CWI Quarterly Volume 9 (4) 1996, pp. 357 – 388

A Crash Course in Stochastic Calculus with Applications to Mathematical Finance

Peter Spreij
Department of Econometrics
Vrije Universiteit
De Boelelaan 1105
1081 HV Amsterdam
pspreij@econ.vu.nl

In this paper the essentials of stochastic integration with respect to continuous martingales are reviewed. The presented results are illustrated by applying them to the theory of option pricing.

1. Introduction

The purpose of this paper is to summarize a number of results from stochastic calculus that are of fundamental importance in advanced mathematical finance. The emphasis will be on the concepts from probability theory, stochastic processes and stochastic integration theory. Throughout the paper we will hint at applications of the presented results to problems in mathematical finance, that are treated in Section 12.

A variety of good textbooks on the subject is available, each with its own flavour. We mention Rogers and Williams [10], Revuz and Yor [9], Chung and Williams [2], Øksendal [7], Protter [8]. For the preparation of this paper I heavily leaned on the book by Karatzas and Shreve [5], which was my main source of inspiration. Readers are supposed to be familiar with some measure theory and the measure theoretic foundation of probability. In these notes proofs are usually not given, although sometimes key steps of a proof are indicated to give the reader an idea of the flavour of the methods that are used.

The probably most famous model that is used in the theory of option pricing is that of Black and Scholes [1]. One of its ingredients is an equation that describes the evolution of the price P_t of a risky asset over time. We give the equation first, and after that we discuss what is meant by it and we explain also unknown quantities. Let b be a (measurable) function and σ a constant. The equation reads

$$dP_t = b(t)P_t dt + \sigma P_t dW_t,\ P_0. \tag{1.1}$$

Sometimes this equation is also written in its integral form

$$P_t = P_0 + \int_0^t b(s)P_s ds + \sigma \int_0^t P_s dW_s. \tag{1.2}$$

The first equation (1.1) is called a stochastic differential equation, whereas the second one (1.2) can be called a stochastic integral equation. It is common practice to work with the first one, having in mind the interpretation that it is an abbreviation of the second equation.

Looking at equation (1.1) one is faced with a number of questions. First, what is meant by W, dW_t and $\int_0^t P_s dW_s$?

W is a stochastic process called (standard) Brownian motion, by dW_t we mean an infinitesimally small increment of this process over a time interval $(t, t+dt)$ and finally $\int_0^t P_s dW_s$ is called the stochastic integral of P with respect to W. Furthermore P obviously looks as a solution of (1.1). But what is the solution concept of a stochastic differential equation?

Below we will define in this order what Brownian motion is, what a stochastic integral is and what (a solution of) a stochastic differential equation is. The theoretical framework in which this will take place is that of the general theory of stochastic processes, and (sub)martingales in continuous time. Occasionally we will use examples from discrete time, when this leads to better understanding. Some fundamental issues of this theory will be discussed in the next section.

2. General setting

As usual in probability theory we assume that we work with a probability space $(\Omega, \mathcal{F}, P)$. Since we will deal with (continuous time) stochastic processes we will also assume to have a filtration $\mathbb{F}$, that is a family of sub σ-algebras $\{\mathcal{F}_t : t \geq 0\}$ of $\mathcal{F}$ with the property that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s < t$. The interpretation is that as time proceeds, we have a growing information pattern to our disposal. The quadruple $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is called a *filtered probability space*.

We also frequently need the assumption that $\mathbb{F}$ satisfies the *usual conditions*, a technical term meaning two things: $\mathbb{F}$ is *right continuous*, i.e. $\bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} = \mathcal{F}_t$ for all $t \geq 0$ and that $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$.

We will also use the notation $\mathcal{F}_{t-}$ for the smallest σ-algebra that contains all $\mathcal{F}_s$ for $s < t$, and $\mathcal{F}_\infty$ for the smallest σ-algebra that contains all $\mathcal{F}_s$ for all s. A filtration that is right continuous and satisfies $\mathcal{F}_{t-} = \mathcal{F}_t$ for all t is called *continuous*.

By $\mathcal{B}$ or $\mathcal{B}(\mathbb{R})$ we denote the Borel σ-algebra on $\mathbb{R}$. In general we denote by $\mathcal{B}(E)$ the Borel sets of a topological space E. A real valued stochastic process X on $[0, \infty)$ is a collection $\{X_t : t \geq 0\}$ of random variables. So all the maps

$X_t : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ are measurable. In particular X is a function of two variables, ω and t. Its values are denoted by $X(\omega, t)$ or by $X_t(\omega)$.

This definition suffices in discrete time for most purposes. However, if we view X as a map from $\Omega \times [0, \infty)$ into $\mathbb{R}$ then, in general, that this map is not measurable as a function of the pair of variables (ω, t), where measurability refers to the product σ-algebra $\mathcal{F} \otimes \mathcal{B}([0, \infty))$ on $\Omega \times [0, \infty)$. Here the heart of the problem is of course that we deal with an uncountable set of random variables and, as is known, uncountably many operations easily destroy certain measurability properties.

Here are some other relevant measurability concepts. Consider a stochastic process X. Then

- X is called *measurable* if $X : (\Omega \times [0, \infty), \mathcal{F} \otimes \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ is measurable.
- X is called *adapted* to $\mathbb{F}$ if for all t the random variable X_t is $\mathcal{F}_t$-measurable.
- X is called *progressively measurable* (or progressive) with respect to $\mathbb{F}$ if for all t the map $(\Omega \times [0, t], \mathcal{F} \otimes \mathcal{B}([0, t]) \to (\mathbb{R}, \mathcal{B})$ is measurable.

If there is no ambiguity about the filtration at hand, we will simply speak of an adapted or a progressive process.

Adaptedness can be interpreted as follows. Suppose that the information comes to us in the form of observations of a stochastic process Z, meaning that the $\mathcal{F}_t$ are generated by Z, so $\mathcal{F}_t = \sigma\{Z_s, s \leq t\}$, for all $t \geq 0$. In such a case we often write $\mathcal{F}_t^Z$ instead of $\mathcal{F}_t$ and $\mathbb{F}^Z$ instead of $\mathbb{F}$. If X is adapted to $\mathbb{F}^Z$, then X_t is a functional of the collection $\{Z_s, s \leq t\}$. Progressiveness is a stronger property, a technical concept that one needs to ensure that certain dynamic transformations of adapted processes remain adapted. It is assured in a number of cases, see below.

The definition of X as a stochastic process was in terms of the random variables (measurable functions) X_t. Alternatively we can freeze a variable ω and look at the function $X_{\cdot}(\omega) : [0, \infty) \to \mathbb{R}$. The functions $X_{\cdot}(\omega)$ are called the sample paths of X. Some properties of a process X refer to its sample paths. For instance, X is said to be (left, right) continuous if all the sample paths $X_{\cdot}(\omega)$ are (left, right) continuous.

PROPOSITION 2.1. *Let X be a stochastic process. Then*
(i) X is measurable and adapted if it is progressive.
(ii) X is progressive, if it is adapted and left (or right) continuous.

EXAMPLE. Assume that for a progressive process X the expectations $\mathrm{E}\int_0^t |X_s| ds$ are finite for each t. We can then consider the process $\{\int_0^t X_s ds, t \geq 0\}$. Progressiveness of X ensures by application of Fubini's theorem that this process is adapted again, and even progressive in view of Proposition 2.1 by continuity of its sample paths.

In the theory of stochastic process there are two important concepts of describing in what sense two stochastic processes are the same. Two stochastic processes X and Y are called

– modifications of each other if $\forall t \geq 0 : P(X_t = Y_t) = 1$.
– indistinguishable if $P(\forall t \geq 0 : X_t = Y_t) = 1$.

We now give an example that illustrates why one sometimes needs that the filtration satisfies the usual conditions. Suppose that this is the case and that X is an adapted process. Let Y be a modification of X, then Y is again adapted, since sets $\{Y_t \in B\}$ for $B \in \mathcal{B}$ differ from $\{X_t \in B\}$ by a P-null set.

The following proposition relates the two properties of 'sameness'.

PROPOSITION 2.2. *(i) If X and Y are indistinguishable, then they are modifications of each other.*
(ii) If X and Y are modifications of each other and they are left (or right) continuous, then they are indistinguishable.

EXAMPLE. A number of difficulties with continuous time processes can be illustrated by analyzing properties of the process C defined for $\Omega = [0, \infty)$ by $C(\omega, t) = 1$ if $\omega = t$ and zero else. In this case we also use the Borel σ-field on Ω as $\mathcal{F}$ and for P we take a measure that is absolutely continuous with respect to Lebesgue measure.
Clearly, the zero process is a modification of C, whereas $P(\forall t : C_t = 0) = P(\emptyset) = 0$.

3. Brownian motion

Stochastic process are often specified by attributing properties to them in distributional terms. This happens already in the following definition. The general setting is that of the previous section.

DEFINITION 3.1. *A stochastic process W on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is called a standard Brownian motion with respect to $\mathbb{F}$, if the following properties hold.*

(i) $W_0 = 0$, *a.s.*
(ii) $W_t - W_s$ *is independent of* $\mathcal{F}_s$ *for all* $t \geq s$.
(iii) $W_t - W_s$ *has a normal distribution with zero mean and variance equal to* $t - s$ *for all* $t \geq s$.
(iv) W *is adapted to* $\mathbb{F}$ *and has continuous sample paths.*

The interpretation of W is better understood if we look at its discrete time analogue w. Let $\xi_1, \xi_2, \ldots$ be a sequence of iid standard normal random variables. Then we define for integers t the random variable $w_t = \sum_{k \leq t} \xi_k$. Taking a suitable probability space and $\mathcal{F}_t = \sigma\{\xi_k : k \leq t\}$, one easily verifies that the properties $(i) - (iii)$ hold as well as adaptedness to $\mathbb{F}$. The process w is called a discrete time random walk and W is its continuous time counterpart with continuous sample paths.

The immediate question is of course: does Brownian motion exist? The answer, not unexpectedly, is yes. There are a number of ways to show existence of Brownian motion. One is based on the Daniel-Kolmogorov theorem. We refer for instance to [5] for the precise formulation of this theorem and the terminology underlying it.

So we introduce the space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)}))$, where $\mathcal{B}(\mathbb{R}^{[0,\infty)})$ is the smallest σ-algebra that makes the finite dimensional projections measurable. Let for each natural n $\mathcal{Q}^n$ be a family of distributions on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, indexed by $t \in \mathbb{R}^n$. Denote by $\mathcal{Q}$ the union of all the $\mathcal{Q}^n$. The Daniel-Kolmogorov theorem states that if this family is *consistent*, then on the space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)}))$ there exists a probability measure P such that the stochastic process X defined by $X_t(\omega) = \omega(t)$ (the coordinate process) has exactly the family $\mathcal{Q}$ as its family of finite dimensional distributions.

Take $\mathbb{F}$ defined by $\mathcal{F}_t = \sigma\{X_s, s \leq t\}$. Then the finite dimensional distributions implicitly defined by properties (i)-(iii) of Definition 3.1 are consistent and we obtain the existence of a process X that satisfies these properties and also satisfies the adaptedness requirement of (iv). However, continuity of the sample paths of X is not established. Here a technical problem appears, because it can be shown that $C[0,\infty)$ does not belong to $\mathcal{B}(\mathbb{R}^{[0,\infty)})$. Fortunately, there is a way out. It follows from the Kolmogorov-Čentsov theorem that X has a modification with continuous paths. This gives us the existence of Brownian motion, since modifications don't change the finite dimensional distributions.

