Appendix A

Background theory for stochastic processes

A.1 Brief history of stochastic calculus

The development of general theory of stochastic integration has the following important milestones

- Ville (1939) defines the concept of *martingale* for the first time; this is not to say that martingale-like objects did not appear before but Ville's definition on the level of processes (rather than increments) is significant; to see how Paul Lévy fits into this picture, consult Mazliak (2009);
- Doob (1940) shows that a martingale $\{X_n\}_{n\in\mathbb{N}}$ bounded in L^1 has a limit X_∞; furthermore, X is uniformly integrable if and only if $X_n = E_n[X_\infty]$
- Itô (1944) develops $L^2(P)$–integration theory for Brownian motion;
- Itô (1951) proposes his change of variables formula;
- Doob (1953)
  - uses what we now call a *filtration*
  - defines what is now known as *stopping time*
  - uses the notion of a stopped process (here called *optional stopping* transformation);
  - defines submartingale (which he calls semi-martingale; supermartingale is called 'lower semi-martingale');
  - shows that a submartingale bounded in L^1 has an integrable limit at ∞;
  - states what is now known as the Doob decomposition of a submartingale;

  - proves what is now known as Doob's maximal inequality: if X is a positive submartingale and $p > 1$

$$E\left[\sup_{0\le t\le T} |X_t|^p\right] \le \left(\frac{p}{p-1}\right)^p (E[X_T^p]).$$

- Girsanov (1960) presents a formula for drift of an Itô process under an equivalent measure driven by another Itô process;

- Meyer (1962, 1963) introduces processes of class (D), *defines and proves* unique decomposition for supermartingales of class (D) into an increasing 'natural' process and a martingale. Later Doléans (1967) shows that increasing natural process is the same as increasing (therefore finite variation) predictable process;

- Itô and Watanabe (1965) introduce the concept of local martingale and show any positive supermartingale decomposes into a local martingale and an increasing natural process;

- Kunita and Watanabe (1967) extend stochastic integration to all (locally) square-integrable martingales and define $\langle X, Y\rangle$ for two (locally) square-integrable martingales X, Y;

- Meyer (1966) elaborates on his previous results and on the limit theorems of Doob (1953); in particular, in continuous time he replaces the notion of *separable* stochastic process with the simpler notion of process with *right-continuous paths*;

- Meyer (1967) defines $[X, Y]$ for two local martingales X, Y and shows this is well-defined even though $\langle X, Y\rangle$ might not be. The object $[X, X]$ is interpreted by Doléans (1969) as quadratic variation of X;

- McKean (1969) writes the classical Itô formula in a measure-invariant way;

- Doléans-Dade and Meyer (1969) define the modern notion of semimartingale (local martingale + FV process), construct the stochastic integral of a semimartingale for locally bounded predictable integrands, and give change of variables formula (Itô–Meyer formula). All this is done in more detail in Doléans-Dade and Meyer, 1970.

- Meyer (1976) proves that the notions of semimartingale and stochastic integral are invariant to equivalent change of measure; furthermore he writes the change of variables formula (the Itô–Meyer formula) in a measure-invariant way

A.2 Expanded chronology of stochastic calculus

1. Doob (1940) studies discrete-time martingales;
2. Itô (1944) develops $L^2(P)$ stochastic integral for Brownian motion;
3. Itô (1951) establishes the first change of variables formula, for Itô processes;
4. Doob (1953)
   (a) has a definition of stopping time (without giving it a name, p. 300; modern terminology appears e.g. in Hunt, 1956);
   (b) uses the notion of a stopped process (which he calls "optional stopping transformation", pp. 300, 366); also uses the notion of time change (which he calls "optional sampling");
   (c) defines a submartingale (which he calls semi-martingale); proves that a stopped (sub)martingale is a (sub)martingale (Theorems 2.1 and 11.6, the latter under a technical condition)
   (d) introduces what is now called the Doob decomposition of a submartingale (discrete time, eq. (1.5′), p. 297)

   submartingale = increasing predictable process + martingale;

5. Meyer (1962, 1963)
   (a) Meyer (1962) introduces processes of class (D);
   (b) defines *potential* as a positive supermartingale X with $\lim_{t\to\infty} E[X_t] = 0$.
   (c) in Proposition 1 shows every martingale is of class (DL);
   (d) Meyer (1963) proves uniqueness of the following decomposition (now called Doob-Meyer decomposition) for class (D) supermartingales;

   $$\underbrace{\text{increasing natural process}}_{\text{(same as increasing predictable FV process; Doléans, 1967)}} + \text{martingale}$$

   (e) deduces that any square integrable martingale X
   (f) characterises accessible/totally inaccessible stopping times, the latter are typified by the jumps of a Poisson (or even Lévy) process
6. Fisk (1965) studies quasi-martingale = martingale + FV process
   (a) the only continuous FV martingale is constant

7. Itô and Watanabe (1965)

   (a) introduce the notion of local martingale (credited in Dellacherie and Meyer, 1978, p. 86 and Meyer, 2009)

   (b) in Lemma 2, p. 20, prove every positive supermartingale has the Doob-Meyer decomposition (with local martingale in place of martingale). Positivity assumption is easily removed by stopping at fixed times to obtain that every supermartingale is a special semimartingale (Dellacherie and Meyer, 1978, Theorem VII.12, p. 198)

8. Meyer (1966, Theorems 19, 49) has a new characterisation of natural integrable increasing process (see also Protter, 2005, Theorems III.12-16)

9. Doléans (1967) identifies increasing natural processes with increasing predictable processes

   (a) credited in Dellacherie and Meyer (1978, p. 429), cf. their items VI.61–62

10. Kunita and Watanabe (1967) freely use the notion of localization

    (a) extend Itô formula to all (locally) square-integrable martingales

    (b) define $\langle X, Y\rangle$ for two (locally) square-integrable martingales X, Y. Observe that $\langle X, X\rangle$ is the increasing natural process in the Doob–Meyer decomposition of the submartingale X^2 so the novelty is in extending this to XY by polarization, i.e. exploiting

$$XY = \frac{1}{4}\left((X+Y)^2 - (X-Y)^2\right).$$

11. Meyer (1967)

    (a) defines "processus très-bien-mesurable" which is later (Doléans-Dade and Meyer, 1969) renamed "prévisible", i.e. predictable

    (b) uses $\mathcal{M}_{\text{loc}}$ for what we now denote $\mathcal{M}^2_{\text{loc}}$ and $\mathcal{L}$ for what we now call $\mathcal{M}_{\text{loc}}$. In entirely modern way, defines $\mathcal{A}_{\text{loc}}$ (processes of locally integrable variation). Later, in Meyer (1976, Theorem IV.12) or Dellacherie and Meyer (1978, Theorem VI.80) it is also shown that predictable FV processes are in $\mathcal{A}_{\text{loc}}$.

    (c) has a remark (p. 97) that a local martingale is a martingale if and only if it is of class (DL);

    (d) shows for $M \in \mathcal{M}^2_{\text{loc}}$ process M^2 has a drift denoted by $\langle M, M\rangle$ such that $M^2 - \langle M, M\rangle \in \mathcal{M}_{\text{loc}}$;

(e) in Proposition 4 (p. 104) defines square bracket of a process in $\mathcal{M}_{\text{loc}}$ ($[M, M] := \langle M^c, M^c\rangle + \sum(\Delta M)^2$) and shows $M^2 - [M, M]$ is a local martingale.

(f) extends stochastic integral to processes in $\mathcal{M}_{\text{loc}}$

(g) defines semimartingale as an element of $\mathcal{M}_{\text{loc}} + \mathcal{A}_{\text{loc}}$

(h) on p. 108 defines $[X, X]$ as a quadratic variation of $X = M + A \in \mathcal{M}_{\text{loc}} + \mathcal{A}_{\text{loc}}$ and notes something like $[X, X] = [M, M] + [A, A]$.

(i) It does not appear that invariance under change of measure has ever been a motivator for this definition of semimartingale.

12. Doléans (1969) shows every martingale has finite quadratic variation in the sense that quadratic variation on partitions converges in probability to something that can be chosen right-continuous and increasing.

    (a) Millar (1968) shows one can define quadratic variation of a continuous martingale;

13. Doléans-Dade and Meyer (1969) define semimartingale the way we know it today, i.e. local martingale + FV process;

    (a) do not impose integrability on the finite variation part, which is an improvement on Meyer (1967);

    (b) point out every supermartingale (not necessarily of class D) is a semimartingale, as is every Lévy process;

    (c) use localisation in the modern way, talk about locally bounded processes, decompose local martingale into a square-integrable local martingale and integrable variation local martingale

    (d) produce the modern Itô formula, what we might call Itô–Meyer formula which is written almost entirely in a measure-invariant way, except they use $\langle X^c, X^c\rangle$ where X^c stands for continuous martingale part of X, instead of $[X, X]^c$.

14. Doléans-Dade and Meyer (1970) elaborate on the results announced in Doléans-Dade and Meyer (1969);

    (a) interpret Doob-Meyer decomposition as

        class (D) submartingale = predictable increasing integrable process + martingale

(b) decompose a square-integrable local martingale M uniquely into three square-integrable martingale components

$$M = M^{\mathrm{c}} + M^{\mathrm{dq}} + M^{\mathrm{dp}},$$

where M^c is continuous, M^{dq} is quasi-left-continuous, and M^{dp} is the sum of its jumps at predictable times.

15. Van Schuppen and Wong (1974) have Girsanov formula and make prodigious use of the $[\bullet,\bullet]$ bracket. Further Girsanov-type results are given in Jacod and Mémin (1976), Lenglart (1977), Yoeurp (1976)

16. Yoeurp (1976) defines canonical decomposition of a special semimartingale and shows uniqueness by proving that any predictable local martingale of FV is necessarily constant

17. Meyer (1976)

(a) introduces the notion of a predictable compensator of a process with integrable variation (I.8)

(b) States in III.3 the following two forms of "change of variables formula"

$$\begin{aligned}
df(X) =& f'(X_-)dX + \frac{1}{2}f''(X_-)d\left\langle X^c, X^c\right\rangle \\
&+ \left(f(X_- + \Delta X) - f(X_-) - f'(X_-)\Delta X\right) \\
=& f'(X_-)dX + \frac{1}{2}f''(X_-)d[X,X] \\
&+ \left(f(X_- + \Delta X) - f(X_-) - f'(X_-)\Delta X - \frac{1}{2}f''(X_-)\left(\Delta X\right)^2\right)
\end{aligned}$$

(c) shows predictable FV process has locally integrable variation (IV.12)

(d) defines canonical decomposition of a special semimartingale (IV.32)

(e) In Theorem VI.4 it is shown $[X, X]$ is the limit of quadratic variation sums

(f) In VI.5 it is shown for $\mathcal{C}^1$ (!) function f

$$d[f(X), f(X)] = \left(f'(X_-)\right)^2 d\left\langle X^c, X^c\right\rangle + \left(\Delta f(X)\right)^2$$

(g) VI.22 gives Girsanov-type results and shows P–semimartingale is equivalent to Q–semimartingale for arbitrary $Q \sim P$.

