An introduction to
stochastic analysis
Lectures at EURANDOM
1999 May 3-5

Peter Spreij
University of Amsterdam
Korteweg-de Vries Institute for Mathematics
Plantage Muidergracht 24
1018 TV Amsterdam
spreij@wins.uva.nl

April 1999

Many papers in mathematical finance contain an equation like

$$dP_t = b(t)P_t\,dt + \sigma P_t\,dW_t,\ P_0. \qquad (1)$$

Sometimes this equation is also written in its integral form

$$P_t = P_0 + \int_0^t b(s)P_s\,ds + \sigma\int_0^t P_s\,dW_s. \qquad (2)$$

In this course we give a meaning to this equation and answer the questions: what is meant by W, dW_t and $\int_0^t P_s dW_s$?
It turns out that 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), a stochastic differential equation. But what is the solution concept of a SDE?

Standard setting starts with a probability space $(\Omega, \mathcal{F}, P)$ and 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 $s < t$.

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

$\mathbb{F}$ satisfies the *usual conditions* if:
(a) $\mathbb{F}$ is *right continuous*, i.e. $\bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} = \mathcal{F}_t$ for all $t \geq 0$ and
(b) $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$.

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 $X_t(\omega)$.

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 $X : (\Omega \times [0, t], \mathcal{F}_t \otimes \mathcal{B}([0, t]) \to (\mathbb{R}, \mathcal{B})$ is measurable.

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. *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.

There are two important notions of 'sameness':

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 $\{\omega : \forall t \geq 0 : X_t(\omega) = Y_t(\omega)\}^c$ is a P-null set.

Proposition. *(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.

Brownian motion

Definition. *A stochastic process W on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is called a standard Brownian motion w.r.t. $\mathbb{F}$, 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.

Look at its discrete time analogue w:
Let $\xi_1, \xi_2, \ldots$ be a sequence of iid standard normal random variables. Define for integers t the random variable $w_t = \sum_{k \leq t} \xi_k$ and $\mathcal{F}_t = \sigma\{\xi_k : k \leq t\}$. Then 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.

Existence of Brownian motion

Consider $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)}))$, 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 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. 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 $\mathcal{F}_t = \sigma\{X_s, s \le t\}$. Then the finite dimensional distributions implicitly defined by properties (i)-(iii) of definition of Brownian motion 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, because it can be shown that $C[0,\infty)$ does not belong to $\mathcal{B}(\mathbb{R}^{[0,\infty)})$.

But, it follows from the Kolmogorov-Čentsov theorem that X has a modification (call it W) with continuous paths.
Then W is a Brownian motion, since modifications don't change the finite dimensional distributions.

Another construction (weak convergence).

Let $\xi_1, \xi_2, \ldots$ be a sequence of *iid* random variables on some probability space with $E\xi_i = 0$ and $E\xi_i^2 = 1$. Let $S_k = \sum_{j \le k} \xi_j$ and for each n

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

Each X^n induces a measure P^n on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$. Donsker's invariance principle is

Proposition. *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.*

Some properties of BM

W is a Markov process and a martingale (see below), nice probabilistic properties.

On the other hand, allthough it has continuous sample paths, they are very irregular. For instance, almost all paths of Brownian motion are nowhere differentiable.

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 $\max\{|t_i - t_{i-1}| : i = 1, \ldots, n\}$.

Some sample path properties

- 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,

  $$V_n := \sum_{\Pi_n} (W_{t_i} - W_{t_{i-1}})^2 \xrightarrow{P} b - a$$

  $$\text{if } m(\Pi_n) \to 0.$$

  In fact, one can show that $E(V_n - (b-a))^2 = 2\sum_{\Pi_n}(t_i - t_{i-1})^2 \le 2m(\Pi_n)(b-a)$. 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.

Brownian motion with an initial condition different from zero.

Let W be a standard Brownian motion on some $(\Omega, \mathcal{F}, P)$ with respect to its own filtration and 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 the definition of BM 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 with initial distribution μ 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 the notation P^x is used 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 obtain 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 w.r.t. this family of σ-algebras. So the filtration $\{\mathcal{F}_t^\mu, t \geq 0\}$ satisfies the usual conditions.

Martingales

Definition. *A stochastic process X is called a martingale w.r.t. $\mathbb{F}$ 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 $\geq$.

Example. Let W be a standard Brownian motion w.r.t. $\mathbb{F}$, 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 w.r.t. $\mathbb{F}$, 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 and since the increments $W_t - W_s$ have variance equal to $t - s$, also $\{W_t^2 - t : t \geq 0\}$ is a martingale.

