STOCHASTIC CALCULUS

Honours Project Report

Laird Breyer, B.Sc., B.A.
Department of Mathematics
University of Queensland

Supervisor: Dr. P. K. Pollett

November, 1994

Preface

> The importance of the martingale concept cannot be overemphasized. [...] Martingales, Markov dependence and stationarity are the only three dependence concepts so far isolated which are sufficiently general and sufficiently amenable to investigation yet with a great number of deep properties.
>
> M. Loève. *Ann. Probab.* (1973) vol. 1, no. 1, p.6

This project report deals with the stochastic calculus of semimartingales. In Doob's classical book [Doo53], the semimartingales are processes which we nowadays call submartingales. The change in terminology was heralded in the preface of Loève's third edition (1962) of his book *Probability Theory* [Loè77].

A quarter of a century after Doob's book, the term semimartingale reappears on the scene, this time in Meyer's classic "Course on Stochastic Integrals" [Mey76]. Had Loève written the lines at the top of this page a few years later, he would have doubtlessly included semimartingales in his list.

Historically speaking, there are two ways of dealing with stochastic processes. One is via their transition probabilities, and the other is by directly dealing with their sample paths. The first approach was favoured in the early days; researchers in stochastic processes routinely deal with paths which would horrify a typical nineteenth-century analyst! With the advent of Stochastic Calculus however, the sample path approach gained currency.

The first part of this report deals with the construction of the stochastic integral. We look at what it means to write

$$\int_a^b H_s dX_s$$

when both H and X are random functions of s: stochastic processes. Wiener was the first on the scene; having spectacularly constructed a mathematical model of Brownian motion in [Wie23], he went on to define the integral when H is an ordinary deterministic function and X is Brownian motion. The interpretation is as one would expect: a weighted sum (by H) of small increments of X. To do this, he surmounted a fundamental difficulty: the paths of X are of unbounded variation on every time interval, so that the integral above is most definitely not a Riemann-Stieltjes integral. But the real breakthrough came with Itô (see [Itô51]) who recognized that to integrate random functions H, the process H must not be allowed to anticipate the process X. Itô went on to prove his famous theorem, a kind of fundamental theorem of calculus for his integral.

Suddenly, stochastic integration became useful. By interpreting the Brownian motion as the cumulative effect of many small random perturbations, an integral of "white noise", it became possible to study systems under the influence of randomness in the same way that Newton had studied systems under the influence of nonrandom forces. Itô's integral with respect to Brownian motion spawned new eras in population dynamics, filtering and control, and mathematical finance, to name but a few.

But probabilists did not stand still. Doob recognized that the integral depended on the martingale property of Brownian motion. Later work by Kunita-Watanabe showed that the integral could be localized, so as to increase the available number of integrating processes. When Meyer wrote his course on stochastic integrals, the theory had essentially taken its definitive form. But it was technically very complicated. It took a hundred and fifty pages to explain, and presumed known some important theorems about stochastic processes.

The following years were spent gradually simplifying the theory (and, much more importantly, applying it with spectacular success). The first part of this thesis presents the theory in a way which I believe is about as short and as simple as possible, while still keeping it very powerful. There is a simpler way, due to Protter in [Pro86] (recommended reading!), but his exposition is at the expense of developing a "Riemann integral" instead of a "Lebesgue integral".

Here's an outline of what we'll do in Part I: Processes are introduced as functions on $\mathbb{R}_+ \times \Omega$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is our probability space. Next, the space $\mathbb{R}_+ \times \Omega$ is endowed with a σ-algebra, called the predictable σ-algebra $\mathcal{P}$. The $\mathcal{P}$-measurable functions represent the nonanticipating processes of Itô's theory. Sets of the form $]S, T]$, where S, T are random stopping times, are seen to belong to $\mathcal{P}$. An integrating process is any process X which induces an $L^1(\Omega, \mathcal{F}, \mathbb{P})$-valued measure μ_X on $\mathcal{P}$ such that

$$\mu_X(]S, T]) = X_T - X_S.$$

The integration theory of such measures is recalled. It is the Dunford and Schwarz vector integration theory ([DS58], section IV.10). Processes which induce a measure are characterized. Although their sample paths are quite often of unbounded variation, they have a bounded "probabilistic" variation. They are the semimartingales, and may also be thought of as processes which have a decomposition into a "signal" and a (generalized) "noise". The fundamental theorem of calculus (Itô's formula) for semimartingales is investigated. It takes the form

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \text{“deviation”}.$$

Attention is focused on the "deviation" and methods for computing it are found. Its existence is directly related to the unboundedness of the variation of the sample paths of X.

The idea of considering stochastic integrals as Dunford-Schwarz vector integrals is not new. The classic reference for this seems to be [Kus77], but its simplicity appears to be lost to probabilists. When pressed for a quick development of stochastic integration, all seem to fall back on the "isometry approach" (see notes at end of Part I) and mention in passing that using vector measures is nice, but much more complicated. This is of course true when the vector measures are considered with values in the Orlicz space $L^0(\Omega, \mathcal{F}, \mathbb{P})$. This is why Protter considers the integral as a continuous linear functional on random processes with values in $L^0(\Omega, \mathcal{F}, \mathbb{P})$. See [Pro86, Pro92]. I will attempt to show that the vector integral approach can be very simple when the measures are restricted to have values in the Banach space $L^1(\Omega, \mathcal{F}, \mathbb{P})$ instead of $L^0(\Omega, \mathcal{F}, \mathbb{P})$. The approach was inspired by the book [KK76], and is quite close to [Kus77], though I believe simpler.

We will consider a semimartingale $X = N + A$, where N is a generalized "noise", and A a "signal". The process $f(X)$ turns out to be again a semimartingale, with decomposition $N' + A'$. Furthermore, Itô's formula gives an explicit expression for N' and A' in terms of f and X! This is best seen through a very simple example: suppose X is a Brownian motion. We've earlier interpreted it as $X_t = \int_0^t \xi_s ds$, where ξ_s represents a background random fluctuation at time s. If we apply the function $f(x) = x^2$ to X, Itô's formula gives

$$X_t^2 = X_0^2 + 2\int_0^t X_s dX_s + t.$$

Here the deterministic process t is the "deviation" mentioned earlier, whereas the middle term is again generalized "noise". By our interpretation of X, we have $X_0 = 0$. Thus if we take expectations, we get

$$\mathbb{E}X_t^2 = t,$$

a familiar result for Brownian motion. We will now have some idea about the behavior of the process X_t^2. It should essentially be "noise" superimposed on the increasing function $t \mapsto t$. Furthermore, in Part II we will see that the "noise" process $\int_0^t X_s dX_s$ is in fact the same Brownian motion X, running on a different (random) time scale. This is essentially Lévy's celebrated characterization of Brownian motion. In Part II we will prove that any martingale with continuous paths is a time-changed Brownian motion.

Some of the sections in the first part are marked with the symbol $\star$. This is to indicate that they are vital to a proper understanding of stochastic integration. The symbol $\star\star$ is reserved for the Itô formula, without which stochastic calculus is at best an exercise in futility. The impatient reader should at least go through these sections, perhaps skipping at first the proofs of the theorems. The other sections can be omitted at your own risk and peril! The first three sections on stochastic processes and stopping times are only intended as a review (or a crash course, depending on the reader...) of the concepts which pervade the rest of this report. They can be quickly scanned and later referred to. Most other sections, especially in Part II, are written so that they can be read more or less independently, although obviously to a lesser extent in Part I.

The origins of the proofs presented are indicated in the statements of the theorems, except in a few rare cases where the proof is so simple that I give my own.

Exercises, as P. A. Meyer once wrote, have been placed throughout the text in the form of errors.

It is a pleasure to thank my supervisor, Phil Pollett, for letting me roam around the literature in complete freedom. When I started the year, I knew nothing about stochastic processes. Now, I like to think that I have a minuscule idea of what is happening. I certainly believe that random processes are now ready to, and should, form an integral part of the paradigms of dynamics. He also deserves thanks for pointing out to me some of the idiosyncrasies of the English language; the present copy of this report undoubtedly still contains a number of obscurities. Thanks go also to Michael Nielsen, who consistently reminded me to make the theory *understandable*, and to many others in the Honours Room for patiently listening to my expositions of the theories I encountered.

It is time to begin. We follow the advice K.D. Elworthy gives in his book [Elw82]:

> "DON'T PANIC"

Hitchhiker's Guide To The Galaxy

Contents

Part I

Stochastic integration

$\star$ Introduction

This chapter is about the concept of *stochastic integral.* The basic problem we will look at is to try and make sense of an integral where both the integrand and integrator are stochastic processes. Motivation for this problem comes from a large variety of sources. An especially simple example is the following, taken from population dynamics: let $N(t)$ denote the size of a population at time t, and $a(t)$ be the relative rate of growth. A simple model for the size of this population is

$$\frac{dN}{dt} = a(t)N(t), \quad N(0) = A,$$

the solution of which is given by

$$\log N(t) - \log A = \int_0^t a(s)ds.$$

What if, however, $a(t)$ is subject to various random environmental effects, so that $a(t) = r(t) + \text{"noise"}(t)$, say? The obvious (formal) answer is to write

$$\log N(t) - \log A = \int_0^t r(s)ds + \int_0^t \text{"noise"}(s)ds.$$

The second integral is an example of a stochastic integral. The way we interpret it determines the properties of the "solution" $N(t)$.

Stochastic processes

Our first task is to make precise the notion of "noise"(t). As this will obviously involve the theory of stochastic processes, we first need to set up a proper framework.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Technically, we will always assume that it is *complete*, i.e. if $B \subseteq A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$, then $B \in \mathcal{F}$ with $\mathbb{P}(B) = 0$. This will simplify the presentation, and it is well known that every probability space can be completed.

Let T be a set, $(E, \mathcal{B})$ be a measurable space. A *stochastic process* is a function $X : T \times \Omega \to E$, such that $X(t, \cdot)$ is $(\mathcal{B}, \mathcal{F})$-measurable for each $t \in T$. Unless otherwise stated, we will take $E = \mathbb{R}$ with $\mathcal{B} = \mathcal{B}(\mathbb{R})$ being the Borel sets. Thus $X(t, \cdot)$ is a real-valued random variable for each $t \in T$. The set T of course represents time, and will be taken as $\mathbb{R}_+ = [0, \infty[$, unless an explicit statement to the contrary is made.

If we fix $\omega \in \Omega$, then the function $X_t : t \mapsto X(t, \omega)$ is called a *sample path* of X. It is often convenient to denote the process X by $(X_t)_{t \in \mathbb{R}_+}$ or simply (X_t).

In the same way that we often identify functions which are equal almost everywhere, we say that two processes X and Y are *modifications* (or *versions*) of each other if $\mathbb{P}(\omega : X_t(\omega) = Y_t(\omega)) = 1$ for all $t \in \mathbb{R}_+$.

A stronger condition is that X and Y be *indistinguishable*: we require that $\mathbb{P}(X_t = Y_t \quad \text{for all } t \in \mathbb{R}_+) = 1$. When two processes are indistinguishable, they have a.s. the same sample paths. If they are versions of each other, their sample paths need not be the same. Indeed, we will often assume a process to have right continuous paths; we then say that the *process* is right continuous. Similarly, an *increasing process* is one which has almost all paths increasing (the word increasing is always used in the weak sense of nondecreasing). In such a case, it will be understood that we have chosen a version with right continuous paths. More generally, most equations in the rest of this chapter involving random variables will hold only a.s., and equations involving processes will only hold up to evanescence (a set $A \subset \mathbb{R}_+ \times \Omega$ is called *evanescent* if $\mathbb{P}(\omega : (t, \omega) \in A) = 0$).

Given a process X with right/left-hand limits, we are also often interested in the following processes:

- The *right/left continuous* version $X_\pm$ defined by $(X_\pm)_t = X_{t\pm}$, with the convention $X_{0-} = X_0$.
- The *jump process* $\Delta X = X_+ - X_-$.
- The *continuous part* $X^c = X - \Delta X$.
- The *maximal process* $X_t^* = \sup_{s \leq t} |X_s|$.
- The *limit* X_∞, defined by $X_\infty(\omega) = \lim_{t \to \infty} X_t(\omega)$ on $\{\omega : \lim X_t(\omega) \text{ exists}\}$, and $X_\infty(\omega) = 0$ otherwise. Note that X_∞ is a well-defined random variable, since both $\limsup X_t$ and $\liminf X_t$ are random variables and hence $\{\omega : \limsup X_t = \liminf X_t\} \in \mathcal{F}$.

Filtrations and stopping times

A *filtration* $(\mathcal{F}_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family of sub-σ-algebras of $\mathcal{F}$, i.e. such that for all s, $t \in \mathbb{R}_+$, the inclusion $\mathcal{F}_s \subseteq \mathcal{F}_t$ holds whenever $s \leq t$. If the following two conditions are also satisfied, we say that $(\mathcal{F}_t)$ is a *standard filtration*:

- The $\mathcal{F}_t$ are *right-continuous*: $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$.
- $\mathcal{F}_0$ is complete: it contains all the $\mathbb{P}$-null sets of $\mathcal{F}$.

It is sometimes useful to write $\mathcal{F}_\infty = \vee_{t\geq 0}\mathcal{F}_t$, the smallest σ-algebra containing the whole filtration. The filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ (also called a *stochastic basis*) is said to satisfy the *usual conditions* if $(\mathcal{F}_t)$ is standard. Henceforth, we shall assume that the usual conditions hold.

We say that a process (X_t) is *adapted* to the filtration $(\mathcal{F}_t)$ if $X_t(\cdot)$ is $\mathcal{F}_t$-measurable for every $t \in \mathbb{R}_+$. Thus the complete history of the process up to time t is contained in the σ-algebra $\mathcal{F}_t$. More precisely, by associating an "observer" with $(\mathcal{F}_t)$, the collection $\mathcal{F}_t$ contains all the questions about X that this observer may ask up to time t. In particular, note that if X is $(\mathcal{F}_t)$-adapted, then X_∞ is $\mathcal{F}_\infty$-measurable.

A random variable $T : \Omega \to \overline{\mathbb{R}}_+$ is called a *stopping time* with respect to $(\mathcal{F}_t)$ if the sets $\{T \leq t\} \in \mathcal{F}_t$ for each $t \in \mathbb{R}_+$. Since $(\mathcal{F}_t)$ is a standard filtration, this is equivalent to requiring $\{T < t\} = \bigcup_{s\in\mathbb{Q}\cap[0,t[}\{T \leq s\} \in \mathcal{F}_t$.

Thus, the filtration must "know" by time t whether T has occurred or not. Associated with every stopping time T is a σ-algebra $\mathcal{F}_T$, consisting of all sets $A \in \mathcal{F}$ such that

$$A \cap \{T \leq t\} \in \mathcal{F}_t \quad \text{for all } t \in \mathbb{R}_+.$$

Thus $\mathcal{F}_T$ consists of all events which may be known to occur up to the time T.

Of course, if T is a stopping time, then $X_T 1_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable. This is seen by noting that the r.v. X_T is for each t the composition of φ : $\{T \leq t\} \to \mathbb{R}_+ \times \Omega$, $\omega \mapsto (T(\omega), \omega)$, which is $(\mathcal{F}_t, \mathcal{B}([0,t]) \otimes \mathcal{F}_t)$-measurable, and $X : \mathbb{R}_+ \times \Omega \to \mathbb{R}$, which is $(\mathcal{B}([0,t]) \otimes \mathcal{F}_t, \mathcal{B}(E))$-measurable. Thus $\{X_T \in B\} \cap \{T \leq t\} \in \mathcal{F}_t$ for each $B \in \mathcal{B}(E)$.

If T is the constant time $T(\omega) = s$ for all ω, then we clearly have $\mathcal{F}_T = \mathcal{F}_s$.

Processes and stopping times

The reason such a random variable is called a stopping time is that it can be used to decide whether a process has been stopped. More precisely, if X is an adapted process, define

$$X_t^{|T}(\omega) = X_{t\wedge T(\omega)}(\omega) = X_t 1_{\{t<T\}} + X_T 1_{\{T\leq t\}},$$

and

$$X_t^{|T-}(\omega) = X_{t\wedge T(\omega)-}(\omega) = X_t 1_{\{t<T\}} + X_{T-} 1_{\{T\leq t\}}.$$

Then $X^{|T}$ is again adapted to $(\mathcal{F}_t)$ as can be checked and it is called the process *stopped at time* T. We call $X^{|T-}$ the process stopped strictly before time T. Note that it is common in stochastic processes to use the notations $x \wedge y = \inf(x, y)$ and $x \vee y = \sup(x, y)$.

An important class of stopping times are the *first hitting times* of a set. For example, let B be an open set and define $T = \inf\{t : X_t \in B\}$. If X happens to have right-continuous sample paths, then writing

$$\{T < t\} = \bigcup_{s\in\mathbb{Q}\cap[0,t[} \{X_s \in B\},$$

and noting that $\{X_s \in B\} \in \mathcal{F}_s$, we see that T is a stopping time. We will often be using this result for the special case

$$T = \inf\{t : |X_t| > n\}.$$

We can also manufacture new stopping times from old ones: If S and T are stopping times, then $S \wedge T$, $S \vee T$, and $S + T$ are stopping times, a result which is customarily left to the reader.

Finally, we need to mention stochastic intervals. Given two stopping times S and T, we can define the *stochastic interval*

$$]S, T] = \{(t, \omega) \in \mathbb{R}_+ \times \Omega : S(\omega) < t \leq T(\omega)\},$$

and similarly $[S, T]$, $[S, T[$, and $]S, T[$. The reason for mentioning this is that by our definition, stochastic intervals are always subsets of $\mathbb{R}_+ \times \Omega$. So when dealing with stochastic intervals, we have, for example, $[0, \infty] = \mathbb{R}_+ \times \Omega$, *not* as we might expect $[0, \infty] = \overline{\mathbb{R}}_+ \times \Omega$.

We say that a stopping time T is *predictable* if there is a sequence of stopping times (T_n) such that $T_n \uparrow T$ and $T_n < T$ on $\{T > 0\}$. For example, the time $T = \infty$ is predictable. But the time τ_1 of the first jump of a $PP(\lambda)$ (see later) is not predictable.

$\star$ Brownian motion

Now that we have the proper language to discuss stochastic processes, let us return to our noise process, $n(t,\omega)$ say. It is now obvious that an integral

$$\int_0^t n(s,\omega)ds$$

should represent some stochastic process. We will see later that it can be identified with a well-known process: the Wiener process, or Brownian motion (BM). Formally, we say that a process $B : \mathbb{R}_+ \times \Omega \to \mathbb{R}^d$ is a *d-dimensional Brownian motion starting at $x \in \mathbb{R}^d$* if

- B is adapted to $(\mathcal{F}_t)$, has continuous sample paths and $B_0 = x$ a.s.
- For every $s < t$, the increment $(B_t - B_s)(\omega)$ is independent of $\mathcal{F}_s$ and has a normal distribution with mean zero and covariance matrix $(t-s)C$ for some fixed C.

To simplify things, we will henceforth assume C is the identity matrix. A process B as above will be referred to by the symbol $BM^x(\mathbb{R}^d)$. The construction of $BM^x(\mathbb{R}^d)$ is given in all standard textbooks on stochastic processes, so we will not repeat it here.

Note however that given an arbitrary stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, a $BM^x(\mathbb{R}^d)$ may not exist. In fact, the "minimal" stochastic basis is given by $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$, the space of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^d$, $\mathcal{F}_t = \sigma(\omega(s) : s \leq t, \quad \omega \in \Omega)$, $\mathcal{F} = \mathcal{F}_\infty$, and $\mathbb{P} = \mathbb{W}$, called the *Wiener measure*, which is the unique measure on $(\Omega, \mathcal{F})$ such that for any $0 = t_0 < t_1 < \ldots < t_k \in \mathbb{R}_+$, $B_0, \ldots, B_k \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{W}(\omega(t_0) \in B_0, \ldots, \omega(t_k) \in B_k) = \int_{B_0} \delta_x(dx_0) \int_{B_1} p(0, x_0, t_1, dx_1) \ldots \int_{B_k} p(t_{k-1}, x_{k-1}, t_k, dx_k),$$

where

$$p(s,x,t,B) = \int_B (2\pi|t-s|)^{-d/2} e^{-|y-x|^2/2|t-s|} dy,$$

and $\delta_x(B) = 1_B(x)$ is the *Dirac measure*. The process B is then defined by

$$B_t(\omega) = \omega(t).$$

This is called the *canonical Brownian motion.*

If we write $B_t(\omega) = \int_0^t n(s,\omega)ds$, we see that we can formally solve stochastic differential equations such as the one in the introduction by having a theory of integrals $\int_0^t H_s(\omega)dB_s(\omega)$ where H and B are processes. In particular, the equation in the introduction makes sense as an integral equation

$$\int_0^t dN_s(\omega) = \int_0^t r(s)N_s(\omega)ds + \int_0^t N_s(\omega)dB_s(\omega),$$

where B is a $BM^0(\mathbb{R})$.

The Poisson process

$BM^x(\mathbb{R}^d)$ is a continuous process. It will be useful, mostly for examples, to consider a process which doesn't have continuous paths. A process N is called a *Poisson process with rate* $\lambda > 0$ relative to $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ if

- It is adapted to $(\mathcal{F}_t)$, has right-continuous paths and $N_0 = 0$ a.s.
- For every $s < t$, the increment $(N_t - N_s)(\omega)$ is independent of $\mathcal{F}_s$ and has a Poisson distribution with mean λ.

Such a process will be referred to by the symbol $PP(\lambda)$. As for the $BM^x(\mathbb{R}^d)$, there exists a canonical representation of $PP(\lambda)$. Choose $\Omega = D(\mathbb{R}_+, \mathbb{Z}_+)$, the set of right-continuous functions from $\mathbb{R}_+$ to $\mathbb{Z}_+$, with left limits at every $t > 0$. Take $\mathcal{F}_t = \sigma(\omega(s) : s \leq t, \omega \in \Omega)$, $\mathcal{F} = \mathcal{F}_\infty$ and $\mathbb{P}$ is generated, as for $BM^x(\mathbb{R}^d)$, by

$$\mathbb{P}(\omega(t_0) \in B_0, \ldots, \omega(t_k) \in B_k) = \int_{B_0} \delta_0(dx_0) \int_{B_1} p(0, x_0, t_1, dx_1) \ldots \int_{B_k} p(t_{k-1}, x_{k-1}, t_k, dx_k),$$

except that now, for any $B \subseteq \mathbb{Z}_+$,

$$p(s,x,t,B) = \sum_{k \in (B-x) \cap \mathbb{Z}_+} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}.$$

The process $N_t(\omega) = \omega(t)$ is the canonical $PP(\lambda)$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$.

⋆ Lebesgue-Stieltjes integrals

Now that we know that we want to integrate processes with respect to processes, how do we go about it? If (H_t) is the *integrand* and (X_t) is the *integrator*, what is the meaning of the expression

$$\int_a^b H_t(\omega) dX_t(\omega) \ ?$$

At first glance, the answer seems simple: for each $\omega \in \Omega$, integrate separately and then combine the results into one random variable. Since $X_t(\omega)$ is a "path" for each ω, the obvious approach is to take a Lebesgue-Stieltjes integral (w.r.t. t) of $H_t(\omega)$ over the path $X_t(\omega)$.

Unfortunately, this doesn't always work, for in Stieltjes integration, the paths $t \mapsto X_t(\omega)$ must be of bounded variation. When this is not the case, the approximating sums

$$\sum_{\Delta_n} H_{\tau_i} \left(X_{t_{i+1}} - X_{t_i} \right),$$

taken over partitions $\Delta_n = \{a = t_0 \leq \tau_0 \leq t_1 \leq \cdots \leq \tau_n \leq t_{n+1} = b\}$ with $|\Delta_n| = \sup_i |t_{i+1} - t_i| \to 0$, do not converge.

Well, the one process we certainly do want to integrate (and use as integrator), the BM, does *not* have paths of bounded variation on *any* interval $[a, b]$. To prove this, we first need the following result, which will also be very useful later:

The quadratic variation of $BM^0(\mathbb{R})$

THEOREM **1.** [Pro92] Let (Δ_n) be a sequence of refining partitions of $[0, t]$ with mesh $|\Delta_n| \to 0$, then $\lim_n \sum_{\Delta_n} (B_{t_{i+1}} - B_{t_i})^2 = t$, where convergence is in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

PROOF:

$$\sum_{\Delta_n} (B_{t_{i+1}} - B_{t_i})^2 - t = \sum_{\Delta_n} \left((B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i) \right) = \sum_{\Delta_n} Y_i,$$

where the Y_i are independent with $\mathbb{E} Y_i = 0$. So

$$\mathbb{E} \left(\sum_{\Delta_n} (B_{t_{i+1}} - B_{t_i})^2 - t \right)^2 = \mathbb{E} \left(\sum_{\Delta_n} Y_i \right)^2 = \sum_{\Delta_n} \mathbb{E} Y_i^2.$$

Next, notice that if we set $Z_i^2 = (B_{t_{i+1}} - B_{t_i})^2/(t_{i+1} - t_i)$, then the Z_i are all $N(0,1)$ variables, so $\mathbb{E}(Z_i^2 - 1)^2 = K$ say, for all i. Then

$$\mathbb{E}\left(\sum_{\Delta_n}(B_{t_{i+1}} - B_{t_i})^2 - t\right)^2 = K\sum_{\Delta_n}(t_{i+1} - t_i)^2 \leq K|\Delta_n|t,$$

and since this tends to zero as $n \to \infty$, we have our result. □

The variation of $BM^0(\mathbb{R})$

THEOREM 2. [Pro92] The paths of Brownian motion are a.s. of infinite variation on any time interval.