(h) VI.25 shows that $\langle X^c, X^c\rangle$ does not depend on measure, i.e. definitely identifies $[X, X]^c$ with quadratic variation of X^c and uses invariance of $[X, X]$ to conclude

$$\langle X^c, X^c \rangle = \langle \tilde{X}^c, \tilde{X}^c \rangle$$

where $\tilde{X}^c$ is continuous martingale part under an equivalent measure

(i) Finally in VI.26 it is shown that for locally bounded integrands stochastic integral does not depend on measure (the same in Dellacherie and Meyer, 1978, Theorem VIII.12)

18. Lépingle (1976) describes convergence of p–power variations.

    (a) Monroe (1972) has a statement on p–variation of a Lévy process;

    (b) Monroe (1976) has an example where quadratic variation diverges in some sense as partitions become finer.

A.3 Probability space

Probability space is a triplet $(\Omega, \mathcal{F}, P)$. The set Ω contains all elementary outcomes. For example, if we were simply rolling a dice we could take

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

$\mathscr{F}$ is the σ–field of events, in the dice example it would be the set of all subsets of Ω. P is a measure with total mass 1 assigning probabilities to events in $\mathscr{F}$.

It is natural to ask why one needs $\mathcal{F}$. Would it not be enough to assign probabilities directly to the elements of Ω? In the dice example, this of course does work, and it is easy to work out the probability of the more complicated events in $\mathcal{F}$ from the probability of the individual elementary outcomes $1, 2, \ldots, 6$, which on a fair dice is $1/6$. For example, the probability of the event $\{1, 2\}$, that is, the probability that we roll 1 or 2, has to be $1/3$. In this way the probability measure is extended naturally from all elements of Ω to $\mathcal{F}$. This approach will continue to work for countable Ω. In these simple cases $\mathcal{F}$ is the set of all subsets of Ω.

But consider now elementary outcomes taking values in the whole interval $\Omega = [0, 1]$. Suppose that the probability distribution is uniform, so that each outcome has equal probability. Naturally, the probability of each individual outcome is zero. However, this tells us nothing about the probability of the event $[0, 1/4]$, which should be $1/4$. So in this case it is pointless to specify the probability of individual outcomes and instead we need to define the probability on intervals $[0, x]$ for $0 < x \leq 1$ and then extend the probability to arbitrary intersections and their countable unions to add more events to $\mathcal{F}$.

How large is the σ-algebra $\mathcal{F}$ formed by arbitrary intersections and countable unions of intervals? In particular, does it contain all subsets of $[0, 1]$? The answer is *no*. Why not? Here we come against something very strange. There are subsets of $[0,1]$ which simply cannot be assigned *any* probability. Mathematically we say that these subsets, which do not belong to $\mathcal{F}$, are *not measurable*.[1]

A.4 Surely and almost surely

In the introductory example the outcome of our random experiment is one number in the interval $[0,1]$. By definition of our experiment, we are sure that it has to be one such number. Often, when making statements about random outcomes, it is *only* possible to assert that something happens with probability 1. In such case we say that the outcome happens *almost surely*. For example, the outcome of our experiment will almost surely be an irrational number. This does not mean that the oucome cannot be a fraction (a rational number), but the probability of getting a rational number is zero (because there are only countably many rationals). The sets of measure zero are called *null sets*.

It does not sound like much of a sacrifice to know an outcome "almost surely" instead of "surely". Yet, it is not entirely innocuous either. Using the example above, on an interval [0,1] with a uniform distribution *all* numbers in the interval are almost surely irrational. An even more striking example is provided by the so-called *Nikodym set* which represents a unit square whose every point is painted either black or white. This can be done in such a way that the square is almost surely white, but at the same time every line intersecting the square is completely black, with only one white point.

A.4.1 Complete probability space

One would naturally think that if a certain property $\mathcal{A}$ holds almost surely (that is on a set of probability 1), then, conversely, $\mathcal{A}$ fails to hold *with probability zero*. Strictly speaking, this is false. It may happen that the set where $\mathcal{A}$ fails to hold is, in fact, not even measurable. This can particularly happen in the Borel σ–algebra because Borel null sets can have subsets that are themselves not Borel-measurable!

So, to avoid such unpleasantness, when dealing with a probability space $(\Omega, \mathcal{F}, P)$ it is customary to enlarge $\mathcal{F}$ to include all the subsets of P-null sets already contained in $\mathcal{F}$ and extend P to all these subsets by assigning zero measure to them. This does

[1] There is a subset A of $[0,1]$ (so-called Vitali set) with the following properties: i) for any two rational numbers $r \neq q$ the sets $r + A$ and $q + A$ are disjoint; ii) the union of sets $r + A$ over rational $-1 \leq r \leq 1$ covers the whole interval $[0,1]$ and it is a subset of $[-1,2]$. It follows that A cannot be Lebesgue-measurable, because there are only countably many rationals and the Lebesgue measure is translation-invariant and countably additive.

In another example, it is possible to divide a unit ball in $\mathbb{R}^3$ into four parts and then by rotating and translating these four parts (so without stretching or squeezing anything!) to produce *two disjoint unit balls*. This is the so-called Banach-Tarski theorem.

not create any contradictions in the general theory. With these extra null sets the probability space $(\Omega, \mathcal{F}, P)$ is said to be *complete*. For example, the completion of Borel σ-algebra is called the Lebesgue σ–algebra and the corresponding extension of the Borel measure is the familiar Lebesgue measure.

On a complete probability space one can then use the statements "property $\mathcal{A}$ holds almost surely" and "$\mathcal{A}$ does not hold on a null set" interchangeably.

A.5 Random variable

Random variable X is a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. *Measurable* in this context means that the event $X \leq x$ belongs to $\mathcal{F}$ for all $x \in \mathbb{R}^d$, so it can be assigned a probability. The probability measure P on $\mathcal{F}$ and the mapping $X : \Omega \to \mathbb{R}^d$ give rise to a new measure P_X on $\mathcal{B}(\mathbb{R}^d)$ through the formula

$$P_X(B) = P(X^{-1}(B)).$$

In probability theory P_X is more commonly called the distribution of X.

In other words, the probability space $(\Omega, \mathcal{F}, P)$ is used as an enormous warehouse capable of storing the vast number of random variables that are needed to construct stochastic processes. However, once we are given a random variable X and we know that no other random variable will be required we can restrict our attention to the σ-algebra generated by the sets $X^{-1}((-\infty, x])$. This is called "the smallest σ–algebra generated by X" and it is denoted $\sigma(X)$.

A.6 When are two random variables the same?

When we say that two random variables X and Y are the same we mean they live on the same probability space and $X = Y$ almost surely. This is a much stronger statement than saying the distributions of X and Y are the same, $P_X = P_Y$. To understand the distinction, think of two objects from your daily life that are "the same" but are not the same object.

We will later see how to perform an absolutely continuous change of measure. Suppose now that $Q \ll P$. Then if $X = Y$ P–a.s. then also $X = Y$ Q–a.s. so the identification of random variables does not depend on the probability measure. So if $X = Y$ P–a.s. then necessarily $P_X = P_Y$ and also $Q_X = Q_Y$. In contrast $P_X = P_Y$ in no way guarantees $Q_X = Q_Y$!

To complete this discussion (and possibly confuse you further) we have $X = X$ everywhere (not just a.s.) but P_X is in general different from Q_X.

A.7 Distribution as an image measure*

Speaking more generally, the object P_X is an *image measure* (a.k.a. push-forward measure) of P via the mapping X. We will us this notation consistently; for a measure μ and a mapping f the symbol μ_f will denote the resulting image measure

$$\mu_f(B) = \mu(f^{-1}(B)).$$

In this section we will mention two constructive results on image measures which are used outside the narrow application of distributions and which will save us a lot of work later on.

The first result states that given a measurable map $g : (B, \mathcal{B}) \to (C, \mathcal{C})$ and a measure $\tilde{\mu}$ on $(B, \mathcal{B})$ the resulting image measure $\tilde{\mu}_g$ satisfies

$$\int_B \tilde{h}(g(y))\tilde{\mu}(dy) = \int_C \tilde{h}(x)\tilde{\mu}_g(dx), \tag{A.1}$$

for *any* measurable map $\tilde{h} : (C, \mathcal{C}) \to (C, \mathcal{C})$. In particular, the map g does not have to be one-to-one, so this covers situations where for example g maps x to x^2. The result is very useful when we know the marginal distribution P_X of X, in this example represented by measure $\tilde{\mu}$ and want to calculate the mean of $g(X)$ without evaluating the marginal distribution of $g(X)$ which corresponds to $\tilde{\mu}_g$ in this example. In such case we would evaluate the lhs of (A.1) with $\tilde{h}(x) = x$.

The second important result tells us what happens when we concatenate several transformations together. Let us consider another measurable map $f : (A, \mathcal{A}) \to (B, \mathcal{B})$ and a measure μ on $(A, \mathcal{A})$. The image measure μ_f lives on $(B, \mathcal{B})$ so we can use it instead of $\tilde{\mu}$ in formula (A.1), which now reads

$$\int_B h(g(y))\mu_f(dy) = \int_C h(x) \left(\mu_f\right)_g (dx), \tag{A.2}$$

for every measurable $h : (C, \mathcal{C}) \to (C, \mathcal{C})$. If we also identify the lhs of (A.2) with the rhs of (A.1) by taking $x \equiv y, \tilde{h} \equiv h \circ g$ the formula (A.1) then yields

$$\int_A h(g(f(z)))\mu(dz) = \int_B h(g(y))\mu_f(dy) = \int_C h(x) \left(\mu_f\right)_g (dx), \tag{A.3}$$

for every measurable $h : (C, \mathcal{C}) \to (C, \mathcal{C})$ which on comparing the first and last expressions in (A.3) means that

$$\mu_{g \circ f} = \left(\mu_f\right)_g. \tag{A.4}$$

How is this useful? Suppose we are given a random variable X which, let us assume for the sake of the argument, has a standard normal distribution with density,

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

so that

$$P_X(dx) = p_X(x)dx.$$

Suppose later on we need to work with a new random variable $Y = e^X = g(X)$.

One way to do this is to actually compute the distribution P_Y. In our particular example P_Y will be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. The change of variable formula[2] gives

$$p_Y(y) = \left|\frac{dg^{-1}(y)}{dy}\right| p_X(g^{-1}(y)) = \left|\frac{d\ln y}{dy}\right| p_X(\ln y) = \frac{1}{\sqrt{2\pi}y} e^{-(\ln y)^2/2}.$$

Now suppose the task is to evaluate the characteristic function of Y,

$$\begin{aligned}\varphi_Y(u) =& E[e^{iuY}] \\ =& \int_{\mathbb{R}} e^{iuy} P_Y(dy) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}y} e^{iuy-(\ln y)^2/2} dy.\end{aligned}$$

The concatenation rule for image measures (A.3) tells us

$$P_Y = P_{g(X)} = (P_X)_g, \tag{A.5}$$

that is, the measure P_Y is the image measure of P_X via the mapping g and it is also the image measure of the original measure P via the composed mapping $g(X)$. This is again true for *any* g and X.