Another construction of Brownian motion is via weak convergence. The approach is as follows. Let $\xi_1, \xi_2, \ldots$ be a sequence of *iid* (not necessarily normal) random variables on some probability space with $E\xi_i = 0$ and $E\xi_i^2 = 1$. Let $S_k = \sum_{j \leq k} \xi_j$ and for each n

$$X_t^n = n^{-\frac{1}{2}}\{S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)\xi_{\lfloor nt \rfloor + 1}\}.$$

Then the X^n are continuous processes with $X^n_{m/n} = n^{-\frac{1}{2}} S_m$ for integers m. Hence each X^n induces a measure, P^n say, on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$. The following result (Donsker's invariance principle) holds.

Proposition 3.2. *There exists a measure P^W on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$ such that $\{P^n\}$ weakly converges to P^W and under which the coordinate process W defined by $W_t(\omega) = \omega(t)$ is a standard Brownian motion. P^W is called the Wiener measure.*

Brownian motion has a number of attractive probabilistic properties. For instance it is an example of a strong Markov process. On the other hand its paths exhibit a very irregular behaviour. We mention a few examples. First we need some notation. If I is a compact interval, $[a,b]$ say, we denote by Π_n or $\Pi_n(I)$ a partition $\{t_0, \ldots, t_n\}$ of I with $t_0 = a$ and $t_n = b$. Its *mesh* $m(\Pi_n)$ is then $\max\{|t_i - t_{i-1}| : i = 1, \ldots, n\}$.

- Almost all paths of Brownian motion are nowhere differentiable.
- The paths of Brownian motion are not of bounded variation over (non-empty) bounded intervals, so $\sum_{\Pi_n} |W_{t_i} - W_{t_{i-1}}|$ tends to infinity a.s. if $m(\Pi_n)$ tends to zero. But

– The paths of Brownian motion are of bounded quadratic variation over bounded intervals $[a,b]$. More precisely, $V_n := \sum_{\Pi_n}(W_{t_i}-W_{t_{i-1}})^2 \stackrel{P}{\to} b-a$ if $m(\Pi_n)\to 0$. In fact, one can show that $E(V_n-(b-a))^2 = \sum_{\Pi_n}(t_i-t_{i-1})^2 \leq m(\Pi_n)(b-a)$. If $\sum m(\Pi_n) < \infty$ we also have a.s. convergence of V_n to $b-a$.
– Almost all zero sets $Z(\omega) = \{t : W_t(\omega) = 0\}$ have Lebesgue measure zero, are closed and unbounded and have no isolated points in $(0,\infty)$. In particular $\int_0^\infty 1_{\{Z(\omega)\}}(t)dt = 0$ a.s.

The first two items of this list can heuristically justified by using the observation that for all t,h we can write $W_{t+h} - W_t = \sqrt{h}N(0,1)$, where we denote by $N(0,1)$ any random variable that is a standard normal one. Then we see that, loosely speaking, the difference quotient $\frac{W_{t+h}-W_t}{h}$ is of order $h^{-\frac{1}{2}}$, so we cannot expect to have a finite limit.

Now look at sums $\sum_{i=1}^n |W_{\frac{i}{n}} - W_{\frac{i-1}{n}}|$. A similar way of reasoning suggests that this is of order $n^{\frac{1}{2}}$, which makes it understandable that the paths are not of bounded variation. We will return to the quadratic variation in the next section.

In Definition 3.1 we imposed that $W_0 = 0$ a.s. Instead of standard Brownian motion, we will occasionally consider Brownian motion with an initial condition different from zero. This is done as follows. Let W be a standard Brownian motion on some $(\Omega,\mathcal{F},P)$ with respect to its own filtration. Let B_0 be a random variable with distribution μ, independent of W (extend Ω if necessary) and define $B_t = B_0 + W_t$. Then B satisfies the conditions of Definition 3.1 with $\mathbb{F} = \mathbb{F}^B$ with the exception of the zero initial condition. We say that μ is the initial distribution of B.

The measure that B induces on $(C[0,\infty),\mathcal{B}(C[0,\infty)))$ is usually denoted by P^μ. If μ is a Dirac measure concentrated at $x\in\mathbb{R}$ (in this case B is called Brownian motion starting at x), then we use the notation P^x instead of P^μ.

It is a fact that if we augment the filtration $\mathbb{F}^B$ with the null sets $\mathcal{N}^\mu$ of P^μ, so we consider the σ-fields $\mathcal{F}_t^\mu = \mathcal{F}_t^B \vee \mathcal{N}^\mu$, then we get a right continuous filtration and B is still a Brownian motion with respect to this family of σ-algebras. So the filtration $\{\mathcal{F}_t^\mu, t\geq 0\}$ satisfies the usual conditions.

4. Some martingale theory

Definition 4.1. *A stochastic process X is called a martingale with respect to $\mathbb{F}$ if*

(i) X is adapted to $\mathbb{F}$,
(ii) $E|X_t| < \infty$, for all $t\geq 0$,
(iii) $E[X_t|\mathcal{F}_s] = X_s$, a.s. for all $t\geq s$. X is called a submartingale if we replace the equality sign in (iii) by $\geq$.

The interpretation of property (iii) of a martingale X is that the best guess of the future value X_t given the information up to the present time s is the current value X_s.

EXAMPLE. Let W be a standard Brownian motion with respect to $\mathbb{F}$, then W is a martingale. Indeed, since $W_t - W_s$ is independent of $\mathcal{F}_s$ for $t > s$, we have $E[W_t - W_s|\mathcal{F}_s] = 0$.

EXAMPLE. Let M be a martingale with respect to $\mathbb{F}$, such that $EM_t^2 < \infty$ for all t, then $X_t = M_t^2$ defines a submartingale. In particular $X = W^2$ is a submartingale. Notice that since the increments $W_t - W_s$ have variance equal to $t - s$, we see by an easy calculation that $\{W_t^2 - t : t \geq 0\}$ is a martingale again.

As a preparation for the Doob-Meyer theorem below, we consider a submartingale X in discrete time. Define $A_0 = 0$ and recursively

$$A_k = A_{k-1} + E[X_k|\mathcal{F}_{k-1}] - X_{k-1}.$$

From the submartingale inequality we obtain that A is an increasing process ($A_k \geq A_{k-1}$ a.s) and we also see that A_k is $\mathcal{F}_{k-1}$-measurable (one says that A is a *predictable* discrete time process). If we define $M_k = X_k - A_k$, then one easily sees that M is a martingale. Hence we have the *Doob decomposition* $X = M+A$, with an increasing predictable process A and a martingale M. Such a decomposition is unique. In continuous time, the proof of the corresponding result is rather deep. We give this result for nonnegative submartingales only.

THEOREM 4.2. *Let X be a nonnegative right continuous submartingale, adapted to a filtration $\mathbb{F}$ satisfying the usual conditions. Then X can be decomposed as $X = A+M$, with a martingale M and a right continuous increasing process A, such that $A_0 = 0$ and every A_t is $\mathcal{F}_{t-}$-measurable. The process A can be chosen to satisfy $E\int_{(0,t]} m_s dA_s = E\int_{(0,t]} m_{s-} dA_s$ for every bounded right continuous martingale m. Moreover, given that A satisfies this last property it is unique (up to indistinguishability).*

We apply this theorem to the special case where X is the square of a martingale N, satisfying $EN_t^2 < \infty$ for every t. Martingales N with this property are called square integrable and the class of these martingales is denoted by $\mathcal{M}_2$. The class of square integrable martingales with continuous paths is denoted by $\mathcal{M}_2^c$.

So, let $N \in \mathcal{M}_2$ and $X = N^2$. For this case we use the special notation $\langle N \rangle$ for the process A that appears in the Doob-Meyer decomposition of X. The process $\langle N \rangle$ is called the quadratic variation process, or the predictable quadratic variation process of N.

This process can be viewed as a cumulative conditional variance of N. To see this, use that $N_s = E[N_t|\mathcal{F}_s]$ and consider

$$E[(N_t - N_s)^2|\mathcal{F}_s] = E[N_t^2 - N_s^2|\mathcal{F}_s] = E[\langle N \rangle_t - \langle N \rangle_s|\mathcal{F}_s]. \tag{4.1}$$

The last equality follows from the definition of $\langle N \rangle$.

If we apply this to standard Brownian motion, we get, using previous results, $\langle W \rangle_t \equiv t$.

The name quadratic variation is explained by the following fact. If Π_n are partitions of $[0, t]$ as before whose meshes tend to zero for $n \to \infty$ and N is a *continuous* martingale with $EN_t^2 < \infty$ (so $N \in \mathcal{M}_2^c$), then

$$\sum_{\Pi_n}(N_{t_k}-N_{t_{k-1}})^2 \xrightarrow{P} \langle N\rangle_t. \tag{4.2}$$

We have already seen this result for N a Brownian motion in Section 3 and one could take it as the definition of the quadratic variation process. However, since this characterization is only true for martingales in $\mathcal{M}_2^c$, we prefer to define the quadratic variation process via the Doob-Meyer decomposition, which applies to all martingales in $\mathcal{M}_2$. A natural question is then to ask for what happens with the sum in (4.2) if we drop the continuity assumption on N. One can show that again the limit in probability exists (see [3] for instance). In the theory of martingales it is usually denoted by $[N]_t$ or $[N,N]_t$ and in general it differs from $\langle N\rangle_t$. We illustrate this by an example.

Consider a standard Poisson process, that is a process X defined on some filtered probability space which satisfies conditions (i) and (ii) of Definition 3.1 and with (iii) replaced with the condition that the increments $X_t - X_s$ have a Poisson distribution with parameter $t-s$. In particular we have $E(X_t - X_s) = t-s$ and $\mathrm{var}(X_t - X_s) = t-s$. Define then the process N by $N_t = X_t - t$ for all t. From the independence of the increments of X one sees that N is a martingale. It also belongs to $\mathcal{M}_2$, but obviously doesn't have continuous paths. As a matter of fact, its paths are piecewise linear and show upward jumps of size $+1$. We take all the paths to be right continuous, the standard convention.

Now we turn to equation (4.1). Put $\langle N\rangle_t \equiv t$. Then indeed (4.1) is satisfied, as it follows from the independence of the increments of X and hence of N and the expression of the variance of Poisson random variables that $E[(N_t - N_s)^2|\mathcal{F}_s] = t-s$. On the other hand the limit of the sums in (4.2) is X_t, even a.s., which can be seen as follows. Fix a typical sample path of X and the corresponding one of N. Since the mesh of the considered partition tends to zero, we assume that it is small enough to have that each time interval $J_i = (t_{i-1}, t_i]$ contains at most one of the time instants where the path of X has a jump, there are exactly X_t of these jump times in $[0,t]$. Split the sum in (4.2) in two parts, one such that all the J_i contain a jump time and one such that none of them contains one. The former one then equals $\sum(1-(t_i-t_{i-1}))^2$ with limit X_t, the other one being equal to $\sum(t_i - t_{i-1})^2$ has limit zero.

From the definition of the quadratic variation process of a martingale, we can derive the quadratic covariation process of two martingales. So let M and N be two martingales in $\mathcal{M}_2$. Then the quadratic covariation process of M and N is defined via the polarization formula

$$\langle M,N\rangle = \frac{1}{2}\{\langle M+N\rangle - \langle M\rangle - \langle N\rangle\}.$$

Notice that $\langle M,N\rangle$ is a process of bounded variation over bounded intervals, that $\langle M,M\rangle = \langle M\rangle$ and that

$$MN - \langle M,N\rangle \text{ is a martingale} \tag{4.3}$$

again.

Example. Here is an example of the use of $\langle M\rangle$. Let $T>0$ and assume that $\langle M\rangle_T=0$ a.s., then $M_t=0$ a.s. for all $t\leq T$. This follows easily from Chebychev's inequality and $EM_t^2=E\langle M\rangle_t\leq E\langle M\rangle_T$. We will need this result in Section 10.