Doob decomposition of submartingales

Let X be a submartingale 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:
A is an increasing process ($A_k \geq A_{k-1}$ a.s).
And A_k is $\mathcal{F}_{k-1}$-measurable (a *predictable* discrete time process).
Define $M_k = X_k - A_k$, then M is a martingale. We obtained 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.

Doob Meyer decomposition (special case)

Theorem. *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).*

A is an example of a *predictable* process. One defines the predicatable σ-field $\mathcal{P}$ on $\Omega \times (0,\infty)$ as the one generated by the adapted left continuous processes. A process that is measurable w.r.t. $\mathcal{P}$ is then called predictable. In particular, all (left)continuous processes are predictable.

Martingales N with the property $EN_t^2 < \infty$ for every t 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 $\mathcal{M}_2^c$.

Example Let $X = N^2$, $N \in \mathcal{M}_2$. 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.

The process $\langle N \rangle$ 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

$$\begin{aligned} E[(N_t - N_s)^2|\mathcal{F}_s] &= E[N_t^2 - N_s^2|\mathcal{F}_s] &(3)\\ &= E[\langle N\rangle_t - \langle N\rangle_s|\mathcal{F}_s]. &(4)\end{aligned}$$

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

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

$$\sum_{\Pi_n} (N_{t_k} - N_{t_{k-1}})^2 \xrightarrow{P} \langle N \rangle_t. \qquad (5)$$

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

Example in discrete time

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.

The quadratic covariation process of two martingales 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

$$MN - \langle M,N\rangle \text{ is a martingale} \qquad (6)$$

Fact: $\langle M,N\rangle$ is the unique predictable process of bounded variation such that $MN - \langle M,N\rangle$ is a martingale.

Discrete time stochastic integral

Let M be a martingale in discrete time, $M_0 = 0$ and let $\xi_0, \xi_1, \ldots$ 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.
Define the process I 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

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

So we see how to express the quadratic variation of a discrete time stochastic integral in terms of that of M.

Stochastic integrals

The purpose is to *define* the integral $\int_0^1 \xi_s dM_s$ for a martingale M. If for instance $M = W$, then there is a problem if one wants to mimick the definition of a Stieltjes integral, because existence of Stieltjes integrals is guaranteed only for integrators of bounded variation:

Proposition. *(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 $\int_0^1 h d\alpha$ exists and for partitions Π_n of $[0,1]$ whose meshes tend to zero for $n \to \infty$ we have*

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

(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 $[0,1]$.

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 sensible choice is

$$\int_0^t X_s(\omega) dM_s(\omega) = \xi(\omega)(M_{t\wedge b}(\omega) - M_{t\wedge a}(\omega)).$$

Below we drop in the notation the dependence on ω.

Assume that ξ is $\mathcal{F}_a$-measurable and bounded. View the integral as a stochastic process indexed by t.
Then this process is a martingale if M is one, and continuous if M is.
Moreover its quadratic variation at time t is $\xi^2(\langle M\rangle_{t\wedge b} - \langle M\rangle_{t\wedge a}) = \int_0^t X_s^2 d\langle M\rangle_s$.

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

$$X_t = \xi_0 1_{\{0\}}(t) + \sum_{i=0}^{n} \xi_i 1_{(t_i,t_{i+1}]}(t), \qquad (7)$$

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.

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

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

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

Consider also 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 $\mathcal{L}_T^*$.
The following properties hold true.

- $\mathcal{L}_T^*$ 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 (7), then we define the stochastic integral $\int_0^t X_s dM_s$ of X w.r.t. M for $t \leq T$, also often denoted by $I_t(X)$, as

$$I_t(X) = \sum_{i=0}^{n} \xi_i (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}). \qquad (8)$$

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

Proposition. *The stochastic integral defined for simple processes in equation (8) enjoys the following properties.*
(i) $I.(X)$ *is a continuous process with* $I_0(X) = 0$.
(ii) $I.$ *is a martingale on* $[0,T]$.
(iii) $\langle I.(X)\rangle_t = \int_0^t X_u^2 d\langle M\rangle_u$.
(iv) $I_t(\cdot)$ *is for each* t *a linear operator on* $\mathcal{L}_{0,T}$.
(v) $||I.(X)|| = \mu_M(X)$, *in particular* $I.$ *is an isomorphism from* $\mathcal{L}_{0,T}$ *into* $\mathcal{M}_2^c$.