Now we can see how one can compute $\varphi_Y(u)$ easily without evaluating the distribution of Y. Namely, formula (A.1) with $\tilde{\mu} \equiv P_X, g(x) = e^x$ and $\tilde{h}(y) \equiv e^{iuy}$ asserts

$$\int_{\mathbb{R}} e^{iuy} P_Y(dy) = \int_{\mathbb{R}} e^{iue^x} P_X(dx).$$

It turns out that the results (A.1) and (A.5) are very handy when working with Lévy processes and their transformations.

[2] The change of variable formula hinges on the fact that in this case g^{-1} is a again a measurable map. In general g^{-1} may not be measurable unless g maps null sets into null sets.

A.8 Filtration

Prices change over time. To capture the time evolution of prices we work with nested σ-algebras indexed by time, $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$. This collection of σ–algebras is called a *filtration*. We require that information is not lost over time, $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$. Additionally we require that the filtration is right-continuous, meaning

$$\mathcal{F}_s = \bigcap_{s<t} \mathcal{F}_t.$$

Intuitively, this corresponds to a situation where surprises are revealed at a specific time (as opposed to being revealed in an arbitrarily short interval *after* a specific time). Finally we need all null sets of $\mathcal{F}_T$ to be inside $\mathcal{F}_0$. These conditions will become important in the next section.

A.9 Discrete vs continuous time

In applications we often talk about *discrete-time* models in contrast to *continuous-time* models. This terminology is inaccurate and potentially misleading. *All* dynamic models can be written down in continuous time. The attribute *discrete-time* does not refer to time being discrete but rather to randomness being revealed at a finite number of time points.[3] This means the filtration is piecewise constant, i.e. there is a partition $\{t_k\}_{k=0}^n$ with $t_0 = 0$ and $t_n = T$ such that

$$\mathcal{F}_t = \mathcal{F}_{t_k} \text{ for } t_k \leq t < t_{k+1}.$$

In the terminology of stochastic processes these special times are called *fixed times of jump*, meaning there is *a-priori* a non-zero probability that a discrete change will occur at t_k. It is customary to identify the fixed jump times of a discrete-time model with non-negative integers, i.e. $t_k \equiv k$ $k = 0, 1, 2, \ldots$.

On the opposite side of the spectrum we have continuous processes, that is processes whose value *never* jumps. Between the two extremes there are processes that jump at random times. These processes are classified into two subcategories. In the first subcategory there are processes that jump at predictable times (for example, the jump may only occur at the time when a Brownian motion hits a prespecified barrier). In the second subcategory we have processes whose jumps occur only at totally

[3]We could even allow a countable number of time points as long as each point has a well-defined predecessor and successor. For example the sequence of time points 0, $\{1/n\}_{n=1}^{\infty}$ is countable but 0 does not have a successor, so this would not count as a discrete-time model in the sense we use here. For a model with countably many time points to qualify as a discrete-time model it must be the case that the only accumulation point is $+\infty$.

inaccessible times (like a Poisson process). The latter (those jumping only at totally inaccessible times) are called *quasi-left-continuous* processes.

The significance of discrete-time models, versus continuous processes, and the two shades in between, lies in different appearance of the drift process and degrees of mathematical difficulty with which various mathematical properties (such as martingale property) can be verified. For example, in discrete time a local martingale X that satisfies $E[|X_T|] < \infty$ is automatically a true martingale on $[0, T]$. Outside discrete-time environment a local martingale may not be a true martingale on $[0, T]$ even if it square integrable, $\sup_{t\in[0,T]} E[X_t^2] < \infty$. Inbetween, continuous processes, and in particular diffusions, allow for specific tailor-made criteria to check martingale property of local martingales. The drift of all quasi-left-continuous processes is continuous (if it exists).

Because the mathematical handling of discrete-time models is substantially simpler (and here I mean the checking of technical conditions, such as the martingale property) the terminology *"continuous-time model"* really means everything other than discrete-time models. So continuous-time models include i) continuous price processes, ii) quasi-left continuous processes and iii) processes with jumps at random but predictable times.

A.10 When are two stochastic processes the same?

A.10.1 Modification

One possible definition of a stochastic process is that it is a collection of random variables indexed by time. Two processes X and Y are then effectively the same if it is true that

$$P(X_t = Y_t) = 1 \text{ for all } t \in [0, T].$$

We then say that Y is a *modification* of X.

In Finance we work with *adapted* processes. Process $X := \{X_t\}_{t\in[0,T]}$ is said to be adapted to filtration $\mathbb{F}$ if X_t is $\mathcal{F}_t$-measurable for all $t \in [0, T]$. This just means that the randomness in X_t does not use any information revealed *after* time t. Now, the condition that $\mathcal{F}_0$ contains all the null sets of $\mathcal{F}_T$ is required so that we can be sure that *any* modification of an adapted process is again adapted.

A.10.2 Indistinguishability

It is also possible to view a stochastic process X as a collection of paths $X_\bullet(\omega)$. If we want to make sure that the *paths* of X and Y are the same with probability 1 then we need a stronger notion of equality between processes. This is because for different t the inequality $X_t \neq Y_t$ may occur on different null sets of paths and the union of these null sets over the uncountable $[0, T]$ may in fact add up to the whole Ω. We say that X and Y are *indistinguishable* if

$$P(\{\omega \in \Omega : X_t(\omega) \neq Y_t(\omega) \text{ for some } t \in [0, T]\}) = 0.$$

As we have noted above, two modifications of the same process are not necessarily indistinguishable. One needs to assume more structure to assign a unique set of paths to a collection of random variables. The default choice is to model prices as right-continuous processes with left limits. We say that a process X is *cádlág*[4], meaning its paths are right-continuous with left limits, if the paths $X_{\bullet}(\omega)$ are cádlág almost surely. It then follows that if X has two cádlág modifications then they are in fact indistinguishable, in other words a cádlág modification, if it exists, is unique (up to indistinguishability).

The right-continuity of filtration $\mathbb{F}$ is assumed to ensure that stochastic integrals and martingales do have an adapted cádlág modification. The right-continuity is needed to obtain results such as the Doob maximal inequality in continuous time.

A.11 Radon-Nikodym theorem

We say that measure $P^\star$ is absolutely continuous with respect to P (writing $P^\star \ll P$) if $P(A) = 0$ implies $P^*(A) = 0$, for all events $A \in \mathcal{F}_T$. The Radon-Nikodym[5] theorem states that $P^* \ll P$ if and only if there is a random variable Z such that

$$E^{P^\star}[1_A] = E[Z1_A] \text{ for all } A \in \mathcal{F}_T.$$

This random variable is unique P-almost surely and it is denoted $dP^\star/dP$. As an immediate consequence

$$E^{P^\star}[X] = E[ZX], \tag{A.6}$$

whenever one of the expressions in (A.6) is well defined.

If, additionally, we have $P \ll P^*$ then we say that P and $P^\star$ are equivalent, writing $P \sim P^\star$. In such case the notions of P-almost surely and P^*-almost surely coincide. Furthermore, it holds that

$$dP/dP^\star = 1/(dP^\star/dP).$$

[4]From the French *continue à droite avec des limites à gauche*.

[5]Johann Radon (b. 1887, Děčín, Czech Republic, d. 1957, Vienna) and Otton Marcin Nikodym (b. 1887 Zablotow, Ukraine, d. 1974 Utica, USA)

A.12 Conditional expectation

A.12.1 Absolutely integrable case

Let $\mathscr{G} \subset \mathscr{F}$. For any random variable Y on $(\Omega, \mathscr{F}, P)$ such that $\mathrm{E}[|Y|] < \infty$ there is (an almost surely unique) $\mathscr{G}$–measurable random variable denoted $E[Y|\mathscr{G}]$ such that

$$\mathrm{E}[1_B(Y - \mathrm{E}[Y|\mathscr{G}])] = 0 \qquad \text{for all } B \in \mathscr{G}.$$

The proof goes as follows. Without loss of generality we may assume $Y \geq 0$, otherwise write $Y = Y^+ - Y^-$ and work with $Y^+ \geq 0, Y^- \geq 0$ separately. Define a new measure $P^\star_{\mathscr{G}}$ on $(\Omega, \mathscr{G})$ by setting

$$P^\star_{\mathscr{G}}(B) = \frac{\mathrm{E}[1_B Y]}{E[Y]} \qquad \text{for all } B \in \mathscr{G}. \tag{A.7}$$

Letting $P_{\mathscr{G}}$ denote P restricted to $\mathscr{G}$ one has $P^\star_{\mathscr{G}} \ll P_{\mathscr{G}}$ on $(\Omega, \mathscr{G})$ and the Radon-Nikodym theorem shows there is an P–almost surely unique $\mathscr{G}$–measurable random variable $dP^\star_{\mathscr{G}}/dP_{\mathscr{G}}$ such that

$$P^\star_{\mathscr{G}}(B) = \mathrm{E}^{P^\star}_{\mathscr{G}}[1_B] = \mathrm{E}\left[1_B \frac{dP^\star_{\mathscr{G}}}{dP_{\mathscr{G}}}\right]. \tag{A.8}$$

Equating (A.7) with (A.8) shows that

$$\frac{1}{\mathrm{E}[Y]} \frac{dP^\star_{\mathscr{G}}}{dP_{\mathscr{G}}}$$

satisfies the conditions required by the conditional expectation. The almost sure uniqueness follows from the Radon-Nikodym theorem.

A.12.2 Generalized conditional expectation

When $\mathrm{E}[|Y|] = \infty$ there may still be a perfectly well-defined conditional expectation. To obtain it we proceed in two steps.

1. Define $\mathrm{E}[|Y|\,|\mathscr{G}] = \lim_{n\to\infty} \mathrm{E}[|Y|\mathbf{1}_{\{|Y|\leq n\}}|\mathscr{G}]$. This is a random variable that possibly takes the value of $+\infty$ but still has all the properties of conditional expectation claimed earlier.

2. On the set where $\mathrm{E}[|Y|\,|\mathscr{G}] < \infty$, Define $\mathrm{E}[Y|\mathscr{G}]$ by setting
$$\mathrm{E}[Y|\mathscr{G}] = \mathrm{E}[Y^+|\mathscr{G}] - \mathrm{E}[Y^-|\mathscr{G}].$$

This definition appears in Jacod and Shiryaev (2003, p. I.1.1).