PROOF: For any interval $[a,b]$, the variation of BM on $[a,b]$ is

$$V_a^b(\omega) = \sup_{\Delta}\sum_{\Delta}|B_{t_{i+1}}(\omega) - B_{t_i}(\omega)|,$$

where Δ ranges over all finite partitions of $[a,b]$. Our goal is to show

$$\mathbb{P}(V_a^b = \infty \quad \text{for all } a,\, b \in \mathbb{R}_+) = 1.$$

But $\{V_a^b = \infty \quad \text{for all } a,\, b \in \mathbb{R}_+\} = \bigcap_{a,b\in\mathbb{Q}_+}\{V_a^b = \infty\}$, which is a countable intersection. Thus it suffices to show $\mathbb{P}(V_a^b = \infty) = 0$ for $a,\, b \in \mathbb{Q}_+$. By the previous result, there are partitions (Δ_n) such that

$$\begin{aligned} b - a &= \lim_n \sum_{\Delta_n}\left(B_{t_{i+1}} - B_{t_i}\right)^2 \\ &\leq \lim_n \sup_{t_i\in\Delta_n}|B_{t_{i+1}} - B_{t_i}| \cdot V_a^b. \end{aligned}$$

By the a.s. continuity of the sample paths of BM, this implies $b - a \leq 0$ on the set $\{\omega : V_a^b(\omega) < \infty\}$. But we assumed $b > a$, so $\mathbb{P}(V_a^b < \infty) = 0$. □

An obvious corollary of the above is that the paths of $BM^0(\mathbb{R})$ are a.s. nondifferentiable everywhere. By contrast, since the paths of the $PP(\lambda)$ are increasing, they have finite variation on any interval.

$\star$ Naïve stochastic integration is impossible

Why exactly can't we use Stieltjes integrals $\int_a^b h(x)dF(x)$? Why does the assumption of bounded variation on F become so important? Can we bypass it? The proof of the following result points the way. Recall that for continuous integrands on a finite interval, the Lebesgue-Stieltjes integral can be approximated by Riemann-type sums.

THEOREM **3.** [Pro92] Let (Δ_n) be a refining sequence of partitions of $[a, b]$. If the partial sums

$$S_n(h) = \sum_{\Delta_n} h(t_k)\left(F(t_{k+1}) - F(t_k)\right)$$

converge to a limit for every continuous function h, then F is of finite variation on $[a, b]$.

PROOF: Let X be the Banach space $C[a, b]$ of continuous functions with norm $||f||_\infty = \max\{|f(x)| : x \in [a, b]\}$. Let $S_n : X \to \mathbb{R}$ be the linear operator given by

$$S_n(f) = \sum_{\Delta_n} f(t_k)\left(F(t_{k+1}) - F(t_k)\right).$$

It is easy to construct a function $h_{\Delta_n} \in X$ with $||h_{\Delta_n}||_\infty = 1$ and such that

$$h_{\Delta_n}(t_k) = \operatorname{sgn}\left(F(t_{k+1}) - F(t_k)\right).$$

This means that

$$||S_n|| \geq |S_n(h_{\Delta_n})| = \sum_{\Delta_n} |F(t_{k+1}) - F(t_k)|,$$

and hence $\sup_n ||S_n|| \geq$ (total variation of F on $[a, b]$).

But since by assumption $\lim_n S_n(h)$ exists for each $h \in X$, we have that, for each h, $\sup_n |S_n(h)| < \infty$. By the uniform boundedness principle, this means that $\sup_n ||S_n|| < \infty$ which proves the theorem. □

Although this theorem paints a bleak picture, notice that the function h_{Δ_n} which creates the trouble depends on the exact behaviour of $F(t_{k+1})$ for its value at time t_k. It "knows" what F is going to do in the future! The insight of stochastic integration is to ignore any process H such that H_s can predict X_t for $t > s$. Thus H must be adapted to the filtration generated by X. This is essentially what we will do next.

$\star$ Predictable σ-algebras and processes

Our goal is to define an integral $\int_0^t H_s dX_s$ for processes X and H defined on $\mathbb{R}_+ \times \Omega$, or more generally, on a stochastic interval. We start with a few σ-algebras.

Let $\mathcal{P}_0$ be the collection of sets in $\mathbb{R}_+ \times \Omega$ of the form

$$[S_0, 0] \cup]S_1, T_1] \cup \cdots \cup]S_n, T_n],$$

where the S_i and T_i are stopping times. It is easy to see that $\mathcal{P}_0$ is an algebra on $\mathbb{R}_+ \times \Omega$, called the predictable algebra. We naturally let $\mathcal{P} = \sigma(\mathcal{P}_0)$, and call it the predictable σ-algebra. Note that both $\mathcal{P}_0$ and $\mathcal{P}$ are closely related to the filtration on ω.

A function $H : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ is called predictable if it is $\mathcal{P}$-measurable. Note that if S, T are arbitrary stopping times, then

$$[S, T] = \bigcup_n]S + 1/n, T] \in \mathcal{P},$$

and if T is predictable with announcing sequence $T_n \uparrow T$, then

$$[S, T[= \bigcup_n [S, T_n] \in \mathcal{P}.$$

The following theorem is of much practical value:

THEOREM 4. [RY90, CW83] $\mathcal{P}$ is generated by any of the following:

- $\mathcal{P}_0$: sets of the form $[S_0, 0] \cup]S_1, T_1] \cup \cdots \cup]S_n, T_n]$,
- $\mathcal{P}_1$: stochastic intervals $[S, T]$,
- $\mathcal{P}_2$: stochastic intervals $[S, T[$ where T is predictable,
- $\mathcal{P}_3$: sets of the form $\{0\} \times F_0 \cup]s_1, t_1] \times F_1 \cup \cdots \cup]s_n, t_n] \times F_n$ where s_i, $t_i \in \mathbb{R}_+$ and $F_i \in \mathcal{F}_{s_i}$,
- $\mathcal{P}_4$: adapted processes which are continuous on $[0, \infty[$,
- $\mathcal{P}_5$: adapted processes which are left-continuous on $]0, \infty[$.

PROOF: $\mathcal{P} = \sigma(\mathcal{P}_1)$ because $[S,T] \in \mathcal{P}$ and $]S,T] = [0,T]\backslash[0,S]$.

$\mathcal{P} = \sigma(\mathcal{P}_2)$ because $[S,T[\in \mathcal{P}$ and $[S,T] = \bigcup_n [S, T+1/n[$, where the $T+1/n$ are clearly predictable.

$\sigma(\mathcal{P}_0) \subseteq \sigma(\mathcal{P}_3)$ because the random variables defined by $T(\omega) = t1_F(\omega)$ and $S(\omega) = s1_F(\omega) + \infty 1_{\Omega\backslash F}(\omega)$ are stopping times with $]S,T] =]s,t] \times F$, and similarly $S(\omega) = \infty 1_{\Omega\backslash G}(\omega)$ yields $[S,0] = \{0\} \times G$.

$\sigma(\mathcal{P}_3) \subseteq \sigma(\mathcal{P}_0)$ because first $]S,T] =]0,T]$, then $]0,S]$, $]0,T] = \bigcap_n]0, T + 1/n]$, and finally

$$]0, T+1/n] = \bigcup_{k=0}^{\infty}]k/n, (k+1)/n] \times \{T \geq k/n\},$$

where $\{T \geq k/n\} = \{T < k/n\}^c \in \mathcal{F}_{(k/n)}$.

$\sigma(\mathcal{P}_3) \subseteq \sigma(\mathcal{P}_4)$ because it is easy to find continuous functions f_n on $\mathbb{R}_+$ which converge pointwise to $1_{]u,v]}$, so that $1_F f_n \to 1_{]u,v]\times F}$, and similarly for $1_{\{0\}\times G}$.

$\sigma(\mathcal{P}_4) \subseteq \sigma(\mathcal{P}_5)$ is trivial.

$\sigma(\mathcal{P}_5) \subseteq \sigma(\mathcal{P}_3)$ since if X is a left-continuous process, it is the pointwise limit of

$$X_s{}^n(\omega) = X_0(\omega) 1_{\{0\}}(s) + \sum_{k=0}^{\infty} X_{(k/n)}(\omega) 1_{]k/n,(k+1)/n]}(s).$$

□

The term "predictable σ-algebra" comes from the fact that it is generated by left-continuous processes, i.e. processes which can be predicted at time t if they are known at all earlier times $s < t$.

If A is a predictable stochastic interval, we can also define $\mathcal{P}_0 A = \mathcal{P} \cap A$, and then $\mathcal{P}A = \sigma(\mathcal{P}_0 A) = \mathcal{P} \cap A$. For example, if $A = [S,T]$, we have

$$\mathcal{P}_0[S,T] = \{[U_0, S] \cup]U_1, V_1] \cup \cdots \cup]U_n, V_n] : S \leq U_i, V_i \leq T\},$$

as expected.

⋆ Processes generate elementary integrals

Let X be an $(\mathcal{F}_t)$-adapted process. For every stopping time T, X_T is thus an element of $L^0(\Omega, \mathcal{F}_T, \mathbb{P})$, the vector space of ($\mathbb{P}$-equivalences of) real $\mathcal{F}_T$-measurable random variables.

Now consider the additive set function $\mu_X : \mathcal{P}_0 \to L^0(\Omega, \mathcal{F}_\infty, \mathbb{P})$ (respectively $\mu_X : \mathcal{P}_0[S,T] \to L^0(\Omega, \mathcal{F}_T, \mathbb{P})$ etc., we won't keep track of the possibilities) defined by the formulae

$$\mu_X(]U,V]) = X_V - X_U, \qquad \mu_X([U,0]) = 0.$$

This looks suspiciously like a measure, albeit one with values in $L^0(\Omega, \mathcal{F}_\infty, \mathbb{P})$. If f is a $\mathcal{P}_0[S,T]$-simple function, say $f = \sum_i a_i 1_{A_i}$ where the $a_i \in \mathbb{R}$, $A_i \in \mathcal{P}_0[S,T]$, then we can easily write

$$\int f d\mu_X = \sum_i a_i \mu_X(A_i),$$

and as usual this does not depend upon the representation of f. This elementary integral is now a random variable, not a single real number anymore. A little bit of thought and an application of the previous theorem shows that the function f is a $\mathcal{P}$-simple process, that is

$$f(s,\omega) = H_s(\omega) = h_0(\omega) 1_{[T_0,0]}(s,\omega) + \sum_{1 \le i \le n} h_i(\omega) 1_{]T_i,T_{i+1}]}(s,\omega),$$

where each $h_i(\omega)$ is a simple $\mathcal{F}_{T_i}$-measurable random variable (approximate f by sets in $\mathcal{P}_3$). Then for any stopping time T, we clearly have

$$\begin{aligned} \int_0^T H_s dX_s &= \int_{[0,T]} H_s(\omega) d\mu_X(s,\omega) \\ &= \sum_{1 \le i \le n} h_i(\omega) \left(X_{T \wedge T_{i+1}}(\omega) - X_{T \wedge T_i}(\omega) \right). \end{aligned}$$

Thus we can define an $(\mathcal{F}_t)$-adapted process $(H \cdot X)$ on $\mathbb{R}_+ \times \Omega$ by

$$(H \cdot X)_t = \int_0^t H_s dX_s.$$

Now that we've seen where we want to go, let's go back and look at our "vector measure" μ_X. At the very least, we'd like to integrate arbitrary predictable processes, i.e. $\mathcal{P}$-measurable functions $f : \mathbb{R}_+ \times \Omega \to \mathbb{R}$. Naturally, we will require the set function μ_X to be σ-additive. But what kind of process X produces a σ-additive function μ_X? Before we attempt to answer this question, we jump ahead and develop a little integration theory.

Topological matters

Let $(\Xi, \mathcal{A})$ be a measurable space, E a Banach space. With a slight abuse of notation, we will also denote by $\mathcal{A}$ the vector space of $\mathcal{A}$-measurable real-valued functions on Ξ. The $\mathcal{A}$-measurable, bounded real functions form a subspace, denoted $b\mathcal{A}$, of $\mathcal{A}$. Equipped with the sup-norm

$$\|f\|_\infty = \sup_{x \in \Xi} |f(x)|, \quad f \in b\mathcal{A},$$

$b\mathcal{A}$ becomes a Banach space. This norm topologizes uniform convergence.

The spaces E we are mainly interested in are the cases $E = L^p(\Omega, \mathcal{F}, \mathbb{P})$, where $1 \leq p < \infty$. This will allow us to develop a theory of integration for measures $\mu_X : (\Xi, \mathcal{A}) \to L^p(\Omega, \mathcal{F}, \mathbb{P})$ when X is an *L^p-process*, that is, $X_t \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for each t.

As we saw previously, the natural, and most general, space to consider would be $E = L^0(\Omega, \mathcal{F}, \mathbb{P})$, where the topology would be induced by convergence in probability. There are two reasons why we restrict ourselves to $E = L^1(\Omega, \mathcal{F}, \mathbb{P})$ where $p \geq 1$.

One is that when $p = 0$, E is more difficult to handle: it is not normable, and in fact has trivial dual $E' = \{0\}$. The best approach then is to do a Daniell type (vector) integration theory. What we'll do instead is to exploit the duality of our Banach spaces in such a way as to reduce nearly the whole theory of vector measures to that of real measures.

The other reason is that by a probabilistic procedure called localization, we cover all L^0-integrators as "local" L^1-integrators anyway. So we do not loose much by neglecting the case $E = L^0(\Omega, \mathcal{F}, \mathbb{P})$.

So let's assume from now on that E is Banach:

$\star$ Vector measures

An *E-valued measure* on $(\Xi, \mathcal{A})$ is an additive set function $\mu : \mathcal{A} \to E$ such that whenever (A_i) is a sequence of disjoint elements of $\mathcal{A}$, and $A = \bigcup_i A_i$, then

$$\left(\mu(A) - \sum_{i=0}^{n} \mu(A_i) \right) \to 0 \quad \text{in } E \text{ as } n \to \infty.$$

So μ must be σ-additive as usual.

As with a real measure, we can associate with μ its *variation* $|\mu|$, defined for $A \in \mathcal{A}$ by

$$|\mu|(A) = \sup \sum_i \|\mu(A_i)\|,$$

where the supremum is taken over all disjoint partitions (A_i) of A such that $A_i \in \mathcal{A}$. Note that if $E = \mathbb{R}$, $\|\cdot\| = |\cdot|$, we get the usual variation measure $|\mu|$. However, for general E, the variations are usually infinite and much less useful than in the real case.

To remedy this situation, we introduce the *semivariation* $\|\mu\|$ defined for $A \in \mathcal{A}$ by

$$\|\mu\|(A) = \sup \left\| \sum_i a_i \mu(A_i) \right\|,$$

where the supremum is taken over all finite disjoint partitions (A_i) of A with $A_i \in \mathcal{A}$ and real numbers (a_i) with $|a_i| \leq 1$.

As will be seen a little later, the semivariation gives indications about the range in E of the set function μ. For now, we note that obviously for any $A \in \mathcal{A}$,

$$0 \leq \|\mu(A)\| \leq \|\mu\|(A) \leq \|\mu\|(\Xi).$$

⋆ Vector integration

Now let E' be the topological dual of E. For a continuous linear functional $x' \in E'$, we will write $\langle x', x \rangle$ for its value at a point $x \in E$.

When μ is an E-valued vector measure, the set functions $\langle x', \mu \rangle$, defined for $A \in \mathcal{A}$ by

$$\langle x', \mu \rangle(A) = \langle x', \mu(A) \rangle,$$

are certainly bounded, real-valued measures, for every $x' \in E'$.

We say that an $\mathcal{A}$-measurable $f : \Xi \to \mathbb{R}$ is μ-integrable if it is integrable with respect to every measure $\langle x', \mu \rangle$, $x' \in E'$ and if for every $A \in \mathcal{A}$, there exists an element $x_A \in E$ such that for all $x' \in E'$,

$$\langle x', x_A \rangle = \int_A f d\langle x', \mu \rangle.$$

For each A, the element x_A is clearly unique when it exists. When f is μ-integrable, we write $\int_A f d\mu = x_A$. We also write as usual

$$\int f d\mu = \int_\Xi f d\mu.$$

When $f = \sum_i a_i 1_{A_i}$ is an $\mathcal{A}$-simple function, it is easy to check that for all $A \in \mathcal{A}$,

$$\int_A f d\mu = \sum_i a_i \mu(A_i \cap A).$$

The vector integral is clearly linear in both the integrand and the measure. Thus the collection of integrable functions forms a vector space, denoted $\mathcal{L}^1(\mu)$. As usual, if we consider equivalence classes of μ-a.e. equal functions, we call the vector space $L^1(\mu)$. But hereafter, we will indulge in the usual confusion between $\mathcal{L}^1(\mu)$ and $L^1(\mu)$.

When $f \in L^1(\mu)$, it induces a new vector measure $f \cdot \mu$, defined for $A \in \mathcal{A}$ by

$$(f \cdot \mu)(A) = \int_A f d\mu.$$

The natural topology on $L^1(\mu)$ is given by

$$\|f\| = \|f \cdot \mu\|(\Xi)$$

which makes $L^1(\mu)$ into a complete topological vector space, as will become clear shortly.

In fact, vector measures again form a vector space, written $\mathrm{Meas}(\Xi, \mathcal{A}; E)$ with topology induced by the semivariations $\|\mu\|(\Xi)$.

Familiar examples

This definition of vector measure and integral via duality is perhaps not the most obvious one. But suppose that we had taken the alternative definition that f be integrable whenever there is a sequence of step functions (f_n) converging μ-a.e. to f and such that the sequence of integrals $(\int_A f_n d\mu)$ is convergent in E for each $A \in \mathcal{A}$. Then such an f would also satisfy our duality definition of integrablility.

Now suppose $E = \mathbb{R}$, so that μ is a signed measure. Then $E' = \mathbb{R}$ also and a function f is μ-integrable (in the vector sense) if and only if for all $a \in \mathbb{R}$,

$$\left| a \int f d\mu \right| < \infty,$$

so that f is integrable in the usual sense.

Note that the measure μ does not take infinite values. This is a restriction in the real-valued case, but then in a general Banach space there is no "element ∞". Thus we expect our integration theory to be similar to that of finite, signed, real-valued measures.

When $E = \mathbb{C}$, we can get a larger class of integrable functions by admitting complex-valued functions. However, this relies on being able to multiply two "vectors" of the space $\mathbb{C}$. When E is a general Banach space, there is no "natural" multiplication of vectors, which means that the natural integrable functions must be real-valued. We will come back to this problem when we discuss integration on manifolds.

The range of a vector measure

Semivariations are useful if they turn out to be finite, that is why they are introduced in the first place. We noted earlier that $\|\mu\|(\Xi) \geq \|\mu\|(A) \geq |\mu(A)|_\alpha$ for all $A \in \mathcal{A}$. So if $\|\mu\|(\Xi) < \infty$, this means that the whole *range of* μ, the set $\{\mu(A) : A \in \mathcal{A}\}$, is bounded in E. So μ is really a "finite" measure.

Conversely, if μ has bounded range in E, then it must have finite total semivariation: for $x \in E$, recall that there is $x' \in E'$ with $\|x'\| = 1$ and $\langle x', x\rangle = \|x\|$. So this means

$$\begin{aligned}\|\mu\|(\Xi) &= \sup\{|\langle x', \mu\rangle(\Xi)| : \|x'\| \leq 1\} \\ &\leq \sup\{2 \sup_{A\in\mathcal{A}} |\langle x', \mu\rangle(A)| : \|x'\| \leq 1\} \\ &\leq 2 \sup_{A\in\mathcal{A}} |\mu(A)| < \infty.\end{aligned}$$

The rest of our development of vector integration will depend on the boundedness of the range of μ. Luckily, any measure with values in a Banach space has bounded range. This is seen as follows:

Let $\mathcal{T}$ be the collection of all finite disjoint partitions (A_i) of Ξ with $A_i \in \mathcal{A}$ together with scalars (a_i) such that $|a_i| \leq 1$. Then for each $x' \in E$,

$$\sup_{\mathcal{T}} |\langle x', \sum_i a_i\mu(A_i)\rangle| \leq |\langle x', \mu\rangle|(A) < \infty.$$

Each element $\sum_i a_i\mu(A_i)$ may be considered as an operator on E', so by the uniform boundedness principle,

$$\sup_{\mathcal{T}} \|\sum_i a_i\mu(A_i)\| = \|\mu\|(\Xi) < \infty.$$

The boundedness of our vector measures means that their integration theory is closely related to that of finite measures. Here is a first result:

Bounded functions are integrable

THEOREM 5. [KK76] Every bounded $\mathcal{A}$-measurable function f is μ-integrable, and

$$\left\| \int_A f d\mu \right\| \leq \|f\|_\infty \cdot \|\mu\|(A).$$

PROOF: If f is a simple function, the inequality is obvious from the definition of the semivariations of μ. In general, take a sequence (f_k) of simple functions such that $\|f_k - f\|_\infty < 1/k$. Then

$$\left\| \int_A f_k d\mu - \int_A f_l d\mu \right\| \leq (1/k + 1/l) \, \|\mu\|(A) < \epsilon,$$

provided K and l are sufficiently large. Thus $(\int_A f_k d\mu)$ is a Cauchy sequence in E an hence must converge to some element x_A. Since f is moreover clearly $\langle x', \mu\rangle$-integrable for each $x' \in E'$, we see that $f \in L^1(\mu)$, and

$$\int_A f d\mu = \lim_k \int_A f_k d\mu.$$

Finally, we have

$$\begin{aligned} \left\| \int f d\mu \right\| &\leq \liminf_k \left\| \int f_k d\mu \right\| \\ &\leq \liminf_k \|f\|_\infty \cdot \|\mu\|(A) \\ &\leq \|f\|_\infty \cdot \|\mu\|(A). \end{aligned}$$

□

Dominated functions are integrable

THEOREM 6. [KK76] Suppose f and g are $\mathcal{A}$-measurable functions with $|f| \leq g$ on Ξ. If g is μ-integrable, then so is f.

PROOF: Define $h(t) = f(t)/g(t)$ whenever $g(t) \neq 0$, and $h(t) = 0$ otherwise. Then by the previous result, h is $(g \cdot \mu)$-integrable. But from real measure theory, f is $\langle x', \mu\rangle$-integrable for every $x' \in E'$, since g is. Now if $x_A = \int_A hd(g \cdot \mu)$, then

$$\langle x', x_A\rangle = \int_A hd\langle x', g \cdot \mu\rangle = \int_A hgd\langle x', \mu\rangle = \int_A fd\langle x', \mu\rangle,$$

and hence f is integrable. □

Note that this result implies that $f \in L^1(\mu) \Leftrightarrow |f| \in L^1(\mu)$.

$\star$ The dominated convergence theorem

Finally, we can now easily prove the DCT:

THEOREM 7. Let f_n and g be in $L^1(\mu)$ with $|f_n| \leq g$ and $f_n \to f$ μ-a.e., then $f \in L^1(\mu)$ and

$$\lim_n \int f_n d\mu = \int f d\mu.$$

PROOF: For any $x' \in E'$,

$$\lim_n \langle x', \int f_n d\mu \rangle = \lim_n \int f_n d\langle x', \mu \rangle = \int f d\langle x', \mu \rangle = \langle x', \int f d\mu \rangle.$$

Note that the last equality holds only because f is μ-integrable, which follows from the previous result. □

On integrability

It is easy to rederive a lot of results from real measure theory for the vector integral. However, it is crucial to remember that it is not enough to show that a function is $\langle x', \mu \rangle$-integrable for every $x' \in E'$ for it to be μ-integrable.

Indeed, suppose $\Xi = \mathbb{N}$, $\mathcal{A} = 2^{\mathbb{N}}$, $E = c_0$, the space of sequences of real numbers converging to zero. If $\varphi(t) = 1/t$ and $\mu(A)(t) = \varphi(t) 1_A(t)$ for every $A \in \mathcal{A}$, then μ is a vector measure. Now $E' = l^1$, the space of sequences $x' = (x_i)$ with $\sum_i x_i < \infty$. The function $f(t) = t$ is $\langle x', \mu \rangle$-integrable for all $x' \in E'$, with integral

$$\int_A f d\langle x', \mu \rangle = \sum_{i \in A} x_i < \infty,$$

but f is *not* μ-integrable, for

$$\int f d\mu = (1, 1, 1, \ldots),$$

the unit sequence which is not an element of c_0.

$\star$ The stochastic integral

Let's now again consider our process X with the additive set function on $\mathcal{P}_0$

$$\mu_X(]S,T]) = X_T - X_S, \qquad \mu_X([S,0]) = 0.$$

To apply the theory we just developed, we first consider L^1-processes.

If μ_X extends to a vector measure (with values in $L^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$) on $\mathcal{P}$, will say that X is *summable*. Such processes obviously form a vector space.

If the extension μ_X exists, it must be unique. To see this, suppose $\overline{\mu_X}$ is another extension. Then μ_X and $\overline{\mu_X}$ agree on the algebra $\mathcal{P}_0$. Thus for each $x' \in E'$, the bounded signed measures $\langle x', \mu_X\rangle$ and $\langle x', \overline{\mu_X}\rangle$ agree on $\mathcal{P}_0$. By the Caratheodory extension theorem, they must agree on $\mathcal{P}$, which means that μ_X and $\overline{\mu_X}$ must also agree on $\mathcal{P}$.

The same argument works also for $\mathcal{P}_0[S,T]$ and $\mathcal{P}_0]S,T]$. But $\mathcal{P}_0[S,T[$ and $\mathcal{P}_0]S,T[$ are not algebras, only rings. So there could be a slight problem with uniqueness here. But since T is assumed to be predictable in these cases, we have $[S,T[= \bigcup_n [S,T_n]$ and each $\mathcal{P}_0[S,T_n]$ is an algebra. Thus for any $A \in \mathcal{P}_0[S,T[$,

$$\langle x', \mu_X\rangle(A \cap [S,T_n]) = \langle x', \overline{\mu_X}\rangle(A \cap [S,T_n]),$$

and by dominated convergence,$\langle x', \mu_X\rangle(A) = \langle x', \overline{\mu_X}\rangle(A)$. Thus the extension μ_X is again unique on $\mathcal{P}[S,T[$ and $\mathcal{P}]S,T[$.

Here is the vector version of the Carathéodory extension theorem:

THEOREM 8. [Yor78, DU77] An additive set-function μ, defined on an algebra $\mathcal{A}_0$, with values in a weakly sequentially complete Banach space E, has an extension to a measure on $\mathcal{A} = \sigma(\mathcal{A}_0)$ if and only if for every $x' \in E'$, the real function $\langle x', \mu\rangle : \mathcal{A}_0 \to \mathbb{R}$ is σ-additive and bounded. The extension is unique if it exists.

We will not prove this theorem, though we won't really be using it either. Because of the integration theory we developed, if we simply check that $\langle x', \mu\rangle$ are measures on $\mathcal{A}$ and $||\mu||(\Xi) < \infty$, we can then integrate bounded functions, which is what we will be interested in doing most of the time. If we were to apply the theorem, then checking these same conditions would show that μ is actually a vector measure.