We close this section with some examples for discrete time martingales, that serve to illustrate the results of this section as well as to prepare for the next one. Let M be a martingale in discrete time, $M_0=0$, and let $\Delta M_t=M_t-M_{t-1}$. An example is obtained by taking the ΔM_t to be independent variables with zero mean. If moreover all ΔM_t have finite variance σ_t^2, then we get the Doob decomposition of the submartingale M^2 as $M_t^2=m_t+\sum_{i=1}^t\sigma_i^2$ with m another martingale, as is easily verified. So $\langle M\rangle_t=\sum_{i=1}^t\sigma_i^2$. This again illustrates that the quadratic variation process is a generalization of the cumulative variance process.

Let now $\xi_0,\xi_1,\ldots$ be another process, that for simplicity is taken to be bounded. Assume that this process is predictable, so ξ_t is $\mathcal{F}_{t-1}$-measurable. Consider the process I defined by $I_t=\sum_{k=1}^t\xi_k\Delta M_k$. Predictability and boundedness of ξ yields I a martingale. I is also called a martingale transform of M or a *discrete time stochastic integral* of ξ with respect to M. Furthermore

$$\begin{aligned}\langle I\rangle_t-\langle I\rangle_{t-1} &= E[(I_t-I_{t-1})^2|\mathcal{F}_{t-1}]=\xi_t^2E[(M_t-M_{t-1})^2|\mathcal{F}_{t-1}]\\ &= \xi_t^2(\langle M\rangle_t-\langle M\rangle_{t-1}).\end{aligned}$$

So we see how to express the quadratic variation of a discrete time stochastic integral in terms of that of M. This properties will be encountered in a continuous time setting in the next section.

5. Stochastic integrals

In this section we will define *stochastic integrals* $\int_0^T X_s dM_s$ for a suitable class of stochastic processes X and a *continuous martingale* M. It should be clear that we are faced with a problem of definition (see also Proposition 5.1 below). Take for example $M=W$, standard Brownian motion. We know that the paths of W have unbounded variation over bounded intervals, a property that is shared by all nonconstant continuous martingales, which excludes a naive approach by trying to mimic the definition of Stieltjes integral. We recall a fundamental result from Stieltjes integration theory.

Proposition 5.1. *(i) Let $\alpha:[0,1]\to\mathbb{R}$ be a function of bounded variation. Then for all $h\in C[0,1]$ the Stieltjes integral $\int_0^1 hd\alpha$ exists and for partitions Π_n of $[0,1]$ whose meshes tend to zero for $n\to\infty$ we have*

$$\sum_{\Pi_n}h(t_k)(\alpha_{t_k}-\alpha_{t_{k-1}})\to\int_0^1 hd\alpha.$$

(ii) Conversely, if $\alpha : [0,1] \to \mathbb{R}$ and if for all $h \in C[0,1]$ the above limit exists, then α is of bounded variation over $[0,1]$.

Especially assertion (ii) above suggests that is impossible to define stochastic integrals with respect to a continuous martingale for a reasonably wide class of stochastic processes or even functions.

The proof of this assertion of the theorem (see [8], page 40) involves an application of the Banach-Steinhaus theorem and for a given partition an argument that involves a continuous function h that satisfies

$$h(t_k) = \operatorname{sgn}(\alpha(t_{k+1}) - \alpha(t_k)). \tag{5.1}$$

In the construction below we will exclude functions h of this type, since at a time instant t_k such an h uses the *future* values of the integrator. This exclusion together with a clever use of the fact that the paths of square integrable martingales are of bounded quadratic variation saves us.

Whatever definition of integral one would choose, there should be no confusion about how to integrate step functions. So, whatever we take for M if $X_t(\omega) = \xi(\omega)1_{(a,b]}(t)$, then the only logical definition of $\int_0^t X_s(\omega)dM_s(\omega)$ is $\xi(\omega)(M_{t\wedge b}(\omega) - M_{t\wedge a}(\omega))$. Below we drop in the notation the dependence on ω.

Next we assume that ξ is $\mathcal{F}_a$-measurable and bounded. View the integral as a stochastic process indexed by t. It is then a straightforward computation to see that this process is a martingale if M is one, and continuous if M is. Moreover we can easily compute (compare with the discrete time situation of the previous section) its quadratic variation at time t as $\xi^2(\langle M\rangle_{t\wedge b} - \langle M\rangle_{t\wedge a}) = \int_0^t X_s^2 d\langle M\rangle_s$.
By imposing linearity we now also have a definition of $\int_0^t X_s dM_s$ for step functions X with finitely many values.

We now give the construction of the stochastic integral over a finite interval, $[0,T]$ say. We first define the class of simple processes, denoted by $\mathcal{L}_{0,T}$. It is the class of all processes X that can be written as a finite sum

$$X_t = \xi_0 1_{\{0\}}(t) + \sum_{i=0}^{n} \xi_i 1_{(t_i,t_{i+1}]}(t), \tag{5.2}$$

where the t_i are increasing numbers in $[0,T]$ and the ξ_i are random variables, such that each ξ_i is $\mathcal{F}_{t_i}$-measurable and $\sup_{i,\omega}|\xi_i(\omega)| < \infty$. Notice that $X_{t_{i+1}}$ is $\mathcal{F}_{t_i}$-measurable and so X doesn't peek into the future, in contrast with the function h in equation (5.1) above.

In the rest of this section we will work on a probability space with a filtration that satisfies the usual conditions and with a *continuous* martingale M, that is square integrable over $[0,T]$, so $EM_T^2 < \infty$. As a matter of fact, one thus defines a norm, $||\cdot||$, on the space of (continuous) martingales on the interval $[0,T]$: $||M|| = (EM_T^2)^{\frac{1}{2}}$. Under this norm both $\mathcal{M}_2$ and $\mathcal{M}_2^c$ are Hilbert spaces, if we restrict ourselves to processes defined on the interval $[0,T]$ only, instead of $[0,\infty)$.

Let now X be a measurable process, then we define (what is going to be interpreted as a norm)

$$\mu_M(X) = [E\int_0^T X_s^2 d\langle M\rangle_s]^{\frac{1}{2}}.$$

Notice that this defines a measure on $(\Omega \times [0,T], \mathcal{F} \times \mathcal{B}([0,T]))$ by $\mu_M(A) := \mu_M(1_A)$.

The next object to define is the pre-Hilbert space of stochastic processes

$$\mathcal{L}_T^* = \{X : X \text{ progressive}, \mu_M(X) < \infty\}.$$

Clearly, this space depends on the choice of M. When necessary, we express this by the notation $\mathcal{L}_T^*(M)$.

From now on we will identify two processes X and Y if $\mu_M(X - Y) = 0$. Of course this defines an equivalence relation on $\mathcal{L}_T^*$, whose quotient space is again denoted by $\mathcal{L}_T^*$. Then we have the following properties.

- $\mathcal{L}_T^*$ is a Hilbert space with an inner product that generates the norm μ_M.
- The space of simple processes $\mathcal{L}_{0,T}$ is a dense subset of $\mathcal{L}_T^*$ under the norm μ_M.

Let X be an element of $\mathcal{L}_{0,T}$ of the form as in equation (5.2), then we define the stochastic integral $\int_0^t X_s dM_s$ of X with respect to M for $t \le T$, in this section also often denoted by $I_t(X)$, as

$$I_t(X) = \sum_{i=0}^{n} \xi_i (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}). \tag{5.3}$$

When there are more martingales around, we will express the dependence on M by writing $I_t(X; M)$. We have the following properties.

Proposition 5.2. *The stochastic integral defined for simple processes in equation (5.3) enjoys the following properties.*

(i) $I_\cdot(X)$ *is a continuous process with* $I_0(X) = 0$.
(ii) $I_\cdot(X)$ *is a martingale;* $E[I_t(X)|\mathcal{F}_s] = I_s(X)$ *for* $0 \le s \le t \le T$.
(iii) $E[(I_t(X) - I_s(X))^2|\mathcal{F}_s] = E[\int_s^t X_u^2 d\langle M\rangle_u|\mathcal{F}_s]$, *or* $\langle I_\cdot(X)\rangle_t = \int_0^t X_u^2 d\langle M\rangle_u$.
(iv) $I_t(\cdot)$ *is for each* t *a linear operator on* $\mathcal{L}_{0,T}$.
(v) $||I_\cdot(X)|| = \mu_M(X)$, *in particular* $I_T(\cdot)$ *is an isomorphism from* $\mathcal{L}_{0,T}$ *into* $\mathcal{M}_2^c$.

Having defined the stochastic integral with respect to M for $X \in \mathcal{L}_{0,T}$, we extend its definition to the closure of $\mathcal{L}_{0,T}$, which is $\mathcal{L}_T^*$, as follows.

Let $X \in \mathcal{L}_T^*$, then there exist a sequence $\{X^n\}$ in $\mathcal{L}_{0,T}$ such that $\mu_M(X^n - X) \to 0$. Using the isometry property we then have $||I_\cdot(X^m) - I_\cdot(X^n)|| = \mu_M(X^m - X^n) \to 0$, for $m, n \to \infty$. So $\{I_\cdot(X^n)\}$ is a Cauchy sequence in $\mathcal{M}_2^c$. Hence it has a limit, denoted by $I_\cdot(X)$. It is easy to show, by looking at mixed sequences, that this limit doesn't depend on the particular choice of the sequence $\{X^n\}$, which makes the definition of $I_\cdot(X)$ unambiguous. We will use the following notations $I_T(X) = \int_0^T X_s dM_s = \int_0^T X dM$.

Clearly $I_t(X) = \int_0^t X_s dM_s$ is defined similarly for $t \le T$ and $I_t(X) = I_T(1_{(0,t]}X)$.

All the properties of Proposition 5.2 remain valid except that $I_T(\cdot)$ now acts as an operator on $\mathcal{L}_T^*$ and it is an isometry from this set onto its image. Moreover one can show that for $X, Y \in \mathcal{L}_T^*$

$$\langle I_\cdot(X), I_\cdot(Y)\rangle_t = \int_0^t X_s Y_s d\langle M\rangle_s,\ t \le T. \tag{5.4}$$

The properties above, in particular equation (5.4) can be extended by looking also at other martingales in $\mathcal{M}_2^c$. For instance we have

$$\langle I_\cdot(X), N\rangle_t = \int_0^t X_s d\langle M, N\rangle_s,\ t \le T \tag{5.5}$$

for $X \in \mathcal{L}_T^*(M)$ and $N \in \mathcal{M}_2$. And

$$\langle I_\cdot(X; M), I_\cdot(Y; N)\rangle_t = \int_0^t X_s Y_s d\langle M, N\rangle_s,\ t \le T \tag{5.6}$$

for $M, N \in \mathcal{M}_2^c$, $X \in \mathcal{L}_T^*(M)$ and $Y \in \mathcal{L}_T^*(N)$.

In fact equation (5.5) provides us with a characterization of the $I_\cdot(X; M)$:

Proposition 5.3. *Let $\Phi \in \mathcal{M}_2^c$ be such that $\forall N \in \mathcal{M}_2$ the quadratic covariation process $\langle \Phi, N\rangle$ is equal to $\int_0^\cdot X_s d\langle M, N\rangle_s$ on $[0,T]$. Then Φ and $I_\cdot(X; M)$ are indistinguishable on $[0,T]$.*

Furthermore we have the important chain rule.

Proposition 5.4. *If $M \in \mathcal{M}_2^c$, $X \in \mathcal{L}_T^*(M)$ and $Y \in \mathcal{L}_T^*(I_\cdot(X; M))$, then $XY \in \mathcal{L}_T^*(M)$ and $I_\cdot(YX; M) = I_\cdot(Y; I_\cdot(X; M))$, which in differential notation reads $(YX)dM = Y(XdM)$.*

6. Extension to local martingales

The definition of stochastic integral as we have seen it in the previous section will be extended in the present one. It had been experienced that there was need for a generalization of the martingale property. The resulting class of *local martingales* proves to be closed under a wider set of operations than the class of martingales.