Having defined the stochastic integral w.r.t. 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 $t \leq T$ and $I_t(X) = I_T(1_{(0,t]}X)$.
All the properties of the proposition on slide 27 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 $X, Y \in \mathcal{L}_T^*$

$$\langle I_\cdot(X), I_\cdot(Y)\rangle_t = \int_0^t X_s Y_s d\langle M\rangle_s, \ t \leq T. \quad (9)$$

The properties above, in particular equation (9) can be extended by looking also at other martingales in $\mathcal{M}_2^c$. For instance we have

$$\langle I_\cdot(X), N\rangle_t = \int_0^t X_s d\langle M, N\rangle_s, \ t \leq T \quad (10)$$

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

$$\langle I_\cdot(X; M), I_\cdot(Y; N)\rangle_t = \int_0^t X_s Y_s d\langle M, N\rangle_s, \ t \leq T \quad (11)$$

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

The next proposition gives a characterization of the stochastic integral.

Proposition. *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. *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

$$(YX)dM = Y(XdM).$$

Stopping times

A stopping time T (for $\mathbb{F}$) is a random variable with values in $[0, \infty]$ such that the set $\{T \leq t\}$ belongs to $\mathcal{F}_t$ for all $t \in [0, \infty)$.
A *fundamental* or *localizing* sequence of stopping times is an increasing sequence $T_1, T_2, \dots$ such that $\lim_{n\to\infty} T_n = \infty$.
By 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. *A stochastic process X is called a local martingale (w.r.t. $\mathbb{F}$) if there exists a localizing sequence $T_1, T_2, \dots$ 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 $T_n = \infty$.

Notice also the following important property. If $X \in \mathcal{M}^c_{loc}$, then there exists a fundamental sequence $\{T_n\}$ such that the processes $\{M_{T_n \wedge t}$ belong to $\mathcal{M}^c_2$. (Use the stopping times defined by $\inf\{t : |M_t| \geq n\}$.)

Proposition. *Let $M, N \in \mathcal{M}^c_{loc}$. Then there exists a unique (up to indistinguishability) continuous (hence predictable) process of bounded variation denoted by $\langle M, N\rangle$ such that $MN - \langle M, N\rangle \in \mathcal{M}^c_{loc}$ (compare with (6)). If $M = N$ we simply write $\langle M\rangle$ and this process has non-decreasing paths.*

By definition the class $\mathcal{P}^*_T$, also denoted by $\mathcal{P}^*_T(M)$, for given $M \in \mathcal{M}^c_{loc}$, 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 $M \in \mathcal{M}^c_{loc}$ 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 $\mathcal{M}^c_2$ 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 (10) and (11) are still valid for continuous local martingales M and N and X and Y such that the stochastic integrals exist.

Semimartingales

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

$$X_t = X_0 + M_t + B_t, \ \forall t \geq 0, \qquad (12)$$

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 (12) 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 w.r.t. a semimartingale.
Let X be a continuous semimartingale with decomposition (12) and let Y be a progressive process (in particular measurable).
Assume that the pathwise defined Lebesgue-Stieltjes integral $\int_0^T |Y_s| dB_s$ is finite a.s. and also that $Y \in \mathcal{L}^*_T(M)$, then the stochastic integral of Y with respect to X is defined as

$$\int_0^t Y_s dX_s := \int_0^t Y_s dM_s + \int_0^t Y_s dB_s, \ t \leq T. \quad (13)$$

Here the first integral is of course a stochastic integral.

The Itô rule

Theorem. *Let X be a continuous semimartingale as in (12) and $f \in C^2(\mathbb{R},\mathbb{R})$. Then $f(X)$ is again a continuous semimartingale and for all $t \geq 0$*

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_s)dX_s \\ &\quad + \frac{1}{2}\int_0^t f''(X_s)d\langle M\rangle_s \\ &= f(X_0) + \int_0^t f'(X_s)dM_s \\ &\quad + \int_0^t f'(X_s)dB_s + \frac{1}{2}\int_0^t f''(X_s)d\langle M\rangle_s. \end{aligned} \tag{14}$$

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.
For a semimartingale X with decomposition (12) one often writes $\langle X\rangle$ instead of $\langle M\rangle$. With this convention equation (14) may be written in differential form

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

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 see clearly that we can't do without the correction term, since $EW_T^2 = T$, whereas the stochastic integral, being a martingale, has expectation zero.

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 $\mu_W(W - W^n)^2 = \frac{1}{2}\sum_{i=0}^{n-1}(t_{i+1}-t_i)^2$, which tends to zero for partitions whose meshes tend to zero. Elementary algebra shows that

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

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, 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 $Z_t = \exp(M_t - \frac{1}{2}\langle M\rangle_t)$. Application of Itô's rule yields that Z satisfies

$$Z_t = 1 + \int_0^t Z_s dM_s. \tag{15}$$

The process Z thus defined is known as the Doléans exponential of M, and is often denoted by $Z = \mathcal{E}(M)$.
In general Z is a local martingale only. Important is the case in which it becomes a martingale (with expectation 1). We will return to it, when we discuss Girsanov's theorem.