When X is summable, the process

$$(H \cdot X)_t = \int_0^t H_s dX_s = \int_{[0,t]} H_s(\omega) d\mu_X(s,\omega)$$

makes sense for any $\mathcal{P}$-measurable (predictable) H which is μ_X-integrable on every stochastic interval $\Xi = [0, t]$. Such an H will be called X*-integrable*, and this will be written $H \in L^1(X) = L^1(\mu_X)$.

Each of the random variables $(H \cdot X)_T$, T a stopping time, is almost surely unique. This comes from the fact that $E' = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, the space of bounded random variables, for if $G \in \mathcal{F}$, then $1_G = x' \in E'$, and

$$\mathbb{E}1_G(H \cdot X)_T = \langle x', \int_{[0,T]} H d\mu_X \rangle = \int_{[0,T]} H d\langle x', \mu_X \rangle,$$

so that two random variables which cannot be distinguished by E' also cannot be distinguished by $\mathbb{P}$, and conversely.

Note that the process $(H \cdot X)$ is automatically $(\mathcal{F}_t)$-adapted. Since for each t, $(H \cdot X)_t \in L^1(\Omega, \mathcal{F}_t, \mathbb{P})$, the whole process $(H \cdot X)$ is only defined up to a modification.

Note also that a sufficient condition for $H \in L^1(X)$ *for all summable* X is that the process H be bounded on every stochastic interval $[0, t]$.

The process $(H \cdot X)$ is the probabilistic equivalent of an "indefinite integral". We will now see that it always has a modification which is right-continuous with left limits. This modification is clearly unique up to indistinguishability (compare two processes on the rationals). It is called the *stochastic integral*.

⋆ Càdlàg processes

Suppose X is summable and $H \in L^1(X)$. Then $(H \cdot X)$ has a right continuous version, for if $a, b \in \mathbb{R}_+$ with $a < b$, then

$$(H \cdot X)_b - (H \cdot X)_{a+} = \lim_n \int_{]a+1/n, b]} H dX = \lim_n \int_0^b H^n dX,$$

where $H^n = H 1_{]a+1/n, b]}$. By applying the vector DCT, we conclude that

$$(H \cdot X)_b - (H \cdot X)_{a+} = \int_a^b H dX = (H \cdot X)_b - (H \cdot X)_a,$$

where the equality is on a set of probability 1 (which may depend on b). By letting b run through $\mathbb{Q}_+$, we can conclude that

$$\mathbb{P}\left((H \cdot X)_{a+} = (H \cdot X)_a\right) = 1 \quad \text{for all } a \in \mathbb{R}_+.$$

By an entirely similar reasoning, the process $(H \cdot X)$ must have left limits, though it won't be left continuous in general (since $H^n = H1_{]a,b-1/n]}$ converges to $H1_{]a,b[}$, the vector DCT implies that $(H \cdot X)_{b-1/n} - (H \cdot X)_a$ must converge to something. Call it $(H \cdot X)_{b-} - (H \cdot X)_a$).

If we choose $H_t(\omega) = 1$ in the above, we see that X itself must have a version which is right continuous with left limits.

Note that if the integrator X has left limits, then the process (X_{t-}) is predictable, and so potentially an *integrand*. From now on, we will assume our integrators to be processes with a version which is right continuous with left-hand limits, called *càdlàg* (the acronym is from the French "continu à droite avec limites à gauche", pronounced "kohntynew ah droaht ahvehk limit ah gohsh", which is a considerable improvement over the English r.c.l.l.).

On a similar note, a *càglàd process* is one for which the sample paths are left continuous with right limits ("continu à gauche avec limites à droite", "kohntynew ah gohsh ahvehk limit ah droaht", l.c.r.l.)

The sample paths of càdlàg processes have some nice properties. For example, let $f : \mathbb{R} \to \mathbb{R}$ be a càdlàg function. Then f is bounded on compact intervals. To see this, let $K \subset \mathbb{R}$ be compact. For $x \in K$, both $f(x+)$ and $f(x-)$ exist and are finite. Thus f is bounded on a small neighborhood of x. Since K is compact, it can be covered by a finite number of such neighborhoods, showing that f must be bounded on all of K.

Another useful property is the following: for any $\epsilon > 0$, the function f has only finitely many jumps of size greater than ϵ on any compact interval. The proof is in the same spirit as that of the previous property. This also means that f has only countably many jumps of any size on the whole of $\mathbb{R}$.

The smallest σ-algebra which makes all càdlàg processes on $[0, \infty[$ measurable is called the *optional* σ-algebra, written $\mathcal{O}$. Since continuous processes are càdlàg we have $\mathcal{P} \subset \mathcal{O}$. The reverse inclusion however holds only when all "martingales" (see later, and also [RY90]) with respect to $(\mathcal{F}_t)$ are continuous.

Before we consider the question of which càdlàg processes are locally summable, we will look at some simple properties of integrators in general.

$\star$ The module associativity property

Recall that if X induces the measure μ_X, then for any $f \in L^1(\mu_X)$, $f \cdot \mu_X$ is again a measure. So if K is X-integrable, the process $(K \cdot H)$ is again an integrator. Now suppose H and K are processes such that HK, $K \in L^1(X)$

and $H \in L^1(K \cdot X)$. Then

$$H \cdot (K \cdot X) = (HK) \cdot X$$

In particular, this is true if both H and K are bounded on each $[0,t]$. The proof is simple: for any $G \in \mathcal{F}_t$, $1_G = x' \in E'$. Then

$$\begin{aligned}
\mathbb{E}1_G\left(\int Hd\left(\int Kd\mu_X\right)\right) &= \langle x', \int Hd\left(\int Kd\mu_X\right)\rangle \\
&= \int Hd\langle x', \int Kd\mu_X\rangle \\
&= \int Hd\left(\int Kd\langle x', \mu_X\rangle\right) \\
&= \int HKd\langle x', \mu_X\rangle \\
&= \langle x', \int HKd\mu_X\rangle = \mathbb{E}1_G\left(\int HKd\mu_X\rangle\right),
\end{aligned}$$

which shows that $H \cdot (K \cdot X)_t = (HK \cdot X)_t$ almost surely.

Changing the stochastic basis

There are many things to say about this very important topic. We just note two simple results.

(i) Suppose $(\mathcal{G}_t)$ is a subfiltration of $(\mathcal{F}_t)$, and that X is $(\mathcal{G}_t)$-adapted. If X is an $(\mathcal{F}_t)$-integrator, then it is also a $(\mathcal{G}_t)$-integrator. This is obvious, for we are simply shrinking the predictable σ-algebra $\mathcal{P}[S,T]$, and thus making it "easier" for a process to be summable. In particular, this means that the filtration generated by X is the smallest for which X is an integrator.

(ii) Suppose $\mathbb{Q}$ is another probability on $(\Omega, \mathcal{F})$, absolutely continuous with respect to $\mathbb{P}$ and such that the Radon-Nikodym derivative $d\mathbb{Q}/d\mathbb{P} < K < \infty$. Then if X is summable under $\mathbb{P}$, it is also summable under $\mathbb{Q}$. This is very easy to see as follows: for any $y' \in E' = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, and any $A \in \mathcal{P}[S,T]$,

$$\langle y', \mu_X\rangle_{\mathbb{Q}}(A) = \int y' \frac{d\mathbb{Q}}{d\mathbb{P}} \mu_X(A) d\mathbb{P} = \langle x', \mu_X\rangle_{\mathbb{P}}(A),$$

and

$$\|\mu_X\|_{\mathbb{Q}}(\Xi) = \sup \int \frac{d\mathbb{Q}}{d\mathbb{P}} \left| \sum_i a_i \mu_X(A_i) \right| d\mathbb{P} < K\|\mu_X\|_{\mathbb{P}}(\Xi) < \infty,$$

where $x' = y' d\mathbb{Q}/d\mathbb{P} \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, and $\Xi = [S, T]$.

⋆ Stability under stopping

Suppose X is an integrator, and let μ_X be the vector measure induced on $\mathcal{P}$. If T is a stopping time, then for any U, $V \leq T$,

$$\mu_X(]U, V]) = X_V - X_U = X_{V\wedge T} - X_{U\wedge T} = \mu_{X^{|T}}(]U, V])$$

and of course $\mu_X([U, 0]) = 0 = \mu_{X^{|T}}([U, 0]$. Thus μ_X and $\mu_{X^{|T}}$ agree on $[0, T]$, and also on $[0, t \wedge T]$ for all t. Hence we have almost surely

$$(H \cdot X)_t{}^{|T} = \int_{[0,T\wedge t]} H d\mu_X = (1_{[0,T]} H \cdot X)_t = \int_0^t H d\mu_{X^{|T}} = (H \cdot X^{|T})_t.$$

The importance of this little result should not be underestimated. It allows, as we will see next, an extension of the concept of stochastic integral beyond summable and integrable processes.

⋆ Localization

A *localizing sequence* is an increasing sequence (T_n) of stopping times with $T_n \uparrow \infty$ a.s. Such sequences are useful to weaken various properties of processes.

We say that a process X has property P *locally* (resp. *prelocally*) if for some localizing sequence (T_n), $X^{|T_n}$ (resp. $X^{|T_n-}$) has property P for each n.

Thus X is *prelocally summable* if $X^{|T_n-}$ is summable for each n; H is *locally bounded* if $H^{|T_n}$ is bounded for each n; H is *locally (resp. prelocally) X-integrable* if $H^{|T_n}$ (resp. H^{T_n-} is X-integrable). This will be written $H \in L^1_{\text{loc}}(X)$ (resp. $H \in L^1_{\text{ploc}}(X)$). Note that if X is locally summable/integrable/bounded, then it is also prelocally summable/integrable/bounded.

$\star$ Localizing the stochastic integral

The point of the above definitions is that we can define a "(pre)localized" stochastic integral of (pre)locally integrable processes with respect to (pre)locally summable processes:

Consider first the case $H \in L^1_{\mathrm{loc}}(X)$ where X is summable. There are $T_n \uparrow \infty$ with $H = H^{|T_n}$ on $[0, T_n]$. Since $H^{|T_n} \in L^1(X)$, $(H 1_{[0,T_n]} \cdot X)_S$ is a.s. uniquely defined for any $S \leq T_n$. Call this $(H \cdot X)_S$, and since $T_n \uparrow \infty$, this defines a.s. $(H \cdot X)_S$ for any stopping time S. The definition does not depend on the sequence (T_n), for if (R_n) is another localizing sequence, $R_n \wedge T_n \uparrow \infty$ and for any $S \leq R_n \wedge T_n$,

$$(H 1_{[0,R_n]} \cdot X)_S = (H 1_{[0,R_n \wedge T_n]} \cdot X)_S = (H 1_{[0,T_n]} \cdot X)_S.$$

If X is locally summable with localizing sequence (T_n) and $H \in L^1_{\mathrm{loc}}(X^{|T_n})$, define similarly $(H \cdot X)_S = (X \cdot X^{|T_n})_S$ for any $S \leq T_n$, and again this a.s. yields $(H \cdot X)_S$ for any stopping time.

Now suppose X is only prelocally summable (resp. $H \in L^1_{\mathrm{ploc}}(X)$ etc.), then X agrees with $X^{|T_n -}$ (resp. $H = H^{|T_n -}$) on $[0, T_n[$, so that provided $S < T_n$ on $\{T_n > 0\}$, $(H \cdot X)_S = (H^{|T_n -} \cdot X^{|T_n -})_S$ is uniquely defined. Letting $n \to \infty$, this a.s. provides $(H \cdot X)_S$ for any $S < \infty$.

So we see that in any case, random variables $(H \cdot X)_t$ exist a.s. for any $t \in \mathbb{R}_+$, and define a process, which we again call the *stochastic integral*. Henceforth, we'll abuse notation and write $L^1_{\mathrm{loc}}(X)$, $L^1_{\mathrm{ploc}}(X)$ for the X-integrable processes, even if X is only (pre)locally summable.

Clearly this new stochastic integral has (pre)locally the same properties as the previous integral. In particular, $(H \cdot X)$ is again (pre)locally summable when X is.

When H is locally bounded, we get a bonus: $H \in L^1_{\mathrm{loc}}(X)$ for *all* locally summable processes X. Furthermore, if H is prelocally bounded, it also is locally bounded, because it is the limit of continuous processes H^n, which are locally bounded if and only if they are prelocally bounded.

Similarly, if H is càglàd then it is locally integrable (w.r.t anything). To prove this, we show it must be locally bounded. Let $T_n = \inf\{t : |H_t| > n\}$. Then $T_n \uparrow \infty$ a.s. since each sample path is bounded on the compact time interval $[0, n]$, and $H^{|T_n}$ is bounded.

Finally, note that any locally summable process is prelocally summable, and $(H \cdot X)_S$ is the same in both cases, when $S < \infty$.

The stochastic dominated convergence theorem (SDCT)

Now that we have an integral $(H \cdot X)$ for quite general processes H and X, we would like a specifically stochastic DCT. When Z is a process, recall that we write $Z_t^* = \sup_{s \le t} |Z_s|$ to denote its *maximal process.*

THEOREM **9.** Let X be prelocally summable and $H^n \to H$ (a.s.) be a sequence of processes with $|H^n| \le K$ for some $K \in L^1_{\text{loc}}(X)$ (resp. $L^1_{\text{ploc}}(X)$). Then $H \in L^1_{\text{loc}}(X)$ (resp. $L^1_{\text{ploc}}(X)$) and $(H^n \cdot X)$ converges to $(H \cdot X)$ uniformly on compact time intervals in probability (in short: ucp convergence).

PROOF: By (pre)localization, we can assume $K \in L^1(X)$ and X is summable. To apply the vector DCT (which requires convergence in $\langle x', \mu_X \rangle$-measure for *each* x')), note that $H^n \to H$ $\mathbb{P}$-a.s. implies that with probability 1, $H^n \to H$ $\langle x', \mu_X \rangle$-a.e. This follows because whenever $B \in \mathcal{P}$ and $\mathbb{P}(\omega : (t, \omega) \in B \text{ some } t) = 0$, then $\mathbb{P}(\langle x', \mu_X \rangle (B) = 0) = 1$ as is easily checked on members of $\mathcal{P}_0$. So we can use the vector DCT.

It remains to show convergence in ucp, i.e. for fixed c, ϵ and t,

$$\mathbb{P}\left((H^n \cdot X - H \cdot X)_t^* > c\right) < \epsilon$$

for large enough n.

Let (T_k) and (S_j) be (pre)localizing sequences for X and K respectively. Then

$$\begin{aligned} &\mathbb{P}\left((H^n \cdot X - H \cdot X)_t^* > c\right) \\ &\qquad \le \; \mathbb{P}\left((H^n \cdot X - H \cdot X)_t^* > c | T_k \wedge S_j > t\right) + \mathbb{P}(T_k \wedge S_j \le t) \\ &\qquad = \; \mathbb{P}(Z_t^* > c) + \mathbb{P}(T_k \wedge S_j \le t), \end{aligned}$$

where $Z = (H^n - H)^{|S_j-} \cdot X^{|T_k-})$ on $[0, S_j \wedge T_k[$.

Let $R = \inf\{s : |Z_s| > c\}$. Then $\{Z_t^* > c\} \subseteq \{R \le t\}$. So $c1_{\{Z_t^* > c\}} \le |Z_R| 1_{\{R \le t\}}$, and

$$\begin{aligned} c\mathbb{P}(Z_t^* > c) &\le \mathbb{E}[|Z_R||R \le t]\, \mathbb{P}(R \le t) \\ &\le ||\mu_Z||([0,t] \times \Omega) \\ &\le \sup_{s \le t} ||(H_s^n - H_s)^{|S_j-}||_\infty \cdot ||\mu_{X^{T_k-}}||([0,t] \times \Omega). \end{aligned}$$

Thus if n, k, j are large enough,

$$\mathbb{P}(Z_t^* > c) + \mathbb{P}(T_k \wedge S_j \le t) < \epsilon/2 + \epsilon/2 = \epsilon.$$

□

$\star$ Finite variation processes are integrators

Let A be a càdlàg adapted process with $A_0 = 0$. If the sample paths $t \mapsto A_t(\omega)$ are a.s. of bounded variation on every compact interval of $\mathbb{R}_+$, we call A a *finite variation (FV) process.* For such a process, the expression

$$\int_0^t dA_s(\omega)$$

makes sense as an ω-by-ω Lebesgue-Stieltjes integral, for almost every ω.

The variation of A,

$$\int_0^t |dA_s| = \sup_n \sum_{k=0}^{\infty} |A_{t\wedge(k+1)2^{-n}} - A_{t\wedge k2^{-n}}|,$$

also makes sense a.s.

When A is an FV process, we say that it is an *integrable variation* (IV) process if

$$\mathbb{E}\int_0^\infty |dA_s| < \infty.$$

Our first few integrators are easy to find: any càdlàg IV process A is summable. To see this, just define for any $B \in \mathcal{P}$

$$\mu_A(B) = \int_0^\infty 1_B(s,\omega)dA_s(\omega) \quad \text{(Stieltjes } \omega\text{-by-}\omega).$$

Then we clearly have a.s.

$$\begin{aligned} \mu_A(]S,T]) &= \int_{S(\omega)}^{T(\omega)} dA_s(\omega) = A_T - A_S, \\ \mu_A([S,0]) &= 0. \end{aligned}$$

Also, when (B_i) are disjoint predictable sets with $B = \bigcup_i B_i$, then

$$\begin{aligned} \lim_k \mathbb{E}|\mu_A(B) - \sum_{i=0}^k \mu_A(B_i)| &= \lim_k \mathbb{E}\left|\int_0^\infty (1_B - \sum_{i=0}^k 1_{B_i})dA_s\right| \\ &\leq \mathbb{E}\lim_k |1_B - \sum_{i=0}^k 1_{B_i}| \cdot \int_0^\infty |dA_s| \\ &= 0, \end{aligned}$$

where the DCT applies, since the functions $f_k = |1_B - \sum_{i=0}^k 1_{B_i}| \cdot \int_0^\infty |dA_s|$ are dominated by $\int_0^\infty |dA_s|$, which is integrable by definition.

Note that this argument works for *any* $B \in \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}$, not just predictable ones.

Now let A be any FV process. Setting $T_n = \inf\{t : \int_0^t |dA_s| > n\}$, we see that (T_n) is a prelocalizing sequence and $A^{|T_n-}$ is an IV process for each n. Thus A is an integrator. Note that if A is a predictable FV process, then it is even locally summable.

If N is a $PP(\lambda)$ its sample paths are a.s. increasing. So it is an FV process and

$$\int_0^t H_s dN_s = \sum_{s \le t} H_s \Delta N_s = \sum_{k=1}^{N_t} H_{\tau_k},$$

where $\tau_k = \inf\{t : N_t = k\}$ is the time of the k-th jump.

Stochastic integrals reduce to Stieltjes integrals

Although we have used the Lebesgue-Stieltjes integral to show that an FV process is prelocally summable, we still have an *a priori* choice in interpreting the integral

$$\int_0^t H_s dA_s$$

path-by-path as a Stieltjes integral or as a stochastic integral. That both interpretations give the same result when H is locally bounded is easily seen by the SDCT: When H is a simple process, both integrals clearly give the same result. In the general case, we stop H so that it remains bounded. Taking simple processes $H^n \to H$ and using SDCT then shows both integrals still give a.s. the same result.

Since conversely we know that X must be a.s FV when the stochastic integral reduces to a Stieltjes integral, this means that any other integrators we find will have a "new" integration theory.

$\star$ Martingales

Another important class of processes (not only) in stochastic integration are martingales. We say that an adapted process M is a *martingale* with respect to the filtration $(\mathcal{F}_t)$ and probability measure $\mathbb{P}$ if the following two conditions are satisfied:

1. For all t, $X_t \in L^1(\Omega, \mathcal{F}, \mathbb{P})$,

2. For all $s < t$, $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$.

If we replace $L^1(\Omega, \mathcal{F}, \mathbb{P})$ with $L^p(\Omega, \mathcal{F}, \mathbb{P})$ in the first condition, we end up with *L^p-martingales*. The second condition represents the behaviour of the martingale: since $\mathcal{F}_s$ contains the complete past history of the process X up to time s, this condition means that the best possible predictions of the future evolution of the process involve no discernible trend whatsoever. The process consists of purely random fluctuations (though we must be careful: the fluctuation patterns are often quite different from what we intuitively expect. See Feller's famous chapter on cointossing in [Fel68]). Notice however that the martingale property is not intrinsic to the process. It depends crucially on the filtration and the probability measure. Changing them ("the observer") may result in the loss of the property.

Simple examples of martingales in discrete time $\mathbb{Z}_+$ are sequences of (integrable) independent random variables $(X_n)_{n\in\mathbb{Z}_+}$, and also their partial sums $S_n = \sum_i X_i$. The filtration $(\mathcal{F}_n)$ is then defined by $\mathcal{F}_n = \sigma(X_n, X_{n-1}, \cdots, X_0)$, the smallest σ-algebra making the X_i measurable.

In continuous time, the most important example for us is the $BM^x(\mathbb{R})$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$:

$$\begin{aligned}\mathbb{E}[B_t|\mathcal{F}_s] &= \mathbb{E}[B_t - B_s|\mathcal{F}_s] + \mathbb{E}[B_s|\mathcal{F}_s] \\ &= \mathbb{E}(B_t - B_s) + B_s \\ &= B_s.\end{aligned}$$

If N is a $PP(\lambda)$, the process $(N_t - \lambda t)$ is a martingale, called the *Poisson martingale* on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$:

$$\begin{aligned}\mathbb{E}[N_t - \lambda t|\mathcal{F}_s] &= \mathbb{E}[N_t - N_s|\mathcal{F}_s] + \mathbb{E}[N_s|\mathcal{F}_s] - \lambda t \\ &= \lambda(t-s) + N_s - \lambda t \\ &= N_s - \lambda s.\end{aligned}$$

Another important example is as follows: let X be a fixed (integrable) random variable. Then successive predictions $\widehat{X}_t = \mathbb{E}[X|\mathcal{F}_t]$ form a martingale $(\widehat{X}_t)$ with respect to $(\mathcal{F}_t)$. In fact, on any bounded interval $[s, t]$, any martingale simply predicts its final value X_t by definition. What if $t = \infty$? Then there is the *martingale convergence theorem*:

THEOREM **10.** [RY90] Let (M_t) be a càdlàg L^p-martingale. Then:

- for $p = 1$, $M_t \to M_\infty$ a.s. if $||M||_1 = \sup_t ||M_t||_1 < \infty$. Convergence is also in L^1 if and only if M is uniformly integrable, i.e.

$$\lim_{n\to\infty} \sup_t \mathbb{E}1_{\{|M_t|>n\}}|M_t| = 0.$$

- for $p > 1$, $M_t \to M_\infty$ both a.s. and in L^p if $||M||_p = \sup_t ||M_t||_p < \infty$.
- when $M_t \to M_\infty$ in L^p, then $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]$.

(When $||M||_p < \infty$, we say that the process is L^p-bounded.)
Furthermore, there are *Doob's inequalities*:

THEOREM **11.** [RY90] Let M be a càdlàg L^p-martingale, and let $M_t^* = \sup_{s\leq t} |M_s|$ be its maximal process. Then

- for $p \geq 1$, $\mathbb{P}(M_\infty^* > c) \leq (c^{-1}||M||_p)^p$
- for $p > 1$, $||M^*||_p \leq q||M||_p$, where $p^{-1} + q^{-1} = 1$

Last but not least, martingales obey the *optional stopping theorem*:

THEOREM **12.** [RY90] Let M be a càdlàg martingale and S, T two stopping times with $S \leq T$. Suppose that T is bounded. Then

$$M_S = \mathbb{E}[M_T|\mathcal{F}_S].$$

If also $M_t \to M_\infty$ in L^1, then for any T,

$$M_S = \mathbb{E}[M_T|\mathcal{F}_S] = \mathbb{E}[M_\infty|\mathcal{F}_S].$$

In particular, $M^{|T}$ is again a martingale.

$\star$ Super/submartingales

But martingales don't cover all the types of process we are interested in. As a first generalization, consider a process which has an *upward trend*. Such a process would be increasing on average. We call an adapted process (M_t) an L^p*-submartingale* if:

1. $M_t \in L^p(\Omega, \mathcal{F}, \mathbb{P})$,

2. $\mathbb{E}[M_t|\mathcal{F}_s] \geq M_s$ for all $t \geq s$.

Of course a supermartingale is a process with a *downward trend*, i.e. M is a *supermartinagle* if $(-M)$ is a submartingale (terminology comes from potential theory). Martingales are both super- and submartingales.

Exactly the same inequalities and stopping/convergence theorems hold as for martingales.

Finally, the following theorem explains why we always assume *standard* filtrations:

THEOREM **13.** [RY90] If M is a submartingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, where $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ satisfies the usual conditions, then M has a càdlàg version if $t \mapsto \mathbb{E}M_t$ is a right-continuous function.

In particular, if M is a martingale, the map $t \mapsto \mathbb{E}M_t$ is constant, so M always has a càdlàg modification.

A submartingale representation

A (u.i.) martingale M is generated by the r.v. M_∞, that is, $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]$. Is a similar result true for submartingales? Clearly, one r.v. is not enough, but is there an increasing process which generates it? The following theorem partially answers this question:

THEOREM **14.** [Kry90] Let M be a càdlàg, L^1-bounded, positive submartingale. Then there exists an increasing process N such that a.s. for all t,

$$M_t = \mathbb{E}[N_t|\mathcal{F}_t].$$

PROOF: Let $Q = \{q_0, q_1, q_2, \ldots\}$ be a countable dense subset of $\overline{\mathbb{R}}_+$ with $q_0 = 0$, $q_1 = \infty$, and such that it contains all points of discontinuity of the increasing function $t \mapsto \mathbb{E}M_t$.