A.13 Optional projection

Given an integrable random variable $Y \in L^1(\Omega, \mathcal{F}_T, P)$ it is natural to want to construct a new process

$$X_t = E_t[Y], \qquad t \in [0, T].$$

The problem with the construction as stated above is that for each fixed t the variable X_t is unique up to the null sets of the information algebra $\mathcal{F}_t$. For different values of t these null sets may be different and together they may cover the whole Ω so that in the end X as a process might not be defined properly on any path.

This problem is remedied by the notion of optional projection of non-adapted processes. We first construct the (non-adapted) constant process $Y_t := Y$. The optional projection theorem says there is a unique (up to indistinguishability) optional process denoted ${}^oY$ such that for any stopping time τ one has

$${}^oY_\tau 1_{\tau<\infty} = E_\tau[Y_\tau 1_{\tau<\infty}].$$

In particular, for any fixed $t \leq T$

$${}^oY_t = E_t[Y]$$

so ${}^oY$ is precisely the right object to represent the process $E_t[Y]$.

A.14 Density process of a change of measure

Given a probability measure on a large information set, say $\mathcal{F}_T$, we can always consider the same measure on a smaller information set. In our case the smaller information set will be $\mathcal{F}_t$ with $t < T$. The restriction of P to $\mathcal{F}_t$ is denoted $P|_{\mathcal{F}_t}$ or just P_t for short. Consider the following 3-period binomial example. The large information set is given by the partition

$$\mathcal{P}_3 = \{\{uuu\}, \{uud\}, \{udu\}, \{udd\}, \{duu\}, \{dud\}, \{ddu\}, \{ddd\}\}.$$

Suppose the measure P is defined by

$$P(\{uuu\}) = \frac{1}{8}, \ P(\{uud\}) = \frac{1}{16}, \ P(\{udu\}) = \frac{1}{4}, \ P(\{udd\}) = \frac{1}{16},$$
$$P(\{duu\}) = \frac{1}{8}, \ P(\{dud\}) = \frac{1}{8}, \ P(\{ddu\}) = \frac{1}{16}, \ P(\{ddd\}) = \frac{3}{16}.$$

On the smaller information set $\mathcal{F}_2$ generated by the partition

$$\mathcal{P}_2 = \{\{uuu, uud\}, \{udu, udd\}, \{duu, dud\}, \{ddu, ddd\}\}$$

the restricted measure P_2 reads

$$P_2(\{uuu, uud\}) = \frac{3}{16}, \; P_2(\{udu, udd\}) = \frac{5}{16},$$
$$P_2(\{duu, dud\}) = \frac{1}{4}, \; P_2(\{ddu, ddd\}) = \frac{1}{4}.$$

At $t = 1$ the information set is generated by the two-element partition

$$\mathcal{P}_1 = \{\{uuu, uud, udu, udd\}, \{duu, dud, ddu, ddd\}\}$$

and the restricted measure P_1 reads

$$P_1(\{uuu, uud, udu, udd\}) = \frac{1}{2},$$
$$P_1(\{duu, dud, ddu, ddd\}) = \frac{1}{2}.$$

Let us define a separate measure $P^\star$ by reversing the role of u and d in the definition of P. Once again we can compute the restrictions $P_2^\star$ and $P_1^\star$. In general, it is easily seen that if $P^\star$ is absolutely continuous with respect to P then $P_t^\star \ll P_t$ and hence we can compute the Radon-Nikodym derivatives

$$\frac{dP_t^\star}{dP_t}.$$

We will call the process $\left\{\frac{dP_t^\star}{dP_t}\right\}_{t\in[0,\infty)}$ the *density process* of the change of measure $dP^\star/dP$. Now, quite amazingly, the change of measure $dP_t^\star/dP_t$ can be computed directly from $dP^\star/dP$ as a conditional expectation, i.e.

$$\frac{dP_t^\star}{dP_t} = E_t\left[\frac{dP^\star}{dP}\right]. \tag{A.9}$$

Exercise A.1. *Compute $P_t^\star$ for $t = 1,2,3$ and using this result evaluate $dP_t^\star/dP_t$ for $t = 1,2,3$. Separately compute $\mathrm{E}_t\left[\frac{dP^\star}{dP}\right]$ for $t = 1,2$ and verify that equality (A.9) holds.*

A.15 Stopping time

A $[0,\infty]$–valued random variable τ is called a stopping time if the event $\tau \leq t$ belongs to $\mathcal{F}_t$ for all $t \in [0,T]$. Intuitively, τ is a stopping time if we can decide in every contingency whether we have stopped or not without having to wait for extra information that will only become available later. Typical example of a stopping time is "the first time you rode a motorbike". An example of a random time that is not a stopping time would be "the time you had your last ever motorbike ride".

In stochastic analysis it is common to define $s \wedge t := \min(s,t)$ and $s \vee t := \max(s,t)$. Let τ be a stopping time and X an adapted process. We define X^τ, the process stopped at τ, by

$$X_t^\tau = X_{t\wedge\tau}.$$

A.16 Localization

An increasing sequence of stopping times $\{\tau_n\}_{n=1}^\infty$ such that $\lim_{n\to\infty}\tau_n = \infty$ is called a localizing sequence. The idea of localization is that while process X itself may lack a certain desirable property, it may be possible that there is a localizing sequence $\{\tau_n\}_{n=1}^\infty$ such that each of the stopped processes X^{τ_n} does have this property.

Definition A.1. *Let $\mathcal{C}$ be a class of processes. We say that process X is locally in $\mathcal{C}$ and write $X \in \mathcal{C}_{\mathrm{loc}}$ if there is a localizing sequence of stopping times $\{\tau_n\}_{n=1}^\infty$ such that $X^{\tau_n} \in \mathcal{C}$ for each n.*

A.17 Martingales and UI martingales

Process X is a *martingale* if for all $s \le t$ one has $X_s = E_s[X_t]$. This means that the process does not change "on average" between any two time points $s \le t$, starting from *any* contingency at *any* time s.

Definition A.2. *We say that a collection of random variables $\{X_i\}_{i\in I}$ is* uniformly integrable (UI) under P *if for each $\varepsilon > 0$ there is K such that*

$$E[|X_i|1_{|X_i|>K}] < \varepsilon, \qquad \textit{for all } i \in I.$$

We say that a process X is uniformly integrable if the collection of random variables $\{X_t\}_{t\in[0,\infty)}$ is uniformly integrable.

Now, it can be shown that if X is a martingale and T is a fixed time then X^T (which is process X stopped at T) is uniformly integrable. This is achieved using the property $X_t = E_t[X_T]$ for all $t \in [0,T]$. In fact, more can be said. It is known that a UI collection of random variables is automatically bounded in $L^1(P)$.

Doob (1940) has shown that a martingale X bounded in $L^1(P)$ necessarily possesses an integrable limit $\lim_{t\to\infty} X_t = X_\infty$ and if X is UI then furthermore

$$\lim_{t\to\infty} E|X_t - X_\infty| = 0$$

and $X_t = E_t[X_\infty]$. We say that the UI martingale X is *closed* by X_∞. In Finance we typically work with a finite time horizon T and therefore can consider the stopped process

X^T which will be a UI martingale (with terminal value X_T) if X is a martingale. So on a finite time horizon there is no distinction between a martingale and UI martingale.

Example A.3. *Brownian motion is a martingale (and therefore UI martingale on every finite time interval) but it is not a UI martingale on the whole time line because its L^1–norm explodes,*

$$E[|W_t|] = \sqrt{\frac{2t}{\pi}} \overset{t\to\infty}{\to} \infty.$$

A.18 Martingales and local martingales

Let us now see how the concepts of the previous section relate to local martingales. Recall that a *local martingale* is a process which becomes a martingale when suitably stopped, as per Definition A.1.

We have seen above that every martingale is UI on every finite time interval. It is however *not* true that a *local* martingale uniformly integrable on a finite interval is already a martingale on that interval. Here one needs a stronger notion, supplied by the following definition due to Meyer (1962).

Definition A.4. *We say that process X is of class (D) if the set of random variables*

$$\{X_\tau : \tau \text{ is a } \textbf{finite} \text{ stopping time}\}$$

is uniformly integrable. We say that X is of class (DL) if the set

$$\{X_\tau : \tau \text{ is a } \textbf{bounded} \text{ stopping time}\}$$

is uniformly integrable.

Theorem A.5. *A local martingale is a uniformly integrable martingale if and only if it is of class (D). A local martingale is a martingale iff it is of class (DL).*

A.19 Predictable processes

We come to a concept which is probably hardest to get used to because it is based on thinking of a stochastic process as a map from $\mathbb{R}_+ \times \Omega \to \mathbb{R}$. The positive real line $\mathbb{R}_+$ represents the time while the elements of Ω label individual paths of a process. We equip $\mathbb{R}_+ \times \Omega$ with its own σ-algebra generated by sets $[0, t] \times A$ where $t > 0$ and $A \in \mathcal{F}$. Now recall from section A.5 that a random variable X generates its own σ-algebra. In the same way a *process* $X : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ generates its σ-algebra. Now consider the σ-algebra generated by *all left-continuous processes*. This algebra on $\mathbb{R}_+ \times \Omega$ is called the *predictable* σ–algebra, denoted $\mathcal{P}$. We say that process X is predictable iff X as a mapping from $\mathbb{R}_+ \times \Omega$ to $\mathbb{R}$ is $\mathcal{P}$–measurable.

The label *predictable* may seem to suggests that the given process does not have much short-term randomness. This intuition is wrong because a Brownian motion is continuous and therefore predictable.

A.20 Semimartingales

As we have seen in Section A.10.1 a process is determined uniquely only once its path properties have been fixed. The default choice is to model asset prices as right-continuous with left limits (càdlàg). This brings us to the notion of a semimartingale. Process X is a *semimartingale* if its paths are almost surely *cadlag* and if there is a decomposition

$$X = A + M$$

where A is a process of finite variation and M is a local martingale. This decomposition is not unique because there are many local martingales with paths of finite variation and these can be moved from A to M, or vice versa. A prototypical example of a process that belongs to both A and M is a compensated Poisson process.

For a semimartingale X we denote by X_- the process of left limits $X_{t-}(\omega) = \lim_{s\nearrow t} X_t(\omega)$. Jumps of X are captured as the difference between the original process and its left limit, $\Delta X = X - X_-$. Note that while ΔX is again a semimartingale, process X_- is not (it is not right-continuous to begin with). We know however that X_- is predictable (automatically, because X_- is left-continuous) and locally bounded (thanks to left-continuity the supremum process of $X-$ is finite).

We say that X is a *special semimartingale* if

$$X = X_0 + B^X + M^X$$

where B^X is a *predictable* process of finite variation and M^X is a local martingale, $B_0^X = M_0^X = 0$. This decomposition *is* unique and it is called the *canonical decomposition of* X. Not all semimartingales are special, but those with bounded jumps always are. In particular all continuous semimartingales are special.

