,\end{aligned}$$

because $A_S \leq A_T$ implies that $A_S \leq \mathrm{E}(A_T|\,\mathcal{F}_S)$ almost surely. This concludes the proof of (ii). The statement (i) (with S playing the role of T) is the special case that $T = \infty$. ∎

One consequence of the preceding theorem is that $\mathrm{E}X_T = \mathrm{E}X_0$, whenever T is a stopping time and X a uniformly integrable martingale.

Warning. The condition that X be uniformly integrable cannot be omitted.

2.9 Maximal Inequalities

A maximal inequality for a stochastic process X is a bound on some aspect of the distribution of $\sup_n X_n$. Suprema over stochastic processes are usually hard to control, but not so for martingales. Somewhat remarkably, we can bound the norm of $\sup_n X_n$ by the supremum of the norms, up to a constant.

We start with a probability inequality.

2.43 Lemma. *If X is a submartingale, then for any $x \geq 0$ and every $n \in \mathbb{Z}_+$,*

$$x\mathbb{P}\Bigl(\max_{0\leq i\leq n} X_i \geq x\Bigr) \leq \mathrm{E}X_n 1_{\max_{0\leq i\leq n} X_i \geq x}.$$

Proof. We can write the event in the left side as the disjoint union $\cup_{i=0}^n F_i$ of the events

$$F_0 = \{X_0 \geq x\}, \quad F_1 = \{X_0 < x, X_1 \geq x\},$$
$$F_2 = \{X_0 < x, X_1 < x, X_2 \geq x\}, \ldots.$$

Because $F_i \in \mathcal{F}_i$, the submartingale property gives $\mathrm{E}X_n 1_{F_i} \geq \mathrm{E}X_i 1_{F_i} \geq x\mathbb{P}(F_i)$, because $X_i \geq x$ on F_i. Summing this over $i = 0, 1, \ldots, n$ yields the result. ∎

2.44 Corollary. *If X is a nonnegative submartingale, then for any $p > 1$ and $p^{-1} + q^{-1} = 1$, and every $n \in \mathbb{Z}_+$,*

$$\Bigl\| \max_{0\le i\le n} X_i \Bigr\|_p \le q \|X_n\|_p.$$

If X is bounded in $L_p(\Omega, \mathcal{F}, \mathbb{P})$, then $X_n \to X_\infty$ in L_p for some random variable X_∞ and

$$\Bigl\| \sup_n X_n \Bigr\|_p \le q \|X_\infty\|_p = q \sup_n \|X_n\|_p.$$

Proof. Set $Y_n = \max_{0\le i\le n} X_i$. By Fubini's theorem (or partial integration),

$$\mathrm{E} Y_n^p = \int_0^\infty p x^{p-1} \mathbb{P}(Y_n \ge x)\, dx \le \int_0^\infty p x^{p-2} \mathrm{E} X_n 1_{Y_n \ge x}\, dx,$$

by the preceding lemma. After changing the order of integration and expectation, we can write the right side as

$$p \mathrm{E}\Bigl(X_n \int_0^{Y_n} x^{p-2}\, dx \Bigr) = \frac{p}{p-1} \mathrm{E} X_n Y_n^{p-1}.$$

Here $p/(p-1) = q$ and $\mathrm{E} X_n Y_n^{p-1} \le \|X_n\|_p \|Y_n^{p-1}\|_q$ by Hölder's inequality. Thus $\mathrm{E} Y_n^p \le \|X_n\|_p \|Y_n^{p-1}\|_q$. If $Y_n \in L_p(\Omega, \mathcal{F}, \mathbb{P})$, then we can rearrange this inequality to obtain the result.

This rearranging is permitted only if $\mathrm{E} Y_n^p < \infty$. By the submartingale property $0 \le X_i \le \mathrm{E}(X_n \,|\, \mathcal{F}_i)$, whence $\mathrm{E} X_i^p \le \mathrm{E} X_n^p$, by Jensen's inequality. Thus $\mathrm{E} Y_n^p$ is finite whenever $\mathrm{E} X_n^p$ is finite, and this we can assume without loss of generality.

Because X is a nonnegative submartingale, so is X^p and hence the sequence $\mathrm{E} X_n^p$ is nondecreasing. If X is L_p-bounded (for $p > 1$), then it is uniformly integrable and hence $X_n \to X_\infty$ almost surely for some random variable X_∞, by Theorem 2.23. Taking the limit as $n \to \infty$ in the first assertion, we find by the monotone convergence theorem that

$$\mathrm{E} \sup_n X_n^p = \mathrm{E} Y_\infty^p = \lim_{n\to\infty} \mathrm{E} Y_n^p \le q^p \lim_{n\to\infty} \mathrm{E} X_n^p = q^p \sup_n \mathrm{E} X_n^p.$$

The supremum on the left does not increase if we extend it to $n \in \bar{\mathbb{Z}}_+$. Because $|X_n - X|$ is dominated by $2Y_\infty$, we find that $X_n \to X_\infty$ also in L_p and hence $\mathrm{E} X_\infty^p = \lim_{n\to\infty} \mathrm{E} X_n^p$. ■

The results of this section apply in particular to the submartingales formed by applying a convex function to a martingale. For instance, $|X|$, X^2 or $e^{\alpha X}$ for some $\alpha > 0$ and some martingale X. This yields a wealth of useful inequalities. For instance, for any martingale X,

$$\Bigl\| \sup_n |X_n| \Bigr\|_2 \le 2 \sup_n \|X_n\|_2.$$

2.45 EXERCISE. Let $Y_1, Y_2, \ldots$ be an i.i.d. sequence of random variables with mean zero. set $S_n = \sum_{i=1}^n Y_i$. Show that $\mathrm{E}\max_{1\le i\le n} S_i^2 \le 4\mathrm{E}S_n^2$.

3
Discrete Time Option Pricing

In this chapter we discuss the binary tree model for the pricing of "contingent claims" such as options, due to Cox, Ross and Rubinstein. In this model the price S_n of a stock is evaluated and changes at the discrete time instants $n = 0, 1, \ldots$ only and it is assumed that its increments $S_n - S_{n-1}$ can assume two values only. (This is essential; the following would not work if the increments could assume e.g. three values.) We assume that S is a stochastic process on a given probability space and let $\mathcal{F}_n$ be its natural filtration.

Next to stock the model allows for bonds. A bond is a "risk-free investment", comparable to a deposit in a savings account, whose value increases deterministically according to the relation

$$R_n = (1 + r_n)R_{n-1}, \qquad R_0 = 1,$$

the constant $r_n > 0$ being the "interest rate" in the time interval $(n-1, n)$. A general name for both stock and bond is "asset".

A "portfolio" is a combination of bonds and stocks. Its contents may change over time. A portfolio containing A_n bonds and B_n stocks at time n possesses the value

$$V_n = A_n R_n + B_n S_n. \tag{3.1}$$

A pair of processes (A, B), giving the contents over time, is an "investment strategy" if the processes are predictable. We call a strategy "self-financing" if after investment of an initial capital at time 0, we can reshuffle the portfolio according to the strategy without further capital import. Technically this requirement means that, for every $n \geq 1$,

$$A_n R_{n-1} + B_n S_{n-1} = A_{n-1} R_{n-1} + B_{n-1} S_{n-1}. \tag{3.2}$$

Thus the capital V_{n-1} at time $n - 1$ (on the right side of the equation) is used in the time interval $(n - 1, n)$ to exchange bonds for stocks or vice versa at the current prices R_{n-1} and S_{n-1}. The left side of the equation gives the value of the portfolio after the reshuffling. At time n the value changes to $V_n = A_n S_n + B_n S_n$, due to the changes in the values of the underlying assets.

A "derivative" is a financial contract that is based on the stock. A popular derivative is the option, of which there are several varieties. A "European call option" is a contract giving the owner of the option the right to buy the stock at some fixed time N (the "term" or "expiry time" of the option) in the future at a fixed price K (the "strike price"). At the expiry time the stock is worth S_N. If $S_N > K$, then the owner of the option will exercise his right and buy the stock, making a profit of $S_N - K$. (He could sell off the stock immediately, if he wanted to, making a profit of $S_N - K$.) On the other hand, if $S_N < K$, then the option is worthless. (It is said to be "out of the money".) If the owner of the option would want to buy the stock, he would do better to buy it on the regular market, for the price S_N, rather than use the option.

What is a good price for an option? Because the option gives a right and no obligation it must cost money to get one. The value of the option at expiry time is, as seen in the preceding discussion, $(S_N - K)^+$. However, we want to know the price of the option at the beginning of the term. A reasonable guess would be $\mathrm{E}(S_N - K)^+$, where the expectation is taken relative to the "true" law of the stock price S_N. We don't know this law, but we could presumably estimate it after observing the stock market for a while.

Wrong! Economic theory says that the actual distribution of S_N has nothing to do with the value of the option at the beginning of the term. This economic reasoning is based on the following theorem.

Recall that we assume that possible values of the stock process S form a binary tree. Given its value S_{n-1} at time $n-1$, there are two possibilities for the value S_n. For simplicity of notation assume that

$$S_n \in \{a_n S_{n-1}, b_n S_{n-1}\},$$

where a_n and b_n are known numbers. We assume that given $\mathcal{F}_{n-1}$ each of the two possibilities is chosen with fixed probabilities $1 - p_n$ and p_n. We do not assume that we know the "true" numbers p_n, but we do assume that we know the numbers (a_n, b_n). Thus, for $n \geq 1$,

$$\begin{aligned} \mathbb{P}(S_n = a_n S_{n-1} \mid \mathcal{F}_{n-1}) &= 1 - p_n, \\ \mathbb{P}(S_n = b_n S_{n-1} \mid \mathcal{F}_{n-1}) &= p_n. \end{aligned} \tag{3.3}$$

(Pretty unrealistic, this, but good exercise for the continuous time case.) It follows that the complete distribution of the process S, given its value S_0 at time 0, can be parametrized by a vector $p = (p_1, \ldots, p_n)$ of probabilities.

3.4 Theorem. *Suppose that $0 < a_n < 1 + r_n < b_n$ for all n and nonzero numbers a_n, b_n. Then there exists a unique self-financing strategy (A, B) with value process V (as in (3.1)) such that*
(i) $V \geq 0$.
(ii) $V_N = (S_N - K)^+$.
This strategy requires an initial investment of
(iii) $V_0 = \tilde{\mathrm{E}} R_N^{-1}(S_N - K)^+$,
where $\tilde{\mathrm{E}}$ is the expectation under the probability measure defined by (3.3) with $p = (\tilde{p}_1, \ldots, \tilde{p}_n)$ given by

$$\tilde{p}_n := \frac{1 + r_n - a_n}{b_n - a_n}.$$

The values $\tilde{p}$ are the unique values in $(0, 1)$ that ensure that the process $\tilde{S}$ defined by $\tilde{S}_n = R_n^{-1} S_n$ is a martingale.

Proof. By assumption, given $\mathcal{F}_{n-1}$, the variable S_n is supported on the points $a_n S_{n-1}$ and $b_n S_{n-1}$ with probabilities $1 - p_n$ and p_n. Then

$$\mathrm{E}(\tilde{S}_n |\, \mathcal{F}_{n-1}) = R_n^{-1}\big((1 - p_n)a_n + p_n b_n\big)S_{n-1}.$$

This is equal to $\tilde{S}_{n-1} = R_{n-1}^{-1} S_{n-1}$ if and only if

$$(1 - p_n)a_n + p_n b_n = \frac{R_n}{R_{n-1}} = 1 + r_n, \qquad \leftrightarrow \quad p_n = \frac{1 + r_n - a_n}{b_n - a_n}.$$

By assumption this value of p_n is contained in $(0, 1)$. Thus there exists a unique martingale measure, as claimed.

The process $\tilde{V}_n = \tilde{\mathrm{E}}\big(R_N^{-1}(S_N - K)^+ |\, \mathcal{F}_n\big)$ is a $\tilde{p}$-martingale. Given $\mathcal{F}_{n-1}$ the variables $\tilde{V}_n - \tilde{V}_{n-1}$ and $\tilde{S}_n - \tilde{S}_{n-1}$ are both functions of S_n/S_{n-1} and hence supported on two points (dependent on $\mathcal{F}_{n-1}$). (Note that the possible values of S_n are S_0 times a product of the numbers a_n and b_n and hence are nonzero by assumption.) Because these variables are martingale differences, they have conditional mean zero under $\tilde{p}_n$. Together this implies that there exists a unique $\mathcal{F}_{n-1}$-measurable variable B_n (given $\mathcal{F}_{n-1}$ this is a "constant") such that (for $n \geq 1$)

$$\tilde{V}_n - \tilde{V}_{n-1} = B_n(\tilde{S}_n - \tilde{S}_{n-1}). \tag{3.5}$$

Given this process B, define a process A to satisfy

$$A_n R_{n-1} + B_n S_{n-1} = R_{n-1}\tilde{V}_{n-1}. \tag{3.6}$$

Then both the processes A and B are predictable and hence (A, B) is a strategy. (The values (A_0, B_0) matter little, because we change the portfolio to (A_1, B_1) before anything happens to the stock or bond at time 1; we can choose $(A_0, B_0) = (A_1, B_1)$.)

The preceding displays imply

$$
\begin{aligned}
A_n + B_n\tilde S_{n-1} &= \tilde V_{n-1},\\
A_n + B_n\tilde S_n &= \tilde V_{n-1} + B_n(\tilde S_n - \tilde S_{n-1}) = \tilde V_n, \qquad \text{by (3.5)},\\
R_nA_n + B_nS_n &= R_n\tilde V_n.
\end{aligned}
$$

Evaluating the last line with $n-1$ instead of n and comparing the resulting equation to (3.6), we see that the strategy (A,B) is self-financing.

By the last line of the preceding display the value of the portfolio (A_n,B_n) at time n is

$$V_n = R_n\tilde V_n = R_n\tilde{\mathrm{E}}\bigl(R_N^{-1}(S_N-K)^+|\,\mathcal{F}_n\bigr).$$

At time N this becomes $V_N = (S_N-K)^+$. At time 0 the value is $V_0 = R_0\tilde{\mathrm{E}}R_N^{-1}(S_N-K)^+$. That $V\geq 0$ is clear from the fact that $\tilde V\geq 0$, being a conditional expectation of a nonnegative random variable.

This concludes the proof that a strategy as claimed exists. To see tat it is unique, suppose that (A,B) is an arbitrary self-financing strategy satisfying (i) and (ii). Let $V_n = A_nR_n + B_nS_n$ be its value at time n, and define $\tilde S_n = R_n^{-1}S_n$ and $\tilde V_n = R_n^{-1}V_n$, all as before. By the first paragraph of the proof there is a unique probability measure $\tilde p$ making $\tilde S$ into a martingale. Multipyling the self-financing equation (3.2) by R_{n-1}^{-1}, we obtain (for $n\geq 1$)

$$\tilde V_{n-1} = A_n + B_n\tilde S_{n-1} = A_{n-1} + B_{n-1}\tilde S_{n-1}.$$

Replacing $n-1$ by n in the second representation of $\tilde V_{n-1}$ yields $\tilde V_n = A_n + B_n\tilde S_n$. Subtracting from this the first representation of $\tilde V_{n-1}$, we obtain that

$$\tilde V_n - \tilde V_{n-1} = B_n(\tilde S_n - \tilde S_{n-1}).$$

Because $\tilde S$ is a $\tilde p$-martingale and B is predictable, $\tilde V$ is a $\tilde p$-martingale as well. In particular, $\tilde V_n = \tilde{\mathrm{E}}(\tilde V_N|\,\mathcal{F}_n)$ for every $n\leq N$. By (ii) this means that $\tilde V$ is exactly as in the first part of the proof. The rest must also be the same. ■

A strategy as in the preceding theorem is called a "hedging strategy". Its special feature is that given an initial investment of V_0 at time zero (to buy the portfolio (A_0,B_0)) it leads with certainty to a portfolio with value $(S_N-K)^+$ at time N. This is remarkable, because S is a stochastic process. Even though we have limited its increments to two possibilities at every time, this still allows 2^N possible sample paths for the process $S_1,\ldots,S_N$, and each of these has a probability attached to it in the real world. The hedging strategy leads to a portfolio with value $(S_N-K)^+$ at time N, no matter which sample path the process S will follow.

The existence of a hedging strategy and the following economic reasoning shows that the initial value $V_0 = \tilde{\mathrm{E}}R_N^{-1}(S_N-K)^+$ is the only right price for the option.

First, if the option were more expensive than V_0, then nobody would buy it, because it would cost less to buy the portfolio (A_0, B_0) and go through the hedging strategy. This is guaranteed to give the same value $(S_N - K)^+$ at the expiry time, for less money.

On the other hand, if the option could be bought for less money than V_0, then selling a portfolio (A_0, B_0) and buying an option at time 0 would yield some immediate cash. During the term of the option we could next implement the inverse hedging strategy: starting with the portfolio $(-A_0, -B_0)$ at time 0, we reshuffle the portfolio consecutively at times $n = 1, 2, \ldots, N$ to $(-A_n, -B_n)$. This can be done free of investment and at expiry time we would possess both the option and the portfolio $(-A_N, -B_N)$, i.e. our capital would be $-V_N + (S_N - K)^+$, which is zero. Thus after making an initial gain of V_0 minus the option price, we would with certainty break even, no matter the stock price: we would be able to make money without risk. Economists would say that the market would allow for "arbitrage". But in real markets nothing comes free; real markets are "arbitrage-free".

Thus the value $V_0 = \tilde{\mathrm{E}} R_N^{-1}(S_N - K)^+$ is the only "reasonable price". As noted before, this value does not depend on the "true" values of the probabilities $(p_1, \ldots, p_n)$: the expectation must be computed under the "martingale measure" given by $(\tilde{p}_1, \ldots, \tilde{p}_n)$. It depends on the steps $(a_1, b_1, \ldots, a_n, b_n)$, the interest rates r_n, the strike price K and the value S_0 of the stock at time 0. The distribution of S_N under $\tilde{p}$ is supported on at most 2^N values, the corresponding probabilities being (sums of) products over the probabilities $\tilde{p}_i$. We can write out the expectation as a sum, but this is not especially insightful. (Below we compute a limiting value, which is more pleasing.)

The martingale measure given by $\tilde{p}$ is the unique measure (within the model (3.3)) that makes the "discounted stock process" $R_n^{-1} S_n$ into a martingale. It is sometimes referred to as the "risk-free measure". If the interest rate were zero and the stock process a martingale under its true law, then the option price would be exactly the expected value $\tilde{\mathrm{E}}(S_N - K)^+$ of the option at expiry term. In a "risk-free world we can price by expectation".

The discounting of values, the premultiplying with $R_n^{-1} = \prod_{i=1}^n (1+r_i)$, expresses the "time value of money". A capital v at time 0 can be increased to a capital $R_n v$ at time n in a risk-free manner, for instance by putting it in a savings account. Then a capital v that we shall receive at time n in the future is worth only $R_n^{-1} v$ today. For instance, an option is worth $(S_N - K)^+$ at expiry time N, but only $R_N^{-1}(S_N - K)^+$ at time 0. The right price of the option is the expectation of this discounted value "in the risk-free world given by the martingale measure".

The theorem imposes the condition that $a_n < 1 + r_n < b_n$ for all n. This condition is reasonable. If we had a stock at time $n - 1$, worth S_{n-1}, and kept on to it until time n, then it would change in value to either $a_n S_{n-1}$ or $b_n S_{n-1}$. If we sold the stock and invested the money in bonds, then this capital would change in value to $(1+r_n)S_{n-1}$. The inequality $1+r_n<a_n<b_n$ would mean that keeping the stock would always be more advantageous; nobody would buy bonds. On the other hand, the inequality $a_n<b_n<1+r_n$ would mean that investing in bonds would always be more advantageous. In both cases, the market would allow for arbitrage: by exchanging bonds for stock or vice versa, we would have a guaranteed positive profit, no matter the behaviour of the stock. Thus the condition is necessary for the market to be "arbitrage-free".

3.7 EXERCISE. See if the theorem can be extended to the cases that:
(i) the numbers (a_n,b_n) are predictable processes.
(ii) the interest rates r_n form a predictable process.

3.8 EXERCISE. Let $\varepsilon_1,\varepsilon_2,\ldots$ be i.i.d. random variables with the uniform distribution on $\{-1,1\}$ and set $X_n=\sum_{i=1}^n\varepsilon_i$. Suppose that Y is a martingale relative to $\mathcal{F}_n=\sigma(X_1,\ldots,X_n)$. Show that there exists a predictable process C such that $Y=Y_0+C\cdot X$.

We might view the binary stock price model of this section as arising as a time discretization of a continuous time model. Then the model should become more realistic by refining the discretization. Given a fixed time $T>0$, we might consider the binary stock price model for $(S_0,S_1,\ldots,S_N)$ as a discretization on the grid $0,T/N,2T/N,\ldots,T$. Then it would be reasonable to scale the increments (a_n,b_n) and the interest rates r_n, as they will reflect changes on infinitesimal intervals as $N\to\infty$. Given $T>0$ consider the choices

$$\begin{aligned} a_{n,N}&=e^{\mu T/N-\sigma\sqrt{T/N}},\\ b_{n,N}&=e^{\mu T/N+\sigma\sqrt{T/N}},\\ 1+r_{n,N}&=e^{rT/N}. \end{aligned} \tag{3.9}$$

These choices can be motivated from the fact that the resulting sequence of binary tree models converges to the continuous time model that we shall discuss later on. Presently, we can only motivate them by showing that they lead to nice formulas.

Combining (3.3) and (3.9) we obtain that the stock price is given by

$$S_N=S_0\exp\Bigl(\mu T+\sigma\sqrt{T}\frac{(2X_N-N)}{\sqrt{N}}\Bigr),$$

where X_N is the number of times the stock price goes up in the time span $1,2,\ldots,N$.

It is thought that a realistic model for the stock market has jumps up and down with equal probabilities. Then X_N is binomially $(N,\frac12)$-distributed and the "log returns" satisfy

$$\log\frac{S_N}{S_0}=\mu T+\sigma\sqrt{T}\frac{X_N-N/2}{\sqrt{N}/2}\rightsquigarrow N(\mu T,\sigma^2T),$$

by the Central limit theorem. Thus in the limit the log return at time T is log normally distributed with drift μT and variance $\sigma^2 T$.

As we have seen the true distribution of the stock prices is irrelevant for pricing the option. Rather we need to repeat the preceding calculation using the martingale measure $\tilde{p}$. Under this measure X_N is binomially$(N, \tilde{p}_N)$ distributed, for

$$\begin{aligned}\tilde{p}_N &= \frac{e^{rT/N} - e^{\mu T/N - \sigma\sqrt{T/N}}}{e^{\mu T/N + \sigma\sqrt{T/N}} - e^{\mu T/N - \sigma\sqrt{T/N}}} \\ &= \tfrac{1}{2} - \tfrac{1}{2}\sqrt{\frac{T}{N}}\Big(\frac{\mu + \frac{1}{2}\sigma^2 - r}{\sigma}\Big) + O\Big(\frac{1}{N}\Big),\end{aligned}$$

by a Taylor expansion. Then $\tilde{p}_N(1 - \tilde{p}_N) \to 1/4$ and

$$\begin{aligned}\log\frac{S_N}{S_0} &= \mu T + \sigma\sqrt{T}\Big(\frac{X_N - N\tilde{p}_N}{\sqrt{N}/2} - \sqrt{T}\Big(\frac{\mu + \frac{1}{2}\sigma^2 - r}{\sigma}\Big)\Big) + O\Big(\frac{1}{\sqrt{N}}\Big) \\ &\rightsquigarrow N\big((r - \tfrac{1}{2}\sigma^2)T, \sigma^2 T\big).\end{aligned}$$

Thus, under the martingale measure, in the limit the stock at time T is log normally distributed with drift $(r - \frac{1}{2}\sigma^2)T$ and variance $\sigma^2 T$.

Evaluating the (limiting) option price is now a matter of straightforward integration. For an option with expiry time T and strike price K it is the expectation of $e^{-rT}(S_T - K)^+$, where $\log(S_T/S_0)$ possesses the log normal distribution with parameters $(r - \frac{1}{2}\sigma^2)T$ and variance $\sigma^2 T$. This can be computed to be

$$S_0\Phi\Big(\frac{\log(S_0/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\Big) - Ke^{-rT}\Phi\Big(\frac{\log(S_0/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\Big).$$

This is the formula found by Black and Scholes in 1973 using a continuous time model. We shall recover it later in a continuous time set-up.

4
Continuous Time Martingales

In this chapter we extend the theory for discrete time martingales to the continuous time setting. Besides much similarity there is the important difference of dealing with uncountably many random variables, which is solved by considering martingales with cadlag sample paths.

4.1 Stochastic Processes

A *stochastic process* in continuous time is a collection $X = \{X_t: t \geq 0\}$ of random variables indexed by the "time" parameter $t \in [0, \infty)$ and defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Occasionally we work with the extended time set $[0, \infty]$ and have an additional random variable X_∞.

The *finite-dimensional marginals* of a process X are the random vectors $(X_{t_1}, \ldots, X_{t_k})$, for $t_1, \ldots, t_k$ ranging over the time set and $k \in \mathbb{N}$, and the *marginal distributions* of X are the distributions of these vectors. The maps $t \mapsto X_t(\omega)$, for $\omega \in \Omega$, are called *sample paths.* Unless stated otherwise the variables X_t will be understood to be real-valued, but the definitions apply equally well to vector-valued variables.

Two processes X and Y defined on the same probability space are *equivalent* or each other's *modification* if $(X_{t_1}, \ldots, X_{t_k}) = (Y_{t_1}, \ldots, Y_{t_k})$ almost surely. They are *indistinguishable* if $\mathbb{P}(X_t = Y_t, \forall t) = 1$. Both concepts express that X and Y are the "same", but indistinguishability is quite a bit stronger in general, because we are working with an uncountable set of random variables. However, if the sample paths of X and Y are determined by the values on a fixed countable set of time points, then the concepts agree. This is the case, for instance, if the sample paths are continuous, or more generally left- or right continuous. Most of the stochastic processes that we shall be concerned with possess this property. In particular, we often consider *cadlag* processes (from "continu à droite, limite à gauche"): processes with sample paths that are right-continuous and have limits from the left at every point $t > 0$. If X is a left- or right-continuous process, then

$$X_{t-} = \lim_{s \uparrow t, s<t} X_s, \qquad \text{and} \qquad X_{t+} = \lim_{s \downarrow t, s>t} X_s$$

define left- and right-continuous processes. These are denoted by X_- and X_+ and referred to as the *left- or right-continuous version* of X. The difference $\Delta X := X_+ - X_-$ is the *jump process* of X. The variable X_{0-} can only be defined by convention; it will usually be taken equal to 0, giving a jump $\Delta X_0 = X_0$ at time 0.

A *filtration* in continuous time is a collection $\{\mathcal{F}_t\}_{t\geq 0}$ of sub σ-fields of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s \leq t$. A typical example is the *natural filtration* $\mathcal{F}_t = \sigma(X_s : s \leq t)$ generated by a stochastic process X. A stochastic process X is *adapted* to a filtration $\{\mathcal{F}_t\}$ if X_t is $\mathcal{F}_t$-measurable for every t. The natural filtration is the smallest filtration to which X is adapted. We define $\mathcal{F}_\infty = \sigma(\mathcal{F}_t : t \geq 0)$. As in the discrete time case, we call a probability space equipped with a filtration a *filtered probability space* or a "stochastic basis". We denote it by $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, where it should be clear from the notation or the context that t is a continuous parameter.

Throughout, without further mention, we assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is *complete*. This means that every subset of a null set (a null set being a set $F \in \mathcal{F}$ with $\mathbb{P}(F) = 0$) is contained in $\mathcal{F}$ (and hence is also a null set). This is not a very restrictive assumption, because we can always extend a given σ-field and probability measure to make it complete. (This will make a difference only if we would want to work with more than one probability measure at the same time.)

We also always assume that our filtration satisfies the *usual conditions*: for all $t \geq 0$:

(i) (completeness): $\mathcal{F}_t$ contains all null sets.
(ii) (right continuity): $\mathcal{F}_t = \cap_{s>t} \mathcal{F}_s$.

The first condition can be ensured by completing a given filtration: replacing a given $\mathcal{F}_t$ by the σ-field generated by $\mathcal{F}_t$ and all null sets. The second condition is more technical, but turns out to be important for certain arguments. Fortunately, the (completions of the) natural filtrations of the most important processes are automatically right continuous. Furthermore, if a given filtration is not right continuous, then we might replace it by the the filtration $\cap_{s>t} \mathcal{F}_s$, which can be seen to be right-continuous.

Warning. The natural filtration of a right-continuous process is not necessarily right continuous.

Warning. When completing a filtration we add all null sets in $(\Omega, \mathcal{F}, \mathbb{P})$ to every $\mathcal{F}_t$. This gives a bigger filtration than completing the space $(\Omega, \mathcal{F}_t, \mathbb{P})$ for every $t \geq 0$ separately.

4.1 EXERCISE **(Completion)**. Given an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\tilde{\mathcal{F}}$ be the collection of all sets $F \cup N$ for F ranging over

$\mathcal{F}$ and N ranging over all subsets of null sets, and define $\tilde{\mathbb{P}}(F \cup N) = \mathbb{P}(F)$. Show that $(\Omega, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ is well defined and a probability space.

* **4.2** EXERCISE. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $\mathcal{F}_0 \subset \mathcal{F}$ a sub σ-field. Show that the σ-field generated by $\mathcal{F}_0$ and the null sets of $(\Omega, \mathcal{F}, \mathbb{P})$ is the collection of all $F \in \mathcal{F}$ such that there exists $F_0 \in \mathcal{F}_0$ with $\mathbb{P}(F \triangle F_0) = 0$; equivalently, all $F \in \mathcal{F}$ such that there exists $F_0 \in \mathcal{F}_0$ and null sets N, N' with $F_0 - N \subset F \subset F_0 \cup N'$.

* **4.3** EXERCISE. Show that the completion of a right-continuous filtration is right continuous.

* **4.4** EXERCISE. Show that the natural filtration of the Poisson process is right continuous. (More generally, this is true for any counting process.)

4.2 Martingales

The definition of a martingale in continuous time is an obvious generalization of the discrete time case. We say that a process X is *integrable* if $\mathrm{E}|X_t| < \infty$ for every t.

4.5 Definition. *An adapted, integrable stochastic process X on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is a*
(i) martingale if $\mathrm{E}(X_t |\, \mathcal{F}_s) = X_s$ a.s. *for all* $s \le t$.
(ii) submartingale if $\mathrm{E}(X_t |\, \mathcal{F}_s) \ge X_s$ a.s. *for all* $s \le t$.
(ii) supermartingale if $\mathrm{E}(X_t |\, \mathcal{F}_s) \le X_s$ a.s. *for all* $s \le t$.

The (sub/super) martingales that we shall be interested in are cadlag processes. It is relatively straightforward to extend results for discrete time martingales to these, because given a (sub/super) martingale X:
(i) If $0 \le t_1 < t_2 < \cdots$, then $Y_n = X_{t_n}$ defines a (sub/super) martingale relative to the filtration $\mathcal{G}_n = \mathcal{F}_{t_n}$.
(ii) If $t_0 > t_1 > \cdots \ge 0$, then $Y_n = X_{t_n}$ defines a reverse (sub/super) martingale relative to the reverse filtration $\mathcal{G}_n = \mathcal{F}_{t_n}$.
Thus we can apply results on discrete time (sub/super) martingales to the discrete time "skeletons" X_{t_n} formed by restricting X to countable sets of times. If X is cadlag, then this should be enough to study the complete sample paths of X.

The assumption that X is cadlag is not overly strong. The following theorem shows that under the simple condition that the mean function $t \mapsto \mathrm{E}X_t$ is cadlag, a cadlag modification of a (sub/super) martingale always exists. Because we assume our filtrations to be complete, such a modification is automatically adapted. Of course, it also satisfies the (sub/super) martingale property and hence is a (sub/super) martingale relative to the original filtration. Thus rather than with the original (sub/super) martingale we can work with the modification.

We can even allow filtrations that are not necessarily right-continuous. Then we can both replace X by a modification and the filtration by its "right-continuous version" $\mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s$ and still keep the (sub/super) martingale property, provided that X is right continuous in probability. (This is much weaker than right continuous.) In part (ii) of the following theorem, suppose that the filtration is complete, but not necessarily right-continuous.

4.6 Theorem. *Let X be a (sub/super) martingale relative to the complete filtration $\{\mathcal{F}_t\}$.*
(i) If the filtration $\{\mathcal{F}_t\}$ is right continuous and the map $t \mapsto \mathrm{E}X_t$ is right continuous, then there exists a cadlag modification of X.
(ii) If X is right continuous in probability, then there exists a modification of X that is a cadlag (sub/super) martingale relative to the filtration $\{\mathcal{F}_{t+}\}$.

Proof. Assume without loss of generality that X is a supermartingale. Then $X_s \geq \mathrm{E}(X_t|\,\mathcal{F}_s)$ almost surely for every $s \leq t$, whence $X_s^- \leq \mathrm{E}(X_t^-\,|\,\mathcal{F}_s)$ almost surely and hence $\{X_s^-: 0 \leq s \leq t\}$ is uniformly integrable. Combined with the fact that $t \mapsto \mathrm{E}X_t$ is decreasing and hence bounded on compacts, if follows that $\mathrm{E}|X_t|$ is bounded on compacts.

For fixed T and every $a < b$, define the event

$$F_{a,b} = \Big\{\omega: \exists t \in [0,T): \liminf_{s\uparrow\uparrow t, s\in\mathbb{Q}} X_s(\omega) < a < b < \limsup_{s\uparrow\uparrow t, s\in\mathbb{Q}} X_s(\omega), \quad \text{or} \quad \liminf_{s\downarrow\downarrow t, s\in\mathbb{Q}} X_s(\omega) < a < b < \limsup_{s\downarrow\downarrow t, s\in\mathbb{Q}} X_s(\omega)\Big\}$$

(The symbol $s \uparrow\uparrow t$ denotes a limit as $s \uparrow t$ with s restricted to $s < t$.) Let $\mathbb{Q} \cap [0,T) = \{t_1, t_2, \ldots\}$ and let $U_n[a,b]$ be the number of upcrossings of $[a,b]$ by the process $X_{t_1}, \ldots, X_{t_n}$ put in its natural time order. If $\omega \in F_{a,b}$, then $U_n[a,b] \uparrow \infty$. However, by Doob's upcrossings lemma $\mathrm{E}U_n[a,b] < \sup_{0\leq t\leq T} \mathrm{E}|X_t| + |a|$. We conclude that $\mathbb{P}(F_{a,b}) = 0$ for every $a < b$ and hence the left and right limits

$$X_{t-} = \lim_{s\uparrow\uparrow t, s\in\mathbb{Q}} X_s, \qquad X_{t+} = \lim_{s\downarrow\downarrow t, s\in\mathbb{Q}} X_s$$

exist for every $t \in [0,T)$, almost surely. If we define these processes to be zero whenever one of the limits does not exist, then X_{t+} is $\mathcal{F}_{t+}$-adapted. Moreover, from the definitions X_{t+} can be seen to be right-continuous with left limits equal to X_{t-}. By Fatou's lemma X_{t+} is integrable.

We can repeat this for a sequence $T_n \uparrow \infty$ to show that the limits X_{t-} and X_{t+} exist for every $t \in [0, \infty)$, almost surely. Setting X_{t+} equal to zero on the exceptional null set, we obtain a cadlag process that is adapted to $\mathcal{F}_{t+}$.

By the supermartingale property $\mathrm{E}X_s 1_F \geq \mathrm{E}X_t 1_F$ for every $F \in \mathcal{F}_s$ and $s \leq t$. Given a sequence of rational numbers $t_n \downarrow\downarrow t$, the sequence $\{X_{t_n}\}$ is a reverse super martingale. Because $\mathrm{E}X_{t_n}$ is bounded above, the sequence is uniformly integrable and hence $X_{t_n} \to X_{t+}$ both almost surely (by construction) and in mean. We conclude that $\mathrm{E}X_s 1_F \geq \mathrm{E}X_{t+} 1_F$ for every $F \in \mathcal{F}_s$ and $s \leq t$. Applying this for every $s = s_n$ and s_n a sequence of rational numbers decreasing to some fixed s, we find that $\mathrm{E}X_{s+} 1_F \geq \mathrm{E}X_{t+} 1_F$ for every $F \in \mathcal{F}_{s+} = \cap_n \mathcal{F}_{s_n}$ and $s < t$. Thus $\{X_{t+}: t \geq 0\}$ is a supermartingale relative to $\mathcal{F}_{t+}$.

Applying the first half of the argument of the preceding paragraph with $s = t$ we see that $\mathrm{E}X_t 1_F \geq \mathrm{E}X_{t+} 1_F$ for every $F \in \mathcal{F}_t$. If $\mathcal{F}_{t+} = \mathcal{F}_t$, then $X_t - X_{t+}$ is $\mathcal{F}_t$-measurable and we conclude that $X_t - X_{t+} \geq 0$ almost surely. If, moreover, $t \mapsto \mathrm{E}X_t$ is right continuous, then $\mathrm{E}X_t = \lim_{n\to\infty} \mathrm{E}X_{t_n} = \mathrm{E}X_{t+}$, because $X_{t_n} \to X_{t+}$ in mean. Combined this shows that $X_t = X_{t+}$ almost surely, so that X_{t+} is a modification of X. This concludes the proof of (i).

To prove (ii) we recall that X_{t+} is the limit in mean of a sequence X_{t_n} for $t_n \downarrow\downarrow t$. If X is right continuous in probability, then $X_{t_n} \to X_t$ in probability. Because the limits in mean and in probability must agree almost surely, it follows that $X_t = X_{t+}$ almost surely. ∎

In particular, every martingale (relative to a "usual filtration") possesses a cadlag modification, because the mean function of a martingale is constant and hence certainly continuous.

4.7 Example. If for a given filtration $\{\mathcal{F}_t\}$ and integrable random variable ξ we "define" $X_t = \mathrm{E}(\xi|\, \mathcal{F}_t)$, then in fact X_t is only determined up to a null set, for every t. The union of these null sets may have positive probability and hence we have not defined the process X yet. Any choice of the conditional expectations X_t yields a martingale X. By the preceding theorem there is a choice such that X is cadlag. □

4.8 EXERCISE. Given a standard Poisson process $\{N_t: t \geq 0\}$, let $\mathcal{F}_t$ be the completion of the natural filtration $\sigma(N_s: s \leq t)$. (This can be proved to be right continuous.) Show that:

(i) The process N_t is a submartingale.
(ii) The process $N_t - t$ is a martingale.
(iii) The process $(N_t - t)^2 - t$ is a martingale.

4.9 EXERCISE. Show that every cadlag supermartingale is right continuous in mean. [Hint: use reverse supermartingale convergence, as in the proof of Theorem 4.6.]

4.3 Martingale Convergence

The martingale convergence theorems for discrete time martingales extend without surprises to the continuous time situation.

4.10 Theorem. *If X is a uniformly integrable, cadlag (sub/super) martingale, then there exists an integrable random variable X_∞ such that $X_t \to X_\infty$ almost surely and in L_1 as $t \to \infty$.*
(i) If X is a martingale, then $X_t = \mathrm{E}(X_\infty | \mathcal{F}_t)$ a.s. for all $t \geq 0$.
(ii) If X is a submartingale, then $X_t \leq \mathrm{E}(X_\infty | \mathcal{F}_t)$ a.s. for $t \geq 0$.
Furthermore, if X is L_p-bounded for some $p > 1$, then $X_t \to X_\infty$ also in L_p.

Proof. In view of Theorems 2.23 and 2.25 every sequence X_{t_n} for $t_1 < t_2 < \cdots \to \infty$ converges almost surely, in L_1 or in L_p to a limit X_∞. Then we must have that $X_t \to X_\infty$ in L_1 or in L_p as $t \to \infty$. Assertions (i) and (ii) follow from Theorem 2.23 as well.

The almost sure convergence of X_t as $t \to \infty$ requires an additional argument, as the null set on which a sequence X_{t_n} as in the preceding paragraph may not converge may depend on the sequence $\{t_n\}$. In this part of the proof we use the fact that X is cadlag. As in the proof of Theorem 2.21 it suffices to show that for every fixed numbers $a < b$ the event

$$F_{a,b} = \Bigl\{\omega: \liminf_{t\to\infty} X_t(\omega) < a < b < \limsup_{t\to\infty} X_t(\omega)\Bigr\}$$

is a null set. Assume that X is a supermartingale and for given $t_1, \ldots, t_n$ let $U_n[a,b]$ be the number of upcrossings of $[a,b]$ by the process $X_{t_1}, \ldots, X_{t_n}$ put in its natural time order. By Doob's upcrossing's inequality, Lemma 2.19,

$$(b-a)\mathrm{E}U_n[a,b] \leq \sup_t \mathrm{E}|X_t| + |a|.$$

If we let $\mathbb{Q} = \{t_1, t_2, \ldots\}$, then $U_n[a,b] \uparrow \infty$ on $F_{a,b}$, in view of the right-continuity of X. We conclude that $\mathbb{P}(F_{a,b}) = 0$. ∎

4.4 Stopping

The main aim of this section is to show that a stopped martingale is a martingale, also in continuous time, and to extend the optional stopping theorem to continuous time.

4.11 Definition. *A random variable* $T: \Omega \to [0, \infty]$ *is a stopping time if* $\{T \le t\} \in \mathcal{F}_t$ *for every* $t \ge 0$.

Warning. Some authors use the term *optional time* instead of stopping time. Some authors define an optional time by the requirement that $\{T < t\} \in \mathcal{F}_t$ for every $t \ge 0$. This can make a difference if the filtration is not right-continuous.

4.12 EXERCISE. Show that $T: \Omega \to [0, \infty]$ is a *stopping time* if and only if $\{T < t\} \in \mathcal{F}_t$ for every $t \ge 0$. (Assume that the filtration is right-continuous.)

4.13 Definition. *The* σ*-field* $\mathcal{F}_T$ *is defined as the collection of all* $F \subset \Omega$ *such that* $F \cap \{T \le t\} \in \mathcal{F}_t$ *for all* $t \in [0, \infty]$. *(This includes* $t = \infty$*, where* $\mathcal{F}_\infty = \sigma(\mathcal{F}_t: t \ge 0)$*.)*

The collection $\mathcal{F}_T$ is indeed a σ-field, contained in $\mathcal{F}_\infty \subset \mathcal{F}$, and $\mathcal{F}_T = \mathcal{F}_t$ if $T \equiv t$. Lemma 2.41 on comparing the σ-fields $\mathcal{F}_S$ and $\mathcal{F}_T$ also remains valid as stated. The proofs are identical to the proofs in discrete time. However, in the continuous time case it would not do to consider events of the type $\{T = t\}$ only. We also need to be a little more careful with the definition of stopped processes, as the measurability is not automatic. The stopped process X^T and the variable X_T are defined exactly as before:

$$(X^T)_t(\omega) = X_{T(\omega) \wedge t}(\omega), \qquad X_T(\omega) = X_{T(\omega)}(\omega).$$

In general these maps are not measurable, but if X is cadlag and adapted, then they are. More generally, it suffices that X is "progressively measurable". To define this concept think of X as the map $X: [0, \infty) \times \Omega \to \mathbb{R}$ given by

$$(t, \omega) \mapsto X_t(\omega).$$

The process X is *measurable* if X is measurable relative to the product σ-field $\mathcal{B}_\infty \times \mathcal{F}$, i.e. if it is "jointly measurable in (t, ω)" relative to the product σ-field. The process X is *progressively measurable* if, for each $t \ge 0$, the restriction $X: [0, t] \times \Omega \to \mathbb{R}$ is measurable relative to the product σ-field $\mathcal{B}_t \times \mathcal{F}_t$. This is somewhat stronger than being adapted.

4.14 EXERCISE. Show that a progressively measurable process is adapted.

4.15 Lemma. *If the process* X *is progressively measurable and* T *is a stopping time, then:*
(i) X^T *is progressively measurable (and hence adapted).*
(ii) X_T *is* $\mathcal{F}_T$*-measurable (and hence a random variable).*

(In (ii) it is assumed that X_∞ is defined and $\mathcal{F}_\infty$-measurable if T assumes the value ∞.)

Proof. For each t the map $T \wedge t: \Omega \to [0, \infty]$ is $\mathcal{F}_t$ measurable, because $\{T \wedge t > s\} = \{T > s\} \in \mathcal{F}_s \subset \mathcal{F}_t$ if $s < t$ and $\{T \wedge t > s\}$ is empty if $s \geq t$. Then the map

$$(s, \omega) \mapsto (s, T(\omega) \wedge t, \omega) \mapsto \big(s \wedge T(\omega), \omega\big),$$
$$[0, t] \times \Omega \to [0, t] \times [0, t] \times \Omega \to [0, t] \times \Omega,$$

is $\mathcal{B}_t \times \mathcal{F}_t - \mathcal{B}_t \times \mathcal{B}_t \times \mathcal{F}_t - \mathcal{B}_t \times \mathcal{F}_t$-measurable. The stopped process X^T as a map on $[0, t] \times \Omega$ is obtained by composing $X: [0, t] \times \Omega \to \mathbb{R}$ to the right side and hence is $\mathcal{B}_t \times \mathcal{F}_t$-measurable, by the chain rule. That a progressively measurable process is adapted is the preceding exercise.

For assertion (ii) we must prove that $\{X_T \in B\} \cap \{T \leq t\} \in \mathcal{F}_t$ for every Borel set B and $t \in [0, \infty]$. The set on the left side can be written as $\{X_{T \wedge t} \in B\} \cap \{T \leq t\}$. For $t < \infty$ this is contained in $\mathcal{F}_t$ by (i) and because T is a stopping time. For $t = \infty$ we note that $\{X_T \in B\} = \cup_t \{X_{T \wedge t} \in B\} \cap \{T \leq t\} \cup \{X_\infty \in B\} \cap \{T = \infty\}$ and this is contained in $\mathcal{F}_\infty$. ∎

4.16 Example (Hitting time). Let X be an adapted, progressively measurable stochastic process, B a Borel set, and define

$$T = \inf\{t \geq 0: X_t \in B\}.$$

(The infimum of the empty set is defined to be ∞.) Then T is a stopping time.

Here $X = (X_1, \ldots, X_d)$ may be vector-valued, where it is assumed that all the coordinate processes X_i are adapted and progressively measurable and B is a Borel set in $\mathbb{R}^d$.

That T is a stopping time is not easy to prove in general, and does rely on our assumption that the filtration satisfies the usual conditions. A proof can be based on the fact that the set $\{T < t\}$ is the projection on Ω of the set $\{(s, \omega): s < t, X_s(\omega) \in B\}$. (The *projection* on Ω of a subset $A \subset T \times \Omega$ of some product space is the set $\{\omega: \exists t > 0: (t, \omega) \in A\}$.) By the progressive measurability of X this set is measurable in the product σ-field $\mathcal{B}_t \times \mathcal{F}_t$. By the projection theorem (this is the hard part), the projection of every product measurable set is measurable in the completion. See Elliott, p50.

Under special assumptions on X and B the proof is more elementary. For instance, suppose that X is continuous and that B is closed. Then, for $t > 0$,

$$\{T \leq t\} = \bigcap_n \bigcup_{s < t, s \in \mathbb{Q}} \{d(X_s, B) < n^{-1}\}.$$

The right side is clearly contained in $\mathcal{F}_t$. Furthermore, by the continuity of X and the closedness of B we have $\{T = 0\} = \{X_0 \in B\}$ and this is contained in $\mathcal{F}_0$.

To establish the preceding display, we note first that the event $\{T = 0\}$ is contained in both sides of the equation. Furthermore, it is easy to see the inclusion of right side in left side; we now prove the inclusion of left in right side. By the definition of T and continuity of X, the function $t \mapsto d(X_t, B)$ must vanish at $t = T$ and be strictly positive on $[0, T)$ if $T > 0$. By continuity this function assumes every value in the interval $[0, d(X_0, B)]$ on the interval $[0, T]$. In particular, for every $n \in \mathbb{N}$ there must be some rational number $s \in (0, T)$ such that $d(X_s, B) < n^{-1}$. □

4.17 EXERCISE. Give a direct proof that $T = \inf\{t: X_t \in B\}$ is a stopping time if B is open and X is right-continuous. [Hint: consider the sets $\{T < t\}$ and use the right-continuity of the filtration.]

4.18 EXERCISE. Let X be a continuous stochastic process with $X_0 = 0$ and $T = \inf\{t \geq 0: |X_t| \geq a\}$ for some $a > 0$. Show that T is a stopping time and that $|X^T| \leq a$.

4.19 Lemma. *If X is adapted and right continuous, then X is progressively measurable. The same is true if X is adapted and left continuous.*

Proof. We give the proof for the case that X is right continuous. For fixed $t \geq 0$, let $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ be a sequence of partitions of $[0, t]$ with mesh widths tending to zero as $n \to \infty$. Define X_n to be the discretization of X equal to $X_{t_i^n}$ on $[t_{i-1}^n, t_i^n)$ and equal to X_t at $\{t\}$. By right continuity of X, $X_s^n(\omega) \to X_s(\omega)$ as $n \to \infty$ for every $(s, \omega) \in [0, t] \times \Omega$. Because a pointwise limit of measurable functions is measurable, it suffices to show that every of the maps $X^n: [0, t] \times \Omega \to \mathbb{R}$ is $\mathcal{B}_t \times \mathcal{F}_t$-measurable. Now $\{X^n \in B\}$ can be written as the union of the sets $[t_{i-1}^n, t_i^n) \times \{\omega: X_{t_i^n}(\omega) \in B\}$ and the set $\{t\} \times \{\omega: X_t(\omega) \in B\}$ and each of these sets is certainly contained in $\mathcal{B}_t \times \mathcal{F}_t$. ■

Exactly as in the discrete time situation a stopped (sub/super) martingale is a (sub/super) martingale, and the (in)equalities defining the (sub/super) martingale property remain valid if the (sub/super) martingale is uniformly integrable and the times are replaces by stopping times. At least if we assume that the (sub/super) martingale is cadlag.

4.20 Theorem (Stopped martingale). *If X is a cadlag (sub/super) martingale and T is a stopping time, then X^T is a (sub/super) martingale.*

Proof. We can assume without loss of generality that X is a submartingale. For $n \in \mathbb{N}$ define T_n to be the upward discretization of T on the grid $0 < 2^{-n} < 22^{-n} < \cdots$; i.e. $T_n = k2^{-n}$ if $T \in [(k-1)2^{-n}, k2^{-n})$ (for $k \in \mathbb{N}$) and $T_n = \infty$ if $T = \infty$. Then $T_n \downarrow T$ as $n \to \infty$ and by right continuity $X_{T_n \wedge t} \to X_{T \wedge t}$ for all t, pointwise on Ω. For fixed $t > 0$ let $k_{n,t}2^{-n}$ be the biggest point $k2^{-n}$ on the grid smaller than or equal to t.

For fixed t the sequence

$$X_0, X_{2^{-n}}, X_{22^{-n}}, \ldots, X_{k_{n,t}2^{-n}}, X_t$$

is a submartingale relative to the filtration

$$\mathcal{F}_0 \subset \mathcal{F}_{2^{-n}} \subset \mathcal{F}_{22^{-n}} \subset \cdots \subset \mathcal{F}_{k_{n,t}2^{-n}} \subset \mathcal{F}_t.$$

Here the indexing by numbers $k2^{-n}$ or t differs from the standard indexing by numbers in $\mathbb{Z}_+$, but the interpretation of "submartingale" should be clear. Because the submartingale has finitely many elements, it is uniformly integrable. (If you wish, you may also think of it as an infinite sequence, by just repeating X_t at the end.)

Both $T_n \wedge t$ and $T_{n-1} \wedge t$ can be viewed as stopping times relative to this filtration. For instance, the first follows from the fact that $\{T_n \leq k2^{-n}\} = \{T < k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$ for every k, and the fact that the minimum of two stopping times is always a stopping time. For T_{n-1} we use the same argument and also note that the grid with mesh width 2^{-n+1} is contained in the grid with mesh width 2^{-n}. Because $T_{n-1} \wedge t \geq T_n \wedge t$, the optional stopping theorem in discrete time, Theorem 2.42, gives $\mathrm{E}(X_{T_{n-1}\wedge t}|\, \mathcal{F}_{T_n \wedge t}) \geq X_{T_n \wedge t}$ almost surely. Furthermore, $\mathrm{E}(X_{T_n\wedge t}|\, \mathcal{F}_0) \geq X_0$ and hence $\mathrm{E}X_{T_n \wedge t} \geq \mathrm{E}X_0$.