The following result is known as Lévy's characterization of Brownian motion.

Proposition. *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 for martingales is to apply Itô's rule with the function $f(x) = \exp(iux)$. The result is

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

Multiplication of this equation by $\exp(-iuM_s)$ and taking conditional expectation w.r.t. $\mathcal{F}_s$ 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 $\exp(-\frac{1}{2}(t-s)u^2)$.

The Itô rule can be extended to multivariate processes. One important instance is the product rule for semimartingales.

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

$$X_T Y_T = X_0 Y_0 + \int_0^T X_t dY_t + \int_0^T Y_t dX_t + \langle X, Y\rangle_T, \tag{16}$$

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

We also often write

$$d(XY) = XdY + YdX + d\langle X, Y\rangle.$$

If Y happens to be a continuous semimartingale of bounded variation, then this formula reduces to

$$d(XY) = XdY + YdX.$$

Another important case occurs when one of the two semimartingales is a continuous function or process V that is of bounded variation. Then

$$\begin{aligned} df(V_t, X_t) &= f_v(V_t, X_t)dV_t \\ &\quad + f_x(V_t, X_t)dX_t + \frac{1}{2}f_{xx}(V_t, X_t)d\langle X\rangle_t. \end{aligned} \tag{17}$$

The general Itô formula for a vector
$X = (X^1, \dots, X^d)$ of continuous semimartingales reads

$$\begin{aligned} df(X_t) &= \sum_{i=1}^d f_{x^i}(X_t)dX_t^i \\ &\quad + \frac{1}{2}\sum_{i,j=1}^d f_{x^i x^j}(X_t)d\langle X^i, X^j\rangle_t. \end{aligned}$$

Another (univariate) extension of the Itô formula occurs, when f is not in $\mathcal{C}^2$, but only in $\mathcal{C}^1$ with an absolutely continuous derivative f'. In this case there still exists a locally integrable function f'' such that $f'(b)-f'(a) = \int_a^b f''(x)dx$. Itô's formula continues to hold.

This result is of particular importance, when f' is differentiable in all x, except finitely many, e.g. if $f(x) = x|x|$. Applied to W this gives $W_t|W_t| = 2\int_0^t |W_s|dW_s + \int_0^t \text{sgn}(W_s)ds$.

Exercise: compute the semimartingale decomposition of $W_t|W_t|$ by applying the product rule and local time (see below).

Stochastic differential equations

Consider a probability 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.

Consider 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.

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

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \qquad (18)$$

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

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \qquad (19)$$

The function b is called the drift coefficient of equation (18) and σ is called the diffusion coefficient. Under appropriate conditions (see below) the Lebesgue-Stieltjes and stochastic integral on the right hand side of (19) are well defined.

We have the following interpretation for a process X that can be represented by (19).
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 $\sigma^2(t, X_t)dt$.

First some **Examples.** Consider the equation

$$dX_t = -\alpha X_t dt + \sigma dW_t,\ X_0, \qquad (20)$$

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

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

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

A multivariate extension of the above equation is

$$dX_t = A(t)X_t dt + \Sigma dW_t, X_0 = x_0.$$

Here A is a locally integrable matrix valued function, Σ a matrix and W a multivariate Brownian motion.
Let $\Phi(\cdot, s)$ be the fundamental solution of the ODE

$$\dot{Y}(t) = A(t)Y(t), Y(s) = I.$$

Then $X_t = \Phi(t, 0)x_0 + \int_0^t \Phi(t, s)\Sigma dW_s$.
The proof is as above.

NB. If $A(\cdot) = A$, constant, then $\Phi(t, s) = \exp A(t - s)$.

Example. Let X be a time varying *geometric Brownian motion*:

$$X_t = X_0 \exp(B(t) + \int_0^t \sigma(s)dW_s),$$

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

$$dX_t = X_t(\beta(t) + \frac{1}{2}\sigma^2(t))dt + X_t\sigma(t)dW_t. \quad (21)$$

If $B(0) = 0$ and X_0 is 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.