For each n, let $Q_n = \{r_1 < r_2 < \cdots < r_n\}$ consist of the first n points of Q, in *increasing order*. Since M is a submartingale, functions $f_1 \in \mathcal{F}_{r_1}, \ldots, f_{n-1} \in \mathcal{F}_{r_{n-1}}$ exist such that $0 \leq f_i \leq 1$ and

$$M_{r_i} = f_i \mathbb{E}[M_{r_{i+1}}|\mathcal{F}_{r_i}].$$

Define an increasing process Z^n by

$$Z_t^n = \begin{cases} f_i f_{i+1} \cdots f_{n-1} & \text{if } t \in [r_i, r_{i+1}[\ , \\ 1 & \text{if } t = \infty. \end{cases}$$

By recursively iterating the defining formula for f_i, we see that

$$M_q = \mathbb{E}[M_\infty Z_q^n | \mathcal{F}_q] \quad \text{for all } q \in Q_n.$$

Since now the unit ball of the Hilbert space $L^2(\Omega, \mathcal{F}, M_\infty(\omega)\mathbb{P}(d\omega))$ is sequentially compact, there exists a subsequence (n_k) such that for all $q \in Q$, $Z_q^{n_k}(\omega)$ converges to some $Z_q(\omega)$. Then for any $B \in \mathcal{F}_q$,

$$\mathbb{E}1_B M_q = \mathbb{E}1_B \mathbb{E}[M_\infty Z_q^{n_k} | \mathcal{F}_q] = \mathbb{E}1_B M_\infty Z_q^{n_k},$$

and letting $k \to \infty$ shows that $M_q = \mathbb{E}[M_\infty Z_q | \mathcal{F}_q]$ a.s. for each $q \in Q$. Since Q is countable, this is also true for all q outside a common null set.

If we then write $G_t = \inf\{Z_q : q \geq t, q \in Q\}$, the process G is clearly an increasing right-continuous process and

$$\mathbb{E}M_t = \lim_{q \downarrow t} \mathbb{E}M_q = \lim_{q \downarrow t} \mathbb{E}M_\infty Z_q = \mathbb{E}M_\infty G_t.$$

Choosing $N_t = M_\infty G_t$ gives the result. □

⋆ A partial Doob-Meyer decomposition

If M is a submartingale, an important problem is to try and *compensate* it, that is, to find a nondecreasing process A with $A_0 = 0$ such that $N = M - A$ is trendless: a martingale. In discrete time, this is easy:

- Let $A_0 = 0$.
- Write $A_{n+1} - A_n = \mathbb{E}[M_{n+1} | \mathcal{F}_n] - M_n \geq 0$.
- Define N by $M_n = N_n + A_n$ and check that it is a martingale.

This is termed the *Doob decomposition* of a discrete-time submartingale. It is unique under the condition $A_n \in \mathcal{F}_{n-1}$.

In continuous time however, things are not quite so easy. The following result is a partial version of an important theorem: the *Doob-Meyer decomposition* of a submartingale.

THEOREM **15.** [Kry90] Let M be a càdlàg, L^1-bounded, positive submartingale. Then there exists a decomposition

$$M_t = M_0 + N_t + A_t$$

such that A is nondecreasing with $A_0 = 0$, and N is a martingale with $N_0 = 0$. Moreover, A is natural: it satisfies

$$\mathbb{E}X_t A_t = \mathbb{E}\int_0^t X_{s-} dA_s$$

for any bounded martingale X.

PROOF: By considering $(M - M_0)$ if necessary, we can assume that $M_0 = 0$. Let Q be as in the previous theorem. For any bounded random variable X, $\mathbb{E}[X|\mathcal{F}_t]$ defines a martingale. We will write $\mathbb{E}[X|\mathcal{F}_{t-}] = \lim_{q\uparrow t} \mathbb{E}[X, \mathcal{F}_q]$ over $q \in Q$. Now let's prove the theorem.

There exists a right-continuous increasing process Z such that $M_t = \mathbb{E}[Z_t|\mathcal{F}_t]$. For each $q \in Q$ and $B \in \mathcal{F}$, define

$$\mu_q(B) = \mathbb{E}\int_0^t \mathbb{E}[1_B|\mathcal{F}_{r-}] dZ_r.$$

Since $\mathbb{E}[1_B|\mathcal{F}_{r-}] \leq 1$ and $\mathbb{E}Z_\infty = \mathbb{E}M_\infty < \infty$, the DCT shows that μ_q is a bounded measure, absolutely continuous with respect to $\mathbb{P}$, on $(\Omega, \mathcal{F})$. Define

$$A_q = \frac{d\mu_q}{d\mathbb{P}}, \qquad A_t = \inf_{q \geq t} A_q.$$

Then A is a.s. an increasing right-continuous process and $A_t = d\mu_t/d\mathbb{P}$ a.s. for each and every t. It remains to check that $(M - A)$ is a martingale.

For any $B \in \mathcal{F}_s$, since $\mathbb{E}[1_B|\mathcal{F}_{r-}] = 1_B$ whenever $r > s$, we have

$$\begin{aligned}
\mathbb{E}1_B(A_t - A_s) &= \int 1_B\left(\frac{d\mu_t}{d\mathbb{P}} - \frac{d\mu_s}{d\mathbb{P}}\right) d\mathbb{P} \\
&= \mu_t(B) - \mu_s(B) \\
&= \mathbb{E}\int_s^t \mathbb{E}[1_B|\mathcal{F}_{r-}] dZ_r \\
&= \mathbb{E}1_B(Z_t - Z_s) = \mathbb{E}1_B(M_t - M_s),
\end{aligned}$$

This means $\mathbb{E}[M_t - A_t|\mathcal{F}_s] = \mathbb{E}[M_s - A_s|\mathcal{F}_s]$. The result will follow if we show that A_s is $\mathcal{F}_s$-measurable.

For any $B \in \mathcal{F}$, since $\mathbb{E}[\mathbb{E}[1_B|\mathcal{F}_t]|\mathcal{F}_{r-}] = \mathbb{E}[1_B|\mathcal{F}_{r-}]$ whenever $r \leq t$,

$$\begin{aligned}
\mathbb{E}1_B A_t = \mu_t(B) &= \mathbb{E}\int_0^t \mathbb{E}[\mathbb{E}[1_B|\mathcal{F}_t]|\mathcal{F}_{r-}] dZ_r \\
&= \mathbb{E}(\mathbb{E}[1_B|\mathcal{F}_t] A_t) \\
&= \mathbb{E}(\mathbb{E}[1_B|\mathcal{F}_t]\mathbb{E}[A_t|\mathcal{F}_t]) \\
&= \mathbb{E}1_B \mathbb{E}[A_t|\mathcal{F}_t].
\end{aligned}$$

Now let's check the "naturalness" of A. If X is any bounded martingale, we can use to dominated convergence theorem to get

$$
\begin{aligned}
\mathbb{E}X_t A_t &= \mathbb{E}\int_0^t X_{s-} dZ_s \\
&= \lim_{|\Delta|\to 0} \sum_{\Delta} \mathbb{E}X_{t_i -}(Z_{t_{i+1}} - Z_{t_i}) \\
&= \lim_{|\Delta|\to 0} \sum_{\Delta} \mathbb{E}X_{t_i -}(A_{t_{i+1}} - A_{t_i}) \\
&= \mathbb{E}\int_0^t X_{s-} dA_s.
\end{aligned}
$$

□

Clearly, this result extends in an obvious way to *bounded super/submartingales.*

Just out of interest, here is the full Doob-Meyer decomposition. We will not make use of it.

THEOREM **16.** [Mét82] Let M be a càdlàg submartingale. Then there exists a unique predictable increasing process A and local martingale N with $N_0 = A_0 = 0$ such that $M_t = M_0 + N_t + A_t$

⋆ Square integrable martingales are integrators

Let M be a càdlàg L^2-bounded martingale. The typical element of $\mathcal{P}$ can clearly be written

$$R = [S_0, 0] \cup]S_1, T_1] \cup \cdots \cup]S_n, T_n],$$

where $0 \le S_0 \le S_1 \le T_1 \le S_2 \le \cdots \le S_n \le T_n$. If $i > j$, we find by the optional stopping theorem (M is uniformly integrable)

$$\mathbb{E}(M_{T_i} - M_{S_i})(M_{T_j} - M_{S_j}) = \mathbb{E}(M_{T_i} - M_{S_i})\mathbb{E}[M_{T_j} - M_{S_j} | \mathcal{F}_{T_j}] = 0.$$

(This means that for any A, $B \in \mathcal{P}$, with $A \cap B = \emptyset$, $\mathbb{E}\mu_M(A) \cdot \mu_M(B) = 0$. The set function μ_M is called *orthogonally scattered*). We can now use the above to get

$$\mathbb{E}|\mu_M(R)|^2 = \mathbb{E}|\sum_i (M_{T_i} - M_{S_i})|^2$$

$$
\begin{aligned}
&= \sum_i \mathbb{E}(M_{T_i} - M_{S_i})^2 \\
&= \sum_i \mathbb{E}(M_{T_i}^2 - 2M_{T_i}M_{S_i} + M_{S_i}^2) \\
&= \sum_i \mathbb{E}(M_{T_i}^2 - M_{S_i}^2) \\
&\leq \mathbb{E}(M_{T_n}^2 - M_{S_1}^2) \leq 2\mathbb{E}M_\infty^2 .
\end{aligned}
$$

This computation shows that $||\mu_M||(\mathbb{R}_+ \times \Omega) < \infty$. Furthermore, we also see that $\mathbb{E}|\mu_M(dt, d\omega)|^2 = \mathbb{E}\mu_{M^2}(dt, d\omega) = \nu(dt, d\omega)$ on $\mathcal{P}_0$. Thus $\langle x', \mu_M \rangle$ is σ-additive there if μ_M is if and only if ν is. Now M^2 is a càdlàg L^1-bounded, positive submartingale (apply Jensen's inequality to the convex function $x \mapsto x^2$) and hence it can be written $M^2 = N + A$, where N is a martingale with $N_0 = 0$ and A is increasing. So on $\mathcal{P}[0, T]$, where $T < \infty$ is any stopping time, the measure $\nu(dt, d\omega) = \mathbb{E}\mu_A(dt, d\omega)$ is σ-additive since A is an FV process. This shows that an L^2-bounded martingale is an integrator. The real measure $\mathbb{E}\mu_A(dt, d\omega)$ on $\mathcal{P}[0, T]$ is called a control measure for μ_M.

As opposed to the case of FV processes, the fact that we worked only with predictable sets was crucial in the above.

$\star$ $BM^0(\mathbb{R})$ is an integrator

As a consequence, we can show that if B is a $BM^0(\mathbb{R})$, then it is an integrator: by continuity of the sample paths, the stopping times $T_n = \inf\{t : |B_t - x| > n\}$ are a.s. increasing to infinity and $B^{|T_n}$ is a bounded and hence a square integrable martingale for each n (this argument works for any continuous martingale). In fact, putting $x = 0$ for simplicity, the decomposition $B^2 = N + A$ is given explicitely by $A_t = t$ and $N_t = B_t^2 - t$. Also, the control measure is given explicitely by

$$\mathbb{E}\mu_{B^2}(dt, d\omega) = m \otimes \mathbb{P}(dt, d\omega)$$

where m is Lebesgue measure.

(Super/Sub)Martingales are integrators

We start with the following result:

THEOREM 17. [Bic81] A bounded, positive supermartingale M which vanishes after a stopping time T is an integrator.

PROOF: The conditions on M are sufficient for the existence of a decomposition $M = N + A$, where N is a martingale, and A is decreasing with $A_0 = 0$ (so that $N_0 = M_0$). Let $K < \infty$ be a bound for M. The theorem will follow if we show that N is L^2-bounded. But since $M_t = 0$ for $t \geq T$, we have by the naturalness of A,

$$\begin{aligned} \mathbb{E}N_t^2 = \mathbb{E}A_t^2 &= 2\mathbb{E}A_t^2 - \mathbb{E}\int_0^t A_s + A_{s-}dA_s \\ &\leq -2\mathbb{E}\int_0^t M_s dA_s \\ &\leq -2K\mathbb{E}A_t, \end{aligned}$$

and $-2K\mathbb{E}A_t = 2K\mathbb{E}N_t = 2K\mathbb{E}N_0 = 2K\mathbb{E}N_0 = 2K\mathbb{E}M_0 \leq 2K^2$. So the increasing function $t \mapsto \mathbb{E}N_t^2$ is eventually bounded, and the result follows. □

The above result allows a large extension of the possible integrators as we now demonstrate.

Let M be a super/sub/martingale. To show it is (pre)locally summable, it suffices to show that submartingales are integrators, for if M is a supermartingale, recall that $(-M)$ is a submartingale, and if M is a martingale, it is automatically a submartingale.

So let M be a submartingale. Since $M = M^+ - M^-$, Jensen's inequality applied to the convex functions $x \mapsto x^+ = x \vee 0$ and $x \mapsto x^- = -(x \wedge 0)$ shows that $M^\pm$ are positive submartingales. So we can assume $M \geq 0$.

Now let $T_n = \inf\{t : M_t > n\} \wedge n$. Then T_n is a finite stopping time and $T_n \uparrow \infty$ as n increases. Moreover,

$$M^{|T_n} = (M1_{[0,T_n[} + n1_{[T_n,\infty[}) + (n1_{[0,T_n[} + M_{T_n}1_{[T_n,\infty[}).$$

The first process is a positive, bounded submartingale and hence an integrator, whereas the second process is constant except for a single jump at time T_n, and hence is an IV process.

The Poisson martingale is an integrator

If N is a $PP(\lambda)$, then $M_t = N_t - \lambda t$ is a (discontinuous) martingale, and hence an integrator. In fact, since the paths of N are increasing, the process

M is actually an FV process. N is also a positive (unbounded however) submartingale and has a decomposition $N_t = M_t + \lambda t$. Its control measure acts on predictable sets by

$$\mu_M(]s,t] \times F) = \mathbb{E}1_F(N_t - N_s) = \mathbb{E}1_F\lambda(t-s) = \lambda m \otimes \mathbb{P}(]s,t] \times F).$$

Thus $\mu_M = \lambda m \otimes \mathbb{P}$ on $\mathcal{P}$.

However, if $\tau_1 = \inf\{t : N_t = 1\}$ is the time of the first jump, we have

$$\mu_M(N=0) = \mu_M([0,\tau_1[) = 0,$$

whereas

$$\lambda m \otimes \mathbb{P}(N=0) = \lambda m \otimes \mathbb{P}([0,\tau_1[) = \lambda \int_\Omega \int_{\mathbb{R}_+} 1_{[0,\tau_1[}(s,\omega) ds d\mathbb{P}(\omega) = \lambda E\tau_1 = 1.$$

This means that the set $\{N=0\}$ is not predictable, i.e. *the* $PP(\lambda)$ *is not a predictable process.*

Local martingales are integrators

A local martingale is a process M such that $M^{|T_n}$ is a martingale for some localizing sequence (T_n). *Such processes need not be martingales*, although any martingale is a local martingale, localized by the stopping times $T_n = n$. Note that by choosing $(T_n \wedge n)$ instead, we can always assume $M^{|T_n}$ is uniformly integrable. By the previous results, local martingales are integrators. Moreover, all the integrators we have found so far can clearly be decomposed into a (nonunique) sum of a local martingale and an FV process.

When is a local martingale actually a martingale? Here is a simple sufficient condition:

THEOREM **18.** [Pro92] Let M be a local martingale. If $\mathbb{E}M_t^* < \infty$ for every $t \in \mathbb{R}_+$, then M is a martingale. If also $\mathbb{E}M_\infty^* < \infty$, then M is a uniformly integrable martingale.

PROOF: Let (T_n) be a localizing sequence such that $M^{|T_n}$ is a uniformly integrable martingale. When $s \leq t$, $\mathbb{E}[M_{t\wedge T_n}|\mathcal{F}_s] = M_{s \wedge T_n}$. Letting $n \to \infty$ and using the ordinary DCT, this means $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$. If furthermore $\mathbb{E}M_\infty^* < \infty$, then since $|M_t| \leq M_\infty^*$, the DCT also shows that $M_t \to M_\infty$ in L^1, which implies M is uniformly integrable. $\square$

Quasimartingales are integrators

Let X be càdlàg and $(\mathcal{F}_t)$-adapted. The process X is called a *quasimartingale* if the following holds: for any finite partition Δ of $[0,t]$,

$$\sum_{\Delta} \mathbb{E}\left|\mathbb{E}[X_{t_{i+1}}|\mathcal{F}_{t_i}] - X_{t_i}\right| \leq K(t) \leq K(\infty) < \infty.$$

Such processes are in some sense the probabilistic version of a function of bounded variation.

First note that any submartingale (and hence supermartingale) is trivially a quasimartingale, because $\mathbb{E}[X_{t_{i+1}}|\mathcal{F}_{t_i}] \geq X_{t_i}$ means we can take out the absolute value signs, so that

$$\sum_{\Delta} \mathbb{E}\left(\mathbb{E}[X_{t_{i+1}}|\mathcal{F}_{t_i}] - X_{t_i}\right) = \mathbb{E}X_t - \mathbb{E}X_0 < \infty.$$

It is also easy to see that quasimartingales form a vector space.

Just as real functions of bounded variation may be written as a difference of two increasing functions, we have the following theorem in the stochastic case:

THEOREM **19.** [Pro92] Any quasimartingale is the difference of two submartingales.

PROOF: Let $\Delta = \{q = t_0 < \cdots < t_n = \infty\}$ be a finite subdivision of $[q,\infty]$, where $q \geq 0$ is rational. Define the random variables

$$\begin{aligned} Y_q^{\Delta} &= \mathbb{E}[\sum_{\Delta} \mathbb{E}[X_{t_{i+1}} - X_{t_i}|\mathcal{F}_{t_i}]^+|\mathcal{F}_q] \\ Z_q^{\Delta} &= \mathbb{E}[\sum_{\Delta} \mathbb{E}[X_{t_{i+1}} - X_{t_i}|\mathcal{F}_{t_i}]^-|\mathcal{F}_q] \end{aligned}$$

Then if $\Delta \subseteq \Delta'$, we have $Y_q^{\Delta} \leq Y_q^{\Delta'}$, $Z_q^{\Delta} \leq Z_q^{\Delta'}$. To see this, consider what happens when $\Delta' = \Delta \cup \{t\}$, where $t_i < t < t_{i+1}$:

$$\begin{aligned} \mathbb{E}[X_{t_{i+1}} - X_{t_i}|\mathcal{F}_{t_i}]^{\pm} &\leq \mathbb{E}[\mathbb{E}[X_{t_{i+1}} - X_t|\mathcal{F}_t]|\mathcal{F}_{t_i}]^{\pm} + \mathbb{E}[X_t - X_{t_i}|\mathcal{F}_{t_i}]^{\pm} \\ &\leq \mathbb{E}[\mathbb{E}[X_{t_{i+1}} - X_t|\mathcal{F}_t]^{\pm}|\mathcal{F}_{t_i}] + \mathbb{E}[X_t - X_{t_i}|\mathcal{F}_{t_i}]^{\pm} \end{aligned}$$

by Jensen's inequality applied to $x \mapsto x \vee 0 = x^+$ and $x \mapsto -(x \wedge 0) = x^-$. Taking conditional expectations with respect to $\mathcal{F}_q$ then yields that $Y_q^{\Delta} \leq Y_q^{\Delta'}$ and $Z_q^{\Delta} \leq Z_q^{\Delta'}$.

Now the collection of all partitions $\Delta = \{q = t_0 < \cdots < t_n = \infty\}$ of $[q, \infty]$ forms a directed partially ordered set under inclusion. Since $\mathbb{E}Y_q^\Delta \leq K(\infty) < \infty$, we may take limits in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ along the net $\Delta \mapsto Y_q^\Delta, Z_q^\Delta$ and set

$$Y_q = \lim_\Delta Y_q^\Delta, \quad Z_q = \lim_\Delta Z_q^\Delta.$$

We get right-continuous processes Y and Z by

$$Y_t = \lim_{q \downarrow t} Y_q, \qquad Z_t = \lim_{q \downarrow t} Z_q.$$

Then both processes are right-continuous submartingales, and because for all Δ, the equation $Y_q^\Delta - Z_q^\Delta = \mathbb{E}[\sum_\Delta \mathbb{E}[X_{t_{i+1}} - X_{t_i} | \mathcal{F}_{t_i}] | \mathcal{F}_q] = X_q$ holds, we also see that $X = Y - Z$. □

⋆ Semimartingales

We have now seen two quite different types of processes which can be used as integrators: FV processes, for which the integration theory reduces to Lebesgue-Stieltjes integration, and martingale-type processes, for which the integration theory is "new". Are there any more? The following theorem says no.

THEOREM **20.** [Yor78, Kus77] An adapted, càdlàg process X is locally (resp. prelocally) summable if and only if it can be written $X = M + A$ where M is a local martingale and A is locally an IV process (resp. an FV process).

PROOF: Since we've already seen that any process with such a decomposition is locally summable, we need only show the converse. But this is easy, for if X is summable, then as Δ ranges over all finite partitions of $[0, t]$, we let

$$s_i(\Delta) = \operatorname{sgn}\left(\mathbb{E}[X_{t_{i+1}} - X_{t_i} | \mathcal{F}_{t_i}]\right) \in \mathcal{F}_{t_i},$$

so that

$$\begin{aligned} \sup_\Delta \sum_\Delta \mathbb{E}|\mathbb{E}[X_{t_{i+1}} - X_{t_i} | \mathcal{F}_{t_i}]| &\leq \sup_\Delta \mathbb{E}\left|\sum_\Delta s_i(\Delta)(X_{t_{i+1}} - X_{t_i})\right| \\ &\leq \|\mu_X\|([0, t] \times \Omega) < \infty. \end{aligned}$$

Thus any summable process is a quasimartingale, hence a difference of submartingales, hence decomposable. For the general case, we conclude by localization. □

Thus we have a complete (up to modifications) characterization of the processes we may use as integrators. They are called, not so surprisingly, *semimartingales*. What is much more surprising is the following theorem:

THEOREM **21.** [Bic81, Mey79, Del80] An adapted, càdlàg process X induces a vector measure μ_X on $\mathcal{P}$ with values in $L^0(\Omega, \mathcal{F}, \mathbb{P})$ (topologized by convergence in probability), if and only if it is a semimartingale.

So we see that there really aren't any other integrators, even if we weaken considerably the requirements. Rather than prove this result, which is admirably done in the cited references, let's look next at some of the marvellous properties of stochastic integrals.

$\star$ Stability under semimartingale decomposition

From Lebesgue-Stieltjes integration theory, we know that if we write

$$G(t) = \int_0^t h(s)dF(s),$$

for some function F of bounded variation and $h \in L^1(F)$, then the function G is again of bounded variation. In other words, the class of Lebesgue-Stieltjes integrators is stable under integration. Since the stochastic integral of an FV process may be evaluated path-by-path, we see that FV processes are stable under stochastic integration also. Is a similar result true also for local martingales? The next theorem is a partial yes:

THEOREM **22.** If M is a (local) martingale, H a (locally) bounded process, then the process $(H \cdot M)$ is again a (local) martingale.

PROOF: By localization, we can assume that H is bounded and M is a martingale. If H is a simple $\mathcal{P}_0$-process, so that

$$(H \cdot M)_t = \sum_i h_i \left(M_{T_{i+1}\wedge t} - M_{T_i \wedge t}\right)$$

for some $h_i \in \mathbb{R}$, then $(H \cdot M)$ is easily seen to be a martingale if M is (the $T_i \wedge t$ are finite stopping times, so the optional stopping theorem applies).

For the general case, approximate H by bounded simple $\mathcal{P}_0$-measurable functions (H^n), in such a way that $(H^n \cdot M) \to (H \cdot M)$ in ucp. Pick a subsequence (n_k) such that $(H^{n_k} \cdot M)$ converges to $(H \cdot M)$ a.s. uniformly on every $[0, t]$. Then for $s \leq t$,

$$\mathbb{E}[(H \cdot M)_t | \mathcal{F}_s] = \mathbb{E}[\lim_k (H^{n_k} \cdot M)_t | \mathcal{F}_s] = \lim_k (H^{n_k} \cdot M)_s = (H \cdot M)_s,$$

where the uniform convergence justifies interchanging limits with expectation.

Now a martingale is a local martingale. Thus we have shown the theorem. □

⋆ Riemann approximations

An *adapted subdivision* is a finite sequence $\tau = (T_n)$ of stopping times satisfying $0 = T_0 \leq T_1 \leq \cdots \leq T_k < \infty$. We say that a sequence $(\tau_n = (T_{nm}))$ of adapted subdivisions is a *Riemann sequence* provided that $\sup_m |T_{n(m+1)} \wedge t - T_{nm} \wedge t| \to 0$ for all $t \in \mathbb{R}_+$, and $\lim_n \sup_m T_{nm} = \infty$.

THEOREM **23.** [JS87] Let X be a semimartingale, H a càglàd process, $(\tau_n = (T_{nm}))$ a Riemann sequence. Then the approximations

$$\tau_n (H \cdot X)_t = \sum_m H_{T_{nm}} (X_{T_{n(m+1)} \wedge t} - X_{T_{nm} \wedge t})$$

converge to $(H_- \cdot X)$ in ucp.

PROOF: The processes H^{τ_n}, defined by

$$H^{\tau_n} = \sum_m H_{T_{nm}} 1_{]T_{nm}, T_{n(m+1)}]},$$

converge to H_- as $n \to \infty$. Since H is càglàd its maximal process $H_t^* = \sup_{s \leq t} |H_s|$ is in $L^1_{\text{loc}}(X)$ and by SDCT, $H^{\tau_n} \cdot X = \tau_n (H \cdot X) \to H_- \cdot X$ in ucp. □

⋆ The quadratic covariation process

Since martingales have paths of unbounded variation (such as $BM^0(\mathbb{R})$), do standard results such as integration by parts, the fundamental theorem of calculus still hold? Here, the answer is no, and finding out what replaces these is truly the most beautiful part of stochastic integration.

Let X and Y be semimartingales. Consider the expression

$$[X,Y]_t = X_tY_t - X_0Y_0 - \int_0^t Y_{s-}dX_s - \int_0^t X_{s-}dY_s.$$

Here X_{t-}, Y_{t-} are left-continuous locally bounded processes, so the above expression makes sense.

The process $[X,Y]$ measures the degree to which the classical formula of integration by parts applies to the stochastic calculus of two given semimartingales. It is called the *quadratic covariation of* X *and* Y, or simply the *square bracket*.