This being true for every n it follows that $X_{T_1 \wedge t}, X_{T_2 \wedge t}, \ldots$ is a reverse submartingale relative to the reverse filtration $\mathcal{F}_{T_1 \wedge t} \supset \mathcal{F}_{T_2 \wedge t} \supset \cdots$ with mean bounded below by $\mathrm{E}X_0$. By Lemma 2.34 $\{X_{T_n \wedge t}\}$ is uniformly integrable. Combining this with the first paragraph we see that $X_{T_n \wedge t} \to X_{T \wedge t}$ in L_1, as $n \to \infty$.

For fixed $s < t$ the sequence

$$X_0, X_{2^{-n}}, \ldots, X_{k_{s,n}2^{-n}}, X_s, \ldots, X_{k_{t,n}2^{-n}}, X_t$$

is a submartingale relative to the filtration

$$\mathcal{F}_0 \subset \mathcal{F}_{2^{-n}} \subset \cdots \subset \mathcal{F}_{k_{s,n}2^{-n}} \subset \mathcal{F}_s \subset \cdots \subset \mathcal{F}_{k_{t,n}2^{-n}} \subset \mathcal{F}_t.$$

The variable $T_n \wedge t$ is a stopping time relative to this set-up. By the extension of Theorem 2.13 to submartingales the preceding process stopped at $T_n \wedge t$ is a submartingale relative to the given filtration. This is the process

$$X_0, X_{2^{-n}\wedge T_n}, \ldots, X_{k_{s,n}2^{-n}\wedge T_n}, X_{s\wedge T_n}, \ldots, X_{k_{t,n}2^{-n}\wedge T_n}, X_{t \wedge T_n}.$$

In particular, this gives that

$$\mathrm{E}(X_{T_n \wedge t}|\, \mathcal{F}_s) \geq X_{T_n \wedge s}, \qquad \text{a.s..}$$

As $n \to \infty$ the left and right sides of the display converge in L_1 to $\mathrm{E}(X_{T\wedge t}|\, \mathcal{F}_s)$ and $X_{T \wedge s}$. Because L_1-convergence implies the existence of an almost surely converging subsequence, the inequality is retained in the limit in an almost sure sense. Hence $\mathrm{E}(X_{T\wedge t}|\, \mathcal{F}_s) \geq X_{T \wedge s}$ almost surely. ∎

A uniformly integrable, cadlag (sub/super) martingale X converges in L_1 to a limit X_∞, by Theorem 4.10. This allows to define X_T also if T takes the value ∞.

4.21 Theorem (Optional stopping). *If X is a uniformly integrable, cadlag submartingale and $S \leq T$ are stopping times, then X_S and X_T are integrable and $\mathrm{E}(X_T | \mathcal{F}_S) \geq X_S$ almost surely.*

Proof. Define S_n and T_n to be the discretizations of S and T upwards on the grid $0 < 2^{-n} < 22^{-n} < \cdots$, defined as in the preceding proof. By right continuity $X_{S_n} \to X_S$ and $X_{T_n} \to X_T$ pointwise on Ω. Both S_n and T_n are stopping times relative to the filtration $\mathcal{F}_0 \subset \mathcal{F}_{2^{-n}} \subset \cdots$, and $X_0, X_{2^{-n}}, \ldots$ is a uniformly integrable submartingale relative to this filtration. Because $S_n \leq T_n$ the optional stopping theorem in discrete time, Theorem 2.42, yields that X_{S_n} and X_{T_n} are integrable and $\mathrm{E}(X_{T_n} | \mathcal{F}_{S_n}) \geq X_{S_n}$ almost surely. In other words, for every $F \in \mathcal{F}_{S_n}$,

$$\mathrm{E}X_{T_n} 1_F \geq \mathrm{E}X_{S_n} 1_F.$$

Because $S \leq S_n$ we have $\mathcal{F}_S \subset \mathcal{F}_{S_n}$ and hence the preceding display is true for every $F \in \mathcal{F}_S$. If the sequences X_{S_n} and X_{T_n} are uniformly integrable, then we can take the limit as $n \to \infty$ to find that $\mathrm{E}X_T 1_F \geq \mathrm{E}X_S 1_F$ for every $F \in \mathcal{F}_S$ and the proof is complete.

Both T_{n-1} and T_n are stopping times relative to the filtration $\mathcal{F}_0 \subset \mathcal{F}_{2^{-n}} \subset \cdots$ and $T_n \leq T_{n-1}$. By the optional stopping theorem in discrete time $\mathrm{E}(X_{T_{n-1}} | \mathcal{F}_{T_n}) \geq X_{T_n}$, since X is uniformly integrable. Furthermore, $\mathrm{E}(X_{T_n} | \mathcal{F}_0) \geq X_0$ and hence $\mathrm{E}X_{T_n} \geq \mathrm{E}X_0$. It follows that $\{X_{T_n}\}$ is a reverse submartingale relative to the reverse filtration $\mathcal{F}_{T_1} \supset \mathcal{F}_{T_2} \supset \cdots$ with mean bounded below. Therefore, the sequence $\{X_{T_n}\}$ is uniformly integrable by Lemma 2.34. Of course, the same proof applies to $\{X_{S_n}\}$. ∎

If X is a cadlag, uniformly integrable martingale and $S \leq T$ are stopping times, then $\mathrm{E}(X_T | \mathcal{F}_S) = X_S$, by two applications of the preceding theorem. As a consequence the expectation $\mathrm{E}X_T$ of the stopped process at ∞ is equal to the expectation $\mathrm{E}X_0$ for every stopping time T. This property actually characterizes uniformly integrable martingales.

4.22 Lemma. *Let $X = \{X_t : t \in [0, \infty]\}$ be a cadlag adapted process such that X_T is integrable with $\mathrm{E}X_T = \mathrm{E}X_0$ for every stopping time T. Then X is a uniformly integrable martingale.*

Proof. For a given $F \in \mathcal{F}_t$ define the random variable T to be t on F and to be ∞ otherwise. Then T can be seen to be a stopping time, and

$$\mathrm{E}X_T = \mathrm{E}X_t 1_F + \mathrm{E}X_\infty 1_{F^c},$$
$$\mathrm{E}X_0 = \mathrm{E}X_\infty = \mathrm{E}X_\infty 1_F + \mathrm{E}X_\infty 1_{F^c}.$$

We conclude that $\mathrm{E}X_t 1_F = \mathrm{E}X_\infty 1_F$ for every $F \in \mathcal{F}_t$ and hence $X_t = \mathrm{E}(X_\infty | \mathcal{F}_t)$ almost surely. ∎

4.23 EXERCISE. Suppose that X is a cadlag process such that $X_t = \mathrm{E}(\xi|\,\mathcal{F}_t)$ almost surely, for every t. Show that $X_T = \mathrm{E}(\xi|\,\mathcal{F}_T)$ almost surely for every stopping time T.

4.5 Brownian Motion

Brownian motion is a special stochastic process, which was first introduced as a model for the "Brownian motion" of particles in a gas or fluid, but has a much greater importance, both for applications and theory. It could be thought of as the "standard normal distribution for processes".

4.24 Definition. *A stochastic process B is a (standard) Brownian motion relative to the filtration $\{\mathcal{F}_t\}$ if:*
(i) B is adapted.
(ii) all sample paths are continuous.
(iii) $B_t - B_s$ is independent of $\mathcal{F}_s$ for all $0 \le s \le t$.
(iv) $B_t - B_s$ is $N(0, t-s)$-distributed.
(v) $B_0 = 0$.

The model for the trajectory in $\mathbb{R}^3$ of a particle in a gas is a process (B_t^1, B_t^2, B_t^3) consisting of three independent Brownian motions defined on the same filtered probability space. Property (ii) is natural as a particle cannot jump through space. Property (iii) says that given the path history $\mathcal{F}_s$ the displacement $B_t - B_s$ in the time interval $(s, t]$ does not depend on the past. Property (iv) is the only quantitative property. The normality can be motivated by the usual argument that, even in small time intervals, the displacement should be a sum of many infinitesimal movements, but has some arbitrariness to it. The zero mean indicates that there is no preferred direction. The variance $t-s$ is, up to a constant, a consequence of the other assumptions if we also assume that it may only depend on $t - s$. Property (v) is the main reason for the qualification "standard". If we replace 0 by x, then we obtain a "Brownian motion starting at x".

We automatically have the following properties:
(vi) (independent increments) $B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_k} - B_{t_{k-1}}$ are independent for every $0 \le t_1 < t_2 < \cdots < t_k$.
(vii) $(B_{t_1}, \ldots, B_{t_k})$ is multivariate-normally distributed with mean zero and covariance matrix $\mathrm{cov}(B_{t_i}, B_{t_j}) = t_i \wedge t_j$.

It is certainly not immediately clear that Brownian motion exists, but it does.

4.25 Theorem. *There exists a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable maps $B_t\colon \Omega \to \mathbb{R}$ such that the process B satisfies (i)–(v) relative* *to the completion of the natural filtration generated by B (which is right-continuous).*

There are many different proofs of this theorem, but we omit giving a proof altogether. It is reconforting to know that Brownian motion exists, but, on the other hand, it is perfectly possible to work with it without worrying about its existence.

The theorem asserts that a Brownian motion exists relative to its (completed) natural filtration, whereas the definition allows a general filtration. In fact, there exist many Brownian motions. Not only can we use different probability spaces to carry them, but, more importantly, we may use another than the natural filtration.

Warning. Some authors always use the natural filtration, or its completion. Property (iii) is requires more if $\{\mathcal{F}_t\}$ is a bigger filtration.

Brownian motion is "the" example of a continuous martingale.

4.26 Theorem. *Any Brownian motion is a martingale.*

Proof. Because $B_t - B_s$ is independent of $\mathcal{F}_s$, we have $\mathrm{E}(B_t - B_s | \mathcal{F}_s) = \mathrm{E}(B_t - B_s)$ almost surely, and this is 0. ∎

4.27 EXERCISE. Show that the process $\{B_t^2 - t\}$ is a martingale.

Brownian motion has been studied extensively and possesses many remarkable properties. For instance:

(i) Almost every sample path is nowhere differentiable.

(ii) Almost every sample path has no point of increase. (A point of increase of a function f is a point t that possesses a neighbourhood such $f(s) \leq f(t)$ for $s < t$ and $f(s) \geq f(t)$ for $s > t$ in the neighbourhood.)

(iii) For almost every sample path the set of points of local maximum is countable and dense in $[0, \infty)$. (A point of local maximum of a function f is a point t that possesses a neighbourhood such that on this neighbourhood f is maximal at t.)

(iv) $\limsup_{t\to\infty} B_t/\sqrt{2t \log\log t} = 1$ a.s..

These properties are of little concern in the following. A weaker form of property (i) follows from the following theorem, which is fundamental for the theory of stochastic integration.

4.28 Theorem. *If B is a Brownian motion and $0 < t_0^n < t_n^n < \cdots < t_{k_n}^n = t$ is a sequence of partitions of $[0, t]$ with mesh widths tending to zero, then*

$$\sum_{i=1}^{k_n} (B_{t_i} - B_{t_{i-1}})^2 \xrightarrow{\mathrm{P}} t.$$

Proof. We shall even show convergence in quadratic mean. Because $B_t - B_s$ is $N(0, t-s)$-distributed, the variable $(B_t - B_s)^2$ has mean $t - s$ and variance

$2(t-s)^2$. Therefore, by the independence of the increments and because $t = \sum_i (t_i - t_{i-1})$

$$\mathrm{E}\Big[\sum_{i=1}^{k_n}(B_{t_i} - B_{t_{i-1}})^2 - t\Big]^2 = \sum_{i=1}^{k_n} \mathrm{var}(B_{t_i} - B_{t_{i-1}})^2 = 2\sum_{i=1}^{k_n}(t_i - t_{i-1})^2.$$

The right side is bounded by $2\delta_n \sum_{i=1}^{k_n} |t_i - t_{i-1}| = 2\delta_n t$ for δ_n the mesh width of the partition. Hence it converges to zero. ∎

A consequence of the preceding theorem is that for any sequence of partitions with mesh widths tending to 0

$$\limsup_{n\to\infty} \sum_{i=1}^{k_n} |B_{t_i} - B_{t_{i-1}}| = \infty, \qquad \text{a.s.}.$$

Indeed, if the lim sup were finite on a set of positive probability, then on this set we would have that $\sum_{i=1}^{k_n}(B_{t_i} - B_{t_{i-1}})^2 \to 0$ almost surely, because $\max_i |B_{t_i} - B_{t_{i-1}}| \to 0$ by the (uniform) continuity of the sample paths. This would contradict the convergence in probability to t.

We conclude that the sample paths of Brownian motion are of unbounded variation. In comparison if $f\colon [0,t] \to \mathbb{R}$ is continuously differentiable, then

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \big|f(t_i) - f(t_{i-1})\big| = \int_0^t |f'(s)|\, ds.$$

It is the roughness (or "randomness") of its sample paths that makes Brownian motion interesting and complicated at the same time.

Physicists may even find that Brownian motion is too rough as a model for "Brownian motion". Sometimes this is alleviated by modelling velocity by a Brownian motion, rather than location.

4.6 Local Martingales

In the definition of a stochastic integral L_2-martingales play a special role. A Brownian motion is L_2-bounded if restricted to a compact time interval, but not if the time set is $[0,\infty)$. Other martingales may not even be square-integrable.

Localization is a method to extend definitions or properties from processes that are well-behaved, often in the sense of integrability properties, to more general processes. The simplest form is to consider a process X in turn on the intervals $[0,T_1], [0,T_2], \ldots$ for numbers $T_1 \le T_2 \le \cdots$ increasing to infinity. Equivalently, we consider the sequence of stopped processes X^{T_n}. More flexible is to use stopping times T_n for this purpose. The following definition of a "local martingale" is an example.

4.29 Definition. *An adapted process X is a local (sub/super) martingale in L_p if there exists a sequence of stopping times $0 \leq T_1 \leq T_2 \leq \cdots$ with $T_n \uparrow \infty$ almost surely such that X^{T_n} is a (sub/super) martingale in L_p for every n.*

In the case that $p = 1$ we drop the "in L_1" and speak simply of a local (sub/super) martingale. Rather than "martingale in L_p" we also speak of "L_p-martingale". Other properties of processes can be localized in a similar way, yielding for instance, "locally bounded processes" or "locally L_2-bounded martingales". The appropriate definitions will be given when needed, but should be easy to guess. (Some of these classes actually are identical. See the exercises at the end of this section.)

The sequence of stopping times $0 \leq T_n \uparrow \infty$ is called a *localizing sequence*. Such a sequence is certainly not unique. For instance, we can always choose $T_n \leq n$ by truncating T_n at n.

Any martingale is a local martingale, for we can simply choose the localizing sequence equal to $T_n \equiv \infty$. Conversely, a "sufficiently integrable" local (sub/super) martingale is a (sub/super) martingale, as we now argue. If X is a local martingale with localizing sequence T_n, then $X_t^{T_n} \to X_t$ almost surely for every t. If this convergence also happens in L_1, then the martingale properties of X^{T_n} carries over onto X and X itself is a martingale.

4.30 EXERCISE. Show that a dominated local martingale is a martingale.

Warning. A local martingale that is bounded in L_2 need not be a martingale. A fortiori, a uniformly integrable local martingale need not be a martingale. See Chung and Williams, pp20–21, for a counterexample. Remember that we say a process M is bounded in L_2 if $\sup_t \mathrm{E} M_t^2 < \infty$. For a cadlag martingale, this is equivalent to $\mathrm{E} \sup_t M_t^2 < \infty$, but not for a local martingale!

Warning. Some authors define a local (sub/super) martingale in L_p by the requirement that the process $X - X_0$ can be localized as in the preceding definition. If $X_0 \in L_p$, this does not make a difference, but otherwise it may. Because $(X^{T_n})_0 = X_0$ our definition requires that the initial value X_0 of a local (sub/super) martingale in L_p be in L_p.

We shall mostly encounter the localization procedure as a means to reduce a proof to bounded stochastic processes. If X is adapted and continuous, then

$$T_n = \inf\{t: |X_t| \geq n\} \tag{4.31}$$

is a stopping time. On the set $T_n > 0$ we have $|X^{T_n}| \leq n$. If X is a continuousm local martingale, then we can always use this sequence as the localizing sequence.

4.32 Lemma. *If X is a continuous, local martingale, then T_n given by (4.31) defines a localizing sequence. Furthermore, X is automatically a local L_p-martingale for every $p \geq 1$ such that $X_0 \in L_p$.*

Proof. If $T_n = 0$, then $(X^{T_n})_t = X_0$ for all $t \geq 0$. On the other hand, if $T_n > 0$, then $|X_t| < n$ for $t < T_n$ and there exists $t_m \downarrow T_n$ with $|X_{t_m}| \geq n$. By continuity of X it follows that $|X_{T_n}| = n$ in this case. Consequently, $|X^{T_n}| \leq |X_0| \vee n$ and hence X^{T_n} is even dominated by an element of L_p if $X_0 \in L_p$. It suffices to prove that T_n is a localizing sequence.

Suppose that S_m is a sequence of stopping times with $S_m \to \infty$ as $m \to \infty$ and such that X^{S_m} is a martingale for every m. Then $X^{S_m \wedge T_n} = (X^{S_m})^{T_n}$ is a martingale for each m and n, by Theorem 4.20. For every fixed n we have $|X^{S_m \wedge T_n}| \leq |X_0| \vee n$ for every m, and $X^{S_m \wedge T_n} \to X^{T_n}$ almost surely as $m \to \infty$. Because $X_0 = (X^{S_m})_0$ and X^{S_m} is a martingale by assumption, it follows that X_0 is integrable. Thus $X_{S_m \wedge T_n \wedge t} \to X_{T_n \wedge t}$ in L_1 as $m \to \infty$, for every $t \geq 0$. Upon taking limits on both sides of the martingale equality $\mathrm{E}(X_{S_m \wedge T_n \wedge t} | \mathcal{F}_s) = X_{S_m \wedge T_n \wedge s}$ of $X^{S_m \wedge T_n}$ we see that X^{T_n} is a martingale for every n.

Because X is continuous, its sample paths are bounded on compacta. This implies that $T_n \to \infty$ as $n \to \infty$. ∎

4.33 EXERCISE. Show that a local martingale X is a uniformly integrable martingale if and only if the set $\{X_T : T \text{ finite stopping time}\}$ is uniformly integrable. (A process with this property is said to be *of class D*.)

4.34 EXERCISE. Show that a local L_1-martingale X is also a *locally uniformly integrable martingale*, meaning that there exists a sequence of stopping times $0 \leq T_n \uparrow \infty$ such that X^{T_n} is a uniformly integrable martingale.

4.35 EXERCISE. Show that (for $p > 1$) a local L_p-martingale X is *locally bounded in L_p*, meaning that there exists a sequence of stopping times $0 \leq T_n \uparrow \infty$ such that X^{T_n} is a martingale that is bounded in L_p, for every n.

4.36 EXERCISE. Show that a local martingale that is bounded below is a supermartingale. [Hint: use the conditional Fatou lemma.]

4.7 Maximal Inequalities

The maximal inequalities for discrete time (sub/super) martingales carry over to continuous time cadlag (sub/super) martingales, without surprises. The essential observation is that for a cadlag process a supremum $\sup_t X_t$

over $t \geq 0$ is equal to the supremum over a countable dense subset of $[0, \infty)$, and a countable supremum is the (increasing) limit of finite maxima.

4.37 Lemma. *If X is a nonnegative, cadlag submartingale, then for any $x \geq 0$ and every $t \geq 0$,*

$$x\mathbb{P}\Bigl(\sup_{0\leq s\leq t} X_s > x\Bigr) \leq \mathrm{E}X_t 1_{\sup_{0\leq s\leq t} X_t \geq x} \leq \mathrm{E}X_t.$$

4.38 Corollary. *If X is a nonnegative, cadlag submartingale, then for any $p > 1$ and $p^{-1} + q^{-1} = 1$, and every $t \geq 0$,*

$$\Bigl\|\sup_{0\leq s\leq t} X_s\Bigr\|_p \leq q\|X_t\|_p.$$

If X is bounded in $L_p(\Omega, \mathcal{F}, \mathbb{P})$, then $X_t \to X_\infty$ in L_p for some random variable X_∞ and

$$\Bigl\|\sup_{t\geq 0} X_t\Bigr\|_p \leq q\|X_\infty\|_p = q\sup_{t\geq 0}\|X_t\|_p.$$

The preceding results apply in particular to the absolute value of a martingale. For instance, for any martingale X,

$$\Bigl\|\sup_t |X_t|\Bigr\|_2 \leq 2\sup_t \|X_t\|_2. \qquad (4.39)$$

5
Stochastic Integrals

In this chapter we define integrals $\int X\,dM$ for pairs of a "predictable" process X and a martingale M. The main challenge is that the sample paths of many martingales of interest are of infinite variation. We have seen this for Brownian motion in Section 4.5; this property is in fact shared by all martingales with continuous sample paths. For this reason the integral $\int X\,dM$ cannot be defined using ordinary measure theory. Rather than defining it "pathwise for every ω", we define it as a random variable through an L_2-isometry.

In general the predictability of the integrand (defined in Section 5.1) is important, but in special cases, including the one of Brownian motion, the definition can be extended to more general processes.

The definition is carried out in several steps, each time including more general processes X or M. After completing the definition we close the chapter with Itô's formula, which is the stochastic version of the chain rule from calculus, and gives a method to manipulate stochastic integrals.

Throughout the chapter $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is a given filtered probability space.

5.1 Predictable Sets and Processes

The product space $[0,\infty) \times \Omega$ is naturally equipped with the product σ-field $\mathcal{B}_\infty \times \mathcal{F}$. Several sub σ-fields play an important role in the definition of stochastic integrals.

A stochastic process X can be viewed as the map $X\colon [0,\infty) \times \Omega \to \mathbb{R}$ given by $(t,\omega) \mapsto X_t(\omega)$. We define σ-fields by requiring that certain types of processes must be measurable as maps on $[0,\infty) \times \Omega$.

5.1 Definition. *The predictable σ-field $\mathcal{P}$ is the σ-field on $[0,\infty)\times\Omega$ generated by the left-continuous, adapted processes $X\colon[0,\infty)\times\Omega\to\mathbb{R}$. (It can be shown that the same σ-field is generated by all continuous, adapted processes $X\colon[0,\infty)\times\Omega\to\mathbb{R}$.)*

5.2 Definition. *The optional σ-field $\mathcal{O}$ is the σ-field on $[0,\infty)\times\Omega$ generated by the cadlag, adapted processes $X\colon[0,\infty)\times\Omega\to\mathbb{R}$.*

5.3 Definition. *The progressive σ-field $\mathcal{M}$ is the σ-field on $[0,\infty)\times\Omega$ generated by the progressively measurable processes $X\colon[0,\infty)\times\Omega\to\mathbb{R}$.*

We call a process $X\colon[0,\infty)\times\Omega\to\mathbb{R}$ predictable or optional if it is measurable relative to the predictable or optional σ-field.

It can be shown that the three σ-fields are nested in the order of the definitions:

$$\mathcal{P}\subset\mathcal{O}\subset\mathcal{M}\subset\mathcal{B}_\infty\times\mathcal{F}.$$

The predictable σ-field is the most important one to us, as it defines the processes X that are permitted as integrands in the stochastic integrals. Because, obviously, left-continuous, adapted processes are predictable, these are "good" integrands. In particular, continuous, adapted processes.

Warning. Not every predictable process is left-continuous.

The term "predictable" as applied to left-continuous processes expresses the fact that the value of a left-continuous process at a time t is (approximately) "known" just before time t. In contrast, a general process may jump and hence be "unpredictable" from its values in the past. However, it is not true that a predictable process cannot have jumps. The following exercise illustrates this.

5.4 EXERCISE. Show that any measurable function $f\colon[0,\infty)\to\mathbb{R}$ defines a predictable process $(t,\omega)\mapsto f(t)$. "Deterministic processes are predictable".

There are several other ways to describe the various σ-fields. We give some of these as a series of lemmas. For proofs, see Chung and Williams p25–30 and p57–63.

5.5 Lemma. *The predictable σ-field is generated by the collection of all subsets of $[0,\infty)\times\Omega$ of the form*

$$\{0\}\times F_0,\quad F_0\in\mathcal{F}_0,\qquad \textit{and}\qquad (s,t]\times F_s,\quad F_s\in\mathcal{F}_s, s<t.$$

We refer to the sets in Lemma 5.5 as *predictable rectangles.*

Given two functions $S,T\colon\Omega\to[0,\infty]$, the subset of $[0,\infty)\times\Omega$ given by

$$[S,T]=\bigl\{(t,\omega)\in[0,\infty)\times\Omega\colon S(\omega)\le t\le T(\omega)\bigr\}$$

is a *stochastic interval*. In a similar way, we define the stochastic intervals $(S,T]$, $[S,T)$ and (S,T). The set $[T] = [T,T]$ is the *graph* of T. By definition these are subsets of $[0,\infty) \times \Omega$, even though the right endpoint T may assume the value ∞. If S and/or T is degenerate, then we use the same notation, yielding, for instance, $[0,T]$ or $(s,t]$.

Warning. This causes some confusion, because notation such as $(s,t]$ may now denote a subset of $[0,\infty]$ or of $[0,\infty) \times \Omega$.

We are especially interested in stochastic intervals whose boundaries are stopping times. These intervals may be used to describe the various σ-fields, where we need to single out a special type of stopping time.

5.6 Definition. *A stopping time $T: \Omega \to [0,\infty]$ is predictable if there exists a sequence T_n of stopping times such that $0 \leq T_n \uparrow T$ and such that $T_n < T$ for every n on the set $\{T > 0\}$.*

A sequence of stopping times T_n as in the definition is called an *announcing sequence*. It “predicts” that we are about to stop. The phrase “predictable stopping time” is often abbreviated to “predictable time”.

Warning. A hitting time of a predictable process is not necessarily a predictable time.

5.7 Lemma. *Each of the following collections of sets generates the predictable σ-field.*

(i) All stochastic intervals $[T,\infty)$, where T is a predictable stopping time.

(ii) All stochastic intervals $[S,T)$, where S is a predictable stopping time and T is a stopping time.

(iii) All sets $\{0\} \times F_0$, $F_0 \in \mathcal{F}_0$ and all stochastic intervals $(S,T]$, where S and T are stopping times.

(iv) All sets $\{0\} \times F_0$, $F_0 \in \mathcal{F}_0$ and all stochastic intervals $[0,T]$, where T is a stopping time.

Furthermore, a stopping time T is predictable if and only if its graph $[T]$ is a predictable set.

5.8 Lemma. *Each of the following collections of sets generates the optional σ-field.*

(i) All stochastic intervals $[T,\infty)$, where T is a stopping time.

(ii) All stochastic intervals $[S,T]$, $[S,T)$, $(S,T]$, (S,T), where S and T are stopping times.

5.9 Example. If T is a stopping time and $c > 0$, then $T+c$ is a predictable stopping time. An announcing sequence is the sequence $T + c_n$ for $c_n < c$ numbers with $0 \leq c_n \uparrow c$. Thus there are many predictable stopping times. □

5.10 Example. Let X be an adapted process with continuous sample paths and B be a closed set. Then $T = \inf\{t \geq 0: X_t \in B\}$ is a predictable time. An announcing sequence is $T_n = \inf\{t \geq 0: d(X_t, B) < n^{-1}\} \wedge n$. The proof of this is more or less given already in Example 4.16. (We take the minimum with n to ensure that $T_n < T$ on the set $T = \infty$.) □

5.11 Example. It can be shown that any stopping time relative to the natural filtration of a Brownian motion is predictable. See Chung and Williams, p30–31. □

5.12 Example. The left-continuous version of an adapted cadlag process is predictable, by left continuity. Then so is the jump process ΔX of a predictable process X. It can be shown that this jump process is nonzero only on the union $\cup_n [T_n]$ of the graphs of countably many predictable times T_n. (These predictable times are said to "exhaust the jumps of X".) Thus a predictable process has "predictable jumps". □

5.13 Example. Every measurable process that is indistinguishable from a predictable process is predictable. This means that we do not need to "worry about null sets" too much.

Our assumption that the filtered probability space satisfies the usual conditions is essential for this to be true.

To verify the claim it suffices to show that every measurable process X that is indistinguishable from the zero process (an *evanescent process*) is predictable. By the completeness of the filtration a process of the form $1_{(u,v]\times N}$ is left-continuous and adapted for every null set N, and hence predictable. The product σ-field $\mathcal{B}_\infty \times \mathcal{F}$ is generated by the sets of the form $(u, v] \times F$ with $F \in \mathcal{F}$ and hence for every fixed null set N its trace on the set $[0, \infty) \times N$ is generated by the collection of sets of the form $(u, v] \times (F \cap N)$. Because the latter sets are predictable the traces of the product σ-field and the predictable σ-field on the set $[0, \infty)\times N$ are identical for every fixed null set N. We apply this with the null set N of all ω such that there exists $t \geq 0$ with $X_t(\omega) \neq 0$. For every Borel set B in $\mathbb{R}$ the set $\{(t, \omega): X_t(\omega) \in B\}$ is $\mathcal{B}_\infty \times \mathcal{F}$-measurable by assumption, and is contained in $[0, \infty) \times N$ if B does not contain 0. Thus it can be written as $A \cap \big([0, \infty) \times N\big)$ for some predictable set A and hence it is predictable, because $[0, \infty) \times N$ is predictable. The set $B = \{0\}$ can be handled by taking complements. □

5.2 Doléans Measure

In this section we prove that for every cadlag martingale M in L_2 there exists a σ-finite measure μ_M on the predictable σ-field such that

$$\begin{aligned} \mu_M(0 \times F_0) &= 0, & F_0 \in \mathcal{F}_0, \\ \mu_M\big((s,t] \times F_s\big) &= \mathrm{E}1_{F_s}(M_t^2 - M_s^2), & s < t, F_s \in \mathcal{F}_s. \end{aligned} \tag{5.14}$$

The right side of the preceding display is nonnegative, because M^2 is a submartingale. We can see this explicitly by rewriting it as

$$\mathrm{E}1_{F_s}(M_t - M_s)(M_t + M_s) = \mathrm{E}1_{F_s}(M_t - M_s)^2,$$

which follows because $\mathrm{E}1_{F_s}(M_t - M_s)M_s = 0$ by the martingale property, so that we can change "+" into "−". The measure μ_M is called the *Doléans measure* of M.

5.15 Example (Brownian motion). If $M = B$ is a Brownian motion, then by the independence of $B_t - B_s$ and F_s,

$$\begin{aligned} \mu_B\big((s,t] \times F_s\big) &= \mathrm{E}1_{F_s}\mathrm{E}(B_t^2 - B_s^2) = \mathbb{P}(F_s)(t-s) \\ &= (\lambda \times \mathbb{P})\big((s,t] \times F_s\big). \end{aligned}$$

Thus the Doléans measure of Brownian motion is the product measure $\lambda \times \mathbb{P}$. This is not only well defined on the predictable σ-field, but also on the bigger product σ-field $\mathcal{B}_\infty \times \mathcal{F}$. □

5.16 EXERCISE. Find the Doléans measure of the compensated Poisson process.

In order to prove the existence of the measure μ_M in general, we follow the usual steps of measure theory. First we extend μ_M by additivity to disjoint unions of the form

$$A = \{0\} \times F_0 \bigcup \bigcup_{i=1}^k (s_i, t_i] \times F_i, \qquad F_0 \in \mathcal{F}_0, F_i \in \mathcal{F}_{s_i},$$

by setting

$$\mu_M(A) = \sum_{i=1}^k \mathrm{E}1_{F_i}(M_{t_i}^2 - M_{s_i}^2).$$

It must be shown that this is well defined: if A can be represented as a disjoint, finite union of predictable rectangles in two different ways, then the two numbers $\mu_M(A)$ obtained in this way must agree. This can be shown by the usual trick of considering the common refinement. Given two disjoint, finite unions that are equal,

$$A = \{0\} \times F_0 \bigcup \bigcup_{i=1}^k (s_i, t_i] \times F_i = \{0\} \times F_0 \bigcup \bigcup_{j=1}^l (s_j', t_j'] \times F_j',$$

we can write A also as the disjoint union of $\{0\}\times F_0$ and the sets

$$\bigl((s_i,t_i]\times F_i\bigr)\cap\bigl((s_j',t_j']\times F_j'\bigr)=(s_{i,j}'',t_{i,j}'']\times F_{i,j}''.$$

Thus we have represented A in three ways. Next we show that $\mu_M(A)$ according to the third refined partition is equal to $\mu_M(A)$ defined through the other partitions. We omit the details of this verification. Once we have verified that the measure μ_M is well defined in this way, it is clear that it is finitely additive on the collection of finite disjoint unions of predictable rectangles.

The set of all finite disjoint unions of predictable rectangles is a ring, and generates the predictable σ-field. The first can be proved in the same way as it is proved that the cells in $\mathbb{R}^2$ form a ring. The second is the content of Lemma 5.5. We take both for facts. Next Carathéodory's theorem implies that μ_M is extendible to $\mathcal{P}$ provided that it is countably additive on the ring. This remains to proved.

5.17 Theorem. *For every cadlag martingale M in L_2 there exists a unique measure μ_M on the predictable σ-field such that (5.14) holds.*

Proof. See Chung and Williams, p50–53. ∎

5.18 EXERCISE. Show that $\mu_M\bigl([0,t]\times\Omega\bigr)<\infty$ for every $t\geq 0$ and conclude that μ_M is σ-finite.

5.3 Square-integrable Martingales

Given a square-integrable martingale M we define an integral $\int X\,dM$ for increasingly more general processes X. If X is of the form $1_{(s,t]}Z$ for some (time-independent) random variable Z, then we want to define

$$\int 1_{(s,t]}Z\,dM=Z(M_t-M_s).$$

Here $1_{(s,t]}Z$ is short-hand notation for the map $(u,\omega)\mapsto 1_{(s,t]}(u)Z(\omega)$ and the right side is the random variable $\omega\mapsto Z(\omega)\bigl(M_t(\omega)-M_s(\omega)\bigr)$. We now agree that this random variable is the "integral" written on the left. Clearly this integral is like a Riemann-Stieltjes integral for fixed ω.

We also want the integral to be linear in the integrand, and are lead to define

$$\int\sum_{i=1}^k a_i 1_{(s_i,t_i]\times F_i}\,dM=\sum_{i=1}^k a_i 1_{F_i}(M_{t_i}-M_{s_i}).$$

By convention we choose "to give measure 0 to 0" and set

$$\int a_0 1_{\{0\}\times F_0}\, dM = 0.$$

We can only postulate these definitions if they are consistent. If $X = \sum_{i=1}^k a_i 1_{(s_i,t_i]\times F_i}$ has two representations as a linear combination of predictable rectangles, then the right sides of the second last display must agree. For this it is convenient to restrict the definition initially to linear combinations of disjoint predictable rectangles. The consistency can then be checked using the joint refinements of two given representations. We omit the details.

5.19 Definition. *If $X = a_0 1_{\{0\}\times F_0} + \sum_{i=1}^k a_i 1_{(s_i,t_i]\times F_i}$ is a linear combination of disjoint predictable rectangles, then the stochastic integral of X relative to M is defined as $\int X\, dM = \sum_{i=1}^k a_i 1_{F_i}(M_{t_i} - M_{s_i})$.*

In this definition there is no need for the restriction to predictable processes. However, predictability is important for the extension of the integral. We extend by continuity, based on the following lemmas.

5.20 Lemma. *Every uniformly continuous map defined on a dense subset of a metric space with values in another metric space extends in a unique way to a continuous map on the whole space. If the map is a linear isometry between two normed spaces, then so is the extension.*

5.21 Lemma. *The collection of simple processes X as in Definition 5.19 is dense in $L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$. Every bounded $X \in L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$ is a limit in this space of a uniformly bounded sequence of simple processes.*

5.22 Lemma. *For every X as in Definition 5.19 we have $\int X^2\, d\mu_M = \mathrm{E}(\int X\, dM)^2$.*

Proofs. The first lemma is a standard result from topology.

Because any function in $L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$ is the limit of a sequence of bounded functions, for Lemma 5.21 it suffices to show that any bounded element of $L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$ can be obtained as such a limit. Because $1_{[0,t]}X \to X$ in $L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$ as $t\to\infty$, we can further restrict ourselves to elements that vanish off $[0,t]\times\Omega$.

Let $\mathcal{H}$ be the set of all bounded, predictable X such that $X1_{[0,t]}$ is a limit in $L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$ of a sequence of linear combinations of indicators of predictable rectangles, for every $t\geq 0$. Then $\mathcal{H}$ is a vector space and contains the constants. A "diagonal type" argument shows that it is also closed under bounded monotone limits. Because $\mathcal{H}$ contains the indicators of predictable rectangles (the sets in Lemma 5.5) and this collection of sets is intersection stable, Lemma 5.21 follows from the monotone class theorem, Theorem 1.23.

Using the common refinement of two finite disjoint unions of predictable rectangles, we can see that the minimum of two simple processes is again a simple process. This implies the second statement of Lemma 5.21.

Finally consider Lemma 5.22. Given a linear combination X of disjoint predictable rectangles as in Definition 5.19, its square is given by $X^2 = a_0^2 1_{\{0\}\times F_0} + \sum_{i=1}^k a_i^2 1_{(s_i,t_i]\times F_i}$. Hence, by (5.14),

$$\int X^2\, d\mu_M = \sum_{i=1}^k a_i^2 \mu_M\bigl((s_i,t_i]\times F_i\bigr) = \sum_{i=1}^k a_i^2 \mathrm{E} 1_{F_i}(M_{t_i} - M_{s_i})^2. \tag{5.23}$$

On the other hand, by Definition 5.19,

$$\begin{aligned}\mathrm{E}\Bigl(\int X\, dM\Bigr)^2 &= \mathrm{E}\Bigl(\sum_{i=1}^k a_i 1_{F_i}(M_{t_i} - M_{s_i})\Bigr)^2\\ &= \sum_{i=1}^k\sum_{j=1}^k a_i a_j \mathrm{E} 1_{F_i} 1_{F_j}(M_{t_i} - M_{s_i})(M_{t_j} - M_{s_j}).\end{aligned}$$

Because the rectangles are disjoint we have for $i \neq j$ that either $1_{F_i}1_{F_j} = 0$ or $(s_i,t_i]\cap(s_j,t_j] = \emptyset$. In the first case the corresponding term in the double sum is clearly zero. In the second case it is zero as well, because, if $t_i \le s_j$, the variable $1_{F_i}1_{F_j}(M_{t_i} - M_{s_i})$ is $\mathcal{F}_{s_j}$-measurable and the martingale difference $M_{t_j} - M_{s_j}$ is orthogonal to $\mathcal{F}_{s_j}$. Hence the off-diagonal terms vanish and the expression is seen to reduce to the right side of (5.23). ∎

Lemma 5.22 shows that the map

$$X \mapsto \int X\, dM,$$
$$L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr) \to L_2(\Omega,\mathcal{F},\mathbb{P}),$$

is an isometry if restricted to the linear combinations of disjoint indicators of predictable rectangles. By Lemma 5.21 this class of functions is dense in $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr)$. Because an isometry is certainly uniformly continuous, this map has a unique continuous extension to $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr)$, by Lemma 5.20. We define this extension to be the stochastic integral $\int X\, dM$.

5.24 Definition. *For M a cadlag martingale in L_2 and X a predictable process in $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr)$, the stochastic integral $X \mapsto \int X\, dM$ is the unique continuous extension to $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr)$ of the map defined in Definition 5.19 with range inside $L_2(\Omega,\mathcal{F},\mathbb{P})$.*

Thus defined a stochastic integral is an element of the Hilbert space $L_2(\Omega,\mathcal{F},\mathbb{P})$ and therefore an equivalence class of functions. We shall also consider every representative of the class to be "the" stochastic integral $\int X\,dM$. In general, there is no preferred way of choosing a representative.

If X is a predictable process such that $1_{[0,t]}X \in L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$, then $\int 1_{[0,t]}X\,dM$ is defined through the preceding definition. A short-hand notation for this is $\int_0^t X\,dM$. By linearity of the stochastic integral we then have

$$\int 1_{(s,t]}X\,dM = \int_0^t X\,dM - \int_0^s X\,dM, \qquad s<t.$$

We abbreviate this to $\int_s^t X\,dM$. The equality is understood in an almost sure sense, because all three integrals are equivalence classes.

If $1_{[0,t]}X \in L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$ for every $t\geq 0$, then we can define a process $X\cdot M$ satisfying

$$(X\cdot M)_t = \int_0^t X\,dM \equiv \int 1_{[0,t]}X\,dM.$$

Because for every $t\geq 0$ the stochastic integral on the right is defined only up to a null set, this display does not completely define the process $X\cdot M$. However, any specification yields a martingale $X\cdot M$ and there always exists a cadlag version of $X\cdot M$.

5.25 Theorem. *Suppose that M is a cadlag martingale in L_2 and that X is a predictable process with $\int 1_{[0,t]}X^2\,d\mu_M < \infty$ for every $t\geq 0$.*

(i) Any version of $X\cdot M = \{\int_0^t X\,dM\colon t\geq 0\}$ is a martingale in L_2.
(ii) There exists a cadlag version of $X\cdot M$.
(iii) If M is continuous, then there exists a continuous version of $X\cdot M$.
(iv) The processes $\Delta(X\cdot M)$, where $X\cdot M$ is chosen cadlag, and $X\Delta M$ are indistinguishable.

Proof. If X is a finite linear combination of predictable rectangles, of the form as in Definition 5.19, then so is $1_{[0,t]}X$ and hence $\int 1_{[0,t]}X\,dM$ is defined as

$$\int 1_{[0,t]}X\,dM = \sum_{i=1}^k a_i 1_{F_i}(M_{t_i\wedge t} - M_{s_i\wedge t}).$$

As a process in t, this is a martingale in L_2, because each of the stopped processes M^{t_i} or M^{s_i} is a martingale, so that $M^{t_i}-M^{s_i}$ is martingale whence $1_{F_i}(M^{t_i}-M^{s_i})$ is a martingale on the time set $[s_i,\infty)$, while this process is zero on $[0,s_i]$; furthermore, a linear combination of martingales is a martingale. The stochastic integral $X\cdot M$ of a general integrand X is defined as an L_2-limit of stochastic integrals of simple predictable processes. Because the martingale property is retained under convergence in L_1, the process $X\cdot M$ is a martingale.

Statement (ii) is an immediate consequence of (i) and Theorem 4.6, which implies that any martingale possesses a cadlag version.

To prove statement (iii) it suffices to show that the cadlag version of $X \cdot M$ found in (ii) is continuous if M is continuous. If X is elementary, then this is clear from the explicit formula for the stochastic integral used in (i). In general, the stochastic integral $(X \cdot M)_t$ is defined as the L_2-limit of a sequence of elementary stochastic integrals $(X_n \cdot M)_t$. Given a fixed $T > 0$ we can use the same sequence of linear combinations of predictable rectangles for every $0 \leq t \leq T$. Each process $X \cdot M - X_n \cdot M$ is a cadlag martingale in L_2 and hence, by Corollary 4.38, for every $T > 0$,

$$\Bigl\| \sup_{0 \leq t \leq T} \bigl|(X \cdot M)_t - (X_n \cdot M)_t\bigr| \Bigr\|_2 \leq 2\bigl\|(X \cdot M)_T - (X_n \cdot M)_T\bigr\|_2.$$

The right side converges to zero as $n \to \infty$ and hence the variables in the left side converge to zero in probability. There must be a subsequence $\{n_i\}$ along which the convergence is almost surely, i.e. $(X_{n_i} \cdot M)_t \to (X \cdot M)_t$ uniformly in $t \in [0, T]$, almost surely. Because continuity is retained under uniform limits, the process $X \cdot M$ is continuous almost surely. This concludes the proof of (iii).

Let $\mathcal{H}$ be the set of all bounded predictable processes X for which (iv) is true. Then $\mathcal{H}$ is a vector space that contains the constants, and it is readily verified that it contains the indicators of predictable rectangles. If $0 \leq X_n \uparrow X$ for a uniformly bounded X, then $1_{[0,t]}X_n \to 1_{[0,t]}X$ in $L_2\bigl([0,\infty) \times \Omega, \mathcal{P}, \mu_M\bigr)$. As in the preceding paragraph we can select a subsequence such that, for the cadlag versions, $X_{n_i} \cdot M \to X \cdot M$ uniformly on compacta, almost surely. Because $|\Delta Y| \leq 2\|Y\|_\infty$ for any cadlag process Y, the latter implies that $\Delta(X_{n_i} \cdot M) \to \Delta(X \cdot M)$ uniformly on compacta, almost surely. On the other hand, by pointwise convergence of X_n to X, $X_{n_i}\Delta M \to X\Delta M$ pointwise on $[0, \infty) \times \Omega$. Thus $\{X_n\} \subset \mathcal{H}$ implies that $X \in \mathcal{H}$. By the monotone class theorem, Theorem 1.23, $\mathcal{H}$ contains all bounded predictable X. A general X can be truncated to the interval $[-n, n]$, yielding a sequence X_n with $X_n \to X$ pointwise on $[0, \infty) \times \Omega$ and $1_{[0,t]}X_n \to 1_{[0,t]}X$ in $L_2\bigl([0,\infty) \times \Omega, \mathcal{P}, \mu_M\bigr)$. The latter implies, as before, that there exists a subsequence such that, for the cadlag versions, $X_{n_i} \cdot M \to X \cdot M$ uniformly on compacta, almost surely. It is now seen that (iv) extends to X. ∎

The following two lemmas gives further properties of stochastic integrals. Here we use notation as in the following exercise.

5.26 EXERCISE. Let $S \leq T$ be stopping times and let X be an $\mathcal{F}_S$-measurable random variable. Show that the process $1_{(S,T]}X$ defined as $(t, \omega) \mapsto 1_{(S(\omega),T(\omega)]}(t)X(\omega)$ is predictable.

5.27 Lemma. *Let M be a cadlag martingale in L_2 and let $S \leq T$ be bounded stopping times.*

(i) $\int 1_{(S,T]}X\,dM = X(M_T - M_S)$ almost surely, for every bounded $\mathcal{F}_S$-measurable random variable X.

(ii) $\int 1_{(S,T]} XY\, dM = X \int 1_{(S,T]} Y\, dM$ *almost surely, for every bounded $\mathcal{F}_S$-measurable random variable X and bounded predictable process Y.*

(iii) $\int 1_{(S,T]} X\, dM = N_T - N_S$ *almost surely, for every bounded predictable process X, and N a cadlag version of $X \cdot M$.*

(iv) $\int 1_{\{0\}\times\Omega} X\, dM = 0$ *almost surely for every predictable process X.*

Proof. Let S_n and T_n be the upward discretizations of S and T on the grid $0 < 2^{-n} < 22^{-n} < \cdots < k_n 2^{-n}$, as in the proof of Theorem 4.20, for k_n sufficiently large that $k_n 2^{-n} > S \vee T$. Then $S_n \downarrow S$ and $T_n \downarrow T$, so that $1_{(S_n,T_n]} \to 1_{(S,T]}$ pointwise on Ω. Furthermore,

$$1_{(S_n,T_n]} = \sum_{k=0}^{k_n} 1_{(k2^{-n},(k+1)2^{-n}]\times\{S<k2^{-n}\le T\}}. \tag{5.28}$$

If we can prove the lemma for $(S_n, T_n]$ taking the place of $(S,T]$ and every n, then it follows for $(S,T]$ upon taking limits. (Note here that μ_M is a finite measure on sets of the form $[0,K]\times\Omega$ and all $(S_n,T_n]$ are contained in a set of this form.)

For the proof of (i) we first consider the case that $X = 1_F$ for some $F \in \mathcal{F}_S$. In view of (5.28) and because $\{S < k2^{-n} \le T\} \cap F = (\{S < k2^{-n}\} \cap F) \cap \{k2^{-n} \le T\}$ is contained in $\mathcal{F}_{k2^{-n}}$, the process $1_{(S_n,T_n]}X = 1_{(S_n,T_n]}1_F$ is a linear combination of predictable rectangles. Hence, by Definition 5.19,

$$\begin{aligned}\int 1_{(S_n,T_n]} 1_F\, dM &= \sum_{k=0}^{k_n} 1_{\{S<k2^{-n}\le T\}\cap F}(M_{(k+1)2^{-n}} - M_{k2^{-n}})\\ &= 1_F(M_{T_n} - M_{S_n}).\end{aligned}$$

This proves (i) in the case that $X = 1_F$. By linearity (i) is then also true for X that are simple over $\mathcal{F}_S$. A general, bounded $\mathcal{F}_S$-measurable X can be approximated by a uniformly bounded sequence of simple X. Both sides of the equality in (i) then converge in L_2 and hence the equality is valid for such X.

For the proof of (ii) first assume that $X = 1_F$ for some $F \in \mathcal{F}_S$ and that $Y = 1_{(u,v]\times F_u}$ for some $F_u \in \mathcal{F}_u$. In view of (5.28),

$$1_{(S_n,T_n]} 1_F 1_{(u,v]\times F_u} = \sum_{\substack{k=0\\ k2^{-n}\vee u<(k+1)2^{-n}\wedge v}}^{k_n} 1_{(k2^{-n}\vee u,(k+1)2^{-n}\wedge v]\times\{S<k2^{-n}\le T\}\cap F\cap F_u}$$

is a linear combination of predictable rectangles, whence, by Definition 5.19, with the summation index k ranging over the same set as in the preceding display,

$$\begin{aligned}\int 1_{(S_n,T_n]}1_F 1_{(u,v]\times F_u}\,dM \\ &= \sum_k 1_{\{S<k2^{-n}\le T\}\cap F\cap F_u}(M_{(k+1)2^{-n}\wedge v} - M_{k2^{-n}\vee u}) \\ &= 1_F \sum_k 1_{\{S<k2^{-n}\le T\}\cap F_u}(M_{(k+1)2^{-n}\wedge v} - M_{k2^{-n}\vee u}) \\ &= 1_F \int 1_{(S_n,T_n]}1_{(u,v]\times F_u}\,dM.\end{aligned}$$

This proves (ii) for X and Y of the given forms. The general case follows again by linear extension and approximation.

For (iii) it suffices to show that $N_{T_n} = \int 1_{(0,T_n]}X\,dM$ almost surely. Since $N_0 = 0$,

$$\begin{aligned}N_{T_n} &= \sum_k 1_{\{k2^{-n}\le T\}}(N_{(k+1)2^{-n}} - N_{k2^{-n}}) \\ &= \sum_k \int 1_{\{k2^{-n}\le T\}}1_{(k2^{-n},(k+1)2^{-n}]}X\,dM = \int 1_{(0,T_n]}X\,dM,\end{aligned}$$

where the second equality follows from (ii), and the last equality by (5.28) after changing the order of summation and integration.

Because μ_M does not charge $\{0\}\times\Omega$, $1_{\{0\}\times\Omega}X = 0$ in $L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$ for any X and hence $0 = \int 1_{\{0\}\times\Omega}X\,dM$ in L_2, by the isometry. This proves (iv). ∎

The preceding lemma remains valid for unbounded processes X,Y or unbounded stopping times S,T, provided the processes involved in the statements are appropriately square-integrable. In each case this is true under several combinations of conditions on X,Y,S,T and M.

5.29 Lemma (Substitution). *Let M be a cadlag martingale in L_2 and let $N = Y\cdot M$ be a cadlag version of the stochastic integral of a predictable process Y with $1_{[0,t]}Y \in L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_M\big)$ for every $t\ge 0$. Then*

(i) μ_N is absolutely continuous relative to μ_M and $d\mu_N = Y^2\,d\mu_M$.

(ii) $\int X\,dN = \int XY\,dM$ almost surely for every $X \in L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_N\big)$.