We can also look at the above example in a different way. Start out with equation (21) and write $b(t) = \beta(t) + \frac{1}{2}\sigma^2(t)$, so

$$dX_t = X_t(b(t)dt + \sigma(t)dW_t), X_0.$$

Define for any continuous semimartingale Y its Doléans exponential $\mathcal{E}(Y)$ to be the process $\exp(Y - \frac{1}{2}\langle Y\rangle)$. An application of Itô's rule shows that $\mathcal{E}(Y)$is a continuous semimartingale with $d\mathcal{E}(Y)_t = \mathcal{E}(Y)_t dY_t$.
Let $Y_t = \int_0^t (b(s)ds + \sigma(s)dW_s)$, so $dX_t = X_t dY_t$.
Then Y is a continuous semimartingale with $\langle Y\rangle = \int_0^\cdot \sigma^2(s)ds$.
Hence we can write $X = X_0\mathcal{E}(Y)$, which is equal to the expression we encountered above.

We continue with introducing solution concepts. The setting is that of slide 45.

Strong solutions

Definition. *A stochastic process is called a strong solution of equation (18) 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 (19) holds a.s. for all $t \geq 0$.
One says that strong uniqueness holds for (18) 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 (19) 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 (19) as a machine that produces at time t the random variable X_t 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 (18) 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. *Assume that* $\forall n \in IN : \exists K_n > 0 : \forall t \geq 0, |x|, |y| \leq n$ *it holds that*

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

Then strong uniqueness for (18) holds.

Some remarks on the proof of this proposition: Assume that X and $\tilde{X}$ are two strong solutions. Then a localization procedure is followed: define stopping times S_n by $S_n = \inf\{t \geq 0 : |X_t| \vee |\tilde{X}_t| > n\}$ and subtract the two equations for X and $\tilde{X}$.

After some manipulations (using that b and σ are locally Lipschitz) one obtains the inequality

$$E|X_{t\wedge S_n} - \tilde{X}_{t\wedge S_n}|^2$$

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

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

Intermezzo (Gronwall's inequality)

If $x(t) \leq \beta(t) + C\int_0^t x(s)ds$ for all $t \in [0,T]$, then $x(t) \leq \beta(t) + C\int_0^t \exp(C(t-s))\beta(s)ds$.

Proof. Write $x(t) = \beta(t) + C\int_0^t x(s)ds - p(t)$. Then p is a positive function. Solving this (linear) equation results in $x(t) = \beta(t) - p(t) + C\int_0^t \exp(C(t-s))(\beta(s)-p(s))ds$. Use that $p(t) \geq 0$.

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. *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 (18) 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. More precisely, for all $T > 0$ there exists $C = C(K,T)$ such that

$$E|X_t|^2 \leq C(1 + E\xi^2)e^{Ct}, \forall t \in [0, T].$$

Moreover the solution has the Markov property:

$$E[f(X_t)|\mathcal{F}_s] = E[f(X_t)|X_s],$$

whenever $Ef(X_t)$ exists.

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

$$E \max_{s \leq t} |X_s^{k+1} - X_s^k|^2 \leq C^*(Lt)^k / k!$$

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 $\{X_\cdot^k(\omega)\}$ is a Cauchy sequence in $C[0,T]$ (with sup-norm) for each $\omega \in \Omega^*$.
The resulting (continuous) limit process satisfies equation (18) on every interval $[0, T]$ and hence on $[0, \infty)$.

Weak solutions

Definition. *A weak solution of equation (18) 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) above are satisfied. The measure $\mu := P(X_0 \in \cdot)$ *is called the initial distribution.*
One says that uniqueness in law holds for (18) if for any two weak solutions $((\Omega, \mathcal{F}, P), \mathbb{F}, X, W)$ *and* $((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}), \tilde{\mathbb{F}}, \tilde{X}, \tilde{W})$, *satisfying* $P(X_0 \in \cdot) = \tilde{P}(\tilde{X}_0 \in \cdot)$, *it holds that* X *and* $\tilde{X}$ *have the same law.*

Clearly every strong solution is a weak solution too.
The 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 probabilistic modelling.
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. *Consider the following stochastic differential equation*

$$dX_t = b(t, X_t)dt + dW_t, t \in [0, T]. \qquad (22)$$

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* $|b(t, x)| \leq K(1 + |x|)$. *Then there exists a weak solution with initial distribution* μ *on* $[0, T]$.

The proof of this proposition is based on Girsanov's theorem and will be discussed later on.

Connection with PDEs

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*

$$\frac{\partial u}{\partial t} - \frac{1}{2}\frac{\partial^2 u}{\partial x^2} = 0.$$

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 $u(0, x) = f(x)$.

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 $v(t,x) = E^x f(W_{T-t})$, or in integral form,

$$v(t,x) = \int_{\mathbb{R}} f(y)p(T-t,x-y)dy.$$

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 w.r.t. this density by $E^{t,x}$, we get $v(t,x) = E^{t,x} f(W_T)$.