First we define *stopping times.* A stopping time T (for $\mathbb{F}$) is a random variable with values in $[0, \infty]$ such that the set $\{T \le t\}$ belongs to $\mathcal{F}_t$ for all $t \in [0, \infty)$. A *fundamental* or *localizing* sequence of stopping times is an increasing sequence $T_1, T_2, \ldots$ such that $\lim_{n\to\infty} T_n = \infty$.

By a localization one means that one can find a suitable localizing sequence of stopping times and that one studies a process on each of the random intervals $[0, T_n]$.

DEFINITION 6.1. *A stochastic process X is called a local martingale (with respect to $\mathbb{F}$) if there exists a localizing sequence $T_1, T_2, \ldots$ of stopping times such that every process $\{X_{T_n \wedge t}, t \geq 0\}$ is a martingale. A continuous local martingale is a local martingale with continuous paths.*

The class of local martingales is denoted by $\mathcal{M}_{loc}$ and the one of continuous local martingales by $\mathcal{M}_{loc}^c$.

Clearly, every martingale is a local martingale, because we can take $T_n = \infty$. Having extended the class of martingales, we now also extend the class of integrands $\mathcal{L}_T^*$ that we had in the previous section. But before doing so we need the following result. Let $M, N \in \mathcal{M}_{loc}^c$. Then there exists a unique (up to indistinguishability) process of bounded variation denoted by $\langle M, N \rangle$ such that $MN - \langle M, N \rangle \in \mathcal{M}_{loc}^c$ (compare with (4.3)).

If $M = N$ we simply write $\langle M \rangle$ and this process has nondecreasing paths. This process shows up in the definition of the class of integrands.

By definition the class $\mathcal{P}_T^*$, also denoted by $\mathcal{P}_T^*(M)$, for given $M \in \mathcal{M}_{loc}^c$, is the class of progressive processes X such that $\int_0^T X_t^2 d\langle M \rangle_t$ is finite a.s.

By localization and truncation it is possible to define for $t \leq T$ the stochastic integrals $I_t(X; M)$ for $M \in \mathcal{M}_{loc}^c$ and $X \in \mathcal{P}_T^*(M)$, as an a.s. limit of certain stochastic integrals $I_T(X^n; M^n)$, where all the M^n belong to $\mathcal{M}_2^c$ and $X^n \in \mathcal{L}_T^*(M^n)$ for all n. One can show that this stochastic integral, viewed as a process in t, is again a continuous local martingale on $[0, T]$. Another important property is that equations (5.5) and (5.6) are still valid for continuous local martingales M and N and X and Y such that the stochastic integrals exist. Also sample path properties of stochastic integrals with respect to an $M \in \mathcal{M}_2^c$ as a rule carry over to the stochastic integrals with respect to continuous local martingales. However properties involving (conditional) expectations are in general lost.

As in the previous section $I(X; M)$ can be characterized as the unique (up to indistinguishability) continuous local martingale Φ, such that $\langle \Phi, N \rangle = \int_0^{\cdot} X_s d\langle M, N \rangle_s$ (compare with proposition 5.3).

7. THE ITÔ RULE

The Itô rule, also called the change-of-variable rule, could be called the most important operational rule in stochastic calculus. To prepare for it we need an auxiliary concept, fundamental in the theory of stochastic processes.

DEFINITION 7.1. *A stochastic process X is called a continuous semimartingale if the following decomposition holds*

$$X_t = X_0 + M_t + B_t, \ \forall t \geq 0, \tag{7.1}$$

where M is a continuous local martingale with $M_0 = 0$ and B a continuous adapted process of bounded variation over bounded intervals with $B_0 = 0$.

We have already encountered an example of a semimartingale, namely when X is the square of a martingale N, in which case the process B in equation (7.1) becomes the quadratic variation process $\langle N \rangle$, or more general when X is a submartingale.

The decomposition of a continuous semimartingale is unique. Therefore we are ready to define the stochastic integral with respect to a semimartingale. Let X be a continuous semimartingale with decomposition (7.1) and let Y be a progressive process (in particular measurable) and assume that the pathwise defined Lebesgue-Stieltjes integral $\int_0^T |Y_s| dB_s$ is finite a.s. Assume also that $Y \in \mathcal{L}_T^*(M)$, then the stochastic integral of Y with respect to X is defined as

$$\int_0^t Y_s dX_s := \int_0^t Y_s dM_s + \int_0^t Y_s dB_s, \ t \leq T. \tag{7.2}$$

Here the first integral is of course a stochastic integral.

THEOREM 7.2. *Let X be a continuous semimartingale as in (7.1) and $f \in C^2(\mathbb{R}, \mathbb{R})$. Then $f(X)$ is again a continuous semimartingale and for all $t \geq 0$*

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle M \rangle_s \\ &= f(X_0) + \int_0^t f'(X_s) dM_s + \int_0^t f'(X_s) dB_s \\ &\quad + \frac{1}{2} \int_0^t f''(X_s) d\langle M \rangle_s. \end{aligned} \tag{7.3}$$

Notice that the semimartingale decomposition of $f(X)$ has $\int_0^\cdot f(X_s) dM_s$ as its local martingale part and that this stochastic integral is well defined since by continuity of f' and X almost all sample paths of $f'(X)$ are bounded over bounded intervals, hence $\int_0^t f'(X_s)^2 d\langle M \rangle_s$ is finite a.s. (cf. Section 6).

For a semimartingale X with decomposition (7.1) one usually writes $\langle X \rangle$ instead of $\langle M \rangle$. With this convention equation (7.3) is often written in the differential form

$$df(X_t) = f'(X_t) dX_t + \frac{1}{2} f''(X_t) d\langle X \rangle_t.$$

Application of ordinary calculus rules would only give the first integral in the right hand side of (7.3). The second one is sometimes referred to as Itô's correction term. The next example gives some insight in and explains the presence of the correction term.

EXAMPLE. Take $X = M = W$, with W standard Brownian motion. Application of the Itô rule with $f(x) = x^2$ yields $W_T^2 = 2 \int_0^T W_t dW_t + T$. We clearly see that we can't do without the correction term, since $EW_T^2 = T$, whereas the stochastic integral, being a martingale, has expectation zero. Now we consider a discrete time version of the integral. Let $\Pi_n = \{0 = t_0, \ldots, t_n = T\}$ be a partition of $[0, T]$ and define $W_t^n = \sum_{i=0}^{n-1} W_{t_i} 1_{(t_i, t_{i+1}]}(t)$. Then $\mu_W(W - W^n)^2 = \sum_{i=0}^{n-1}(t_{i+1} - t_i)^2$, which tends to zero for partitions whose meshes tend to zero. Elementary algebra shows that

$$2\int_0^T W_t^n dW_t = 2\sum_{i=0}^{n-1} W_{t_i}(W_{t_{i+1}} - W_{t_i}) = W_T^2 - \sum_{i=0}^{n-1}(W_{t_{i+1}} - W_{t_i})^2.$$

The term in the left hand side of this equation tends to $2\int_0^T W_t dW_t$, and the sum in the right hand side to T in probability (see Section 3), which gives the desired result.

EXAMPLE. Let $f(x) = e^x$ and $M \in \mathcal{M}_2^c$, with $M_0 = 0$. Define Z by $Z_t = \exp(M_t - \langle M \rangle_t)$. Application of Itô's rule yields that Z satisfies

$$Z_t = 1 + \int_0^t Z_s dM_s. \tag{7.4}$$

The process Z thus defined is known as the Doléans exponential of M, and is often denoted by $Z = \mathcal{E}(M)$. We will return to it, when we discuss Girsanov's theorem in Section 9.

Itô's rule is also a convenient tool to prove Lévy's characterization of Brownian motion, which is stated as

PROPOSITION 7.3. *Let M be a continuous local martingale relative to a filtration $\mathbb{F}$. If $\langle M \rangle_t \equiv t$, then M is standard Brownian motion relative to $\mathbb{F}$.*

The key to the proof, of which we present only the main steps, is to apply Itô's rule with the function $f(x) = \exp(iux)$. The result is

$$\exp(iuM_t) = \exp(iuM_s) + \int_s^t \exp(iuM_\tau) dM_\tau - \frac{1}{2}u^2 \int_s^t \exp(iuM_\tau) d\tau.$$

Multiplication of this equation by $\exp(-iuM_s)$ and taking conditional expectation with respect to $\mathcal{F}_s$ gives an integral equation for the conditional characteristic function $E[\exp(iu(M_t - M_s))|\mathcal{F}_s]$, which is then shown to be equal to $\exp(-\frac{1}{2}(t - s))$. This proves the proposition. □

The assertion of Theorem 7.2 can be extended to multivariate processes. We mention one important application, the product rule for semimartingales.

PROPOSITION 7.4. *Let X and Y be (real valued) continuous semimartingales. Then*

$$X_T Y_T = X_0 Y_0 + \int_0^T X_t dY_t + \int_0^T Y_t dX_t + \langle X, Y \rangle_T, \tag{7.5}$$

where $\langle X, Y\rangle$ stands for the quadratic covariation of the (local) martingale parts of X and Y.

This result formally follows from the Itô-rule for multivariate processes by considering the bivariate process (X, Y) and $f : (x, y) \to xy$.

8. Representation theorem

In many areas of applied probability, the martingale representation Theorem (theorem 8.3 below) for so called Brownian martingales is a key tool in obtaining certain results. For instance in mathematical finance it is used to show that in the Black Scholes framework the market is complete, i.e., in the language of finance, every contingent claim is attainable (see Section 12). Another application is in the derivation of the equation of the optimal filter for diffusion observations. First a weak result.

Proposition 8.1. *Let M be a continuous local martingale with respect to a filtration $\mathbb{F}$ and assume that $\langle M\rangle = \int_0^{\cdot} \alpha_t^2 dt$ for a strictly positive process α. Then there exists a Brownian motion W with respect to $\mathbb{F}$ such that $M = I_{\cdot}(\alpha; W)$.*

The proof of this result is rather simple. α can be shown to be a progressive process and α^{-1} belongs to $\mathcal{L}_T^*(M)$ for every $T > 0$. Hence the stochastic integral $W := I_{\cdot}(\alpha^{-1}; M)$ is well defined and $\langle W\rangle_t \equiv t$. Hence we obtain from Levy's characterization that W is a standard Brownian motion.

Actually, the requirement that α is strictly positive can be dropped at the cost of getting a Brownian motion on an enlarged probability space.

Below we present a stronger result after a general statement. Assume that the filtration $\mathbb{F}$ satisfies the usual conditions. Let $N \in \mathcal{M}_2$. We denote by $\mathcal{M}^*(N)$ the class of all stochastic integrals $I(X; N)$ for X belonging to $\mathcal{L}_T^*(N)$ for all $T \geq 0$.

Proposition 8.2. *Let $M \in \mathcal{M}_2$. Then there exists a unique (up to indistinguishability) element $M^* \in \mathcal{M}^*(N)$ ('orthogonal projection') such that $M - M^*$ is orthogonal to $\mathcal{M}^*(N)$ in the sense that its quadratic covariation with every element of $\mathcal{M}^*(N)$ is the zero process.*

If we take in this proposition N to be a standard Brownian motion W, and we take $\mathbb{F}$ to be the filtration generated by W augmented with the P-null sets of $\mathcal{F}$, then we obtain a much stronger result, that tells us that the 'projection error is zero'. In this case a martingale $M \in \mathcal{M}_2$ is also called *Brownian*, because of the special choice of the filtration.