Proof. By Lemma 5.27(ii), for every bounded predictable process Y and every $s<t$ and $F_s\in\mathcal{F}_s$,

$$1_{F_s}\int 1_{(s,t]}Y\,dM = \int 1_{(s,t]\times F_s}Y\,dM. \tag{5.30}$$

This can be extended to predictable Y as in the statement of the lemma by approximation. Specifically, if Y_n is Y truncated to the interval $[-n,n]$, then

$1_{(s,t]}Y_n \to 1_{(s,t]}Y$ in $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr)$ and hence also $1_{F_s}1_{(s,t]}Y_n \to 1_{F_s}1_{(s,t]}Y$ in this space. By the isometry property of the stochastic integral it follows that $\int 1_{(s,t]}Y_n\,dM$ and $\int 1_{(s,t]\times F_s}Y_n\,dM$ converge in L_2 to the corresponding expressions with Y instead of Y_n, as $n\to\infty$. Therefore, if (5.30) is valid for Y_n instead of Y for every n, then it is valid for Y.

We can rewrite the left side of (5.30) as $1_{F_s}(N_t-N_s)$. Therefore, for every predictable rectangle $(s,t]\times F_s$,

$$\mu_N\bigl((s,t]\times F_s\bigr) = \mathrm{E}1_{F_s}(N_t-N_s)^2 = \mathrm{E}\Bigl(\int 1_{(s,t]\times F_s}Y\,dM\Bigr)^2$$
$$= \int 1_{(s,t]\times F_s}Y^2\,d\mu_M,$$

by the isometry property of the stochastic integral. The predictable rectangles are an intersection stable generator of the predictable σ-field and $[0,\infty)\times\Omega$ is a countable union of predictable rectangles of finite measures under μ_N and $Y^2\cdot\mu_M$. Thus these measures must agree on all predictable sets, as asserted in (i).

For the proof of (ii) first assume that $X=1_{(s,t]\times F_s}$ for $F_s\in\mathcal{F}_s$. Then the equality in (ii) reads

$$1_{F_s}(N_t-N_s) = \int 1_{(s,t]\times F_s}Y\,dM, \qquad \text{a.s.}.$$

The left side of this display is exactly the left side of (5.30) and hence (ii) is correct for this choice of X. By linearity this extends to all X that are simple over the predictable rectangles.

A general $X\in L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_N\bigr)$ can be approximated in this space by a sequence of simple X_n. Then by (i)

$$\int |X_nY-XY|^2\,d\mu_M = \int |X_n-X|^2\,d\mu_N \to 0.$$

Thus, by the isometry property of the stochastic integral, we can take limits as $n\to\infty$ in the identities $\int X_nY\,dM = \int X_n\,dN$ to obtain the desired identity for general X and Y. ∎

5.4 Locally Square-integrable Martingales

In this section we extend the stochastic integral by localization to more general processes X and M.

Given a cadlag local L_2-martingale M we allow integrands X that are predictable processes and are such that there exists a sequence of stopping times $0\le T_n\uparrow\infty$ such that, for every n,

(i) M^{T_n} is a martingale in L_2,
(ii) $1_{[0,t\wedge T_n]}X \in L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_{M^{T_n}}\big)$ for every $t \geq 0$.
A sequence of stopping times T_n of this type is called a *localizing sequence for the pair* (X, M). If such a localizing sequence exists, then

$$\int 1_{[0,t\wedge T_n]}X\, dM^{T_n}$$

is a well-defined element of $L_2(\Omega,\mathcal{F},\mathbb{P})$, for every n, by Definition 5.24. We define $\int_0^t X\,dM$ as the almost sure limit as $n\to\infty$ of these random variables. The set of all X for which a localizing sequence exists for the pair (X, M) is denoted $L_{2,loc}(M)$.

5.31 Definition. *Given a cadlag local L_2-martingale M and a predictable process X for which there exists a localizing sequence T_n for the pair (X, M), the stochastic integral $\int_0^t X\,dM$ is defined as the almost sure limit of the sequence of random variables $\int 1_{[0,t\wedge T_n]}X\,dM^{T_n}$, as $n\to\infty$.*

We denote the stochastic process $t\mapsto \int_0^t X\,dM$ by $X\cdot M$. If the limit as $t\to\infty$ of $(X\cdot)_t$ exists, then we denote this by $\int_0^\infty X\,dM$ or $(X\cdot M)_\infty$.

It is not immediately clear that this definition is well posed. Not only do we need to show that the almost sure limit exists, but we must also show that the limit does not depend on the localizing sequence. This issue requires scrutiny of the definitions, but turns out to be easily resolvable. An integral of the type $\int 1_{[0,S]}X\,dM^T$ ought to depend only on $S\wedge T$ and the values of the processes X and M on the set $[0, S\wedge T]$, because the integrand $1_{[0,S]}X$ vanishes outside $[0, S]$ and the integrator M^T is constant outside $[0, T]$. In analogy with the ordinary integral, a nonzero integral should require both a nonzero integrand and a nonconstant integrator.

This reasoning suggests that, for every $n\geq m$, on the event $\{t\leq T_m\}$, where $t\wedge T_m = t\wedge T_n$, the variable $\int 1_{[0,t\wedge T_m]}X\,dM^{T_m}$ is the same as the variable $\int 1_{[0,t\wedge T_n]}X\,dM^{T_n}$. Then the limit as $n\to\infty$ trivially exists on the event $\{t\leq T_m\}$. Because $\cup_m\{t\leq T_m\}=\Omega$ the limit exists everywhere.

The following lemma makes these arguments precise.

5.32 Lemma. *Let M be a cadlag process and X a predictable process, and let S, T, U, V be stopping times such that S and U are bounded, M^T and M^V are martingales in L_2 and such that $1_{[0,S]}X$ and $1_{[0,U]}X$ are contained in $L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_{M^T}\big)$ and $L_2\big([0,\infty)\times\Omega,\mathcal{P},\mu_{M^V}\big)$, respectively. Then $\int 1_{[0,S]}X\,dM^T = \int 1_{[0,U]}X\,dM^V$ almost surely on the event $\{S\wedge T = U\wedge V\}$.*

Proof. First assume that X is a predictable rectangle of the form $X =$

$1_{(s,t]\times F_s}$. By Lemma 5.27(ii) and next (i),

$$\int 1_{[0,S]} 1_{(s,t]\times F_s}\, dM^T = 1_{F_s} \int 1_{[0,S]} 1_{(s,t]}\, dM^T = 1_{F_s}(M^T_{S\wedge t} - M^T_{S\wedge s})$$
$$= 1_{F_s}(M_{S\wedge t\wedge T} - M_{S\wedge s\wedge T}).$$

The right side depends on on (S,T) only through $S\wedge T$. Clearly the same calculation with the stopping times U and V gives the same result on the event $\{S\wedge T = U\wedge V\}$.

Next let X be a bounded predictable process. Then, for every given $t \geq 0$, the process $1_{[0,t]}X$ is automatically contained in $L_2\bigl([0,\infty)\times \Omega, \mathcal{P}, \mu_{M^T} + \mu_{M^V}\bigr)$ and by (a minor extension of) Lemma 5.21 there exists a bounded sequence of simple processes X_n with $X_n \to 1_{[0,t]}X$ in $L_2\bigl([0,\infty)\times \Omega, \mathcal{P}, \mu_{M^T} + \mu_{M^V}\bigr)$. If $t\geq S$, then this implies that $1_{[0,S]}X_n \to 1_{[0,S]}X$ in $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_{M^T}\bigr)$ and hence $\int 1_{[0,S]}X_n\, dM^T \to \int 1_{[0,S]}X\, dM^T$ in L_2, by the isometry. We can argue in the same way with S and T replaced by U and V. Thus the equality of $\int 1_{[0,S]}X_n\, dM^T$ and $\int 1_{[0,U]}X_n\, dM^V$ for every n on the event $\{S\wedge T = U\wedge V\}$ carries over onto X.

A general X as in the lemma can be truncated to $[-n,n]$ and next we take limits. ∎

Thus the reasoning given previously is justified and shows that the almost sure limit of $\int 1_{[0,t\wedge T_n]}X\, dM^{T_n}$ exists. To see that the limit is also independent of the localizing sequence, suppose that S_n and T_n are two localizing sequences for the pair of processes (X,M). Then the lemma implies that on the event $A_n = \{t\wedge S_n = t\wedge T_n\}$, which contains $\{t\leq S_n\wedge T_n\}$,

$$\int 1_{[0,t\wedge S_n]}X\, dM^{S_n} = \int 1_{[0,t\wedge T_n]}X\, dM^{T_n}, \qquad \text{a.s..}$$

It follows that the almost sure limits of left and right sides of the display, as $n\to\infty$, are the same almost surely on the event A_n for every n, and hence on the event $\cup_n A_n = \Omega$. Thus the two localizing sequences yield the same definition of $\int_0^t X\, dM$.

In a similar way we can prove that we get the same stochastic integral if we use separate localizing sequences for X and M. (See Exercise 5.33.) In particular, if M is an L_2-martingale, X is a predictable process, and $0\leq T_n\uparrow\infty$ is a sequence of stopping times such that $1_{[0,t\wedge T_n]}X \in L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr)$ for every t and every n, then

$$\int 1_{[0,t\wedge T_n]}X\, dM,$$

which is well defined by Definition 5.24, converges almost surely to $\int_0^t X\, dM$ as defined in Definition 5.31. So "if it is not necessary to localize M, then not doing so yields the same result".

5.33 EXERCISE. Suppose that M is a local L_2-martingale with localizing sequence T_n, X a predictable process, and $0 \leq S_n \uparrow \infty$ are stopping times such that $1_{[0,t\wedge S_n]}X \in L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_{M^{T_n}}\bigr)$ for every $t\geq 0$ and n. Show that $\lim_{n\to\infty}\int 1_{[0,t\wedge S_n]}X\,dM^{T_n}$ exists almost surely and is equal to $\int_0^t X\,dM$. (Note that $S_n\wedge T_n$ is a localizing sequence for the pair (X,M), so that $\int_0^t X\,dM$ is well defined in view of Exercise 5.34.)

5.34 EXERCISE. Let M be a cadlag process and S and T stopping times such that M^S and M^T are L_2-martingales. Show that
(i) $\mu_{M^S}\bigl(A\cap[0,S\wedge T]\bigr)=\mu_{M^T}\bigl(A\cap[0,S\wedge T]\bigr)$ for every $A\in\mathcal{P}$.
(ii) if M is an L_2-martingale, then $\mu_{M^S}(A)=\mu_M\bigl(A\cap[0,S]\bigr)$ for every $A\in\mathcal{P}$.

The present extension of the stochastic integral possesses similar properties as in the preceding section.

5.35 Theorem. *Suppose that M is a cadlag local L_2-martingale and X a predictable process for which there exists a localizing sequence T_n for the pair (X,M).*
(i) There exists a cadlag version of $X\cdot M$.
(ii) Any cadlag version of $X\cdot M$ is a local L_2-martingale relative to the localizing sequence T_n.
(iii) If M is continuous, then there exists a continuous version of $X\cdot M$.
(iv) The processes $\Delta(X\cdot M)$, where $X\cdot M$ is chosen cadlag, and $X\Delta M$ are indistinguishable.

Proof. For every n let Y_n be a cadlag version of the process $t\mapsto\int 1_{[0,t\wedge T_n]}X\,dM^{T_n}$. By Theorem 5.25 such a version exists; it is an L_2-martingale; and we can and do choose it continuous if M is continuous. For fixed $t\geq 0$ the variable $T_m\wedge t$ is a stopping time and hence by Lemma 5.27(iii)

$$Y_{n,T_m\wedge t}=\int 1_{[0,T_m\wedge t\wedge T_n]}X\,dM^{T_n},\qquad \text{a.s.}$$

By Lemma 5.32 the right side of this display changes at most on a null set if we replace M^{T_n} by M^{T_m}. For $m\leq n$ we have $T_m\wedge T_n=T_m$ and hence the integrand is identical to $1_{[0,t\wedge T_m]}X$. If we make both changes, then the right side becomes $Y_{m,t}$. We conclude that $Y_{n,T_m\wedge t}=Y_{m,t}$ almost surely, for every fixed t and $m\leq n$. This shows that the stopped martingale $Y_n^{T_m}$ is a version of the stopped martingale $Y_m^{T_m}$, for $m\leq n$. Because both martingales possess cadlag sample paths, the two stopped processes are indistinguishable. This implies that Y_n and Y_m agree on the set $[0,T_m]$ except possibly for points (t,ω) with ω ranging over a null set. The union of all null sets attached to some pair (m,n) is still a null set. Apart from points (t,ω) with ω contained in this null set, the limit Y as $n\to\infty$ of $Y_{n,t}(\omega)$ exists and agrees with $Y_{m,t}(\omega)$ on $[0,T_m]$. The latter implies that it is cadlag, and Y^{T_m} is indistinguishable of Y_m. Furthermore, the jump process of Y is indistinguishable of the jump process of Y_m on the set $[0,T_m]$ and hence is equal to $1_{[0,T_m]}X\Delta M^{T_m}=X\Delta M$ on the set $[0,T_m]$, by Theorem 5.25(iv).

By definition this limit Y is a version of $X\cdot M$. ∎

The properties as in Lemmas 5.27 and 5.29 also extend to the present more general integral. For instance, in a condensed notation we have, for T a stopping time and for processes X, Y and M for which the expressions are defined,

$$\begin{aligned}(X\cdot M)^T&=X\cdot M^T=(1_{[0,T]}X)\cdot M,\\ X\cdot(Y\cdot M)&=(XY)\cdot M,\\ \Delta(X\cdot M)&=X\Delta M.\end{aligned}\tag{5.36}$$

We shall formalize this later, after introducing the final extension of the stochastic integral.

5.37 Example (Continuous processes). The stochastic integral $X\cdot M$ is defined for every pair of a continuous process X and a continuous local martingale M with $M_0=0$.

Such a pair can be localized by the stopping times

$$T_n=\inf\{t\ge 0\colon |X_t|\ge n, |M_t|\ge n\}.$$

If $0<t\le T_n$, then $|X_t|\le n$ and $|M_t|\le n$, by the continuity of the sample paths of the processes. It follows that M^{T_n} is an L_2-bounded martingale and

$$\begin{aligned}\bigl|1_{(0,T_n]}X\bigr|&\le n,\\ \mu_{M^{T_n}}(0,\infty)&=\mathrm{E}(M_\infty^{T_n}-M_0^{T_n})^2\le n^2.\end{aligned}$$

Therefore $\mu_{M^{T_n}}$ is a finite measure and $1_{(0,T_n]}X$ is bounded and hence in $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_{M^{T_n}}\bigr)$. Trivially $1_{\{0\}}X\in L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_{M^{T_n}}\bigr)$, because $\mu_{M^{T_n}}(\{0\}\times\Omega)=0$, and hence $1_{[0,T_n\wedge t]}X\in L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_{M^{T_n}}\bigr)$ for every $t\ge 0$. □

5.38 EXERCISE. Extend the preceding example to processes that may have jumps, but of jump sizes that are uniformly bounded.

5.39 Example (Locally bounded integrators). The stochastic integral $X\cdot M$ is defined for every pair of a local L_2-martingale M and a locally bounded predictable process X.

Here "locally bounded" means that there exists a sequence of stopping times $0\le T_n\uparrow\infty$ such that X^{T_n} is uniformly bounded, for every n. We can choose this sequence of stopping times to be the same as the localizing sequence for M. (Otherwise, we use the minimum of the two localizing sequences.) Then $1_{[0,T_n\wedge t]}X$ is uniformly bounded and hence is contained in $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_{M^{T_n}}\bigr)$ for every t and n. Thus Definition 5.24 applies. ◻

5.5 Brownian Motion

The Doléans measure of Brownian motion is the product measure $\lambda\times\mathbb{P}$ and hence exists as a measure on the product σ-field $\mathcal{B}_\infty\times\mathcal{F}$, which is bigger than the predictable σ-field. This can be used to define the stochastic integral $\int X\,dB$ relative to a Brownian motion B also for non-predictable integrands. The main aim of this section is to define the stochastic integral $\int_0^t X\,dB$ for all measurable, adapted processes X such that $\int_0^t X_s^2\,ds$ is finite almost surely.

Going from predictable to adapted measurable processes may appear an important extension. However, it turns out that any measurable, adapted process X is almost everywhere equal to a predictable process $\tilde X$, relative to $\lambda\times\mathbb{P}$. Because we want to keep the isometry relationship of a stochastic integral, then the only possibility is to define $\int_0^t X\,dM$ as $\int_0^t \tilde X\,dM$. From this perspective we obtain little new.

The key in the construction is the following lemma.

5.40 Lemma. *For every measurable, adapted process $X\colon[0,\infty)\times\Omega\to\mathbb{R}$ there exists a predictable process $\tilde X$ such that $X=\tilde X$ almost everywhere under $\lambda\times\mathbb{P}$.*

Proof. The proof is based on two facts:

(i) For every bounded, measurable process X there exists a bounded optional process $\bar X$ such that $\mathrm{E}(X_t|\,\mathcal{F}_t)=\bar X_t$ almost surely, for every $t\ge 0$.

(ii) For every bounded, optional process $\bar X$ there exists a predictable process $\tilde X$ such that the set $\{\bar X\ne\tilde X\}$ is contained in the union $\cup_n[T_n]$ of the graphs of countably many stopping times.

If we accept (i)–(ii), then the lemma can be proved as follows. For every bounded measurable process X, facts (i) and (ii) yield processes $\bar X$ and $\tilde X$. If X is adapted, then $X_t=\mathrm{E}(X_t|\,\mathcal{F}_t)=\bar X_t$ almost surely for every $t\ge 0$, by (i). Consequently, by Fubini's theorem

$$\lambda\times\mathbb{P}(X\ne\bar X)=\int\mathbb{P}\bigl(\omega\colon X_t(\omega)\ne\bar X_t(\omega)\bigr)\,d\lambda(t)=0.$$

Because the sections $\{t: (\omega, t) \in G\}$ of the set $G = \cup_n [T_n]$ contain at most countably many points, they have Lebesgue measure zero and hence $\lambda \times \mathbb{P}(\bar{X} \neq \tilde{X}) = 0$, by another application of Fubini's theorem. Combining (i) and (ii), we see that $\lambda \times \mathbb{P}(X \neq \tilde{X}) = 0$. This proves the lemma for bounded, measurable, adapted processes X. We can treat general processes X by truncating and taking limits. Specifically, if X_n is X truncated to $[-n, n]$, then $X_n \to X$ on $[0, \infty) \times \Omega$. If $\tilde{X}_n$ is predictable with $\tilde{X}_n = X_n$ except on a null set B_n, then $\tilde{X}_n$ converges to a limit at least on the complement of $\cup_n B_n$. We can define $\tilde{X}$ to be $\lim \tilde{X}_n$ if this exists and 0 otherwise.

We prove (i) by the monotone class theorem, Theorem 1.23. Let $\mathcal{H}$ be the set of all bounded, measurable processes X for which there exists an optional process $\bar{X}$ as in (i). Then $\mathcal{H}$ is a vector space and contains the constants. If $X_n \in \mathcal{H}$ with $0 \leq X_n \uparrow X$ for some bounded measurable X and $\bar{X}_n$ are the corresponding optional processes as in (i), then the process $\bar{X}$ defined as $\liminf \bar{X}_n$ if this liminf is finite, and as 0 if not, is optional. By the monotone convergence theorem for conditional expectations $(\bar{X}_n)_t = \mathrm{E}((X_n)_t | \mathcal{F}_t) \uparrow \mathrm{E}(X_t | \mathcal{F}_t)$ almost surely, for every $t \geq 0$. Hence for each $t \geq 0$, we have that $\bar{X}_t = \mathrm{E}(X_t | \mathcal{F}_t)$ almost surely.

In view of Theorem 1.23 it now suffices to show that the indicators of the sets $[0, s) \times F$, for $s \geq 0$ and $F \in \mathcal{F}$, which form an intersection stable generator of $\mathcal{B}_\infty \times \mathcal{F}$, are in $\mathcal{H}$. By Example 2.6 there exists a cadlag process Y such that $Y_t = \mathrm{E}(1_F | \mathcal{F}_t)$ almost surely, for every $t \geq 0$. Then $\bar{X} = 1_{[0,s)} Y$ is right continuous and hence optional. It also satisfies $\bar{X}_t = \mathrm{E}(1_{[0,s) \times F} | \mathcal{F}_t)$ almost surely. The proof of (i) is complete.

To prove (ii) we apply the monotone class theorem another time, this time with $\mathcal{H}$ equal to the set of bounded, optional processes $\bar{X}$ for which there exists a predictable process $\tilde{X}$ as in (ii). Then $\mathcal{H}$ is a vector space that contains the constants. It is closed under taking bounded monotone limits, because if $\bar{X}_n = \tilde{X}_n$ on G_n and $\bar{X}_n \to \bar{X}$, then $\lim \tilde{X}_n$ must exist at least on $\cap_n G_n$ and be equal to $\bar{X}$ there. We can define $\tilde{X}$ to be $\lim \tilde{X}_n$ if this exists and 0 otherwise. Because the stochastic integral $(S, T]$ for two given stopping times S, T is predictable, $\mathcal{H}$ clearly contains all indicators of stochastic intervals $[S, T)$, $[S, T]$, $(S, T]$ and (S, T). These intervals form an intersection stable generator of the optional σ-field by Lemma 5.8. ∎

Let X be a measurable, adapted process for which there exists a sequence of stopping times $0 \leq T_n \uparrow \infty$ such that, for every $t \geq 0$ and n,

$$1_{[0, t \wedge T_n]} X \in L_2\big([0, \infty) \times \Omega, \mathcal{B}_\infty \times \mathcal{F}, \lambda \times \mathbb{P}\big). \tag{5.41}$$

By the preceding lemma there exists a predictable process $\tilde{X}$ such that $X = \tilde{X}$ almost everywhere under $\lambda \times \mathbb{P}$. Relation (5.41) remains valid if we replace X by $\tilde{X}$. Then we can define a stochastic integral $\int 1_{[0, t \wedge T_n]} \tilde{X}\, dB$ as in Definition 5.24 and the discussion following it. We define $\int_0^t X\, dB$ as the almost sure limit of these variables as $n \to \infty$.

5.42 Definition. *Given a measurable, adapted process X for which there exists a localizing sequence T_n satisfying (5.41) the stochastic integral $\int_0^t X\,dB$ is defined as the almost sure limit of the sequence of cadlag processes $t \mapsto \int 1_{[0,t\wedge T_n]}\tilde X\,dB$.*

The verification that this definition is well posed is identical to the similar verification for stochastic integrals relative to local martingales.

Condition (5.41) is exactly what is needed, but it is of interest to have a more readily verifiable condition for a process X to be a good integrand.

5.43 Lemma. *Let X be a measurable and adapted process.*
(i) If $\int_0^t X_s^2\,ds < \infty$ almost surely for every $t \geq 0$, then there exists a sequence of stopping times $0 \leq T_n \uparrow \infty$ such that (5.41) is satisfied and hence $\int_0^t X\,dB$ can be defined as a continuous local martingale.
(ii) If $\int_0^t \mathrm{E}X_s^2\,ds < \infty$, then $\int_0^t X\,dB$ can be defined as a continuous martingale in L_2.

Proof. There exists a predictable process $\tilde X$ with $X = \tilde X$ almost everywhere under $\lambda \times \mathbb{P}$. By Fubini's theorem the sections $\{t\colon X_t(\omega) \neq \tilde X_t(\omega)\}$ of the set $\{X \neq \tilde X\}$ are Lebesgue null sets for $\mathbb{P}$-almost every ω. Therefore, the conditions (i) or (ii) are also satisfied with $\tilde X$ replacing X.

Because $\tilde X$ is predictable, it is progressive. This means that $\tilde X\colon [0,t] \times \Omega \to \mathbb{R}$ is a $\mathcal{B}_t \times \mathcal{F}_t$-measurable map and so is $\tilde X^2$. Consequently, by the measurability part of Fubini's theorem, the map $\omega \mapsto Y_t(\omega) := \int_0^t \tilde X_s^2(\omega)\,ds$ is $\mathcal{F}_t$-measurable for every $t \geq 0$, which means that the process Y is adapted. The variables $T_n = \inf\{t \geq 0\colon Y_t \geq n\}$ are stopping times, with $0 \leq T_n \uparrow \infty$ on the event where Y_t is finite for every t, by the continuity of the sample paths of Y. This is a set of probability one by assumption (i), and hence we can redefine T_n such that $0 \leq T_n \uparrow \infty$ everywhere. Furthermore,

$$\int 1_{[0,t\wedge T_n]}\tilde X^2\,d(\lambda \times \mathbb{P}) = \mathrm{E}Y_{T_n\wedge t} \leq n.$$

Thus the process $\tilde X$ satisfies (5.41), concluding the proof of (i).

For (ii) it suffices to prove that $1_{[0,t]}X \in L_2\bigl([0,\infty)\times\Omega, \mathcal{B}_\infty \times \mathcal{F}, \lambda \times \mathbb{P}\bigr)$ for every $t \geq 0$. Then the same is true for $\tilde X$, and the result follows from Theorem 5.25(iii). (The localization applied in Definition 5.42 is unnecessary in this situation. Equivalently, we can put $T_n \equiv \infty$.) But $\int 1_{[0,t]}X^2\,d\lambda \times \mathbb{P} = \int_0^t \mathrm{E}X_s^2\,ds$, by Fubini's theorem. ∎

5.6 Martingales of Bounded Variation

We recall that the *variation* of a cadlag function $A\colon \mathbb{R} \to \mathbb{R}$ over the interval $(a, b]$ is defined as

$$\int_a^b |dA_s| := \sup_{a=t_0<t_1<\cdots<t_k=b} \sum_{i=1}^k |A_{t_i} - A_{t_{i-1}}|,$$

where the supremum is taken over all partitions $a = t_0 < t_1 < \cdots < t_k = b$ of the interval. The function A is called of "locally bounded variation" if its variation over every compact interval is finite. It can be shown that this is equivalent to the existence of two nondecreasing cadlag functions A_1 and A_2 such that $A = A_1 - A_2$. Thus every function of locally bounded variation defines a signed measure $B \mapsto \int_B dA$ defined as the difference of the measures defined by the functions A_1 and A_2. It can be shown that there is a unique decomposition, written as $A = A^+ - A^-$, such that the measures defined by A^+ and A^- are orthogonal. The sum of the corresponding measures is denoted $|A| = A^+ + A^-$ and is called the *total variation* of A. It can be shown that $\int_{(a,b]} d|A|_s$ is equal to the variation over $(a, b]$ as defined in the preceding display. In particular, the expressions

$$\int_a^b |dA_s|, \qquad \text{and} \qquad \int_{(a,b]} d|A|_s$$

denote the same number.

If the sample paths of the martingale M are of bounded variation, then we can define an integral $\int X\,dM$ based on the usual Lebesgue-Stieltjes integral. Specifically, if for a given $\omega \in \Omega$ the variation $\int |dM_t|(\omega)$ of the function $t \mapsto M_t(\omega)$ is finite, then $B \mapsto \int_B dM_t(\omega)$ defines a signed measure on the Borel sets (a difference of two ordinary measures) and hence we can define an integral

$$\int X_t(\omega)\,dM_t(\omega)$$

for every process X and ω such that the function $t \mapsto X_t(\omega)$ is Borel measurable and integrable relative to the measure $B \mapsto \int_B d|M_t|(\omega)$. (All integrals are relative to t, for fixed ω, although this is not easily seen from the notation.)

If this is true for every ω, then we have two candidates for the integral $\int X\,dM$, the "pathwise" Lebesgue-Stieltjes integral and the stochastic integral. These better be the same. Under some conditions they are indeed. For clarity of the following theorem we denote the Lebesgue-Stieltjes integral by $\int X_s\,dM_s$ and the stochastic integral by $\int X\,dM$.

A process X is said to be *locally bounded* if there exists a sequence of stopping times $0 \leq T_n \uparrow \infty$ such that X^{T_n} is uniformly bounded on $[0, \infty) \times \Omega$, for every n. A process X is said to be *of locally bounded variation* if there exists a sequence of stopping times $0 \leq T_n \uparrow \infty$ such that every of the sample paths of X^{T_n} is of bounded variation on $[0,\infty)$, for every n. This can be seen to be identical to the variation of every sample path of X on every compact interval $[0,t]$ being finite, which property is well described as *locally of bounded variation.*

Warning. "Locally bounded" is defined to mean "locally uniformly bounded". This appears to be stronger than existence of a localizing sequence such that each of the sample paths of every of the stopped processes is bounded. On the other hand, "locally of bounded variation" is to be understood in a nonuniform way; it is weaker than existence of a sequence of stopping times such that all sample paths of X^{T_n} are of variation bounded by a fixed constant, depending only on n.

5.44 Theorem. *Let M be a cadlag local L_2-martingale of locally bounded variation, and let X be a locally bounded predictable process. Then for every $t \geq 0$ the stochastic integral $\int_0^t X\,dM$ and the Lebesgue-Stieltjes integral $\int_{(0,t]} X_s\,dM_s$ are both well defined and agree almost surely.*

Proof. If X is a measurable process, then the Lebesgue-Stieltjes integral $\int X_s\,dM_s$ is well defined (up to integrability of the integrand $s \mapsto X_s$ relative to the measure $s \mapsto |M_s|$), because the map $t \mapsto X_t(\omega)$ is measurable for every ω. The integral $\int X_s\,dM_s$ is then also measurable as a map on Ω. This is clear if X is the indicator function of a product set in $[0,\infty)\times\Omega$. Next we can see it for a general X by an application of the monotone class theorem, Theorem 1.23.

By assumption there exist sequences of stopping times $0 \leq T_n \uparrow \infty$ such that M^{T_n} is an L_2-martingale and such that X^{T_n} is uniformly bounded, for every n. It is not a loss of generality to choose these two sequences the same; otherwise we use the minimum of the two sequences. We may also assume that M^{T_n} is L_2-bounded. If not, then we replace T_n by $T_n \wedge n$; the martingale $M^{T_n\wedge n}$ is bounded in L_2, because $\mathrm{E}M^2_{T_n\wedge t\wedge n} \leq \mathrm{E}(M^{T_n})^2_n < \infty$ for all $t \geq 0$, by the submartingale property of $(M^{T_n})^2$.

The process $1_{[0,T_n]}X$ is uniformly bounded and hence is contained in the Hilbert space $L_2\bigl([0,\infty)\times\Omega, \mathcal{P}, \mu_{M^{T_n}}\bigr)$. Therefore, the stochastic integral $\int_0^t X\,dM$ is well defined according to Definition 5.24 as the almost sure limit of the sequence $\int 1_{[0,T_n\wedge t]}X\,dM^{T_n}$.

Because $\int_{(0,t]}|dM_s|$ is finite for every t, and the process $1_{[0,t]}X$ is uniformly bounded on the event $A_n = \{t \leq T_n\}$, the Lebesgue-Stieltjes integral $\int_{(0,t]}|X_s|\,|dM_s|$ if finite on this event, and hence almost surely on $\Omega = \cup_n A_n$, for every given t. We conclude that $\int_{(0,t]} X_s\,dM_s$ is well defined and finite, almost surely. By dominated convergence it is the limit as $n \to \infty$ of the sequence $\int 1_{(0,T_n\wedge t]}(s)X_s\,dM_s$, almost surely.

We conclude that it suffices to show that $\int 1_{[0,T_n\wedge t]}X\,dM^{T_n}$ and $\int 1_{(0,T_n\wedge t]}(s)X_s\,dM_s$ agree almost surely, for every n. For simplicity of notation, we drop the localization and prove that for any L_2-bounded martingale M with $\int |dM_s| < \infty$ almost surely, and every bounded, predictable process X the stochastic integral $\int X\,dM$ and Lebesgue-Stieltjes integral $\int X_s\,dM_s$ are the same almost surely, where we interprete the mass that $s \mapsto M_s$ puts at 0 to be zero.

We apply the monotone class theorem, with $\mathcal{H}$ the set of all bounded predictable X for which the integrals agree almost surely. Then $\mathcal{H}$ contains all indicators of predictable rectangles, because both integrals agree with the Riemann-Stieltjes integral for such integrands. Because both integrals are linear, $\mathcal{H}$ is a vector space. Because $\mu_M\bigl([0,\infty)\times\Omega\bigr) = \mathrm{E}(M_\infty - M_0)^2 < \infty$, the Doléans measure of M is finite, and hence the constant functions are integrable. If $0 \le X_n \uparrow X$ for a bounded X and $\{X_n\} \subset \mathcal{H}$, then $X_n \to X$ in $L_2\bigl([0,\infty)\times\Omega, \mathcal{P}, \mu_M\bigr)$ by the dominated convergence theorem, and hence $\int X_n\,dM \to \int X\,dM$ in L_2. Furthermore, $\int X_{n,s}\,dM_s \to \int X_s\,dM_s$ pointwise on Ω, by the dominated convergence theorem, because $\int |dM_s| < \infty$. Because L_2-limits and pointwise limits must agree, it follows that the two integrals agree almost surely. The unit function is a limit of a sequence of indicators of predictable rectangles and hence we can infer that the constant functions are in $\mathcal{H}$. Next an application of Theorem 1.23 shows that $\mathcal{H}$ contains all bounded predictable processes. ∎

As a corollary of the preceding theorem we see that the Lebesgue-Stieltjes integral of a locally bounded predictable process relative to a cadlag local L_2-martingale of locally bounded variation is a local martingale. Indeed, under these conditions the two types of integrals coincide and the stochastic integral is a local martingale. In the next section we want to drop the "L_2" from the conditions and for this reason we now give a direct proof of this martingale property for integrators that are only local martingales.

5.45 Lemma. *If M is a cadlag local martingale of locally bounded variation and X is a locally bounded predictable process, then the Lebesgue-Stieltjes integrals $(X\cdot M)_t{:}= \int_{(0,t]} X_s\,dM_s$ define a cadlag local martingale $X\cdot M$.*

Proof. Write $\int_0^t$ for $\int_{(0,t]}$. Let $0 \le T_n \uparrow \infty$ be a sequence of stopping times such that M^{T_n} is a martingale and such that X^{T_n} is uniformly bounded, for every n. Because

$$(X\cdot M)_t^{T_n} = \int_0^t X_s^{T_n}\,dM_s^{T_n},$$

the lemma will follow if $t \mapsto \int_0^t X_s\,dM_s$ is a cadlag martingale for every given pair of a bounded predictable process X and martingale of locally bounded variation M.

This is clear if X is the indicator of a predictable rectangle. In that case the Lebesgue-Stieltjes integral is a Riemann-Stieltjes integral, and coincides with the elementary stochastic integral, which is a martingale. The set $\mathcal{H}$ of all bounded predictable X for which $X \cdot M$ is a martingale is a vector space and contains the constants. If $0 \leq X_n \uparrow X$ for a uniformly bounded process X, then $\int_0^t X_{n,s}\,dM_s \to \int_0^t X_s\,dM_s$ pointwise on Ω and in L_1, for every $t \geq 0$, by two applications of the dominated convergence theorem. We conclude that the set $\mathcal{H}$ is closed under bounded monotone limits. It contains the constants and is a generator of the predictable processes. Hence $\mathcal{H}$ contains all bounded predictable processes, by the monotone class theorem, Theorem 1.23. ∎

Warning. The predictability of the integrand is important. For instance, if N is a standard Poisson process and T is the time of its first jump, then the process M defined by $M_t = N_t - t$ and the process M^T are martingales. The Lebesgue-Stieltjes integral $\int_0^t N_s\,dM_s^T = 1_{t\geq T} N_T = 1_{t\geq T}$ is certainly not a martingale (as can be seen from the fact that $\mathrm{E}1\{t \geq T\} = 1 - e^{-t}$ is not constant) and hence this Lebesgue-Stieltjes integral lacks the most striking property of the stochastic integral. In comparison N_- is a predictable process and $\int_0^t N_{s-}\,dM_s^T = 0$ is certainly a martingale.

Warning. The mere existence of the Lebesgue-Stieltjes integrals (i.e. $\int_0^t |X_s|\,|dM_s| < \infty$ almost surely) is not enough to render $\int_0^t X_s\,dM_s$ a local martingale. For example, let S and T be independent exponential variables and $M = 1_{[T,\infty)}\bigl(1_{T\leq S} - 1_{T>S}\bigr)$?? The process M is a martingale relative to its natural filtration, and the LS-integral $Y_t = \int_0^t s^{-1}\,dM_s$ is well defined, but $\mathrm{E}|Y_U| = \infty$ for every stopping time U which is not identically zero, so that Y is not a local martingale.

The most important example of a continuous martingale is Brownian motion and this has sample paths of unbounded variation. The latter property is not special to Brownian motion, but is shared by all continuous martingales, or more generally all predictable local martingales. We can prove this important and interesting result by a comparison of stochastic and Lebesgue-Stieltjes integrals.

5.46 Theorem. *Let M be a cadlag predictable process that is both a local martingale and a process of locally bounded variation, and 0 at 0. Then $M = 0$ up to indistinguishability.*

Proof. First assume that M is continuous. By assumption there exists a sequence $0 \leq T_n \uparrow \infty$ of stopping times such that M^{T_n} is both a martingale and of bounded variation. If necessary we can replace T_n by the minimum of T_n and $\inf\{t \geq 0\colon |M_t| \geq n\}$ to ensure also that M^{T_n} is bounded, and hence in L_2. Because M^{T_n} is of bounded variation, the integration by parts formula for Lebesgue-Stieltjes integrals yields (with $\int_0^t$ denoting $\int_{(0,t]}$)

$$(M^{T_n})_t^2 = \int_0^t M_-^{T_n}\,dM^{T_n} + \int_0^t M^{T_n}\,dM^{T_n}.$$

Under the present assumption that M is continuous, the integrands in these integrals are continuous and hence predictable. (The two integrals are also identical, but we write them differently because the identity is valid even for discontinuous M, and we need it in the second part of the proof.) Therefore, the integrals on the right can be viewed equivalently as Lebesgue-Stieltjes or stochastic integrals, by Theorem 5.44. The interpretation as stochastic integrals shows that the right side is a martingale. This implies that $\mathrm{E}M^2_{T_n\wedge t} = 0$ and hence $M_t = 0$ almost surely, for every t.

The proof if M is not continuous is similar, but requires additional steps, and should be skipped at first reading. A stopped predictable process is automatically predictable. (This is easy to verify for indicators of predictable rectangles and next can be extended to general predictable processes by a monotone class argument.) Therefore, the integrands in the preceding display are predictable also if M is not continuous. On the other hand, if M is not continuous, then M^{T_n} as constructed previously is not necessarily bounded and we cannot apply Theorem 5.44 to conclude that the Lebesgue-Stieltjes integral $\int_0^t M^{T_n}\,dM^{T_n}$ is a martingale. We can solve this by "stopping earlier", if necessary. The stopping time $S_n = \inf\{t \ge 0\colon |M_t| \ge n\}$ is predictable, as $[S_n] = [0, S_n] \cap M^{-1}([-n,n]^c)$ is predictable. (See the last assertion of Lemma 5.7.) Thus S_n is the monotone limit of a sequence of stopping times $\{S_{m,n}\}_{m=1}^\infty$ strictly smaller than S_n on $\{S_n > 0\} = \Omega$. Then $R_n = \max_{i,j\le n} S_{i,j}$ defines a sequence of stopping times with $0 \le R_n \uparrow \infty$ and $|M^{R_n}| \le n$ for every n by the definition of S_n and the fact that $R_n < S_n$. Now we may replace the original sequence of stopping times T_n by the minimum of T_n and R_n, and conclude the argument as before. ∎

5.47 EXERCISE. Show that a cadlag predicatble process is locally bounded. [See the preceding proof.]

5.7 Semimartingales

The ultimate generalization of the stochastic integral uses "semimartingales" as integrators. Because these are defined as sums of local martingales and bounded variation processes, this does not add much to what we have already in place. However, the concept of a semimartingale does allow some unification, for instance in the statement of Itô's formula, and it is worth introducing it.

5.48 Definition. *A cadlag adapted stochastic process X is a semimartingale if it has a representation of the form $X = X_0 + M + A$ for a cadlag local* *martingale M and a cadlag adapted process of locally bounded variation A.*

The representation $X = X_0 + M + A$ of a semimartingale is nonunique. It helps to require that $M_0 = A_0 = 0$, but this does not resolve the nonuniqueness. This arises because there exist martingales that are locally of bounded variation. The compensated Poisson process is a simple example.

We would like to define a stochastic integral $Y \cdot X$ as $Y \cdot M + Y \cdot A$, where the first integral $Y \cdot M$ is a stochastic integral and the second integral $Y \cdot A$ can be interpreted as a Lebesgue-Stieltjes integral. If we restrict the integrand Y to locally bounded, predictable processes, then $Y \cdot M$ is defined as soon as M is a local L_2-martingale, by Definition 5.31 and Example 5.39. In the given decomposition $X = X_0 + M + A$, the martingale is not required to be locally in L_2, but one can always achieve this by proper choice of M and A, in view of the following lemma. The proof of this lemma is long and difficult and should be skipped at first reading. It suffices to remember that "local martingale" in the preceding definition may be read as "local L_2-martingale", without any consequence; and that a continuous semimartingale can be decomposed into continuous processes M and A. The latter means that a continuous semimartingale can equivalently be defined as a process that is the sum of a continuous local martingale and a continuous adapted process of locally bounded variation.

5.49 Lemma. *For any cadlag semimartingale X there exists a decomposition $X = X_0 + M + A$ such that M is a cadlag local L_2-martingale and A is a cadlag adapted process of locally bounded variation. Furthermore, if X is continuous, then M and A can be chosen continuous.*

* **Proof.** We may without loss of generality assume that X is a local martingale. Define a process Z by $Z_t = \sum_{s\le t} \Delta X_s 1_{|\Delta X_s|>1}$. This is well defined, because a cadlag function can have at most finitely many jumps of absolute size bigger than some fixed constant on any given compact interval. We show below that there exists a cadlag predictable process B of locally bounded variation such that $Z-B$ is a local martingale. Next we set $A = Z - B$ and $M = X - X_0 - A$ and show that $|\Delta M| \le 2$. Then M is a locally bounded martingale and hence certainly a local L_2-martingale, and hence the first assertion of the lemma is proved.

In order to show the existence of the process B define a process Z^u by $Z_t^u = \sum_{s\le t} \Delta X_s 1_{\Delta X_s>1}$. This is clearly nondecreasing. We claim that it is locally in L_1 and hence a local submartingale. To see this, let $0 \le S_n \uparrow \infty$ be a sequence of stopping times such that X^{S_n} is a uniformly integrable martingale, for every n, and define $T_n = \inf\{t \ge 0\colon Z_t^u > n, |X_t| > n\} \wedge S_n$. Then $Z_t^u \vee |X^{T_n}| \le n$ on $[0, T_n)$ and

$$0 \le Z_{T_n\wedge t}^u \le Z_{T_n}^u \le n + |\Delta X_{T_n}| \le 2n + |X_{T_n}|.$$

The right side is integrable by the optional stopping theorem, because $T_n \leq S_n$ and X^{S_n} is uniformly integrable.

Being a local submartingale, the process Z^u possesses a compensator B^u by the Doob-Meyer decomposition, Lemma 5.74. We can apply a similar argument to the process of cumulative jumps of X less than -1, and take differences to construct a process B with the required properties.

The proof that $|\Delta M| \leq 2$ is based on the following facts:

(i) For every cadlag predictable process X there exists a sequence of predictable times T_n such that $\{(t,\omega)\colon \Delta X_t(\omega) \neq 0\} = \cup_n [T_n]$. (The sequence T_n is said to *exhaust the jumps of* X. See e.g. Jacod and Shiryaev, I2.24.)

(ii) If X is a predictable process and T a stopping time, then X_T is $\mathcal{F}_{T-}$-measurable, where we define X_∞ to be 0. (See e.g. Jacod and Shiryaev, I2.4; and I1.11 for the definition of $\mathcal{F}_{T-}$.)

(iii) For any cadlag martingale X and predictable stopping time T we have $\mathrm{E}(X_T|\,\mathcal{F}_{T-}) = X_{T-}$ almost surely on $\{T < \infty\}$. (See e.g. Jacod and Shiryaev, I2.27.)

(iv) For any cadlag martingale X and predictable stopping time T we have $\mathrm{E}(\Delta X_T|\,\mathcal{F}_{T-}) = 0$ almost surely on the set $\{T < \infty\}$. This follows by applying (ii) to the predictable process X_- to see that X_{T-} is $\mathcal{F}_{T-}$-measurable and combining this with (iii) to compute the conditional expectation of $\Delta X_T = X_T - X_{T-}$.

The processes X, $A = Z - B$ and $M = X - X_0 - A$ are local martingales. If we can show that $|\Delta M^{T_n}| \leq 2$ for every T_n in a localizing sequence $0 \leq T_n \uparrow \infty$, then it follows that $|\Delta M| \leq 2$ and the proof is complete. For simplicity of notation assume that X, M and A are martingales. The process Z has been constructed so that $|\Delta(X - Z)| \leq 1$ and hence $\bigl|\mathrm{E}\bigl(\Delta(X-Z)_T|\,\mathcal{F}_{T-}\bigr)\bigr| \leq 1$ almost surely, for every stopping time T. By (iv) $\mathrm{E}(\Delta M_T|\,\mathcal{F}_{T-}) = 0$ almost surely on $\{T < \infty\}$, for every predictable time T. Because $\Delta M = \Delta(X - Z) + \Delta B$, it follows that $\bigl|\mathrm{E}(\Delta B_T|\,\mathcal{F}_{T-})\bigr| \leq 1$ for every predictable time T. Since B and B_- are predictable, ΔB_T is $\mathcal{F}_{T-}$-measurable by (ii) and hence $|\Delta B_T| \leq 1$ almost surely. Consequently $|\Delta B| \leq 1$ by (i), and hence $|\Delta M| \leq \bigl|\Delta(X - Z)\bigr| + |\Delta B| \leq 2$.

This concludes the proof of the first assertion of the theorem. Next, we prove that a continuous semimartingale X permits a decomposition $X = X_0 + M + A$ such that M and A are continuous.

Suppose that X is continuous and let $X = X_0 + M + A$ be a given decomposition in a local L_2-martingale M and a process of locally bounded variation A. Let $0 \leq S_n \uparrow \infty$ be a sequence of stopping times such that M^{S_n} is a martingale for every n and define

$$T_n = \inf\Bigl\{t \geq 0\colon |M_t| > n, \int_0^t |dA_s| > n\Bigr\} \wedge n \wedge S_n.$$

Then the process M^{T_n} is a uniformly integrable martingale, is bounded in absolute value by n on $[0, T_n)$, and

$$\int_0^{T_n} |dA_s| = \int_{[0,T_n)} |dA_s| + |\Delta A_{T_n}| \leq n + |\Delta X_{T_n}| + |\Delta M_{T_n}| \leq 2n + |M_{T_n}|.$$

The right side is integrable by the optional stopping theorem, Theorem 4.21, whence the process A is locally integrable. We conclude that the positive and negative variation processes corresponding to A are both locally integrable. Because they are nondecreasing, they are submartingales, and permit Doob-Meyer decompositions as in Lemma 5.74. We conclude that there exists a cadlag predictable process $\bar{A}$ such that $A - \bar{A}$ is a local martingale. Now $X = X_0 + (M + A - \bar{A}) + \bar{A}$ is a decomposition of X into a local martingale $\bar{M} = M + A - \bar{A}$ and a predictable process of locally bounded variation $\bar{A}$. We shall show that these processes are necessarily continuous.

By predictability the variable $\Delta\bar{A}_T$ is $\mathcal{F}_{T-}$-measurable for every stopping T, by (ii). If M is integrable, then $\mathrm{E}(\Delta M_T|\, \mathcal{F}_{T-}) = 0$ for every predictable stopping time, by (iv) because M is a martingale. Since $\Delta X = 0$, it then follows that $\Delta\bar{A}_T = \mathrm{E}(\Delta\bar{A}_T|\, \mathcal{F}_{T-}) = 0$ for every predictable time T, whence the process $\bar{A}$ and hence M are continuous, by (i). If ΔM_T is not integrable, we can first localize the processes and apply the argument to stopped processes. ∎

5.50 Definition. *The integral $Y \cdot X$ of a locally bounded, predictable process Y relative to a cadlag semimartingale X with decomposition $X = X_0 + M + A$ as in Lemma 5.49 is defined as $Y \cdot M + Y \cdot A$, where the first integral $Y \cdot M$ is a stochastic integral defined according to Definition 5.31 and the second integral $Y \cdot A$ is a Lebesgue-Stieltjes integral.*

We use the notations $(Y \cdot X)_t$, $\int_0^t Y\, dX$ or $\int_{[0,t]} Y\, dX$, for the stochastic integral interchangeably. To stress that an integral is to be understood as a Lebesgue-Stieltjes integral we indicate the integration variable explicitly, as in $\int Y_s\, dA_s$.

Warning. Rather than "the integral" as in the definition, it is customary to say "the stochastic integral", in particular in the mixed case. This is reasonable terminology if we are integrating stochastic processes. In this general sense the stochastic integral may well be a Lebesgue-Stieltjes integral!

Because the decomposition of Lemma 5.49 is not unique, we must verify that the preceding definition is well posed. This follows from the fact that for any other decomposition $X = X_0 + \bar{M} + \bar{A}$ as in Lemma 5.49 the process $M - \bar{M} = \bar{A} - A$ is a cadlag local L_2-martingale that is locally of bounded variation. Therefore, the Lebesgue-Stieltjes integral and the stochastic integral of a locally bounded predictable process Y relative to this process coincide by Theorem 5.44 and hence $Y \cdot M + Y \cdot A = Y \cdot \bar{M} + Y \cdot \bar{A}$, if the integrals $Y \cdot M$, $Y \cdot A$, $Y \cdot \bar{M}$, and $Y \cdot \bar{A}$ are interpreted as stochastic or Lebesgue-Stieltjes integrals, as in the definition.

5.51 EXERCISE. Suppose that M is a cadlag local martingale that is locally of bounded variation, and Y is a locally bounded process. Show that the integral $Y \cdot M$ as defined by the preceding definition coincides with the Lebesgue-Stieltjes integral $\int_0^t Y_s \, dM_s$. [Hint: this is a trivial consequence of the fact that the definition is well posed. Don't be confused by the fact that M is a martingale.]

5.52 Theorem. *If X is a cadlag semimartingale and Y is a predictable locally bounded process, then:*
(i) There exists a cadlag version of $Y \cdot X$.
(ii) This version is a semimartingale.
(iii) If X is a local martingale, then this version is a local martingale.
(iv) If X is continuous, then there exists a continuous version of $Y \cdot X$.
(v) The processes $\Delta(Y \cdot X)$, where $Y \cdot X$ is a cadlag version, and $Y\Delta X$ are indistinguishable.

Proof. Let $X = X_0 + M + A$ be an arbitrary decomposition in a cadlag local L_2-martingale M and a cadlag adapted process of locally bounded variation A. By definition $Y \cdot X = Y \cdot M + Y \cdot A$, where the first is a stochastic integral and the second a Lebesgue-Stieltjes integral. By Theorem 5.35 the stochastic integral $Y \cdot M$ permits a cadlag version and this is a local L_2-martingale; it permits a continuous version if M is continuous; and its jump process is $Y\Delta M$. The Lebesgue-Stieltjes integral $Y \cdot A$ is of locally bounded variation and cadlag; it is continuous if A is continuous; it is a local martingale if A is a local martingale, by Lemma 5.45; and its jump process is $Y\Delta A$. Finally, if X is continuous, then the processes M and A can be chosen continuous. ∎

Now that we have completely dressed up the definition of the stochastic integral, it is useful to summarize some properties. The first is a kind of dominated convergence theorem for stochastic integrals. The localization technique is very helpful here.

5.53 Lemma. *If X is a cadlag semimartingale, and Y_n is a sequence of predictable processes such that $Y_n \to Y$ pointwise on $[0, \infty) \times \Omega$ and $|Y_n| \le K$ for a locally bounded predictable process K and every n, then the cadlag versions of $Y_n \cdot X$ and $Y \cdot X$ satisfy $\sup_{s \le t} \bigl|(Y_n \cdot X)_s - (Y \cdot X)_s\bigr| \xrightarrow{\mathrm{P}} 0$, for every $t \ge 0$.*

Proof. We can decompose $X = X_0 + M + A$ for a cadlag local L_2-martingale M and a cadlag process of locally bounded variation A, 0 at 0.