A.21 Stochastic integration

Given an arbitrary semimartingale X stochastic integral $\int \eta dX$ can be defined for any locally bounded predictable process η. By default the paths of $\int \eta dX$ are taken to be cádlág and the integral process then becomes a new semimartingale. Note that η itself need not be a semimartingale because it may not be right-continuous. In particular

X_-, which is not a semimartingale in general, can be used to integrate any other semimartingale.

If X itself has finite variation then it is possible to define $\int \eta dX$ also for integrands that are not predictable, but in general such integration is not well defined.

A.22 Stochastic differential equations

Let X be a (multidimensional) semimartingale. The SDE $dY = a(Y_-)dX$ has a unique solution if function a is locally Lipschitz. The same SDE will have a solution that exists at all times (does not explode) if function a grows at most linearly. For example the SDE

$$\begin{aligned} dr_t &= (\alpha - \beta r_t)dt + \sqrt{r_t}dW_t, \\ r_0 &> 0, \end{aligned}$$

with $X \equiv (t, W)$ and $a(r) = (\alpha - \beta r, \sqrt{r})$ satisfies the linear growth condition and also the local Lipschitz condition on the set $(0, \infty)$.

Another ubiquitous example is the solution of

$$dY_t = Y_{t-}dX,$$

which is denoted by $\mathscr{E}(X)$ and called the stochastic exponential of X. In finance dX could represent the rate of return on a particular asset and $\mathscr{E}(X)$ would then be the value of a fund whose gains are being continuously reinvested in the asset with return dX.

References

Dellacherie, C. and P.-A. Meyer (1978). *Probabilities and potential*. Vol. 29. North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York, pp. viii+189. ISBN: 0-7204-0701-X.

Doléans-Dade, C. and P.-A. Meyer (1969). "Intégrales stochastiques par rapport aux martingales locales". In: *C. R. Acad. Sci. Paris Sér. A-B* 269, A144–A147.

— (1970). "Intégrales stochastiques par rapport aux martingales locales". In: *Séminaire de Probabilités IV, Strasbourg*. Vol. 124. Lecture Notes in Mathematics. Springer, Berlin, pp. 77–107.

Doléans, C. (1967). "Processus croissants naturels et processus croissants très-bien-mesurables". In: *C. R. Acad. Sci. Paris Sér. A* 264, pp. 874–876. ISSN: 0151-0509.

— (1969). "Variation quadratique des martingales continues à droite". In: *Ann. Math. Statist.* 40, pp. 284–289. ISSN: 0003-4851.

Doob, J. L. (1940). "Regularity properties of certain families of chance variables". In: *Transactions of the American Mathematical Society* 47, pp. 455–486. ISSN: 0002-9947.
— (1953). *Stochastic processes*. John Wiley & Sons, Inc., New York, pp. viii+654.
Fisk, D. L. (1965). "Quasi-martingales". In: *Trans. Amer. Math. Soc.* 120, pp. 369–389. ISSN: 0002-9947.
Girsanov, I. V. (1960). "On transforming a class of stochastic processes by absolutely continuous substitution of measures". In: *Theory of Probability and its Applications* 5.3, pp. 285–330. ISSN: 0040-361x.
Hunt, G. A. (1956). "Some theorems concerning Brownian motion". In: *Trans. Amer. Math. Soc.* 81, pp. 294–319. ISSN: 0002-9947.
Itô, K. and S. Watanabe (1965). "Transformation of Markov processes by multiplicative functionals". In: *Annales de l'Institut Fourier (Grenoble)* 15.1, pp. 13–30. ISSN: 0373-0956.
Itô, K. (1944). "Stochastic integral". In: *Proc. Imp. Acad. Tokyo* 20, pp. 519–524.
— (1951). "On a formula concerning stochastic differentials". In: *Nagoya Math. J.* 3, pp. 55–65. ISSN: 0027-7630.
Jacod, J. and J. Mémin (1976). "Caractéristiques locales et conditions de continuité absolue pour les semi-martingales". In: *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* 35.1, pp. 1–37.
Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes*. 2nd. Vol. 288. Comprehensive Studies in Mathematics. Springer-Verlag, Berlin, pp. xx+661. ISBN: 3-540-43932-3.
Kunita, H. and S. Watanabe (1967). "On square integrable martingales". In: *Nagoya Math. J.* 30, pp. 209–245. ISSN: 0027-7630.
Lenglart, E. (1977). "Transformation des martingales locales par changement absolument continu de probabilités". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 39.1, pp. 65–70.
Lépingle, D. (1976). "La variation d'ordre p des semi-martingales". In: *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* 36.4, pp. 295–316.
Mazliak, L. (2009). "How Paul Lévy saw Jean Ville and martingales". In: *J. Électron. Hist. Probab. Stat.* 5.1, p. 24. ISSN: 1773-0074.
McKean Jr., H. P. (1969). *Stochastic Integrals*. Probability and Mathematical Statistics, No. 5. Academic Press, New York-London, pp. xiii+140.
Meyer, P.-A. (1962). "A decomposition theorem for supermartingales". In: *Illinois Journal of Mathematics* 6, pp. 193–205. ISSN: 0019-2082.
— (1963). "Decomposition of supermartingales: the uniqueness theorem". In: *Illinois Journal of Mathematics* 7, pp. 1–17. ISSN: 0019-2082.
— (1966). *Probability and potentials*. Blaisdell Publishing Co., Waltham MA, pp. xiii+266.
— (1967). "Intégrales stochastiques. I, II, III, IV". In: *Séminaire de Probabilités I, Strasbourg*. Lecture Notes in Math. Springer, Berlin, pp. 72–94, 95–117, 118–141, 142–162.

Meyer, P.-A. (1976). "Un cours sur les intégrales stochastiques". In: *Séminaire de Probabilités X, Strasbourg*. Vol. 511. Lecture Notes in Math. Springer, Berlin, pp. 245–400.

— (2009). "Stochastic processes from 1950 to the present". In: *Journal Électronique d'Histoire des Probabilités et de la Statistique* 5.1, p. 42. ISSN: 1773-0074.

Millar, P. W. (1968). "Martingale integrals". In: *Trans. Amer. Math. Soc.* 133, pp. 145–166. ISSN: 0002-9947.

Monroe, I. (1972). "On the γ-variation of processes with stationary independent increments". In: *Ann. Math. Statist.* 43, pp. 1213–1220. ISSN: 0003-4851.

— (1976). "Almost sure convergence of the quadratic variation of martingales: a counterexample". In: *Ann. Prob.* 4.1, pp. 133–138.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations*. 2nd. Vol. 21. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, pp. xiv+419. ISBN: 3-540-00313-4.

Van Schuppen, J. H. and E. Wong (1974). "Transformation of local martingales under a change of law". In: *Annals of Probability* 2, pp. 879–888.

Ville, J. (1939). *Étude critique de la notion de collectif*. NUMDAM, Paris, p. 116.

Yoeurp, C. (1976). "Décompositions des martingales locales et formules exponentielles". In: *Séminaire de Probabilités X, Strasbourg*. Vol. 511. Lecture Notes in Math. Springer, Berlin, pp. 432–480.

Chapter 1

(Mostly) continuous-time stochastic calculus with some applications at the end

The material in this section is reasonably standard an you will have seen it already. The novelty, such as it is, lies in the representation of stochastic processes. We write (represent) each stochastic process in a measure-invariant way. For Itô processes this approach goes back to McKean (1969) and is by now quite standard, see for example Grigoriu (2002), although it has yet to appear in standard asset pricing textbooks. The measure-invariant representation has many advantages as it is easier to write down and memorise, especially when it comes to measure changes. As a background reading you may wish to download Černý and Ruf (2019) discussing measure-invariant calculus with jumps.

1.1 The time process

Strictly speaking X_t is not a process, it is just one random variable from the collection which defines the process X. The worst abuse of notation which often leads to students' confusion occurs for the time process. Suppose we denote the time process by I, then $I_t = t$. This would be the correct way to write things. However, everyone denotes the time process by t so the symbol t is at once the whole process and also one specific value. This bad notation is so pervasive that one has to learn to live with it.

1.2 Integral notation

Consider a filtered probability space and on this space consider a semimartingale X and a *locally bounded* predictable process η. We can then define a new semimartingale

Y by setting

$$Y_t = Y_0 + \int_0^t \eta_s dX_s. \tag{1.1}$$

This follows from Dellacherie and Meyer (1978, Theorem VIII.3) by localization. Historically, the stochastic integral (1.1) is constructed by decomposing X into two components, one of which is a (locally) square-integrable martingale. Such decomposition is clearly measure-specific. However, the resulting stochastic integral is measure-invariant (Meyer, 1976, Theorem VI.26). It therefore makes sense to write the integral (1.1) without decomposing X.

There is a helpful criterion to decide local boundedness of a predictable process (Larsson and Ruf, 2014, Proposition 3.2(vi)):

$$\text{predictable process } \eta \text{ is locally bounded} \iff \sup_{0\leq s\leq t} |\eta_s| < \infty \quad \text{for all } t > 0.$$

By this criterion, **any left-continuous finite-valued process is locally bounded**.

> Notice that η_s in (1.1) must not depend on t by construction. If it did the calculus below would not be valid any longer (see Example 1.2).

We write (1.1) in short

$$dY_t = \eta_t dX_t. \tag{1.2}$$

We will mostly use the differential form (1.2) but every now and then we will need to convert it back to the integral. This is done by means of the (trivial) statement

$$Y_t = Y_0 + \int_0^t dY_s \tag{1.3}$$

which says that terminal value Y_t is obtained from the initial value Y_0 by adding all the increments of process Y in the time interval $[0, t]$.

Example 1.1. *Suppose*

$$dY_t = tdt, \tag{1.4}$$

with $Y_0 = 1$. Evaluate Y_t.

Solution: *This is a good example of a possible mixup in notation. In the expression for dY_t symbol t represents a process, not a fixed date. The correct interpretation of (1.4) is*

$$dY_s = sds, \tag{1.5}$$

and not $dY_s = tds$. On using (1.5) in equation (1.3) we obtain the correct result

$$Y_t = Y_0 + \int_0^t dY_s = 1 + \int_0^t sds = 1 + t^2/2. \tag{1.6}$$

Example 1.2. *Suppose process X is given and we define Y as an exponentially-weighted moving average of X values*

$$Y_t = \int_0^t e^{-\alpha(t-s)} X_s ds.$$

Find the expression for dY_t.
Solution: *In this case we cannot write an expression for dY_t directly because the integrand $e^{-\alpha(t-s)}X_s$ depends on t. Such situation is not covered by stochastic integration directly. In this case we are lucky enough to be able to write*

$$Y_t = e^{-\alpha t} \int_0^t e^{\alpha s} X_s ds,$$

and now the integral has the right form. On dividing by $e^{-\alpha t}$ we obtain

$$e^{\alpha t} Y_t = \int_0^t e^{\alpha s} X_s ds,$$

In this way, we have obtained an expression for $d(e^{\alpha t} Y_t)$

$$d(e^{\alpha t} Y_t) = e^{\alpha t} X_t dt.$$

To come up with an explicit expression for dY_t we would need the Itô–Meyer formula which is explained below in Proposition 1.3. We anticipate slightly by saying that the Itô–Meyer formula yields

$$d(e^{\alpha t} Y_t) = e^{\alpha t} dY_t + \alpha e^{\alpha t} Y_t dt$$

and after rearranging one obtains $dY_t = (X_t - \alpha Y_t)dt$.