Consider again the SDE (18), with the difference that we take as initial condition $X_t = x$ for some t and x and we look for solutions on $[t,\infty)$.
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 family of operators $\mathcal{A}_t$ is called the generator of X. The generator has the following property: M defined by

$$M_t = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}_s f)(X_s)ds$$

is a martingale for bounded twice continuously differentiable functions f.

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

$$\frac{\partial v}{\partial t} + \mathcal{A}_t v = kv - g, \tag{23}$$

with terminal condition

$$v(T,x) = f(x),\ x \in \mathbb{R} \tag{24}$$

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 (in particular linear growth) to have a weak solution of the SDE (18).

Feynman-Kac formula

Theorem. *If v is a solution of the Cauchy problem and X the unique in law weak solution of (18) with $X_t = x$, then v can be represented as*

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

Hence, the solution v is unique.

In a financial applications g is often zero and then

$$v(t,x) = E^{t,x}[f(X_T)\exp(-\int_t^T k(s,X_s)ds).$$

Here we discuss the case for $g = 0$. Let X be a (weak) solution of equation (18) with initial condition $X_t = x$ and let v be a sufficiently smooth function of t and x. Consider the process m on $[t, T]$ defined by

$$m_\tau = \exp(-\int_t^\tau k(s, X_s)ds)v(\tau, X_\tau). \qquad (25)$$

Application of Itô's rule to m yields the following

$$\begin{aligned} dm_\tau &= \exp(-\int_t^\tau k(s, X_s)ds)[(-k(\tau, X_\tau)v(\tau, X_\tau) \\ &\quad + v_t(\tau, X_\tau) + b(\tau, X_\tau)v_x(\tau, X_\tau) \\ &\quad + \frac{1}{2}\sigma^2(\tau, X_\tau)v_{xx}(\tau, X_\tau))d\tau \\ &\quad + v_x(\tau, X_\tau)\sigma(\tau, X_\tau)dW_\tau] \\ &= \exp(-\int_t^\tau k(s, X_s)ds)[(-k(\tau, X_\tau)v(\tau, X_\tau) \\ &\quad + v_t(\tau, X_\tau) + \mathcal{A}_\tau v(\tau, X_\tau)d\tau \\ &\quad + v_x(\tau, X_\tau)\sigma(\tau, X_\tau)dW_\tau]. \end{aligned}$$

Hence we see that if we take v to be a solution of equation (23) with $g = 0$, the process m becomes a (local) martingale,

$$dm_\tau = \exp(-\int_t^\tau k(s, X_s)ds)v_x(\tau, X_\tau)\sigma(\tau, X_\tau)dW_\tau.$$

Assuming moreover the martingale property of m, so $m_t = E^{t,x}M_T$ we get from

$$m_t = v(t, X_t) = v(t, x)$$

and

$$\begin{aligned} m_T &= \exp(-\int_t^T k(s, X_s)ds)v(T, X_T) \\ &= \exp(-\int_t^T k(s, X_s)ds)f(X_T) \end{aligned}$$

the desired result.

Local time

Let W be Brownian motion. The following limit exists a.s. and in L^2-sense:

$$L(t, x) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t 1_{(x-\varepsilon, x+\varepsilon)}(W_s)ds,$$

and is called the local time of W at x. It has the intuitively clear property that

$$\int_a^b L(t, x)dx = \int_0^t 1_{(a,b)}(W_s)ds.$$

More general it holds that

$$\int_{-\infty}^\infty L(t, x)f(x)dx = \int_0^t f(W_s)ds,$$

for locally integrable Borel functions f.

Local time also has the following properties. It is a.s. continuous in both t and x, it is a.s increasing in t or each fixed x and

$$\int_0^\infty 1_{\{x\}^c}(W_t)L(dt, x) = 0.$$

Extensions of Itô's formula

Consider the function $f_x(y) = |y - x|$. Formal differentiation gives

$$f_x'(y) = \text{sgn}(y - x)|, f_x''(y) = \delta_x(y).$$

If we apply Itô's formula to the function f_x and interpret $\delta_x(y)$ as $\lim_{\varepsilon\downarrow} \frac{1}{2\varepsilon} 1_{(x-\varepsilon, x+\varepsilon)}(y)$, then one obtains

$$|W_t - x| = |W_0 - x| + \int_0^t \text{sgn}(W_s - x)dW_s + L(t, x).$$

This result (known as Tanaka's formula) can be proven rigourously. The key trick is to approximate $|\cdot|$ with a smooth function, for instance convolution with the density of a normal $N(0, \sigma^2)$ distribution. Call the resulting function f_σ. It has derivative $f_\sigma'(x) = 2\Phi(x/\sigma) - 1$ and second derivative $f_\sigma''(x) = 2\phi(x/\sigma)/\sigma$. Notice that $f_\sigma'(x) \to \text{sgn}(x)$ and $f_\sigma''(x) \to \delta(x)$. Tanaka's formula follows by applying the Itô rule to $f_\sigma(W_t)$ and then sending σ to 0.