That A is of locally bounded variation implies that $\int_0^t |dA_s|(\omega) < \infty$ and that K is locally bounded implies that $\sup_{s \le t} K_s(\omega) < \infty$, both for every fixed ω and t. It follows that $\int_0^t K_s(\omega) \, |dA_s|(\omega) < \infty$ (the integral is relative to s). Because, for fixed ω, the map $s \mapsto Y_{n,s}(\omega)$ is dominated by the map $s \mapsto K_s(\omega)$, the dominated convergence theorem implies that $\int_0^t \bigl|Y_{n,s}(\omega) - Y_s(\omega)\bigr|\,|dA_s|(\omega) \to 0$. This being true for every ω, we conclude that $\sup_{s\le t}\bigl|(Y_n\cdot A)_s - (Y\cdot A)_s\bigr|$ converges to zero almost surely and hence in probability.

There exists a sequence of stopping times $0 \le T_m \uparrow \infty$ such that M^{T_m} is an L_2-bounded martingale and K^{T_m} is a uniformly bounded process, for every m. Then, because $1_{[0,T_m\wedge t]}Y_n$ is bounded by K^{T_m}, the dominated convergence theorem yields, as $n\to\infty$, for every fixed m,

$$\int \bigl(1_{[0,T_m\wedge t]}Y_n - 1_{[0,T_m\wedge t]}Y\bigr)^2\,d\mu_{M^{T_m}} \to 0.$$

On the set $[0,T_m]$ the stochastic integral $Y_n\cdot M$ can be defined as $s \mapsto \int 1_{[0,T_m\wedge s]}Y_n\,dM^{T_m}$, and similarly with Y instead of Y_n. (See Lemma 5.32 or the proof of Theorem 5.35.) For the cadlag versions of these processes, the maximal inequality (4.39) yields, for every fixed m, as $n\to\infty$,

$$\begin{aligned}
\mathrm{E}\sup_{s\le t}\bigl|(Y_n\cdot M)_s - (Y\cdot M)_s\bigr|^2 1_{t\le T_m}&\\
&\le \mathrm{E}\sup_{s\le t}\Bigl|\int 1_{[0,T_m\wedge s]}(Y_n - Y)\,dM^{T_m}\Bigr|^2,\\
&\le 4\mathrm{E}\Bigl|\int 1_{[0,T_m\wedge t]}(Y_n - Y)\,dM^{T_m}\Bigr|^2 \to 0,
\end{aligned}$$

by the isometry and the preceding display. This being true for every m implies that $\sup_{s\le t}\bigl|(Y_n\cdot M)_s - (Y\cdot M)_s\bigr|$ converges to zero in probability. ∎

5.54 EXERCISE. Show that every left-continuous adapted process that is 0 at 0 is locally bounded.

5.55 Lemma. *For every locally bounded predictable processes X and Y, cadlag semimartingale Z and stopping time T, up to indistinguishability:*
(i) $(Y\cdot Z)^T = Y\cdot Z^T = (1_{[0,T]}Y)\cdot Z$.
(ii) $X\cdot(Y\cdot Z) = (XY)\cdot Z$.
(iii) $\Delta(Y\cdot Z) = Y\Delta Z$, *if* $Y\cdot Z$ *is chosen cadlag.*
(iv) $(V1_{(S,T]})\cdot Z = V(1_{(S,T]}\cdot Z)$ *for every* $\mathcal{F}_S$*-measurable random variable* V.

Proof. The statements follow from the similar statements on stochastic integrals, properties of Lebesgue-Stieltjes integrals, and localization arguments. We omit the details. ∎

5.8 Quadratic Variation

To every semimartingale or local L_2-martingale X correspond processes $[X]$ and $\langle X\rangle$, which play an important role in stochastic calculus. They are known as the "quadratic variation process" and "predictable quadratic variation process", and are also referred to as the *square bracket process* and the *angle bracket process.* In this section we discuss the first of the two. In the next section we shall see that the two processes are the same for continuous L_2-local martingales.

5.56 Definition. *The quadratic covariation of two cadlag semimartingales X and Y is a cadlag version of the process*

$$[X,Y] = XY - X_0Y_0 - X_- \cdot Y - Y_- \cdot X. \tag{5.57}$$

The process $[X,X]$, abbreviated to $[X]$, is called the quadratic variation of X.

As usual we need to check that the definition is well posed. In this case this concerns the semimartingale integrals $X_- \cdot Y$ and $Y_- \cdot X$; these are well defined by Definition 5.50, because a left-continuous adapted process that is 0 at 0 (such as X_- and Y_-) is predictable and locally bounded.

We refer to the formula (5.57) as the *integration-by-parts formula.* The ordinary integration-by-parts formula for processes X and Y of locally bounded variation, from Lebesgue-Stieltjes theory, asserts that

$$X_tY_t - X_0Y_0 = \int_{(0,t]} X_-\,dY + \int_{(0,t]} Y_-\,dX + \sum_{0<s\le t} \Delta X_s \Delta Y_s.$$

Comparing this to (5.57) we see that in this case the quadratic variation $[X,Y]$ is the last term on the right. (Cf. Example 5.67 for more details.) One way of looking at the quadratic variation process for general semimartingales is to view it as the process that "makes the integration-by-parts formula true". Many semimartingales are not locally of bounded variation, and then the quadratic covariation does not reduce to a function of the jump processes, as in the preceding display. In particular, the quadratic covariation of a continuous semimartingale is typically nonzero.

The name "quadratic covariation" is better explained by the following theorem, which may also be viewed as an alternative definition of this process.

5.58 Theorem. *For any pair of cadlag semimartingales X and Y, any sequence of partitions $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ of mesh widths tending to zero, and any $t \ge 0$, as $n \to \infty$,*

$$\sum_{i=1}^{k_n} (X_{t_i^n} - X_{t_{i-1}^n})(Y_{t_i^n} - Y_{t_{i-1}^n}) \xrightarrow{\mathrm{P}} [X,Y]_t. \tag{5.59}$$

Proof. Because $4xy = (x+y)^2 - (x-y)^2$ for any numbers x, y, the case of two semimartingales X and Y can be reduced to the case that $X = Y$. For simplicity of notation we only consider the latter case. By the identity $(y-x)^2 = y^2 - x^2 - 2x(y-x)$ we can write

$$\sum_{i=1}^{k_n}(X_{t_i^n} - X_{t_{i-1}^n})^2 = \sum_{i=1}^{k_n}(X_{t_i^n}^2 - X_{t_{i-1}^n}^2) - 2\sum_{i=1}^{k_n} X_{t_{i-1}^n}(X_{t_i^n} - X_{t_{i-1}^n})$$

$$= X_t^2 - X_0^2 - 2(X_n \cdot X)_t, \tag{5.60}$$

for X_n the simple predictable process defined by

$$X_n = \sum_{i=1}^{k_n} X_{t_{i-1}^n} 1_{(t_{i-1}^n, t_i^n]}.$$

The sequence of processes X_n converges pointwise on $[0,t]\times\Omega$ to the process X_- (where $X_{0-} = 0$). The process K defined by $K_t = \sup_{s\leq t} X_{s-}$ is adapted and left continuous and hence predictable and locally bounded, and it dominates X_n. Lemma 5.53 implies that the sequence $(X_n \cdot X)_t$ converges in probability to $(X_- \cdot X)_t$. ∎

5.61 Example (Brownian motion). The quadratic variation process of Brownian motion is computed in Theorem 4.28 and is given by $[B]_t = t$. This is special, because it is a deterministic process. We shall see later that Brownian motion is the only continuous local martingale with quadratic variation process the identity function.

In view of the representation in (5.57) and the continuity of Brownian motion,

$$B_t^2 = 2\int_0^t B\,dB + t.$$

Compare this to the formula $f^2(t) = 2\int_0^t f(s)\,df(s)$ (where $df(s) = f'(s)\,ds$) for a continuously differentiable function f, and be at least a little bit surprised. Itô's formula in Section 5.10 is the generalization of this result and has a similar "correction term" relative to ordinary calculus. □

5.62 EXERCISE. Show that the quadratic variation process of both the Poisson process N and the compensated Poisson process $\{N_t - t: t \geq 0\}$ is N itself. [Hint: subtraction of the smooth function t does not change the limit of the sum of squares; N is a jump process of jump sizes $1 = 1^2$.]

5.63 EXERCISE. Show that $4[X,Y] = [X+Y] - [X-Y]$.

5.64 Example (Multivariate Brownian motion). The quadratic covariation between the coordinates of a multivariate Brownian motion

$(B^1,\dots,B^d)$ is given by $[B^i,B^j]_t=\delta_{ij}t$, for $\delta_{ij}=0$ or 1 if $i=j$ or $i\neq j$ the Kronecker delta.

This can be seen in a variety of ways. For instance, the covariation between two independent martingales is zero in general. A simple proof, which makes use of the special properties of Brownian motion, is to note that $(B^i-B^j)/\sqrt{2}$ and $(B^i+B^j)/\sqrt{2}$ are both Brownian motions in their own right and hence $[B^i-B^j]=[B^i+B^j]$, whence $[B^i,B^j]=0$ by Exercise 5.63, for $i\neq j$. □

It is clear from relation (5.59) that the quadratic variation process $[X]$ can be chosen nondecreasing almost surely. By the "polarization identity" of Exercise 5.63, the quadratic covariation process $[X,Y]$ is the difference of two nondecreasing processes and hence is of locally bounded variation. The following lemma lists some further properties.

5.65 Lemma. *Let X and Y be cadlag semimartingales.*
(i) $[X^T,Y]=[X,Y]^T=[X^T,Y^T]$ *for every stopping time T.*
(ii) If X and Y are local martingales, then $XY-[X,Y]$ is a local martingale.
(iii) If X and Y are L_2-martingales, then $XY-[X,Y]$ is a martingale.
(iv) If X and Y are L_2-bounded martingales, then $[X,Y]$ is L_1-bounded.
(v) If X or Y is continuous, then $[X,Y]$ is continuous.
(vi) The processes $\Delta[X,Y]$ and $\Delta X\Delta Y$ are indistinguishable.

Proof. Assertion (i) can be proved using (5.57), or from (5.59) after verifying that this relation remains true for partitions with a random endpoint. Assertion (vi) is a consequence of assertion (v). Assertion (ii) is a consequence of the representation (5.57) of $XY-[X,Y]$ in terms of the stochastic integrals $X_-\cdot Y$ and $Y_-\cdot X$ and Theorem 5.52(iii).

For statements (iii)–(iv) it suffices to consider the case that $X=Y$.

If X is a square-integrable martingale, then the term $(X_n\cdot X)_t$ in (5.60) has mean zero by the orthogonality of the martingale increment $X_{t_i^n}-X_{t_{i-1}^n}$ to $\mathcal{F}_{t_{i-1}^n}$. Then, by Fatou's lemma and (5.60),

$$\mathrm{E}[X]_t\le\liminf_{n\to\infty}\mathrm{E}\sum_{i=1}^{k_n}(X_{t_i^n}-X_{t_{i-1}^n})^2=\mathrm{E}(X_t^2-X_0^2).$$

This proves (iv) and also that the process $[X]$ is in L_1 if X is in L_2. To see that in the latter case $X^2-[X]$ is a martingale, as claimed in (iii), it suffices to show that $X_-\cdot X$ is a martingale. By (ii) it is a local martingale. If T_n is a localizing sequence, then, by (5.57) and (i),

$$2\Big|(X_-\cdot X)_t^{T_n}\Big|=\big|X_{T_n\wedge t}^2-X_0^2-[X]_{T_n\wedge t}\big|\le X_{T_n\wedge t}^2+X_0^2+[X]_t,$$

because $[X]$ is nondecreasing. Because $X_{T_n\wedge t}=\mathrm{E}(X_t|\,\mathcal{F}_{T_n\wedge t})$ by the optional stopping theorem, Jensen's inequality yields that $X_{T_n\wedge t}^2\le$

$E(X_t^2 | \mathcal{F}_{T_n \wedge t})$ and hence the sequence $\{X_{T_n \wedge t}^2\}_{n=1}^\infty$ is uniformly integrable, for every fixed $t \geq 0$. We conclude that the right side and hence the left side of the preceding display is uniformly integrable, and the sequence of processes $(X_- \cdot X)^{T_n}$ converges in L_1 to the process $X_- \cdot X$, as $n \to \infty$. Then the martingale property of the processes $(X_- \cdot X)^{T_n}$ carries over onto the process $X_- \cdot X$. This concludes the proof of (iii).

For assertion (vi) we write $X = X_- + \Delta X$ and $Y = Y_- + \Delta Y$ to see that the jump process of the process XY is given by $\Delta(XY) = X_- \Delta Y + Y_- \Delta X + \Delta X \Delta Y$. Next we use (5.57) to see that $\Delta[X, Y] = \Delta(XY) - \Delta(X_- \cdot Y) - \Delta(Y_- \cdot X)$, and conclude by applying Lemma 5.55(iii). ∎

* **5.66** EXERCISE. Show that the square root $\sqrt{[M]}$ of the quadratic variation of a local martingale is locally integrable. [Hint: decompose $M = N + A$ for a local L_2-martingale N and a local martingale of locally bounded variation A. A local L_2-martingale is locally integrable by Lemma 5.65(iii). Let S_n be a localizing sequence such that A^{S_n} is uniformly integrable (this exists, cf. J&S I3.11) and set $T_n = S_n \wedge \inf\{t > 0[A]_t > n, |A_t| > n\}$. Then $[A]_{T_n} \leq n + |\Delta A_{T_n}|^2$ and $|\Delta A_{T_n}| \leq n + |A_{T_n}|$, which is integrable by optional stopping.]

5.67 Example (Bounded variation processes). The quadratic variation process of a cadlag semimartingale X that is locally of bounded variation is given by $[X]_t = \sum_{0<s\leq t} (\Delta X_s)^2$.

An intuitive explanation of the result is that for a process of locally bounded variation the sums of infinitesimal absolute increments converges to a finite limit. Therefore, for a continuous process of locally bounded variation the sums of infinitesimal *square* increments, as in (5.59), converges to zero. On the other hand, the squares of the jumps in the discrete part of a process of locally bounded variation remain.

The claim can be proved directly from the definition of $[X]$ as the sum of infinitesimal square increments in equation (5.59) of Theorem 5.58, but the following indirect proof is easier. The integration-by-parts formula for cadlag functions of bounded variation shows that

$$X_t^2 - X_0^2 = 2 \int_{(0,t]} X_{s-} \, dX_s + \sum_{0<s\leq t} (\Delta X_s)^2.$$

Here the integral on the right is to be understood as a pathwise Lebesgue-Stieltjes integral, and is equal to the Lebesgue-Stieltjes integral $\int_{(0,t]} X_{s-} \, d(X_s - X_0)$. Because the decomposition $X = X_0 + M + A$ of the semimartingale X can be chosen with $M = 0$ and $A = X - X_0$, the latter Lebesgue-Stieltjes integral is by definition the semimartingale integral $(X_- \cdot X)_t$, as defined in Definition 5.50. Making this identification and comparing the preceding display to (5.57) we conclude that $[X]_t = \sum_{0<s\leq t} (\Delta X_s)^2$. □

5.68 EXERCISE. Let X be a continuous semimartingale and Y a cadlag semimartingale that is locally of bounded variation. Show that $[X,Y]=0$. [Hint: one possibility is to use (5.59).]

5.69 EXERCISE. Show that for any cadlag semimartingales X and Y, almost surely:
(i) $[X,Y]^2 \le [X][Y]$.
(ii) $\sqrt{[X+Y]} \le \sqrt{[X]} + \sqrt{[Y]}$.
(iii) $\bigl|\sqrt{[X]} - \sqrt{[Y]}\bigr| \le \sqrt{[X-Y]}$.

5.70 EXERCISE. Let X be a cadlag semimartingale and Y a continuous semimartingale of locally bounded variation. Show that $[X,Y]=0$. [Hint: one possibility is to use the preceding exercise and $[Y]=0$.]

The quadratic variation process of a stochastic integral $X\cdot Y$ is the limit in probability along a sequence of partitions of

$$\sum_{i=1}^{k_n}\Bigl(\int_{t_{i-1}^n}^{t_i^n} X\,dY\Bigr)^2.$$

For sufficiently regular X and Y it is clear that the terms of the sum are close to $X_{t_{i-1}^n}^2(Y_{t_i^n}-Y_{t_{i-1}^n})^2$ and hence we can expect that $[X\cdot Y]=X^2\cdot[Y]$, where the right side is the pathwise Lebesgue-Stieltjes integral of $s\mapsto X_s^2$ with respect to the bounded variation function $[Y]$. This turns out to be true in complete generality.

5.71 Lemma. *For every locally bounded predictable process X and cadlag semimartingales Y and Z we have $[X\cdot Y,Z]=X\cdot[Y,Z]$.*

Proof. In view of Lemmas 5.55 and 5.65 if the result is true for stopped processes X^{T_n} and Y^{T_n} for a sequence of stopping times $T_n\uparrow\infty$ and every n, then it is true for X and Y. We can choose the localizing sequence such that X^{T_n} is uniformly bounded and Y^{T_n} is the sum of an L_2-bounded martingale and a process of bounded variation, for every n. For simplicity of notation we assume that X and Y themselves have these properties. Because the assertion is linear in Y, it also suffices to prove the lemma separately under the assumptions that Y is an L_2-bounded martingale or a bounded varation process.

For X equal to an indicator of a predictable rectangle $(u,v]\times F_u$ we may use (5.59) with a sequence of partitions that include $u<v$ to see that $[X\cdot Y,Z]=1_{F_u}\bigl([Y,Z]_v-[Y,Z]_u\bigr)$. This is identical to the integral $X\cdot[Y,Z]$. Thus the lemma is true for all indicators of predictable rectangles, and by linearity also for all simple predictable processes.

For any sequence of processes X_n Exercise 5.69 gives that $\big[(X_n - X)\cdot Y, Z\big]^2 \le \big[(X_n - X)\cdot Y\big][Z]$. Therefore, if the lemma holds for a sequence X_n and $[(X_n - X)\cdot Y] \xrightarrow{\mathrm{P}} 0$, then the lemma is valid for X.

For a given bounded predictable process X we choose a uniformly bounded sequence of predictable processes X_n that converge pointwise to X. If Y is an L_2-martingale, then by Lemma 5.65 the process $((X_n - X)\cdot Y)^2 - [(X_n - X)\cdot Y]$ is a martingale for every n, whence

$$\mathrm{E}[(X_n - X)\cdot Y]_t = \mathrm{E}((X_n - X)\cdot Y)_t^2 = \int 1_{[0,t]}(X_n - X)^2\, d\mu_Y \to 0,$$

by the L_2-isometry and dominated convergence. If Y has bounded variation, then $X\cdot Y$ is a Lebesgue-Stieltjes integral and for any partition $0 = t_0 < \cdots < t_k = t$

$$\begin{aligned}\sum_{i=1}^k \Big(\int_{t_{i-1}}^{t_i} (X_n - X)_s\, dY_s\Big)^2 &\le \sum_{i=1}^k \int_0^t (X_n - X)_s^2\, d|Y|_s \int_{t_{i-1}}^{t_i} d|Y|_s\\ &= \int_0^t (X_n - X)_s^2\, d|Y|_s \int_0^t d|Y|_s.\end{aligned}$$

The right side converges to zero pointwise as $n \to \infty$ by the dominated convergence theorem. The quadratic variation $[X\cdot Y]_t$ is bounded above by the supremum of the left side over all partitions, in view of (5.59), and hence also converges to zero almost surely as $n \to \infty$. ∎

5.9 Predictable Quadratic Variation

The "angle bracket process" $\langle M\rangle$ is defined for the smaller class of local L_2-martingales M, unlike the square bracket process, which is defined for general semimartingales. If M is continuous, we can define $\langle M\rangle$ simply to be identical to $[M]$. For general local L_2-martingales, we define the angle bracket process by reference to the *Doob-Meyer decomposition*. This decomposition, given in Lemma 5.74, implies that for any local L_2-martingale M there exists a unique predictable process A such that $M^2 - A$ is a local martingale. We define this process as the predictable quadratic variation of M.

5.72 Definition. *The* predictable quadratic variation *of a cadlag local L_2-martingale M is the unique cadlag nondecreasing predictable process $\langle M\rangle$, 0 at 0, such that $M^2 - \langle M\rangle$ is a local martingale. The* predictable quadratic covariation *of a pair of cadlag local L_2-martingales M and N is the process $\langle M, N\rangle$ defined by $4\langle M, N\rangle = \langle M + N\rangle - \langle M - N\rangle$.*

5.73 EXERCISE. Show that $MN - \langle M, N\rangle$ is a local martingale.

If M is a local martingale, then the process $M^2 - [M]$ is a local martingale, by Lemma 5.65(ii). Consequently, if $[M]$ is predictable, in particular if M is continuous, then $\langle M\rangle = [M]$. However, the process $[M]$ is not always predictable, and hence is not always equal to the process $\langle M\rangle$.

To see that Definition 5.72 is well posed, we use the Doob-Meyer decomposition. The square of a local martingale is a local submartingale, by Jensen's inequality, and hence existence and uniqueness of $\langle M\rangle$ follows from (ii) of the following lemma.

A process Z is said to be *of class D*, if the collection of all random variable Z_T with T ranging over all finite stopping times, is uniformly integrable.

5.74 Lemma (Doob-Meyer).

(i) Any cadlag submartingale Z of class D can be written uniquely in the form $Z = Z_0 + M + A$ for a cadlag uniformly integrable martingale M and a cadlag predictable nondecreasing process A with $\mathrm{E}A_\infty < \infty$, *both 0 at 0. The process A is continuous if and only if* $\mathrm{E}Z_{T-} = \mathrm{E}Z_T$ *for every finite predictable time T.*

(ii) Any cadlag local submartingale Z can be written uniquely in the form $Z = Z_0 + M + A$ for a cadlag local martingale M and a cadlag predictable nondecreasing process A, both 0 at 0.

Proof. For a proof of (i) see e.g. Rogers and Williams, VI-29.7 and VI-31.1. The uniqueness of the decomposition follows also from Theorem 5.46, because given two decompositions $Z = Z_0 + M + A = Z_0 + \bar M + \bar A$ of the given form, the process $M - \bar M = \bar A - A$ is a cadlag predictable process of bounded variation, 0 at 0, and hence is 0.

Given (i) we can prove (ii) by localization as follows. Suppose $0 \le T_n \uparrow \infty$ is a sequence of stopping times such that Z^{T_n} is a submartingale of class D for every n. Then by (i) it can be written as $Z^{T_n} = Z_0 + M_n + A_n$ for a uniformly integrable martingale M_n and a cadlag, nondecreasing integrable predictable process A_n. For $m \le n$ we have $Z^{T_m} = (Z^{T_n})^{T_m} = Z_0 + M_n^{T_m} + A_n^{T_m}$. By uniqueness of the decomposition it follows that $M_n^{T_m} = M_m$ and $A_n^{T_m} = A_m$. This allows us to define processes M and A in a consistent manner by specifying their values on the set $[0, T_m]$ to be M_m and A_m, for every m. Then $M^{T_n} = M_n$ and hence M is a local martingale. Also $Z = Z_0 + M + A$ on $[0, T_m]$ for every m and hence on $[0, \infty) \times \Omega$.

We still need to show the existence of the stopping times T_n. By assumption there are stopping times $0 \le S_n \uparrow \infty$ such that Z^{S_n} is a submartingale. Define

$$T_n = S_n \wedge n \wedge \inf\{t \ge 0 : |Z_t^{S_n}| \ge n\}.$$

Then $|Z_t^{S_n}| \leq |Z_{T_n}^{S_n}| \vee n$ for $t \in [0, T_n]$ and hence $|Z_T^{T_n}| \leq |Z_{T_n}^{S_n}| \vee n$ for every stopping time T. The right side is integrable because T_n is bounded and Z^{S_n} is a submartingale (and hence is in L_1) by Theorem 4.20. ■

The nondecreasing, predictable process A in the Doob-Meyer decomposition given by Lemma 5.74(i)–(ii) is called the *compensator* or "dual predictable projection" of the submartingale Z.

5.75 Example (Poisson process). The standard Poisson process is nondecreasing and integrable and hence trivially a local submartingale. The process M defined by $M_t = N_t - t$ is a martingale, and the identity function $t \mapsto t$, being a deterministic process, is certainly predictable. We conclude that the compensator of N is the identity function.

The process $t \mapsto M_t^2 - t$ is also a martingale. By the same reasoning we find that the predictable quadratic variation of M is given by $\langle M \rangle_t = t$. In contrast, the quadratic variation is $[M] = N$. (See Exercise 5.62.) □

5.76 EXERCISE. Show that the compensator of $[M]$ is given by $\langle M \rangle$.

5.77 EXERCISE. Show that $\langle M^T \rangle = \langle M \rangle^T$ for every stopping time T. [Hint: a stopped predictable process is predictable.]

* **5.78** EXERCISE. Show that $M^2 - \langle M \rangle$ is a martingale if M is an L_2-martingale. [Hint: if M is L_2-bounded, then M^2 is of class D and we can apply (i) of the Doob-Meyer lemma; a general M can be stopped.]

Both quadratic variation processes are closely related to the Doléans measure. The following lemma shows that the Doléans measure can be disintegrated as,

$$d\mu_M(s,\omega) = d[M]_s(\omega)\, d\mathbb{P}(\omega) = d\langle M \rangle_s(\omega)\, d\mathbb{P}(\omega).$$

Here $d[M]_s(\omega)$ denotes the measure on $[0,\infty)$ corresponding to the nondecreasing, cadlag function $t \mapsto [M]_s(\omega)$, for given ω, and similarly for $d\langle M \rangle_s(\omega)$. The three measures in the display agree on the predictable σ-field, where the Doléans measure was first defined. (See (5.14)). Off the predictable σ-field the two disintegrations offer possible extensions, which may be different.

5.79 Lemma. *If M is an L_2-martingale, then, for all $A \in \mathcal{P}$,*

$$\mu_M(A) = \int\int_0^\infty 1_A(s,\omega)\, d[M]_s(\omega)\, d\mathbb{P}(\omega) = \int\int_0^\infty 1_A(s,\omega)\, d\langle M \rangle_s(\omega)\, d\mathbb{P}(\omega).$$

Proof. Because the predictable rectangles form an intersection stable generator of the predictable σ-field, it suffices to verify the identity for every set of the form $A = (s,t] \times F_s$ with $F_s \in \mathcal{F}_s$. Now

$$\mathrm{E}\int_0^\infty 1_{(s,t]\times F_s}(u,\omega)\,d[M]_u = \mathrm{E}1_{F_s}\big([M]_t - [M]_s\big).$$

Because $M^2 - [M]$ is a martingale, by Lemma 5.65(iii), the variable $(M_t^2 - [M]_t) - (M_s^2 - [M]_s)$ is orthogonal to $\mathcal{F}_s$. This implies that we may replace $[M]_t - [M]_s$ in the display by $M_t^2 - M_s^2$. The resulting expression is exactly $\mu_M\big((s,t]\times F_s\big)$.

The argument for $\langle M\rangle$ is identical, if we note that $M^2 - \langle M\rangle$ is a martingale if M is in L_2. (Cf. Exercise 5.78.) ■

5.80 EXERCISE. Consider the compensated Poisson process. Show that the three measures in Lemma 5.79 are given by $ds d\mathbb{P}(\omega)$, $dN_s(\omega)d\mathbb{P}(\omega)$ and $ds d\mathbb{P}(\omega)$. Show by direct calculation that they are the same on the predictable σ-field. Are they the same on the optional s-field or the product σ-field?

5.81 Example (Integration with Continuous Integrators). We have seen in Example 5.37 that a continuous local martingale M, 0 at 0, is a local L_2-martingale, and hence can act as an integrator. It can now be seen that any predictable process X with, for every $t \geq 0$,

$$\int_0^t X_s^2\,d[M]_s < \infty, \qquad \text{a.s.}$$

is a good integrand relative to M. This is to say that under this condition there exists a localizing sequence $0 \leq T_n \uparrow \infty$ for the pair (X, M) and hence Definition 5.31 of the stochastic integral applies. An appropriate localizing sequence is

$$T_n = \inf\Big\{t \geq 0\colon |M_t| > n, \int_0^t X_s^2\,d[M]_s > n\Big\}.$$

For this sequence we have that M^{T_n} is bounded and $1_{[0,t\wedge T_n]}X$ is contained in $L_2\big([0,\infty)\times\Omega, \mathcal{P}, \mu_{M^{T_n}}\big)$ in view of Lemma 5.79, because $\int 1_{[0,T_n]}(s)X_s^2\,d[M]_s \leq n$ and hence has expectation $\int 1_{[0,T_n]}X^2\,d\mu_{M^{T_n}}$ bounded by n. □

We have seen that the square and angle bracket processes coincide for continuous local L_2-martingales. In the noncontinuous case they may differ, because the square bracket process is not necessarily predictable. The following lemma characterizes the square bracket process as a "nonpredictable compensator" of M^2 with jumps $(\Delta M)^2$.

5.82 Lemma. *The quadratic variation process of a cadlag local martingale M is the unique adapted process A of locally bounded variation, 0 at 0, such that $M^2 - A$ is a local martingale and $\Delta A = (\Delta M)^2$.*

Proof. The quadratic variation process $[M]$ possesses the listed properties, by Lemma 5.65(ii) and (vi). Given another process A with these properties the process $[M] - A$ is the difference of two local martingales and hence a local martingale. It is also of locally bounded variation and 0 at 0. Moreover, it is continuous, because $\Delta[M] = (\Delta M)^2 = \Delta A$. Theorem 5.46 shows that $[M] - A = 0$. ∎

Because the quadratic covariation process $[X, Y]$ is of locally bounded variation, integrals of the type $\int_0^t Z_s\, d[X,Y]_s$ can be defined as Lebesgue-Stieltjes integrals, for every measurable (integrable) process Z. (The s in the notation is to indicate that the integral is a Lebesgue-Stieltjes integral relative to s, for every fixed pair of sample paths of Z and $[X, Y]$.) The integrals in the following lemmas can be understood in this way.

5.83 Lemma. *Let M and N be local L_2-martingales and let X and Y be locally bounded predictable processes.*
(i) $[X \cdot M, Y \cdot N]_t = \int_0^t X_s Y_s\, d[M, N]_s$.
(ii) $\langle X \cdot M, Y \cdot N\rangle_t = \int_0^t X_s Y_s\, d\langle M, N\rangle_s$.

Proof. Assertion (i) is already proved in Lemma 5.71 (in greater generality). It is proved here in a different way in parallel to the proof of (ii). For simplicity of notation we give the proof in the case that $X = Y$ and $M = N$. Furthermore, we abbreviate the process $t \mapsto \int_0^t X_s^2\, d\langle M\rangle_s$ to $X^2 \cdot \langle M\rangle$, and define $X^2 \cdot [M]$ similarly.

Because a compensator of a local submartingale is unique, for (ii) it suffices to show that the process $X^2 \cdot \langle M\rangle$ is predictable and that the process $(X \cdot M)^2 - X^2 \cdot \langle M\rangle$ is a local martingale. Similarly, for (i) it suffices to show that the process $(X \cdot M)^2 - X^2 \cdot [M]$ is a local martingale and that $\Delta(X^2 \cdot [M]) = \big(\Delta(X \cdot M)\big)^2$.

Now any integral relative to a predictable process of locally bounded variation is predictable, as can be seen by approximation by integrals of simple integrands. Furthermore, by properties of the Lebesgue-Stieltjes integral $\Delta(X^2 \cdot [M]) = X^2\, \Delta[M] = X^2 (\Delta M)^2$, by Lemma 5.65(vi), while $\big(\Delta(X \cdot M)\big)^2 = (X \Delta M)^2$, by Lemma 5.55. We are left with showing that the processes $(X \cdot M)^2 - X^2 \cdot \langle M\rangle$ and $(X \cdot M)^2 - X^2 \cdot [M]$ are local martingales.

Suppose first that M is L_2-bounded and that X is a predictable process with $\int X^2\, d\mu_M < \infty$. Then $X \cdot M$ is an L_2-bounded martingale, and for every stopping time T, by Lemma 5.55(i),

$$\begin{aligned} \mathrm{E}(X\cdot M)_T^2 &= \mathrm{E}\Bigl(\int X1_{[0,T]}\,dM\Bigr)^2 = \int X^2 1_{[0,T]}\,d\mu_M \\ &= \mathrm{E}\int_0^T X_s^2\,d[M]_s = \mathrm{E}(X^2\cdot[M])_T, \end{aligned}$$

where we use Lemma 5.79 for the first equality on the second line of the display. We can conclude that the process $(X\cdot M)^2 - X^2\cdot[M]$ is a martingale by Lemma 4.22.

For a general local L_2-martingale we can find a sequence of stopping times $0\le T_n\uparrow\infty$ such that M^{T_n} is L_2-bounded and such that $1_{[0,T_n]}X\in L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_{M^{T_n}}\bigr)$ for every n. By the preceding argument the process $(1_{[0,T_n]}X\cdot M^{T_n})^2 - 1_{[0,T_n]}X^2\cdot[M^{T_n}]$ is a martingale for every n. But this is the process $(X\cdot M)^2 - X^2\cdot[M]$ stopped at T_n and hence this process is a local martingale.

The proof for the process $(X\cdot M)^2 - X^2\cdot\langle M\rangle$ is similar. ∎

5.84 Example (Counting processes). A *counting process* is a stochastic process N with cadlag, nondecreasing sample paths $t\mapsto N_t$ that increase by jumps of size 1 only starting from $N_0=0$. Thus N_t can be thought of as the "number of events" in the interval $[0,t]$.

If the sample paths of N are bounded on bounded intervals, then N is a process of locally bounded variation. Consequently, it is a semimartingale relative to any filtration to which it is adapted. By Example 5.67 its quadratic variation process is given by its cumulative square jump process $[N]_t=\sum_{s\le t}(\Delta N_s)^2$. Because the jump sizes are 0 or 1, the square is superfluous and hence we obtain that $[N]=N$.

For $t\ge s$ the monotonicity of the sample paths yields $\mathrm{E}(N_t|\,\mathcal{F}_s)\ge \mathrm{E}(N_s|\,\mathcal{F}_s)=N_s$, for any filtration to which N is adapted, provided the conditional expectations are defined. Thus a counting process with $\mathrm{E}N_t<\infty$ for every t is trivially a submartingale, and hence possesses a compensator A, by the Doob-Meyer decomposition, Lemma 5.74.

If the sample paths of N are bounded on bounded intervals, then the stopping times $T_n=\inf\{t\ge 0\colon N_t=n\}$ (where the infimum over the empty set is ∞) increase to infinity, and the stopped process N^{T_n} is uniformly bounded by n. Because the stopped processes are also counting processes, they are submartingales by the preceding argument. Hence N is a local submartingale and possesses a compensator, also under the weaker condition that the number of events in finite intervals is finite.

The process $N-A$ is a local L_2-martingale of locally bounded variation. By Example 5.67 its square bracket process is given by

$$[N-A]_t=\sum_{s\le t}\bigl(\Delta(N-A)_s\bigr)^2 = N_t - 2\int_{[0,t]}\Delta A_s\,dN_s + \int_{[0,t]}\Delta A_s\,dA_s.$$

The predictable quadratic variation process $\langle N-A\rangle$ is the compensator of $[N-A]$. The third term on the far right in the preceding display is predictable and hence its own compensator, whereas the compensators of the first two terms are obtained by replacing N by its compensator A. Thus we obtain

$$\langle N-A\rangle_t = A_t - \int_{[0,t]} \Delta A_s\, dA_s = \int_{[0,t]} \bigl(1 - A\{s\}\bigr)\, dA_s.$$

If A is continuous, then this reduces to A and hence in this case the predictable quadratic variation of $N-A$ is simply the compensator of N.

A *multivariate counting process* is a vector $(N_1,\ldots,N_k)$ of counting processes N_i with the property that at most one of the processes N_i can jump at every given time point: $(\Delta N_i)_t(\Delta N_j)_t = 0$ for every $i\neq j$ and every t.

An immediate consequence of the latter restriction on the jump times is that $[N_i,N_j]=0$ if $i\neq j$.

The vector $(A_1,\ldots,A_k)$ of compensators of the processes N_i, relative to a filtration to which all N_i are adapted, is called the compensator of the multivariate counting process. By similar arguments as previously we see that $[N_i-A_i, N_j-A_j] = -\Delta A_i\circ N_j - \Delta A_j\circ N_i + \Delta A_i\circ A_j$, for $i\neq j$. As before the the angle bracket process is obtained by replacing the counting processes N_i in this expression by their compensators. Consequently, the predictable cross variation process can be expressed in the compensators as $d\langle N_i,N_j\rangle = -\delta_{i,j}\Delta A_i\Delta A_j$. If the compensators are continuous, then $\langle N_i,N_j\rangle = 0$ for every $i\neq j$. □

5.85 EXERCISE. Let (M,N) be a multivariate counting process with compensator (A,B). Find the predictable covariation processes of $M+N$ and $M-N$. [Warning: the process $M+N$ is a counting process and hence the answer is obvious from the preceding; the case of $M-N$ appears to require a calculation.]

The quadratic variation process $[M,N]$ of two martingales is of locally bounded variation and hence defines a random measure on $(0,\infty)$. The following lemma, which will not be used in the remainder, bounds the total variation of this measure.

* **5.86 Lemma (Kunita-Watanabe).** *If M and N are cadlag local martingales and X and Y are predictable processes, then*

$$\Bigl(\int_s^t |d[M,N]_u|\Bigr)^2 \le \int_s^t d[M]_u \int_s^t d[N]_u, \qquad \text{a.s.},$$

$$\Bigl(\mathrm{E}\int |X_uY_u|\,|d[M,N]|_u\Bigr)^2 \le \int X^2\, d\mu_M \int Y^2\, d\mu_N.$$

Proof. For $s < t$ abbreviate $[M,N]_t - [M,N]_s$ to $[M,N]_s^t$. Let $s = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ be a sequence of partitions of $[s,t]$ of mesh widths tending to zero as $n \to \infty$. Then, by Theorem 5.58 and the Cauchy-Schwarz inequality,

$$\begin{aligned}\big|[M,N]_s^t\big|^2 &= \lim_{n\to\infty}\Big|\sum_{i=1}^{k_n}(M_{t_i^n} - M_{t_{i-1}^n})(N_{t_i^n} - N_{t_{i-1}^n})\Big|^2 \\ &\le \lim_{n\to\infty}\sum_{i=1}^{k_n}(M_{t_i^n} - M_{t_{i-1}^n})^2\sum_{i=1}^{k_n}(N_{t_i^n} - N_{t_{i-1}^n})^2 = [M]_s^t[N]_s^t.\end{aligned}$$

Here the limits may be interpreted as limits in probability, or, by choosing an appropriate subsequence of $\{n\}$, as almost sure limits. By applying this inequality to every partitioning interval (t_{i-1}, t_i) in a given partition $s = t_0 < t_1 < \cdots < t_k = t$ of $[s,t]$, we obtain

$$\sum_{i=1}^{k}\big|[M,N]_{t_{i-1}}^{t_i}\big| \le \sum_{i=1}^{k}\sqrt{[M]_{t_{i-1}}^{t_i}[N]_{t_{i-1}}^{t_i}} \le \sqrt{\sum_{i=1}^{k}[M]_{t_{i-1}}^{t_i}\sum_{i=1}^{k}[N]_{t_{i-1}}^{t_i}},$$

by the Cauchy-Schwarz inequality. The right side is exactly the square root of $\int_s^t d[M]_u \int_s^t d[N]_u$. The supremum of the left side over all partitions of the interval $[s,t]$ is $\int_s^t |d[M,N]_u|$. This concludes the proof of the first inequality in Lemma 5.86.

To prove the second assertion we first note that by the first, for any measurable processes X and Y,

$$\Big(\int |X_u||Y_u|\,|d[M,N]|_u\Big)^2 \le \int |X_u|^2\,d[M]_u \int |Y_u|^2\,d[N]_u, \qquad \text{a.s..}$$

Next we take expectations, use the Cauchy-Schwarz inequality on the right side, and finally rewrite the resulting expression in terms of the Doléans measures, as in Lemma 5.79. ■

5.10 Itô's Formula for Continuous Processes

Itô's formula is the cornerstone of stochastic calculus. In this section we present it for continuous processes, which allows some simplification. In the first statement we also keep the martingale and the bounded variation process separated, which helps to understand the essence of the formula. The formulas for general semimartingales are more symmetric, but also more complicated at first.

For a given function $f:\mathbb{R}^d \to \mathbb{R}$ write $D_i f$ for its ith partial derivative and $D_{i,j} f$ for its $(i,j)th$ second degree partial derivative.

5.87 Theorem (Itô's formula). *Let M be a continuous local martingale and A a continuous process that is locally of bounded variation. Then, for every twice continuously differentiable function $f\colon \mathbb{R}^2 \to \mathbb{R}$,*

$$f(M_t, A_t) - f(M_0, A_0) = \int_0^t D_1 f(M, A)\, dM + \int_0^t D_2 f(M_s, A_s)\, dA_s + \tfrac{1}{2} \int_0^t D_{11} f(M_s, A_s)\, d[M]_s, \qquad \text{a.s..}$$

The special feature of Itô's formula is that the martingale M gives two contributions on the right hand side (the first and third terms). These result from the linear and quadratic approximations to the function on the left. An informal explanation of the formula is as follows. For a given partition $0 = t_0 < t_1 < \cdots < t_k = t$, we can write the left side of the theorem as

$$\sum_i \bigl(f(M_{t_{i+1}}, A_{t_{i+1}}) - f(M_{t_{i+1}}, A_{t_i})\bigr) + \sum_i \bigl(f(M_{t_{i+1}}, A_{t_i}) - f(M_{t_i}, A_{t_i})\bigr)$$

$$(5.88) \qquad \approx \sum_i D_2 f(M_{t_{i+1}}, A_{t_i})(A_{t_{i+1}} - A_{t_i})$$

$$+ \sum_i D_1 f(M_{t_i}, A_{t_i})(M_{t_{i+1}} - M_{t_i}) + \tfrac{1}{2} \sum_i D_{11} f(M_{t_i}, A_{t_i})(M_{t_{i+1}} - M_{t_i})^2.$$

We have dropped the quadratic approximation involving the terms $(A_{t_{i+1}} - A_{t_i})^2$ and all higher order terms, because these should be negligible in the limit if the mesh width of the partition converges to zero. On the other hand, the quadratic approximation coming from the martingale part, the term on the far right, does give a contribution. This term is of magnitude comparable to the quadratic variation process on the left side of (5.59).

5.89 EXERCISE. Apply Theorem 5.87 to the function $f(m, a) = m^2$. Compare the result to Theorem 5.58.

If we apply Theorem 5.87 with the function $f(m, a) = g(m + a)$, then we find the formula

$$g(M_t + A_t) - g(M_0 + A_0) = \int_0^t g'(M + A)\, d(M + A) + \tfrac{1}{2} \int_0^t g''(M_s + A_s)\, d[M]_s.$$

Here $X = M + A$ is a semimartingale. The quadratic variation $[X]$ of the continuous semimartingale X can be shown to be equal to $[M]$ (See Exercise 5.92). Thus we can also write this as

$$(5.90) \qquad g(X_t) - g(X_0) = \int_0^t g'(X)\, dX + \tfrac{1}{2} \int_0^t g''(X_s)\, d[X]_s.$$

This pleasantly symmetric formula does not permit the study of transformations of pairs of processes (M, A), but this can be remedied by studying functions $g(X_{1,t},\ldots,X_{d,t})$ of several semimartingales $X_i=\{X_{i,t}: t\geq 0\}$. In the present section we restrict ourselves to continuous semimartingales. It was shown in Lemma 5.49 that the processes M and A in the decomposition $X=X_0+M+A$ of a continuous semimartingale can always be chosen continuous. The following definition is therefore consistent with the earlier definition of a semimartingale.

5.91 Definition. *A continuous semimartingale X is a process that can be written as the sum $X=X_0+M+A$ of a continuous local martingale M and a continuous process A of locally bounded variation, both 0 at 0.*

The decomposition $X=X_0+M+A$ of a continuous semimartingale in its continuous martingale and bounded variation parts M and A is unique, because a continuous local martingale that is of locally bounded variation is necessarily constant, by Theorem 5.46.

5.92 EXERCISE. Show that the quadratic variation of a continuous semimartingale $X=X_0+M+A$, as defined in (5.59), is given by $[M]$, i.e. the contributions of the bounded variation part is negligible. [Hint: the continuity of the processes is essential; one method of proof is to use that $[M,A]=0=[A]$.]

5.93 Theorem (Itô's formula). *Let $X=(X_1,\ldots,X_d)$ be a vector of continuous semimartingales. Then, for every twice continuously differentiable function $f:\mathbb{R}^d\to\mathbb{R}$,*

$$f(X_t)-f(X_0)=\sum_{i=1}^d\int_0^t D_if(X)\,dX_i+\tfrac{1}{2}\sum_{i=1}^d\sum_{j=1}^d\int_0^t D_{ij}f(X)\,d[X_i,X_j],\quad \text{a.s..}$$

Proofs. For a proof of Theorem 5.87 based directly on the Taylor approximation (5.88), see Chung and Williams, pp94–97. Here we give a proof of the more general Theorem 5.93, but following the "convention" stated by Rogers and Williams, p61: "Convention dictates that Itô's formula should only be proved for $d=1$, the general case being left as an exercise, amid bland assurances that only the notation is any more difficult."

The proof proceeds by first establishing the formula for all polynomials f and next generalization to general smooth functions by approximation. The formula is trivially true for the polynomials $f(x)=1$ and $f(x)=x$. Next we show that the formula is correct for the function fg if it is correct for the functions f and g. Because the set of functions for which it is correct is also a vector space, we then can conclude that the formula is correct for all polynomials.

An essential step in this argument is the defining equation (5.57) for the quadratic variation process, which can be viewed as the Itô formula for polynomials of degree 2 and can be written in the form

$$X_tY_t - X_0Y_0 = (X\cdot Y)_t + (Y\cdot X)_t + [X,Y]_t. \tag{5.94}$$

Then suppose that Itô's formula is correct for the functions f and g. This means that (5.90) is valid for g (as it stands) and for f in the place of g. The formula implies that the processes $f(X)$ and $g(X)$ are semimartingales. For instance, if $X = X_0 + M + A$ then the process $g(X)$ has decomposition $g(X) = g(X_0) + \bar M + \bar A$ given by

$$\bar M_t = \int_0^t g'(X)\,dM, \qquad \bar A_t = \int_0^t g'(X_s)\,dA_s + \tfrac12\int_0^t g''(X_s)\,d[X]_s.$$

In view of Exercise 5.92, the quadratic covariation $[f(X),g(X)]$ is the quadratic covariation between the martingale parts of $f(X)$ and $g(X)$, and is equal to $f'(X)g'(X)\cdot[X]$, by Lemma 5.83. Applying (5.94) with X and Y there replaced by $f(X)$ and $g(X)$, we find

$$\begin{aligned} f(X_t)g(X_t) &- f(X_0)g(X_0)\\ &= \bigl(f(X)\cdot g(X)\bigr)_t + \bigl(g(X)\cdot f(X)\bigr)_t + \bigl[f(X),g(X)\bigr]_t\\ &= \bigl(f(X)g'(X)\cdot X\bigr)_t + \tfrac12 f(X)g''(X)\cdot[X]_t\\ &+ \bigl(g(X)f'(X)\cdot X\bigr)_t + \tfrac12 g(X)f''(X)\cdot[X]_t + f'(X)g'(X)\cdot[X]_t,\end{aligned}$$

where we have used (5.90) for f and g, and the substitution formula of Lemma 5.55(ii). By regrouping the terms this can be seen to be the Itô formula for the function fg.

Finally, we extend Itô's formula to general functions f by approximation. Because f'' is continuous, there exists a sequence of polynomials f_n with $f_n''\to f''$, $f_n'\to f'$ and $f_n\to f$ pointwise on $\mathbb{R}$ and uniformly on compacta, by an extension of the Weierstrass approximation theorem. (See Lemma 5.95 below.) Then the sequences $f_n(X)$, $f_n'(X)$ and $f_n''(X)$ converge pointwise on $\Omega\times[0,\infty)$ to $f(X)$, $f'(X)$ and $f''(X)$. The proof of the theorem is complete, if we can show that all terms of Itô's formula applied with f_n converge to the corresponding terms with f instead of f_n, as $n\to\infty$. This convergence is clear for the left side of the formula. For the proof of the convergence of the integral terms, we can assume without loss of generality that the process X in the integrand satisfies $X_0 = 0$; otherwise we replace the integrand by $X1_{(0,\infty)}$.

The process $K = \sup_n|f_n'(X)|$ is predictable and is bounded on sets where $|X|$ is bounded. If $T_m = \inf\{t\ge 0\colon |X_t| > m\}$, then, as we have assumed that $X_0 = 0$, $|X|\le m$ on the set $[0,T_m]$ and hence K^{T_m} is bounded. We conclude that K is locally bounded, and hence, by Lemma 5.53, $f_n'(X)\cdot X \xrightarrow{\mathrm{P}} f'(X)\cdot X$, as $n\to\infty$.

Finally, for a fixed m on the event $\{t\le T_m\}$, the processes $s\mapsto f_n''(X)$ are uniformly bounded on $[0,t]$. On this event $\int_0^t f_n''(X_s)\,d[X]_s \to$

$\int_0^t f''(X_s)\,d[X]_s$, as $n \to \infty$, by the dominated convergence theorem, for fixed m. Because the union over m of these events is Ω, the second terms on the right in the Itô formula converge in probability. ∎

Itô's formula is easiest to remember in terms of differentials. For instance, the one-dimensional formula can be written as

$$df(X_t) = f'(X_t)\,dX_t + \tfrac{1}{2}f''(X_t)\,d[X]_t.$$

The definition of the quadratic variation process suggests to think of $[X]_t$ as $\int (dX_t)^2$. For this reason Itô's rule is sometimes informally stated as

$$df(X_t) = f'(X_t)\,dX_t + \tfrac{1}{2}f''(X_t)\,(dX_t)^2.$$

Since the quadratic variation of a Brownian motion B is given by $[B]_t = t$, a Brownian motion then satisfies $(dB_t)^2 = dt$. A further rule is that $(dB_t)(dA_t) = 0$ for a process of bounded variation A, expressing that $[B, A]_t = 0$. In particular $dB_t dt = 0$.

5.95 Lemma. *For every twice continuously differentiable function* $f\colon \mathbb{R} \to \mathbb{R}$ *there exist polynomials* $p_n\colon \mathbb{R} \to \mathbb{R}$ *such that* $\sup_{|x|\le n}\bigl|p_n^{(i)}(x) - f^{(i)}(x)\bigr| \to 0$ *as* $n \to \infty$*, for* $i = 0, 1, 2$.

Proof. For every $n \in \mathbb{N}$ the function $g_n\colon [0,1] \to \mathbb{R}$ defined by $g_n(x) = f''(xn)$ is continuous and hence by Weierstrass' theorem there exists a polynomial r_n such that the uniform distance on $[-1, 1]$ between g_n and r_n is smaller than n^{-3}. This uniform distance is identical to the uniform distance on $[-n, n]$ between f'' and the polynomial q_n defined by $q_n(x) = r_n(x/n)$. We now define p_n to be the polynomial with $p_n(0) = f(0)$, $p_n'(0) = f'(0)$ and $p_n'' = q_n$. By integration of $f'' - p_n''$ it follows that the uniform distance between f' and p_n' on $[-n, n]$ is smaller than n^{-2}, and by a second integration it follows that the uniform distance between f and p_n on $[-n, n]$ is bounded above by n^{-1}. ∎

* 5.11 Space of Square-integrable Martingales

Recall that we call a martingale M *square-integrable* if $\mathrm{E}M_t^2 < \infty$ for every $t \ge 0$ and L_2*-bounded* if $\sup_{t\ge 0}\mathrm{E}M_t^2 < \infty$. We denote the set of all cadlag L_2-bounded martingales by $\mathcal{H}^2$, and the subset of all continuous L_2-bounded martingales by $\mathcal{H}_c^2$.