Theorem 8.3. *Let W be a standard Brownian motion adapted to its own filtration augmented with the P-null sets. Let M be a square integrable martingale, adapted to the same filtration with $M_0 = 0$ (a Brownian martingale), then there exists a progressive process $Y \in \mathcal{L}_T^*(W)$ for all $T \geq 0$, such that*

$$M_t = \int_0^t Y_s dW_s, \ \forall t \geq 0. \tag{8.1}$$

In particular M is a continuous martingale. Y is essentially unique in the sense that if $\widetilde{Y}$ is another process that satisfies equation (8.1), then $E\int_0^\infty (Y_t - \widetilde{Y}_t)^2 dt = 0$.
If M is a local martingale, then (8.1) is still valid for some $Y \in \mathcal{P}_T^*(W)$, but the property $Y \in \mathcal{L}_T^*(W)$ is lost in general.

We omit the proof of this theorem, but emphasize its content. Every martingale that at any time t can be considered as a 'functional' of $W_s, s \leq t$, is actually a stochastic integral.

This feature has no counterpart in discrete time. For instance if $\xi_1, \xi_2, \ldots$ is a sequence of iid standard normals, then $M_t = \sum_{k=1}^t \xi_k^3$ defines a martingale, but there is clearly no (predictable) process Y such that $M_t = \sum_{k=0}^t Y_k \xi_k$.

9. Girsanov's theorem

The theorem, known under the name of Girsanov's theorem, that we treat in this section gives an expression for the Radon-Nikodym derivative of two measures on the σ-algebra $\mathcal{F}_T$. Of particular interest is the case, in which this σ-algebra is generated by a Brownian motion up to time T. First we consider a simple situation common in statistics, in which we in fact deal with a finite family of random variables. Let $Y_1, \ldots, Y_n$ be real random variables that are assumed to be independent and such that each Y_i has a normal $N(\theta_i, \delta_i)$ distribution. Keep the δ_i fixed and denote the probability distribution of the vector $Y = (Y_1, \ldots, Y_n)$ on $\mathbb{R}^n$ by P^θ. Then the density of P^θ (with respect to the Lebesgue measure on $\mathbb{R}^n$) is given by

$$g(y; \theta) := \prod_{i=1}^n (2\pi\delta_i)^{-\frac{1}{2}} \exp(-\frac{1}{2}\sum_{i=1}^n \frac{(y_i - \theta_i)^2}{\delta_i}).$$

Consider two possibilities. In the first one the θ_i equal some real numbers u_i, in the second one all the θ_i are zero. Then we can compute the *likelihood ratio* (which is a Radon- Nikodym derivative)

$$g(Y; u)/g(Y; 0) = \exp(\sum_{i=1}^n \frac{u_i}{\delta_i} Y_i - \frac{1}{2}\sum_{i=1}^n \frac{u_i^2}{\delta_i}).$$

Thus, we obtained an expression for the Radon-Nikodym derivative of two (Gaussian) probability measures. Below we construct the Y_i from a Brownian motion under a given measure and we then define a new measure by selecting a certain random variable as its Radon-Nikodym derivative with respect to the given measure.

So let $(\Omega, \mathcal{F}, P)$ be a probability space and let W be a standard Brownian motion on this space with respect to its own filtration. Fix a time interval $[0, T]$. Let $\Pi_n = \{0 = t_0, \ldots, t_n = T\}$ be a partition of $[0, T]$. Consider the random variables $Y_i = W_{t_i} - W_{t_{i-1}}$, $i = 1, \ldots, n$. Then the Y_i are independent random variables with each a $N(0, \delta_i)$ distribution, where $\delta_i = t_i - t_{i-1}$.

Define now another measure $\widetilde{P}$, absolutely continuous with respect to P, on $\mathcal{F}_T^W$ by its Radon-Nikodym derivative $Z_T = \exp(\sum_{i=1}^n \frac{u_i}{\delta_i} Y_i - \frac{1}{2} \sum_{i=1}^n \frac{u_i^2}{\delta_i})$, so for $F \in \mathcal{F}_T^W$ we have $\widetilde{P}(F) = E[1_F Z_T]$ and $\frac{d\widetilde{P}}{dP} = Z_T$ (here and below, as before, E means expectation under the measure P). Then under $\widetilde{P}$ the Y_i are still independent, but now they have a $N(u_i, \delta_i)$ distribution. In fact, Z_T equals the likelihood ratio of the independent normal random variables with $N(u_i, \delta_i)$ and $N(0, \delta_i)$ distributions respectively that we have seen above.

Define now the process X on $[0, T]$ by $X_t = \sum_{i=1}^n \frac{u_i}{\delta_i} 1_{(t_{i-1}, t_i]}(t)$ and then the process L by $L_t = \int_0^t X_s dW_s$. Then L is a (Gaussian) martingale under P with quadratic variation given by $\langle L \rangle_t = \int_0^t X_s^2 ds$. In particular at time T we get $L_T = \sum_{i=1}^n \frac{u_i}{\delta_i} Y_i$ and $\langle L \rangle_T = \sum_{i=1}^n \frac{u_i^2}{\delta_i}$. Hence we have $Z_T = \exp(L_T - \frac{1}{2} \langle L \rangle_T)$ and with $\mathcal{E}(L)$ the Doléans exponential of L, $Z_T = \mathcal{E}(L)_T$. We will also write Z to denote the process $\mathcal{E}(L)$.

We can also compute the covariation process $\langle L, W \rangle$ (under P) and we get $\langle L, W \rangle_t = \int_0^t X_s ds$. In particular $\langle L, W \rangle_{t_k} = \sum_{i=1}^k u_i$. Define now the process $\widetilde{W}$ by $\widetilde{W}_t = W_t - \langle L, W \rangle_t$. Then $\widetilde{W}_{t_i} - \widetilde{W}_{t_{i-1}} = W_{t_i} - W_{t_{i-1}} - u_i = Y_i - u_i$. Hence, under the measure $\widetilde{P}$ the increments $\widetilde{W}_{t_i} - \widetilde{W}_{t_{i-1}}$ are independent with a $N(0, \delta_i)$ distribution, just as the Y_i had under the measure P. Notice that the variances remain the same. Since $\widetilde{W}$ has continuous paths, this suggests (and it is true in fact, see Corollary 9.2 below) that $\widetilde{W}$ is a Brownian motion under $\widetilde{P}$, obtained from W by subtracting a process (in this example even a function) of bounded variation. Notice that Z is a martingale under P with $EZ_t = EZ_T = 1$.

We will now generalize the situation mentioned above. Consider a continuous local martingale L on the probability space $(\Omega, \mathcal{F}, P)$, adapted to a filtration $\mathbb{F}$. Let $Z = \mathcal{E}(L)$, so $Z_t = \exp(L_t - \frac{1}{2} \langle L \rangle_t)$ and assume that $EZ_T = 1$ (and hence $EZ_t = 1, \forall t \in [0, T]$), then Z becomes a martingale on $[0, T]$. A sufficient condition for this to be true is $E \exp(\frac{1}{2} \langle L \rangle_T) < \infty$, which is known as Novikov's condition.

Define a new measure $\widetilde{P}$ on $\mathcal{F}_T$ by $\widetilde{P}(F) = E[1_F Z_T]$. So Z_T is its Radon-Nikodym derivative with respect to P.

Let now M be another continuous local martingale (adapted to $\mathbb{F}$) and define the process $\widetilde{M}$ by $\widetilde{M} = M - \langle M, L \rangle$.

THEOREM 9.1. *The process $\widetilde{M}$ is under the new measure $\widetilde{P}$ a continuous local martingale on the interval $[0, T]$. Moreover the quadratic variation process of $\widetilde{M}$ under $\widetilde{P}$ coincides with the quadratic variation process of M under P.*

The main ingredients in the proof of this theorem are localization and the fact that $\widetilde{M}$ is a martingale under $\widetilde{P}$ *iff* $Z\widetilde{M}$ is a martingale under P, which is a consequence of the Itô rule for products (7.4).

We specialize to the Brownian case. Let $X \in \mathcal{L}^*_T(W)$ and $L = \int_0^\cdot X_s dW_s$. With this choice of L we get

$$Z_T = \mathcal{E}(L)_T = \exp(\int_0^T X_s dW_s - \frac{1}{2}\int_0^T X_s^2). \tag{9.1}$$

Then, defining $\widetilde{P}$ as before, we obtain

Corollary 9.2. *The process* $\widetilde{W} = W - \langle W, L\rangle = W - \int_0^\cdot X_s ds$ *is a Brownian motion under* $\widetilde{P}$ *on the interval* $[0,T]$.

This corollary immediately follows from Theorem 9.1, since it gives us that $\widetilde{W}$ is a local martingale under $\widetilde{P}$ with quadratic variation (under $\widetilde{P}$) at time t equal to t. The Brownian character then follows from Lévy's characterization, Proposition 7.3.

The application of Girsanov's theorem in finance is the construction of the 'equivalent martingale measure', see Section 12. In problems of statistical inference for stochastic processes this theorem lies at the basis of maximum likelihood estimation. This is not surprising in view of the simple situation that we described in the beginning of this section. To be a bit more precise (we omit the technical details), consider that one wants to model the probabilistic behaviour of some observed continuous process Y on a time interval $[0,T]$. One model is to assume that the probability measure, call it P, on the background is such that Y is a standard Brownian motion. In another model we assume that the probability measure is such that Y is a Brownian motion plus drift, meaning that we have a process X such that $Y - \int_0^\cdot X_s ds$ is Brownian motion. If one assumes that the process X is parametrized by a parameter θ, so $X = X(\theta)$, then we write P^θ for the measure that yields $Y - \int_0^\cdot X_s(\theta) ds$ a Brownian motion. The likelihood $\frac{dP^\theta}{dP}$ (which is assumed to exist) is then just the Z_T of (9.1), with X replaced with $X(\theta)$ and W replaced with Y (recall that Y was Brownian motion under P). The maximizer of this quantity is then by definition the maximum likelihood estimator of θ.

10. Stochastic differential equations

In this section we work on a space $(\Omega, \mathcal{F}, P)$ and a Brownian motion adapted to its own filtration $\mathbb{F}^W$. Moreover on this space lives a random variable ξ that is independent of W. Below we will work with the filtration $\mathbb{F}$, with each member $\mathcal{F}_t$ generated by W_s for all $s \leq t$, ξ and augmented by the P-null sets of $\mathcal{F}^W_\infty \vee \sigma(\xi)$. This filtration satisfies the usual conditions, and W is still a Brownian motion with respect to it, cf. Section 3.

Let $b, \sigma : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be two measurable functions. Consider the following *stochastic differential equation*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \ , \ X_0 \tag{10.1}$$

This equation is understood as a shorthand notation for the corresponding integral version

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{10.2}$$

The function b is called the drift coefficient of equation (10.1) and σ is called the diffusion coefficient. Under appropriate conditions (see below) the Lebesgue-Stieltjes and stochastic integral on the right hand side of (10.2) are well defined.

Heuristically we have the following for a process X that can be represented by (10.2). Consider the displacement of X over an infinitesimally small interval $[t, t+dt]$. Then the conditional mean of the displacement is $E[X_{t+dt} - X_t|\mathcal{F}_t] = b(t, X_t)dt$ and the conditional variance equals $\sigma^2(t, X_t)dt$.

We now introduce the first solution concept for the stochastic differential equation (10.1).

DEFINITION 10.1. *A stochastic process is called a strong solution of equation (10.1) with initial condition ξ, if the following requirements hold.*
(a) X is adapted to $\mathbb{F}$.
(b) $P(X_0 = \xi) = 1$.
(c) $\int_0^t (|b(s, X_s)| + \sigma^2(s, X_s))ds < \infty$ a.s. for all $t \geq 0$.
(d) Equation (10.2) holds a.s. for all $t \geq 0$.
One says that strong uniqueness holds for (10.1) if, given $(\Omega, \mathcal{F}, P)$, W and $\mathbb{F}$ as above any two strong solutions with the same initial conditions are indistinguishable.