Let f be a real convex function. It is known that f is continuous, that the right-hand and left-hand derivatives D^+f and D^-f are defined everywhere and coincide except on a (at most) countable set. Both D^+f and D^-f are increasing functions.

Define for convex f a measure μ on $(\mathbb{R}, \mathcal{B})$ by $\mu((a,b]) = D^+f(b) - D^+f(a)$. Then

Proposition.

$$\begin{aligned} f(W_t) &= f(W_0) + \int_0^t D^+f(W_s)ds \\ &\quad + \frac{1}{2}\int_{-\infty}^{\infty} L(t,x)d\mu(x). \end{aligned}$$

With the special choice $f(y) = |y-x|$ we get Tanaka's formula back.

More on SDE

Of course having seemingly different solutions concepts for SDEs 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 $\text{sgn}(x) = 1_{[0,\infty)}(x) - 1_{(-\infty,0)}(x)$ and consider the equation

$$dX_t = \text{sgn}(X_t)dW_t. \tag{26}$$

Notice that 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 (26) 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. Hence, this martingale is indistinguishable from zero. Combination of these observations with $d|X_t|^2 = 2X_tdX_t + dt$ (from Itô's rule) yields the alternative way of writing equation (26) 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 the definition of a strong solution 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 (26) does not exist.

Another way of reaching the same conclusion by means of local time is to notice that for a strong solution we would have by Tanaka's formula

$$|X_t| = W_t + L^X(t,0),$$

which also implies $\mathcal{F}_t^W \subset \mathcal{F}_t^{|X|}$ and the same contradiction as above arises when a strong solution is assumed to exist.

Extensions of SDE theory

We mention two extensions of the above presented theory. One concerns replacement of the functions b and σ with random functions, the other one concerns functional equations in which $b(t, X_t)$ is replaced with $b(t, X^t)$, where X^t denotes the whole past of the process X up to time t.

The first case is of particular importance for stochastic control problems, where along with the process X also a control process u is considered, that influences $X = X(u)$ in a certain way. The stochastic control problem is then to choose u in such a way that some criterion on X is optimized, given appropriate constraints on u.

The SDE we consider is

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, X_0 = x_0 \in \mathbb{R}. \tag{27}$$

Here for all fixed x the *processes* $b(\cdot, x)$ and $\sigma(\cdot, x)$ are assumed to be progressively measurable. In stochastic control one chooses $b(t, x, \omega) = f(t, x, u(t, \omega))$ and $\sigma(t, x, \omega) = \gamma(t, x, u(t, \omega))$.

Proposition. *Assume that the following conditions are satisfied.*

Let there be a constant K and a stochastic process g such that

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

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

and such that $\alpha_m = E \int_0^T g(t)^m dt < \infty$ for some $m \geq 2$.

Then the SDE (27) has a unique strong solution on $[0, T]$ satisfying

$$E|X_t - x_0|^m \leq C[|x_0|^m + T^{\frac{m}{2}-1}e^{CT}(|x_0|^m + \alpha_m)],$$

for a constant $C = C(K, m, T)$

Measure transformations (discrete time)

Let $Y_1, \ldots, Y_n$ be independent random variables, 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 the *likelihood ratio* $\frac{dP^u}{dP^0}$ (which is a Radon- Nikodym derivative) becomes

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

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.

Let W be a standard Brownian motion on $(\Omega, \mathcal{F}, P)$ with respect to its own filtration.

Fix a time interval $[0, T]$ and 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 with $N(0, \delta_i)$ distributions, where $\delta_i = t_i - t_{i-1}$.

Define $\tilde{P}$, absolutely continuous with respect to P, on $\mathcal{F}_T^W$ by its Radon-Nikodym derivative $\frac{d\tilde{P}}{dP} = Z_T$,

$$Z_T = \exp(\sum_{i=1}^n \frac{u_i}{\delta_i}Y_i - \frac{1}{2}\sum_{i=1}^n \frac{u_i^2}{\delta_i}),$$

so for $F \in \mathcal{F}_T^W$ we have $\tilde{P}(F) = E[1_F Z_T]$.

Under $\tilde{P}$ 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 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 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

$$Z_T = \exp(L_T - \frac{1}{2}\langle L \rangle_T) = \mathcal{E}(L)_T$$

We will also write Z to denote the process $\mathcal{E}(L)$.