By Theorem 4.10 every L_2-bounded martingale $M = \{M_t\colon t \ge 0\}$ converges almost surely and in L_2 to a "terminal variable" M_∞ and

$M_t = \mathrm{E}(M_\infty |\, \mathcal{F}_t)$ almost surely for all $t \geq 0$. If we require the martingale to be cadlag, then it is completely determined by the terminal variable (and the filtration, up to indistinguishability). This permits us to identify a martingale M with its terminal variable M_∞, and to make $\mathcal{H}^2$ into a Hilbert space, with inner product and norm

$$(M, N) = \mathrm{E} M_\infty N_\infty, \qquad \|M\| = \sqrt{\mathrm{E} M_\infty^2}.$$

The set of continuous martingales $\mathcal{H}_c^2$ is closed in $\mathcal{H}^2$ relative to this norm. This follows by the maximal inequality (4.39), which shows that $M_\infty^n \to M_\infty$ in L_2 implies the convergence of $\sup_t |M_t^n - M_t|$ in L_2, so that continuity is retained when taking limits in $\mathcal{H}^2$. We denote the orthocomplement of $\mathcal{H}_c^2$ in $\mathcal{H}^2$ by $\mathcal{H}_d^2$, so that

$$\mathcal{H}^2 = \mathcal{H}_c^2 + \mathcal{H}_d^2, \qquad \mathcal{H}_c^2 \perp \mathcal{H}_d^2.$$

The elements of $\mathcal{H}_d^2$ are referred to as the *purely discontinuous martingales* bounded in L_2.

Warning. The sample paths of a purely discontinuous martingale are not "purely discontinuous", as is clear from the fact that they are cadlag by definition. Nor is it true that they change by jumps only. The compensated Poisson process (stopped at a finite time to make it L_2-bounded) is an example of a purely discontinuous martingale. (See Example 5.98.)

5.96 EXERCISE. Show that $\|M\|^2 = \mathrm{E}[M]_\infty + \mathrm{E} M_0^2 \leq 2\|M\|^2$.

The quadratic covariation processes $[M, N]$ and $\langle M, N\rangle$ offer another method of defining two martingales to be "orthogonal": by requiring that their covariation process is zero. For the decomposition of a martingale in its continuous and purely discontinuous part this type of orthogonality is equivalent to orthogonality in the inner product $(\cdot, \cdot)$.

5.97 Lemma. *For every $M \in \mathcal{H}^2$ the following statements are equivalent.*
(i) $M \in \mathcal{H}_d^2$.
(ii) $M_0 = 0$ *almost surely and* MN *is a uniformly integrable martingale for every* $N \in \mathcal{H}_c^2$.
(iii) $M_0 = 0$ *almost surely and* MN *is a local martingale for every continuous local martingale* N.
(iv) $M_0 = 0$ *almost surely and* $[M, N] = 0$ *for every continuous local martingale* N.
(v) $M_0 = 0$ *almost surely and* $\langle M, N\rangle = 0$ *for every* $N \in \mathcal{H}_c^2$.
Furthermore, statements (iii) and (iv) are equivalent for every local martingale M.

Proof. If M and N are both in $\mathcal{H}^2$, then $|M_t N_t| \leq M_t^2 + N_t^2 \leq \sup_t (M_t^2 + N_t^2)$, which is integrable by (4.39). Consequently, the process

MN is dominated and hence uniformly integrable. If it is a local martingale, then it is automatically a martingale. Thus (iii) implies (ii). Also, that (ii) is equivalent to (v) is now immediate from the the definition of the predictable covariation. That (iv) implies (v) is a consequence of Lemma 5.65(ii) and the fact that the zero process is predictable. That (iv) implies (iii) is immediate from Lemma 5.65(ii).

(ii) $\Rightarrow$ (i). If MN is a uniformly integrable martingale, then $(M,N) \equiv \mathrm{E}M_\infty N_\infty = \mathrm{E}M_0 N_0$ and this is zero if $M_0 = 0$.

(i) $\Rightarrow$ (ii). Fix $M \in \mathcal{H}^2_d$, so that $\mathrm{E}M_\infty N_\infty = 0$ for every $N \in \mathcal{H}^2_c$. The choice $N \equiv 1_F$ for a set $F \in \mathcal{F}_0$ yields, by the martingale property of M that $\mathrm{E}M_0 1_F = \mathrm{E}M_\infty 1_F = \mathrm{E}M_\infty N_\infty = 0$. We conclude that $M_0 = 0$ almost surely.

For an arbitrary $N \in \mathcal{H}^2_c$ and an arbitrary stopping time T, the process N^T is also contained in $\mathcal{H}^2_c$ and hence, again by the martingale property of M combined with the optional stopping theorem, $\mathrm{E}M_T N_T = \mathrm{E}M_\infty N_T = \mathrm{E}M_\infty (N^T)_\infty = 0$. Thus MN is a uniformly integrable martingale by Lemma 4.22.

(i)+(ii) $\Rightarrow$ (iii). A continuous local martingale N is automatically locally L_2-bounded and hence there exists a sequence of stopping times $0 \le T_n \uparrow \infty$ such that N^{T_n} is an L_2-bounded continuous martingale, for every n. If M is purely discontinuous, then $0 = [N^{T_n}, M] = [N^{T_n}, M^{T_n}]$. Hence $(MN)^{T_n} = M^{T_n} N^{T_n}$ is a martingale by Lemma 5.65(ii), so that MN is a local martingale.

(iii) $\Rightarrow$ (iv) By Lemma 5.65(ii) the process $MN - [M,N]$ is always a local martingale. If MN is a local martingale, then $[M,N]$ is also a local martingale. The process $[M,N]$ is always locally of bounded variation. If N is continuous this process is also continuous in view of Lemma 5.65(vi). Therefore $[M,N] = 0$ by Theorem 5.46. ∎

The quadratic covariation process $[M,N]$ is defined for processes that are not necessarily L_2-bounded, or even square-integrable. It offers a way of extending the decomposition of a martingale into a continuous and a purely discontinuous part to general local martingales. A local martingale M is said to be *purely discontinuous* if $M_0 = 0$ and $[M,N] = 0$ for every continuous local martingale N. By the preceding lemma it is equivalent to say that M is purely discontinuous if and only if MN is a local martingale for every continuous local martingale N, and hence the definition agrees with the defintion given earlier in the case of L_2-bounded martingales.

5.98 Example (Bounded variation martingales). Every local martingale that is of locally bounded variation is purely discontinuous.

To see this, note that if N is a continuous process, 0 at 0, then $\max_i |N_{t_i^n} - N_{t_{i-1}^n}| \to 0$ almost surely, for every sequence of partitions as in Theorem 5.58. If M is a process whose sample paths are of bounded variation on compacta, it follows that the left side in the definition (5.59) of the quadratic covariation process converges to zero, almost surely. Thus $[M,N]=0$ and MN is a local martingale by Lemma 5.65(ii). □

The definition of $\mathcal{H}^2_d$ as the orthocomplement of $\mathcal{H}^2_c$ combined with the projection theorem in Hilbert spaces shows that any L_2-bounded martingale M can be written uniquely as $M = M^c + M^d$ for $M^c \in \mathcal{H}^2_c$ and $M^d \in \mathcal{H}^2_d$. This decomposition can be extended to local martingales, using the extended definition of orthogonality.

5.99 Lemma. *Any cadlag local martingale M possesses a unique decomposition $M = M_0 + M^c + M^d$ into a continuous local martingale M^c and a purely discontinuous local martingale M^d, both 0 at 0. (The uniqueness is up to indistinguishability.)*

Proof. In view of Lemma 5.49 we can decompose M as $M = M_0 + N + A$ for a cadlag local L_2-martingale N and a cadlag local martingale A of locally bounded variation, both 0 at 0. By Example 5.98 A is purely discontinuous. Thus to prove existence of the decomposition it suffices to decompose N. If $0 \le T_n \uparrow \infty$ is a sequence of stopping times such that N^{T_n} is an L_2-martingale for every n, then we can decompose $N^{T_n} = N^c_n + N^d_n$ in $\mathcal{H}^2$ for every n. Because this decomposition is unique and both $\mathcal{H}^2_c$ and $\mathcal{H}^2_d$ are closed under stopping (because $[M^T,N]=[M,N]^T$), and $N^{T_m} = (N^{T_n})^{T_m} = (N^c_n)^{T_m} + (N^d_n)^{T_m}$ for $m \le n$, it follows that $(N^c_n)^{T_m} = N^c_m$ and $(N^d_n)^{T_m} = N^d_m$. This implies that we can define N^c and N^d consistently as N^c_m and N^d_m on $[0,T_m]$. The resulting processes satisfy $(N^c)^{T_m} = N^c_m$ and $(N^d)^{T_m} = N^d_m$. The first relation shows immediately that N^c is continuous, while the second shows that N^d is purely discontinuous, in view of the fact $[N^d,K]^{T_m} = [(N^d)^{T_m},K] = 0$ for every continuous $K \in \mathcal{H}^2$.

Given two decompositions $M = M_0 + M^c + M^d = M_0 + N^c + N^d$, the process $X = M^c - N^c = N^d - M^d$ is a continuous local martingale that is purely discontinuous, 0 at 0. By the definition of "purely discontinuous" it follows that X^2 is a local martingale as well. Therefore there exist sequences of stopping times $0 \le T_n \uparrow \infty$ such that $Y = X^{T_n}$ and $Y^2 = (X^2)^{T_n}$ are uniformly integrable martingales, for every n. It follows that $t \mapsto \mathrm{E}Y_t^2$ is constant on $[0,\infty]$ and at the same time $Y_t = \mathrm{E}(Y_\infty|\, \mathcal{F}_t)$ almost surely, for every t. Because a projection decreases norm, this is possible only if $Y_t = Y_\infty$ almost surely for every t. Thus X is constant. ■

Warning. For a martingale M of locally bounded variation the decomposition $M = M_0 + M^c + M^d$ is not the same as the decomposition of M in its continuous and jump parts in the "ordinary" sense of variation, i.e. M^d_t is not equal to $\sum_{s\le t} \Delta M_s$. For instance, the compensated Poisson process is purely discontinuous and hence has continuous part zero. Many (or "most") purely discontinuous martingales contain a nontrivial continuous part in the sense of variation. Below the decomposition of a purely discontinuous martingale into its continuous and jump parts is shown to correspond to a decomposition as a "compensated sum of jumps".

The local martingale M in the decomposition $X = X_0 + M + A$ of a given semimartingale X can be split in its continuous and purely discontinuous parts M^c and M^d. Even though the decomposition of X is not unique, the continuous martingale part M^c is the same for every decomposition. This is true because the difference $M - \bar{M} = \bar{A} - A$ resulting from the given decomposition and another decomposition $X = X_0 + \bar{M} + \bar{A}$ is a local martingale of locally bounded variation, whence it is purely discontinuous by Example 5.98. The process M^c is called the *continuous martingale part* of the semimartingale X, and denoted by X^c.

The decomposition of a semimartingale in its continuous martingale and remaining (purely discontinuous and bounded variation) parts makes it possible to describe the relationship between the two quadratic variation processes.

5.100 Theorem. *For any semimartingales X and Y and $t \geq 0$,*

$$[X, Y]_t = \langle X^c, Y^c \rangle_t + \sum_{s \leq t} \Delta X_s \Delta Y_s.$$

Proof. For simplicity we give the proof only in the case that $X = Y$. The general case can be handled by the polarization identities.

We can decompose the semimartingale as $X = X^c + M + A$, where M is a purely discontinuous local martingale and the process A is cadlag, adapted and locally of bounded variation. By the bilinearity of the quadratic variation process, this decomposition implies that $[X] = [X^c] + [M] + 2[M, A] + [A] + 2[X^c, M] + 2[X^c, A]$. The first term on the right is equal to $\langle X^c \rangle$. The last two terms on the right are zero, because X^c is continuous, M is purely discontinuous, and A is of locally bounded variation. We need to show that the process $[M] + 2[M, A] + [A]$ is equal to the square jump process

$$\sum_{s \leq t} (\Delta X_s)^2 = \sum_{s \leq t} (\Delta M_s)^2 + 2 \sum_{s \leq t} \Delta M_s \Delta A_s + \sum_{s \leq t} (\Delta A_s)^2.$$

We shall prove that the three corresponding terms in the two sums are identical.

If f is a cadlag function and A is a cadlag function of bounded variation, then, for any partition of the interval $[0, t]$ with meshwidth tending to zero,

$$\sum_i f(t_{i+1}^n) \big(A(t_{i+1}^n) - A(t_i^n) \big) \to \int_{[0,t]} f_s \, dA_s,$$

$$\sum_i f(t_i^n) \big(A(t_{i+1}^n) - A(t_i^n) \big) \to \int_{[0,t]} f_{s-} \, dA_s.$$

Consequently, the difference of the left sides tends to $\int_{s\le t}\Delta f_s\,dA_s = \sum_{s\le t}\Delta f_s\Delta A_s$. This observation applied in turn with $f = M$ and $f = A$ shows that $[M,A]$ and $[A]$ possess the forms as claimed, in view of Theorem 5.58.

Finally, we show that $[M]_t = \sum_{s\le t}(\Delta M_s)^2$ for every purely discontinuous local martingale M. By localization it suffices to show this for every $M\in\mathcal{H}^2_d$. If M is of locally bounded variation, then the equality is immediate from Example 5.67. The space $\mathcal{H}^2_d$ is the closure in $\mathcal{H}^2$ relative to the norm induced by the inner product $(\cdot,\cdot)$ of the set of all $M\in\mathcal{H}^2$ that are of locally bounded variation, by Theorem 5.103. This implies that for any given $M\in\mathcal{H}^2_d$ there exists a sequence of cadlag bounded variation martingales M_n such that $\mathrm{E}|M_{n,\infty}-M_\infty|^2 = \mathrm{E}[M_n-M]_\infty\to 0$. By the triangle inequality (Exercise 5.69),

$$\sup_t\Bigl|\sqrt{[M_n]_t}-\sqrt{[M]_t}\Bigr|\le\sqrt{[M_n-M]_\infty}.$$

The right and hence the left side converges to zero in L_2, whence the sequence of processes $[M_n]$ converges, uniformly in t, to the process $[M]$ in L_1.

Because $\Delta[M] = (\Delta M)^2$, it follows that $\sum_t(\Delta M_t)^2\le[M]_\infty$, and similarly for M_n and M_n-M. For M_n the inequality is even an equality, by Example 5.67. By Cauchy-Schwarz' inequality,

$$\Bigl(\mathrm{E}\sum_t\bigl|(\Delta M_{n,t})^2-(\Delta M_t)^2\bigr|\Bigr)^2\le\mathrm{E}\sum_t\bigl(\Delta(M_n-M)_t\bigr)^2\mathrm{E}\sum_t\bigl(\Delta(M_n+M)_t\bigr)^2.$$

The right side is bounded by $\mathrm{E}[M_n-M]_\infty 2(\mathrm{E}[M_n]_\infty+\mathrm{E}[M]_\infty)$ and hence converges to zero. We conclude that the sequence of processes $[M_n]_t = \sum_t(\Delta M_{n,t})^2$ converges in L_1 to the corresponding square jump process of M. Combination with the result of the preceding paragraph gives the desired representation of the process $[M]$.† ∎

5.101 EXERCISE. If M is a local martingale and H is a locally bounded predictable process, then $H\circ M$ is a local martingale by Theorem 5.52. Show that $H\circ M$ is purely continuous if M is purely discontinuous. [Hint: by the preceding theorem the angle bracket process of $(H\circ M)^c$ is given by $[H\circ M]-\sum\Delta(H\circ M)^2 = H^2\circ[M]-\sum H^2\Delta M^2$. Use the preceding theorem again to see that this is zero.]

† For another proof see Rogers and Williams, pp384–385, in particular the proof of Theorem 36.5.

5.11.1 Compensated Jump Martingales

If M and N are both contained in $\mathcal{H}^2_d$ and possess the same jump process $\Delta M = \Delta N$, then $M - N$ is contained in both $\mathcal{H}^2_d$ and $\mathcal{H}^2_c$ and hence is zero (up to evanescence). This remains true if M and N are purely discontinuous local martingales. We can paraphrase this by saying that a purely discontinuous local martingale is completely determined by its jump process. This property can be given a more concrete expression as follows.

If M is a martingale of integrable variation (i.e. $\mathrm{E}\int |dM| < \infty$), then it is purely discontinuous by Example 5.98. Its cumulative jump process $N_t = \sum_{s\le t} \Delta M_s$ is well defined and integrable, and hence possesses a compensator A, by the Doob-Meyer decomposition. The process $M-(N-A)$ is the difference of two martingales and hence is a martingale itself, which is predictable, because $M - N$ is continuous and A is predictable. By Theorem 5.46 the process $M - (N - A)$ is zero. We conclude that any martingale M of integrable variation can be written as

$$M = N - A, \qquad\qquad N_t = \sum_{s\le t} \Delta M_s,$$

with A the compensator of N. Thus M is a "compensated sum of jumps" or *compensated jump martingale*. Because the compensator $A = N - M$ is continuous, the decomposition $M = N - A$ is at the same time the decomposition of M into its jump and continuous parts (in the ordinary measure-theoretic sense). The compensated Poisson process is an example of this type of martingale, with N the "original" Poisson process and A its (deterministic!) compensator.

5.102 EXERCISE. The process $M_t = 1_{T\le t} - \Lambda(T\wedge t)$ for T a nonnegative random variable with cumulative hazard function Λ is a martingale relative to the natural filtration. Find the decomposition as in the preceding paragraph. [Warning: if Λ possesses jumps, then $N_t \neq 1_{T\le t}$.]

General elements of $\mathcal{H}^2_d$ are more complicated than the "compensated jump martingales" of the preceding paragraph, but can be written as limits of sequences of such simple martingales. The following theorem gives a representation as an infinite series of compensated jump martingales, each of which jumps at most one time.

We shall say that a sequence of stopping times T_n *covers the jumps* of a process M if $\{\Delta M \neq 0\} \subset \cup_n [T_n]$.[‡]

5.103 Theorem. *For every $M \in \mathcal{H}^2_d$ there exists a sequence of stopping times T_n with disjoint graphs that covers the jumps of M such that*

(i) each process $t \mapsto N_{n,t} := \Delta M_{T_n} 1_{T_n\le t}$ is bounded and possesses a continuous compensator A_n.

[‡] It is said to *exhaust the jumps* of M if the graphs are disjoint and $\{\Delta M \neq 0\} = \cup_n [T_n]$.

(ii) $M = M_0 + \sum_n (N_n - A_n)$.

Proof. For simplicity assume that $M_0 = 0$. Suppose first that there exists a sequence of stopping times as claimed. The variation of the process $N_n - A_n$ is equal to $|\Delta M_{T_n}| 1_{T_n<\infty}$, which is bounded by the root of $\sum_s (\Delta M_s)^2 \leq [M]_\infty$ and hence is integrable. It follows that each process $N_n - A_n$ is a martingale of integrable variation (which ostensibly is already written in the compensated jump form as in the preceding discussion) and hence so are the partial sums $M_n = \sum_{k\leq n}(N_k - A_k)$ and the differences $M_n - M_m$. By Example 5.67 (or Theorem 5.100) the quadratic variation process of $M_n - M_m$ is given by, for $m \leq n$,

$$[M_n - M_m]_t = \sum_{s\leq t} \bigl(\Delta(M_n - M_m)_s\bigr)^2 = \sum_{k:m<k\leq n, T_k\leq t} (\Delta M_{T_k})^2,$$

because the processes A_k are continuous and the graphs $[T_k]$ are disjoint. For every $m \leq n$ and $t > 0$ this process is bounded by the series $\sum_k (\Delta M_{T_k})^2 = \sum_s (\Delta M_s)^2 \leq [M]_\infty$, which is integrable. Invoking the dominated convergence theorem we see that $\mathrm{E}[M_n - M_m]_\infty \to 0$ as $n \geq m \to \infty$, and hence the sequence M_n is a Cauchy sequence in $\mathcal{H}^2$. By completeness of this Hilbert space there exists $M' \in \mathcal{H}^2$ such that $\mathrm{E}(M_n - M')^2_\infty = \mathrm{E}[M_n - M']_\infty \to 0$. This implies that $\mathrm{E}\sup_t (M_n - M')^2_t \to 0$, and hence that $\sup_t \bigl|\Delta M_{n,t} - \Delta M'_t\bigr| \to 0$ almost surely along some sequence. By construction the jump process of M_n is equal to the jump process of $t \mapsto \sum_{k\leq n} \Delta M_{T_k} 1_{T_k\leq t}$ and converges to the jump process of M as $n \to \infty$. We conclude that $\Delta M' = \Delta M$ up to evanescence, whence $M' - M$ is a continuous martingale. Because this process is purely discontinuous by construction and assumption, this implies that $M' = M$ and consequently the sequence M_n converges in $\mathcal{H}^2$ to M.

A suitable sequence of stopping times T_n can be constructed to be a sequence of stopping times with disjoint graphs that covers the jumps of M such that every T_n is either predictable or totally inaccessible.[♭] That the corresponding processes N_n possess continuous compensators can be seen as follows.

If T_n is a predictable time, then $\mathrm{E}(\Delta M_{T_n} |\, \mathcal{F}_{T_n-}) = 0$ (e.g. Jacod and Shiryaev, 2.27) and hence $\mathrm{E}N_{n,T} = \mathrm{E}\Delta M_{T_n} 1_{T_n\leq T} = 0$ for every stopping time T, as $\{T_n \leq T\} \in \mathcal{F}_{T_n-}$ (easy, Jacod and Shiryaev, 1.17). Thus the process N_n is a martingale by Lemma 4.22, and hence possesses compensator $A_n = 0$, which is certainly continuous.

If T_n is totally inaccessible, then by definition the graph of T_n is disjoint with the graph of any predictable time T and hence $\Delta N_{n,T} 1_{T<\infty} = 0$ ("N_n is quasi left-continuous"). Then $0 = \mathrm{E}\Delta N_{n,T} 1_{T<\infty} = \mathrm{E}\int 1_{[T]}\, dN_n =$

[♭] A stopping time T is called *totally inaccessible* if $\mathbb{P}(S = T < \infty) = 0$ for every predictable time S.

$\mathrm{E}\int 1_{[T]}\,dA_n = \mathrm{E}\Delta A_{n,T}1_{T<\infty}$, where the second last equality follows because the process $1_{[T]}\circ(N_n - A_n)$ is a martingale. Because ΔA_n is a predictable process, the equality $\mathrm{E}\Delta A_{n,T}1_{T<\infty} = 0$ for every predictable time T implies that $\Delta A_n = 0$, by the section theorem, whence A_n is continuous.

Finally, we prove the existence of the stopping times. The points in $[0,\infty)$ where a given cadlag function jumps more than a given positive number ε are isolated. Thus, given a sequence $\varepsilon_0 = \infty > \varepsilon_1 > \varepsilon_2 \cdots \downarrow 0$ and given $n,k \in \mathbb{N}$, we can define $T_{n,k}$ to be the kth jump of M of size in $[\varepsilon_n, \varepsilon_{n-1})$. We can write these times also in the form

$$T_{n,0} = 0, \qquad T_{n,k} = \inf\bigl\{t > 0\colon |\Delta M_t| 1_{t>T_{n,k-1}} \in [\varepsilon_n, \varepsilon_{n-1})\bigr\}.$$

Because the process ΔM is progressively measurable and $T_{n,k}$ is the hitting time of the interval $[\varepsilon_n, \varepsilon_{n-1})$ by the process $|\Delta M| 1_{(T_{n,k-1},\infty)}$, it follows that the $T_{n,k}$ are stopping times. Their graphs are disjoint and exhaust the jump times of M.

The next step is to decompose each $T_{n,k}$ into a predictable and a totally inaccessible part. Drop the index (n,k) and consider a fixed stopping time T. For a given predictable time S, set $F_S = \{S = T < \infty\}$, an event contained in $\mathcal{F}_T$. Let $p = \sup \mathbb{P}(\cup_i F_{S_i})$, where the supremum is taken over all countable unions of sets F_{S_i} for predictable times S_i. There exists a sequence of predictable times $S_{n,i}$ with $\mathbb{P}(\cup_i F_{S_{n,i}}) \uparrow p$ as $n \to \infty$. If $S_1, S_2, \ldots$ is the collection of all $S_{n,i}$ and $F = \cup_i F_{S_i}$, then $\mathbb{P}(F) = p$, which implies that $F_S \subset F$ up to a null set, for every predictable time S. For any stopping time T and $A \in \mathcal{F}_T$, the random variable $T_A\colon = T1_A + \infty 1_{A^c}$ is a stopping time. Using this notation, we can partition $[T] = [T_F] \cup [T_{F^c}]$. Because $\mathbb{P}(S = T_{F^c} < \infty) = \mathbb{P}(F_S \cap F^c) = 0$ for every predictable time S, the stopping time T_{F^c} is totally inaccessible. By construction $[T_F] \subset \cup_i [S_i]$.

The graphs of the predictable times S_i thus obtained are not necessarily disjoint. Furthermore, it appears that in general we cannot arrange it so that $[T_F] = \cup_i [S_i]$, so that the graphs of S_i corresponding to different stopping times $T_{n,k}$ may overlap. The last step is to replace the collection of all predictable times associated through the preceding construction to some $T_{n,k}$ by a sequence of predictable times with disjoint graphs. For any pair of predictable stopping times S, S' the set $\{S \neq S'\}$ is contained in $\mathcal{F}_{S-}$ and S_F is predictable for every $F \in \mathcal{F}_{S-}$. Thus given a sequence of predictable times S_i, we can replace S_i by $(S_i)_{F_i}$ for $F_i = \cap_{j<i}\{S_i \neq S_j\}$ to obtain a sequence of predictable times with disjoint graphs.

Finally, the jumps ΔM_T at the totally inaccessible times are bounded in view of the construction of the stopping times $T_{n,k}$. The graphs of the predictable times S can be split further, if necessary, to ensure boundedness of the associated jumps ΔM_S: replace S by the sequence S_{G_n} for $G_n = \{|\Delta M_S| \in [\varepsilon_n, \varepsilon_{n-1})\}$. ■

* 5.12 Itô's Formula

The change of variables formula extends to arbitrary semimartingales. The difference with Itô's formula for continuous semimartingales is, of course, caused by the presence of the purely discontinuous martingale part and possible jumps in the bounded variation process. In Theorem 5.100 the purely discontinuous martingale was seen to contribute to the quadratic process only through its jumps. It is the same for the second order (quadratic) term in Itô's formula. Thus the additional complication in Itô's formula is no greater than the complication that already arises in the change of variables formula for deterministic processes of bounded variation. Taking into account that the latter type of formula is rarely included in a course on measure theory, we may consider the following theorem a little esoteric. However, in siutations where processes with jumps arise together with continuous semimartingales, the unification offered by the theorem is important. Such situations arise for instance in limiting theory, where processes with jump may have continuous approximations.

5.104 Theorem (Itô's formula). *For any vector $X = (X_1, \ldots, X_d)$ of cadlag semimartingales and every twice continuously differentiable function $f: \mathbb{R}^d \to \mathbb{R}$,*

$$f(X_t) - f(X_0) = \sum_{i=1}^d \int_0^t D_i f(X_-)\, dX_i + \tfrac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \int_0^t D_{ij} f(X_{s-})\, d[X_i^c, X_j^c]_s$$
$$+ \sum_{s \le t} \Big(\Delta\big(f(X_s)\big) - \sum_{i=1}^d D_i f(X_{s-}) \Delta X_{i,s} \Big), \qquad \text{a.s..}$$

6
Stochastic Calculus

In this chapter we discuss some examples of “stochastic calculus”, the manipulation of stochastic integrals, mainly by the use of the Itô formula. The more substantial application to stochastic differential equations is discussed in Chapter 7.

We recall the *differential notation* for stochastic integrals. For processes X, Y, Z we write

$$dX = Y\,dZ, \qquad \text{iff} \qquad X = X_0 + Y \cdot Z.$$

In particular $d(Y \cdot Z) = Y\,dZ$. By the substitution rule, Lemma 5.55(ii), it follows that $dZ = Y^{-1}\,dX$ if $dX = Y\,dZ$ for a strictly positive process Y, provided the stochastic integrals are well defined.

For notational convenience we use complex-valued processes in some of the proofs. A complex-valued random variable Z on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $Z\colon \Omega \to \mathbb{C}$ of the form $Z = U + iV$ for ordinary, real-valued random variables U and V. Its expectation is defined as $\mathrm{E}Z = \mathrm{E}U + i\mathrm{E}V$, if U and V are integrable. Conditional expectations $\mathrm{E}(Z|\,\mathcal{F}_0)$ are defined similarly from the conditional expectations of the real and imaginary parts of Z. A complex-valued stochastic process is a collection $Z = \{Z_t\colon t \geq 0\}$ of complex-valued random variables. A *complex-valued martingale* Z is a complex-valued process whose real and imaginary parts are martingales. Given the preceding definitions of (conditional) expectations, this is equivalent to the process satisfying the martingale property $\mathrm{E}(Z_t|\,\mathcal{F}_s) = Z_s$ for $s \leq t$.

With these definitions it can be verified that Itô's formula extends to twice continuously differentiable complex-valued functions $f\colon \mathbb{R}^d \to \mathbb{C}$. We simply apply the formula to the real and imaginary parts of f and next combine.

6.1 Lévy's Theorem

The (predictable) quadratic variation process of a Brownian motion is the identity function. Lévy's theorem asserts that Brownian motion is the only continuous local martingale with this quadratic variation process. It is a useful tool to show that a given process is a Brownian motion. The continuity is essential, because the compensated Poisson process is another example of a martingale with predictable quadratic variation process equal to the identity.

6.1 Theorem (Lévy). *Let M be a continuous local martingale, 0 at 0, such that $[M]$ is the identity function. Then M is a Brownian motion process.*

Proof. For a fixed real number θ consider the complex-valued stochastic process

$$X_t = e^{i\theta M_t + \frac{1}{2}\theta^2 t}.$$

By application of Itô's formula to $X_t = f(M_t, t)$ with the complex-valued function $f(m,t) = \exp(i\theta m + \frac{1}{2}\theta^2 t)$, we find

$$dX_t = X_t i\theta\, dM_t + \tfrac{1}{2} X_t (i\theta)^2\, d[M]_t + X_t \tfrac{1}{2}\theta^2 dt = X_t i\theta\, dM_t,$$

since $[M]_t = t$ by assumption. It follows that $X = X_0 + i\theta X \cdot M$ and hence X is a (complex-valued) local martingale. Because $|X_t|$ is actually bounded for every fixed t, X is a martingale. The martingale relation $\mathrm{E}(X_t \,|\, \mathcal{F}_s) = X_s$ can be rewritten in the form

$$\mathrm{E}\Big(e^{i\theta(M_t - M_s)} \,|\, \mathcal{F}_s\Big) = e^{-\frac{1}{2}\theta^2(t-s)}, \qquad \text{a.s.}, \qquad s < t.$$

This implies that $M_t - M_s$ is independent of $\mathcal{F}_s$ and possesses the normal distribution with mean zero and variance $t - s$. (Cf. Exercise 6.2.) ∎

6.2 EXERCISE. Let X be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}_0 \subset \mathcal{F}$ a sub σ-field such that $\mathrm{E}(e^{i\theta X} \,|\, \mathcal{F}_0)$ is equal to a constant $c(\theta)$ for every $\theta \in \mathbb{R}$. Show that X is independent of $\mathcal{F}_0$.

Lévy's theorem may be interpreted in the sense that among the continuous local martingales Brownian motion is determined by its quadratic variation process. Actually, every continuous local martingale is "determined" by its quadratic variation process, in a certain sense. The following theorem shows that we can generate an arbitrary continuous local martingale from a Brownian motion by transforming the time scale using the inverse process of the quadratic variation. In the words of Rogers and Williams, p64, any such continuous local martingale "has delusions of grandeur: it thinks it is a Brownian motion" running on a different clock.

6.3 Theorem. *Let M be a continuous local martingale relative to a filtration $\{\mathcal{F}_t\}$ such that $[M]_t \uparrow \infty$ almost surely, as $t \uparrow \infty$. Let $T_t = \inf\{s \geq 0\colon [M]_s > t\}$. Then the process $B_t = M_{T_t}$ is a Brownian motion relative to the filtration $\{\mathcal{F}_{T_t}\}$ and $M_t = B_{[M]_t}$.*

Proof. For every fixed t the variable T_t is a stopping time relative to the filtration $\{\mathcal{F}_t\}$, and the maps $t \mapsto T_t$ are right continuous. It follows from this that $\{\mathcal{F}_{T_t}\}$ is a right continuous filtration. Indeed, if $A \in \mathcal{F}_{T_q}$ for every rational number $q > t$, then $A \cap \{T_q < u\} \in \mathcal{F}_u$ for every $u \geq 0$, by the definition of $\mathcal{F}_{T_q}$. Hence $A \cap \{T_t < u\} = \cup_{q>t} A \cap \{T_q < u\} \in \mathcal{F}_u$ for every $u \geq 0$, whence $A \in \mathcal{F}_{T_t}$. The filtration $\{\mathcal{F}_{T_t}\}$ is complete, because $\mathcal{F}_{T_t} \supset \mathcal{F}_0$ for every t.

For simplicity assume first that the sample paths $s \mapsto [M]_s$ of $[M]$ are strictly increasing. Then the maps $t \mapsto T_t$ are their true inverses and, for every $s, t \geq 0$,

$$T_{t\wedge[M]_s} = T_t \wedge s. \tag{6.4}$$

In the case that $t < [M]_s$, which is equivalent to $T_t < s$, this is true because both sides reduce to T_t. In the other case, that $t \geq [M]_s$, the identity reduces to $T_{[M]_s} = s$, which is correct because T is the inverse of $[M]$.

The continuous local martingale M can be localized by the stopping times $S_n = \inf\{s \geq 0\colon |M_s| \geq n\}$. The stopped process M^{S_n} is a bounded martingale, for every n. By the definition $B_t = M_{T_t}$ and (6.4),

$$\begin{aligned} B_{t\wedge[M]_{S_n}} &= M_{T_t\wedge S_n}, \\ B^2_{t\wedge[M]_{S_n}} - t \wedge [M]_{S_n} &= M^2_{T_t\wedge S_n} - [M]_{T_t\wedge S_n}, \end{aligned}$$

where we also use the identity $t = [M]_{T_t}$. The variable $R_n = [M]_{S_n}$ is an $\mathcal{F}_{T_t}$-stopping time, because, for every $t \geq 0$,

$$\{[M]_{S_n} > t\} = \{S_n > T_t\} \in \mathcal{F}_{T_t}.$$

The last inclusion follows from the fact that for any pair of stopping times S, T the event $\{T < S\}$ is contained in $\mathcal{F}_T$, because its intersection with $\{T < t\}$ can be written in the form $\cup_{q<t}\{T < q \leq t < S\} \in \mathcal{F}_t$, where the union is restricted to rational numbers $q \geq 0$.

By the optional stopping theorem the processes $t \mapsto M_{T_t\wedge S_n}$ and $t \mapsto M^2_{T_t\wedge S_n} - [M]_{T_t\wedge S_n}$ are martingales relative to the filtration $\{\mathcal{F}_{T_t}\}$. Because they are identical to the processes $t \mapsto B_t$ and $t \mapsto B_t^2 - t$ stopped at R_n, we conclude that the latter two processes are local martingales. From the local martingale property of the process $t \mapsto B_t^2 - t$ it follows that $\langle B\rangle$ is the identity process. Because M and T are continuous, so is B. By Lévy's theorem, Theorem 6.1, we conclude that B is a Brownian motion. This concludes the proof if $[M]$ is strictly increasing.

For the proof in the general case we may still assume that the sample paths of $[M]$ are continuous and nondecreasing, but we must allow them to possess intervals of constant value, which we shall refer to as "flats". The maps $t \mapsto T_t$ are "generalized inverses" of the maps $s \mapsto [M]_s$ and map a value t to the largest time s with $[M]_s = t$, i.e. the right end point of the flat at height t. The function $s \mapsto [M]_s$ is constant on each interval of the form $[s, T_{[M]_s}]$, the time $T_{[M]_s}$ being the right end point of the flat at height $[M]_s$. The inverse maps $t \mapsto T_t$ are cadlag with jumps at the values t that are heights of flats of nonzero length. For every $s, t \geq 0$,

$$\begin{aligned} &T_t < s \text{ iff } t < [M]_s, \\ &[M]_{T_t} = t, \\ &T_{[M]_s} \geq s, \end{aligned}$$

with, in the last line, equality unless s is in the interior or on the left side of a flat of nonzero length.

These facts show that (6.4) is still valid for every s that is not in the interior or on the left side of a flat. Then the proof can be completed as before provided that the stopping time S_n is never in the interior or on the left of a flat and the sample paths of B are continuous.

Both properties follow if M is constant on every flat. (Then S_n cannot be in the interior or on the left of a flat, because by its definition M increases immediately after S_n.) It is sufficient to show that the stopped process M^{S_n} has this property, for every n. By the martingale relation, for every stopping time $T \geq s$,

$$\begin{aligned} \mathrm{E}\big((M_T^{S_n} - M_s^{S_n})^2 |\, \mathcal{F}_s\big) &= \mathrm{E}\big(M_{S_n \wedge T}^2 - M_{S_n \wedge s}^2 |\, \mathcal{F}_s\big) \\ &= \mathrm{E}\big([M]_{S_n \wedge T} - [M]_{S_n \wedge s} |\, \mathcal{F}_s\big). \end{aligned}$$

For T equal to the stopping time $\inf\{t \geq s : [M]_{S_n \wedge t} > [M]_{S_n \wedge s}\}$, the right side vanishes. We conclude that for every $s \geq 0$, the process M takes the same value at s as at the right end point of the flat containing s, almost surely. For ω not contained in the union of the null sets attached to some rational s, the corresponding sample path of M is constant on the flats of $[M]$. ∎

The filtration $\{\mathcal{F}_{T_t}\}$ may be bigger than the completed natural filtration generated by B and the variables $[M]_t$ may not be stopping times for the filtration generated by B. This hampers the interpretation of M as a *time-changed Brownian motion*, and the Brownian motion may need to have special properties. The theorem is still a wonderful tool to derive properties of general continuous local martingales from properties of Brownian motion.

The condition that $[M]_t \uparrow \infty$ cannot be dispensed of in the preceding theorem, because if $[M]_t$ remains bounded, then the process B is not defined on the full time scale $[0, \infty)$. However, the theorem may be adapted to cover more general local martingales, by piecing B as defined together with an additional independent Brownian motion that starts at time $[M]_\infty$. For this, see Chung and Williams, p??, or Rogers and Williams, p64-67.

Both theorems allow extension to multidimensional processes. The multivariate version of Lévy's theorem can be proved in exactly the same way. We leave this as an exercise. Extension of the time-change theorem is harder.

6.5 EXERCISE. For $i = 1, \ldots, d$ let M_i be a continuous local martingale, 0 at 0, such that $[M_i, M_j]_t = \delta_{ij} t$ almost surely for every $t \geq 0$. Show that $M = (M_1, \ldots, M_d)$ is a vector-valued Brownian motion, i.e. for every $s < t$ the random vector $M_t - M_s$ is independent of $\mathcal{F}_s$ and normally distributed with mean zero and covariance matrix $(t - s)$ times the identity matrix.

6.2 Brownian Martingales

Let B be a Brownian motion on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and denote the completion of the natural filtration generated by B by $\{\mathcal{F}_t\}$. Stochastic processes on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ that are martingales are referred to as *Brownian martingales*. Brownian motion itself is an example, and so are all stochastic integrals $X \cdot B$ for predictable processes X that are appropriately integrable to make the stochastic integral well defined.

The following theorem shows that these are the only Brownian martingales.

One interesting corollary is that every Brownian martingale can be chosen continuous, because all stochastic integrals relative to Brownian motion have a continuous version.

6.6 Theorem. *Let $\{\mathcal{F}_t\}$ be the completion of the natural filtration of a Brownian motion process B. If M is a cadlag local martingale relative to $\{\mathcal{F}_t\}$, then there exists a predictable process X with $\int_0^t X_s^2\, ds < \infty$ almost surely for every $t \geq 0$ such that $M = M_0 + X \cdot B$, up to indistinguishability.*

Proof. We can assume without loss of generality that $M_0 = 0$.

First suppose that M is an L_2-bounded martingale, so that $M_t = \mathrm{E}(M_\infty |\, \mathcal{F}_t)$ almost surely, for every $t \geq 0$, for some square-integrable variable M_∞. For a given process $X \in L_2\bigl([0, \infty) \times \Omega, \mathcal{P}, \mu_M\bigr)$ the stochastic integral $X \cdot B$ is an L_2-bounded martingale with L_2-limit $(X \cdot B)_\infty = \int X\, dB$, because $\int (X 1_{[0,t]} - X)^2\, d\mu_M \to 0$ as $t \to \infty$. The map $I\colon X \to (X \cdot B)_\infty$ is an isometry from $L_2\bigl([0, \infty) \times \Omega, \mathcal{P}, \mu_M\bigr)$ into $L_2(\Omega, \mathcal{F}, \mathbb{P})$. If M_∞ is contained in the range range(I) of this map, then $M_t = \mathrm{E}(M_\infty |\, \mathcal{F}_t) = \mathrm{E}\bigl((X \cdot B)_\infty |\, \mathcal{F}_t\bigr) =$

$(X\cdot B)_t$, almost surely, because $X\cdot B$ is a martingale. Therefore, it suffices to show that range(I) contains all square-integrable variables M_∞ with mean zero.

Because the map I is an isometry on a Hilbert space, its range is a closed linear subspace of $L_2(\Omega,\mathcal{F},\mathbb{P})$. It suffices to show that 0 is the only element of mean zero that is orthogonal to range(I).

Given some process $X\in L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr)$ and a stopping time T, the process $X1_{[0,T]}$ is also an element of $L_2\bigl([0,\infty)\times\Omega,\mathcal{P},\mu_M\bigr)$ and $(X1_{[0,T]}\cdot B)_\infty=(X\cdot B)_T$, by Lemma 5.27(iii). If $M_\infty\perp$ range(I), then it is orthogonal to $(X1_{[0,T]}\cdot B)_\infty$ and hence $0=\mathrm{E}M_\infty(X\cdot B)_T=\mathrm{E}M_T(X\cdot B)_T$, because M is a martingale and $(X\cdot B)_T$ is $\mathcal{F}_T$-measurable. By Lemma 4.22 we conclude that the process $M(X\cdot B)$ is a uniformly integrable martingale.

The process $X_t=\exp(i\theta B_t+\frac12\theta^2 t)$ satisfies $dX_t=i\theta X_t\,dB_t$, by Itô's formula (cf. the proof of Theorem 6.1), and hence $X=1+i\theta X\cdot B$. The process X is not uniformly bounded and hence is not an eligible choice in the preceding paragraph. However, the process $X1_{[0,T]}$ is uniformly bounded for every fixed constant $T\ge 0$ and hence the preceding shows that the process $MX^T=M+i\theta M(X1_{[0,T]}\cdot B)$ is a uniformly integrable martingale. This being true for every $T\ge0$ implies that MX is a martingale. The martingale relation for the process MX can be written in the form

$$\mathrm{E}\Bigl(M_t e^{i\theta(B_t-B_s)}\,|\,\mathcal{F}_s\Bigr)=M_s e^{-\frac12\theta^2(t-s)},\qquad \text{a.s.},\quad s\le t.$$

Multiplying this equation by $\exp(i\theta'(B_s-B_u))$ for $u\le s$ and taking conditional expectation relative to $\mathcal{F}_u$, we find, for $u\le s\le t$,

$$\mathrm{E}\Bigl(M_t e^{i\theta(B_t-B_s)+i\theta'(B_s-B_u)}\,|\,\mathcal{F}_u\Bigr)=M_u e^{-\frac12\theta^2(t-s)-\frac12\theta'^2(u-s)},\ \text{a.s.}.$$

Repeating this operation finitely many times, we find that for an arbitrary partition $0=t_0\le t_1\le\cdots\le t_k=t$ and arbitrary numbers $\theta_1,\ldots,\theta_k$,

$$\mathrm{E}\mathrm{E}\Bigl(M_t e^{i\sum_j\theta_j(B_{t_j}-B_{t_{j-1}})}\,|\,\mathcal{F}_0\Bigr)=\mathrm{E}M_0 e^{-\frac12\sum_j\theta_j^2(t_j-t_{j-1})}=0.$$

We claim that this shows that $M=0$, concluding the proof in the case that M is L_2-bounded.

The claim follows essentially by the uniqueness theorem for characteristic functions. In view of the preceding display the measures $\mu^+_{t_1,\ldots,t_k}$ and $\mu^-_{t_1,\ldots,t_k}$ on $\mathbb{R}^k$ defined by

$$\mu^\pm_{t_1,\ldots,t_k}(A)=\mathrm{E}M_t^\pm 1_A(B_{t_1-t_0},\ldots,B_{t_k-t_{k-1}}),$$

possess identical characteristic functions and hence are identical. This shows that the measures μ^+ and μ^- on $(\Omega,\mathcal{F})$ defined by $\mu^\pm(F)=\mathrm{E}M_t^\pm 1_F$ agree on the σ-field generated by $B_{t_1-t_0},\ldots,B_{t_k-t_{k-1}}$. This being true for every partition of $[0,t]$ shows that μ^+ and μ^- also agree on the algebra generated by $\{B_s: 0 \leq s \leq t\}$ and hence, by Carathéodory's theorem, also on the σ-field generated by these variables. Thus $\mathrm{E} M_t 1_F = 0$ for every F in this σ-field, whence $M_t = 0$ almost surely, because M_t is measurable in this σ-field.

Next we show that any local martingale M as in the statement of the theorem possesses a continuous version. Because we can localize M, it suffices to prove this in the case that M is a uniformly integrable martingale. Then $M_t = \mathrm{E}(M_\infty |\, \mathcal{F}_t)$ for an integrable variable M_∞. If we let M_∞^n be M_∞ truncated to the interval $[-n, n]$, then $M_t^n := \mathrm{E}(M_\infty^n |\, \mathcal{F}_t)$ defines a bounded and hence L_2-bounded martingale, for every n. By the preceding paragraph this can be represented as a stochastic integral with respect to Brownian motion and hence it possesses a continuous version. The process $|M^n - M|$ is a cadlag submartingale, whence by the maximal inequality given by Lemma 4.37,

$$\mathbb{P}\Bigl(\sup_t |M_t^n - M_t| \geq \varepsilon\Bigr) \leq \frac{1}{\varepsilon} \mathrm{E}|M_\infty^n - M_\infty|.$$

The right side converges to zero as $n \to \infty$, by construction, whence the sequence of suprema in the left side converges to zero in probability. There exists a subsequence which converges to zero almost surely, and hence the continuity of the processes M^n carries over onto the continuity of M.

Every continuous local martingale M is locally L_2-bounded. Let $0 \leq T_n \uparrow \infty$ be a sequence of stopping times such that M^{T_n} is an L_2-bounded martingale, for every n. By the preceding we can represent M^{T_n} as $M^{T_n} = X_n \cdot B$ for a predictable process $X_n \in L_2\bigl([0, \infty) \times \Omega, \mathcal{P}, \mu_M\bigr)$, for every n. For $m \leq n$,

$$X_m \cdot B = M^{T_m} = (M^{T_n})^{T_m} = (X_n \cdot B)^{T_m} = X_n 1_{[0, T_m]} \cdot B,$$

by Lemma 5.27(iii) or Lemma 5.55. By the isometry this implies that, for every $t \geq 0$,

$$0 = \mathrm{E}\bigl(X_m \cdot B - X_n 1_{[0,T_m]} \cdot B\bigr)_t^2 = \mathrm{E} \int_0^t (X_m - X_n 1_{[0,T_m]})^2 \, d\lambda.$$

We conclude that $X_m = X_n$ on the set $[0, T_m]$ almost everywhere under $\lambda \times \mathbb{P}$. This enables to define a process X on $[0, \infty) \times \Omega$ in a consistent way, up to a $\lambda \times \mathbb{P}$-null set, by setting $X = X_m$ on the set $[0, T_m]$. Then $(X \cdot B)^{T_m} = X 1_{[0,T_m]} \cdot B = X_m \cdot B = M^{T_m}$ for every m and hence $M = X \cdot B$. The finiteness of $\mathrm{E} \int X_m^2 \, d\lambda$ for every m implies that $\int_0^t X^2 \, d\lambda < \infty$ almost surely, for every $t \geq 0$. ∎

The preceding theorem concerns processes that are local martingales relative to a filtration generated by a Brownian motion. This is restrictive in terms of the local martingales it can be applied to, but at the same time determines the strength of the theorem, which gives a representation as a stochastic integral relative to the given Brownian motion.

If we are just interested in representing a local martingale as a stochastic integral relative to *some* Brownian motion, then we need not restrict the filtration to a special form. Then we can define a Brownian motion in terms of the martingale, and actually the proof of the representation can be much simpler. We leave one result of this type as an exercise. See e.g. Karatzas and Shreve, p170–173 for slightly more general results.

6.7 EXERCISE. Let M be a continuous local martingale with quadratic variation process $[M]$ of the form $[M]_t = \int_0^t \lambda_s\,ds$ for a continuous, strictly positive stochastic process λ. Show that $B = \lambda^{-1/2} \cdot M$ is a Brownian motion, and $M = \sqrt{\lambda} \cdot B$. [Hint: don't use the preceding theorem!]

For an intuitive understanding of the meaning of Theorem 6.6 it helps to think in terms of differentials. The martingale representation says that the infinitesimal increments of any Brownian local martingale M satisfy $dM_t = X_t\,dB_t$ for some predictable process X. In terms of differentials the (local) martingale property could be interpreted as saying that $\mathrm{E}(dM_t|\,\mathcal{F}_t) = 0$. This is pure intuition, as we have not agreed on a formalism to interpret this type of statement concerning differentials. Continuing in this fashion we see that for a predictable process X the value X_t is "known just before t" and hence $\mathrm{E}\big(X_t\,dB_t|\,\mathcal{F}_t\big) = X_t\mathrm{E}(dB_t|\,\mathcal{F}_t) = 0$, by the martingale property of B. Theorem 6.6 says that the increments of any Brownian martingale are constructed in this way: the increment dB_t of Brownian motion times a quantity that can be considered a "known constant" at time t. Thus a Brownian local martingale M is built up of infinitesimal increments dM_t, which all are "deterministic multiples" of the increments of the underlying Brownian motion. The requirements that the increments of M are both mean zero given the past and adapted to the filtration generated by B apparently leave no other choice but the trivial one of multiples of the increments of B. It is clear that the requirement of being adapted to the filtration of B is crucial, because given a much bigger filtration it would be easy to find other ways of extending the sample paths of M through martingale increments dM_t.

6.3 Exponential Processes

The *exponential process* corresponding to a continuous semimartingale X is the process $\mathcal{E}(X)$ defined by

$$\mathcal{E}(X)_t = e^{X_t - \frac{1}{2}[X]_t}.$$

The name "exponential process" would perhaps suggest the process e^X rather than the process $\mathcal{E}(X)$ as defined here. The additional term $\frac{1}{2}[X]$ in the exponent of $\mathcal{E}(X)$ is motivated by the extra term in the Itô formula. An application of this formula to the right side of the preceding display yields

$$d\mathcal{E}(X)_t = \mathcal{E}(X)_t \, dX_t. \tag{6.8}$$

(Cf. the proof of the following theorem.) If we consider the differential equation $df(x) = f(x)\,dx$ as the true definition of the exponential function $f(x) = e^x$, then $\mathcal{E}(X)$ is the "true" exponential process of X, not e^X.

Besides that, the exponentiation as defined here has the nice property of turning local martingales into local martingales.

6.9 Theorem. *The exponential process $\mathcal{E}(X)$ of a continuous local martingale X with $X_0 = 0$ is a local martingale. Furthermore,*

(i) If $\mathrm{E}e^{\frac{1}{2}[X]_t} < \infty$ for every $t \geq 0$, then $\mathcal{E}(X)$ is a martingale.

(ii) If X is an L_2-martingale and $\mathrm{E}\int_0^t \mathcal{E}(X)_s^2 \, d[X]_s < \infty$ for every $t \geq 0$, then $\mathcal{E}(X)$ is an L_2-martingale.

Proof. By Itô's formula applied to the function $f(X_t, [X]_t) = \mathcal{E}(X)_t$, we find that

$$d\mathcal{E}(X)_t = \mathcal{E}(X)_t \, dX_t + \tfrac{1}{2}\mathcal{E}(X)_t \, d[X]_t + \mathcal{E}(X)_t \, (-\tfrac{1}{2}) \, d[X]_t.$$

This simplifies to (6.8) and hence $\mathcal{E}(X) = 1 + \mathcal{E}(X) \cdot X$ is a stochastic integral relative to X. If X is a local martingale, then so is $\mathcal{E}(X)$. Furthermore, if X is an L_2-martingale and $\int 1_{[0,t]} \mathcal{E}(X)^2 \, d\mu_X < \infty$ for every $t \geq 0$, then $\mathcal{E}(X)$ is an L_2-martingale, by Theorem 5.25. This condition reduces to the condition in (ii), in view of Lemma 5.79.