We comment a bit on this definition. Condition (b) expresses that ξ is the initial condition. Condition (c) is a technical condition on X to make the integrals in (10.2) well defined. Condition (d) justifies that X is called a solution. Finally condition (a) refers to the 'strongness' of the solution. It expresses that we can interpret (10.2) as a machine that produces at time t the random variable as an output, if we use $W_s, s \leq t$ and ξ as inputs.

We know from the theory of ordinary differential equations, that equations with non-unique solutions exist. Consider for example

$$dX_t = |X_t|^\alpha dt, X_0 = 0.$$

Then we have uniqueness of solutions if $\alpha \geq 1$, whereas for $0 < \alpha < 1$ we have that for any $s \geq 0$ the function defined by $X_t = \{(1-\alpha)(t-s)^+\}^{\frac{1}{1-\alpha}}$ is a solution. Just as for ordinary differential equations we have to impose (local) Lipschitz conditions on the coefficients of (10.1) to have uniqueness. Notice that in the last example the function $x \mapsto |x|^\alpha$ is locally Lipschitz for $\alpha \geq 1$, but not for $0 < \alpha < 1$. More precisely we have

PROPOSITION 10.2. *Assume that $\forall n \in \mathbb{N} : \exists K_n > 0 : \forall t \geq 0, |x|, |y| \leq n$ it holds that*

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq K_n|x-y|.$$

Then strong uniqueness for (10.1) holds.

The key of the proof of this proposition is the following. Assume that X and $\widetilde{X}$ are two strong solutions. Then a localization procedure is followed: define stopping times S_n by $S_n = \inf\{t \geq 0 : |X_t| \vee |\widetilde{X}_t| > n\}$ and subtract the two equations for X and $\widetilde{X}$. After some manipulations (using that b and σ are locally Lipschitz) one obtains the inequality

$$E|X_{t\wedge S_n} - \widetilde{X}_{t\wedge S_n}|^2 \leq 4(t+1)K_n^2 E\int_0^t |X_{s\wedge S_n} - \widetilde{X}_{s\wedge S_n}|^2 ds.$$

An application of Gronwall's inequality gives $E|X_{t\wedge S_n} - \widetilde{X}_{t\wedge S_n}|^2 = 0$, from which the assertion of the proposition follows by letting $S_n \to \infty$.

We also know from the theory of ordinary differential equations that a local Lipschitz condition on the coefficients is insufficient to guarantee the existence of a solution for all $t > 0$. For example, the equation

$$X_t = 1 + \int_0^t X_s^2 ds$$

has a solution on $[0, 1)$ only: $X_t = \frac{1}{1-t}$.

A uniform Lipschitz condition together with a growth condition on the coefficients is sufficient for the existence of strong solutions for all $t \in [0, \infty)$.

PROPOSITION 10.3. *Assume that there exists a constant $K > 0$ such that for all $t \geq 0, x, y \in \mathbb{R}$ the following two inequalities are valid:*

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq K|x-y|,$$

and

$$|b(t,x)| + |\sigma(t,x)| \leq K(1+|x|).$$

Then the stochastic differential equation (10.1) has a unique strong solution defined for all $t \in [0, \infty)$. If moreover $E\xi^2 < \infty$, then also $EX_t^2 < \infty$ for all t.

We highlight some key steps in the proof of this proposition. The idea is to mimic the Picard-Lindelöf iteration procedure known for ordinary differential equations. So we define a sequence of processes X^k by $X^0 \equiv \xi$ and recursively

$$X_t^{k+1} = \xi + \int_0^t b(s, X_s^k)ds + \int_0^t \sigma(s, X_s^k)dW_s.$$

Then, under the condition $E\xi^2 < \infty$, several computations lead to the estimate that for all $T > 0$ there exists a constant L such that

$$E\max_{s\leq t} |X_s^{k+1} - X_s^k|^2 \leq L\int_0^t E|X_s^k - X_s^{k-1}|^2 ds.$$

Iteration of this equality leads to $E\max_{s\le t}|X^{k+1}_s - X^k_s|^2 \le C^*(Lt)^k/k!$ for some finite C^*. An application of the Borel- Cantelli lemma yields the existence of a set $\Omega^* \in \mathcal{F}$ with $P(\Omega^*) = 1$ such that $\{X^k_\cdot(\omega)\}$ is for each $\omega \in \Omega^*$ a Cauchy sequence in $C[0,T]$ (with sup-norm). The resulting (continuous) limit process satisfies equation (10.1) on every interval $[0,T]$ and hence on $[0,\infty)$.

We now turn to another solution concept for a stochastic differential equation, that of a *weak* solution.

DEFINITION 10.4. *A weak solution of equation (10.1) consists of*

- *a probability space* $(\Omega, \mathcal{F}, P)$ *together with a filtration* $\mathbb{F}$ *that satisfies the usual conditions.*
- *a process* X *that is adapted to* $\mathbb{F}$.
- *a Brownian motion* W *that is adapted to* $\mathbb{F}$,

such that conditions (c) and (d) of Definition 10.1 are satisfied. The measure $\mu := P(X_0 \in \cdot)$ *is called the initial distribution.*
One says that uniqueness in law holds for (10.1) if for any two weak solutions $((\Omega, \mathcal{F}, P), \mathbb{F}, X, W)$ *and* $((\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}), \widetilde{\mathbb{F}}, \widetilde{X}, \widetilde{W})$*, satisfying*

$$P(X_0 \in \cdot) = \widetilde{P}(\widetilde{X}_0 \in \cdot),$$

it holds that X *and* $\widetilde{X}$ *have the same law.*

Clearly every strong solution is a weak solution too. Understanding the concept of a weak solution is a little more problematic. A major difference with a strong solution is that the probability space is now part of the solution, and that it is no longer required that X is the output of a machine that uses W as an input. However the distributional properties of a weak solution that is unique in law are completely fixed, which is sufficient for most purposes as long as we are only interested in probabilities that certain events (e.g. that $\max_{t\in[0,1]}|X_t| \le 1$) takes place. For questions of this kind it is sufficient that we can model a process X as (part of) the weak solution of a certain stochastic differential equation. For similar reasons, the concept of uniqueness in law is a reasonable one. There are however other uniqueness concepts.

The following proposition concerns a stochastic differential equation with constant diffusion coefficient. Existence of a weak solution is established under a growth condition on the drift only, Lipschitz conditions are not needed. The proof of this proposition illuminates the absence of adaptiveness in the definition of weak solution as well as the reason why we take a probability space as part of a weak solution.

PROPOSITION 10.5. *Consider the following stochastic differential equation*

$$dX_t = b(t, X_t)dt + dW_t, t \in [0,T]. \tag{10.3}$$

Let μ *be a probability distribution on* $\mathbb{R}$. *Assume that there is a constant* $K > 0$ *such that for all* $t \in [0,T]$ *and for all* $x \in \mathbb{R}$ *it holds that* $|b(t,x)| \le K(1 + |x|)$. *Then there exists a weak solution with initial distribution* μ *on* $[0,T]$.

The proof of this proposition proceeds along the following lines. First we pick a measurable space $(\Omega, \mathcal{F})$ together with a filtration $\mathbb{F} = \{\mathcal{F}_t, t \in [0,T]\}$, a family of measures $\{P^x, x \in \mathbb{R}\}$ and a process X that is Brownian and adapted to $\mathbb{F}$ under each P^x and $P^x(X_0 = x) = 1$. The growth condition on b ensures that the process Z defined on $[0,T]$ by

$$Z_t = \exp(\int_0^t b(s, X_s) dW_s - \frac{1}{2} \int_0^t b(s, X_s)^2 ds)$$

is a martingale under each P^x. Hence we can apply Girsanov's theorem to get that W defined on $[0,T]$ by

$$W_t = X_t - X_0 - \int_0^t b(s, X_s) ds$$

is a standard Brownian motion under the measure Q^x defined on $\mathcal{F}_T$ by $\frac{dQ^x}{dP^x} = Z_T$. But this is just another way of writing (10.3)! The weak solution then is complete by taking the probability measure $P = Q^\mu$ with Q^μ defined by $Q^\mu(\cdot) = \int_{-\infty}^{\infty} Q^x(\cdot)\mu(dx)$.

Notice that from the construction of this solution we get the inclusions $\mathcal{F}_t^W \subset \mathcal{F}_t^X$, just the opposite of the ones for strong solutions! One can also establish uniqueness in law.

Of course introduction of two seemingly different solutions concepts is senseless if there exists no example of a stochastic differential equation that admits a weak solution, but fails to have a strong one. Below we present such an, admittedly somewhat artificial, example.

Let sgn be the real function defined by $\text{sgn}(x) = 1_{[0,\infty)}(x) - 1_{(-\infty)}(x)$ and consider the equation

$$dX_t = \text{sgn}(X_t) dW_t. \tag{10.4}$$

Notice that we cannot apply Theorem 10.3 since the function sgn lacks any Lipschitz property. Suppose that a weak solution exists. Then the solution process X is a continuous square integrable martingale with $\langle X \rangle_t \equiv t$. Hence X is a Brownian motion, because of Lévy's characterization and consequently uniqueness in law holds. But for a Brownian X by the same argument the process W defined by $dW_t = \text{sgn}(X_t) dX_t$ is again Brownian. Hence a weak solution of (10.4) exists. Observe, as a side remark, that along with a weak solution process X also the process $-X$ is a weak solution and that these processes have different paths.

Now we observe that $\text{sgn}(x) = \frac{x}{|x|} 1_{\{0\}^c}(x) + 1_{\{0\}}(x)$ and that $\int_0^\cdot 1_{\{0\}}(X_t) dX_t$ is a square integrable martingale with covariation process $\int_0^\cdot 1_{\{0\}}(X_t) dt$ which is zero by a property of the zero sets of Brownian motion, see Section 3. Hence, from the last example of Section 4 we obtain that this martingale is indistinguishable from zero. Combination of these observations with $d|X_t|^2 = 2X_t dX_t + dt$ (from Itô's rule) yields the alternative way of writing equation (10.4) as

$$dW_t = \frac{1}{2|X_t|} 1_{\{0\}^c}(|X_t|)d(|X_t|^2 - t).$$

From this we obtain the inclusions $\mathcal{F}_t^W \subset \mathcal{F}_t^{|X|}$ for all t.

Suppose now that X were also a strong solution, then we would obtain from condition (a) of Definition 10.1 the inclusions $\mathcal{F}_t^X \subset \mathcal{F}_t^{|X|}$ for a Brownian motion X, in other words, the paths of X are determined by those of $|X|$, which is absurd. A strong solution of equation (10.4) does not exist.

Example. Consider the equation

$$dX_t = -\alpha X_t dt + \sigma dW_t,\ X_0, \tag{10.5}$$

with W standard Brownian motion, α and σ real constants. This equation has a unique strong solution, given by

$$X_t = X_0 + \sigma \int_0^t e^{-\alpha(t-s)} dW_s.$$

Indeed application of the product rule (proposition 7.4) to $e^{-\alpha t} \int_0^t e^{\alpha s} dW_s$ shows this.

If $m(t)$ denotes EX_t, then $m(t) = m(0)e^{-\alpha t}$ and if $V(t)$ denotes the variance of X_t, then $V(t) = \frac{\sigma^2}{2\alpha} + (V(0) - \frac{\sigma^2}{2\alpha})e^{-\alpha t}$.

One can show that X is a Gaussian process, when X_0 has a Gaussian distribution.

If $\alpha > 0$ and X_0 has a normal $N(0, \frac{\sigma^2}{\alpha})$ distribution, then all X_t have the same distribution. Moreover, X is then a stationary Gaussian process, with covariance function $EX_tX_s = \frac{\sigma^2}{2\alpha} e^{-\alpha|t-s|}$. This process is known as the Ornstein-Uhlenbeck process.