The bracket is clearly bilinear, symmetric, has $[X,Y]_0 = 0$, and obeys the following convenient *polarization formula*:

$$[X,Y] = \frac{1}{2}\Big([X+Y,X+Y] - [X,X] - [Y,Y]\Big).$$

Why is it called the quadratic covariation? Apply a Riemann approximation: Let $(\tau_n = (T_{nm})_{m=0}^{k(n)})$ be a Riemann sequence and define

$$\tau_n(XY)_t = X_0Y_0 + \sum_m (X_{T_{n(m+1)}\wedge t}Y_{T_{n(m+1)}\wedge t} - X_{T_{nm}\wedge t}Y_{T_{nm}\wedge t}) = (XY)_t^{|T_nk(n)}.$$

Then

$$\begin{aligned}
\tau_n[X,Y]_t &= \tau_n(XY)_t - X_0Y_0 + \tau_n(X_-\cdot Y)_t - \tau_n(Y_-\cdot X)_t \\
&= \sum_m (X_{T_{n(m+1)}\wedge t} - X_{T_{nm}\wedge t})(Y_{T_{n(m+1)}\wedge t} - Y_{T_{nm}\wedge t}),
\end{aligned}$$

and as $n\to\infty$, these partial sums converge to $[X,Y]$ in ucp.

How important is the bracket really? When X and Y are FV processes, so that the classical integration by parts applies for almost each ω, we expect the equation $[X,Y]_t = \sum_{s\le t}\Delta X_s \Delta Y_s$. This doesn't really warrant the fancy bracket notation. But suppose $X = Y = B$ is a $BM^0(\mathbb{R})$. Then $\Delta B = 0$, but we've seen earlier that

$$[B,B]_t = \lim_n \sum_{\Delta_n} (B_{t_{i+1}\wedge t} - B_{t_i\wedge t})^2 = t,$$

which is definitely not zero! Thus $[X,Y]$ is really quite an exciting process.

From the approximation, we read off that the bracket $[X,X]$ is always an increasing process, and by polarization, $[X,Y]$ is always an FV process, hence a semimartingale.

Note that since all the other terms in the defining equation for $[X,Y]$ are semimartingales, the process XY must also be one. So the space of semimartingales forms an *algebra*.

$\star$ Some properties of the bracket

By the Riemann approximation

$$[X,Y]_t = \lim_{\tau_n} \sum_m (X^{|T_{n(m+1)}} - X_t^{|T_{nm}})(Y_t^{|T_{n(m+1)}} - Y_t^{|T_{nm}}),$$

we trivially deduce that for any stopping time T,

$$[X,Y]^{|T} = [X^{|T}, Y] = [X, Y^{|T}] = [X^{|T}, Y^{|T}].$$

THEOREM **24.** [Pro92, Del80] Let X, Y be semimartingales, H, K be locally bounded processes. Then

$$[H \cdot X, K \cdot Y] = (HK) \cdot [X,Y].$$

PROOF: By polarization, it will be sufficient to show $[H \cdot X, X] = H \cdot [X,X]$. If $H = 1_{]S,T]}$, then clearly $H \cdot X = X^{|T} - X^{|S}$, and

$$H \cdot [X,X] = [X,X]^{|T} - [X,X]^{|S} = [X^{|T} - X^{|S}, X] = [H \cdot X, X].$$

Similarly, if $H = 1_{[S,0]}$, then $H \cdot X = 0$, and $[H \cdot X, H \cdot X] = 0 = H \cdot [X,X]$. Thus by linearity, the theorem is true for $H \in \mathcal{P}_0$.

If H is a locally bounded, predictable process, we can assume it is bounded by localization. Then there are $H^n \in \mathcal{P}_0$ with $H^n \to H$. So $H^n \cdot X \to H \cdot X$ in ucp. So there is a subsequence (n_k) such that $H^{n_k} \cdot X \to H \cdot X$ a.s., and of course $H^{n_k} \to H$. Then

$$\begin{aligned} H^{n_k} \cdot [X,X] &= [H^{n_k} \cdot X, X] \\ &= (H^{n_k} \cdot X)X - H_0^{n_k} X_0^2 - (H^{n_k} \cdot X)_- \cdot X - X_-(H^{n_k} \cdot X), \end{aligned}$$

and note that $X_- \cdot (H^{n_k} \cdot X) = (X_- H^{n_k}) \cdot X = H^{n_k} \cdot (X_- \cdot X)$. Letting $k \to \infty$ and applying SDCT yields the theorem. $\square$

The bracket and martingales

The quadratic variation yields a very useful sufficient condition for a local martingale to be a true martingale.

THEOREM **25.** [Pro92] Let M be a càdlàg local martingale. If $\mathbb{E}[M,M]_t < \infty$ for all t, then M is an L^2-martingale. Furthermore, $\mathbb{E}M_t^2 = \mathbb{E}[M,M]_t$.

Recall that we need to show $\mathbb{E}M_t^* < \infty$ for all t. Define an increasing sequence of stopping times $T_n = \inf\{t : |M_t| > n\} \wedge n \uparrow \infty$. Then

$$(M^{|T_n})^* \leq n + |\Delta M_{T_n}| \leq n + ([M,M]_n)^{1/2} < \infty,$$

where the last term is in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. So $M^{|T_n}$ is an L^2-bounded martingale. So it is summable and further $(M^{|T_n} \cdot M^{|T_n})$ is a martingale with zero mean. So

$$\mathbb{E}(M_t^{|T_n})^2 = 2\mathbb{E}(M^{|T_n} \cdot M^{|T_n})_t + \mathbb{E}[M^{|T_n}, M^{|T_n}]_t = \mathbb{E}[M,M]_t^{|T_n}.$$

But by Doob's inequality,

$$\mathbb{E}(M^*_{t\wedge T_n})^2 \leq 4\mathbb{E}(M_{t\wedge T_n})^2 = 4\mathbb{E}[M,M]_{t\wedge T_n} \leq 4\mathbb{E}[M,M]_t.$$

So by the monotone convergence theorem, we get as $n \to \infty$

$$\mathbb{E}(M_t^*)^2 \leq 4\mathbb{E}[M,M]_t < \infty,$$

and we have shown that M is an L^2-martingale. □

⋆ Discontinuities

How do we compute the discontinuities of $(H \cdot X)$? Assume H is locally bounded, then here's the answer:

$$\Delta(H \cdot X) = H\Delta X.$$

This is easy to prove: It is clearly true if H is simple predictable. In the general case, apply localization and find a sequence (H^k) of bounded simple processes converging to X; SDCT does the rest.

Another very useful result is the following: for any two semimartingales,

$$\Delta[X,Y] = \Delta X \Delta Y.$$

This is easily seen as follows: by polarization, we can assume $Y = X$. Then

$$\begin{aligned}\Delta[X,X] &= \Delta(X^2) - \Delta(X_0^2) - 2X_-\Delta X \\ &= X^2 - X_-^2 - 2X_-(X - X_-) \\ &= (X - X_-)^2 = (\Delta X)^2.\end{aligned}$$

The above also yields the following decomposition of the bracket into its continuous and discrete parts:

$$[X,Y] = [X,Y]^c + \sum_{s\le\cdot} \Delta X_s \Delta Y_s.$$

Observe how the above implies that $[X,Y]$ is continuous provided one of X,Y is.

When X is a semimartingale and Y is an FV process

Here is a theorem which illustrates the use of random Riemann sequences.

THEOREM **26.** [JS87] Let X be a semimartingale and Y an FV process. Then

$$[X,Y] = \Delta X \cdot Y.$$

PROOF: We define a Riemann sequence by $T_{n0} = 0$, and

$$T_{n(m+1)} = \inf\{t > T_{nm} : |X_t - X_{T_{nm}}| > 1/n\} \wedge (T_{nm} + 1/n).$$

This means that $|X_s - X_{T_{nm}}| \le 1/n$ on $\{T_{nm} < s < T_{n(m+1)}\}$. Then

$$|X_{T_{n(m+1)}} - X_{T_{nm}}| \le \frac{1}{n} + |\Delta X_{T_{n(m+1)}}| \le \frac{3}{n} \quad \text{on } \{|\Delta X_{T_{n(m+1)}}| \le 2/n\},$$

so that the FV process $A^n = (\Delta X 1_{\{|\Delta X|>2/n\}}) \cdot Y$ satisfies

$$|\tau_n([X,Y])_t - A^n_t| \le \frac{3}{n}\sum_m |Y_{t\wedge T_{n(m+1)}} - Y_{t\wedge T_{nm}}| \le \frac{3}{n}\int_0^t |dY_s|,$$

since Y is an FV process. Letting $n \to \infty$ shows that $A^n_t \to (\Delta X \cdot Y)_t$ on the one hand, and $A^n_t \to \tau_n([X,Y])$ on the other. Hence the theorem is proved. □

In particular, if both X and Y are FV processes, we get back the classical formula
$$[X,Y] = \sum_{s\leq \cdot} \Delta X \Delta Y.$$

As an application, suppose N is a $PP(\lambda)$. The process is constant between jumps, and the jumps have size 1 and are always positive. So $N_t = \sum_{s\leq t} \Delta N_s$ and we get
$$[N,N] = \sum_{s\leq \cdot} (\Delta N_s)^2 = \sum_{s\leq \cdot} \Delta N_s = N$$

The Kunita-Watanabe inequality

If we write $[X,Y]_s^t = [X,Y]_t - [X,Y]_s$ then it turns out that
$$|[X,Y]_s^t| \leq ([X,X]_s^t)^{1/2} \cdot [Y,Y]_s^t)^{1/2}.$$

To see this, note that for each s, t, and λ, we have outside a null set (which may depend on s, t, λ)
$$\begin{aligned} 0 &\leq [X+\lambda Y, X+\lambda Y]_s^t \\ &= \lambda^2 [Y,Y]_s^t + 2\lambda [X,Y]_s^t + [X,X]_s^t. \end{aligned}$$

Fixing rational s and t, this quadratic equation must hold for all λ by continuity, which means its discriminant is nonnegative. This gives the required inequality for all rational s, t outside a common null set. Using right-continuity of the bracket, it must hold for all real s, t outside a common null set.

This may now be generalized to the *Kunita-Watanabe inequality*:

THEOREM **27.** [CW83] Let X and Y be two semimartingales, and both H and K be two $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}_\infty$-measurable processes. Then a.s.
$$\int_0^t |H_s K_s|\, |d[X,Y]_s| \leq \left(\int_0^t H_s^2 d[X,X]_s\right)^{1/2} \left(\int_0^t K_s^2 d[Y,Y]_s\right)^{1/2}$$

PROOF: If H and K are $\mathcal{P}_0$-simple, then they can be simultaneously written as $H = h_0 1_{[S_0,0]} + \sum_i h_i 1_{]S_i,T_i]}$, and $K = k_0 1_{[S_0,0]} + \sum_i k_i 1_{]S_i,T_i]}$, where the h_i, $k_i \in \mathbb{R}$. Then for each ω outside some nullset,
$$\int_0^t |H_s K_s|\, |d[X,Y]_s|(\omega) \quad \leq \quad \sum_{\Delta} |h_i k_i|\, |[X,Y]_{S_i(\omega)}^{T_i(\omega)}(\omega)|$$

$$
\begin{aligned}
&\leq \sum_{\Delta} |h_i k_i| \left([X,X]_{S_i(\omega)}^{T_i(\omega)}(\omega)[Y,Y]_{S_i(\omega)}^{T_i(\omega)}(\omega) \right)^2 \\
&\leq \left(\sum_{\Delta} h_i^2 [X,X]_{S_i(\omega)}^{T_i(\omega)}(\omega) \right)^{1/2} \left(\sum_{\Delta} k_i^2 [Y,Y]_{S_i(\omega)}^{T_i(\omega)}(\omega) \right)^{1/2} \\
&= \left(\int_0^t H_s^2(\omega) d[X,X]_s(\omega) \right)^{1/2} \left(\int_0^t K_s^2(\omega) d[Y,Y]_s(\omega) \right)^{1/2}.
\end{aligned}
$$

For general H and K, approximate for almost each ω the real functions $s \mapsto |H_s(\omega)|$ and $s \mapsto |K_s(\omega)|$ by increasing sequences of simple functions and apply the monotone convergence theorem. □

Of course, applying Hölder's inequality yields

$$
\mathbb{E} \int_0^t |H_s K_s| \, |d[X,Y]_s| \leq \left\| \left(\int_0^t H_s^2 d[X,X]_s \right)^{1/2} \right\|_p \left\| \left(\int_0^t K_s^2 d[Y,Y]_s \right)^{1/2} \right\|_q ,
$$

where $p^{-1} + q^{-1} = 1$. This yields an especially nice formula when $p = q = 1/2$.

As a simple application of this inequality, we immediately see that if $[X,X] = 0$ or $[Y,Y] = 0$, then always $[X,Y] = 0$. In particular, if either X of Y is a continuous FV process, then $[X,Y] = 0$.

⋆ ⋆ Itô's formula

We now come to the fundamental reason for studying stochastic integrals: Itô's formula. It is the stochastic analogue of the fundamental theorem of calculus.

THEOREM 28. [JS87] Let X be a semimartingale and $f \in C^2(\mathbb{R}, \mathbb{R})$. Then $f(X)$ is again a semimartingale and

$$
f(X) = f(X_0) + f'(X_-) \cdot X + \frac{1}{2} f''(X_-) \cdot [X,X]^c + \sum_{s \leq \cdot} \widehat{f}(X_s),
$$

where

$$
\widehat{f}(X_s) = f(X_s) - f(X_{s-}) - f'(X_{s-}) \Delta X_s.
$$

PROOF: (i) We first show the result when f is a polynomial. Since products of semimartingales are semimartingales, $f(X)$ is a again a semimartingale.

When f is constant, the result is trivial. Now assume g satisfies Itô's formula and let $f(x) = xg(x)$. Since

$$f(X) = Xg(X) = X_0 g(X_0) + X_- \cdot g(X) + g(X_-) \cdot X + [X, g(X)],$$

and noting that the last two terms in Itô's formula are FV processes, a little bracket computation yields

$$\begin{aligned} [X, g(X)] &= g'(X) \cdot [X,X]^c + \sum_{s \le \cdot} \Delta X_s \left(g(X_s) - g(X_{s-}) \right), \\ X_- \cdot g(X) &= X_- g'(X_-) \cdot X + \frac{1}{2} X_- g''(X_-) \cdot [X,X]^c + \sum_{s \le \cdot} X_{s-} \widehat{g}(X_s). \end{aligned}$$

Hence

$$\begin{aligned} f(X) &= f(X_0) + (X_- g'(X_-) + g(X_-)) \cdot X + \frac{1}{2} \left(2g'(X_) + X_- g''(X_-) \right) \cdot [X,X]^c \\ &\quad + \sum_{s \le \cdot} \left\{ X_{s-} \widehat{g}(X_s) + \Delta X_s \left(g(X_s) - g(X_{s-}) \right) \right\} \\ &= f(X_0) + f'(X_-) \cdot X + \frac{1}{2} f''(X_-) \cdot [X,X]^c \\ &\quad + \sum_{s \le \cdot} \left\{ f(X_s) - f(X_{s-}) - f'(X_{s-}) \Delta X_s \right\}, \end{aligned}$$

and the result is true for polynomials.

(ii) Observe that, in general, Itô's formula is stable under stopping. By the localizing sequence $T_n = \inf\{t : |X_t| > n\}$, we can assume that X takes its values in the compact set $K = \{x : |x| \le n\}$, provided we work on the stochastic interval $[0, T_n[$. The Weierstrass approximation theorem then tells us that there are polynomials g_k which converge (as well as their first and second derivatives) uniformly to f (and its first and second derivative) on K. The g_k satisfy Itô's formula. Now let $k \to \infty$. It remains to show that we can take the various limits under the sums/integrals.

Fix t. The case $g_k(X_t) \to f(X_t)$ as $k \to \infty$ is clear. Further, both $g_k(X_-)$ and $g''(X_-)$ are bounded when X_- is in K, so by SDCT,

$$g'_k(X_-) \cdot X \to f'(X_-) \cdot X$$

and

$$g''_k(X_-) \cdot [X,X]^c \to f''(X_-) \cdot [X,X]^c.$$

Finally, if we define $\widetilde{f}(x,y) = f(x) - f(y) - f'(y)(x-y)$ (so $\widehat{f}(X_s) = \widetilde{f}(X_s, X_{s-})$), then whenever $|x|$, $|y| \le n$,

$$|\widetilde{f}(x,y)| \le C|x-y|^2,$$

and

$$|\widetilde{g}_k(x,y)| \le C|x-y|^2,$$

for some constant C.

Hence $\sum_{s \le t} |\widetilde{f}(X_s, X_{s-})| < \infty$ and

$$\sum_{s \le t} \widetilde{g}_k(X_s, X_{s-}) \to \sum_{s \le t} \widetilde{f}(X_s, X_{s-}),$$

by the ordinary DCT. □

When the semimartingale X is continuous, the formula is often written in "differential form":

$$df(X_t) = f'(X_{t-})dX_t + \frac{1}{2} f''(X_{t-})d[X,X]_t.$$

Itô's original formula dealt with the $BM^0(\mathbb{R})$:

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds.$$

Extensions of Itô's formula

There is a straightforward multidimensional extension of Itô's formula which is proved in the same way:

THEOREM **29.** [JS87] Let $X = (X^1, \dots, X^d)$ be an d-dimensional semimartingale, and $f \in C^2(\mathbb{R}^d, \mathbb{R})$. Then $f(X)$ is a semimartingale and we have:

$$f(X) = f(X_0) + \sum_{i \le d} D_i f(X_-) \cdot X^i + \frac{1}{2} \sum_{i,j \le d} D_{ij} f(X_-) \cdot [X^i, X^j]^c + \sum_{s \le \cdot} \widehat{f}(X_s),$$

where

$$\widehat{f}(X_s) = f(X_s) - f(X_{s-}) - \sum_{i \le d} D_i f(X_{s-}) \Delta X_s^i.$$

The extension to complex functions is now trivial. Say that $Z = X + iY$ is a complex semimartingale if X and Y are real semimartingales. If f is a complex-valued function, we can extend our stochastic calculus in the obvious way by treating real and imaginary parts separately (processes such as $[Z, W]$ then become $\mathbb{C}$-linear).

Recall that we often write $\frac{\partial}{\partial z} = \frac{1}{2}(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y})$ and $\frac{\partial}{\partial \overline{z}} = \frac{1}{2}(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y})$. Then any function $f : \mathbb{C} \to \mathbb{C}$, differentiable as a function of both variables x and y, and satisfying $\partial f/\partial\overline{z} = 0$ is holomorphic, in which case we write $f' = \partial f/\partial z$.

THEOREM **30.** Let $f : \mathbb{C} \to \mathbb{C}$ be twice continuously differentiable (as a function of two real variables), and let Z be a continuous complex semimartingale. Then $f(Z)$ is again a complex semimartingale and

$$\begin{aligned} f(Z) &= f(Z_0) + \frac{\partial f}{\partial z}(Z) \cdot Z + \frac{\partial f}{\partial \overline{z}}(Z) \cdot \overline{Z} + \frac{1}{2}\frac{\partial^2 f}{\partial z^2}(Z) \cdot [Z, Z] \\ &\quad \frac{1}{2}\frac{\partial^2 f}{\partial \overline{z}^2}(Z) \cdot [\overline{Z}, \overline{Z}] + \frac{\partial^2 f}{\partial z \partial \overline{z}}(Z) \cdot [Z, \overline{Z}]. \end{aligned}$$

PROOF: Simply switch from (x, y) coordinates to $(z, \overline{z})$ coordinates. □

In particular, if f is holomorphic, the above reduces to

$$f(Z) = f(Z_0) + f'(Z) \cdot Z + \frac{1}{2}f''(Z) \cdot [Z, Z].$$

Discrete time

Just out of interest, what do all the things we discussed look like in discrete time? If time is indexed by $\mathbb{Z}_+$, processes are just sequences of r.v.'s (X_n). Filtrations are simply increasing sequences $(\mathcal{F}_n)$ of σ-algebras, for which right continuity has no meaning, though $\mathcal{F}_0$ is still assumed to be complete. A map $T : \Omega \to \overline{\mathbb{Z}}_+$ is a stopping time provided $\{T = n\} \in \mathcal{F}_n$ for each n. Similarly, $\mathcal{F}_T = \{A : A \cap \{T = n\} \in \mathcal{F}_n\}$.

Although the notion of càdlàg process has no meaning, we can associate to each X the process X_- given by

$$X_{0-} = X_0, \qquad X_{n-} = X_{n-1},$$

so that

$$\Delta X_n = X_n - X_{n-}.$$

Other notions such as martingales and their properties are valid without change except the obvious discretization. Note that all discrete-time adapted processes are in fact FV processes.

The predictable σ algebras $\mathcal{P}[0,t]$ are the algebras generated by all processes X such that X_0 is $\mathcal{F}_0$-measurable and X_n is $\mathcal{F}_{n-1}$-measurable for each $1 \leq n \leq t$. Every process is naturally summable.

Thus we see that a process X is a semimartingale if and only if it is adapted to the filtration $(\mathcal{F}_n)$.

The stochastic integral $(H \cdot X)$ is defined by

$$(H \cdot X)_n = \sum_{1 \leq p \leq n} H_p(X_p - X_{p-1}) = \sum_{p \leq n} H_p \Delta X_p.$$

Note that when X is a martingale, this process is commonly known as the *martingale transform.*

The quadratic covariation is naturally defined as

$$[X, Y]_n = \sum_{1 \leq p \leq n} (X_p - X_{p-1})(Y_p - Y_{p-1}) = \sum_{p \leq n} \Delta X_p \Delta Y_p,$$

and finally, since for any process $X^c = 0$, Itô's formula reduces to the trivial identity

$$\begin{aligned} f(X_n) &= f(X_0) + \sum_{1 \leq p \leq n} \sum_{i \leq d} D_i f(X_{p-1})(X_p^i - X_{p-1}^i) \\ &\quad + \sum_{1 \leq p \leq n} \left(f(X_p) - f(X_{p-1}) - \sum_{i \leq d} D_i f(X_{p-1})(X_p^i - X_{p-1}^i) \right). \end{aligned}$$

Note that the discrete-time case actually can be imbedded into the continuous-time setup: Suppose $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ is our discrete-time filtered space. Set $\mathcal{F}'_t = \mathcal{F}_n$ whenever $t \in [n, n+1[$. Then $(\mathcal{F}'_t)$ satisfies the usual conditions and $(\Omega, \mathcal{F}, (\mathcal{F}'_t), \mathbb{P})$ is a continuous-time filtered space. With each process (X_n), associate (X'_t), defined by $X'_t = X_n$ whenever $t \in [n, n+1[$. Then X is $(\mathcal{F}_n)$-adapted if and only if X' is $(\mathcal{F}'_t)$-adapted, and X' is clearly right-continuous.

Deterministic processes and the stochastic integral

To end this chapter, here is an interesting question: does the stochastic integral yield more integrators in the deterministic case than the Lebesgue-Stieltjes integral? Unfortunately, the answer is no. Here's the theorem:

THEOREM 31. [JS87] Let $f : \mathbb{R}_+ \to R$. The process $X_t(\omega) = f(t)$ is a semimartingale if and only if f is càdlàg and has finite variation on every compact interval.

PROOF: If X is an FV process, we've seen it is a semimartingale. Conversely, let $X = f(0) + M + A$ be a semimartingale (hence càdlàg), (T_n) a localizing sequence such that $M^{|T_n}$ is a uniformly integrable martingale with $M_0 = 0$, and $A^{|T_n}$ is an IV process. Let also $F_n(dx)$ be the distribution of T_n. Then

$$X_t 1_{\{T_n > t\}} = X_{t \wedge T_n} - X_{T_n} 1_{\{T_n \le t\}},$$

and upon taking expectations, we get

$$\begin{aligned} f(t)\mathbb{P}(T_n > t) &= f(0) + \mathbb{E}M_t{}^{|T_n} + \mathbb{E}A_t{}^{|T_n} - \mathbb{E}\mathbb{E}[X_{T_n}|T_n] \\ &= f(0) + \mathbb{E}A_t{}^{|T_n} - \int_{[0,t]} f(s)dF_n(s). \end{aligned}$$

All terms on the right are functions (in t) of finite variation, while if n is large enough, $\mathbb{P}(T_n > t) > 0$ since $T_n \uparrow \infty$. This shows that $f(t)$ is of finite variation on each $[0, t]$. $\square$

In fact, it is easily seen that a deterministic martingale must be constant. Also, this theorem gives immediately an example of a process which is not a semimartingale: any deterministic process with paths of unbounded variation will do.

Notes and Comments

An excellent reference for measure theory is [Doo93]. The fundamentals of functional analysis can be found in [Rud73].

The definitions of processes, filtrations, stopping times etc. are standard and can be found in any book on stochastic processes. See for example [RY90, Mét82, Wil79, DM78, JS87, Pro92, vWW90, KS88, LS89] to name but a few. The theory associated with these definitions is usually called the *general theory of processes.*

Constructions of $BM^x(\mathbb{R}^d)$ and $PP(\lambda)$ abound, especially that of $PP(\lambda)$ which is much simpler (see [KT81] for instance). The book [Kni81] has many different constructions of $BM^x(\mathbb{R}^d)$, but see also [SV79, Wil79, KS88, RY90]. The proofs that $BM^0(\mathbb{R})$ has paths of infinite variation, along with many other esoteric properties are usually given in those books too. See in particular [Nel67] for a delightful account of the physical theory of Brownian motion.

That naïve stochastic integration is impossible is taken from [Pro92]. A superb, definitely recommended, survey of stochastic calculus is the appendix by P.A. Meyer in [EM89].

There are essentially three different approaches to constructing the stochastic integral. The oldest follows the historical development quite closely. If X is an IV process, $H \cdot X$ is defined as a path-by-path Lebesgue-Stieltjes integral. If X is a square-integrable martingale, the full Doob-Meyer decomposition theorem is used to write $X^2 = M + \langle X, X \rangle$, where $\langle X, X \rangle$ is a predictable, increasing process (when X is continuous, $\langle X, X \rangle = [X, X]$). Writing $\mu(A) = \mathbb{E}[1_A \cdot \langle X, X \rangle]$ on $\mathcal{P}_0$ yields a measure which can be extended to $\mathcal{P}$. Then the equation $\mathbb{E}\langle H \cdot X, H \cdot X \rangle = \mathbb{E} H^2 \cdot \langle X, X \rangle$ yields an isometry from $L^2(\Omega \times \mathbb{R}_+, \mathcal{P}, \mu)$ to the Hilbert space $\mathcal{H}^2$ of all square-integrable martingales, so that we can use $\langle H \cdot X, H \cdot X \rangle = H^2 \cdot \langle X, X \rangle$ for all $H \in L^2(\Omega \times \mathbb{R}_+, \mathcal{P}, \mu)$ to define $H \cdot X$ unambiguously (up to indistinguishability). Proper use of stopping times and decompositions then gives the martingale integral and if $X = M + A$ is a semimartingale, the integral $H \cdot X$ is defined to be $H \cdot M + H \cdot A$. It must then be checked that the process does not depend on the decomposition or the stopping times used. See the classical lecture notes [Mey76], which extended [KW67]. See also [Mét82, JS87, DM80, LS89, vWW90] for the complete theory. The books [CW83, RY90] deal only with continuous semimartingales. I believe the books [RW87, IW81] also belong in this category, though I haven't seen them. The paper [Rog81] is a short, but more detailed review, but see also [Del80], which initiated the third approach to stochastic integration. The original work of Itô used the isometry in the special case of $BM^0(\mathbb{R})$. See his book [Itô51].