The proof of (i) should be skipped at first reading. If $0 \leq T_n \uparrow \infty$ is a localizing sequence for $\mathcal{E}(X)$, then Fatou's lemma gives

$$\mathrm{E}\big(\mathcal{E}(X)_t \,|\, \mathcal{F}_s\big) \leq \liminf_{n\to\infty} \mathrm{E}\big(\mathcal{E}(X)_{t\wedge T_n} \,|\, \mathcal{F}_s\big) = \liminf_{n\to\infty} \mathcal{E}(X)_{s\wedge T_n} = \mathcal{E}(X)_s.$$

Therefore, the process $\mathcal{E}(X)$ is a supermartingale. It is a martingale if and only if its mean is constant, where the constant must be $\mathrm{E}\mathcal{E}(X)_0 = 1$.

In view of Theorem 6.3 we may assume that the local martingale X takes the form $X_t = B_{[X]_t}$ for a process B that is a Brownian motion relative to a certain filtration. For every fixed t the random variable $[X]_t$ is a stopping time relative to this filtration. We conclude that it suffices to prove: if B is a Brownian motion and T a stopping time with $\mathrm{E}\exp(\frac{1}{2}T) < \infty$, then $\mathrm{E}\exp(B_T - \frac{1}{2}T) = 1$.

Because $2B_s$ is normally distributed with mean zero and variance $4s$,

$$\mathrm{E}\int_0^t \mathcal{E}(B)_s^2 \, ds = \int_0^t \mathrm{E}e^{2B_s} e^{-s} \, ds = \int_0^t e^s \, ds < \infty$$

By (ii) it follows that $\mathcal{E}(B)$ is an L_2-martingale. For given $a < 0$ define $S_a = \inf\{t \geq 0\colon B_t - t = a\}$. Then S_a is a stopping time, so that $\mathcal{E}(B)^{S_a}$ is a martingale, whence $\mathrm{E}\mathcal{E}(B)_{S_a \wedge t} = 1$ for every t. It can be shown that S_a is finite almost surely and

$$\mathrm{E}\mathcal{E}(B)^{S_a} = \mathrm{E}e^{B_{S_a} - \frac{1}{2}S_a} = 1.$$

(The distribution of S_a is known in closed form. See e.g. Rogers and Williams I.9, p18-19; because $B_{S_a} = S_a + a$, the right side is the expectation of $\exp(a + \frac{1}{2}S_a)$.) With the help of Lemma 1.22 we conclude that $\mathcal{E}(B)_{S_a \wedge t} \to \mathcal{E}(B)_{S_a}$ in L_1 as $t \to \infty$, and hence $\mathcal{E}(B)^{S_a}$ is uniformly integrable. By the optional stopping theorem, for any stopping time T,

$$1 = \mathrm{E}\mathcal{E}(B)^{S_a}_T = \mathrm{E}1_{T<S_a}e^{B_T - \frac{1}{2}T} + \mathrm{E}1_{T \geq S_a}e^{B_{S_a} - \frac{1}{2}S_a}.$$

Because the sample paths of the process $t \mapsto B_t - t$ are bounded on compact time intervals, $S_a \uparrow \infty$ if $a \downarrow -\infty$. Therefore, the first term on the right converges to $\mathrm{E}\exp(B_T - \frac{1}{2}T)$, by the monotone convergence theorem. The second term is equal to

$$\mathrm{E}1_{T \geq S_a}e^{S_a + a - \frac{1}{2}S_a} \leq e^a \mathrm{E}e^{\frac{1}{2}T}.$$

If $\mathrm{E}\exp(\frac{1}{2}T) < \infty$, then this converges to zero as $a \to -\infty$. ∎

In applications it is important to determine whether the process $\mathcal{E}(X)$ is a martingale, rather than just a local martingale. No simple necessary and sufficient condition appears to be known, although the condition in (i), which is known as *Novikov's condition*, is optimal in the sense that the factor $\frac{1}{2}$ in the exponent cannot be replaced by a smaller constant, in general.

6.10 EXERCISE. Let X be a continuous semimartingale with $X_0 = 0$. Show that $Y = \mathcal{E}(X)$ is the *unique* solution to the pair of equations $dY = Y\,dX$ and $Y_0 = 1$. [Hint: using Itô's formula show that $d\big(Y\mathcal{E}(X)^{-1}\big) = 0$ for every solution Y, so that $Y\mathcal{E}(X)^{-1} \equiv Y_0\mathcal{E}(X)_0^{-1} = 1$.]

6.11 EXERCISE. Show that $\mathcal{E}(X)^T = \mathcal{E}(X^T)$ for every stopping time T.

6.4 Cameron-Martin-Girsanov Theorem

Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ and a probability measure $\tilde{\mathbb{P}}$ that is absolutely continuous relatively to $\mathbb{P}$, let $d\tilde{\mathbb{P}}/d\mathbb{P}$ be a version of the Radon-Nikodym density of $\tilde{\mathbb{P}}$ relative to $\mathbb{P}$. The process L defined by

$$L_t = \mathrm{E}\Big(\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} |\, \mathcal{F}_t\Big)$$

is a nonnegative, uniformly integrable martingale with mean $\mathrm{E}L_t = 1$. Conversely, every nonnegative, uniformly integrable martingale L with mean 1 possesses a terminal variable L_∞, and can be used to define a probability measure $\tilde{\mathbb{P}}$ by $d\tilde{\mathbb{P}} = L_\infty\, d\mathbb{P}$. Thus there is a one-to-one relationship between "absolutely continuous changes of measure" and certain uniformly integrable martingales.

If the restrictions of $\tilde{\mathbb{P}}$ and $\mathbb{P}$ to the σ-field $\mathcal{F}_t$ are denoted by $\mathbb{P}_t$ and $\mathbb{P}_t$, then, for every $F \in \mathcal{F}_t$, by the martingale property of L,

$$\tilde{\mathbb{P}}_t(F) = \tilde{\mathbb{P}}(F) = \mathrm{E}L_\infty 1_F = \mathrm{E}L_t 1_F = \int_F L_t\, d\mathbb{P}_t.$$

This shows that the measure $\tilde{\mathbb{P}}_t$ is absolutely continuous with respect to the measure $\mathbb{P}_t$, with density

$$\frac{d\tilde{\mathbb{P}}_t}{d\mathbb{P}_t} = L_t.$$

For this reason the martingale L is also referred to as the *density process*. Its value at t gives insight in the "change of measure" of events up till time t.

It may happen that the measure $\tilde{\mathbb{P}}_t$ is absolutely continuous relative to the measure $\mathbb{P}_t$ for every $t \geq 0$, but $\tilde{\mathbb{P}}$ is not absolutely continuous relatively to $\mathbb{P}$. To cover this situation it is useful to introduce a concept of "local absolute continuity". A measure $\tilde{\mathbb{P}}$ is *locally absolutely continuous* relatively to a measure $\mathbb{P}$ if $\tilde{\mathbb{P}}_t \ll \mathbb{P}_t$ for every $t \in [0, \infty)$. In this case we can define a process L through the corresponding Radon-Nikodym densities, as in the preceding display. Then $\mathrm{E}L_t 1_F = \tilde{\mathbb{P}}_t(F) = \tilde{\mathbb{P}}_s(F) = \mathrm{E}L_s 1_F$, for every $F \in \mathcal{F}_s$ and $s < t$, and hence the process L is a martingale, with mean $\mathrm{E}L_t = 1$. If this (generalized) density process were uniformly integrable, then it would have a terminal variable, and we would be back in the situation as previously. Thus the difference between absolute continuity and local absolute continuity is precisely the uniform integrability of the density process.

By Theorem 4.6 the martingale L possesses a cadlag version, which we use throughout this section. In the following lemma we collect some properties. Call $\tilde{\mathbb{P}}$ and $\mathbb{P}$ locally equivalent if the pair of measures is locally absolutely continuous in both directions.

6.12 Lemma. *If the measure $\tilde{\mathbb{P}}$ is locally absolutely continuous relative to the measure $\mathbb{P}$, and L is a cadlag version of the corresponding density process, then:*

(i) $\tilde{\mathbb{P}}(F \cap \{T < \infty\}) = \mathrm{E}L_T 1_F 1_{T<\infty}$*, for every* $F \in \mathcal{F}_T$ *and every stopping time* T*.*
(ii) *If* $T_n \uparrow \infty$ $\mathbb{P}$*-almost surely, then* $T_n \uparrow \infty$ $\tilde{\mathbb{P}}$*-almost surely, for any increasing sequence of stopping times* $T_1 \leq T_2 \leq \cdots$*.*
(iii) $L > 0$ *up to* $\tilde{\mathbb{P}}$*-evanescence; and also up to* $\mathbb{P}$*-evanescence if* $\tilde{\mathbb{P}}$ *and* $\mathbb{P}$ *are locally equivalent.*
(iv) *There exists a stopping time* T *such that* $L > 0$ *on* $[0, T)$ *and* $L = 0$ *on* $[T, \infty)$ *up to* $\mathbb{P}$*-evanescence.*

Proof. (i). For every $n \in \mathbb{N}$ the optional stopping theorem applied to the uniformly integrable martingale L^n yields $L_{T\wedge n} = \mathrm{E}(L_n | \mathcal{F}_T)$, $\mathbb{P}$-almost surely. For a given $F \in \mathcal{F}_T$ the set $F \cap \{T \leq n\}$ is contained in both $\mathcal{F}_T$ and $\mathcal{F}_n$. We conclude that $\mathrm{E}L_T 1_F 1_{T\leq n} = \mathrm{E}L_{T\wedge n} 1_F 1_{T\leq n} = \mathrm{E}L_n 1_F 1_{T \leq n} = \tilde{\mathrm{E}} 1_F 1_{T\leq n}$. Finally, we let $n \uparrow \infty$.

(ii). Because $T = \lim T_n$ defines a stopping time, assertion (i) yields that $\tilde{\mathbb{P}}(T < \infty) = \mathrm{E}L_T 1_{T<\infty}$. If $\mathbb{P}(T = \infty) = 1$, then the right side is 0 and hence $\tilde{\mathbb{P}}(T = \infty) = 1$.

(iii). For $n \in \mathbb{N}$ define a stopping time by $T_n = \inf\{t > 0 \colon L_t < n^{-1}\}$. By right continuity $L_{T_n} \leq n^{-1}$ on the event $T_n < \infty$. Consequently property (i) gives $\tilde{\mathbb{P}}(T_n < \infty) = \mathrm{E}L_{T_n} 1_{T_n<\infty} \leq n^{-1}$. We conclude that $\tilde{\mathbb{P}}(\inf_t L_t = 0) \leq n^{-1}$ for every n, and hence $L > 0$ almost surely under $\tilde{\mathbb{P}}$. Equivalently, $T_n \uparrow \infty$ almost surely under $\tilde{\mathbb{P}}$. If $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are locally equivalent, then (ii) implies that $T_n \uparrow \infty$ also $\mathbb{P}$-almost surely, and hence $L > 0$ up to $\mathbb{P}$-evanescence.

(iv). The stopping times T_n defined in the proof of (iii) are strictly increasing and hence possess a limit T. By definition of T_n we have $L_t \geq n^{-1}$ on $[0, T_n)$, whence $L_t > 0$ on $[0, T)$. For any m the optional stopping theorem gives $\mathrm{E}(L_{T\wedge m} | \mathcal{F}_{T_n \wedge m}) = L_{T_n \wedge m} \leq n^{-1}$ on the event $T_n \leq m$. We can conclude that $\mathrm{E}L_{T\wedge m} 1_{T\leq m} \leq \mathrm{E}L_{T\wedge m} 1_{T_n \leq m} = 0$ for every m, and hence $L_T = 0$ on the event $T < \infty$. For any stopping time $S \geq T$ another application of the optional stopping theorem gives $\mathrm{E}(L_{S\wedge m} | \mathcal{F}_{T\wedge m}) = L_{T\wedge m} = 0$ on the event $T \leq m$. We conclude that $L_S = 0$ on the event $S < \infty$. This is true in particular for $S = \inf\{t > T \colon L_t > \varepsilon\}$, for any $\varepsilon > 0$, and hence $L = 0$ on (T, ∞). ■

If M is a local martingale on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, then it typically looses the local martingale property if we use another measure $\tilde{\mathbb{P}}$. The Cameron-Martin-Girsanov theorem shows that M is still a semimartingale under $\tilde{\mathbb{P}}$, and gives an explicit decomposition of M in its martingale and bounded variation parts.

We start with a general lemma on the martingale property under a "change of measure". We refer to a process that is a local martingale under $\mathbb{P}$ as a $\mathbb{P}$-local martingale. For simplicity we restrict ourselves to the case that $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are locally equivalent, i.e. the restrictions $\tilde{\mathbb{P}}_t$ and $\mathbb{P}_t$ are locally absolutely continuous for every t.

6.13 Lemma. *Let $\tilde{\mathbb{P}}$ and $\mathbb{P}$ be locally equivalent probability measures on $(\Omega, \mathcal{F})$ and let L be the corresponding density process. Then a stochastic process M is a $\tilde{\mathbb{P}}$-local martingale if and only if the process LM is a $\mathbb{P}$-local martingale.*

Proof. We first prove the theorem without "local". If M is an adapted $\tilde{\mathbb{P}}$-integrable process, then, for every $s < t$ and $F \in \mathcal{F}_s$,

$$\tilde{\mathrm{E}} M_t 1_F = \mathrm{E} L_t M_t 1_F,$$
$$\tilde{\mathrm{E}} M_s 1_F = \mathrm{E} L_s M_s 1_F,$$

The two left sides are identical for every $F \in \mathcal{F}_s$ and $s < t$ if and only if M is a $\tilde{\mathbb{P}}$-martingale. Similarly, the two right sides are identical if and only if LM is a $\mathbb{P}$-martingale. We conclude that M is a $\tilde{\mathbb{P}}$-martingale if and only if LM is a $\mathbb{P}$-martingale.

If M is a $\tilde{\mathbb{P}}$-local martingale and $0 \le T_n \uparrow \infty$ is a localizing sequence, then the preceding shows that the process LM^{T_n} is a $\mathbb{P}$-martingale, for every n. Then so is the stopped process $(LM^{T_n})^{T_n} = (LM)^{T_n}$. Because T_n is also a localizing sequence under $\mathbb{P}$, we can conclude that LM is a $\mathbb{P}$-local martingale.

Because $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are locally equivalent, we can select a version of L that is strictly positive. Then $d\mathbb{P}_t/d\tilde{\mathbb{P}}_t = L_t^{-1}$ and we can use the argument of the preceding paragraph in the other direction to see that $M = L^{-1}(LM)$ is a $\tilde{\mathbb{P}}$-local martingale if LM is a $\mathbb{P}$-local martingale. ∎

Warning. A sequence of stopping times is defined to be a "localizing sequence" if it is increasing everywhere and has almost sure limit ∞. The latter "almost sure" depends on the underlying probability measure. Thus a localizing sequence for a measure $\mathbb{P}$ need not be localizing for a measure $\tilde{\mathbb{P}}$. In view of Lemma 6.12(ii) this problem does not arise if the measures $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are locally equivalent. In the preceding lemma the implication that LM is a $\mathbb{P}$-local martingale if M is a $\tilde{\mathbb{P}}$ local martingale can be false if $\tilde{\mathbb{P}}$ is locally absolutely continuous relative to $\mathbb{P}$, but not the other way around.

If M itself is a $\mathbb{P}$-local martingale, then generally the process LM will not be a $\mathbb{P}$-local martingale, and hence the process M will not be a $\tilde{\mathbb{P}}$-local martingale. We can correct for this by subtracting an appropriate process. We restrict ourselves to continuous local martingales M. Then a $\mathbb{P}$-local martingale becomes a $\tilde{\mathbb{P}}$ local martingale plus a "drift" $(L_-^{-1})\cdot[L, M]$, which is of locally bounded variation.

6.14 Theorem (Girsanov). *Let $\tilde{\mathbb{P}}$ and $\mathbb{P}$ be locally equivalent probability measures on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\})$ and let L be the density process of $\tilde{\mathbb{P}}$ relative to $\mathbb{P}$. If M is a continuous $\mathbb{P}$-local martingale, then $M - L_-^{-1}\cdot[L, M]$ is a $\tilde{\mathbb{P}}$-local martingale.*

Proof. By Lemma 6.12(ii) the process L_- is strictly positive under both $\tilde{\mathbb{P}}$ and $\mathbb{P}$, whence the process L_-^{-1} is well defined. Because it is left-continuous, it is locally bounded, so that the integral $L_-^{-1} \cdot [L, M]$ is well defined. We claim that the processes

$$LM - [L, M]$$
$$L(L_-^{-1} \cdot [L, M]) - [L, M]$$

are both $\mathbb{P}$-local martingales. Then, taking the difference, we see that the process $LM - LL_-^{-1} \cdot [L, M])$ is a $\mathbb{P}$-local martingale and hence the theorem is a consequence of Lemma 6.13.

That the first process in the display is a $\mathbb{P}$-local martingale is an immediate consequence of Lemma 5.65(ii). For the second we apply the integration-by-parts (or Itô's) formula to see that

$$d\big(L(L_-^{-1} \cdot [L, M])\big) = (L_-^{-1} \cdot [L, M])\, dL + L_-\, d(L_-^{-1} \cdot [L, M]).$$

No "correction term" appears at the end of the display, because the quadratic covariation between the process L and the continuous process of locally bounded variation $L_-^{-1} \cdot [L, M]$ is zero. The integral of the first term on the right is a stochastic integral (of $L_-^{-1} \cdot [L, M]$) relative to the $\mathbb{P}$-martingale L and hence is a $\mathbb{P}$-local martingale. The integral of the second term is $[L, M]$. It follows that the process $L(L_-^{-1} \cdot [L, M]) - [L, M]$ is a local martingale. ■

* **6.15** EXERCISE. In the preceding theorem suppose that M is not necessarily continuous, but the predictable quadratic covariation $\langle L, M\rangle$ is well defined. Show that $M - L_-^{-1} \cdot \langle L, M\rangle$ is a $\tilde{\mathbb{P}}$-local martingale.

The quadratic covariation process $[L, M]$ in the preceding theorem was meant to be the quadratic covariation process under the orginal measure $\mathbb{P}$. Because $\tilde{\mathbb{P}}$ and $\mathbb{P}$ are locally equivalent and a quadratic covariation process can be defined as a limit of inner products of increments, as in (5.59), it is actually also the quadratic variation under $\tilde{\mathbb{P}}$.

Because $L_-^{-1} \cdot [L, M]$ is continuous and of locally bounded variation, the process $M - L_-^{-1} \cdot [L, M]$ possesses the same quadratic variation process $[M]$ as M, where again it does not matter if we use $\mathbb{P}$ or $\tilde{\mathbb{P}}$ as the reference measure. Thus even after correcting the "drift" due to a change of measure, the quadratic variation remains the same.

The latter remark is particularly interesting if M is a $\mathbb{P}$-Brownian motion process. Then both M and $M - L_-^{-1} \cdot [L, M]$ possess quadratic variation process the identity. Because $M - L_-^{-1} \cdot [L, M]$ is a continuous local martingale under $\tilde{\mathbb{P}}$, it is a Brownian motion under $\tilde{\mathbb{P}}$ by Lévy's theorem. This proves the following corollary.

6.16 Corollary. *Let $\tilde{\mathbb{P}}$ and $\mathbb{P}$ be locally equivalent probability measures on $(\Omega, \mathcal{F}\{\mathcal{F}_t\})$ and let L be the corresponding density process. If B is a $\mathbb{P}$-Brownian motion, then $B - L_-^{-1} \cdot [L, B]$ is a $\tilde{\mathbb{P}}$-Brownian motion.*

Many density processes L arise as exponential processes. In fact, given a strictly positive, continuous martingale L, the process $X = L_-^{-1} \cdot L$ is well defined and satisfies $L_-\, dX = dL$. The exponential process is the unique solution to this equation, whence $L = L_0 \mathcal{E}(X)$. (See Section 6.3 for continuous L and Section ?? for the general case.) Girsanov's theorem takes a particularly simple form if formulated in terms of the process X.

6.17 Corollary. *Let $\tilde{\mathbb{P}}$ and $\mathbb{P}$ be locally equivalent probability measures on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\})$ and let the corresponding density process L take the form $L = \mathcal{E}(X)$ for a continuous local martingale X, 0 at 0. If M is a continuous $\mathbb{P}$-local martingale, then $M - [X, M]$ is a $\tilde{\mathbb{P}}$-local martingale.*

Proof. The exponential process $L = \mathcal{E}(X)$ satisfies $dL = L_-\, dX$, or equivalently, $L = 1 + L_- \cdot X$. Hence $L_-^{-1} \cdot [L, M] = L_-^{-1} \cdot [L_- \cdot X, M] = [X, M]$, by Lemma 5.83(i). The corollary follows from Theorem 6.14. ∎

A special case arises if $L = \mathcal{E}(X)$ for X equal to the stochastic integral $X = Y \cdot B$ of a process Y relative to Brownian motion. Then $[X]_t = \int_0^t Y_s^2\, ds$ and

$$\frac{d\tilde{\mathbb{P}}_t}{d\mathbb{P}_t} = e^{\int_0^t Y_s\, dB_s - \frac{1}{2}\int_0^t Y_s^2\, ds} \qquad \text{a.s..} \tag{6.18}$$

By the preceding corollaries the process

$$t \mapsto B_t - \int_0^t Y_s\, ds$$

is a Brownian motion under $\tilde{\mathbb{P}}$. This is the original form of Girsanov's theorem.

It is a fair question why we would be interested in "changes of measure" of the form (6.18). We shall see some reasons when discussing stochastic differential equations or option pricing in later chapters. For now we can note that in the situation that the filtration is the completion of the filtration generated by a Brownian motion any change to an equivalent measure is of the form (6.18).

6.19 Lemma. *Let $\{\mathcal{F}_t\}$ be the completion of the natural filtration of a Brownian motion process B defined on $(\Omega, \mathcal{F}, \mathbb{P})$. If $\tilde{\mathbb{P}}$ is a probability measure on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$, then there exists a predictable process Y with $\int_0^t Y_s^2\, ds < \infty$ almost surely for every $t \geq 0$ such that the restrictions $\tilde{\mathbb{P}}_t$ and $\mathbb{P}_t$ of $\tilde{\mathbb{P}}$ and $\mathbb{P}$ to $\mathcal{F}_t$ satisfy (6.18).*

Proof. The density process L is a martingale relative to the filtration $\{\mathcal{F}_t\}$. Because this is a Brownian filtration, Theorem 6.6 implies that L permits a continuous version. Because L is positive the process L^{-1} is well defined, predictable and locally bounded. Hence the stochastic integral $Z = L^{-1} \cdot L$ is a well-defined local martingale, relative to the Brownian filtration $\{\mathcal{F}_t\}$. By Theorem 6.6 it can be represented as $Z = Y \cdot B$ for a predictable process Y as in the statement of the lemma. The definition $Z = L^{-1} \cdot L$ implies $dL = L\,dZ$. Because $\mathcal{F}_0$ is trivial, the density at zero can be taken equal to $L_0 = 1$. This pair of equations is solved uniquely by $L = \mathcal{E}(Z)$. (Cf. Exercise 6.10.) ∎

6.20 Example. For a given measurable, adapted process Y and constant $T > 0$ assume that

$$\mathrm{E}e^{\frac{1}{2}\int_0^T Y_s^2\,ds} < \infty.$$

Then the process $Y1_{[0,T]} \cdot B = (Y \cdot B)^T$ satisfies Novikov's condition, as its quadratic variation is given by

$$[Y1_{[0,T]} \cdot B]_t = \int_0^{T\wedge t} Y_s^2\,ds.$$

By Theorem 6.9 the process $\mathcal{E}((Y\cdot B)^T)$ is a martingale. Because it is constant on $[T,\infty)$, it is uniformly integrable. Thus according to the discussion at the beginning of this section, we can define a probability measure $\tilde{\mathbb{P}}$ on $\mathcal{F}$ by $d\tilde{\mathbb{P}} = \mathcal{E}(Y\cdot B)^T\,d\mathbb{P}$.

Then the corresponding density process is given by (6.18) with $Y1_{[0,T]}$ replacing Y. We conclude that the process $\{B_t - \int_0^{T\wedge t} Y_s\,ds : t \geq 0\}$ is a Brownian motion under the measure $\tilde{\mathbb{P}}$. In particular, the process $B_t - \int_0^t Y_s\,ds$ is a Brownian motion on the restricted time interval $[0,T]$ relative to the measure $\tilde{\mathbb{P}}_T$ with density $\mathcal{E}(Y\cdot B)_T$ relative to $\mathbb{P}$. □

If a probability measure $\tilde{\mathbb{P}}$ is locally absolutely continuous relative to a probability measure $\mathbb{P}$, then the corresponding density process is a nonnegative $\mathbb{P}$-martingale with mean 1. We may ask if, conversely, every nonnegative martingale L with mean 1 on a given filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ arises as the density process of a measure $\tilde{\mathbb{P}}$ relative to $\mathbb{P}$. In the introduction of this section we have seen that the answer to this question is postive if the martingale is uniformly integrable, but the answer is negative in general.

Given a martingale L and a measure $\mathbb{P}$ we can define for each $t \geq 0$ a measure $\tilde{\mathbb{P}}_t$ on the σ-field $\mathcal{F}_t$ by

$$\frac{d\tilde{\mathbb{P}}_t}{d\mathbb{P}_t} = L_t.$$

If the martingale is nonnegative with mean value 1, then this defines a probability measure for every t. The martingale property ensures that the collection of measures $\tilde{\mathbb{P}}_t$ is consistent in the sense that $\tilde{\mathbb{P}}_s$ is the restriction of $\tilde{\mathbb{P}}_t$ to $\mathcal{F}_s$, for every $s < t$. The remaining question is whether we can find a measure $\tilde{\mathbb{P}}$ on $\mathcal{F}_\infty$ for which $\tilde{\mathbb{P}}_t$ is its restriction to $\mathcal{F}_t$.

Such a "projective limit" of the system $(\tilde{\mathbb{P}}_t, \mathcal{F}_t)$ does not necessarily exist under just the condition that the process L is a martingale. A sufficient condition is that the filtration be generated by some appropriate process. Then we can essentially use Kolmogorov's consistency theorem to construct $\tilde{\mathbb{P}}$.

6.21 Theorem. *Let L be a nonnegative martingale with mean 1 on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. If $\mathcal{F}_t$ is the filtration $\sigma(Z_s: s \le t)$ generated by some stochastic process Z on $(\Omega, \mathcal{F})$ with values in a Polish space, then there exists a probability measure $\tilde{\mathbb{P}}$ on $\mathcal{F}_\infty$ whose restriction to $\mathcal{F}_t$ possesses density L_t relative to $\mathbb{P}$.*

Proof. Define a probability measure $\tilde{\mathbb{P}}_t$ on $\mathcal{F}_\infty$ by its density L_t relative to $\mathbb{P}$, as before. For $0 \le t_1 < t_2 < \cdots < t_k$ let $R_{t_1,\ldots,t_k}$ be the distribution of the vector $(Z_{t_1}, \ldots, Z_{t_k})$ on the Borel σ-field $\mathcal{D}^k$ of the space $\mathbb{D}^k$ if $(\Omega, \mathcal{F}_\infty)$ is equipped with $\tilde{\mathbb{P}}_{t_k}$. This system of distributions is consistent in the sense of Kolmogorov and hence there exists a probability measure R on the space $(\mathbb{D}^{[0,\infty)}, \mathcal{D}^{[0,\infty)})$ whose marginal distributions are equal to the measures $R_{t_1,\ldots,t_k}$.

For a measurable set $B \in \mathcal{D}^{[0,\infty)}$ now define $\tilde{\mathbb{P}}\big(Z^{-1}(B)\big) = R(B)$. If this is well defined, then it is not difficult to verify that $\tilde{\mathbb{P}}$ is a probability measure on $\mathcal{F}_\infty = Z^{-1}(\mathcal{D}^{[0,\infty)})$ with the desired properties.

The definition of $\tilde{\mathbb{P}}$ is well posed if $Z^{-1}(B) = Z^{-1}(B')$ for a pair of sets $B, B' \in \mathcal{D}^{[0,\infty)}$ implies that $R(B) = R(B')$. Actually, it suffices to show that this is true for every pair of sets B, B' in the union $\mathcal{A}$ of all cylinder σ-fields in $\mathbb{D}^{[0,\infty)}$ (the collection of all measurable sets depending on only finitely many coordinates). Then $\tilde{\mathbb{P}}$ is well defined and σ-additive on $\cup_t \mathcal{F}_t = Z^{-1}(A)$, which is an algeba, and hence possesses a unique extension to the σ-field $\mathcal{F}_\infty$, by Carathéodory's theorem.

The algebra $\mathcal{A}$ consists of all sets B of the form $B = \big\{z \in \mathbb{D}^{[0,\infty)}: (z_{t_1}, \ldots, z_{t_k}) \in B_k\big\}$ for a Borel set B_k in $\mathbb{R}^k$. If $Z^{-1}(B) = Z^{-1}(B')$ for sets $B, B \in \mathcal{A}$, then there exist k, coordinates $t_1, \ldots, t_k$, and Borel sets B_k, B_k' such that $\{(Z_{t_1}, \ldots, Z_{t_k}) \in B_k\} = \{(Z_{t_1}, \ldots, Z_{t_k}) \in B_k'\}$ and hence $R_{t_1,\ldots,t_k}(B_k) = R_{t_1,\ldots,t_k}(B_k')$, by the definition of $R_{t_1,\ldots,t_k}$. ∎

The condition of the preceding lemma that the filtration be the natural filtration generated by a process Z does not permit that the filtration is complete under $\mathbb{P}$. In fact, completion may cause problems, because, in general, the measure $\tilde{\mathbb{P}}$ will not be absolutely continuous relative to $\mathbb{P}$. This is illustrated in the following simple problem.

* **6.22 Example (Brownian motion with linear drift).** Let B be a Brownian motion on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, which is assumed to satisfy the usual conditions. For a given constant $\mu > 0$ consider the process L defined by

$$L_t = e^{\mu B_t - \frac{1}{2}\mu^2 t}.$$

The process L can be seen to be a $\mathbb{P}$-martingale, either by direct calculation or by Novikov's condition, and it is nonnegative with mean 1. Therefore, for every $t \geq 0$ we can define a probability measure $\tilde{\mathbb{P}}_t$ on $\mathcal{F}_t$ by $d\tilde{\mathbb{P}}_t = L_t\, d\mathbb{P}$. Because by assumption the Brownian motion B is adapted to the given filtration, the natural filtration $\mathcal{F}_t^o$ generated by B is contained in the filtration $\mathcal{F}_t$. The measures $\tilde{\mathbb{P}}_t$ are also defined on the filtration $\mathcal{F}_t^o$. By the preceding lemma there exists a probability measure $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F}_\infty^o)$ whose restriction to $\mathcal{F}_t^o$ is $\tilde{\mathbb{P}}_t$, for every t.

We shall now show that:

(i) There is no probability measure $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F}_\infty)$ whose restriction to $\mathcal{F}_t$ is equal to $\tilde{\mathbb{P}}_t$.

(ii) The process $B_t - \mu t$ is a Brownian motion on $(\Omega, \mathcal{F}_\infty^o, \{\mathcal{F}_t^o\}, \tilde{\mathbb{P}})$ (and hence also on the completion of this filtered space).

Claim (ii) is a consequence of Girsanov's theorem. In Example 6.20 this theorem was seen to imply that the process $\{B_t - \mu t: 0 \varepsilon t \leq T\}$ is a Brownian motion on the "truncated" filtered space $(\Omega, \mathcal{F}_T, \{\mathcal{F}_t \cap \mathcal{F}_T\}, \tilde{\mathbb{P}}_T)$, for every $T > 0$. Because the process is adapted to the smaller filtration $\mathcal{F}_t^o$, it is also a Brownian motion on the space $(\Omega, \mathcal{F}_T^o, \{\mathcal{F}_t^o \cap \mathcal{F}_T^o\}, \tilde{\mathbb{P}}_T)$. This being true for every $T > 0$ implies (ii).

If there were a probability measure $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F}_\infty)$ as in (i), then the process $B_t - \mu t$ would be a Brownian motion on the filtered space $(\Omega, \mathcal{F}_\infty, \{\mathcal{F}_t\}, \tilde{\mathbb{P}})$, by Girsanov's theorem. We shall show that this leads to a contradiction. For $n \in \mathbb{R}$ define the event

$$F_\nu = \Big\{\omega \in \Omega: \lim_{t \to \infty} \frac{B_t(\omega)}{t} = \nu\Big\}.$$

Then $\mathcal{F}_\nu \in \mathcal{F}_\infty^o$ and $F_\nu \cap F_{\nu'} = \emptyset$ for $\nu \neq \nu'$. Furthermore, by the ergodic theorem for Brownian motion, $\mathbb{P}(F_0) = 1$ and hence $\mathbb{P}(F_\mu) = 0$. Because $B_t - \mu t$ is a Brownian motion under $\tilde{\mathbb{P}}$, also $\tilde{\mathbb{P}}(F_\mu) = 1$ and hence $\tilde{\mathbb{P}}(F_0) = 0$. Every subset F of F_μ possesses $\mathbb{P}(F) = 0$ and hence is contained in $\mathcal{F}_0$, by the (assumed) completeness of the filtration $\{\mathcal{F}_t\}$. If $B_t - \mu t$ would be a Brownian motion on $(\Omega, \mathcal{F}_\infty, \{\mathcal{F}_t\}, \tilde{\mathbb{P}})$, then $B_t - \mu t$ would be independent (relative to $\tilde{\mathbb{P}}$) of $\mathcal{F}_0$. In particular, B_t would be independent of the event $\{B_t \in C\} \cap F_\mu$ for every Borel set C. Because $\tilde{\mathbb{P}}(F_\mu) = 1$, the variable B_t would also be independent of the event $\{\mathcal{B}_t \in C\}$. This is only possible if B_t is degenerate, which contradicts the fact that $B_t - \mu t$ possesses a normal distribution with positive variance. We conclude that $\tilde{\mathbb{P}}$ does not exist on $\mathcal{F}_\infty$.

The problem in this example is caused by the fact that the projective limit of the measures $\tilde{\mathbb{P}}_t$, which exists on the smaller σ-field $\mathcal{F}^o_\infty$, is orthogonal to the measure $\mathbb{P}$. In such a situation completion of a filtration under one of the two measures effectively adds all events that are nontrivial under the other measure to the filtration at time zero. This is clearly undesirable if we wish to study a process under both probability measures. □

7 Stochastic Differential Equations

In this chapter we consider stochastic differential equations of the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dB_t.$$

Here μ and σ are given functions and B is a Brownian motion process. The equation may be thought of as a randomly perturbed version of the first order differential equation $dX_t = \mu(t, X_t)\, dt$. Brownian motion is often viewed as an appropriate "driving force" for such a noisy perturbation.

The stochastic differential equation is to be understood in the sense that we look for a continuous stochastic process X such that, for every $t \geq 0$,

$$X_t = X_0 + \int_0^t \mu(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s, \qquad \text{a.s.}. \tag{7.1}$$

Usually, we add an initial condition $X_0 = \xi$, for a given random variable ξ, or require that X_0 possesses a given law.

It is useful to discern two ways of posing the problem, the strong and the weak one, differing mostly in the specification of what is being given a-priori and of which further properties the solution X must satisfy. The functions μ and σ are fixed throughout, and are referred to as the "drift" and "diffusion coefficients" of the equation.

In the "strong setting" we are given a particular filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, a Brownian motion B and an initial random variable ξ, both defined on the given filtered space, and we search for a continuous adapted process X, also defined on the given filtered space, which satisfies the stochastic differential equation with $X_0 = \xi$. It is usually assumed here that the filtration $\{\mathcal{F}_t\}$ is the smallest one to which B is adapted and for which ξ is $\mathcal{F}_0$-measurable, and which satisfies the usual conditions. The requirement that the solution X be adapted then implies that it can be expressed as $X = F(\xi, B)$ for a suitably measurable map F, and the precise definition of a *strong solution* could include certain properties of F, such as appropriate measurability, or the requirement that $F(x, B')$ is a solution of the stochastic differential equation with initial variable $x \in \mathbb{R}$, for every x and every Brownian motion B' defined on some filtered probability space. Different authors make this precise in different ways; we shall not add to this confusion here.

For a *weak solution* of the stochastic differential equation we search for a filtered probability space, as well as a Brownian motion and an initial random variable ξ, and a continuous adapted process X satisfying the stochastic differential equation, all defined on the given filtered space. The initial variable X_0 is usually required to possess a given law. The filtration is required to satisfy the usual conditions only, so that a weak solution X is not necessarily a function of the pair (X_0, B).

Clearly a strong solution in a given setting provides a weak solution, but the converse is false. The existence of a weak solution does not even imply the existence of a strong solution (depending on the measurability assumptions we impose). In particular, there exist examples of weak solutions, for which it can be shown that the filtration must necessarily be bigger than the filtration generated by the driving Brownian motion, so that the solution X cannot be a function of (ξ, B) alone. (For instance, Tanaka's example, see Chung and Williams, pages 248–250.)

For X to solve the stochastic differential equation, the integrals in (7.1) must be well defined. This is certainly the case if μ and σ are measurable functions and, for every $t \geq 0$,

$$\int_0^t |\mu(s, X_s)|\, ds < \infty, \qquad \text{a.s.},$$

$$\int_0^t |\sigma^2(s, X_s)|\, ds < \infty, \qquad \text{a.s.}.$$

Throughout we shall silently understand that it is included in the requirements for "X to solve the stochastic differential equation" that these conditions are satisfied.

7.2 EXERCISE. Show that $t \mapsto \sigma(t, X_t)$ is a predictable process if $\sigma: \mathbb{R}^2 \to \mathbb{R}$ is measurable and X is predictable. [Hint: consider the map $(t, \omega) \mapsto \big(t, X_t(\omega)\big)$ on $[0, \infty) \times \Omega$ equipped with the predictable σ-field.]

The case that μ and σ depend on X only is of special interest. The stochastic differential equation

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dB_t \tag{7.3}$$

is known as a *diffusion equation*. Under some conditions the solution X of a diffusion equation is a time-homogeneous Markov process. Some authors use the term *diffusion process* to denote any time-homogeneous (strong)

Markov process, while other authors reserve the term for solutions of diffusion equations only, sometimes imposing additional conditions of a somewhat technical nature, or relaxing the differential equation to a statement concerning first and second infinitesimal moments of the type

$$\begin{aligned} \mathrm{E}(X_{t+h}-X_t|\,\mathcal{F}_t) &= \mu(X_t)h+o(h), && \text{a.s.}\\ \mathrm{var}(X_{t+h}-X_t|\,\mathcal{F}_t) &= \sigma^2(X_t)h+o(h), && \text{a.s.}, \qquad h\downarrow 0. \end{aligned}$$

These infinitesimal conditions give an important interpretation to the functions μ and σ, and can be extended to the more general equation (7.1). Apparently, stochastic differential equations were invented, by Itô in the 1940s, to construct processes that are "diffusions" in this vaguer sense.

7.4 EXERCISE. Derive the approximations in the preceding display if μ and σ are bounded functions. [Hint: Use the fact that the process $N_h = X_{t+h}-X_t-\int_t^{t+h}\mu(X_s)\,ds$ is a (local) martingale relative to the filtration $\mathcal{G}_h=\mathcal{F}_{t+h}$, and a similar property for the process $h\mapsto (X_{t+h}-X_t)^2-[N]_h$.]

Rather than simplifying the stochastic differential equation, we can also make it more general, by allowing the functions μ and σ to depend not only on (t,X_t), but on t and the sample path of X until t. The resulting stochastic differential equations can be treated by similar methods. (See e.g. pages 122–124 of Rogers and Williams.)

Another generalization is to multi-dimensional equations, driven by a multivariate Brownian motion $B=(B_1,\ldots,B_l)$ and involving a vector-valued function $\mu\colon[0,\infty)\times\mathbb{R}^k\to\mathbb{R}^k$ and a function $\sigma\colon[0,\infty)\times\mathbb{R}^k\to\mathbb{R}^{kl}$ with values in the $k\times l$-matrices. Then we search for a continuous vector-valued process $X=(X_1,\ldots,X_k)$ satisfying, for $i=1,\ldots,k$,

$$X_{t,i}=X_{0,i}+\int_0^t\mu_i(s,X_s)\,ds+\sum_{j=1}^l\int_0^t\sigma_{i,j}(s,X_s)\,dB_{j,s}.$$

Multivariate stochastic differential equations of this type are not essentially more difficult to handle than the one-dimensional equation (7.1). For simplicity we consider the one-dimensional equation (7.1), or at least shall view the equation (7.1) as an abbreviation for the multivariate equation in the preceding display.

We close this section by showing that Girsanov's theorem may be used to construct a weak solution of a special type of stochastic differential equation, under a mild condition. This illustrates that special approaches to special equations can be more powerful than the general results obtained in this chapter.

7.5 Example. Let ξ be an $\mathcal{F}_0$-measurable random variable and let $X-\xi$ be a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. For a given measurable function μ define a process Y by $Y_t = \mu(t, X_t)$, and assume that the exponential process $\mathcal{E}(Y\cdot X)$ is a uniformly integrable martingale. Then $d\tilde{\mathbb{P}} = \mathcal{E}(Y\cdot X)_\infty$ defines a probability measure and, by Corollary 6.16 the process B defined by $B_t = X_t - \xi - \int_0^t Y_s\,ds$ is a $\tilde{\mathbb{P}}$-Brownian motion process. (Note that $Y\cdot X = Y\cdot(X-\xi)$.) It follows that X together with the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \tilde{\mathbb{P}})$ provides a weak solution of the stochastic differential equation $X_t = \xi + \int_0^t \mu(s, X_s)\,ds + B_t$.

The main condition to make this work is that the exponential process of $Y\cdot X$ is a uniformly integrable martingale. This is easy to achieve on compact time intervals by Novikov's condition. □

7.1 Strong Solutions

Following Itô's original approach we construct in this section strong solutions under Lipschitz and growth conditions on the functions μ and σ. We assume that for every $t \geq 0$ there exists a constant C_t such that, for all $s \in [0,t]$ and for all $x, y \in [-t, t]$,

$$\begin{aligned} \big|\mu(s,x) - \mu(s,y)\big| &\leq C_t|x-y|, \\ \big|\sigma(s,x) - \sigma(s,y)\big| &\leq C_t|x-y|. \end{aligned} \tag{7.6}$$

Furthermore, we assume that for every $t \geq 0$ there exists a constant C_t such that, for all $s \in [0,t]$ and $x \in \mathbb{R}$,

$$\begin{aligned} \big|\mu(s,x)\big| &\leq C_t(1+|x|), \\ \big|\sigma(s,x)\big| &\leq C_t(1+|x|). \end{aligned} \tag{7.7}$$

Then the stochastic differential equation (7.1) possesses a strong solution in every possible setting. The proof of this is based on an iterative construction of processes that converge to a solution, much like the Picard iteration scheme for solving a deterministic differential equation.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ be an arbitrary filtered probability space, and let B be a Brownian motion and ξ an $\mathcal{F}_0$-measurable random variable defined on it.

7.8 Theorem. *Let μ and σ be measurable functions that satisfy (7.6)–(7.7). Then there exists a continuous, adapted process X on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with $X_0 = \xi$ that satisfies (7.1). This process is unique up to indistinguishability, and its distribution is uniquely determined by the distribution of ξ.*

Proof. For a given process X let LX denote the process on the right of (7.1), i.e.

$$(LX)_t = \xi + \int_0^t \mu(s, X_s)\,ds + \int_0^t \sigma(s, X_s)\,dB_s.$$

We wish to prove that the equation $LX = X$ possesses a unique continuous adapted solution X. By assumption (7.7) the absolute values of the integrands are bounded above by $C_t\bigl(1 + |X_s|\bigr)$ and hence the integrals in the definition of LX are well defined for every continuous adapted process X.

First assume that ξ is square-integrable and the Lipschitz condition (7.6) is valid for every $x, y \in \mathbb{R}$ (and not just for $x, y \in [-t, t]$). We may assume without of loss of generality that the constants C_t are nondecreasing in t.

By the triangle inequality, the maximal inequality (4.39), the Cauchy-Schwarz inequality, and the defining isometry of stochastic integrals,

$$\begin{aligned}
&\mathrm{E}\sup_{s\le t}\bigl|(LX)_s - (LY)_s\bigr|^2 \\
&\quad\le 2\mathrm{E}\Bigl|\int_0^t \bigl|\mu(s, X_s) - \mu(s, Y_s)\bigr|\,ds\Bigr|^2 + 8\mathrm{E}\Bigl|\int_0^t \bigl(\sigma(s, X_s) - \sigma(s, Y_s)\bigr)\,dB_s\Bigr|^2 \\
&\quad\le 2t\,\mathrm{E}\int_0^t \bigl|\mu(s, X_s) - \mu(s, Y_s)\bigr|^2\,ds + 8\mathrm{E}\int_0^t \bigl(\sigma(s, X_s) - \sigma(s, Y_s)\bigr)^2\,ds \\
&\quad\le 8(t+1)C_t^2\mathrm{E}\int_0^t |X_s - Y_s|^2\,ds.
\end{aligned}$$

The use of the maximal inequality (in the first $\lesssim$) is justified as soon as the process $t \mapsto \int_0^t \bigl(\sigma(s, X_s) - \sigma(s, Y_s)\bigr)\,dB_s$ is an L_2-martingale, which is certainly the case if the final upper bound is finite.

Define processes $X^{(n)}$ by $X^{(0)} = \xi$ and, recursively, $X^{(n)} = LX^{(n-1)}$, for $n \ge 1$. In particular,

$$X_t^{(1)} = \xi + \int_0^t \mu(s, \xi)\,ds + \int_0^t \sigma(s, \xi)\,dB_s.$$

By similar arguments as previously, but now using the growth condition (7.7), we obtain

$$\begin{aligned}
\mathrm{E}\sup_{s\le t}|X_s^{(1)} - X_s^{(0)}|^2 &\le 2t\,\mathrm{E}\int_0^t \mu^2(s, \xi)\,ds + 8\mathrm{E}\int_0^t \sigma^2(s, \xi)\,ds \\
&\le 8(t+1)^2 C_t^2\,\mathrm{E}(1 + \xi^2).
\end{aligned}$$

Furthermore, for $n \ge 1$, since $X^{(n+1)} - X^{(n)} = LX^{(n)} - LX^{(n-1)}$,

$$\mathrm{E}\sup_{s\le t}|X_s^{(n+1)} - X_s^{(n)}|^2 \le 8(t+1)C_t^2\,\mathrm{E}\int_0^t |X_s^{(n)} - X_s^{(n-1)}|^2\,ds.$$

Iterating this last inequality and using the initial bound for $n = 0$ of the preceding display, we find that, with $M = \mathrm{E}(1 + \xi^2)$,

$$\mathrm{E} \sup_{s \le t} |X_s^{(n)} - X_s^{(n-1)}|^2 \le 8^n \frac{(t+1)^{2n} C_t^{2n} M}{n!}.$$

We conclude that, for $m \le n$, by the triangle inequality, for $t > 1$,

$$\varepsilon_{m,n} := \Bigl\| \sup_{s \le t} |X_s^{(n)} - X_s^{(m)}| \Bigr\|_2 \le \sum_{i=m+1}^{n} 4^i \frac{(t+1)^i}{\sqrt{i!}} C_t^i \sqrt{M}.$$

For fixed t, we have that $\varepsilon_{m,n} \to 0$ as $m, n \to \infty$. We conclude that the variables in the left side of the last display converge to zero in quadratic mean and hence in probability as $m, n \to \infty$. In other words, the sequence of processes $X^{(n)}$ forms a Cauchy sequence in probability in the space $C[0, t]$ of continuous functions, equipped with the uniform norm. Since this space is complete there exists a process X such that, as $n \to \infty$,

$$\sup_{s \le t} |X_s^{(n)} - X_s| \xrightarrow{\mathrm{P}} 0.$$

Being a uniform limit of continuous processes, the process X must be continuous. By Fatou's lemma

$$\varepsilon_m := \Bigl\| \sup_{s \le t} |X_s - X_s^{(m)}| \Bigr\|_2 \le \lim_{n \to \infty} \varepsilon_{m,n}.$$

Because $LX^{(n)} = X^{(n+1)}$, the triangle inequality gives that

$$\begin{aligned}
&\Bigl\| \sup_{s \le t} |(LX)_s - X_s| \Bigr\|_2 \\
&\quad \lesssim \Bigl\| \sup_{s \le t} |(LX)_s - (LX^{(n)})_s| \Bigr\|_2 + \Bigl\| \sup_{s \le t} |X_s^{(n+1)} - X_s| \Bigr\|_2 \\
&\quad \lesssim \sqrt{t+1}\, C_t \sqrt{\mathrm{E} \int_0^t |X_s - X_s^{(n)}|^2 \, ds} + \varepsilon_{n+1} \\
&\quad \lesssim \sqrt{t+1} \sqrt{t}\, C_t\, \varepsilon_n + \varepsilon_{n+1}.
\end{aligned}$$

The right side converges to zero as $n \to \infty$, for fixed t, and hence the left side must be identically zero. This shows that $LX = X$, so that X solves the stochastic differential equation, at least on the interval $[0, t]$.

If Y is another solution, then, since in that case $X - Y = LX - LY$,

$$\mathrm{E} \sup_{s \le t} |X_s - Y_s|^2 \lesssim (t+1) C_t^2 \int_0^t \mathrm{E} \sup_{u \le s} |X_u - Y_u|^2 \, ds.$$

By Gronwall's lemma, Lemma 7.11, applied to the function on the left side and with $A = 0$, it follows that the left side must vanish and hence $X = Y$.

By going through the preceding for every $t \in \mathbb{N}$ we can consistently construct a solution on $[0, \infty)$, and conclude that this is unique.

By the measurability of μ and σ the processes $t \mapsto \mu(t, X_t)$ and $t \mapsto \sigma(t, X_t)$ are predictable, and hence progressively measurable, for every predictable process X. (Cf. Exercise 7.2.) By Fubini's theorem the process $t \mapsto \int_0^t \mu(s, X_s)\, ds$ is adapted, while the stochastic integral $t \mapsto \int_0^t \sigma(s, X_s)\, dB_s$ is a local martingale and hence certainly adapted. Because the processes are also continuous, they are predictable. The process $X^{(0)}$ is certainly predictable and hence by induction the process $X^{(n)}$ is predictable for every n. The solution to the stochastic differential equation is indistinguishable from $\liminf_{n\to\infty} X^{(n)}$ and hence is predictable and adapted.

The remainder of the proof should be skipped at first reading. It consists of proving the theorem without the additional conditions on the functions μ and σ and the variable ξ, and is based on the identification lemma given as Lemma 7.12 below. First assume that μ and σ only satisfy (7.6) and (7.7), but ξ is still square-integrable.

For $n \in \mathbb{N}$ let $\chi_n: \mathbb{R} \to \mathbb{R}$ be continuously differentiable with compact support and be equal to the unit function on $[-n, n]$. Then the functions μ_n and σ_n defined by $\mu_n(t, x) = \mu(t, x)\chi_n(x)$ and $\sigma_n(t, x) = \sigma(t, x)\chi_n(x)$ satisfy the conditions of the first part of the proof. Hence there exists, for every n, a continuous adapted process X_n such that

$$X_{n,t} = \xi + \int_0^t \mu_n(s, X_{n,s})\, ds + \int_0^t \sigma_n(s, X_{n,s})\, dB_s. \tag{7.9}$$

For fixed $m \leq n$ the functions μ_m and μ_n, and σ_m and σ_n agree on the interval $[-m, m]$, whence by Lemma 7.12 the process X_m and X_n are indistinguishable on the set $[0, T_m]$ for $T_m = \inf\{t \geq 0: |X_{m,t}| \geq m \text{ or } |X_{n,t}| \geq m\}$.