1.3 Quadratic covariation

Suppose X, Y are two *arbitrary* semimartingales. Then the process with value

$$X_t Y_t - X_0 Y_0 - \int_0^t X_{s-} dY_s - \int_0^t Y_{s-} dX_s$$

is well defined. We call it the *quadratic covariation* of X and Y, and its value is denoted by $[X, Y]$. The process $[X, X]$ is called the *quadratic variation* of X. In differential notation

$$\begin{aligned} d[X,Y]_t &= d(X_t Y_t) - X_{t-} dY_t - Y_{t-} dX_t \\ [X,Y]_0 &= 0. \end{aligned}$$

After rearranging

$$d(X_t Y_t) = X_{t-} dY_t + Y_{t-} dX_t + d[X,Y]_t. \tag{1.7}$$

- It is not at all obvious from (1.7) but the process $[X, X]$ is increasing. And because $[X, X]$ is well-defined, i.e. finite-valued, this means $[X, X]$ is automatically of **finite variation (FV)**.
- It is much easier to see that $[\cdot, \cdot]$ is bilinear, therefore satisfies the *polarization* identity
$$[X, Y] = \frac{1}{4}\left([X+Y, X+Y] - [X-Y, X-Y]\right).$$
This means $[X, Y]$, too, is of finite variation.

1.3.1 Simplified notation

Some textbooks use $dX_t dY_t$ instead of $d[X, Y]_t$ and we shall adopt this convention here. This has the following advantages. The following box works for all semimartingales.

First, the expression (1.7) evokes a 2nd order Taylor expansion[a]
$$d(X_t Y_t) = X_{t-} dY_t + Y_{t-} dX_t + dX_t dY_t.$$
Secondly, the intuition this creates is the right one; $[X, X]$ does behave like a quadratic expression in that
$$[X+Y, X+Y] = [X, X] + 2[X, Y] + [Y, Y],$$
which in the $(dX_t)^2$ notation leads to a very natural statement
$$(dX_t + dY_t)^2 = (dX_t)^2 + 2dX_t dY_t + (dY_t)^2.$$
Likewise, the homogeneity property
$$\left[\int_0^{\cdot} \eta_t dX_t, \int_0^{\cdot} \xi_t dY_t\right] = \int_0^{\cdot} \eta_t \xi_t d[X, Y]_t,$$
takes a very natural form in
$$(\eta_t dX_t)(\xi_t dY_t) = \eta_t \xi_t dX_t dY_t.$$

[a]This is even more obvious with $Y = X$,
$$dX_t^2 = 2X_{t-} dX_t + (dX_t)^2.$$

1.3.2 Time and quadratic covariation

Time process has a very special property in that its quadratic covariation with any other semimartingale is zero,

$$dtdX_t = 0. \tag{1.8}$$

As a special case we obtain $(dt)^2 = 0$. Proposition 1.10 explores this property in a more general context.

1.4 Itô–Meyer formula for continuous processes

For $g : \mathbb{R}^n \to \mathbb{R}$ we introduce the notation

$$\partial_i g(x) = \frac{\partial g(x_1, x_2, \ldots, x_n)}{\partial x_i}, \qquad \partial_{ij} g(x) = \frac{\partial^2 g(x_1, x_2, \ldots, x_n)}{\partial x_i \partial x_j}, \qquad i, j \in \{1, \ldots, n\}$$

where such derivatives exist.

Proposition 1.3. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function on an open set $\mathcal{U} \subseteq \mathbb{R}^n$ and $X = (X^{(1)}, X^{(2)}, \ldots, X^{(n)})$ a continuous semimartingale with values in $\mathcal{U}$. Then $g(X)$ is a continuous semimartingale and we have*

$$dg(X_t) = \sum_{i=1}^{n} \partial_i g(X_t) dX_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} g(X_t) dX_t^{(i)} dX_t^{(j)}. \tag{1.9}$$

Proof. See Dellacherie and Meyer (1978, Theorem VIII.27) and the paragraph preceding it. □

The idea of writing the Itô formula in this way was first presented in McKean (1969). The formula (1.9) is however more general and does not assume that X is an Itô process, merely that it is continuous. In this generality it appears in Meyer (1967, Theorem II.3) but see also Kunita and Watanabe (1967, Theorem 2.2).

Example 1.4. *Evaluate the expression $d(e^{\alpha t} Y_t)$.*

Method 1: *Set $g(t, Y_t) = e^{\alpha t} Y_t$. Evaluate the necessary partial derivatives*

$$\partial_1 g(t, Y_t) = \alpha e^{\alpha t} Y_t, \qquad \partial_2 g(t, Y_t) = e^{\alpha t},$$
$$\partial_{11} g(t, Y_t) = \alpha^2 e^{\alpha t} Y_t, \qquad \partial_{12} g(t, Y_t) = \alpha e^{\alpha t}, \qquad \partial_{22} g(t, Y_t) = 0.$$

Substitute these into the Itô formula (1.9),

$$
\begin{aligned}
d(e^{\alpha t}Y_t) = dg(t,Y_t) =& \partial_1 g(t,Y_t)dt + \partial_2 g(t,Y_t)dY_t \\
&+ \partial_{11} g(t,Y_t)(dt)^2 + 2\partial_{12} g(t,Y_t)(dtdY_t) + \partial_{22} g(t,Y_t)(dY_t)^2, \\
=& \alpha e^{\alpha t}Y_t dt + e^{\alpha t}dY_t.
\end{aligned}
$$

We have used $(dt)^2 = (dtdY_t) = 0$ from equation (1.8).

Method 2: *It is faster to employ equation (1.7) directly, where we think of $e^{\alpha t}$ as being the process X*

$$d(e^{\alpha t}Y_t) = e^{\alpha t}dY_t + Y_t de^{\alpha t} + dY_t de^{\alpha t}. \tag{1.10}$$

Now use the Itô formula to obtain $de^{\alpha t} = \alpha e^{\alpha t}dt$ and plug this back into (1.10),

$$d(e^{\alpha t}Y_t) = e^{\alpha t}dY_t + \alpha Y_t e^{\alpha t}dt.$$

We have again used the property (1.8) to obtain $dtdY_t = 0$.

Example 1.5. *Consider two continuous processes J, S such that $S > 0$. Write down the Itô–Meyer formula for the process $Y = J/S$. This situation arises in the pricing of Asian options.*

Solution: *Let $g(x) = \frac{x_1}{x_2}$ for $x_1 \in \mathbb{R}$ and $x_2 \neq 0$. Observe the open set $\mathcal{U}$ appearing in Proposition 1.3 in this case is given by the half-plane $\mathcal{U} = \mathbb{R} \times (0,\infty)$. First we evaluate the necessary partial derivatives*

$$
\begin{aligned}
\partial_1 g(x) = \frac{1}{x_2}, \qquad \partial_2 g(x) = -\frac{x_1}{x_2^2}, \\
\partial_{11} g(x) = 0, \qquad \partial_{12} g(x) = -\frac{1}{x_2^2}, \qquad \partial_{22} g(x) = 2\frac{x_1}{x_2^3}.
\end{aligned}
$$

The Itô formula (1.9) reads

$$
\begin{aligned}
dY_t =& d\frac{J_t}{S_t} = dg(J_t,S_t) \\
=& \partial_1 g(J_t,S_t)dJ_t + \partial_2 g(J_t,S_t)dS_t \\
&+ \frac{1}{2}(\partial_{11} g(J_t,S_t)(dJ_t)^2 + 2\partial_{12} g(J_t,S_t)dJ_t dS_t) + \partial_{22} g(J_t,S_t)(dS_t)^2) \\
=& \frac{dJ_t}{S_t} - J_t\frac{dS_t}{S_t^2} + \frac{1}{2}\left(0 \times (dJ_t)^2 - 2\frac{dJ_t dS_t}{S_t^2} + 2J_t\frac{(dS_t)^2}{S_t^3}\right) \\
=& \frac{dJ_t}{S_t} - J\frac{dS_t}{S_t^2} - \frac{dJ_t dS_t}{S_t^2} + J_t\frac{(dS_t)^2}{S_t^3}.
\end{aligned}
$$

1.5 Information structure (filtration)

1.5.1 Filtration

Filtration captures different amounts of information available at different time points. Mathematically it is a collection of increasing σ–fields indexed by time,

$$\mathbb{F} = \{\mathscr{F}_t\}_{t\in[0,T]}.$$

Intuitively, in a discrete setting where all possible outcomes are captured by a non-recombining tree the σ–algebra $\mathscr{F}_t$ is generated by all the nodes in the tree at time t.From here it is obvious that $\mathscr{F}_s \subset \mathscr{F}_t$ for $s \leq t$ since there are necessarily fewer nodes at time s than there are nodes at time t. For more details consult Černý (2009, Chapter 8).

1.5.2 What is known and when

In applications it is important to distinguish time points when the value of a certain random variable is already known and time points when it is still perceived as being random. To express mathematically that a random variable X is known at time t we say that X is $\mathscr{F}_t$–measurable. In a discrete setting this means that X assigns just one value (as opposed to two or more) to each node in the non-recombining tree at time t.

A process Y such that Y_t is known at time t for all $t \in [0, T]$ is called an *adapted process*. In discrete time a process Y is *predictable* if Y_t is known at $t - 1$ for all t. A process Y is said to be deterministic if the value of Y_t is known already at time 0 for all t. The understanding of *when* a particular information is revealed is important when calculating expected values.

1.6 Expectation operator

1.6.1 Conditional expectation

Given a filtration $\mathbb{F} = \{\mathscr{F}_t\}_{t\in[0,T]}$ one denotes the expectation of random variable X conditional on the information at time t by $\mathrm{E}[X|\mathscr{F}_t]$. We will write more compactly $\mathrm{E}_t[X] = \mathrm{E}[X|\mathscr{F}_t]$ when the filtration is fixed.[1] The quantity $\mathrm{E}_t[X]$ is a random variable and it assigns one value to each node in a non-recombining tree at time t. These values

[1]In these notes, we work with one filtration at a time, so the shorter notation $\mathrm{E}_t[Y]$ is perfectly unambiguous. In contrast, in the literature on insider trading one uses one filtration, $\mathbb{F}$, for uninformed traders and another, larger filtration, $\mathbb{G}$, for insiders. Then one has to use the full notation $\mathrm{E}[Y|\mathscr{F}_t], \mathrm{E}[Y|\mathscr{G}_t]$.

may be different from node to node for given t. This means $\mathrm{E}_t[X]$ is known at time t but it is generally not known at time 0.