Example. Let X be a time varying *geometric Brownian motion* of the type $X_t = X_0 \exp(B(t) + \int_0^t \sigma(s) dW_s)$, where $B \in C^1(\mathbb{R})$ with $B'(t) = \beta(t)$, σ a measurable function with $\int_0^t \sigma^2(s)ds < \infty$ for all t and $X_0 \geq 0$. Application of Itô's rule with $f(x) = e^x$ gives, that X satisfies the following stochastic differential equation.

$$dX_t = X_t(\beta(t) + \frac{1}{2}\sigma^2(t))dt + X_t\sigma(t)dW_t \tag{10.6}$$

It follows from theorem 10.3, that for bounded β and σ this equation has a unique strong solution, the one we started this example with.

Without loss of generality we may take $B(0) = 0$. Assume that X_0 is a strictly positive positive random variable, independent of W. Then for $t \geq u$ we have $\log \frac{X_t}{X_u} = B(t) - B(u) + \int_u^t \sigma(s) dW_s$, which is independent of $\mathcal{F}_u^X$ and has a normal distribution with mean $B(t) - B(u)$ and variance $\int_u^t \sigma^2(s)ds$. It also follows that X is a Markov process.

11. Connection with partial differential equations

There exists a wide range of connections between functionals of the (weak) solution of a stochastic differential equation and solutions of partial differential equations, the most simple one being the following. Let W be a standard Brownian motion. Then W_t, which has a $N(0,t)$ distribution for $t > 0$, has density $x \mapsto p(t,x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{1}{2t}x^2)$. One easily shows that p is a solution of the *heat equation* $\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}$.

Other solutions of the heat equation are $u(t,x) = \int_{-\infty}^{\infty} f(y)p(t,x-y)dy$, when the integral is well defined. These solutions can alternatively be written as $E^x f(W_t)$, where E^x means expectation under the measure for which W is a Brownian motion starting at x. These solutions have the property that $u(0,x) = f(x)$.

Something similar holds for the densities at times t (if they exist) of a solution of a stochastic differential equation. The resulting partial differential equation is known as the Fokker-Planck equation. In this section we concentrate on the Cauchy problem.

As an appetizer we consider the heat equation again. Consider a solution $u(t,x) = E^x f(W_t)$. Fix a terminal time T and define $v(t,x) = u(T-t,x)$. Then v solves the backward heat equation $\frac{\partial v}{\partial t} + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} = 0$ with terminal condition $v(T,x) = f(x)$ and $v(t,x) = E^x f(W_{T-t})$. However, if we write $v(t,x)$ in integral form, then $v(t,x) = \int_{\mathbb{R}} f(y)p(T-t,x-y)dy$. But $y \mapsto p(T-t,x-y)$ is for standard Brownian motion also the conditional density of W_T given $W_t = x$. Denoting expectation with respect to this density by $E^{t,x}$, we get $v(t,x) = E^{t,x} f(W_T)$.

Now we turn to a more general situation, that involves a partial differential equation, that reduces to the backward heat equation by a proper choice of the coefficients. Consider again equation (10.1), with the difference that we take as initial condition $X_t = x$ for some t and x and we look for solutions on $[t,\infty)$. We assume that a weak solution exists, that is unique in law and we denote the probability measure P that is part of the weak solution by $P^{t,x}$ and the corresponding expectations by $E^{t,x}$.

Introduce the differential operators $\mathcal{A}_t$ defined by

$$(\mathcal{A}_t f)(x) = b(t,x)f'(x) + \frac{1}{2}\sigma^2(t,x)f''(x).$$

The Cauchy problem is the following. Find a function $v \in C^{1,2}([0,T) \times \mathbb{R})$ that satisfies for given functions $x \mapsto f(x)$, $(t,x) \mapsto g(t,x)$ and $(t,x) \mapsto k(t,x)$ the partial differential equation

$$\frac{\partial v}{\partial t} + \mathcal{A}_t v = kv - g, \tag{11.1}$$

with terminal condition

$$v(T,x) = f(x),\ x \in \mathbb{R}. \tag{11.2}$$

The functions f, g and k are supposed to be not too wild (we don't enter into details, nonnegativity is one possibility), whereas the coefficients of $\mathcal{A}_t$ are supposed to satisfy the requirements of Proposition 10.3. The following proposition (for the case where X is Brownian motion starting at x known as the Feynman-Kac formula) gives an expression for v in terms of the weak solution of (10.1).

THEOREM 11.1. *If v is a solution of the Cauchy problem and X the unique in law weak solution of (10.1) with $X_t = x$, then v can be represented as*

$$\begin{aligned} v(t,x) &= E^{t,x}[f(X_T)\exp(-\int_t^T k(s,X_s)ds) \\ &\quad + \int_t^T g(s,X_s)\exp(-\int_t^s k(\sigma,X_\sigma)d\sigma)ds]. \end{aligned} \tag{11.3}$$

Hence, the solution v is unique.

In Section 12 we will need this formula only for $g = 0$ and $k \geq 0$. We will explain, in heuristic terms, by using Itô's rule the form of the solution (11.3) for this case. Let X be a (weak) solution of equation (10.1) with some initial condition and let v be a sufficiently smooth function of t and x. Consider the process m defined by

$$m_t = \exp(-\int_0^t k(s,X_s)ds)v(t,X_t). \tag{11.4}$$

Application of Itô's rule to m yields the following (where subscripts denote partial derivatives)

$$\begin{aligned} dm_t &= \exp(-\int_0^t k(s,X_s)ds)[(-k(t,X_t)v(t,X_t) + v_t(t,X_t) \\ &+ b(t,X_t)v_x(t,X_t) + \frac{1}{2}\sigma^2(t,X_t)v_{xx}(t,X_t))dt \\ &+ v_x(t,X_t)\sigma(t,X_t)dW_t]. \end{aligned}$$

Hence we see that if we take v to be a solution of equation (11.1) with $g = 0$, the process m becomes a (local) martingale.

Conversely, let $E|f(X_T)| < \infty$ and consider the martingale m on $[0,T]$ defined by

$$m_t = E[f(X_T)\exp(-\int_0^T k(s,X_s)ds)|\mathcal{F}_t]$$

and let $v_t = m_t \exp(\int_0^t k(s,X_s)ds) = E[f(X_T)\exp(-\int_t^T k(s,X_s)ds)|\mathcal{F}_t]$. Due to the Markovian character of X, we can write $v_t = v(t,X_t)$ and then $v(T,X_T) = f(X_T)$. Again applying Itô's rule we get equation (11.1) back with X_t as the second variable of v.

In the theory of option pricing we get for $v(t,x)$ the price at time t of a European call option with maturity time T and exercise price q if x, is the price of the underlying asset at time t by setting $f(x) = (x-q)^+$ (for a real number r we write $r^+ = \max\{r, 0\}$), $g = 0$ and k the interest rate. This will be shown in Section 12.

12. Applications to the theory of Option Pricing

In this section we consider a market in which two assets are traded, one called stock and the other called bond. We assume that the bond price $P_0(t)$ evolves deterministically in time according to the ordinary differential equation

$$dP_0(t) = r(t)P_0(t)dt. \tag{12.1}$$

Here r is the interest rate function, which is assumed to be bounded.

The price $P_1(t)$ of the stock, the risky asset, is random and evolves according to the stochastic differential equation

$$dP_1(t) = b(t)P_1(t)dt + \sigma(t)P_1(t)dW_t. \tag{12.2}$$

Here $b(t)$ is called the mean rate of return of the stock and $\sigma(t)$ the volatility. Both b and σ are assumed to be deterministic and bounded measurable functions. We also assume that $\inf\{\sigma^2(t) : t \geq 0\} > 0$.

W is standard Brownian motion on a space $(\Omega, \mathcal{F}, P)$, and we take the filtration to be the one that is generated by W, augmented with the P-null sets of $\mathcal{F}$. We have seen in Section 3 that this filtration satisfies the usual conditions.

Under the stipulated conditions (12.2) has a unique strong solution, given by the example in Section 10 with $\beta(t) = b(t) - \frac{1}{2}\sigma^2(t)$. Notice that $P_1(t)$ is always nonnegative if the initial value is so.

We will start to work under the assumption that the interest rate is zero. Then we may also assume without loss of generality that $P_0(t) = 1$ for all t, since it is constant. As a matter of fact, one then uses the price of the bond as the unit in which all other prices are expressed. We write P_t for $P_1(t)$.

Consider an agent who owns at time t the number $N_0(t)$ of units of the bond and $N_1(t)$ units of the stock and that his initial endowment is $x \geq 0$. Then his wealth X_t at time t is given by $X_t = N_1(t)P_t + N_0(t)$. We allow the $N_i(t)$ to be real numbers and hence to take on negative values as well. $N_1(t)P_t$ is called the portfolio process (of the stock) and will be denoted by π_t and the bivariate process (N_0, N_1) is called an investment strategy or trading strategy.

We assume that changes in the wealth are only due to changes in the stock price, or if the stock price would be constant over some time interval, the agent may change the numbers of the two assets only in such a way that his wealth remains the same (the investment strategy is then called *self-financing*). Then we have that

$$\begin{aligned} dX_t &= N_1(t)dP_t && (12.3) \\ &= b(t)\pi_t dt + \sigma(t)\pi_t dW_t. && (12.4) \end{aligned}$$

As a matter of fact, equation (12.3) is in mathematical terms the definition of a self-financing strategy if the bond price is constant and equal to 1.

We impose the reasonable condition that for some time T, the endpoint of the interval of trading times, $\int_0^T \pi_t^2 dt = \int_0^T N_1(t)^2 P_t^2 dt$ is a.s. finite. Then the stochastic integral in equation (12.4) is well defined, see Section 6.

Notice that for an empty portfolio, $\pi_t \equiv 0$, the wealth process is constant, $X_t \equiv x \geq 0$. We will only consider investment strategies for which the resulting wealth process is nonnegative. These strategies are called *admissible*.

The wealth process would be a local martingale if $b(t)$ were identically zero. We accomplish the same result by a change of measure, using Girsanov's theorem. Let $\theta(t) = b(t)/\sigma(t)$, then θ is bounded. Define $Z = \mathcal{E}(-\int_0^\cdot \theta(s) dW_s)$ for $t \leq T$. Then Z is a martingale on $[0, T]$ and we can define the measure $\widetilde{P}$ on $\mathcal{F}_T$ by $\frac{d\widetilde{P}}{dP} = Z_T$. One easily sees that these two measures are equivalent on $\mathcal{F}_T$. In the financial literature this measure is called the equivalent martingale measure. Expectation with respect to $\widetilde{P}$ is denoted by $\widetilde{E}$.

Application of Girsanov's theorem, or rather Corollary 9.2, yields that $\widetilde{W}$ with $\widetilde{W}_t = W_t + \int_0^t \theta(s) ds$ is a Brownian motion under $\widetilde{P}$. Using equation (12.4) and the definition of $\widetilde{W}$, we obtain

$$dX_t = \sigma(t) \pi_t d\widetilde{W}_t, \tag{12.5}$$

which makes the wealth process a nonnegative local martingale under the new measure $\widetilde{P}$. This fact we will use to show that $\widetilde{E} X_T \leq x$.

Let $\{T_n\}$ be a fundamental sequence for X, so $\{X_{T_n \wedge t}, t \in [0, T]\}$ is a $\widetilde{P}$-martingale for each n. Then Fatou's lemma gives

$$\widetilde{E} X_T = \widetilde{E} \lim_{n \to \infty} X_{T_n \wedge T} \leq \liminf_{n \to \infty} \widetilde{E} X_{T_n \wedge T} = \widetilde{E} X_0 = x.$$

We define a *contingent claim* as a nonnegative $\mathcal{F}_T$-measurable random variable. The investment strategy in this context is called *hedging* against the contingent claim if it is admissible and if the the resulting (nonnegative) wealth process X with $X_0 = x$ is such that $X_T = f_T$.