In particular, the first times that X_m or X_n leave the interval $[-m, m]$ are identical and hence the possibility "$|X_{n,t}| > m$" in the definition of T_m is superfluous. If $0 \leq T_n \uparrow \infty$, then we can consistently define a process X by setting it equal to X_n on $[0, T_n]$, for every n. Then $X^{T_n} = X_n^{T_n}$ and, by the preceding display and Lemma 5.55(i),

$$X_t^{T_n} = \xi + \int_0^t 1_{(0,T_n]}(s)\mu_n(s, X_{n,s})\, ds + \int_0^t 1_{(0,T_n]}(s)\sigma_n(s, X_{n,s})\, dB_s. \tag{7.10}$$

By the definitions of T_n, μ_n, σ_n and X the integrands do not change if we delete the subscript n from μ_n, σ_n and X_n. We conclude that

$$X_t^{T_n} = \xi + \int_0^{T_n \wedge t} \mu(s, X_s)\, ds + \int_0^{T_n \wedge t} \sigma(s, X_s)\, dB_s.$$

This being true for every n implies that X is a solution of the stochastic differential equation (7.1).

We must still show that $0 \le T_n \uparrow \infty$. By the integration-by-parts formula and (7.9)

$$X_{n,t}^2 - X_{n,0}^2 = 2\int_0^t X_{n,s}\mu_n(s, X_{n,s})\,ds + 2\int_0^t X_{n,s}\sigma_n(s, X_{n,s})\,dB_s + \int_0^t \sigma_n^2(s, X_{n,s})\,ds.$$

The process $1_{(0,T_n]}X_{n,s}\sigma_n(s, X_{n,s})$ is bounded on $[0,t]$ and hence the process $t \mapsto \int_0^{T_n\wedge t} X_{n,s}\sigma_n(s, X_{n,s})\,dB_s$ is a martingale. Replacing t by $T_n \wedge t$ in the preceding display and next taking expectations we obtain

$$\begin{aligned}
1 + \mathrm{E}X_{n,T_n\wedge t}^2 &= 1 + \mathrm{E}\xi^2 + 2\mathrm{E}\int_0^{T_n\wedge t} X_{n,s}\mu_n(s, X_{n,s})\,ds \\
&\quad + \mathrm{E}\int_0^{T_n\wedge t} \sigma_n^2(s, X_{n,s})\,ds \\
&\lesssim 1 + \mathrm{E}\xi^2 + (C_t + C_t^2)\,\mathrm{E}\int_0^{T_n\wedge t} (1 + X_{n,s}^2)\,ds \\
&\lesssim 1 + \mathrm{E}\xi^2 + (C_t + C_t^2)\int_0^t (1 + \mathrm{E}X_{n,T_n\wedge s}^2)\,ds.
\end{aligned}$$

We can apply Gronwall's lemma, Lemma 7.11, to the function on the far left of the display to conclude that this is bounded on $[0,t]$, uniformly in n, for every fixed t. By the definition of T_n

$$\mathbb{P}(0 < T_n \le t)n^2 \le \mathrm{E}X_{n,T_n\wedge t}^2.$$

Hence $\mathbb{P}(0 < T_n \le t) = O(n^{-2}) \to 0$ as $n \to \infty$, for every fixed t. Combined with the fact that $\mathbb{P}(T_n = 0) = \mathbb{P}\bigl(|\xi| > n\bigr) \to 0$, this proves that $0 \le T_n \uparrow \infty$.

Finally, we drop the condition that ξ is square-integrable. By the preceding there exists, for every $n \in \mathbb{N}$, a solution X_n to the stochastic differential equation (7.1) with initial value $\xi 1_{|\xi|\le n}$. By Lemma 7.12 the processes X_m and X_n are indistibguishable on the event $\{|\xi| \le m\}$ for every $n \ge m$. Thus $\lim_{n\to\infty} X_n$ exists almost surely and solves the stochastic differential equation with initial value ξ.

The last assertion of the theorem is a consequence of Lemma 7.13 below, or can be argued along the following lines. The distribution of the triple $(\xi, B, X^{(n)})$ on $\mathbb{R}\times C[0,\infty)\times C[0,\infty)$ is determined by the distribution of $(\xi, B, X^{(n-1)})$ and hence ultimately by the distribution of $(\xi, B, X^{(0)})$, which is determined by the distribution of ξ, the distribution of B being fixed as that of a Brownian motion. Therefore the distribution of X is determined by the distribution of ξ as well. (Even though believable this argument needs to be given in more detail to be really convincing.) ∎

7.11 Lemma (Gronwall). *Let $f\colon [0,T] \to \mathbb{R}$ be an integrable function such that $f(t) \le A + B\int_0^t f(s)\,ds$ for every $t \in [0,T]$ and constants A and $B > 0$. Then $f(t) \le Ae^{Bt}$ on $[0,T]$.*

Proof. We can write the inequality in the form $F'(t) - BF(t) \le A$, for F the primitive function of f with $F(0) = 0$. This implies that $\bigl(F(t)e^{-Bt}\bigr)' \le Ae^{-Bt}$. By integrating and rearranging we find that $F(t) \le (A/B)(e^{Bt}-1)$. The lemma follows upon reinserting this in the given inequality. ∎

* 7.1.1 Auxiliary Results

The remainder of this section should be skipped at first reading.

The following lemma is used in the proof of Theorem 7.8, but is also of independent interest. It shows that given two pairs of functions (μ_i, σ_i) that agree on $[0,\infty)\times[-n,n]$, the solutions X_i of the corresponding stochastic differential equations (of the type (7.1)) agree as long as they remain within $[-n,n]$. Furthermore, given two initial variables ξ_i the corresponding solutions X_i are indistinguishable on the event $\{\xi_1 = \xi_2\}$.

7.12 Lemma. *For $i = 1,2$ let $\mu_i, \sigma_i\colon [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be measurable functions that satisfy (7.6)–(7.7), let ξ_i be $\mathcal{F}_0$-measurable random variables, and let X_i be continuous, adapted processes that satisfy (7.1) with (ξ_i, μ_i, σ_i) replacing (ξ, μ, σ). If $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$ on $[0,\infty) \times [-n,n]$ and $T = \inf\{t \ge 0\colon |X_{1,t}| > n,\ \text{or } |X_{2,t}| > n\}$, then $X_1^T = X_2^T$ on the event $\{\xi_1 = \xi_2\}$.*

Proof. By subtracting the stochastic differential equations (7.1) with $(\xi_i, \mu_i, \sigma_i, X_i)$ replacing (ξ, μ, σ, X), and evaluating at $T \wedge t$ instead of t, we obtain

$$X_{1,t}^T - X_{2,t}^T = \xi_1 - \xi_2 + \int_0^{T\wedge t} \bigl(\mu_1(s, X_{1,s}) - \mu_2(s, X_{2,s})\bigr)\,ds + \int_0^{T\wedge t} \bigl(\sigma_1(s, X_{1,s}) - \sigma_2(s, X_{2,s})\bigr)\,dB_s.$$

On the event $F = \{\xi_1 = \xi_2\} \in \mathcal{F}_0$ the first term on the right vanishes. On the set $(0,T]$ the processes X_1 and X_2 are bounded in absolute value by n. Hence the functions μ_1 and μ_2, and σ_1 and σ_2, agree on the domain involved in the integrands and hence can be replaced by their common values $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$. Then we can use the Lipschitz properties of μ_1 and σ_1, and obtain, by similar arguments as in the proof of Theorem 7.8, that

$$\mathrm{E}\sup_{s\le t} |X_{1,s}^T - X_{2,s}^T|^2 1_F \lesssim (t+1)C_t^2\mathrm{E}\int_0^{T\wedge t} |X_{1,s} - X_{2,s}|^2\,ds 1_F.$$

(Note that given an event $F \in \mathcal{F}_0$ the process $Y1_F$ is a martingale whenever the process Y is a martingale.) By Gronwall's lemma the left side of the last display must vanish and hence $X_1^T = X_2^T$ on F. ■

The next lemma gives a strengthening of the last assertion of Theorem 7.8. The lemma shows that, under the conditions of the theorem, solutions to the stochastic differential equation (7.1) can be constructed in a canonical way as $X = F(\xi, B)$ for a fixed map F in any strong setting consisting of an initial variable ξ and a Brownian motion B defined on some filtered probability space. Because the map F is measurable, it follows in particular that the law of X is uniquely determined by the law of ξ.

The sense of the measurability of F is slightly involved. The map F is defined as a map $F: \mathbb{R} \times C[0,\infty) \to C[0,\infty)$. Here $C[0,\infty)$ is the collection of all continuous functions $x: [0,\infty) \to \mathbb{R}$. The *projection σ-field* Π_∞ on this space is the smallest σ-field making all evaluation maps ("projections") $\pi_t: x \mapsto x(t)$ measurable. The *projection filtration* $\{\Pi_t\}$ is defined by $\Pi_t = \sigma(\pi_s: s \le t)$. (The projection σ-field can be shown to be the Borel σ-field for the topology of uniform convergence on compacta.) A Brownian motion process induces a law on the measurable space $\bigl(C[0,\infty), \Pi_\infty\bigr)$. This is called the *Wiener measure*. We denote the completion of the projection filtration under the Wiener measure by $\{\bar{\Pi}_t\}$.

For a proof of the following lemma, see e.g. Rogers and Williams, pages 125–127 and 136–138.

7.13 Lemma. *Under the conditions of Theorem 7.8 there exists a map $F: \mathbb{R} \times C[0,\infty) \to C[0,\infty)$ such that, given any filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with a Brownian motion B and an $\mathcal{F}_0$-measurable random variable ξ defined on it $X = F(\xi, B)$ is a solution to the stochastic differential equation (7.1). This map can be chosen such that the map $\xi \mapsto F(\xi, x)$ is continuous for every $x \in C[0,\infty)$ and such that the map $x \mapsto F(\xi, x)$ is $\bar{\Pi}_t - \Pi_t$-measurable for every $t \ge 0$ and every $\xi \in \mathbb{R}$. In particular, it can be chosen $\mathcal{B} \times \bar{\Pi}_\infty - \Pi_\infty$-measurable.*

Because the solution to the stochastic differential equation in a given setting $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is unique, it follows that any solution takes the form $F(\xi, B)$ and hence induces the same law on $C[0,\infty)$. The latter property is known as *weak uniqueness*.

The preceding lemma gives much more information than weak uniqueness. Weak uniqueness can also be derived as a direct consequence of the uniqueness of the solution asserted in Theorem 7.8 (known as "pathwise uniqueness"). A famous result by Watanabe asserts that pathwise uniqueness always implies weak uniqueness.

7.2 Martingale Problem and Weak Solutions

If X is a continuous solution to the diffusion equation (7.3), defined on some filtered probability space, and $f: \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function, then Itô's formula yields that

$$df(X_t) = f'(X_t)\sigma(X_t)\,dB_t + f'(X_t)\mu(X_t)\,dt + \tfrac{1}{2}f''(X_t)\sigma^2(X_t)\,dt.$$

Defining the differential operator A by

$$Af = \mu f' + \tfrac{1}{2}\sigma^2 f'',$$

we conclude that the process

$$t \mapsto f(X_t) - f(X_0) - \int_0^t Af(X_s)\,ds \tag{7.14}$$

is identical to the stochastic integral $(f'\sigma)(X) \cdot B$, and hence is a local martingale. If f has compact support, in addition to being twice continuously differentiable, and σ is bounded on compacta, then the function $f'\sigma$ is bounded and the process in (7.14) is also a martingale. It is said that X is a solution to the (local) *martingale problem*. This martingale problem can be used to characterize, study and construct solutions of the diffusion equation: instead of constructing a solution directly, we search for a solution to the martingale problem. The following theorem shows the feasibility of this approach.

7.15 Theorem. *Let X be a continuous adapted process on a given filtered space such that the process in (7.14) is a local martingale for every twice continuously differentiable function with compact support. Then there exists a weak solution to the diffusion equation (7.3) with the law of X_0 as the initial law.*

Proof. For given $n \in \mathbb{N}$ let $T_n = \inf\{t \geq 0: |X_t| \geq n\}$, so that $|X^{T_n}| \leq n$ on $(0, T_n]$. Furthermore, let f and g be twice continuously differentiable functions with compact supports that coincide with the functions $x \mapsto x$ and $x \mapsto x^2$ on the set $[-n, n]$. By assumption the processes (7.14) obtained by setting the function f in this equation equal to the present f and to g are local martingales. On the set $(0, T_n]$ they coincide with the processes M and N defined by

$$M_t = X_t - X_0 - \int_0^t \mu(X_s)\,ds$$

$$N_t = X_t^2 - X_0^2 - \int_0^t \bigl(2X_s\mu(X_s) + \sigma^2(X_s)\bigr)\,ds.$$

At time 0 the processes M and N vanish and so do the processes of the type (7.14). We conclude that the correspondence extends to $[0,T_n]$ and hence the processes M and N are local martingales. By simple algebra

$$\begin{aligned} M_t^2 &= X_t^2 - 2X_tX_0 + X_0^2 - 2(X_t - X_0)\int_0^t \mu(X_s)\,ds + \Bigl(\int_0^t \mu(X_s)\,ds\Bigr)^2 \\ &= N_t + A_t + \int_0^t \sigma^2(X_s)\,ds, \end{aligned}$$

for the process A defined by

$$A_t = -2(X_t - X_0)\Bigl(X_0 + \int_0^t \mu(X_s)\,ds\Bigr) + \Bigl(\int_0^t \mu(X_s)\,ds\Bigr)^2 + \int_0^t 2X_s\mu(X_s)\,ds.$$

By Itô's formula

$$\begin{aligned} dA_t &= -2(X_t - X_0)\mu(X_t)\,dt - 2dX_t\Bigl(X_0 + \int_0^t \mu(X_s)\,ds\Bigr) \\ &\qquad + 2\int_0^t \mu(X_s)\,ds\,\mu(X_t)\,dt + 2\mu(X_t)X_t\,dt \\ &= -2\Bigl(X_0 + \int_0^t \mu(X_s)\,ds\Bigr)\,dM_t. \end{aligned}$$

We conclude that the process A is a local martingale and hence so is the process $t \mapsto M_t^2 - \int_0^t \sigma^2(X_s)\,ds$. This implies that $[M]_t = \int_0^t \sigma^2(X_s)\,ds$.

Define a function $\tilde\sigma: \mathbb{R} \to \mathbb{R}$ by setting $\tilde\sigma$ equal to $1/\sigma$ if $\sigma \neq 0$ and equal to 0 otherwise, so that $\tilde\sigma\sigma = 1_{\sigma\neq 0}$. Furthermore, given a Brownian motion process $\tilde B$ define

$$B = \tilde\sigma(X)\cdot M + 1_{\sigma(X)=0}\cdot \tilde B.$$

Being the sum of two stochastic integrals relative to continuous martingales, the process B possesses a continuous version that is a local martingale. Its quadratic variation process is given by

$$[B]_t = \tilde\sigma^2(X)\cdot[M]_t + 2(\tilde\sigma(X)1_{\sigma(X)=0})\cdot[M,\tilde B]_t + 1_{\sigma(X)=0}\cdot[\tilde B]_t.$$

Here we have linearly expanded $[B] = [B,B]$ and used Lemma 5.83. The middle term vanishes by the definition of $\tilde\sigma$, while the sum of the first and third terms on the right is equal to $\int_0^t (\tilde\sigma^2\sigma^2(X_s) + 1_{\sigma(X_s)=0})\,ds = t$. By Lévy's theorem, Theorem 6.1, the process B is a Brownian motion process. By our definitions $\sigma(X)\cdot B = 1_{\sigma(X)\neq 0}\cdot M = M$, because $[1_{\sigma(X)=0}\cdot M] = 0$ whence $1_{\sigma(X)=0}\cdot M = 0$. We conclude that

$$X_t = X_0 + M_t + \int_0^t \mu(X_s)\,ds = X_0 + \int_0^t \sigma(X_s)\,dB_s + \int_0^t \mu(X_s)\,ds.$$

Thus we have found a solution to the diffusion equation (7.3).

In the preceding we have implicitly assumed that the process X and the Brownian motion $\tilde{B}$ are defined on the same filtered probability space, but this may not be possible on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ on which X is given originally. However, we can always construct a Brownian motion $\tilde{B}$ on some filtered space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{\mathbb{P}})$ and next consider the product space

$$(\Omega \times \tilde{\Omega}, \mathcal{F} \times \tilde{\mathcal{F}}, \{\mathcal{F}_t \times \tilde{\mathcal{F}}_t\}, \mathbb{P} \times \tilde{\mathbb{P}}),$$

with the maps

$$(\omega, \tilde{\omega}) \mapsto X(\omega),$$
$$(\omega, \tilde{\omega}) \mapsto \tilde{B}(\tilde{\omega}).$$

The latter processes are exactly as the original processes X and $\tilde{B}$ and hence the first process solves the martingale problem and the second is a Brownian motion. The enlarged filtered probability space may not be complete and satisfy the usual conditions, but this may be remedied by completion and replacing the product filtration $\mathcal{F}_t \times \tilde{\mathcal{F}}_t$ by its completed right-continuous version. ∎

It follows from the proof of the preceding theorem, that a solution X to the martingale problem together with the filtered probability space on which it is defined yields a weak solution of the diffusion equation if σ is never zero. If σ can assume the value zero, then the proof proceeds by extending the given probability space, and X, suitably defined on the extension, again yields a weak solution. The extension may be necessary, because the given filtered probability space may not be rich enough to carry a suitable Brownian motion process.

It is interesting that the proof of Theorem 7.15 proceeds in the opposite direction of the proof of Theorem 7.8. In the latter theorem the solution X is constructed from the given Brownian motion, whereas in Theorem 7.15 the Brownian motion is constructed out of the given X. This is a good illustration of the difference between strong and weak solutions.

Now that it is established that solving the martingale problem and solving the stochastic differential equation in the weak sense are equivalent, we can prove existence of weak solutions for the diffusion equation from consideration of the martingale problem. The advantage of this approach is the availability of additional technical tools to handle martingales.

7.16 Theorem. *If $\mu, \sigma: \mathbb{R} \to \mathbb{R}$ are bounded and continuous and ν is a probability measure on $\mathbb{R}$, then there exists a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with a Brownian motion and a continuous adapted process X satisfying the diffusion equation (7.3) and such that X_0 has law ν.*

Proof. Let (B, ξ) be a pair of a Brownian motion and an $\mathcal{F}_0$-measurable random variable with law ν, defined on some filtered probability space. For every $n \in \mathbb{N}$ define a process $X^{(n)}$ by

$$X_0^{(n)} = \xi,$$
$$X_t^{(n)} = X_{k2^{-n}}^{(n)} + \mu(X_{k2^{-n}}^{(n)})(t - k2^{-n}) + \sigma(X_{k2^{-n}}^{(n)})(B_t - B_{k2^{-n}}),$$
$$k2^{-n} < t \le (k+1)2^{-n}, k = 0, 1, 2, \ldots.$$

Then, for every n, the process $X^{(n)}$ is a continuous solution of the stochastic differential equation

$$X_t^{(n)} = \xi + \int_0^t \mu_n(s)\, ds + \int_0^t \sigma_n(s)\, dB_s, \tag{7.17}$$

for the processes μ_n and σ_n defined by

$$\mu_n(t) = \mu(X_{k2^{-n}}^{(n)}), \qquad \sigma_n(t) = \sigma(X_{k2^{-n}}^{(n)}), \qquad k2^{-n} < t \le (k+1)2^{-n}.$$

By Lemma 5.83 the quadratic variation of the process M defined by $M_t = (\sigma_n \cdot B)_{s+t} - (\sigma_n \cdot B)_s$ is given by $[M]_t = \int_s^{s+t} \sigma_n^2(u)\, du$. For $s \le t$ we obtain, by the triangle inequality and the Burkholder-Davis-Gundy inequality, Lemma 7.19,

$$\begin{aligned} \mathrm{E}|X_s^{(n)} - X_t^{(n)}|^4 &\lesssim \mathrm{E}\Big|\int_s^t \mu_n(u)\, du\Big|^4 + \mathrm{E}\Big|\int_s^t \sigma_n^2(u)\, dB_u\Big|^2 \\ &\lesssim \|\mu\|_\infty^4 |s-t|^4 + \|\sigma\|_\infty^4 |s-t|^2. \end{aligned}$$

By Kolmogorov's criterion (e.g. Van der Vaart and Wellner, page 104) it follows that the sequence of processes $X^{(n)}$ is uniformly tight in the metric space $C[0,\infty)$, equipped with the topology of uniform convergence on compacta. By Prohorov's theorem it contains a weakly converging subsequence. For simplicity of notation we assume that the whole sequence $X^{(n)}$ converges in distribution in $C[0,\infty)$ to a process X. We shall show that X solves the martingale problem, and then can complete the proof by applying Theorem 7.15.

The variable X_0 is the limit in law of the sequence $X_0^{(n)}$ and hence is equal in law to ξ.

For a twice continuously differentiable function $f: \mathbb{R} \to \mathbb{R}$ with compact support, an application of Itô's formula and (7.17) shows that the process

$$f(X_t^{(n)}) - f(X_0^{(n)}) - \int_0^t \big(\mu_n(s) f'(X_s^{(n)}) + \tfrac{1}{2}\sigma_n^2(s) f''(X_s^{(n)})\big)\, ds \tag{7.18}$$

is a martingale. (Cf. the discussion before the statement of Theorem 7.15.) By assumption the functions μ and σ are uniformly continuous on compacta. Hence for every fixed M the moduli of continuity

$$m(\delta) = \sup_{\substack{|x-y| \le \delta \\ |x| \vee |y| \le M}} \big|\mu(x) - \mu(y)\big|, \qquad s(\delta) = \sup_{\substack{|x-y| \le \delta \\ |x| \vee |y| \le M}} \big|\sigma(x) - \sigma(y)\big|$$

converge to zero as $\delta \downarrow 0$. The weak convergence of the sequence $X^{(n)}$ implies the weak convergence of the sequence $\sup_{s\le t}|X_s^{(n)}|$, for every fixed $t \ge 0$. Therefore, we can choose M such that the events $F_n = \{\sup_{s\le t}|X_s^{(n)}| \le M\}$ possess probability arbitrarily close to one, uniformly in n. The weak convergence also implies that, for every fixed $t \ge 0$,

$$\Delta_n := \sup_{|u-v|<2^{-n}, u\le v\le t} |X_u^{(n)} - X_v^{(n)}| \xrightarrow{\mathrm{P}} 0.$$

On the event F_n

$$\Bigl|\int_0^t \bigl(\mu_n(s) - \mu(X_s^{(n)})\bigr) f'(X_s^{(n)})\,ds\Bigr| \le tm(\Delta_n)\|f'\|_\infty \xrightarrow{\mathrm{P}} 0.$$

Combining this with a similar argument for σ_n^2 we conclude that the sequence of processes in (7.18) is asymptotically equivalent to the sequence of processes

$$M_t^n := f(X_t^{(n)}) - f(X_0^{(n)}) - \int_0^t Af(X_s^{(n)})\,ds.$$

These processes are also uniformly bounded on compacta. The martingale property of the processes in (7.18) now yields that $\mathrm{E}M_t^n g(X_u^{(n)}: u \le s) \to 0$ for every bounded, continuous function $g: C[0,s] \to \mathbb{R}$. Because the map $x \mapsto f(x_t) - f(x_0) - \int_0^t Af(x_s)\,ds$ is also continuous and bounded as a map from $C[0,\infty)$ to $\mathbb{R}$, this implies that

$$\mathrm{E}\Bigl(f(X_t) - f(X_s) - \int_s^t Af(X_u)\,du\Bigr) g(X_u: u \le s) = 0.$$

We conclude that X is a martingale relative to its natural filtration. It is automatically also a martingale relative to the completion of its natural filtration. Because X is right continuous, it is again a martingale relative to the right-continuous version of its completed natural filtration, by Theorem 4.6.

Thus X solves the martingale problem, and there exists a weak solution to the diffusion equation with initial law the law of X_0, by Theorem 7.15. ■

7.19 Lemma (Burkholder-Davis-Gundy). *For every $p \ge 2$ there exists a constant C_p such that $\mathrm{E}|M_t|^p \le C_p \mathrm{E}[M]_t^{p/2}$ for every continuous martingale M, 0 at 0, and every $t \ge 0$.*

Proof. Define $m = p/2$ and $Y_t = cM_t^2 + [M]_t$ for a constant $c > 0$ to be determined later. By Itô's formula applied with the functions $x \mapsto x^{2m}$ and $(x,y) \to (cx^2 + y)^m$ we have that

$$\begin{aligned} dM_t^{2m} &= 2mM_t^{2m-1}\,dM_t + \tfrac12 2m(2m-1)M_t^{2m-2}\,d[M]_t, \\ dY_t^m &= mY_t^{m-1}2cM_t\,dM_t + mY_t^{m-1}\,d[M]_t \\ &\qquad + \tfrac12\Bigl(m(m-1)Y_t^{m-2}4c^2M_t^2 + mY_t^{m-1}2c\Bigr)\,d[M]_t. \end{aligned}$$

Assume first that the process Y is bounded. Then the integrals of the two first terms on the right are martingales. Taking the integrals and next expectations we conclude that

$$\mathrm{E}M_t^{2m} = \mathrm{E}\int_0^t \tfrac{1}{2}2m(2m-1)M_s^{2m-2}\,d[M]_s,$$

$$\begin{aligned}\mathrm{E}Y_t^m = \mathrm{E}\int_0^t mY_s^{m-1}\,d[M]_s + \mathrm{E}\int_0^t \tfrac{1}{2}m(m-1)Y_s^{m-2} \\ + 4c^2M_s^2\,d[M]_s + \mathrm{E}\int_0^t \tfrac{1}{2}mY_s^{m-1}2c\,d[M]_s.\end{aligned}$$

The middle term in the second equation is nonnegative, so that the sum of the first and third terms is bounded above by $\mathrm{E}Y_t^m$. Because $M_t^2 \le Y_t/c$, we can bound the right side of the first equation by a multiple of this sum. Thus we can bound the left side $\mathrm{E}M_t^{2m}$ of the first equation by a multiple of the left side $\mathrm{E}Y_t^m$ of the second equation. Using the inequality $|x+y|^m \le 2^{m-1}(x^m+y^m)$ we can bound $\mathrm{E}Y_t^m$ by a multipe of $c^m\mathrm{E}M_t^{2m} + \mathrm{E}[M]_t^m$. Putting this together, we obtain the desired inequality after rearranging and choosing $c>0$ sufficiently close to 0.

If Y is not uniformly bounded, then we stop M at the time $T_n = \inf\{t \ge 0\colon |Y_t| > n\}$. Then Y^{T_n} relates to M^{T_n} in the same way as Y to M and is uniformly bounded. We can apply the preceding to find that the desired inequality is valid for the stopped process M. Next we let $n \to \infty$ and use Fatou's lemma on the left side and the monotone convergence theorem on the right side of the inequality to see that it is valid for M as well. ■

Within the context of weak solutions to stochastic differential equations "uniqueness" of a solution should not refer to the underlying filtered probability space, but it does make sense to speak of "uniqueness in law". Any solution X in a given setting induces a probability distribution on the metric space $C[0,\infty)$. A solution X is called *unique-in-law* if any other solution $\tilde{X}$, possibly defined in a different setting, induces the same distribution on $C[0,\infty)$. Here X and $\tilde{X}$ are understood to possess the same distribution if the vectors $(X_{t_1},\ldots,X_{t_k})$ and $(\tilde{X}_{t_1},\ldots,\tilde{X}_{t_k})$ are equal in distribution for every $0 \le t_1 \le \cdots \le t_k$. (This corresponds to using on $C[0,\infty)$ the σ-field of all Borel sets of the topology of uniform convergence on compacta.)

The last assertion of Theorem 7.8 is exactly that, under the conditions imposed there, that the solution of the stochastic differential equation is unique-in-law. Alternatively, there is an interesting sufficient condition for uniqueness in law in terms of the *Cauchy problem* accompanying the differential operator A. The Cauchy problem is to find, for a given initial function f, a solution $u\colon [0,\infty)\times\mathbb{R}\to\mathbb{R}$ to the partial differential equation

$$\frac{\partial u}{\partial t} = Au, \qquad u(0,\cdot) = f.$$

Here $\partial u/\partial t$ is the partial derivative relative to the first argument of u, whereas the operator A on the right works on the function $x \mapsto u(t,x)$ for fixed t. We make it part of the requirements for solving the Cauchy problem that the partial derivatives $\partial u/\partial t$ and $\partial^2 u/\partial x^2$ exist on $(0,\infty)\times\mathbb{R}$ and possess continuous extensions to $[0,\infty)\times\mathbb{R}$.

A sufficient condition for solvability of the Cauchy problem, where the solution also satisfies the condition in the next theorem, is that the functions μ and σ^2 are Hölder continuous and that σ^2 is bounded away from zero. See Stroock and Varadhan, Theorem 3.2.1.

For a proof of the following theorem, see Karatzas and Shreve, pages 325–427 or Stroock and Varadhan.

7.20 Theorem. *Suppose that the accompanying Cauchy problem admits for every twice continuous differentiable function f with compact support a solution u which is bounded and continuous on the strips $[0,t]\times\mathbb{R}$, for every $t \ge 0$. Then for any $x\in\mathbb{R}$ the solution X to the diffusion equation with initial law $X_0 = x$ is unique.*

7.3 Markov Property

In this section we consider the diffusion equation

$$X_t = X_0 + \int_0^t \mu(X_u)\,du + \int_0^t \sigma(X_u)\,dB_u.$$

Evaluating this equation at the time points $t+s$ and s, taking the difference, and making the change of variables $u = v+s$ in the integrals, we obtain

$$X_{s+t} = X_s + \int_0^t \mu(X_{s+v})\,dv + \int_0^t \sigma(X_{s+v})\,dB_{s+v}.$$

Because the stochastic integral depends only on the increments of the integrator, the process B_{s+v} can be replaced by the process $\tilde{B}_v = B_{s+v} - B_s$, which is a Brownian motion itself and is independent of $\mathcal{F}_s$. The resulting equation suggests that conditionally on $\mathcal{F}_s$ (and hence given X_s) the process $\{X_{s+t}: t \ge 0\}$ relates to the initial value X_s and the Brownian motion $\tilde{B}$ in the same way as the process X relates to the pair (X_s, B) (with X_s fixed). In particular, the conditional law of the process $\{X_{s+t}: t\ge 0\}$ given $\mathcal{F}_s$ should be the same as the law of X given the initial value X_s (considered fixed).

This expresses that a solution of the diffusion equation is a time-homogeneous Markov process: at any time the process will given its past evolve from its present according to the same probability law that determines its evolvement from time zero. This is indeed true, even though a proper mathematical formulation is slightly involved.

A *Markov kernel* from $\mathbb{R}$ into $\mathbb{R}$ is a map $(x, B) \mapsto Q(x, B)$ such that

(i) the map $x \mapsto Q(x, B)$ is measurable, for every Borel set B;

(ii) the map $B \mapsto Q(x, B)$ is a Borel measure, for every $x \in \mathbb{R}$.

A general process X is called a time-homogeneous *Markov process* if for every $t \geq 0$ there exists a Markov kernel Q_t such that, for every Borel set B and every $s \geq 0$,

$$\mathbb{P}(X_{s+t} \in B |\, X_u : u \leq s) = Q_t(X_s, B), \qquad \text{a.s.}.$$

By the towering property of a conditional expectation the common value in the display is then automatically also a version of $\mathbb{P}(X_{s+t} \in B |\, X_s)$. The property expresses that the distribution of X at the future time $s+t$ given the "past" up till time s is dependent on its value at the "present" time s only. The Markov kernels Q_t are called the *transition kernels* of the process.

Suppose that the functions μ and σ satisfy the conditions of Theorem 7.8. In the present situation these can be simplified to the existence, for every $t \geq 0$ of a constant C_t such that, for all $x, y \in [-t, t]$,

$$\begin{aligned} \big|\mu(x) - \mu(y)\big| &\leq C_t |x - y|, \\ \big|\sigma(x) - \sigma(y)\big| &\leq C_t |x - y|, \end{aligned} \tag{7.21}$$

and the existence of a constant C such that, for all $x \in \mathbb{R}$,

$$\begin{aligned} \big|\mu(x)\big| &\leq C(1 + |x|), \\ \big|\sigma(x)\big| &\leq C(1 + |x|). \end{aligned} \tag{7.22}$$

Under these conditions Theorem 7.8 guarantees the existence of a solution X^x to the diffusion equation with initial value $X_0^x = x$, for every $x \in \mathbb{R}$, and this solution is unique in law. The following theorem asserts that the distribution $Q_t(x, \cdot)$ of X_t^x defines a Markov kernel, and any solution to the diffusion equation is a Markov process with Q_t as its transition kernels.

Informally, given $\mathcal{F}_s$ and $X_s = x$ the distribution of X_{s+t} is the same as the distribution of X_t^x.

7.23 Theorem. *Assume that the functions $\mu, \sigma: \mathbb{R} \to \mathbb{R}$ satisfy (7.21)–(7.22). Then any solution X to the diffusion equation (7.3) is a Markov process with transition kernels Q_t defined by $Q_t(x, B) = \mathbb{P}(X_t^x \in B)$.*

Proof. See Chung and Williams, pages 235–243. These authors (and most authors) work within the canonical set-up where the process is (re)defined as the identity map on the space $C[0, \infty)$ equipped with the distribution induced by X^x. This is immaterial, as the Markov property is a distributional property; it can be written as

$$\mathrm{E}1_{X_{s+t} \in B}\, g(X_u : u \leq s) = \mathrm{E}Q_t(X_s, B) g(X_u : u \leq s),$$

for every measurable set B and bounded measurable function $g: C[0,s] \to \mathbb{R}$. This identity depends on the law of X only, as does the definition of Q_t.

The map $x \mapsto \int f(y)\, Q_t(x, dy)$ is shown to be continuous for every bounded continuous function $f: \mathbb{R} \to \mathbb{R}$ in Lemma 10.9 of Chung and Williams. In particular, it is measurable. By a monotone class argument this can be seen to imply that the map $x \mapsto Q_t(x, B)$ is measurable for every Borel set B. ∎

8
Option Pricing in Continuous Time

In this chapter we discuss the *Black-Scholes model* for the pricing of derivatives. Given the tools developed in the preceding chapters it is relatively straightforward to obtain analogues in continuous time of the discrete time results for the Cox-Ross-Rubinstein model of Chapter 3. The model can be set up for portfolios consisting of several risky assets, but for simplicity we restrict to one such asset.

We suppose that the price S_t of a stock at time $t \geq 0$ satisfies a stochastic differential equation of the form

$$dS_t = \mu_t S_t\,dt + \sigma_t S_t\,dW_t. \tag{8.1}$$

Here W is a Brownian motion process on a given filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, and $\{\mu_t: t \geq 0\}$ and $\{\sigma_t: t \geq 0\}$ are predictable processes. The filtration $\{\mathcal{F}_t\}$ is the completed natural filtration generated by W, and it is assumed that S is continuous and adapted to this filtration. The choices $\mu_t = \mu$ and $\sigma_t = \sigma$, for constants μ and σ, give the original Black-Scholes model. These choices yield a stochastic differential equation of the type considered in Chapter 7, and Theorem 7.8 guarantees the existence of a solution S in this case. (The solution can also be explicitly written as an exponential of Brownian motion with drift. See later.) For many other choices the existence of a solution is guaranteed as well. For our present purpose it is enough to assume that there exist a continuous adapted solution S.

The process σ is called the *volatility* of the stock. It determines how variable or “volatile” the movements of the stock are. We assume that this process is strictly positive. The process μ gives the drift of the stock. It is responsible for the exponential growth of a typical stock price.

Next to stocks our model allows for bonds, which in the simplest case are riskless assets with a predetermined yield, much as money in a savings account. More generally, we assume that the price R_t of a bond at time t

satisfies the differential equation

$$dR_t = r_t R_t\, dt, \qquad\qquad R_0 = 1.$$

Here r_t is some continuous adapted process called the *interest rate process*. (Warning: r is not the derivative of R, as might be suggested by the notation.) The differential equation can be solved to give

$$R_t = e^{\int_0^t r_s\, ds}.$$

This is the "continuously compounded interest" over the interval $[0, t]$. In the special case of a constant interest rate $r_t = r$ this reduces to $R_t = e^{rt}$.

A *portfolio* (A, B) is defined to be a pair of predictable processes A and B. The pair (A_t, B_t) gives the numbers of bonds and stocks owned at time t, giving the *portfolio value*

$$V_t = A_t R_t + B_t S_t. \tag{8.2}$$

The predictable processes A and B can depend on the past until "just before t" and we may think of changes in the content of the portfolio as a reallocation of bonds and stock that takes place just before time t. A portfolio is "self-financing" if such reshuffling can be carried out without import or export of money, whence changes in the value of the portfolio are due only to changes in the values of the underlying assets. More precisely, we call the portfolio (A, B) *self-financing* if

$$dV_t = A_t\, dR_t + B_t\, dS_t. \tag{8.3}$$

This is to be interpreted in the sense that V must be a semimartingale satisfying $V = V_0 + A \cdot R + B \cdot S$. It is implicitly required that A and B are suitable integrands relative to R and S.

A *contingent claim* with expiry time $T > 0$ is defined to be an $\mathcal{F}_T$-measurable random variable. It is interpreted as the value at the expiry time of a "derivative", a contract based on the stock. The European call option, considered in Chapter 3, is an important example, but there are many other contracts. Some examples of contingent claims are:

(i) *European call option*: $(S_T - K)^+$.
(ii) *European put option*: $(K - S_T)^+$.
(iii) *Asian call option*: $\bigl(\int_0^T S_t\, dt - K\bigr)^+$.
(iv) *lookback call option*: $S_T - \min_{0 \le t \le T} S_t$,
(v) *down and out barrier option*: $(S_T - K)^+ 1\{\min_{0 \le t \le T} S_t \ge H\}$.

The constants K and H and the expiry time T are fixed in the contract. There are many more possibilities; the more complicated contracts are referred to as *exotic options*. Note that in (iii)–(v) the claim depends on the history of the stock price throughout the period $[0, T]$. All contingent claims can be priced following the same no-arbitrage approach that we outline below.

A popular option that is not covered in the following is the *American put option*. This is a contract giving the right to sell a stock at any time in $[0, T]$ for a fixed price K. The value of this contract cannot be expressed in a contingent claim, because its value depends on an optimization of the time to exercise the contract (i.e. sell the stock). Pricing an American put option involves optimal stopping theory, in addition to the risk-neutral pricing we discuss below. A bit surprising is that a similar complication does not arise with the *American call option*, which gives the right to buy a stock at any time until expiry time. It can be shown that it is never advantageous to exercise a call option before the expiry time and hence the American call option is equivalent to the European call option.

Because the claims we wish to evaluate always have a finite term T, all the processes in our model matter only on the interval $[0, T]$. We may or must understand the assumptions and assertions accordingly.

In the discrete time setting of Chapter 3 claims are priced by reference to a "martingale measure", defined as the unique measure that turns the "discounted stock process" into a martingale. In the present setting the discounted stock price is the process $\tilde{S}$ defined by $\tilde{S}_t = R_t^{-1} S_t$. By Itô's formula and (8.1),

$$d\tilde{S}_t = -\frac{S_t}{R_t^2}\, dR_t + \frac{1}{R_t}\, dS_t = \frac{\mu_t - r_t}{\sigma_t}\frac{\sigma_t}{R_t} S_t\, dt + \frac{\sigma_t}{R_t} S_t\, dW_t. \tag{8.4}$$

Here and in the following we apply Itô's formula with the function $r \mapsto 1/r$, which does not satisfy the conditions of Itô's theorem as we stated it. However, the derivations are correct, as can be seen by substituting the explicit form for R_t as an exponential and next applying Itô's formula with the exponential function.

Under the true measure $\mathbb{P}$ governing the Black-Scholes stochastic differential equation (8.1) the process W is a Brownian motion and hence $\tilde{S}$ is a local martingale if its drift component vanishes, i.e. if $\mu_t \equiv r_t$. This will rarely be the case in the real world. Girsanov's theorem allows us to eliminate the drift part by a change of measure and hence provides the martingale measure that we are looking for. The process

$$\theta_t = \frac{\mu_t - r_t}{\sigma_t}$$

is called the *market price of risk*. If it is zero, then the real world is already "risk-neutral"; if not, then the process θ measures the deviation from a risk-neutral market relative to the volatility process.

Let $Z = \mathcal{E}(-\theta \cdot W)$ be the exponential process of $-\theta \cdot Z$, i.e.

$$Z_t = e^{-\int_0^t \theta_s\, dW_s - \frac{1}{2}\int_0^t \theta_s^2\, ds}.$$

We assume that the process θ is such that the process Z is a martingale (on $[0, T]$). For instance, this is true under Novikov's condition. We can next define a measure $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ by its density $d\tilde{\mathbb{P}} = Z_T \, d\mathbb{P}$ relative to $\mathbb{P}$. Then the process $\tilde{W}$ defined by

$$\tilde{W}_t = W_t + \int_0^t \theta_s \, ds$$

is a Brownian motion under $\tilde{\mathbb{P}}$, by Corollary 6.16, and, by the preceding calculations,

$$d\tilde{S}_t = \frac{\sigma_t}{R_t} S_t \, d\tilde{W}_t. \tag{8.5}$$

It follows that $\tilde{S}$ is a $\tilde{\mathbb{P}}$-local martingale. As in the discrete time setting the "reasonable price" at time 0 for a contingent claim with pay-off X is the expectation under the martingale measure of the discounted value of the claim at time T, i.e.

$$V_0 = \tilde{\mathrm{E}} R_T^{-1} X,$$

where $\tilde{\mathrm{E}}$ denotes the expectation under $\tilde{\mathbb{P}}$. This is a consequence of economic, no-arbitrage reasoning, as in Chapter 3, and the following theorem.

8.6 Theorem. *Suppose that the process $\mathcal{E}(\theta \cdot Q)^T$ is a martingale and define $d\tilde{\mathbb{P}} = \mathcal{E}(\theta \cdot Q)_T \, d\mathbb{P}$. Let X be a nonnegative contingent claim with $\tilde{\mathrm{E}} R_T^{-1} |X| < \infty$. Then there exists a self-financing strategy with value process V such that*
(i) $V \geq 0$ up to indistinguishability.
(ii) $V_T = X$ almost surely.
(iii) $V_0 = \tilde{\mathrm{E}} R_T^{-1} X$.

Proof. The process $\tilde{S} = R^{-1} S$ is a continuous semimartingale under $\mathbb{P}$ and a continuous local martingale under $\tilde{\mathbb{P}}$, in view of (8.5). Let $\tilde{V}$ be a cadlag version of the martingale

$$\tilde{V}_t = \tilde{\mathrm{E}}\big(R_T^{-1} X | \, \mathcal{F}_t\big).$$

Suppose that there exists a predictable process B such that

$$d\tilde{V}_t = B_t \, d\tilde{S}_t.$$

Then $\tilde{V}$ is continuous, because $\tilde{S}$ is continuous, and hence predictable. Define

$$A = \tilde{V} - B\tilde{S}.$$

Then A is predictable, because $\tilde{V}$, B and $\tilde{S}$ are predictable. The value of the portfolio (A, B) is given by $V = AR + BS = (\tilde{V} - B\tilde{S})R + BS = R\tilde{V}$ and hence, by Itô's formula and (8.4),

$$\begin{aligned} dV_t &= \tilde{V}_t \, dR_t + R_t \, d\tilde{V}_t = (A_t + B_t \tilde{S}_t) \, dR_t + R_t B_t \, d\tilde{S}_t \\ &= (A_t + B_t R_t^{-1} S_t) \, dR_t + R_t B_t \big(-S_t R_t^{-2} \, dR_t + R_t^{-1} \, dS_t\big) \\ &= A_t \, dR_t + B_t \, dS_t. \end{aligned}$$

Thus the portfolio (A, B) is self-financing. Statements (i)–(iii) of the theorem are clear from the definition of $\tilde{V}$ and the relation $V = R\tilde{V}$.

We must still prove the existence of the process B. In view of (8.5) we need to determine this process B such that

$$d\tilde{V}_t = B_t \frac{\sigma_t S_t}{R_t}\, d\tilde{W}_t.$$

The process $\tilde{W}$ is a $\tilde{\mathbb{P}}$-Brownian motion and $\tilde{V}$ is a $\tilde{\mathbb{P}}$-martingale. If the underlying filtration would be the completion of the natural filtration generated by $\tilde{W}$, then the representation theorem for Brownian local martingales, Theorem 6.6, and the fact that $\sigma_t S_t$ is strictly positive would immediately imply the result. By assumption the underlying filtration is the completion of the natural filtration generated by W. Because W and $\tilde{W}$ differ by the process $\int_0^t \theta_s\, ds$, it appears that the two filtrations are not identical and hence this argument fails in general. (In the special case in which μ_t, σ_t and r_t and hence θ_t are deterministic functions the two filtrations are clearly the same and hence the proof is complete at this point.) We can still prove the desired representation by a detour. We first write the $\tilde{\mathbb{P}}$-local martingale $\tilde{V}$ in terms of $\mathbb{P}$-local martingales through

$$\tilde{V}_t = \frac{\mathrm{E}(R_T^{-1} X Z_T |\, \mathcal{F}_t)}{\mathrm{E}(Z_T |\, \mathcal{F}_t)} = \frac{U_t}{Z_t}, \qquad \text{a.s..}$$

Here U, defined as the numerator in the preceding display, is a $\mathbb{P}$-martingale relative to $\{\mathcal{F}_t\}$. By the representation theorem for Brownian martingales the process U possesses a continuous version and there exists a predictable process C such that $U = U_0 + C\cdot W$. The exponential process $Z = \mathcal{E}(-\theta\cdot W)$ satisfies $dZ = Z\, d(-\theta \cdot W) = -Z\theta\, dW$ and hence $d[Z]_t = Z_t^2\theta_t^2\, dt$. Careful application of Itô's formula gives that

$$\begin{aligned} d\tilde{V}_t &= -\frac{U_t}{Z_t^2}\, dZ_t + \frac{dU_t}{Z_t} + \tfrac{1}{2}\frac{2U_t}{Z_t^3}\, d[Z]_t - \frac{1}{Z_t^2}\, d[U,Z]_t \\ &= -\frac{U_t}{Z_t^2}(-Z_t\theta_t)\, dW_t + \frac{C_t\, dW_t}{Z_t} + \frac{U_t}{Z_t^3}\, Z_t^2\theta_t^2\, dt + \frac{1}{Z_t^2} C_t Z_t \theta_t\, dt \\ &= \frac{U_t\theta_t + C_t}{Z_t}\, d\tilde{W}_t. \end{aligned}$$

This gives the desired representation of $\tilde{V}$ in terms of $\tilde{W}$. ∎

We interpret the preceding theorem economically as saying that $V_0 = \tilde{\mathrm{E}} R_T^{-1} X$ is the just price for the contingent claim X. In general it is not easy to evaluate this explicitly, but for Black-Scholes option pricing it is.

First the stock price can be solved explicitly from (8.1) to give

$$S_t = S_0 e^{\int_0^t (\mu_s - \frac{1}{2}\sigma_s^2)\, ds + \int_0^t \sigma_s\, dW_s}.$$

Because we are interested in this process under the martingale measure $\tilde{\mathbb{P}}$, it is useful to write it in terms of $\tilde{W}$ as

$$S_t = S_0 e^{\int_0^t (r_s - \frac{1}{2}\sigma_s^2)\, ds + \int_0^t \sigma_s \, d\tilde{W}_s}.$$

Note that the drift process μ does not make part of this equation: it plays no role in the pricing formula. Apparently the systematic part of the stock price diffusion can be completely hedged away. If the volatility σ and the interest rate r are constant in time, then this can be further evaluated, and we find that, under $\tilde{\mathbb{P}}$,

$$\log \frac{S_t}{S_0} \sim N\bigl((r - \tfrac{1}{2}\sigma^2)t, \sigma^2 t\bigr).$$

This is exactly as in the limiting case for the discrete time situation in Chapter 3. The price of a European call option can be written as, with Z a standard normal variable,

$$e^{-rT}\mathrm{E}\Bigl(S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z} - K\Bigr)^+.$$

It is straightforward calculus to evaluate this explicitly, and the result is given already in Chapter 3.

The exact values of most of the other option contracts mentioned previously can also be evaluated explicitly in the Black-Scholes model. This is more difficult, because the corresponding contingent claims involve the full history of the process S, not just the marginal distribution at some fixed time point.

If the processes σ and r are not constant, then the explicit evaluation may be impossible. In some cases the problem can be reduced to a partial differential equation, which can next be solved numerically.

Assume that the value process V of the replicating portfolio as in Theorem 8.6 can be written as $V_t = f(t, S_t)$ for some twice differentiable function f.[♯] Then, by Itô's formula and (8.1),

$$dV_t = D_1 f(t, S_t)\, dt + D_2 f(t, S_t)\, dS_t + \tfrac{1}{2} D_{22} f(t, S_t) \sigma_t^2 S_t^2 \, dt.$$

By the self-financing equation and the definition of $V = AR + BS$, we have that

$$dV_t = A_t \, dR_t + B_t \, dS_t = (V_t - B_t S_t) r_t \, dt + B_t \, dS_t.$$

The right sides of these two equations are identical if

$$D_1 f(t, S_t) + \tfrac{1}{2} D_{22} f(t, S_t)\sigma_t^2 S_t^2 = \bigl(V_t - B_t S_t\bigr) r_t,$$
$$D_2 f(t, S_t) = B_t.$$

♯ I do not know in what situations this is a reasonable assumption.

We can substitute $V_t = f(t, S_t)$ in the right side of the first equation, and replace B_t by the expression given in the second. If we assume that $\sigma_t = \sigma(t, S_t)$ and $r_t = r(t, S_t)$, then the resulting equation can be written in the form

$$f_t + \tfrac{1}{2} f_{ss} \sigma^2 s^2 = fr - f_s sr,$$

where we have omitted the arguments (t, s) from the functions f_t, f_{ss}, σ, f, f_s and r, and the indices t and s denote partial derivatives relative to t or s of the function $(t, s) \mapsto f(t, s)$. We can now try and solve this partial differential equation, under a boundary condition that results from the pay-off equation. For instance, for a European call option the equation $f(T, S_T) = V_T = (S_T - K)^+$ yields the boundary condition

$$f(T, s) = (s - K)^+.$$

8.7 EXERCISE. Show by an economic argument that the value of a call option at time t is always at least $(S_t - e^{-r(T-t)}K)^+$, where r is the (fixed) interest rate. [Hint: if not, show that any owner of a stock would gain riskless profit by: selling the stock, buying the option and putting $e^{-rt}K$ in a savings account, sitting still until expiry and hence owning an option and money K at time T, which is worth at least S_T.]

8.8 EXERCISE. Show, by "economic reasoning", that the early exercise of an American call option never pays. [Hint: if exercised at time t, then the value at time t is $(S_t - K)^+$. This is less than $(S_t - e^{-r(T-t)}K)^+$.]

8.9 EXERCISE. The *put-call parity* for European options asserts that the values P_t of a put and C_t of a call option at t with strike price K and expiry time T based on the stock S are related as $S_t + P_t = C_t + Ke^{-r(T-t)}$, where r is the (fixed) interest rate. Derive this by an economic argument, e.g. comparing portfolios consisting of one stock and one put option, or one call option and an amount Ke^{-rT} in a savings account. Which one of the two portfolios would you prefer?

9
Random Measures

A random measure is a map from a probability space into the collection of measures on a given measurable space. In this chapter the latter measurable space is the space $[0,\infty)\times\mathbb{D}$ for a given metric space $\mathbb{D}$. Then we obtain stochastic processes in "time" if we view the random measure as a function of the first coordinate. In particular, we are interested in integer-valued random measures, with as special example marked point processes.

9.1 Compensators

Let $(\Omega,\mathcal{F},\{\mathcal{F}_t\},P)$ be a filtered probability space and let $(\mathbb{D},\mathcal{D})$ be a complete separable metric space with its Borel σ-field $\mathcal{D}$. We call $\tilde{\mathcal{P}}=\mathcal{P}\times\mathcal{D}$ and $\tilde{\mathcal{O}}=\mathcal{O}\times\mathcal{D}$ the *predictable σ-field* and *optional σ-field* on the space $[0,\infty)\times\Omega\times\mathbb{D}$, respectively, and call a map $X\colon[0,\infty)\times\Omega\times\mathbb{D}\to\mathbb{R}$ *predictable* or *optional* if it is measurable relative to the σ-field $\tilde{\mathcal{P}}$ or $\tilde{\mathcal{O}}$. We may think of such a map as a stochastic process $(X_t\colon t\geq 0)$ with values in the space $\mathbb{D}$.