1.6.2 Generalized conditional expectation

Example 1.6. *Consider the following model. Let U take the values ± 1 with equal probability. Let V be a random variable, independent of U, with zero mean and infinite variance. Let $X_0 = 0$, $X_1 = V$, and $X_2 = V + UV^2$. Finally, consider the filtration generated by X, i.e., $\mathscr{F}_0$ trivial, $\mathscr{F}_1 = \sigma(V)$, and $\mathscr{F}_2 = \sigma(U, V)$.*

The conditional expectation $\mathrm{E}_1[X_2]$, as described in Appendix A, is not well-defined because $\mathrm{E}[|X_2|] = \infty$. However, the mean of U is zero, hence one should be able to write

$$\mathrm{E}[V + UV^2 | V] = V. \tag{1.11}$$

This, however, needs a more careful definition of conditional expectation, or rather a condition less restrictive than $\mathrm{E}[|X_2|] = \infty$.

The idea of *generalized conditional expectation* is to build $\mathrm{E}_1[X_2]$ piece-wise. Suppose there is an event $\mathcal{P} \in \mathscr{F}_1$ such that $\mathrm{E}[|X_2|\mathbf{1}_{\mathcal{P}}] < \infty$. Then the conditional expectation is well-defined on the subset $\mathcal{P}$ in the standard way and therefore it is defined everywhere if we can find a sequence of such events $\mathcal{P}$ that cover the whole Ω. In the present example one can take $\mathcal{P} = \{\omega \in \Omega : |V(\omega)| \leq n\}$ for $n \in \mathbb{N}$ and this makes the expression (1.11) rigorous.

1.7 Predictable characteristics

1.7.1 Canonical decomposition of special semimartingale

Every semimartingale X with bounded jumps has a unique *canonical decomposition*

$$dX_t = dB_t^X + dM_t^X,$$

where B^X is a predictable process of finite variation and M^X is a *local* P–martingale, both starting at zero. As discussed above B^X corresponds to a no-surprise component while M^X is a purely random component with conditional mean zero. In discrete time this decomposition looks very natural

$$X_t = \underbrace{\mathrm{E}_{t-1}[X_t]}_{\text{predictable part}} + \underbrace{(X_t - \mathrm{E}_{t-1}[X_t])}_{\substack{\text{random part with} \\ \text{zero conditional mean}}} .$$

If we now subtract X_{t-1} on both sides we will obtain the discrete-time version of the canonical decomposition

$$\underbrace{X_t - X_{t-1}}_{\Delta X_t} = \underbrace{E_{t-1}[X_t] - X_{t-1}}_{\Delta B_t^X} + \underbrace{(X_t - E_{t-1}[X_t])}_{\Delta M_t^X}.$$

The canonical decomposition *depends on the chosen probability measure*, but the process X itself does not. When dealing with Itô processes you are free to think of M^X as the "Brownian motion part" of X. We refer to B^X as the cumulative drift. In this sense, a predictable process of finite variation is a "pure drift".

1.7.2 Predictable characteristics of a continuous process

A peculiar property of a continuous semimartingale X is that its quadratic variation $[X, X]$ is always predictable and of finite variation. That is $[X, X]$ is a pure drift. This is generally no longer true when X has jumps.

We call $(B^X, C^X = [X, X])$ the *predictable characteristics* of X. By construction the drift part B^X is sensitive to the chosen probability measure, but C^X is not. We now wish to relate B^X, C^X to the drift and volatility of X.

1.7.3 Differential characteristics, Itô process, drift and volatility

A continuous semimartingale X is called an Itô process if there are predictable processes b^X, c^X such that

$$dB_t^X = b_t^X dt,$$
$$dC_t^X = c_t^X dt.$$

We call b^X the *drift* and $\sqrt{c^X}$ the *volatility* of process X.

Often one needs to consider a multivariate process X, in which case $b^X(t)$ is vector-valued and $c^X(t)$ a positive semidefinite matrix-valued (i.e. c^X behaves like a variance-covariance matrix). In such case $\sqrt{c_{ii}^X}$ is the volatility of $X^{(i)}$ while c_{ij}^X is the covariation rate between $X^{(i)}$ and $X^{(j)}$ i.e.

$$dX^{(i)} dX^{(j)} = c_{ij}^X dt.$$

1.7.4 Itô process with Markov property = diffusion

Proposition 1.7. *If there are deterministic functions f, g such that*

$$b^X(t) = f(t, X_t),$$
$$c^X(t) = g(t, X_t),$$

then X is Markov under P, meaning that for any smooth bounded function h and any $s \leq u$ the conditional expectation

$$\mathrm{E}_s[h(X_u)]$$

is a function of s and X_s only. In such case we call process X a diffusion under *P.*

When the drift and volatility are independent of time

$$b^X(t) = f(X_t),$$
$$c^X(t) = g(X_t),$$

we say that the process X is *time-homogeneous* under P.

1.8 Processes with independent increments

Theorem 1.8. *A continuous semimartingale is a process with independent increments (PII) under measure P if and only if B^X and C^X are deterministic.*

Proof. Jacod and Shiryaev (2003, Theorem II.4.15). □

The next statement shows that continuous PII-s are *normally distributed*. This forms the basis of most results in these notes.

Proposition 1.9. *If a continuous semimartingale X is a PII under P then*

1. *for any $s \leq t$ we have*

$$X_t | \mathscr{F}_s \sim N(X_s + B_t^X - B_s^X, C_t^X - C_s^X).$$

In other words the conditional P–distribution of X_t as of time s is Gaussian.

2. *For each $n \geq 1$ and any times $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$, the random variables $\{X_{t_r} - X_{t_{r-1}}\}_{r=1,\ldots,n}$ are independent.*

Proof. Jacod and Shiryaev (2003, Theorem II.4.4). □

Observe that a continuous PII process is not necessarily an Itô process. For example, $C^X(t)$ could be the cumulative distribution of the Cantor measure on $[0, 1]$ hence not absolutely continuous with respect to t.

1.8.1 Brownian motion: continuous, time-homogeneous PII

An Itô process with zero drift and constant volatility equal to 1 is called the *standard Brownian motion*. Thus an Itô process W is a standard Brownian motion under P iff

$$b^W = 0 \text{ and } c^W = 1.$$

It is customary to write an Itô process X with characteristics b^X, c^X as follows

$$dX_t = b_t^X dt + \sqrt{c_t^X} dW_t. \tag{1.12}$$

However, this has no clear advantages and becomes counterproductive when one deals with several correlated processes.

1.8.2 Computation of drift and volatility from the Itô formula

To illustrate computations *without* Brownian motion consider the univariate form of the Itô formula

$$\begin{aligned} dg(X_t) &= g'(X_t)dX_t + \frac{1}{2}g''(X_t)(dX_t)^2 \\ &= g'(X_t)dX_t + \frac{1}{2}g''(X_t)c_t^X dt. \end{aligned} \tag{1.13}$$

This formula stresses the important fact that $g(X_t)$ does not depend on the probability measure. There is no need to substitute (1.12) for dX_t.

If we wish to calculate the P–drift of $g(X)$ we will simply evaluate it as the P–drift of the right-hand side in (1.13)

$$b_t^{g(X)} = g'(X_t)b_t^X + \frac{1}{2}g''(X)c_t^X.$$

Likewise, to evaluate the volatility of $g(X)$ we use the property (1.8) to obtain from (1.13),

$$\begin{aligned} (dg(X_t))^2 &= (g'(X_t))^2(dX_t)^2 = (g'(X_t))^2 c_t^X dt, \\ c_t^{g(X)} &= (g'(X_t))^2 c_t^X. \end{aligned}$$

1.9 Further properties of quadratic covariation

Proposition 1.10. *The following statements hold:*

1. *Suppose X is a predictable process of finite variation and Y is a general continuous*

process. Then

$$[X, Y] = 0.$$

2. *Suppose M is a local martingale starting at 0. Then $[M, M] = 0$ if and only if $M = 0$.*

Proof. See Jacod and Shiryaev, 2003, I.4.49d) and I.4.50d). □

In the context of Itô processes part 1 of Proposition 1.10 implies

$$dtdY = 0,$$

for any Itô process Y. The 2^{nd} part of Proposition 1.10 shows that an Itô process X has zero quadratic variation if and only if it is of the form $X = \int_0^t b_s^X ds$, i.e. if and only if it is equal to its drift part.

Please note that in our simplified notation $(dY_t)^2 = c_t^Y dt$ which in words means that "quadratic variation of Y grows at the rate c^Y". We are NOT allowed to take a square root here, i.e. it is not true that $dY_t = \sqrt{c_t^Y dt}$.

1.9.1 Portfolio theorem for Itô processes

Theorem 1.11. *Suppose that X is an n-dimensional Itô process with covariation matrix c^X. Consider two processes $dY_t = \eta_t dX_t$ and $dZ = \xi dX$ with η, ξ being $n \times 1$–valued predictable processes. Then*

$$\begin{aligned} dY_t dZ_t &= \sum_{i=1}^n \eta_t^{(i)} dX_t^{(i)} \sum_{j=1}^n \xi_t^{(j)} dX_t^{(j)} \\ &= \sum_{i,j=1}^n \eta_t^{(i)} \xi^{(j)} dX_t^{(i)} dX_t^{(j)} = \sum_{i,j=1}^n \eta_t^{(i)} \xi_t^{(j)} c_{ij}^X dt \\ &= \eta c^X \xi^\top dt. \end{aligned}$$

1.9.2 Quadratic variation and variance

There is a link between conditional variance and quadratic variation of a diffusion. Suppose we have

$$\begin{aligned} b^X &= f(t, X), \\ c^X &= g(t, X), \end{aligned}$$

in other words suppose that X is a diffusion. Under some technical conditions which govern the variability of f and g we have

$$\begin{aligned} \mathrm{E}_t[X_{t+\delta} - X_t] &= b^X_t \delta + o(\delta), \\ \mathrm{Var}_t(X_{t+\delta} - X_t) &= c^X_t \delta + o(\delta), \end{aligned}$$

where the quantity $o(\delta)$ is "much smaller than δ as $\delta \to 0$". This means that for a very short time step δ the conditional variance of the shock to X (shock $= X_{t+\delta} - X_t$) is equal (with error $o(\delta)$) to the increase in the quadratic variation process. Another way to say this is to write

$$\lim_{\delta \to 0} \frac{\mathrm{Var}_t(X_{t+\delta} - X_t)}{\delta} = c^X_t,$$

which means that the variance of the shock per unit of time equals the quadratic variation rate of the process.

> The link between quadratic variation and variance in general only holds over *short* time horizons. Long-term it only holds in the special case of the continuous PII, see Proposition 1.9.