The second method for defining the stochastic integral is closely related to the one presented here. One considers a vector integral in Hilbert space (this applies to L^2-martingales, which yield orthogonally scattered measures, see [Mas68]) or in Banach spaces (see [Kus77, Yor78] and most certainly [Pel73, Mét73], though I haven't seen them) or in the Orlicz space $L^0(\Omega, \mathcal{F}, \mathbb{P})$ (initially in [MP77], and after that in [MP80] and more recently in [Kwa92]). The vector integration theory used always seems to be a suitably adapted version of [DS58], section IV.10. This requires some knowledge of functional analysis, especially with L^0-integration theory. To the best of my knowledge, the simplicity (inspired by [KK76]) of using duality to reduce vector integration to real integration is always overlooked, in the context of the stochastic integral. Note however that using duality doesn't work for L^0-measures, since $L^0(\Omega, \mathcal{F}, \mathbb{P})$ has trivial dual.

At the time of this writing, I also became aware of [Rao93], which promises a unified treatment via vector integrals, though I am not quite sure where it fits in the above.

The third approach to the stochastic integral represents a Daniell-type vector integration theory in $L^0(\Omega, \mathcal{F}, \mathbb{P})$. For simple processes, the elementary integral is defined essentially as we did, but now it is considered as a continuous linear map satisfying the DCT into $L^0(\Omega, \mathcal{F}, \mathbb{P})$, topologized by convergence in probability. It is then extended to larger and larger classes of processes and ultimately to locally bounded predictable processes (and beyond!). See [Pro92] for a complete exposition, although [Pro86] is certainly easier to read. This approach was pioneered in the review paper [Del80]. See also [Bic81], which is harder than [Pro92], but well worth it. When doing vector measure theory in $L^0(\Omega, \mathcal{F}, \mathbb{P})$, it is important to have bounded measures. It turns out that every vector measure into $L^0(\Omega, \mathcal{F}, \mathbb{P})$ is bounded, but this is far from trivial; see [KPR82] for example. The paper [Sch80] explores what happens when the measures μ_X are not defined everywhere, as is the case for Lebesgue measure on $\mathbb{R}$, which is infinite on some sets.

On the topic of changing the stochastic basis, the result about shrinking the filtration was originally discovered in [Str77], but was much more difficult to prove, as it considered semimartingales without reference to vector measures. There are some partial results on augmenting the filtration, but no complete characterization, as far as I know. See [Jeu79, JY85] for example. The result about absolutely continuous changes of probability can be much improved. A locally summable process stays locally summable, and there is no need to bound $d\mathbb{Q}/d\mathbb{P}$ at all in that case. I suspect that the proof presented for bounded $d\mathbb{Q}/d\mathbb{P}$ can be localized, but I don't quite see how at the moment. The way the semimartingale decomposition changes with $\mathbb{Q}$ is also known explicitely. It is called Girsanov's theorem. See [Pro92, Mét82, RY90] and many others.

The stochastic DCT is the best reasonable DCT for semimartingales. The convergence in probability can be improved by putting additional assumptions on the integrator. See [Bic81] for a very complete treatment.

The martingale theorems are standard and can be found in any of the books mentioned earlier. The partial Doob-Meyer decomposition was taken from [Kry90], as was the multiplicative decomposition of a submartingale preceding it.

The fact that $PP(\lambda)$ is not predictable is taken from [vWW90]. The treatment of the bracket process essentially follows [Del80]. See also [Pro86, Pro92].

The proof of Itô's formula comes from [JS87], which was inspired by [DM80]. It shows that Itô's formula is really equivalent to the definition of the quadratic variation. Other popular proofs use Taylor series, see [Pro86] for example. One thing I would like to know is if duality can be used to prove Itô's formula via a few applications of the ordinary fundamental theorem of calculus, much like the vector DCT is proved via duality and the ordinary DCT.

Finally the discrete-time case, as well as the relation between semimartingales and deterministic processes comes from [JS87]. In discrete time, the martingale integral is usually known a the *martingale transform.* See [Bur66].

A few elements are missing in this short exposition of stochastic integration; some were omitted on purpose, such as the theories of $\mathcal{H}^p$ semimartingales, compensators and the angle bracket, others because then, Part I would have grown even more, to the point of leaving no room for the applications in Part II, which is, after all, why stochastic integration is done in the first place. This is the case for topics such as product integration and the stochastic Fubini theorem, Hilbert-valued stochastic processes and *their* calculus, the Girsanov theorem. As it is, Part I is already longer than I expected.

Part II

Applications and nice results

The Doléans exponential

Recall that one can define the exponential function e^x as the solution to the differential equation $f'(x) = f(x)$ with initial condition $f(0) = 1$. Here is the *stochastic exponential*:

THEOREM **32.** [JS87] Let X be a semimartingale. Then there exists a unique semimartingale $\mathcal{E}(X)$ satisfying the equation $\mathcal{E}(X) = 1 + \mathcal{E}(X)_- \cdot X$. An explicit formula for $\mathcal{E}(X)$ is

$$\mathcal{E}(X)_t = \exp\left(X_t - X_0 - \frac{1}{2}[X,X]_t^c\right)\prod_{s\le t}(1+\Delta X_s)\exp(-\Delta X_s).$$

PROOF: (i) First, note that $\mathcal{E}(X)$ is indeed a semimartingale. The first factor is clearly one, and we are going to show that the infinite product is an FV process: it is clearly adapted and càdlàg. Since X is càdlàg, there are (for each ω) only finitely many jumps of size greater than $1/2$ on any interval $[0,t]$. So if we write $\widetilde{\Delta X_s} = \Delta X_s 1_{\{|\Delta X_s|<1/2\}}$, it suffices to check that

$$R_t = \prod_{s\le t}(1+\widetilde{\Delta X_s})\exp(-\widetilde{\Delta X_s})$$

converges and is of finite variation for each t. But $|\log(1+x) - x| \le x^2$ whenever $|x| \le 1/2$. So $\log R_t = \sum_{s\le t}\log(1+\widetilde{\Delta X_s}) - \widetilde{\Delta X_s}$ is an absolutely convergent series, for

$$\sum_{s\le t}\left|\log(1+\widetilde{\Delta X_s}) - \widetilde{\Delta X_s}\right| \le \sum_{s\le t}(\widetilde{\Delta X_s})^2 \le [X,X]_t < \infty.$$

Thus $\log R$, and hence R, is an FV process.

Now let's show that $\mathcal{E}(X)$ indeed satisfies the equation: we start by writing $K_t = X_t - X_0 - \frac{1}{2}[X,X]_t^c$ and $V_t = \prod_{s\le t}(1+\Delta X_s)\exp(-\Delta X_s)$, so that $\mathcal{E}(X) = e^K V$. Note also that $[K,V]^c = [V,V]^c = 0$. By Itô's multidimensional formula,

$$\begin{aligned}
\mathcal{E}(X) &= 1 + \mathcal{E}(X)_- \cdot K + e^{K_-}\cdot V + \frac{1}{2}\mathcal{E}(X)_- \cdot [K,K]^c \\
&\quad + \sum_{s\le \cdot}\left(\mathcal{E}(X)_s - \mathcal{E}(X)_{s-} - \mathcal{E}(X)_{s-}\Delta K_s - e^{K_{s-}}\Delta V_s\right) \\
&= 1 + \mathcal{E}(X)_- \cdot X - \frac{1}{2}\mathcal{E}(X)_- \cdot [X,X]^c + e^{K_-}\cdot V + \frac{1}{2}\mathcal{E}(X)_- \cdot [X,X]^c \\
&\quad + \sum_{s\le \cdot}\left(\mathcal{E}(X)_s - \mathcal{E}(X)_{s-} - \mathcal{E}(X)_{s-}\Delta K_s - e^{K_{s-}}\Delta V_s\right).
\end{aligned}$$

Furthermore, $e^{K_-} \cdot V = \sum_{s\le\cdot} e^{K_{s-}} \Delta V_s$, and also $\mathcal{E}(X)_s = \mathcal{E}(X)_{s-}(1+\Delta X_s)$, and $\mathcal{E}(X)_{s-}\Delta K_s = \mathcal{E}(X)_{s-}\Delta X_s$. Substituting into the equation yields

$$\begin{aligned}
\mathcal{E}(X) &= 1 + \mathcal{E}(X)_- \cdot X + \sum_{s\le\cdot} e^{K_{s-}} \Delta S_s \\
&\quad + \sum_{s\le\cdot} \left(\mathcal{E}(X)_{s-}(1+\Delta X_s) - \mathcal{E}(X)_{s-} - \mathcal{E}(X)_{s-}\Delta X_s - e^{K_{s-}}\Delta V_s\right) \\
&= 1 + \mathcal{E}(X)_- \cdot X.
\end{aligned}$$

(ii) Uniqueness: Suppose the semimartingale Y also satisfies $Y = 1 + Y_- \cdot X$. Apply Itô's formula to $W = e^{-K}Y$:

$$\begin{aligned}
W &= 1 - W_- \cdot K + e^{-K_-} \cdot Y + \frac{1}{2} W_- \cdot [K,K]^c - e^{-K_-} \cdot [K,Y]^c \\
&\quad + \sum_{s\le\cdot} \left(W_s - W_{s-} + W_{s-}\Delta K_s - e^{-K_{s-}}\Delta Y_s\right),
\end{aligned}$$

and since Y is also a solution, we have $\Delta Y = Y_- \Delta X$; also $[K,Y]^c = Y_- \cdot [X,X]^c$ and $\Delta W_s = W_{s-}\left(e^{-\Delta X_s}(1+\Delta X_s) - 1\right)$. Putting all this into the equation yields

$$\begin{aligned}
W &= 1 - W_- \cdot X + e^{-K_-} \cdot Y + \sum_{s\le\cdot} \left(W_{s-}e^{-\Delta X_s}(1+\Delta X_s) - 1 + \Delta K_s - \Delta K_s\right) \\
&= 1 - W_- \cdot X + W_- \cdot X + \sum_{s\le\cdot} (W_{s-}e^{-\Delta X_s}(1+\Delta X_s) - 1) \\
&= 1 + W_- \cdot A,
\end{aligned}$$

where $A = \sum_{s\le\cdot}(e^{-\Delta X_s}(1+\Delta X_s) - 1)$. Taking logarithms as previously for V, we can show that A is an FV process.

Now observe that $Z = W - V$ satisfies the equation $Z = Z_- \cdot A$. Thus if $S = \inf\{t : Z_t \neq 0\}$, then $Z_S = 0$ on $\{S < \infty\}$. Choose $T \ge S$ such that $\{S < \infty\} \subset \{T > S\}$ and

$$\int_{]S,T]} |dA_s| \le 1/2.$$

Then

$$Z_t = Z_S + (Z_- 1_{]S,t]} \cdot A) = \int_{]S,t]} Z_{s-} dA_s,$$

and thus $\sup_{t\le T}|Z_t| \le \frac{1}{2}\sup_{t\le T}|Z_t|$, which clearly implies $\sup_{t\le T}|Z_t| = 0$. Since $T > S$ on $\{S < \infty\}$, it follows that $S = +\infty$. In other words, $W_t = V_t$ for all t, which means $Y = \mathcal{E}(X)$. $\square$

As with matrix exponentials, $e^x e^y \neq e^{(x+y)}$ in general. In fact

$$\begin{aligned}\mathcal{E}(X)\mathcal{E}(Y) &= 1 + \mathcal{E}(X)_- \cdot \mathcal{E}(Y) + \mathcal{E}(Y)_- \cdot \mathcal{E}(X) + [\mathcal{E}(X), \mathcal{E}(Y)] \\ &= 1 + \mathcal{E}(X)_-\mathcal{E}(Y)_- \cdot Y + \mathcal{E}(Y)_-\mathcal{E}(X)_- \cdot X + \mathcal{E}(X)\mathcal{E}(Y) \cdot [X, Y] \\ &= \mathcal{E}(X + Y + [X, Y]),\end{aligned}$$

which is quite easy to remember. As a trivial consequence we have

$$\mathcal{E}(X)^{-1} = \mathcal{E}(-X + [X, X]).$$

If X is a continuous process with $X_0 = 0$, the exponential clearly reduces to

$$\mathcal{E}(X)_t = \exp(X_t - \frac{1}{2}[X, X]_t).$$

Linear SDEs

The next result is as close as we will come to stochastic differential equations. A *linear stochastic differential equation (SDE)* is an integral equation

$$Y_t = H_t + \int_0^t Y_{s-} dX_s,$$

where H and X are two given semimartingales. Such an equation is usually written in differential notation

$$dY_t = dH_t + Y_{t-} dX_t, \qquad Y_0 = H_0,$$

which accounts for its name. A *solution* is any semimartingale Y satisfying the equation. Our first application of the exponential is an explicit solution to the linear SDE when H and X are continuous.

THEOREM **33.** [RY90] The semimartingale

$$Y_t = \mathcal{E}(X)_t \left(H_0 + \int_0^t \mathcal{E}(X)_s^{-1} (dH_s - d[H, X]_s) \right)$$

is a solution to the linear SDE (it can be shown to be the unique solution).

PROOF: By Itô's formula,

$$
\begin{aligned}
Y \cdot X &= H_0 \mathcal{E}(X) \cdot X + \mathcal{E}(X) \left(\mathcal{E}(X)^{-1} \cdot (H - [H,X]) \right) \cdot X \\
&= -H_0 + H_0 \mathcal{E}(X) + \mathcal{E}(X) \left(\mathcal{E}(X)^{-1} \cdot (H - [H,X]) \right) \\
&\quad - \mathcal{E}(X) \left(\mathcal{E}(X)^{-1} \cdot H - \mathcal{E}(X)^{-1} \cdot [H,X] \right) - [\mathcal{E}(X), \mathcal{E}(X)^{-1} \cdot (H - [H,X])] \\
&= Y - H + [H,X] - [\mathcal{E}(X) \cdot X, \mathcal{E}(X)^{-1} \cdot H] \\
&= Y - H,
\end{aligned}
$$

and the result follows. □

Perhaps the first SDE to be studied was the *Langevin equation*

$$dV_t = dB_t - \beta V_t dt,$$

where B is a $BM^0(\mathbb{R})$ and β is a real constant. The process V represents the speed of a physical particle undergoing Brownian motion in a medium with friction coefficient β. The solution starting at v is thus given by

$$V_t = e^{-\beta t} \left(v + \int_0^t e^{\beta s} dB_s \right).$$

Lévy's characterization of $BM^0(\mathbb{R})$

Let's now go back to the mathematical Brownian motion. The following is a famous result due to P. Lévy:

THEOREM **34.** [Pro92] A stochastic process X with $X_0 = x$ is a $BM^x(\mathbb{R})$ if and only if it is a local martingale with $[X,X]_t = t$.

PROOF: By considering the process $B - x$, it suffices to consider the case $x = 0$. We've already seen that any $BM^0(\mathbb{R})$ is a continuous local martingale such that $[X,X]_t = t$. We need to prove the converse.

Since the bracket is continuous, the process X must be too. Now apply Itô's formula to the C^2 function $f(x,y) = \exp(iux + u^2 y/2)$, where $u \in \mathbb{R}$.

$$e^{iuX_t + u^2 t/2} = 1 + iu \int_0^t e^{iuX_s + u^2 s/2} dX_s,$$

where we have used the fact $[X,X]_t = t$. This is just the exponential equation $\mathcal{E}(iuX) = 1 + \mathcal{E}(iuX) \cdot (iuX)$. Since (iuX) is a (complex) continuous local martingale, we see that $\mathcal{E}(iuX)$ is one also. Furthermore,

$$\sup_{s \le t} |\mathcal{E}(iuX)_s| = \sup_{s \le t} |\exp(iuX_s + u^2 s/2)| = e^{u^2 t/2},$$

which shows that $\mathcal{E}(iuX)$ is bounded on each $[0,t]$, and hence a true martingale. Then for any $t \geq s$,

$$\begin{aligned}
\mathbb{E}[e^{iu(X_t - X_s)}|\mathcal{F}_s] &= \mathbb{E}[e^{-u^2(t+s)/2}\mathcal{E}(iuX)_t\mathcal{E}(-iuX)_s|\mathcal{F}_s] \\
&= e^{-u^2(t+s)/2}\mathcal{E}(-iuX)_s\mathbb{E}[\mathcal{E}(iuX)_t|\mathcal{F}_s] \\
&= e^{-u^2(t-s)/2},
\end{aligned}$$

which shows that the increment $(X_t - X_s)$ is independent of $\mathcal{F}_s$ and has a $N(0, t-s)$ distribution. So X is a $BM^0(\mathbb{R})$. □

Lévy's characterization also has a multidimensional version. Suppose B is a $BM^x(\mathbb{R}^d)$. This means that $B_t = (B_t^1, \ldots, B_t^d)$, where the B^i $(i = 1, \ldots, d)$ are independent $BM^{x^i}(\mathbb{R})$s. For this process, we have $[B^i, B^j] = \delta^{ij}t$, where of course $\delta^{ij} = 1$ if $i = j$ and is zero otherwise. This is easy to see from the polarization formula

$$[B^i, B^j] = \frac{1}{4}\left([B^i + B^j, B^i + B^j] - [B^i - B^j, B^i - B^j]\right),$$

once it is observed that $(B^i + B^j)/\sqrt{2}$ and $(B^i - B^j)/\sqrt{2}$ are standard $BM^0(\mathbb{R})$s.

THEOREM **35.** [RY90] A d-dimensional local martingale $X = (X^1, \ldots, X^d)$ such that $X_0 = x$ is a $BM^x(\mathbb{R}^d)$ if and only if $[X^i, X^j] = \delta^{ij}t$.

A characterization of $PP(\lambda)$

We've seen earlier that if N is a $PP(\lambda)$, then $[N, N] = N$. This implies that if $X_t = N_t - \lambda t$ is the Poisson martingale, we have $[X, X] = X + \lambda t$. The following theorem is a converse:

THEOREM **36.** [Mét82] A càdlàg local martingale X with $X_0 = 0$ satisfies

$$[X, X]_t = X_t + \lambda t$$

if and only if $X_t = N_t - \lambda t$, where N is a $PP(\lambda)$.

PROOF: We prove only sufficiency. Let t be a jump time of X. Then

$$\Delta[X, X]_t = (\Delta X_t)^2 = \Delta(X_t + \lambda t) = \Delta X_t,$$

so that $\Delta X_t = +1$. Also, because $X_t = [X,X]_t - \lambda t$ is an FV process, the bracket $[X,X]_t = \sum_{s\le t} \Delta X_s$ is constant between jumps. Let $\tau_n = \inf\{t : [X,X]_t = n\}$ be the time of the nth jump. If we write Itô's formula for e^{iuX_t} when $t = \tau_n$ and $t = \tau_{n-1}$, then subtract and use the relations $[X,X]^c = 0$, $X_{\tau_n -} = X_{\tau_{n-1}} - \lambda(\tau_n - \tau_{n-1})$, we get after a little rearranging

$$e^{iuX_{\tau_{n-1}}}\left[(1+iu)e^{-iu\lambda(\tau_n-\tau_{n-1})} - 1\right] = iu\left[(e^{iuX_-}\cdot X)_{\tau_n} - (e^{iuX_-}\cdot X)_{\tau_{n-1}}\right].$$

Now let S_n be a localizing sequence so that $X^{|S_n}$ is a martingale. We have

$$\mathbb{E}[X,X]_{t\wedge S_n} = \mathbb{E}\left(X_{t\wedge S_n} + \lambda(t\wedge S_n)\right) = \lambda\mathbb{E}(t\wedge S_n),$$

and as $n\to\infty$, we conclude, by monotone convergence, that $\mathbb{E}[X,X]_t = \lambda t < \infty$, and X is a true martingale. Since also $|e^{iuX_-}| \le 1$, we see that $(e^{iuX_-}\cdot X)$ is an L^2-martingale. This means that

$$\mathbb{E}[e^{-iu\lambda(\tau_n-\tau_{n-1})}|\mathcal{F}_{\tau_{n-1}}] = \frac{1}{1+iu},$$

so that the waiting times $(\tau_n - \tau_{n-1})$ are independent of $\mathcal{F}_{\tau_{n-1}}$ and have an exponential distribution with parameter λ. This shows that the process

$$N_t = [X,X]_t = \sum_n 1_{[\tau_n,\infty[}$$

is a $PP(\lambda)$. □

Time changes

Lévy's characterization brings up the following reasoning: Suppose (X_t) is a local martingale with $[X,X]_t = f(t)$ for some increasing function f with inverse g. Then the process $Y_t = X_{g(t)}$ has $[Y,Y]_t = t$ and hence is a $BM^0(\mathbb{R})$. So X was a $BM^0(\mathbb{R})$ running at a different speed. More generally, since $[X,X]$ depends on ω, each sample path of X would be a path of $BM^0(\mathbb{R})$ running at its own speed. This reasoning will now be made precise.

A *time change* C is a family $(C_s)_{s\ge 0}$ of stopping times such that for almost every ω, the map $s\mapsto C_s(\omega)$ is increasing and right continuous. A little bit of thought shows that if we set

$$A_t = \inf\{s : C_s > t\},\quad t\in\mathbb{R}_+,$$

then we have $C_{A_t} = t$ for all t. $(A_t)_{t\geq 0}$ is also a time-change, with jumps occurring when C is flat.

The process X is said to be C*-continuous* if it is constant on each interval $[C_{t-}, C_t]$. In that case, the process X_C is continuous.

We will always assume $C_\infty = \infty$, so that $A_\infty = \infty$ also and hence C_t, $A_t < \infty$ for $t < \infty$. We also assume $C_0 = 0$.

Now let X be an $(\mathcal{F}_t)$-adapted process. We call $(\mathcal{F}_{C_t})_{t\geq 0}$ the *time-changed filtration*, and X_C, defined by

$$(X_C)_t(\omega) = X_{C_t(\omega)}(\omega) = (X \circ \theta)(t, \omega),$$

where $\theta : (t, \omega) \mapsto (C_t(\omega), \omega)$, is called the *time-changed process.*

As was seen earlier, we have $X_{C_t} \in \mathcal{F}_{C_t}$, that is X_C is $(\mathcal{F}_{C_t})$-adapted.

The map $\sigma : (t, \omega) \mapsto (A_t(\omega), \omega)$ is a measurable transformation from the space $([0,t], \mathcal{P}[0,t])$ into $([0, A_{t+}], \mathcal{P}[0, A_{t+}])$, since $\sigma^{-1}(]S,T]) =]C_S, C_T]$ and also $\sigma^{-1}([S,0]) = [C_S, C_0]$. Now for any two stopping times S,$T \leq t$,

$$\begin{aligned}
(\langle x', \mu_{X_C}\rangle \circ \sigma)\,(]S,T]) &= \langle x', \mu_{X_C}\rangle(]A_S, A_T]) \\
&= \langle x', X_T - X_S\rangle \\
&= \langle x', \mu_X\rangle(]S,T]),
\end{aligned}$$

and similarly for stochastic intervals $[S,0]$. So by uniqueness, $\langle x', \mu_{X_C}\rangle \circ \sigma = \langle x', \mu_X\rangle$ on $\mathcal{P}[0,t]$. Hence we have

$$\int_{[0,A_{t+}]} (H \circ \theta) d\langle x', \mu_{X_C}\rangle = \int_{[0,t]} H d\langle x', \mu_X\rangle,$$

which means that $H_C \cdot X_C = (H \cdot X)_C$. The stochastic integral is stable under continuous time changes.

Time changes of Brownian motion

THEOREM 37. [RY90] If M is a continuous local martingale such that $M_0 = x$ and $[M,M]_\infty = \infty$, and if we set

$$T_t = \inf\{s : [M,M]_s > t\},$$

then $B_t = M_{T_t}$ is a $BM^x(\mathbb{R})$ relative to the filtration $(\mathcal{F}_{T_t})$, and $M_t = B_{[M,M]_t}$.

PROOF: Each T_s is a.s. finite since $[M,M]_\infty = \infty$, and if $s_n \downarrow s$, then $T_{s_n} \downarrow T_s$ by continuity of $[M,M]$. so T is a time-change. The process M is T-continuous, for if $[M,M]$ is constant on a stochastic interval $[U,V]$, the Riemann approximation shows that M must be constant on $[U,V]$. Also,

$$[B,B]_t = M_{T_t}^2 - M_0^2 - 2(M_T \cdot M_T)_t = [M,M]_{T_t} = t,$$

and although in general $T_{[M,M]_t} \geq t$, the equation $B_{[M,M]t} = M_{T_{[M,M]_t}} = M_t$ follows because M is constant when $[M,M]$ is.

It remains to check that B is a (fil_{T_t})-local martingale. For this, let (S_n) be a localizing sequence such that $M^{|S_n}$ is a u.i. martingale. If $R_n = \inf\{t : T_t \geq S_n\}$, then R_n is an $(\mathcal{F}_{T_t})$-stopping time and $B_t^{|R_n} = M_{T_{t \wedge R_n}} = M_{T_t}^{|S_n}$ is a martingale by the stopping theorem. But clearly $R_n \uparrow \infty$, which completes the proof. □

What happens if $\mathbb{P}([M,M]_\infty < \infty) > 0$? Intuitively, we expect to get a stopped $BM^x(\mathbb{R})$. This is indeed the case, but there is a slight difficulty. Suppose $\Omega = \{\omega\}$ consists of only one point. Then all our stochastic processes are deterministic, and $\mathbb{R}_+ \times \Omega$ clearly doesn't support a $BM^0(\mathbb{R})$. But any constant process $M_t(\omega) = c$ for some constant c is a local martingale with $[M,M]_t = 0$. How can we say in this case that M is a stopped $BM^c(\mathbb{R})$? The answer is to enlarge the probability space.