The covariation process $\langle L, W \rangle$ (under P) is $\langle L, W \rangle_t = \int_0^t X_s ds$. In particular $\langle L, W \rangle_{t_k} = \sum_{i=1}^k u_i$.
Define now the process $\tilde{W}$ by

$$\tilde{W}_t = W_t - \langle L, W \rangle_t.$$

Then $\tilde{W}_{t_i} - \tilde{W}_{t_{i-1}} = W_{t_i} - W_{t_{i-1}} - u_i = Y_i - u_i$.
Hence, under the measure $\tilde{P}$ the increments $\tilde{W}_{t_i} - \tilde{W}_{t_{i-1}}$ 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*.
It is a fact, see below, that $\tilde{W}$ is a Brownian motion under $\tilde{P}$, 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 $EZ_t = EZ_T = 1$.

The general case

Consider a continuous local martingale L on the probability space $(\Omega, \mathcal{F}, P)$, adapted to a filtration $\mathbb{F}$ with $L_0 = 0$.
Let $Z = \mathcal{E}(L)$, so

$$Z_t = \exp(L_t - \frac{1}{2}\langle L \rangle_t).$$

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 $\tilde{P}$ on $\mathcal{F}_T$ by $\tilde{P}(F) = E[1_F Z_T]$. So Z_T is its Radon-Nikodym derivative w.r.t. P.
Let now M be another continuous local martingale (adapted to $\mathbb{F}$) and define the process $\tilde{M}$ by $\tilde{M} = M - \langle M, L \rangle$.

Girsanov's theorem

Theorem. *The process $\tilde{M} = M - \langle M, L \rangle$ is a continuous local martingale under the new measure $\tilde{P}$ on the interval $[0,T]$.*
Moreover the quadratic variation process of $\tilde{M}$ under $\tilde{P}$ 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

$\tilde{M}$ is a martingale under $\tilde{P}$ *iff* $Z\tilde{M}$ is a martingale under P.

To show that $Z\tilde{M}$ is a martingale under P follows from the Itô rule for products.

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

$$Z_T = \mathcal{E}(L)_T = \exp(\int_0^T X_s dW_s - \frac{1}{2}\int_0^T X_s^2). \tag{28}$$

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

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

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

The application of Girsanov's theorem in finance is the construction of the "equivalent martingale measure".

In statistics for stochastic processes it is the basis of maximum likelihood estimation. Suppose that one observes a continuous process Y on a time interval $[0, T]$. One model is to assume that there is a probability measure P 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 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 (28), 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.

Application: Weak solutions to (22)

The proof of the proposition proceeds along the following lines.

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 (22)!

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 $\mathcal{F}^W_t \subset \mathcal{F}^X_t$, just the opposite of the ones for strong solutions! One can also establish uniqueness in law.

Representation of martingales

First a weak result.

Proposition. *Let M be a continuous local martingale w.r.t. 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 $M = I.(\alpha; W)$.*

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 $W := I.(\alpha^{-1}; M)$ is well defined and $\langle W\rangle_t \equiv t$. Hence we obtain form 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. *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. *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*

$$M_t = \int_0^t Y_s dW_s, \ \forall t \geq 0. \qquad (29)$$

In particular M is a continuous martingale. Y is essentially unique in the sense that if $\tilde{Y}$ is another process that satisfies equation (29), then $E\int_0^\infty (Y_t - \tilde{Y}_t)^2 dt = 0$.

If M is a local martingale, then (29) is still valid for some $Y \in \mathcal{P}_T^*(W)$, but the property $Y \in \mathcal{L}_T^*(W)$ is lost in general.

Literature

I. Karatzas & S. Shreve (1988), *Brownian motion and stochastic calculus*

Ph. Protter (1990), *Stochastic Integration and Stochastic Differential Equations: A new approach*, Springer.

B.K. Øksendal (1995), *Stochastic Differential Equations: an introduction, 4th ed.*, Springer.

K.L. Chung and R.J. Williams (1983), *Introduction to Stochastic Integration*, Birkhäuser.

G. Kallianpur (1980), *Stochastic filtering theory*, Springer.

D. Revuz and M. Yor (1991), *Continuous Martingales and Brownian Motion*, Springer.

L.C.G. Rogers and D. Williams (1994), *Diffusions, Markov Processes and Martingales*, Wiley.

T. Mikosch (1998), *Elementary Stochastic Calculus with Finance in View*, World Scientific.

C. Dellacherie and P.A. Meyer (1980), *Probabilités et potentiel, Theorie des martingales*, Hermann.

J. Jacod (1979), *Calcul Stochastique et Problèmes de Martingales*, LNM 714, Springer.