9.1 Definition. *A random measure on $[0,\infty)\times\mathbb{D}$ is a map $(\omega,B)\mapsto\mu(\omega,B)$ from $\Omega\times(\mathcal{B}_\infty\times\mathcal{D})\to\mathbb{R}$ such that:*
(i) The map $\omega\mapsto\mu(\omega,B)$ is measurable for every $B\in\mathcal{B}_\infty\times\mathcal{D}$.
(ii) The map $B\mapsto\mu(\omega,B)$ is a measure for every $\omega\in\Omega$.
(iii) $\mu(\omega,\{0\}\times\mathbb{D})=0$ for every $\omega\in\Omega$.

The first two requirements characterize a random measure as a transition kernel from Ω into $[0,\infty)\times\mathbb{D}$. If the total mass $\mu\bigl(\omega,[0,\infty)\times\mathbb{D}\bigr)$ were equal to 1 for every ω, then μ would be a Markov kernel from Ω into $[0,\infty)\times\mathbb{D}$. The third requirement corresponds to the usual convention that "nothing happens at time zero".

We shall often think of a random measure as the collection of stochastic processes $(t,\omega)\mapsto \mu\bigl(\omega,[0,t]\times D\bigr)$, for D ranging over $\mathcal{D}$. If these are finite for sufficiently many measurable sets D, then these processes give a complete description of the random measure, but this requirement is not included in the definition of a random measure. We can consider more generally, for a jointly measurable map $X\colon [0,\infty)\times\Omega\times\mathbb{D}\to\mathbb{R}$, the stochastic process $X*\mu\colon [0,\infty)\times\Omega\to\mathbb{R}$ defined by the Lebesgue integrals (if they exist)

$$(X*\mu)_t(\omega)=\int_{[0,t]}\int_{\mathbb{D}} X(s,\omega,y)\,\mu(\omega,ds,dy).$$

The positioning of the differential symbol "d" in such expressions as $\mu(\omega,ds,dy)$ is to indicate the arguments over which integration is carried out. As before, we often leave out the argument ω and write the preceding integral as $\int_0^t\int X_{s,y}\,d\mu_{s,y}$. The expectation of the process $X*\mu$ can be written in the form $\mathrm{E}(X*\mu)_t=\int X\,d(\mu\otimes P)$ for $\mu\otimes P$ the measure on the measurable space $\bigl([0,\infty)\times\Omega\times\mathbb{D},\mathcal{B}_\infty\times\mathcal{F}\times\mathcal{D}\bigr)$ given by

$$d\mu\otimes P(t,\omega,y)=\mu(\omega,dt,dy)\,dP(\omega).$$

This is to say that $\mu\otimes P([0,t]\times F\times D)=\int_F\mu\bigl(\omega,[0,t]\times D\bigr)\,dP(\omega)$, for every $t\ge 0$ and measurable sets $F\in\mathcal{F}$ and $D\in\mathcal{D}$.

The process $X*\mu$ has cadlag sample paths $t\mapsto (X*\mu)_t$. We shall be interested in random measures such that this process is adapted, in which case it is optional, at least for optional processes X. Furthermore, there is a special role for random measures such that the process $X*\mu$ is predictable for every predictable process X.

9.2 Definition. *The random measure μ is called:*
(i) predictable if the process $X*\mu$ is predictable for every nonnegative predictable process X.
(ii) optional if the process $X*\mu$ is optional for every nonnegative optional process X.
(iii) σ-finite if there exists a strictly positive, predictable map V with $\mathrm{E}V*\mu_\infty<\infty$.

9.3 EXERCISE. Suppose that the process $t\mapsto\mu\bigl(\omega,[0,t]\times D\bigr)$ is finite for a countable collection of sets D whose union is $\mathbb{D}$. Show that a random measure is predictable or optional if and only if the process $t\mapsto\mu\bigl(\omega,[0,t]\times D\bigr)$ is predictable for every $D\in\mathcal{D}$. [Hint: for a predictable X of the form $X=1_{[0,T]\times D}$ for a stopping time T and $D\in\mathcal{D}$ the process $(X*\mu)_t=Z_t^T$, for $Z_t=\mu\bigl([0,t]\times D\bigr)$, is predictable, because a stopped predictable process is predictable. Similarly, for an optional process of the type $X=1_{[T,\infty)\times D}$ the process $(X*\mu)_t=1_{[T,\infty)}(Z_t-Z_{T-})$ is optional. Extend by monotone class arguments.]

We shall only deal with σ-finite random measures. An equivalent description of σ-finiteness is that the measure $\mu\otimes P$ on the measurable space $\bigl([0,\infty)\times\Omega\times\mathbb{D},\tilde{\mathcal{P}}\bigr)$ is σ-finite in the usual sense.

Warning. The requirement that V in (iii) be predictable or the restriction of $\mu\otimes P$ to the predictable σ-field $\tilde{\mathcal{P}}$ in the preceding remark make the requirement of σ-finiteness stronger. For this reason other authors use the phrase "predictable σ-finite". Because we shall not consider other types of σ-finiteness, the abbreviation to "σ-finite" will not cause confusion.

If the process $Z = X*\mu$ is optional and locally integrable, then it possesses a compensator A by the Doob-Meyer theorem: a predictable process A such that $X*\mu - A$ is a local martingale. The following theorem shows that this compensator can be written in the form $A = X*\nu$ for a predictable random measure ν. This is called the *predictable projection* or *compensator* of μ.

9.4 Theorem. *For every optional, σ-finite random measure μ there exists a predictable random measure ν such $X*\mu - X*\nu$ is a local martingale for every optional process X such that $|X|*\mu$ is locally integrable.*

Proof. Let V be a strictly positive predictable process such that $V*\mu$ is predictable, and let A be the compensator of the process $V*\mu$. For every bounded measurable process X on $[0,\infty)\times\Omega\times\mathbb{D}$ the expectation $m(X) = \mathrm{E}\bigl((XV)*\mu\bigr)_\infty$ is well defined and defines a measure m on $([0,\infty)\times\Omega\times\mathbb{D},\mathcal{B}_\infty\times\mathcal{F}\times\mathcal{D})$. The martingale property of $V*\mu - A$ gives that, for every $s<t$ and $F_s\in\mathcal{F}_s$,

$$m\bigl((s,t]\times F_s\times\mathbb{D}\bigr) = \mathrm{E}1_{F_s}\bigl[(V*\mu)_t-(V*\mu)_s\bigr] = \mathrm{E}1_{F_s}(A_t-A_s).$$

The sets $\{0\}\times F_0$ and the sets of the form $(s,t]\times F_s$ generate the predictable σ-field $\mathcal{P}$. By assumption $m\bigl(\{0\}\times\Omega\times\mathbb{D}\bigr)=0$. Therefore the display shows that the restriction of the marginal of m on $[0,\infty)\times\Omega$ to $\mathcal{P}$ is given by the measure m_1 defined by $dm_1(t,\omega) = dA_t(\omega)\,dP(\omega)$. Let

$$dm(t,\omega,y) = dm_2(z|\,t,\omega)\,dA_t(\omega)\,dP(\omega)$$

be a disintegration of the restriction of m to $\mathcal{P}\times\mathcal{D}$ relative to its marginal m_1. Next define $\nu(\omega,dt,dz) = V(\omega,t,z)^{-1}dm_2(z|\,t,\omega)\,dA_t(\omega)$. ∎

9.5 Example (Increasing process). A cadlag increasing process A with $A_0=0$ defines a random measure on $[0,\infty)\times\{1\}$ through $\mu\bigl(\omega,[0,t]\times\{1\}\bigr) = A_t(\omega)$. That this defines a random measure for every fixed ω is a fact from measure theory; the measurability of $\omega\mapsto\mu\bigl(\omega,B\times\{1\}\bigr)$ is obvious for $B=[0,t]$, and follows by a monotone class argument for general Borel sets $B\subset[0,\infty)$.

The random measure μ is optional or predictable if and only if the process A is optional or predictable.

If the process A is locally integrable, then the random measure μ is σ-finite. If B is the compensator of A, then $X * B$ is the compensator of $X * A$ for every sufficiently regular process X. It follows that the compensator of μ is attached to the process B in the same way as μ is attached to A. □

Warning. It is not included in the definition of a random measure μ that the process $(t, \omega) \mapsto \mu(\omega, [0, t] \times D)$, for a fixed measurable set D, is finite-valued. Not even σ-finiteness need assure this. Thus we cannot in general identify these processes resulting from a random measure with increasing processes. In that sense the random measures in the preceding example are rather special. The jump measure of a semimartingale (see??) provides an example that motivates to work with random measures in the present generality.

9.2 Marked Point Processes

Random measures with values in the integers form a special class of random measures, which includes point processes and marked point processes. We deal only with σ-finite, integer-valued random measures. These correspond to random measures that, for each ω, are a counting measure on a countable set of points (which often will depend on ω). Here a "counting measure" is a discrete measure with atoms of probability one (or zero) only.

9.6 Definition. *A σ-finite random measure μ is called integer-valued if for almost every ω the measure $B \mapsto \mu(\omega, B)$ is a counting measure on a countable set of points in $(0, \infty) \times \mathbb{D}$ that intersects each set of the form $\{t\} \times \mathbb{D}$ in at most one point.*

* **9.7** EXERCISE. Show that a σ-finite random measure μ is integer-valued iff $\mu(\omega, B) \in \bar{\mathbb{N}}$ and $\mu(\omega, \{t\} \times \mathbb{D}) \in \{0, 1\}$ for every $B \in \mathcal{B}_\infty \times \mathcal{D}$ and every $(\omega, t) \in \Omega \times [0, \infty)$. [Hint: For a σ-finite, integer-valued random measure μ there exists a (predictable) strictly positive map V such that $V * \mu_\infty$ is finite almost surely. This implies that for almost every ω the measure $B \mapsto \mu(\omega, B)$ can have at most countably many atoms.]

Another way of defining a σ-finite integer-valued random measure is to say that it is a *point process* on $[0, \infty) \times \mathbb{D}$ satisfying the further requirement that each set $\{t\} \times \mathbb{D}$ contain at most one point. The latter restriction is somewhat odd. If t is interpreted as a "time" parameter, the property can be paraphrased by saying that "at most one event can happen" at each time point.

For a σ-finite integer-valued measure μ there exists for almost every ω at most countable many values $t = t(\omega) > 0$ such that $\mu(\omega, \{t\} \times \mathbb{D}) > 0$.

For each such t the measure of the set $\{t\}\times\mathbb{D}$ is exactly 1, and there exists a unique point $Z_t(\omega)\in\mathbb{D}$ such that $\mu(\omega,\{t\}\times\{Z_t(\omega)\})=1$. This explains the following representation, which shows that the points $t(\omega)$ can be chosen to be given by a sequence of stopping times if the random measure μ is optional.

9.8 Lemma. *A random measure μ is optional, σ-finite and integer-valued if and only if there exists a sequence of strictly positive stopping times T_n with disjoint graphs and an optional process $Z\colon[0,\infty)\times\Omega\to\mathbb{D}$ such that, up to evanescence,*

$$\mu(\omega,[0,t]\times D)=\sum_n 1_{T_n(\omega)\le t}1_D\big(Z_{T_n(\omega)}\big).$$

The optionality of the process Z in the preceding lemma implies that the variable Z_{T_n} is $\mathcal{F}_{T_n}$-measurable for every stopping time T_n. This suggests an interpretation of a "mark" Z_{T_n} being generated at the "event time" T_n. The random measure μ is the sum of the Dirac measures at all points $(T_n,Z_{T_n})\in(0,\infty)\times\mathbb{D}$.

The stopping times in the lemma cannot necessarily be ordered in a sequence $T_1\le T_2\le\cdots$. There may not be a smallest time T_n and the values T_n may accumulate at several, even infinitely many, points. Integer-valued random measures for which the sequence T_n can be ordered in a sequence are of special interest. We call an integer-valued random measure μ a *marked point process* if it possesses a representation as in the preceding lemma for a strictly increasing sequence of stopping times $0<T_1<T_2<\cdots$. We may then think of these stopping times as the times of a sequence of events and of $Z_{T_1},Z_{T_2},\ldots$ as "marks" that are generated at the consecutive event times.

A *multivariate point process* is the further specialization of a marked point process with a finite mark space. If the mark space is $\mathbb{D}=\{1,\ldots,k\}$ and the number of events in finite intervals is finite, then we may identify the marked point process with the vector $(N_1,\ldots,N_k)$ of processes $(N_i)_t(\omega)=\mu\big(\omega,[0,t]\times\{i\}\big)$.

Even if the times $T_1<T_2<\cdots$ can be ordered in a sequence, the general definition of an integer-valued random measure does not imply that this sequence increases indefinitely. If $T_n\uparrow T$ for a finite limit T, then the corresponding marked point process is said to be *explosive.* A simple example is a process $N\circ\Phi$ for N a Poisson process and Φ an increasing map of $[0,1]$ onto $[0,\infty]$.

Warning. Some authors restrict the term "multivariate point process" to point processes without explosion.

The compensator ν of an integer-valued random measure is typically not integer-valued, and may even have no atoms at all. In the proof of the following lemma we establish the following identity, which characterizes the atoms in the compensator as probabilities of "immediate" jumps in the random measure. For every predictable time T:

$$\mathbb{P}\Big(\mu\big(\{T\}\times\mathbb{D}\big)=1|\,\mathcal{F}_{T_-}\Big)=\nu\big(\{T\}\times\mathbb{D}\big). \tag{9.9}$$

The lemma shows that a compensator can always be chosen such that $\nu(\omega,\{t\}\times\mathbb{D})\le 1$ identically, mimicking this property of μ.

9.10 Lemma. *Every integer-valued random measures possesses a compensator ν with $\nu(\omega,\{t\}\times\mathbb{D})\le 1$ for every (t,ω).*

Proof. Given a predictable time T the process $(t,\omega,y)\mapsto 1_{[T]}(t,\omega)$ is predictable and hence the process $M=1_{[T]}*(\mu-\nu)$ is a local martingale, and even a martingale, because the process $1_{[T]}*\mu$ is bounded (by 1). The process M can be alternatively described as $M_t=(\mu-\nu)\big(\{T\}\times\mathbb{D}\big)1_{T\le t}$ and hence its jump process at T is given by $\Delta M_T=(\mu-\nu)\big(\{T\}\times\mathbb{D}\big)$. The martingale property gives that $\mathrm{E}\big(\Delta M_T|\,\mathcal{F}_{T-}\big)=0$ almost surely. This can be written in the form (9.9).

It follows that $\nu(\omega,\{T\}\times\mathbb{D})\le 1$ almost surely, for every predictable time T. The set $\big\{(t,\omega)\colon\nu(\omega,\{t\}\times\mathbb{D})>0\big\}$ where the compensator possesses atoms are the locations of the jumps of the predictable process $t\mapsto\nu\big(\omega,[0,t]\times\mathbb{D}\big)$ and hence is exhausted by a sequence of predictable stopping times. Thus if we redefine the compensator on the set where $\nu(\omega,\{t\}\times\mathbb{D})>1$, then we redefine it at most on the union of countably many graphs of predictable times and on each graph at most on a null set. Thus this gives another predictable measure that differs by evanescence and possesses the property as in the lemma. ∎

If the filtration

9.3 Jump Measure

Every cadlag function jumps at most a countable number of times, and hence the jumps of a cadlag stochastic process X occur at most at a countable number of "random times". For a cadlag, adapted process X these times are given by a sequence of stopping times: there exist stopping times $S_1,S_2,\ldots$ such that

$$\big\{(t,\omega)\colon\Delta X(t,\omega)\ne 0\big\}=\bigcup_n[S_n].$$

(See below.) The graphs of the stopping times may be taken disjoint without loss of generality. This allows to define an integer-valued random measure on $[0,\infty)\times\mathbb{R}$ by, for D a Borel set in $\mathbb{R}$,

$$\mu^X(\omega,[0,t]\times D)=\sum_n 1_{S_n(\omega)\le t}1_D(\Delta X_{S_n(\omega)}).$$

Because the jump process ΔX of a cadlag, adapted process is optional, the *jump measure* μ^X is an optional random measure by Lemma 9.8. (By convention there is not jump at time 0.)

To construct the stopping times, fix a sequence of numbers $\varepsilon_0=\infty>\varepsilon_1>\varepsilon_2\cdots\downarrow 0$. Because the points in $[0,\infty)$ where a given cadlag function jumps more than a given positive number ε are isolated, we can define, for given $n,k\in\mathbb{N}$, a variable $S_{n,k}$ as the kth jump of X of size in $[\varepsilon_n,\varepsilon_{n-1})$. We can write these variables also in the form: for every $n\in\mathbb{N}$ and $k=1,2,\ldots$,

$$S_{n,0}=0,\qquad S_{n,k}=\inf\bigl\{t>0\colon |\Delta X_t|1_{t>S_{n,k-1}}\in[\varepsilon_n,\varepsilon_{n-1})\bigr\}.$$

Because the process ΔX is progressively measurable and $S_{n,k}$ is the hitting time of the interval $[\varepsilon_n,\varepsilon_{n-1})$ by the process $|\Delta X|1_{(S_{n,k-1},\infty)}$, it follows that the $S_{n,k}$ are stopping times. Their graphs are disjoint and exhaust the jump times of X.

The sets $B_{n,k}=[0,S_{n,k}]\times\bigl([\varepsilon_n,\varepsilon_{n-1})\cup(-\varepsilon_{n-1},-\varepsilon_n]\bigr]$ are predictable, and cover $[0,\infty)\times\Omega\times\mathbb{R}$ if k,n range over $\mathbb{N}$. By construction $\mu^X(B_{n,k})\le k$ for every n,k. Thus the function $V=\sum_{n,k}2^{-n-k}1_{B_{n,k}}$ is strictly positive and $(V*\mu^X)_\infty\le\sum_{n,k}k2^{-n-k}<\infty$. It follows that the jump measure μ^X of a cadlag, adapted process X is σ-finite.

A cadlag process X is called *quasi left-continuous* if for every increasing sequence of stopping times $0\le T_n\uparrow T$ we have that $X_{T_n}\to X_T$ almost surely on the event $\{T<\infty\}$. Because fixed times are stopping times, this requirement includes in particular that $X_{t_n}\to X_t$ almost surely for every deterministic sequence $t_n\uparrow t$ and every t. However, the exceptional null set where convergence fails may depend on t and hence quasi left-continuous can be far from "left-continuous". It can be characterized by the continuity of the compensator of the jump measure of X. This can be derived from the identity (9.9), which in the present case takes the form: for every predictable time T,

$$\nu^X\bigl(\{T\}\times\mathbb{D}\bigr)=\mathbb{P}\bigl(\Delta X_T\ne 0|\,\mathcal{F}_{T-}\bigr).$$

9.11 Lemma. *A cadlag, adapted process X is quasi left-continuous if and only if there exists a version of the compensator ν^X of μ^X such that $\nu^X\bigl(\omega,\{t\}\times\mathbb{R}\bigr)=0$ for all $t\in[0,\infty)$ and $\omega\in\Omega$.*

Proof. The identity stated before the lemma implies that $E\nu^X\bigl(\{T\}\times\mathbb{D}\bigr)=\mathbb{P}(\Delta X_T\ne 0)$ for every predictable time T. It follows the variable $\nu^X\bigl(\{T\}\times\mathbb{D}\bigr)$ is zero almost surely if and only if this is the case for the variable ΔX_T.

We conclude the proof by showing that a process X is quasi left-continuous if and only if $\Delta X_T 1_{T<\infty} = 0$ for every predictable time T.

For any predictable time T there exists a sequence of stopping times $T_n \uparrow T$ with $T_n < T$ whenever $T > 0$. If X is quasi left-continuous, then $X_T = \lim X_{T_n} = X_{T-}$ almost surely on $\{T < \infty\}$ and hence $\Delta X_T = 0$.

Conversely, if T_n are stopping times with $T_n \uparrow T$, then $S_n = T 1_{T_n<T} + \infty 1_{T_n=T}$ are stopping times with $S_n \uparrow S = T 1_F + \infty 1_{F^c}$, for $F = \cap_n \{T_n < T\}$. The stopping times $S_n \wedge n$ also increase to S and satisfy $S_n < S$. It follows that S is a predictable time. If X_{T_n} does not converge to X_T, then clearly $T_n < T$ for all n and hence $X_{S_n} = X_{T_n}$ does not converge to $X_T = X_S$, whence $\Delta X_S 1_{S<\infty} \neq 0$. ∎

9.4 Change of Measure

An optional random measure μ defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ is obviously still an optional random measure on the same filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, Q)$ but equipped with another probability measure Q. In this section we show that if Q is absolutely continuous relative to P, then the compensator ν_Q of μ relative to Q is absolutely continuous relative to its compensator ν_P relative to P, and characterize the density $d\nu_Q/d\nu_P$ as a conditional expectation of the density process of Q relative to P.

Letting Q_t and P_t denote the restrictions of Q and P to $\mathcal{F}_t$, the density process of Q relatively to P is the process L with

$$L_t = \frac{dQ_t}{dP_t}.$$

If $Q \ll P$, then the process L is a uniformly integrable P-martingale. The definition as a Radon-Nikodym density for every t does define the process L only up to a null set for every t, but L is unique up to evanescence if we use the cadlag version of this martingale, as we will.

For a measurable process $Y\colon [0,\infty) \times \Omega \times \mathbb{D} \to [0,\infty]$ let $Y\nu$ denote the random measure

$$(Y\nu)\big(\omega, [0,t] \times \mathbb{D}\big) = \int_{[0,t]\times\mathbb{D}} Y(s,\omega,z)\,\nu(\omega, ds, dz).$$

If Y is a predictable process and ν a predictable random measure, then $Y\nu$ is a predictable random measure. The predictable measure $Y\nu$ is the Q-compensator of μ if and only if the process $X * \mu - X * (Y\nu)$ is a Q-local martingale for every sufficiently integrable, predictable process X. By

Lemma 6.13 this is the case if the process $L\bigl(X*\mu - X*(Y\nu)\bigr)$ is a P-local martingale. We shall show that this is the case for the process Y defined by

$$Y=\frac{\mathrm{E}_{\mu\otimes P}(L|\,\tilde{\mathcal{P}})}{L_-}1_{L_->0}.$$

Here $\mathrm{E}_{\mu\otimes P}(L|\,\tilde{\mathcal{P}})$ is a *generalized conditional expectation* of the random variable L defined on the measure space $\bigl([0,\infty)\times\Omega\times\mathbb{D},\mathcal{B}_\infty\times\mathcal{F}\times\mathcal{D},\mu\otimes P\bigr)$. If the measure $\mu\otimes P$ were a probability measure, then $\mathrm{E}_{\mu\otimes P}(L|\,\tilde{\mathcal{P}})$ would be the ordinary conditional expectation. The "generalized" is because the measure $\mu\otimes P$ may be σ-finite only.

For a process $L\ge 0$ the conditional expectation $L'=\mathrm{E}_{\mu\otimes P}(L|\,\tilde{\mathcal{P}})$ is defined to be a $\tilde{\mathcal{P}}$-measurable process such that $(\mu\otimes P)L'X=(\mu\otimes P)LX$ for every $\tilde{\mathcal{P}}$-measurable, nonnegative process X. The existence of this generalized conditional expectation follows from the Radon-Nikodym theorem, applied to the measure $B\mapsto(\mu\otimes P)L1_B$ on $([0,\infty)\times\Omega\times\mathbb{D},\tilde{\mathcal{P}})$, provided that this measure is σ-finite. This measure is well defined on the larger σ-field $\mathcal{B}_\infty\times\mathcal{F}\times\mathbb{D}$, and is clearly absolutely continuous relative to itself. The σ-finiteness refers to its restriction to the predictable σ-field $\tilde{\mathcal{P}}$, and is not an automatic consequence of the σ-finiteness of μP, which we assume throughout, as L is typically not predictable. If it defined, then the conditional expectation is unique and finite-valued up to $\mu\otimes P$-null sets. For a general, not necessarily nonnegative, process L the conditional expectation $\mathrm{E}_{\otimes P}(L\tilde{\mathcal{P}})$ can be defined by taking differences, provided the measure $B\mapsto(\mu\otimes P)|L|1_B$ is σ-finite on $\tilde{\mathcal{P}}$.

Warning. In the present case the variable L depends on the first two arguments of (t,ω,z) only, but $\mathrm{E}_{\mu\otimes P}(L|\,\tilde{\mathcal{P}})$ is by definition a function of all three.

9.12 Theorem. *If μ is a P-σ-finite, optional random measure with compensator ν and $Q\ll P$, then μ is a Q-σ-finite, optional random measure with compensator $Y\nu$.*

Proof. As explained, it suffices to prove that the process $L\bigl(X*\mu-X*(Y\nu)\bigr)$ is a P-local martingale, for a sufficiently large collection of predictable processes X. This follows by taking differences if we can show that the following two processes are P-local martingales:

$$L(X*\mu)-L_-\cdot\bigl(X*(Y\nu)\bigr),$$
$$L\bigl(X*(Y\nu)\bigr)-L_-\cdot\bigl(X*(Y\nu)\bigr).$$

The definition of the quadratic variation process, or Itô's formula, allows to rewrite the second process as

$$\bigl(X*(Y\nu)\bigr)_-\cdot L+\bigl[L,X*(Y\nu)\bigr]=\bigl(X*(Y\nu)\bigr)\cdot L,$$

because the process $X*(Y\nu)$ is predictable and of locally bounded variation. (Cf. Exercise 9.13.) In view of the predictability of the process $X*(Y\nu)$ the right side is a P-local martingale.

9.13 EXERCISE. Let X be a predictable process of locally bounded variation and M a martingale. Show that $X\cdot M = X.M + [X,M]$, provided that the stochastic integrals exist. [Hint: Show that $[X,M] = \sum_s(\Delta X_s)(\Delta M_s) = \Delta X\cdot M$ and use the linearity of the integral.]

By the same arguments,

$$L(X*\mu) = (X*\mu)_-\cdot L + L\cdot (X*\mu).$$

The first term on the right is a P-local martingale. To conclude the proof we show that the second process, which of locally bounded variation, possesses compensator $L_-\cdot\bigl(X*(Y\nu)\bigr)$. For any $X\geq 0$ and stopping time T,

$$\mathrm{E}\bigl(L\cdot (X*\mu)\bigr)_T = \mathrm{E}_{\mu P}1_{[0,T]}LX = \mathrm{E}_{\mu P}1_{[0,T]}\mathrm{E}_{\mu P}(L|\,\tilde{\mathcal{P}})X.$$

By Lemma 6.12 the sample paths of L remain at zero if they ever reach zero. In particular $L_- = 0$ implies that $L=0$ and hence we may restrict the integral to the set $L_- > 0$. Then the definition of Y allows to rewrite the expectation as

$$\begin{aligned}\mathrm{E}_{\mu P}1_{[0,T]}YL_-X &= \mathrm{E}\bigl[L_-\cdot\bigl((XY)*\mu\bigr)\bigr]_T\\ &= \mathrm{E}\bigl[L_-\cdot\bigl((XY)*\nu\bigr)\bigr]_T = \mathrm{E}\bigl[L_-\cdot\bigl(X*(Y\nu)\bigr)\bigr]_T.\end{aligned}$$

In the one before last step we use that the process $Z*\mu - Z*\nu$, and hence the process $Z'\cdot(Z*\mu - Z*\nu)$ is a local martingale, for every predictable processes Z and Z'. [Need integrability??] We conclude that the process $L\cdot(X*\mu) - L_-\cdot\bigl(X*(Y\nu)\bigr)$ is a local martingale. ∎

9.5 Reduction of Flow

Let μ be a σ-finite, optional random measure on a filtered probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t\},P)$ with compensator ν. If the processes $t\mapsto \mu\bigl([0,t]\times D\bigr)$, for $D\in\mathcal{D}$, are adapted to a smaller filtration $\mathcal{G}_t\subset\mathcal{F}_t$, then μ is also an optional random measure relative to the filtration $\mathcal{G}_t$. Unless the compensator ν is predictable relative to the smaller filtration, it cannot be the compensator of μ relative to $\mathcal{G}_t$. In this section we show that it can be obtained as a conditional expectation. In the general case, this relationship remains somewhat abstract, but the formula takes a simple and intuitive form if ν possesses a density relative to a fixed measure.

9.14 Theorem. *Let μ be a σ-finite random measure on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with compensator ν such that $\nu(\omega, dt, dy) = a_{t,y}(\omega)\, d\lambda(t,y)$ for a nonrandom σ-finite Borel measure λ on $[0,\infty) \times \mathbb{D}$. If the process β is nonnegative and predictable relative to the filtration $\mathcal{G}_t \subset \mathcal{F}_t$, then the measure π given by $\pi(\omega, dt, dy) = b_{t,y}(\omega)\, d\lambda(t,y)$ is the compensator of μ on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{G}_t\}, \mathbb{P})$ if and only if and $\mathrm{E}(a_{t,y}|\, \mathcal{G}_{t-}) = b_{t,y}$ for λ-almost every pair (t,y),.*

Proof. The measure π is a predictable random measure.

For any $\mathcal{G}_t$-optional process X the difference $M = X * \mu - X * \nu$ is an $\mathcal{F}_t$-local martingale and hence $\mathrm{E}(M_t - M_s|\, \mathcal{G}_s) = 0$ for every $s < t$. It suffices to show that $X * \nu - X * \pi$ is an $\mathcal{G}_t$-local martingale, i.e. π is the compensator of ν relative to the filtration $\mathcal{G}_t$.

For any $\mathcal{G}_t$-predictable process X, the variable $X_{t,y}$ is $\mathcal{G}_{t-}$-measurable, for every (t,y). Therefore, by Fubini's theorem, for every sufficiently integrable $\mathcal{G}_t$-predictable process X.

$$\begin{aligned}\mathrm{E}(X * \nu - X * \pi)_\infty &= \int_0^\infty \int_{\mathbb{D}} \mathrm{E} X_{t,y} (a_{t,y} - b_{t,y})\, d\lambda(t,y)\\ &= \int_0^\infty \int_{\mathbb{D}} \mathrm{E} X_{t,y} \bigl(\mathrm{E}(a_{t,y}|\, \mathcal{G}_{t-}) - b_{t,y}\bigr)\, d\lambda(t,y).\end{aligned}$$

If b satisfies the condition of the theorem, then the right side vanishes. If π' is the $\mathcal{G}_t$-compensator of ν, then $X * \nu - X * \pi'$ is a martingale with mean zero and hence we conclude that $\mathrm{E}(X * \pi' - X * \pi)_\infty = 0$ for every $\mathcal{G}_t$-predictable process X. In view of the predictability of $\pi' - \pi$, this implies that $\pi' = \pi$.

Conversely, if π is the compensator of ν, then the left side of the preceding display vanishes for every $\mathcal{G}_t$-predictable X. Choosing X equal to the process $(s,y) \mapsto h(s,y) 1_G 1_{[t,\infty)(s)}$ for measurable functions $h\colon [0,\infty) \times \mathbb{D} \to \mathbb{R}$ and events $G \in \mathcal{G}_{t-}$, we conclude that $\mathrm{E} 1_G \bigl(\mathrm{E}(a_{t,y}|\, \mathcal{G}_{t-}) - b_{t,y}\bigr) = 0$ for λ-almost every (t,y), and every $G \in \mathcal{G}_{t-}$. This implies that b satisfies the condition of the theorem. ∎

The assertion of the theorem is particularly intuitive in the case of multivariate counting processes. Suppose that N is a nonexplosive counting processes with compensator A relative to the filtration $\mathcal{F}_t$ of the form $A_t = \int_0^t a_s\, ds$ for a predictable "intensity process" a. Then the intensity process of N relative to a smaller filtration $\mathcal{G}_t$ satisfies, for Lebesgue almost every t,

$$b_t = \mathrm{E}(a_t|\, \mathcal{G}_{t-}), \qquad \text{a.s.}.$$

Because the intensity a_t can be interpreted as the conditional infinitesimal probability of a jump at t given the past $\mathcal{F}_{t-}$, this expresses the $\mathcal{G}_t$-intensity as the expected value of the intensity given the "smaller past" $\mathcal{G}_{t-}$.

A predictable process b as in the preceding display always exists (up to integrability?). Indeed, the *predictable projection* of the process a relative to the filtration $\mathcal{G}_t$ is a $\mathcal{G}_t$-predictable process b such that, for every $\mathcal{G}_t$-predictable time T,

$$b_T = \mathrm{E}(a_T \,|\, \mathcal{G}_{T-}), \qquad \text{a.s..}$$

Because the constant stopping time $T = t$ is predictable this strengthens the preceding display. The last display determines the process b up to evanescence.

These observations extend to multivariate, nonexplosive counting processes $N = (N_1, \ldots, N_k)$ in an obvious way.

9.6 Stochastic Integrals

For sufficiently regular processes X the processes $X * \mu$ and $X * \nu$ are defined as Lebesgue-Stieltjes integrals (as previously) and hence so is their difference

$$X * (\mu - \nu)_t = \int_{[0,t]} \int_{\mathbb{D}} X_{s,y} \, d(\mu_{s,y} - \nu_{s,y}).$$

It is sometimes useful to define "integrals" $X * (\mu - \nu)$ relative to a compensated random measure for a slightly larger class of predictable processes X. By the definition of ν as a compensator, the process $X * \mu - X * \nu$ is a local martingale for every predictable process $X\colon [0, \infty) \times \Omega \times \mathbb{D} \to \mathbb{R}$. It is of locally bounded variation and hence is a purely discontinuous local martingale. Because the difference of two purely discontinuous local martingales with the same jump process is a continuous local martingale, it is constant and hence a purely discontinuous local martingale is completely determined by its jump process (Cf. Section 5.11.) Therefore, within the class of purely discontinuous local martingales the process $X * \mu - X * \nu$ is uniquely determined by its jump process

$$\Delta(X * \mu - X * \nu)_t(\omega) = \int_{\mathbb{D}} X(t, \omega, y) \, (\mu - \nu)(\omega, \{t\} \times dy) =: \tilde{X}_t(\omega).$$

The right side of this display avoids the integral with respect to the time variable t and may be well defined (as Lebesgue integrals, for each fixed ω) even if the Lebesgue-Stieltjes integrals defining $X * \mu$ and $X * \nu$ are not. This observation can be used to extend the definition of the integral $X * (\mu - \nu)$.

9.15 Definition. *Given a predictable process X such that the right side $\tilde{X}_t$ of the preceding display is well defined and finite for every t and such that the process $\bigl(\sum_{s \le t} \tilde{X}_s^2\bigr)^{1/2}$ is locally integrable, the stochastic integral $X * (\mu - \nu)$ is defined to be the purely discontinuous local martingale with jump process $\Delta\bigl(X * (\mu - \nu)\bigr)$ equal to $\tilde{X}_t$.*

To justify this complicated definition it must be shown that a purely discontinuous local martingale as in the definition exists and is unique. The uniqueness is clear from the fact that purely discontinuous local martingales are uniquely determined by their jump processes. Existence is a more complicated matter and requires a construction. The condition on the process $\tilde{X}$ comes from the fact that the cumulative square jump process $\sum_{s\leq t}(\Delta Y_s)^2$ of a local martingale Y is bounded above by the quadratic variation process $[Y]_t$. The square root of the quadratic variation process can be shown to be locally integrable. (See Exercise 5.66.) Thus the definition goes as far and abstract as possible in terms of relaxing the requirement of integrability of ΔX relative to $\mu - \nu$.

It is good to know the extent to which definitions can be pushed. However, in the following we shall only encounter integrals where the existence follows from the context, so that it is not a serious omission to accept that the definition is well posed without proof.†

The stochastic integral $X * (\mu - \nu)$ is completely determined by its jump process. The following lemma makes this explicit for integer-valued random measures.

9.16 Lemma. *If a local martingale M can be written in the form $M = X * (\mu - \nu)$ for a predictable process X and an integer-valued random measure μ with compensator ν, then it can be written in this form for*

$$X(t,\omega,y) = \mathrm{E}_{\mu P}(\Delta M|\,\tilde{\mathcal{P}})(t,\omega,y) + \frac{\int \mathrm{E}_{\mu P}(\Delta M|\,\tilde{\mathcal{P}})(t,\omega,z)\,\nu\bigl(\omega,\{t\}\times dz\bigr)}{1-\nu\bigl(\omega,\{t\}\times\mathbb{D}\bigr)}.$$

Proof. Define $U = \mathrm{E}_{\mu P}(\Delta M|\,\tilde{\mathcal{P}})$ and set

$$X'_{t,y} = U_{t,y} + \frac{\int U_{t,z}\,\nu\bigl(\{t\}\times dz\bigr)}{1-\nu\bigl(\{t\}\times\mathbb{D}\bigr)} =: U_{t,y} + \frac{\hat{U}_t}{1-a_t},$$

(say.) Straightforward algebra shows that the jump process of the process $X' * (\mu - \nu)$ is given by

$$\begin{aligned}&\int X_{t,y}\,(\mu-\nu)\bigl(\{t\}\times dy\bigr)\\ &\qquad= \int X_{t,y}\,\mu\bigl(\{t\}\times dy\bigr) - \hat{U}_t + \frac{\hat{U}_t}{1-a_t}(\mu-\nu)\bigl(\{t\}\times\mathbb{D}\bigr)\\ &\qquad= \int X_{t,y}\,\mu\bigl(\{t\}\times dy\bigr) - \frac{\hat{U}_t}{1-a_t}\bigl[1-\mu\bigl(\{t\}\times\mathbb{D}\bigr)\bigr].\end{aligned}$$

For a fixed ω the measure $B \mapsto \mu(\omega, B)$ is a counting measure on a countable set of points $\bigl(t, Z_t(\omega)\bigr)$. For every fixed ω the variable $1-\mu\bigl(\omega,\{t\}\times\mathbb{D}\bigr)$ is

† See e.g. Jacod and Shiryaev, II and I .

0 almost surely under this measure and the process $U_{t,y}$ is equal to U_{t,Z_t} almost surely. If follows that under the measure $\otimes P$ the right side of the preceding display is almost surely equal to the process U_{t,Z_t}. It follows that the jump process of $M - X' * (\mu - \nu) = (X - X') * (\mu - \nu)$ is given by $\Delta M_t - U_{t,Z_t}$. Taking again the support of the measure $\mu \otimes P$ into account, we see that, if considered a function on $[0,\infty) \times \Omega \times \mathbb{D}$, this process is $\mu \otimes P$-almost surely equal to the process $(t,\omega,y) \mapsto \Delta M_t(\omega) - U_{t,y}(\omega)$, and hence $\mathrm{E}_{\mu\otimes P}\big(\Delta(M - X' * (\mu - \nu))|\,\tilde{\mathcal{P}}\big) = \mathrm{E}_{\mu\otimes P}\big(\Delta M - U|\,\tilde{P}\big) = 0$ almost surely under $\mu \otimes P$.

The proof is complete if it can be shown that any martingale of the form $N = X * (\mu - \nu)$ with $\mathrm{E}_{\mu\otimes P}(\Delta N|\,\tilde{\mathcal{P}}) = 0$ is evanescent. Because N is purely discontinuous, it suffices to show that its jump process ΔN is evanescent. This jump process is given by

$$\Delta N_t = X_{t,Z_t}\mu\big(\{t\} \times \mathbb{D}\big) - \hat{X}_t.$$

Under the measure $\mu \otimes P$ the process X_{t,Z_t}, if seen as function on $[0,\infty) \times \Omega \times \mathbb{D}$, is almost surely equal to the process $X_{t,y}$, which is predictable by assumption. It follows that $\Delta N_t = X_{t,y} - \hat{X}_t$ almost surely under $\mu \otimes P$ and hence $X_{t,y} - \hat{X}_t$ is a version of $\mathrm{E}_{\mu\otimes P}(\Delta N|\,\tilde{\mathcal{P}}) = 0$. Therefore for P-almost every ω we have that $X_{t,Z_t(\omega)}(\omega) - \hat{X}_t(\omega) = 0$ for every t such that $\mu\big(\omega,\{t\} \times \mathbb{D}\big) > 0$. Because $\mu\big(\omega,\{t\} \times \mathbb{D}\big) = 0$ for other values of t we can conclude that $\Delta N_t(\omega) = \hat{X}_t(\omega)\big[1 - \mu\big(\omega,\{t\} \times \mathbb{D}\big)\big]$ for P-almost every ω.

As N is a martingale, we have that $\mathrm{E}(\Delta N_T|\,\mathcal{F}_{T-}) = 0$ for every predictable time T. Combined with the preceding identity this gives that $\hat{X}_t(1 - a_t) = 0$ almost surely and hence $\Delta N_t 1_{a_t<1} = 0$ almost surely.

Finally it suffices to show that $\Delta N_t 1_{a_t=1} = 0$. Now the set $\{a = 1\} = \{(t,\omega)\colon a_t(\omega) = 1\}$ is a subset of $\{a > 0\}$, which is union of the graphs of countable many predictable times. For any predictable time T () gives $\mathrm{E}\mu\big(\{T\} \times \mathbb{D}\big)1_{a_T=1} = \mathrm{E}a_T 1_{a_T=1} > 0$ unless $\{a_T > 0\}$ is a null set. It follows that the set $\{a = 1\}$ is contained in the set $\{(t,\omega)\colon \mu\big(\omega,\{t\} \times \mathbb{D}\big) = 1\}$. ∎

Warning. The preceding lemma does not claim that a process X satisfying $M = X * (\mu - \nu)$ is uniquely determined.

As indicated before the definition, if $X * \mu$ and $X * \nu$ are well defined as Lebesgue-Stieltjes integrals, then $X*(\mu-\nu)$ is the same as $X*\mu - X*\nu$. Some well-known properties of ordinary integrals also generalize.

9.17 Lemma. *If X is a predictable process such that $X * (\mu - \nu)$ is well defined, then:*

(i) $(X1_{[0,T]}) * (\mu - \nu) = \big(X * (\mu - \nu)\big)^T$ *for every stopping time T.*

(ii) $(YX)*(\mu-\nu) = Y \cdot \big(X*(\mu-\nu)\big)$ *for every bounded predictable process Y.*

10 Stochastic Calculus

In this chapter we continue the calculus for stochastic processes, extending this to general semimartingales.

10.1 Characteristics

Every cadlag function jumps at most a countable number of times, and hence the jumps of a cadlag stochastic process X occur at most at a countable number of "random times". For a cadlag, adapted process X these times can be shown to be given by a sequence of stopping times $T_1, T_2, \ldots$, in the sense that

$$\bigl\{(t,\omega)\colon \Delta X \neq 0\bigr\} = \cup_n [T_n].$$

The graphs of the stopping times may be taken disjoint without loss of generality. This allows to define an integer-valued random measure on $[0,\infty)\times\mathbb{R}$ by, for D a Borel set in $\mathbb{R}$,

$$\mu^X(\omega, [0,t]\times D) = \sum_n 1_{T_n(\omega)\le t} 1_D(\Delta X_{T_n(\omega)}).$$

Because the jump process ΔX of a cadlag, adapted process is optional, the *jump measure* μ^X is an optional random measure by Lemma 9.8. It can be shown to be σ-finite and hence possesses a compensator ν^X.

10.1 EXERCISE. Show that the jump measure μ^X of a cadlag, adapted process X is σ-finite. [Hint: for given n let $S_{n,1}, S_{n,2}, \ldots$ the times of the first, second, etc. jumps of X greater than $1/n$. Then $\mu^X([0, S_{n,k}]) \le k$.]

The compensator ν^X of the jump process of a semimartingale X is called the *third characteristic* of X. The *second characteristic* is the quadratic variation process $[X^c] = \langle X^c \rangle$ of the continuous martingale part X^c of X. The purpose of this section is to define also the first characteristic and establish a "canonical representation" of a semimartingale.

This canonical representation is dependent on a "truncation function" h that will be fixed throughout. Let $h: \mathbb{R} \to \mathbb{R}$ be a function with compact support which agrees with the identity in a neighbourhood of zero, and set $\bar{h}(y) = y - h(y)$. The classical choice of truncation function is

$$h(y) = y1_{|y| \leq 1}, \qquad \bar{h}(y) = y1_{|y|>1}.$$

Every cadlag function possesses at most finitely many jumps of absolute size bigger than a given positive number on every given finite interval. Therefore, the process

$$(\bar{h} * \mu^X)_t = \sum_{s \leq t} \bar{h}(\Delta X_s)$$

is well defined. For the classical truncation function it is the cumulative sum of the "big" jumps of the process X. (Here and in the following h and $\bar{h}$ are considered the functions on the jump space $\mathbb{R}$, and an expression such as $h * \mu_t$ is understood to mean $\int \int 1_{[0,t]}(s) h(y)\, \mu(ds, dy)$.)

The cumulative sum of the remaining, "small" jumps are given, in principle, by the process $h * \mu^X$. However, for a general semimartingale the series $\sum_{s \leq t} |\Delta X_s|$ need not converge, and hence we cannot speak about a process $h * \mu^X$ in general. This may be accomodated by first compensating μ^X and using the abstract definition of an integral relative to a compensated point process, Definition 9.15, to define the "compensated cumulative small jump process" $h * (\mu^X - \nu^X)$. The following theorem implicitly asserts that the integral $h * (\mu^X - \nu^X)$ is always well defined in the sense of Definition 9.15.

10.2 Theorem. *For every semimartingale X there exists a (unique) predictable process B such that $X = X_0 + X^c + h * (\mu^X - \nu^X) + \bar{h} * \mu^X + B$.*

Proof. The uniqueness of B is clear from the fact that the other terms of the decomposition have a clear meaning.

The process $Y = X - X_0 - \bar{h} * \mu^X$ is a semimartingale with uniformly bounded jumps. As seen in the proof of Lemma 5.49 the martingale M in a decomposition $Y = M + A$ of Y into a martingale and bounded variation part possesses uniformly bounded jumps as well and hence so does the process A. If $V_t = \int_{[0,t]} |dA_s|$ is the variation of A and $S_n = \inf\{t > 0: V_{t-} > n\}$, then $V_{S_n} \leq n + |\Delta A_s|$, which is bounded by n plus a constant. It follows that the process A is of locally bounded variation and hence possesses locally integrable variation. By the Doob-Meyer decomposition there exists a predictable process B such that $A - B$ is a local martingale.

This gives the decomposition $X = X_0 + M + (A - B) + \bar{h} * \mu^X + B$. The local martingale $A - B$ is of locally bounded variation and hence is purely discontinuous. It follows that the continuous martingale parts of M and X coincide. We can conclude the proof by showing that $M^d + (A - B) = h * (\mu^X - \nu^X)$.

By definition $h * (\mu^X - \nu^X)$ is the unique purely discontinuous local martingale with jump process $\int_{\mathbb{D}} h(y)\,(\mu^X - \nu^X)(\{t\} \times dy) = h(\Delta X_t) - \int_{\mathbb{D}} h(y)\,\nu^X(\{t\} \times dy)$. The process $N = M^d + (A - B)$ is a purely discontinuous local martingale and hence it suffices to verify that it possesses the same jump process. The jump process of $N = X - X_0 - X^c - \bar{h} * \mu^X - B$ is $\Delta N = \Delta X - \bar{h}(\Delta X) - \Delta B = h(\Delta X) - \Delta B$. It suffices to show that $\Delta B = \int_{\mathbb{D}} h(y)\,\nu^X(\{t\} \times dy)$. Because both these processes are predictable, it suffices to show that they agree at all predictable times T. The predictability of ΔB also gives that ΔB_T is $\mathcal{F}_{T-}$ measurable (for every stopping time T). Combined with the fact that $\mathrm{E}(\Delta N_T|\, \mathcal{F}_{T-}) = 0$ for every predictable time T, because N is a local martingale, this gives that $\Delta B_T = \mathrm{E}\big(h(\Delta X_T)|\, \mathcal{F}_{T-}\big)$ for every predictable time T. The predictability of the map $Z\colon (t, \omega, y) \mapsto h(y) 1_{[T]}(t, \omega)$, for a predictable time T, yields that the process $K = Z * (\mu^X - \nu^X)$ is a local martingale. Its jump process is $\Delta K_t = \int h(y) 1_{[T]}(t)\,(m^X - \nu^X)(\{t\} \times dy)$, which is equal to $\Delta K_T = \int h(y)\,(m^X - \nu^X)(\{T\} \times dy) = h(\Delta X_T) - \int h(y)\,\nu^X(\{T\} \times dy)$ if evaluated at T. By the martingale property we have $\mathrm{E}(\Delta K_T|\, \mathcal{F}_{T-}) = 0$, so that finally we obtain that $\Delta B_T = \mathrm{E}\big(h(\Delta X_T)|\, \mathcal{F}_{T-}\big) = \int h(y)\,\nu^X(\{T\} \times dy)$. ∎

The predictable process B in the decomposition given by the preceding theorem is called the *first characteristic* of the semimartingale X. Thus we obtain a triple $(B, \langle X^c \rangle, \nu^X)$ of characteristics of a semimartingale, each of which is a predictable process.

In general these "predictable characteristics" do not uniquely determine the semimartingale X, or its distribution, and hence are not true characteristics, for instance in the sense that the characteristic function of a probability distribution determines this distribution. However, for several subclasses of semimartingales, including diffusions and counting processes, the characteristics do have this determining property. Furthermore, the characteristics play an important role in formulas for density processes and weak convergence theory for semimartinges. The following examples show that the characteristics are particularly simple for the basic examples of semimartingales.

10.3 Example (Brownian motion). Because a Brownian motion B is a continuous martingale, its "canonical decomposition" possesses only one term and takes the form $B = B^c$. The predictable quadratic variation is the identity. Thus the triple of characteristics of Brownian motion is $(0, id, 0)$. □

10.4 Example (Poisson process). The jump measure μ^N of a standard Poisson process N satisfies $\mu^X([0, t] \times D) = N_t 1_D(1)$. Because the compensated Poisson process $t \mapsto N_t - t$ is a martingale, the compensator of μ^N is the measure $\nu^X([0,t] \times D) = t\lambda(D)$. If we use the canonical truncation function h, then the process of "big" jumps (strictly bigger than 1) is zero, and the process of "small" jumps is $h * \mu^N = N$ It follows that the canonical decomposition is given by $N_t = (N_t - t) + t$. Thus the triple of characteristics of the Poisson process is $(id, 0, id)$. □

The second and third of the characteristics of a semimartingale are independent of the choice of truncation function, but the first characteristic is not. This first characteristic is also somewhat unsatisfactory as it is dependent on the unnatural different treatment of small and big jumps in the decomposition $X = X_0 + X^c + h * (\mu^X - \nu^X) + \bar{h} * \mu^X + B$. In this decomposition the small jumps are compensated, whereas the big jumps are not. A decomposition of the type $X = X_0 + X^c + id * (\mu^X - \nu^X) + B'$, with id the identity function on the jump space, would have been more natural, but this is not well defined in general, because the cumulative big jump process $\bar{h} * \mu^X$ may not have a compensator. The point here is that the Doob-Meyer decomposition guarantees the existence of a compensator only for processes that are locally of integrable variation. Even though the process $\bar{h} * \mu^X$ is well defined and of locally bounded variation, it may lack (local) integrability. If the process $\bar{h} * \mu^X$ is of locally integrable variation, then the more natural decomposition is possible. Semimartingales for which this is true are called "special".

More formally a semimartingale X is called *special* if it possesses a decomposition $X = X_0 + M + A$ into a local martingale M and a process A of locally integrable variation.

10.5 Theorem. *A semimartingale X is special if and only if there exists a (unique) predictable process B' such that $X = X_0 + X^c + id * (\mu^X - \nu^X) + B'$.*

Proof. It was seen in the proof of Theorem 10.2 that any semimartingale can be decomposed as $X = X_0 + M' + A' + \bar{h} * \mu^X$ for a local martingale M' and a process A' that is locally of integrable variation. If X is special and $X = X_0 + M + A$ is a decomposition in a local martingale M and process A of locally integrable variation, then $M - M' = \bar{h} * X + A' - A$ is a local martingale of locally bounded variation, and hence is locally of integrable variation. We conclude that the process $\bar{h} * \mu^X$ is of locally integrable variation and hence possesses compensator $\bar{h} * \nu^X$. We can now reorganize the decomposition given by Theorem 10.2 as $X = X_0 + X^c + h * (\mu^X - \nu^X) + \bar{h}(\mu^X - \nu^X) + \bar{h} * \nu^X + B$. Because $h + \bar{h} = id$ and the compensated jump integral given by Definition 9.15 is linear, this gives a decomposition as desired, with $B' = \bar{h} * \nu^X + B$.

A decomposition as claimed can be written as $X = X_0 + M + B'$ for a local martingale M and predictable process B'. Because a predictable process is automatically locally of integrable variation, it follows that X is special. ■