Take any stochastic basis $(\Omega', \mathcal{F}', (\mathcal{F}'_t), \mathbb{P}')$ which supports a $BM^0(\mathbb{R})$, B'. Let

$$\overline{\Omega} = \Omega \times \Omega', \quad \overline{\mathcal{F}_t} = \mathcal{F}_{T_t} \otimes \mathcal{F}'_t, \quad \overline{\mathbb{P}} = \mathbb{P} \otimes \mathbb{P}',$$

and let $\overline{B_t}(\omega, \omega') = B_t(\omega)$. Then the process $\overline{B}$ is independent of M and we can define

$$B_t = M_{T_t} + \int_0^t 1_{\{s > [M,M]_\infty\}} d\overline{B_s},$$

so that

$$[B,B]_t = [M_T, M_T]_t + \int_0^t 1_{\{s > [M,M]_\infty\}} ds = t.$$

This shows that B is a $BM^0(\mathbb{R})$, and of course we have

$$B^{|[M,M]_\infty} = M_T,$$

which means M_T is a $BM^0(\mathbb{R})$ stopped by $[M,M]_\infty$.

Here again, there is an extension to the multidimensional case:

THEOREM 38. [RY90] Let $M = (M^1, \ldots, M^d)$ be a continuous vector-valued local martingale such that $M_0 = x$, $[M^k, M^k]_\infty = \infty$ for every k and $[M^i, M^j] = 0$ for $i \neq j$. If we set

$$T_t^k = \inf\{s : [M^k, M^k]_s > t\},$$

and $B_t^k = M^k_{T_t^k}$, the process $B = (B^1, \cdots, B^d)$ is a $BM^x(\mathbb{R}^d)$.

Observe that, while we don't assume anything about the components of M, the components of B are independent!

Recurrence of $BM^x(\mathbb{R}^d)$

Now let B be a $BM^{x_0}(\mathbb{R}^d)$ and consider the ball $K = \{x : |x| < R\}$. We would like to know when B first leaves K, assuming $|x_0| < R$. Let $T = \inf\{t : |B_t| > R\}$. By Itô's formula,

$$|B_{T\wedge n}|^2 = |B_0|^2 + 2\sum_{i=1}^{d}(B_-^i \cdot B^i)_{T\wedge n} + d(T \wedge n).$$

Now for each i, the process $(B_-^i \cdot B^i)^{|T}$ is a martingale, since its quadratic variation is bounded by $R^2 t$. Thus it will disappear when we take expectations. So we get

$$\mathbb{E}|B_{T\wedge n}|^2 = |x_0|^2 + d\mathbb{E}(T \wedge n).$$

But because $|B_{T\wedge n}| \leq R$, we have $\mathbb{E}(T \wedge n) \leq d^{-1}(R^2 - |x_0|^2)$ for each n, so letting $n \to \infty$ produces the equation

$$\mathbb{E}T = d^{-1}(R^2 - |x_0|^2).$$

We are also interested in knowing when B reenters K. So suppose $|x_0| > R > 0$, and for each n consider the annulus $A_k = \{x : R < |x| < 2^k R\}$. The stopping times $S_k = \inf\{t : B_t \notin A_k\} \uparrow S$, where S is the first hitting time of K. For any C^2 function f (which may not be defined throughout K), we have by Itô's formula

$$f(B_{S_k}) = f(x_0) + \sum_{i=1}^{d}(D_i f(B_-) \cdot B)_{S_k} + \int_0^{S_k} \Delta f(B_{s-})ds.$$

So if we choose f to be a C^2 function such that for $|x| > R/2$,

$$f(x) = \begin{cases} x & \text{if } d = 1, \\ -\log|x| & \text{if } d = 2, \\ |x|^{2-d} & \text{if } d > 2, \end{cases}$$

then $\Delta f = 0$ on each A_k and the variables $(D_i f(B) \cdot B)_{S_k}$ again have zero expectation. Hence we get for each k

$$\mathbb{E} f(B_{S_k}) = f(x_0).$$

Explicitly for each d, this equation says (where we set $p_k = \mathbb{P}(|B_{S_k}| = R)$ and $q_k = \mathbb{P}(|B_{S_k}| = 2^k R)$)

- $d = 1$: $p_k R + q_k 2^k R = x_0$,
- $d = 2$: $p_k \log R + q_k(\log R + k \log 2) = \log|x_0|$,
- $d > 2$: $p_k R^{2-d} + q_k(2^k R)^{2-d} = |x_0|^{2-d}$,

and as $k \to \infty$, this implies $q_k \to 0$ when $d \leq 2$ and also $p_k \to (|x_0|/R)^{2-d}$ when $d > 2$.

We have shown that $BM^{x_0}(\mathbb{R}^d)$ is neighbourhood recurrent when $d \leq 2$ and transient when $d > 2$! It can be shown, however, that unless $d = 1$, the $BM^x(\mathbb{R}^d)$ never returns to any given single point.

Random generalized functions

Let's recall some terminology from the theory of Schwartz distributions (generalized functions). We denote by $\mathcal{D} = \mathcal{D}(\mathbb{R}) = C_K^\infty(\mathbb{R})$ the space of C^∞ functions $\varphi : \mathbb{R} \to \mathbb{R}$ which have compact support. The following function is easily seen to be in $\mathcal{D}$:

$$\varphi(x) = \begin{cases} \exp(-\frac{1}{1-x^2}) & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases}$$

Note that it clearly satisfies $\int_{\mathbb{R}} \varphi(x)dx < \infty$. So the function ρ, defined by

$$\rho(x) = \frac{\varphi(x)}{\int_{\mathbb{R}} \varphi(x)dx},$$

integrates to one.

$\mathcal{D}$ is a topological vector space with topology such that a sequence (φ_n) converges to φ in $\mathcal{D}$ if and only if there exists a fixed compact subset $K \subset \mathbb{R}$ containing the supports of all the φ_k such that $(\varphi_n - \varphi)$ and its derivatives of all orders converge uniformly to zero on K.

A *Schwartz distribution* T is an element of $\mathcal{D}'$, the topological dual of $\mathcal{D}$, so that whenever $\varphi_n \to 0$ in $\mathcal{D}$, we have $\langle T, \varphi_n \rangle \to 0$. Here as usual, $\langle T, \varphi \rangle$ denotes the value of T on φ.

Here are a few classic examples of Schwartz distributions:

Given a Radon measure (finite on compact sets) μ on $\mathbb{R}$, we get a distribution T_μ by

$$\langle T_\mu, \varphi \rangle = \int_{\mathbb{R}} \varphi d\mu.$$

In particular, if f is a locally Lebesgue-integrable function, we get a distribution T_f by

$$\langle T_f, \varphi \rangle = \int_{\mathbb{R}} \varphi(x) f(x) dx.$$

We usually denote T_f simply by f.

Our third example is the important *Dirac delta function* δ_x. It is the distribution defined by

$$\langle \delta_x, \varphi \rangle = \varphi(x).$$

By analogy with the case $T = T_f$, we usually define the symbol $\delta(x)$ by

$$\int_{\mathbb{R}_+} \delta(x-y)\varphi(y)dy = \langle \delta_x, \varphi \rangle,$$

so that we have symbolically $\delta = \delta_0 = T_\delta$.

Schwartz distributions have an interesting physical interpretation: exact measurements being impossible, if we were to measure the value of a function f on $\mathbb{R}$ in a physical experiment, the instrument used would allow us to get only an average value

$$\langle f, \varphi \rangle = \int_{\mathbb{R}} \varphi(x) f(x) dx,$$

different measuring instruments being characterized by different functions φ.

The functions $\rho_n(x) = n\rho(nx)$ have shrinking support around $x = 0$. So they might represent better and better measuring instruments. The sequence

(ρ_n), called a *mollifier*, has a very useful property: given a locally integrable function f, we can convolve it with ρ_n to get

$$f_n(x) = (\rho_n * f)(x) = \int_{\mathbb{R}} f(x-y)\rho_n(y)dy.$$

The functions f_n are C^∞ and converge to f *in* $\mathcal{D}$.

In contrast to ordinary functions, distributions can be differentiated infinitely often as follows. Let D denote differentiation, then for all $\varphi \in \mathcal{D}$, set

$$\langle DT, \varphi \rangle = -\langle T, D\varphi \rangle.$$

Note that if $T = T_f = f$ is an ordinary function, then

$$\begin{aligned}
\langle Df, \varphi \rangle &= -\langle T_f, D\varphi \rangle \\
&= -\int_{\mathbb{R}} \varphi'(x) f(x) dx \\
&= \int_{\mathbb{R}} \varphi(x) f'(x) dx \\
&= \langle f', \varphi \rangle
\end{aligned}$$

by integration by parts, since φ vanishes outside a compact set. This shows that differentiation in the sense of distributions is a generalization of ordinary differentiation.

Another classic example is the following. Let H_x be the step function defined by $H_x(s) = 1_{[x,\infty[}(s)$. This function is not differentiable at x, but if we consider it as a Schwartz distribution, we find

$$\begin{aligned}
\langle DH_x, \varphi \rangle &= -\langle H_x, D\varphi \rangle \\
&= -\int_x^\infty \varphi'(s) ds \\
&= \varphi(x) \\
&= \langle \delta_x, \varphi \rangle,
\end{aligned}$$

which gives us an interpretation of δ_x: if it were a function, it would be infinite at x, zero outside $\{x\}$, and its total mass would be

$$\int_{\mathbb{R}_+} \delta_x(s) ds = \langle \delta_x, \varphi \rangle = 1,$$

where $\varphi \in \mathcal{D}$ is such that $\varphi(x) = 1$. Note that the function H_x is classically differentiable almost everywhere, with derivative zero, which agrees with δ almost everywhere. Finally, observe that the mollifier (ρ_n) tends to the delta function as n tends to infinity.

Now let X be a semimartingale. Any function $\varphi \in \mathcal{D}$ with support contained in $]0, \infty[$ is obviously a deterministic FV process, so that it can be viewed as a semimartingale. Then we can write

$$[X, \varphi]_t = X_t\varphi(t) - X_0\varphi(0) - \int_0^t \varphi(s)dX_s - \int_0^t X_s\varphi'(s)ds.$$

Since φ is FV, the bracket vanishes and if we let $t \to \infty$, we get

$$\int_0^\infty \varphi(s)dX_s = -\int_0^\infty X_s\varphi'(s)ds.$$

Thus if we regard X as a random Schwartz distribution, then stochastic integration corresponds to differentiation in the sense of distributions. This provides a simple, but of course not very powerful, way of defining the stochastic integral.

White noise and the $BM^0(\mathbb{R})$

It is interesting to look more closely at the case when X is a $BM^0(\mathbb{R})$. By Fubini's theorem, we can define the expectation of X in the sense of distributions (where both φ, $\psi \in \mathcal{D}(]0, \infty[)$) by

$$\langle \mathbb{E}X, \varphi\rangle = \mathbb{E}\langle X, \varphi\rangle = \int_{\mathbb{R}_+} \mathbb{E}X_s\varphi(s)ds = 0,$$

and its covariance $\mathrm{cov}X$ by

$$\begin{aligned}\langle \mathrm{cov}X, \varphi \otimes \psi\rangle &= \mathbb{E}\left(\langle X - \mathbb{E}X, \varphi\rangle\langle X - \mathbb{E}X, \psi\rangle\right)\\ &= \int_{\mathbb{R}_+} ds \int_{\mathbb{R}_+} dt\ \varphi(s)\psi(t)(\mathrm{cov}[X_s, X_t])\\ &= \int_{\mathbb{R}_+} ds \int_{\mathbb{R}_+} dt\ \varphi(s)\psi(t)(s \wedge t).\end{aligned}$$

Now lets look at the distributional derivative $\xi = DX$ of X:

$$\langle \mathbb{E}\xi, \varphi\rangle = \mathbb{E}\langle DX, \varphi\rangle = -\langle \mathbb{E}X, D\varphi\rangle = 0,$$

and

$$
\begin{aligned}
\langle \mathrm{cov}\xi, \varphi \otimes \psi \rangle &= \langle \mathrm{cov}X, D\varphi \otimes D\psi \rangle \\
&= \int_{\mathbb{R}_+} ds \int_{\mathbb{R}_+} dt\ \varphi'(s)\psi'(t)(s \wedge t) \\
&= \int_{\mathbb{R}_+} ds\varphi'(s) \left(\int_0^s \psi'(t)dt + s\int_s^\infty \psi'(t)dt \right) \\
&= -\int_{\mathbb{R}_+} ds\ \varphi'(s) \int_0^s \psi(t)dt \\
&= -\int_{\mathbb{R}_+} dt\ \psi(t) \int_t^s \varphi'(s)ds \\
&= \int_{\mathbb{R}_+} \varphi(t)\psi(t)dt \\
&= \int_{\mathbb{R}_+ \times \mathbb{R}_+} \delta(s-t)\varphi(s)\psi(t)dsdt.
\end{aligned}
$$

In other words, the generalized process $(\xi_t) = DX$ has

$$
\begin{aligned}
\mathbb{E}\xi_t &= 0, \\
\mathrm{cov}[\xi_s, \xi_t] &= \delta(t-s),
\end{aligned}
$$

so it is weakly stationary and uncorrelated. Finally, we can take a look at its "generalized spectral density". For any $\varphi \in \mathcal{D}$, let

$$
\check{\varphi} = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx}\varphi(t)dt.
$$

Then

$$
\langle \check{\delta}, \varphi \rangle = \langle \delta, \check{\varphi} \rangle = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t)dt = \langle \frac{1}{2\pi}, \varphi \rangle
$$

and we see that the spectral density of ξ should be constant. This is why ξ is interpreted as white noise: every spectral frequency is equally represented. We refer to [GV68] for a much more complete treatment of generalized random processes.

Itô's formula for convex functions

If f is not a C^2 function, it is nevertheless often possible to approximate it by C^2 functions, even C^∞ functions for that matter. The question then is whether Itô's formula is still true when passing to the limit. One constraint is certainly that $f(X)$ should again be a semimartingale. The following theorem shows that if f is a difference of convex functions, Itô's formula may be extended.

THEOREM **39.** [RY90] Let $f : \mathbb{R} \to \mathbb{R}$ be convex, and let X be a semimartingale. Then $f(X)$ is again a semimartingale and there exists an adapted, càdlàg increasing process A^f such that (here $D_- f$ is the left derivative of f)

$$f(X) = f(X_0) + D_- f(X_-) \cdot X + A^f.$$

Moreover, the jumps of A^f are given by

$$\Delta A_t = f(X_t) - f(X_{t-}) - D_- f(X_{t-}) \Delta X_t.$$

PROOF: The proof is along the same lines as the proof of the standard Itô formula. By stopping, we may assume that X is bounded during longer and longer stochastic intervals $[0, T_n[$. In that case, both $f(X)$ and $D_- f(X)$ are also bounded.

Now choose a mollifier (φ_n) with compact support in $]-\infty, 0]$ and let

$$f_n = \int_{\mathbb{R}} f(x+y)\varphi_n(y)dy.$$

Each f_n is convex and C^2, and the f'_n increase to $D_- f$. Itô's formula yields

$$f_n(X) = f_n(X_0) + f'_n(X_-) \cdot X + A^{f,n},$$

where

$$A^{f,n} = \sum_{s \leq \cdot} (f_n(X_s) - f_n(X_{s-}) - f'_n(X_{s-})\Delta X_s) + \frac{1}{2} f''_n(X_-) \cdot [X, X]^c.$$

Since f is convex, $A^{f,n}$ is clearly increasing. Applying SDCT now gives the theorem. It remains only to find the jumps of $A^f = \lim_n A^{f,n}$. This is easy: from the equation

$$f(X_t) - f(X_0) = (D_- f(X_-) \cdot X)_t + A^f_t,$$

we get by applying the operator Δ

$$f(X_t) - f(X_{t-}) = D_- f(X_{t-}) \Delta X_t + \Delta A^f_t.$$

$\square$

The functions $x \mapsto x^{\pm}$, $x \mapsto |x| = x^+ - x^-$ are all convex, so that given a semimartingale X, the processes $|X|$, X^+ and X^- are semimartingales again.

Furthermore, since $x \wedge y = \frac{1}{2}(x+y-|x-y|)$ and $x \vee y = \frac{1}{2}(|x-y|+x+y)$, we see that the infimum and supremum of two semimartingales are again semimartingales. Thus semimartingales actually form a *lattice*.

Local times

Fix a semimartingale X. If we take $f(x) = |x-a|$, then f is convex and the convex Itô formula applies. For each a, define the process L^a to be the continuous part of the increasing process A^f. If we assume X continuous, we then have the nice formula

$$|X-a| \;=\; |X_0 - a| + \mathrm{sgn}_-(X-a)\cdot X + L^a,$$

where sgn_- is the left-derivative of the function $x \mapsto |x|$, i.e.

$$\mathrm{sgn}_-(x) = \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x \le 0. \end{cases}$$

The process (L^a_t) is called the *local time of X at a*. Since it is increasing, it represents a random measure $L^a(dt)$ on $\mathbb{R}_+$. The following theorem gives an indication of why L^a is called the local time.

THEOREM **40.** [RY90] The measure $L^a(dt)$ is a.s. carried by the set $\{t : X_t = a\}$.

PROOF: Recall that X is assumed continuous. We know that $|X-a|$ is a semimartingale. By Itô's formula and the definition of L^a,

$$\begin{aligned}(X-a)^2 &= (X_0-a)^2 + 2|X-a|\cdot|X-a| + [|X-a|,|X-a|] \\ &= (X_0-a)^2 + 2|X-a|\cdot\mathrm{sgn}_-(X-a)\cdot X + 2|X-a|\cdot L^a + [X,X].\end{aligned}$$

But Itô's formula also gives directly

$$(X-a)^2 = (X_0-a)^2 + 2(X-a)\cdot X + [X,X],$$

so that by comparing the two formulas, we get a.s.

$$\int_0^t |X_s - a| L^a(ds) = 0,$$

and the result follows. $\square$

Thus if the measure $L^a(dt)$ measures something, it only measures it when $X_t = a$. What happens if t is fixed? Here is a glimpse at an amazing generalization of the Itô formula.

First of all, it is possible to show that the function $\tilde{L}(a,t,\omega) = L_t^a(\omega)$ is actually $\mathcal{B}(\mathbb{R}) \otimes \mathcal{P}$-measurable.

Next, if f is convex, then its derivative in the sense of distributions is

$$\begin{aligned}
\langle Df, \varphi \rangle &= -\int_{\mathbb{R}} \left(\lim_{\epsilon \downarrow 0} \frac{\varphi(x+\epsilon) - \varphi(x)}{\epsilon} \right) f(x)dx \\
&= -\lim_{\epsilon \to 0} \int_{\mathbb{R}} \varphi(x) \left(\frac{f(x-\epsilon) - f(x)}{\epsilon} \right) dx \\
&= \int_{\mathbb{R}} \varphi(x) D_- f(x) dx.
\end{aligned}$$

By integration by parts again, we see that the second derivative $f'' = D^2 f$ of f in the sense of distributions is the positive measure associated with the increasing function $D_- f$.

Now here is the theorem:

THEOREM 41. [Pro92] Let f be the difference of two convex functions, $D_- f$ its left derivative, and f'' the signed measure corresponding to the second derivative of f in the sense of distributions. For any semimartingale, $f(X)$ is again a semimartingale and if (L_t^a) denotes the local time of X at a, then

$$\begin{aligned}
f(X) &= f(X_0) + D_- f(X_-) \cdot X + \frac{1}{2} \int_{\mathbb{R}} L^a f''(da) \\
&\quad + \sum_{s \leq \cdot} \left(f(X_s) - f(X_{s-}) - D_- f(X_{s-}) \Delta X_s \right).
\end{aligned}$$

PROOF: Omitted. $\square$

Here are a few consequences. If f is C^2, then by comparing the formula above with the earlier Itô formula gives

$$\int_{\mathbb{R}} L_t^a f''(a) da = \int_0^t f''(X_{s-}) d[X,X]^c.$$

Note that $f''(x)$ could be any bounded Borel-measurable function. In particular, if $f''(x) = 1$, then

$$[X,X]_t^c = \int_{\mathbb{R}} L_t^a da,$$

whereas if $f(x) = 1_B(x)$ for some $B \in \mathcal{B}(\mathbb{R})$, then

$$\int_0^t 1_{\{X_{s-} \in B\}} d[X, X]_s^c = \int_B L_t^a da.$$

It is noteworthy to consider the special case when X is a $BM^0(\mathbb{R})$. In that case, we see that

$$\int_0^t 1_{\{X_s \in B\}} ds = \int_B L_t^a da,$$

and L_t^a represents precisely the total amount of time that X spent in B up to the instant t. Moreover, if $B = \{a\}$, then L_t^a is the total time spent at the point a so far. Our time-change results can now give a similar interpretation for continuous local martingales.

If we let $a = 0$, then when B is a $BM^0(\mathbb{R})$, we also get *Tanaka's formula*

$$|B_t| = |B_0| + \beta_t + L_t^0,$$

where $\beta_t = \int_0^t \operatorname{sgn}_-(B_s) dB_s$ is a martingale, the quadratic variation of which satisfies $[\beta, \beta]_t = \int_0^t \operatorname{sgn}_-(B_s)^2 d[B, B]_s = [B, B]_t = t$. Thus β is actually another $BM^0(\mathbb{R})$.

Semimartingale functions of Markov processes

It is possible to show that convex functions are essentially the most general functions for which Itô's formula makes sense (i.e. for which $f(X)$ is again a semimartingale).

The next logical step is to look for functions which take an arbitrary process and make semimartingales out of them. The solution of this problem for Markov processes is given in [ÇJPS80].

Recall a few basic definitions from the theory of Markov processes. We fix as usual a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and a state space $(E, \mathcal{B})$. Any mapping $N : E \times \mathcal{B} \to \overline{\mathbb{R}}_+$ is called a *kernel* if both

- for fixed $x \in E$, the map $A \mapsto N(x, A)$ is a positive measure on $\mathcal{B}$,
- for fixed $A \in \mathcal{B}$, the map $x \mapsto N(x, A)$ is $\mathcal{B}$-measurable.

If also $N(x,E)=1$ for every x, the kernel N is called a *transition probability.* Kernels act on $\mathcal{B}$-measurable functions f by

$$Nf(x)=\int_E N(s,dy)f(y).$$

Similarly, the product of two kernels M and N is again a kernel given by

$$MN(x,A)=\int_E M(x,dy)N(y,A).$$

A *transition function* on $(E,\mathcal{B})$ is a family $(P_{s,t}(x,A)=P(s,x,t,A))_{s,t\in\mathbb{R}_+}$ of transition probabilities satisfying the *Chapman-Kolmogorov equations*:

$$\int_E P(s,x,t,dy)P(t,y,u,A)=P(s,x,u,A).$$

Finally, an adapted process X is called a *Markov process* with respect to $(\mathcal{F}_t)$, with transition functions $(P_{s,t})$, if for any $f\in\mathcal{B}_+$ and $s<t$,

$$\mathbb{E}[f(X_t)|\mathcal{F}_s]=P_{s,t}f(X_s).$$

The measure $\mathbb{P}\circ X_0^{-1}$ is called the *initial distribution* of the process.

Here are some examples. The $BM^x(\mathbb{R}^d)$ is a Markov process on $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$, with transition function given by

$$P_{s,t}(x,A)=\int_A (2\pi|t-s|)^{-d/2}e^{-|y-x|^2/2|t-s|}dy.$$

It is *homogeneous*, as $P_{s,t}$ depends only on the difference $(t-s)$. Its initial distribution is δ_x, the Dirac measure centered at x.

The $PP(\lambda)$ is a homogeneous Markov process on $(\mathbb{Z}_+,2^{\mathbb{Z}_+})$, with transition function given by

$$P_{s,t}(x,A)=\sum_{k\in(A-x)\cap\mathbb{Z}_+} e^{-\lambda(t-s)}\frac{(\lambda(t-s))^k}{k!}.$$

Its initial distribution is δ_0.

By extending the state space to include time, one can always assume the transition functions are homogeneous, that is

$$P(s,x,t,A)=P(0,x,t-s,A),$$

in which case we write P_{t-s} instead of $P_{s,t}$, and the Chapman-Kolmogorov equations read

$$P_{t+s}(x, A) = \int_E P_s(s, dy) P_t(y, A).$$

Thus the $(P_t)_{t \in \mathbb{R}_+}$ form a *semi-group*.

Given a Markov process, we now aim to find out when $f(X)$ is a semimartingale. One says that a function f is *excessive* for X (actually, its semi-group) if

- $P_t f \le f$ for every $t > 0$,
- $\lim_{t \downarrow 0} P_t f = f$.

If moreover $P_t f = f$ for all t, then f is called *invariant*.

If f is excessive, then $f(X)$ is a $(\mathcal{F}_t)$-supermartingale, for

$$\mathbb{E}[f(X_t) | \mathcal{F}_s] = P_{t-s} f(X_s) \le f(X_s),$$

and of course if f is invariant, then $f(X)$ is a martingale. We haven't used the second property in the definition of excessive functions, but then we always work with the usual conditions, which is not the usual setup for Markov processes.

Thus excessive functions turn Markov processes into semimartingales. Of course, a composition of a convex function and an excessive function also yields a semimartingale, etc. Such functions are called *semimartingale functions*. The paper [ÇJPS80] has a characterization of such functions.

Approximating the integral

We've seen how Riemann sums may be used to approximate certain stochastic integrals. Those approximations were in terms of "lower sums". Recall that in elementary real analysis, the Riemann (and also the Riemann-Stieltjes) integral is defined as the common limit of upper and lower sums, whenever these are the same. For the stochastic integral however, upper and lower sums cannot have a common limit in general (for if they did, the integral would reduce to a Riemann-Stieltjes integral). The following theorem shows that by choosing our approximations correctly, we can "modify" some properties of the stochastic integral.