We now define the fair price of the contingent claim f_T at time $t = 0$ as the smallest number $x \geq 0$, such that there exists a hedging investment strategy with initial wealth x. That this definition is reasonable can be argued as follows. Suppose that the above minimum is x_0. Then nobody wants to pay a price $x > x_0$, since for x_0 he will already be able to find a hedging strategy resulting in a terminal wealth equal to f_T. The problem we will address now is how to compute this minimum.

Let x be a value of the initial wealth for which a hedging strategy exists. We have seen that for any admissible strategy process it holds for the resulting wealth process that $\widetilde{E} X_T \leq x$. So if the strategy is hedging we obtain $\widetilde{E} f_T \leq x$. We conclude that $\widetilde{E} f_T$ is a lower bound for the fair price of the contingent claim.

Next we show that for an initial wealth equal to $\widetilde{E} f_T$ a hedging strategy exists (the contingent claim is then called *attainable*), thus obtaining that $\widetilde{E} f_T$

equals the fair price. Consider the martingale M defined by $M_t = \widetilde{E}[f_T|\mathcal{F}_t] - \widetilde{E}f_T$. We apply the representation Theorem 8.3 for Brownian martingales: there exists a progressive process such that $M_t = \int_0^t Y_s d\widetilde{W}_s$. We can do this since $\widetilde{W}$ is obtained by a deterministic shift from W, so $\widetilde{W}$ and W generate the same filtration, and because the $\widetilde{P}$-null sets of $\mathcal{F}_T$ are the same as its P-null sets.

Choose $\pi_t = Y_t/\sigma_t$ and consider the corresponding wealth process with some initial value x. From equation (12.5) we find that $dX_t = Y_t d\widetilde{W}_t$. Hence we get $X_t = x + M_t = \widetilde{E}[f_T|\mathcal{F}_t] - \widetilde{E}f_T + x$. This holds for any x, so we take $x = \widetilde{E}f_T$. Then $X_t = \widetilde{E}[f_T|\mathcal{F}_t] \geq 0$ and in particular $X_T = f_T$ a.s. So we found a hedging strategy against f_T for the initial endowment $\widetilde{E}f_T$ and we conclude that $\widetilde{E}f_T$ is the fair price of the contingent claim at $t = 0$. As a matter of fact for each $t \leq T$ the fair price of the contingent claim is given by $f_t = \widetilde{E}[f_T|\mathcal{F}_t]$.

This has not completely solved the problem of finding the fair price of the contingent claim, we only found a characterization of the fair price. An explicit expression is usually not available. The Black-Scholes framework enables one to give such an expression. In this framework the contingent claim at hand is the terminal payoff of a European call option, $f_T = (P_1(T) - q)^+$, where q is the *exercise price at maturity*. We will return to this later in this section.

So far we have assumed that r was zero. In the case where this is not true, we define $P_t = P_1(t)/P_0(t)$. If we now denote by $\widetilde{X}$ the *discounted* wealth process, then $\widetilde{X}_t = P_0(t)^{-1}X_t = P_0(t)^{-1}(N_1(t)P_1(t) + N_0(t)P_0(t)) = N_1(t)P_t + N_0(t)$ as before. Furthermore, we now call the investment strategy self-financing, if the differential of the discounted wealth process is equal to the right hand side of (12.3). This is equivalent to having for the wealth process X itself the relation

$$dX_t = N_1(t)dP_1(t) + N_0(t)dP_0(t),$$

again reflecting the idea that changes in the wealth process are due to changes in the stock or bond prices only.

The rest of the story is as before upon noticing that in equation (12.4) we have to replace X with $\widetilde{X}$, π with the discounted portfolio process $tilde\pi_t = P_0(t)^{-1}\pi_t$, $b(t)$ with $b(t) - r(t)$ and that we apply Girsanov's theorem with $\theta(t) = \frac{b(t)-r(t)}{\sigma(t)}$. The final result is that the fair price f_t of the contingent claim f_T at time t becomes $\widetilde{E}[\frac{P_0(t)}{P_0(T)}f_T|\mathcal{F}_t]$ or, in discounted terms, $\tilde{f}_t := \frac{f_t}{P_0(t)} = \widetilde{E}[\tilde{f}_T|\mathcal{F}_t]$, as before.

Using the Markov property of both P_1 and P we can write $f_t = v(t, P_1(t))$, $t \leq T$ and $\tilde{f}_t = \tilde{v}(t, P(t))$, $t \leq T$.

The functions v and $\tilde{v}$ are related by

$$v(t,x) = \tilde{v}(t, \frac{x}{P_0(t)})P_0(t). \tag{12.6}$$

Using the results of Section 11 we can then write partial differential equations for both v and $\tilde{v}$. We get (indicating partial derivatives by subscripts) using the stochastic differential equations for both P and P_1 under $\widetilde{P}$

$$\tilde{v}_t(t,x) + \frac{1}{2}\sigma^2(t)x^2\tilde{v}_{xx}(t,x) = 0 \tag{12.7}$$

$$v_t(t,x) + \frac{1}{2}\sigma^2(t)x^2 v_{xx}(t,x) + r(t)xv_x(t,x) - r(t)v(t,x) = 0. \tag{12.8}$$

Indeed the relation (12.6) transforms the above two partial differential equations into each other. We concentrate on equation (12.7) and take $\sigma(t) \equiv 1$. Let for this case $u(t,x) = \tilde{v}(T-t,x)$. Then we have

$$u_t - \frac{1}{2}x^2 u_{xx} = 0. \tag{12.9}$$

Let now $\Sigma^2(t) = \int_0^t \sigma^2(s)ds$, and take $\tilde{v}(t,x) = u(\Sigma^2(T) - \Sigma^2(t), x)$. Then this $\tilde{v}$ solves equation (12.7) if u is a solution of equation (12.9) and $\tilde{v}(T,x) = u(0,x)$ and v and u are related by

$$v(t,x) = u(\Sigma^2(T) - \Sigma^2(t), \frac{x}{P_0(t)})P_0(t). \tag{12.10}$$

Hence, instead of solving the rather complicated looking equation (12.8) it suffices to solve equation (12.9). If we consider (12.9) together with the initial condition

$$u(0,x) = (x-k)^+, \tag{12.11}$$

we find by using methods from the theory of partial differential equations (which reduce (12.9) to the heat equation, cf. Section 11) the solution

$$u(t,x) = x\Phi(\rho_+(t,x)) - k\Phi(\rho_-(t,x)), \tag{12.12}$$

where Φ is the standard normal distribution function and $\rho_\pm$ is given by $\rho_\pm(t,x) = \frac{1}{\sqrt{t}}(\log\frac{x}{k} \pm \frac{1}{2}t)$. So equations (12.10), (12.12) and the substitution $k = \frac{q}{P_0(T)}$ give the fair price of a European call option with exercise price q in the Black-Scholes framework with time varying interest function r and time varying volatility function σ.

As a historical note we mention that in their paper [1] Black and Scholes worked with equation (12.8) for constant interest rate r and constant volatility parameter σ. In this case we get the explicit (Black and Scholes option pricing) formula

$$\begin{aligned} f_t &= P_1(t)\Phi(\frac{1}{\sigma\sqrt{T-t}}(\log\frac{P_1(t)}{q} + (T-t)(r+\frac{1}{2}\sigma^2))) + \\ &\quad -e^{-r(T-t)}q\Phi(\frac{1}{\sigma\sqrt{T-t}}(\log\frac{P_1(t)}{q} + (T-t)(r-\frac{1}{2}\sigma^2))). \end{aligned} \tag{12.13}$$

There exists also an alternative way of arriving at equation (12.13). We use the properties of geometric Brownian motion as we have seen them in the example in Section 10. Recall that the fair price f_t of a contingent claim f_T is given by $f_t = \widetilde{E}[\frac{P_0(t)}{P_0(T)}f_T|\mathcal{F}_t]$. In the European call option framework we then get $f_t = e^{-r(T-t)}\widetilde{E}[(P_1(T)-q)^+|\mathcal{F}_t]$. Since under $\widetilde{P}$ the price P_1 satisfies

$$dP_1(t) = P_1(t)(rdt + \sigma d\widetilde{W}),$$

we get

$$f_t = e^{-r(T-t)}\widetilde{E}[(P_1(t)e^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma(\widetilde{W}_T-\widetilde{W}_t)} - q)^+|\mathcal{F}_t].$$

Denote by ϕ the density of a standard normal random variable and let s_t be the solution of $P_1(t)e^{(r-\frac{1}{2}\sigma^2)(T-t)+s_t\sigma\sqrt{T-t}} - q = 0$. Then

$$\begin{aligned} f_t = \quad & e^{-r(T-t)}(\textstyle\int_{s_t}^{\infty} P_1(t)e^{(r-\frac{1}{2}\sigma^2)(T-t)+z\sigma\sqrt{T-t}}\phi(z)dz \\ & \qquad - \textstyle\int_{s_t}^{\infty} q\phi(z)dz). \end{aligned} \tag{12.14}$$

Both integrals in (12.14) can be evaluated explicitly in terms of the cumulative distribution function Φ, which after some calculations again yields (12.13).

13. Final remarks

The theory of stochastic integration and stochastic differential equations that we outlined in these notes used a *continuous* (local) martingale as the basis process. There also exists a similar theory, in which the continuity assumption is dropped. As a result, the class of integrable processes becomes smaller if one wants to keep the property that a stochastic integral with respect to a local martingale is a local martingale again. This smaller class of processes is formed by the *predictable* processes. Directly from the definition of *predictable* process one sees that every left continuous adapted process is predictable (so for left continuous adapted processes the distinction between predictable and progressive disappears). The construction of a stochastic integral in this case parallels to a large extent what we have done in Section 5. In particular the simple processes are again dense in the class of predictable ones under a suitable metric. Obviously the simple processes, defined in Section 5 are predictable, since they are left continuous and adapted.

The extension of the theory to encompass also integration with respect to discontinuous local martingales is quite natural from a practical point of view. Discontinuous processes are widely used, the Poisson process being the most well known. If N is the standard Poisson process, then $M_t = N_t - t$ is a typical example of a discontinuous martingale. In the books [3], [4], [8] or [6] a general theory of stochastic integration is treated.

For clarity of exposition we confined ourselves to real valued process. The extension to multivariate processes, stochastic integrals and stochastic differential equations in higher dimensions is often rather straightforward, modulo the usual complications that pop up in multidimensional analysis.

References

1. F. Black and M. Scholes (1973), The pricing of options and corporate liabilities, *J. Polit. Economy* **81**, pp 637–659.

2. K.L. Chung and R.J. Williams (1983), *Introduction to Stochastic Integration*, Birkhäuser.
3. C. Dellacherie and P.A. Meyer (1980), *Probabilités et potentiel, Theorie des martingales*, Hermann.
4. J. Jacod (1979), *Calcul Stochastique et Problèmes de Martingales*, LNM 714, Springer.
5. I. Karatzas and S. Shreve (1991), *Brownian Motion and Stochastic Calculus, 2nd edition*, Springer.
6. R.S. Liptser and A.N. Shiryaev (1986), *Theory of Martingales*, Kluwer.
7. B.K. Øksendal (1995), *Stochastic Differential Equations: an introduction, 4th ed.*, Springer.
8. Ph. Protter (1990), *Stochastic Integration and Stochastic Differential Equations: A new approach*, Springer.
9. D. Revuz and M. Yor (1991), *Continuous Martingales and Brownian Motion*, Springer.
10. L.C.M. Rogers and D. Williams (1994), *Diffusions, Markov Processes and Martingales*, Wiley.