THEOREM **42.** [RY90] Let X, Y be semimartingales where Y is also continuous. Given a probability measure on $[0,1]$, write $\overline{\mu} = \int \lambda\mu(d\lambda)$. Provided the paths of $[X,Y]$ are absolutely continuous, then for any (deterministic) refining sequence (Δ_n) of finite partitions of $\mathbb{R}_+$ and any C^1 function f, the partial sums

$$\sum_{\Delta_n} \int_0^1 \mu(d\lambda) f(Y_{t_i+\lambda(t_{i+1}-t_i)})(X^{|t_{i+1}} - X^{|t_i})$$

converge in ucp to

$$f(Y)\cdot X + \overline{\mu} f'(Y)\cdot [X,Y].$$

PROOF: By using $f(Y_{\lambda_i}) - f(Y_{t_i}) = \int_0^1 f'(Y_{t_i} + s(Y_{\lambda_i} - Y_{t_i}))ds$ where we are abbreviating $\lambda_i = t_i + \lambda(t_{i+1} - t_i)$, we can write

$$\begin{aligned}\sum_{\Delta_n}(X_{t\wedge t_{i+1}} - X_{t\wedge t_i})\int_0^1 f(Y_{\lambda_i})\mu(d\lambda) &= \sum_{\Delta_n} f(Y_{t_i})(X^{|t_{i+1}} - X^{|t_i}) \\ &+ \sum_{\Delta_n}(X_t{}^{|t_{i+1}} - X_t{}^{|t_i})\int_0^1 f'(Y_{t_i})(Y_{\lambda_i} - Y_{t_i})\mu(d\lambda) \\ &+ \sum_{\Delta_n}(X_t{}^{|t_{i+1}} - X_t{}^{|t_i})\int_0^1 \mu(d\lambda)\int_0^1 ds\,(f'(Y_{t_i} + s(Y_{\lambda_i} - Y_{t_i})) - f'(Y_{t_i}))\,(Y_{\lambda_i} - Y_{t_i}).\end{aligned}$$

Now the first sum is just a Riemann approximation of $f(Y)\cdot X$ which thus converges in ucp. Also if $K_n = \sup_{t_i\in\Delta_n}\sup_{s\in[0,1]}|f'(Y_{t_i} + s(Y_{\lambda_i} - Y_{t_i})) - f'(Y_{t_i})|$ then $K_n \to 0$ a.s. since $f \in C^1(\mathbb{R})$, and the third sum is dominated by

$$\begin{aligned}&\sum_{\Delta_n} |X_t{}^{|t_{i+1}} - X_t{}^{|t_i}| \int_0^1 \mu(d\lambda) K_n |Y_{\lambda_i} - Y_{t_i}| \\ &\qquad\le \int_0^1 \mu(d\lambda) K_n \left(\sum_{\Delta_n}(Y_{\lambda_i} - Y_{t_i})^2\right)^{1/2}\left(\sum_{\Delta_n}(X_t{}^{|t_{i+1}} - X_t{}^{|t_i})^2\right)^{1/2},\end{aligned}$$

which converges to zero in probability on each interval $[0,t]$.

Finally, rewrite the second sum as

$$\begin{aligned}&\sum_{\Delta_n}\int_0^1 \mu(d\lambda) f'(Y_{t_i})(Y_{\lambda_i} - Y_{t_i})(X_t^{|\lambda_i} - X_t{}^{|t_i}) \\ &\qquad+ \sum_{\Delta_n}\int_0^1 \mu(d\lambda) f'(Y_{t_i})(Y_{\lambda_i} - Y_{t_i})(X_t{}^{|t_{i+1}} - X_t^{|\lambda_i}),\end{aligned}$$

where the second part is really $H^n \cdot X$,

$$H^n = \sum_{\Delta_n} f'(Y_{t_i})(Y_{\lambda_i} - Y_{t_i}) 1_{]\lambda_i, t_i+1]}.$$

By stopping at $T_k = \inf\{t : |Y_t| \geq k\}$, we see that H^n is locally bounded, so by SDCT $H^n \cdot X$ converges to zero in ucp.

It remains to look at the first part of this sum. Since $[X, Y]$ has absolutely continuous paths, there is a continuous process A with $[X, Y]_t = \int_0^t A_s ds$. The following steps, explained below, then yield the result:

$$\begin{aligned}
\lim_n \sum_{\Delta_n} f'(Y_{t_i})(Y_t^{|\lambda_i} - Y_t^{|t_i})(X_t^{|\lambda_i} - X_t^{|t_i}) &= \lim_n \sum_{\Delta_n} f'(Y_{t_i})([X, Y]_t^{|\lambda_i} - [X, Y]_t^{|t_i}) \\
&= \lim_n \sum_{\Delta_n} f'(Y_{t_i}) \int_{t_i}^{\lambda_i \wedge t} A_s ds \\
&= \lambda \int_0^t f'(Y_s) d[X, Y]_s.
\end{aligned}$$

The first equality is seen as follows: write $H = f'(Y)$, and $H^n = \sum_{\Delta_n} H_{t_i} 1_{]t_i, \lambda_i]}$. Then in ucp,

$$\begin{aligned}
&\lim_n \sum_{\Delta_n} H_{t_i} \left([X, Y]^{|\lambda_i} - [X, Y]^{|t_i}\right) \\
&= \lim_n H^n \cdot [X, Y] \\
&= \lim_n H^n \cdot (XY) - (H^n X_-) \cdot Y - (H^n Y_-) \cdot X \\
&= \lim_n \sum_{\Delta_n} \left(H_{t_i}(X^{|\lambda_i} Y^{|\lambda_i} - X^{|t_i} Y^{|t_i}) - H_{t_i} X_{t_i}(Y^{|\lambda_i} - Y^{|t_i}) - H_{t_i} Y_{t_i}(X^{|\lambda_i} - X^{|t_i})\right) \\
&= \lim_n \sum_{\Delta_n} H_{t_i}(X^{|\lambda_i} - X^{|t_i})(Y^{|\lambda_i} - Y^{|t_i}).
\end{aligned}$$

For the last equality, note that

$$\begin{aligned}
\left|\sum_{\Delta_n} \int_{t_i}^{\lambda_i \wedge t} A_s ds - \lambda \int_0^t A_s ds\right| &= \left|\sum_{\Delta_n} \int_{t_i}^{\lambda_i \wedge t} A_s ds - \sum_{\Delta_n} A_{t_i} \lambda(t_{i+1} \wedge t - t_i \wedge t)\right| \\
&\leq \sum_{\Delta_n} \int_{t_i}^{\lambda_i \wedge t} |A_s - A_{t_i}| ds,
\end{aligned}$$

which tends to zero by continuity of A. $\square$

Here are now some striking applications of this theorem. They may also be thought of as an important warning for the numerical evaluation of stochastic integrals:

Suppose X is a continuous semimartingale. Then for any C^2 function f, recall that Itô's formula is

$$f(X) = f(X_0) + f'(X_-) \cdot X + \frac{1}{2} f''(X_-) \cdot [X, X].$$

We now would like to naïvely approximate (in probability) the right-hand side via Riemann-Stieltjes sums.

Thus we take a refining sequence of (deterministic) partitions (Δ_n^μ), where we have $\Delta_n^\mu = \{t_{n0} \le \tau_{n0}^\mu \le t_{n1} \le \ldots\}$, and the τ_{ni}^μ are chosen at random according to the probability measure μ. The n-th approximation of an integral $H \cdot X$ will be written

$$\Delta_n^\mu (H \cdot X)_t = \sum_{\Delta_n^\mu} H_{\tau_{ni}^\mu} (X_t^{|t_{n(i+1)}} - X_t^{|t_{ni}}).$$

Applying this to the Itô formula gives

$$\begin{aligned} f^\mu(X) &= f(X_0) + \lim_n \Delta_n^\mu (f'(X_-) \cdot X) + \frac{1}{2} \Delta_n^\mu (f''(X_-) \cdot [X, X]) \\ &= f(X_0) + f'(X_-) \cdot X + (\frac{1}{2} + \overline{\mu}) f''(X_-) \cdot [X, X] \\ &= f(X) + \overline{\mu} f''(X_-) \cdot [X, X]. \end{aligned}$$

Thus in general, unless $[X, X] = 0$ say, our naive approximations will not yield $f^\mu(X) = f(X)$ as expected, but will produce a process which might be very different!

In fact, looking back at the theorem, we see that we get valid Riemann-Stieltjes approximations of (arbitrary) stochastic integrals only when $\mu(d\lambda) = \delta_0(d\lambda)$, the probability measure which is concentrated at 0. Other popular measures are the choice $\mu(d\lambda) = \delta_1(d\lambda)$, which yields approximations converging to the "backward integral", and more importantly, $\mu(d\lambda) = \delta_{1/2}(d\lambda)$, which yields the "Stratonovich integral".

Stratonovich stochastic integrals

The previous theorem allows us to consider other "stochastic integrals". The most ubiquitous is called the *Stratonovich integral*. It is usually defined when

B is a $BM^0(\mathbb{R})$as the limit in probability

$$\begin{aligned}\int_0^t f(B_s) \star dB_s &= \lim_n \sum_{\Delta_n} f(B_{(t_i+t_{i+1})/2}(B_{t_{i+1}} - B_{t_i}) \\ &= (f(B) \cdot B)_t + \frac{1}{2}\int_0^t f(B_s) ds.\end{aligned}$$

For our purposes, given two semimartingales X and Y, we define the *Stratonovich integral* of Y with respect to X as

$$\int_0^t Y_{s-} \star dX_s = (Y_- \cdot X)_t + \frac{1}{2}[Y, X]_t^c.$$

When Y is continuous, this clearly is the ucp limit of $\sum_{\Delta_n} Y_{(t_i+t_{i+1})/2}(X^{|t_{i+1}} - X^{|t_i})$, *provided that* $[X, Y]$ *is absolutely continuous.*

The reason for ever considering Stratonovich integrals lies in the following theorem:

THEOREM **43.** [Pro92] Let X be a semimartingale and f be a C^3 function. Then

$$f(X_t) - f(X_0) = \int_0^t f'(X_{s-}) \star dX_s + \sum_{s \le t} (f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s).$$

PROOF: Given the definition and Itô's formula, it suffices to show

$$\frac{1}{2}[f'(X), X]^c = \frac{1}{2} f''(X_-) \cdot [X, X]^c.$$

By applying Itô's formula to $f'(X)$ yields

$$[f'(X), X]^c = [f''(X_-) \cdot X, X]^c + \frac{1}{2}[f^{(3)}(X_-) \cdot [X, X], X]^c,$$

and the result follows because $[X, X]$ is an FV process. □

Thus Stratonovich integrals follow the usual rules of calculus. On the other hand, there are also drawbacks. The most important one is that if X is a local martingale, $\int_0^t H_s \star dX_s$ isn't one anymore, so that the powerful stability under semimartingale decomposition is lost.

Financial option pricing

On the stock market, it is possible to trade *call options* for stock in some company WXYZA Inc. The holder of such an option is allowed to buy stock in WXYZA at a fixed *striking* price, regardless of the fluctuations of the actual price of the stock. Similarly, a *put option* allows one to sell WXYZA stock at a fixed price. Options have naturally expiry dates, beyond which they become worthless. An *American* option can be exercised at any moment until expiry, whereas a *European* option works only on the expiry date. The natural question to ask is: what are such options worth? Here is Black and Scholes' answer, for the case of a European call option:

We assume no stock dividends, no transaction costs and continuous trading. The stock price S_t at time t is supposed to satisfy the equation (in differential notation, as is customary):

$$dS_t = \mu S_t dt + G S_t dB_t,$$

where μ, G are real constants and B is a standard BM. The filtration $(\mathcal{F}_t)$ used throughout is the smallest standard filtration containing $\sigma(B_s : s \leq t)$.

Pick an arbitrary investor with a portfolio consisting of f_W units of WXYZA call options, all expiring at the same date t_0, and f_S units of WXYZA stock. At time t, the worth I of the portfolio is

$$I_t = f_W(t, S, W) W(t, S) + f_S(t, S, W) S_t.$$

We assume f_W and f_S are at least C^2 and $(\mathcal{F}_t)$-adapted. This means that the portfolio mix can only depend on the past history of the stock price.

What we are interested in is $W(t, S)$: it is the worth of an option at time t (obviously then $W(t, S) = 0$ for $t > t_0$, and it should also be C^2 and $(\mathcal{F}_t)$-adapted). At any moment, the investor is allowed to alter the mix of WXYZA options and stock in his portfolio, thus altering its worth, but it must always contain nothing but WXYZA options and stock. If the portfolio is self-financing, i.e. the investor buys stock with the money he makes by selling options and vice-versa, we have

$$W(t, S, W) df_W + S_t df_S = 0.$$

Applying the Itô formula with this constraint gives

$$dI_t = f_W \left(\frac{\partial W}{\partial t} + \mu S_t \frac{\partial W}{\partial S} + \frac{1}{2} G^2 S_t^2 \frac{\partial^2 W}{\partial S^2} + f_S \mu S_t \right) dt + \left(f_W \frac{\partial W}{\partial S} + f_S \right) G S_t dB_t.$$

If we set

$$f_W \frac{\partial W}{\partial S} + f_S = 0,$$

then the second integral on the right disappears: the portfolio is "riskless".

If now, in flagrant disregard for the rules, the investor were to suddenly decide to convert his whole portfolio into another "riskless" asset, such as treasury bills or a bank account, with constant continuously compounded interest rate r, dI_t would become

$$dI_t = r(f_W W + f_S S_t)dt,$$

for all subsequent t.

The idea now is to say that if $W(t, S)$ is a fair price for the option, then the investor should not be able to make (or lose) money by converting his portfolio to other "riskless" assets. This means we get the equation

$$\frac{1}{2} G^2 S_t^2 \frac{\partial W}{\partial S^2} + \mu S_t \frac{\partial W}{\partial S} - rW - \frac{\partial W}{\partial t} = 0,$$

with the boundary condition

$$W(t_0, S) = 0 \vee (S_{t_0} - E),$$

where E is the striking price.

The above equation can be transformed into the heat equation. Thus the solution becomes

$$W(t, S) = S_t N(0, 1)(] - \infty, d_1]) - EN(0, 1)(] - \infty, d_2])e^{-r(t-t_0)},$$

where $N(0, 1)$ is the normal distribution

$$N(0, 1)(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

and the constants d_1 and d_2 are given by

$$d_1 = \frac{\log(S/E) + (r + \frac{1}{2}G^2)(t - t_0)}{G(t - t_0)^{1/2}},$$

and

$$d_2 = d_1 - G(t - t_0)^{1/2}.$$

Notes and Comments

The Doléans exponential has many uses, quite a few of them in the theory of stochastic differential equations (SDEs) as expected. But it can also be used to prove the Girsanov transformation formula, which is one of our major omissions, and relates the decompositions of a semimartingale under two equivalent probability measures.

We did not look at SDEs more closely, mainly out of concern for the size of this honours thesis. But as with ordinary differential equations (let alone PDEs!) the number of explicitly solvable SDEs is terribly small. Most books with the words "stochastic differential equations" in their titles deal only with equations of the form

$$\int_0^t dX_s = x + \int_0^t a(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s,$$

where B is a $BM^0(\mathbb{R})$. For such equations the original integration theory of Itô is quite sufficient. See for example [Øks85, Sko89, Arn74, Fri76], and more generally, most books written before the late 1980s. The book [KP92] contains a nice list of explicitly solvable equations of this type. For semimartingale SDEs, [Pro92] is excellent. Nonlinear SDEs are considered in [Mao94].

The characterizations of $BM^0(\mathbb{R})$ and $PP(\lambda)$ are by now classics. The result that any continuous local martingale is a time-changed (and possibly stopped) $BM^0(\mathbb{R})$ has a precursor in the following famous theorem of Paul Lévy:

Let B be a $BM^x(\mathbb{C})$ and $f : \mathbb{C} \to \mathbb{C}$ be analytic and non-constant. Then $f(B)$ is a time-changed $BM^x(\mathbb{C})$. More precisely,

$$f(B_t) = f(B_0) + \tilde{B}_{C_t},$$

where $C_t = \int_0^t |f'(B_s)|^2 ds$ and $\tilde{B}$ is a $BM^0(\mathbb{C})$.

The books [RY90, Dur84] contain many more results on the interaction between analysis and probability.

Our treatment of local times is very limited. An excellent survey of what can be done with them and the related theory of excursions is [Rog89]. In particular, the theory of excursions yields a surprising relationship between the $BM^0(\mathbb{R})$ and $PP(\lambda)$. This seems to have been first noticed by Itô in [Itô70].

The interpretation of the $BM^0(\mathbb{R})$ as an integral of white noise is very old. For the random Schwartz distribution treatment, see [GV68]. See also [Hid93].

Approximating the stochastic integral is of course a very important topic. The book [KP92] deals exclusively with this in the case of integrals $(H \cdot B)$ where B is a $BM^0(\mathbb{R})$. It is also interesting to note that most books which deal with the original Itô integral (that is, do not mention semimartingales) define it by Riemann sums. As we saw, this procedure does not generalize to even all local martingales.

Another important issue related to approximations is the matter of modelling. There are many arguments in the scientific literature about which integral, Itô or Stratonovich, is appropriate in models of various systems under the influence of random effects. As we saw, the resulting processes can have quite different qualitative behaviour. Usually, the SDEs in those models are approximations to difference equations anyway, and knowing how approximations converge can be of help in deciding which integral to choose.

Finally, the option-pricing formula first appeared in [BS73]. We follow the review paper [Sha90] which also contains clear explanations of many other topics in finance. We only dealt with European options. The same problem for American options is more difficult. See [DHRW93] for a solution.

Bibliography

[Arn74] L. Arnold. *Stochastic Differential Equations: Theory and Applications.* Wiley, New York, 1974.

[Bic81] K. Bichteler. Stochastic integration and L^p-theory of semimartingales. *Ann. Probab.*, 9(1):49–89, 1981.

[BS73] F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81:637–659, 1973.

[Bur66] D. L. Burkholder. Martingale transforms. *Ann. Math. Statist.*, 37:1494–1504, 1966.

[ÇJPS80] E. Çinlar, J. Jacod, P. Protter, and M. Sharpe. Semimartingales and Markov processes. *Z. W.*, 54:161–220, 1980.

[CW83] K. L. Chung and R. J. Williams. *Introduction to Stochastic Integration.* Birkhäuser, Boston, Basel, Stuttgart, 1983.

[Del80] C. Dellacherie. Un survol de la theorie de l'intégrale stochastique. *Stochastic Process. Appl.*, 10:115–144, 1980.

[DHRW93] J. N. Dewynne, S. D. Howison, I. Rupf, and P. Wilmott. Some mathematical results in the pricing of american options. *European J. Appl. Math.*, 4:381–398, 1993.

[DM78] C. Dellacherie and P. A. Meyer. *Probabilities and Potential.* North Holland, Amsterdam, New York, 1978.

[DM80] C. Dellacherie and P. A. Meyer. *Probabilités et Potentiel II.* Hermann, Paris, 1980.

[Doo53] J. L. Doob. *Stochastic Processes.* Wiley, New York, 1953.

[Doo93] J. L. Doob. *Measure Theory.* Springer, Berlin, Heidelberg, New York, 1993.

[DS58] N. Dunford and J. T. Schwarz. *Linear Operators Part I: General Theory.* Wiley, New York, 1958.

[DU77] J. Diestel and J. J. Uhl. *Vector Measures.* Amer. Math. Soc., 1977.

[Dur84] R. Durrett. *Brownian Motion and Martingales in Analysis.* Wadsworth, Belmont, cal., 1984.

[Elw82] K. D. Elworthy. *Stochastic Differential Equations on Manifolds.* Cambridge University Press, New York, 1982.

[EM89] M. Emery and P. A. Meyer. *Stochastic Calculus in Manifolds.* Springer, Berlin, Heidelberg, New York, 1989.

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, New York, 1968.

[Fri76] A. Friedman. *Stochastic Differential Equations and Applications.* Academic Press, New York, 1976.

[GV68] I. M. Gel'fand and N. Y. Vilenkin. *Generalized Functions*, volume 4. Academic Press, New York, 1968.

[Hid93] T. Hida. *White Noise: an Infinite Dimensional Calculus.* Dordrecht, Boston, 1993.

[Itô51] K. Itô. *Stochastic Differential Equations*, volume 4 of *Memoirs of the AMS.* Amer. Math. Soc., 1951.

[Itô70] K. Itô. Poisson point processes attached to Markov processes. In *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, volume 3, pages 225–239, 1970.

[IW81] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes.* North Holland, Amsterdam, New York, 1981.

[Jeu79] J. Jeulin. *Grossissement d'une filtration et applications*, volume 721 of *Lecture Notes in Math.* Springer, Berlin, Heidelberg, New York, 1979.

[JS87] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes.* Springer, Berlin, Heidelberg, New York, 1987.

[JY85] Th. Jeulin and M. Yor, editors. *Grossissements de Filtrations: Exemples et Applications: Séminaire de Calcul Stochastique 1982/83, Université Paris VI*, volume 1118 of *Lecture Notes in Math.*, Berlin, Heidelberg, New York, 1985. Springer.

[KK76] I. Kluvanek and G. Knowles. *Vector Measures and Control Systems.* North Holland, Amsterdam, New York, 1976.

[Kni81] F. B. Knight. *Essentials of Brownian Motion.* Amer. Math. Soc., 1981.

[KP92] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, Heidelberg, New York, 1992.

[KPR82] N. J. Kalton, N. T. Peck, and J. W. Roberts. L_0-valued vector measures are bounded. *Proc. Amer. Math. Soc.*, 85(4):575–582, 1982.

[Kry90] N. V. Krylov. Une représentation des sousmartingales positives et ses applications. In *Séminaire de Probabilités XXIV*, volume 1426 of *Lecture Notes in Math.*, pages 473–476, Berlin, Heidelberg, New York, 1990. Springer.

[KS88] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, Berlin, Heidelberg, New York, 1988.

[KT81] S. Karlin and H. Taylor. *A Second Course in Stochastic Processes.* Academic Press, New York, 1981.

[Kus77] A. U. Kussmaul. *Stochastic Integration and Generalized Martingales.* Pitman, London, San Francisco, Melbourne, 1977.

[KW67] H. Kunita and S. Watanabe. On square integrable martingales. *Nagoya Math. J.*, 30:209–245, 1967.

[Kwa92] S. Kwapien. *Random Series and Stochastic Integrals: Single and Multiple.* Birkhäuser, Boston, Basel, Stuttgart, 1992.

[Loè77] M. Loève. *Probability Theory.* Springer, Berlin, Heidelberg, New York, 4 edition, 1977.

[LS89] R. S. Lipster and A. N. Shiryayev. *Theory of Martingales.* Dordrecht, Boston, 1989.

[Mao94] X. Mao. *Exponential Stability of Stochastic Differential Equations.* Marcel Dekker, New York, 1994.

[Mas68] P. Masani. Orthogonally scattered measures. *Adv. Math.*, 2:61–117, 1968.

[Mét73] M. Métivier. Stochastic integrals and vector-valued measures. In *Vector and Operator Valued Measures and Applications.* Academic Press, 1973.

[Mét82] M. Métivier. *Semimartingales: a Course on Stochastic Processes.* de Gruyter, Berlin, New York, 1982.

[Mey76] P. A. Meyer. Un cours sur les intégrales stochastiques. In *Séminaire de Probabilités X*, volume 511 of *Lecture Notes in Math.*, pages 246–400, Berlin, Heidelberg, New York, 1976. Springer.

[Mey79] P. A. Meyer. Caractérisation des semimartingales, d'après dellacherie. In *Séminaire de Probabilités XIII*, volume 721 of *Lecture Notes in Math.*, pages 620–623, Berlin, Heidelberg, New York, 1979. Springer.

[MP77] M. Métivier and J. Pellaumail. Mesures stochastiques à valeurs dans des espaces L_0. *Z. W.*, 40:101–114, 1977.

[MP80] M. Métivier and J. Pellaumail. *Stochastic Integration.* Academic Press, New York, 1980.

[Nel67] E. Nelson. *Dynamical Theories of Brownian Motion.* Princeton University Press, 1967.

[Øks85] B. Øksendal. *Stochastic Differential Equations. An Introduction with Applications.* Springer, Berlin, Heidelberg, New York, 1985.

[Pel73] J. Pellaumail. *Sur l'Intégrale Stochastique et la Décomposition de Doob-Meyer.* PhD thesis, 1973.

[Pro86] P. Protter. Stochastic integration without tears (with apology to P.A. Meyer). *Stochastics*, 16:295–325, 1986.

[Pro92] P. Protter. *Stochastic Integration and Differential Equations; A New Approach.* Springer, Berlin, Heidelberg, New York, 1992.

[Rao93] M. M. Rao. An approach to stochastic integration (a generalized and unified treatment). In *Multivariate Analysis: Future directions*, pages 347–374. Elsevier Science Publishers, 1993.

[Rog81] L. C. G. Rogers. Stochastic integrals: basic theory. In *Stochastic Integrals. Proceedings of the LMS Durham Symposium*, volume 851 of *Lecture Notes in Math.*, pages 56–71, Berlin, Heidelberg, New York, 1981. Springer.

[Rog89] L. C. G. Rogers. A guided tour through excursions. *Bull. London Math. Soc.*, 21:305–341, 1989.

[Rud73] W. Rudin. *Functional Analysis.* McGraw-Hill, New York, 1973.

[RW87] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and Martingales 2: Itô calculus.* Wiley, New York, 1987.

[RY90] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer, Berlin, Heidelberg, New York, 1990.

[Sch80] L. Schwartz. Semi-martingales formelles. In *Séminaire de Probabilités XV*, volume 850 of *Lecture Notes in Math.*, pages 413–489, Berlin, Heidelberg, New York, 1980. Springer.

[Sha90] K. P. Sharp. Stochastic differential equations in finance. *Appl. Math. Comput.*, 37:131–148, 1990.

[Sko89] A. V. Skorokhod. *Asymptotic Methods in the Theory of Stochastic Differential Equations.* Amer. Math. Soc., New York, 1989.

[Str77] C. Stricker. Quasimartingales, semimartingales et filtrations naturelles. *Z. W.*, 39:55–65, 1977.

[SV79] D. W. Strook and S. R. S. Varadhan. *Multidimensional Diffusion Processes.* Springer, Berlin, Heidelberg, New York, 1979.

[vWW90] H. v. Weizsäcker and G. Winkler. *Stochastic Integrals.* Vieweg, Braunschweig/Wiesbaden, 1990.

[Wie23] N. Wiener. Differential space. *J. Math. Phys.*, 2:132–174, 1923.

[Wil79] D. Williams. *Diffusions, Markov processes, and Martingales 1: Foundations.* Wiley, New York, 1979.

[Yor78] M. Yor. Quelques interactions entre mesures vectorielles et intégrales stochastiques. In *Séminaire de théorie du potentiel IV*, volume 713 of *Lecture Notes in Math.*, pages 264–281, Berlin, Heidelberg, New York, 1978. Springer.