For the deterministic (PII) process $X_t = t$, we have $[X, X] = 0$ and $\mathrm{Var}(X_t) = 0$. On the other hand, consider $X = \int_0^t W_s ds$ where W is a standard Brownian motion. We still have $[X, X] = 0$ but $\mathrm{Var}(X) = t^3/3 > 0$ as shown in the example below. This means that there are many processes which have zero conditional variance over a short period of time but non-zero variance over long horizons. These processes are made of pure drift (hence zero conditional variance over short horizons) but because the drift is itself random the cumulative effect of that randomness leads to non-zero variance over long time horizons. In financial context a good example is the cumulative return on a bank account with risk-free short rate, when the short rate is itself stochastic.

Example 1.12. *Let us evaluate the variance of* $\int_0^T W_s ds$ *with* W *being the standard Brownian motion starting at* $W_0 = 0$.

Solution: *The solution uses the Itô formula and integration of a Gaussian process. Itô formula tells us*

$$d(tW_t) = tdW_t + W_t dt + dW_t dt = tdW_t + W_t dt.$$

Rearrange to get $W_t dt$ *on the left-hand-side*

$$W_t dt = d(tW_t) - tdW$$

and write this as an integral $\int_0^T$

$$\begin{aligned}\int_0^T W_s ds = &W_T T - W_0 0 - \int_0^T s dW_s \\ =&(W_T - W_0)T - \int_0^T s dW_s \\ =&T\int_0^T dW_s - \int_0^T s dW_s \\ =&\int_0^T (T-s)dW_s.\end{aligned}$$

Consequently proposition 1.9, part 3) yields

$$\mathrm{Var}(\int_0^T W_s ds) = \mathrm{Var}(\int_0^T (T-s)dW_s) = \int_0^T (T-s)^2 ds = T^3/3.$$

1.10 Zero drift and martingales

It follows from the canonical decomposition that X is a *local martingale* under P if and only if

$$b_s^X = 0 \text{ for all } s \in [0,T]. \tag{1.14}$$

We say that X is a true *martingale* if and only if

$$X_s = \mathrm{E}_s[X_u] \text{ for all } s \leq u \leq T. \tag{1.15}$$

Every true martingale has zero drift but not every process with zero drift is a true martingale. There are two important results which characterize true martingales:

Proposition 1.13 (first martingale proposition)**.** *An Itô process X with zero drift is a true martingale provided its volatility satisfies the Novikov condition*

$$\mathrm{E}[\exp(\frac{1}{2}\int_0^T c_s^X ds)] < \infty.$$

Proposition 1.14 (second martingale proposition)**.** *Consider a random variable Y with* $\mathrm{E}[|Y|] < \infty$*. Process X defined by*

$$X_s = \mathrm{E}_s[Y] \textit{ for all } s \leq T, \tag{1.16}$$

is a martingale.

Zero drift is a necessary but *not* a sufficient condition for a process to be a martingale, but one can put technical conditions on c^X which guarantee that (1.14) automatically implies (1.15). In this course we assume that such conditions are always satisfied.

Proposition 1.15. *A local martingale bounded from below is a supermartingale, i.e. it is a process decreasing in expectation*

$$X_s \geq E_s[X_t] \text{ for } s \leq t.$$

Conversely, a local martingale bounded from above is a submartingale, i.e. a process increasing in expectation.

$X_s \leq \mathbb{E}_s[X_t] \ \forall \ s \leq t$

1.11 Change of measure

A probability measure $P^\star$ on $(\Omega, \mathscr{F})$ is absolutely continuous with respect to P,

$$P^\star \ll P,$$

iff there is a non-negative random variable on $(\Omega, \mathscr{F})$, which we denote by $\frac{dP^\star}{dP}$ and call the *change of measure*, such that

$$\mathrm{E}^{P^\star}[X] = \mathrm{E}\left[\frac{dP^\star}{dP}X\right], \tag{1.17}$$

for all non-negative random variables X. Property (1.17) with $X = 1$ yields

$$\mathrm{E}\left[\frac{dP^\star}{dP}\right] = 1,$$

which simply means that the new probabilities $P^\star$ add up to 1 across all possible outcomes.

We have $P^\star(\frac{dP^\star}{dP} > 0) = 1$, however $P(\frac{dP^\star}{dP} > 0) \leq 1$ (intuitively some paths with positive P probability may have zero $P^\star$ probability). If also $P(\frac{dP^\star}{dP} > 0) = 1$ then we say that $P^\star$ is equivalent to P and write $P^\star \sim P$. This means that both $dP^\star/dP$ and $dP/dP^\star$ are well-defined and in fact $dP^\star/dP = 1/(dP/dP^\star)$.

1.11.1 Density process

We say that Z is a *density process* associated with the change of measure $dP^\star/dP$ if there is a constant c such that for all t

$$Z_t = c\mathrm{E}_t[dP^\star/dP]. \tag{1.18}$$

Thus density process Z is determined by $dP^\star/dP$ uniquely up to a multiplicative constant. In the literature one often requires a further normalization $Z_0 = 1$ which corresponds to $c = 1$ in equation (1.18). In practical calculations it is however convenient to allow other values of c, see Section 1.11.2 for an example. This convention will later allow us to interpret Z_t itself as a price of some security.

Intuitively, Z_t/Z_0 provides the ratio of $P_t^\star$ over P_t, where P_t is the P–probability of going along a particular path from 0 to t. This is a key observation from which one can deduce two further important results, assuming $P^\star$ is equivalent to P:

1. The ratio of P over $P^\star$ is given by Z_0/Z_t. Therefore Z^{-1} must be the density process of $dP/dP^\star$. Prove it rigorously, i.e., show that

$$Z_t^{-1} = \mathrm{E}^{P^\star}\left[Z_T^{-1}\right].$$

2. The ratio Z_T/Z_0 is the change of measure between 0 and T; likewise, Z_t/Z_0 is the change of measure between 0 and t. The probability of a given path on the interval $[0,T]$ is the probability of the $[0,t]$–segment *multiplied* the conditional probability of the $(t,T]$–segment conditional on the $[0,t]$–segment. Therefore, the conditional change of measure between t and T is given by $(Z_T/Z_0)/(Z_t/Z_0) = Z_T/Z_t$, so that we obtain

$$\mathrm{E}_t^{P^\star}[X] = \mathrm{E}_t\left[X\frac{Z_T}{Z_t}\right], \tag{1.19}$$

whenever one of the two expressions is well defined.

Equation (1.19) yields the following result, which, in turn, is the key to the Girsanov theorem.

Proposition 1.16 (third martingale proposition). *Suppose $P^\star \sim P$ with density process $Z_t = c\mathrm{E}_t[dP^\star/dP]$ and Y is a P–semimartingale. Then the following statements hold*

1. *Y is a $P^\star$–local martingale $\iff$ ZY is a P–local martingale,*
2. *Y is a $P^\star$–martingale $\iff$ ZY is a P–martingale.*

For further material on the density process see for example Černý (2009, Chapter 9).

1.11.2 Typical example of measure change and its density process

In the Black–Scholes model one typically encounters a change of measure of the form

$$\frac{dP^\star}{dP} = \frac{S_T^\eta}{\mathrm{E}[S_T^\eta]}, \qquad \eta \in \mathbb{R}.$$

We can therefore take density process in the form

$$Z_t = \mathrm{E}_t[S_T^\eta].$$

Most often P will be interpreted as the money-market risk-neutral measure. Using the fact that $\ln S$ is an Itô PII with constant drift and volatility,

$$b^{\ln S} = r - \hat{\delta} - \frac{\sigma^2}{2}, \qquad c^{\ln S} = \sigma^2,$$

Proposition 1.9 yields

$$\ln S_T|\mathscr{F}_t \sim N(\ln S_t + b^{\ln S}(T-t), c^{\ln S}(T-t)),$$

and therefore

$$\eta \ln S_T|\mathscr{F}_t \sim N(\eta \ln S_t + \eta b^{\ln S}(T-t), \eta^2 c^{\ln S}(T-t)).$$

From the moment generating function of a normal distribution we finally obtain

$$\begin{aligned} Z_t &= \mathrm{E}_t[S_T^\eta] = \mathrm{E}_t[e^{\eta \ln S_T}] = \mathrm{e}^{\eta \ln S_t + \eta b^{\ln S}(T-t) + \frac{1}{2}\eta^2 c^{\ln S}(T-t)} \\ &= S_t^\eta \exp\left(\eta b^{\ln S}(T-t) + \frac{1}{2}\eta^2 c^{\ln S}(T-t)\right). \end{aligned}$$

1.11.3 Interpretation of measure change in a pricing formula

Let us take P to be the T–forward risk-neutral measure. Let X_T be the pay-off of a contingent claim (for example a call option). The martingale X defined by

$$X_t = \mathrm{E}_t[X_T] \tag{1.20}$$

is then the *forward price* of X_T (there is no discounting in (1.20)). Consider a new measure $P^\star$ with a density process Z, $dP^*/dP = Z_T/Z_0$. By construction, Z_t can be interpreted as the forward price of Z_T. Hence Z can be understood as another derivative contract and the change of measure $dP^*/dP = Z_T/Z_0$ is simply the total return on the buy-and-hold position in this new derivative.

Recall that Z^{-1} is the density of P relative to $P^\star$. Hence we can write

$$X_t = E_t[X_T] = E_t^{P^\star}\left[X_T \frac{Z_T^{-1}}{Z_t^{-1}}\right]$$

and after rearranging

$$\frac{X_t}{Z_t} = E_t^{P^\star}\left[\frac{X_T}{Z_T}\right].$$

This means that $P^\star$ can again be seen as a pricing measure, but the pay-offs and prices have to be expressed in the units of Z. In this context we say that Z is a new *numéraire* (unit of account).

1.11.4 More general construction of the density process

Consider an $\mathbb{R}^d$-valued Itô process X with martingale part $M^X = X - B^X$. Consider an $\mathbb{R}^d$-valued predictable *process* η (no longer a constant) and define a P-local martingale Z

$$\frac{dZ_t}{Z_t} = \eta_t(dX_t - b_t^X dt), \tag{1.21}$$

$$Z_0 > 0. \tag{1.22}$$

Z is our candidate for the density process.[2] On application of the Itô formula with $g(Z) = \ln Z$ one obtains

$$d \ln Z_t = \eta(dX_t - b_t^X dt) - \frac{1}{2}\eta c_t^X \eta^\top dt.$$

On writing this expression in integral form (cf. equation 1.3) we have

$$\ln Z_t = \ln Z_0 + \int_0^t \eta_s dX_s - \int_0^t (\eta_s b_s^X + \frac{1}{2}\eta_s c_s^X \eta_s^\top) ds$$

$$Z_t = Z_0 e^{\int_0^t \eta_s dX_s - \int_0^t (\eta_s b_s^X + \frac{1}{2}\eta_s c_s^X \eta_s^\top) ds}. \tag{1.23}$$

This shows $Z > 0$ as long as $Z_0 > 0$. Finally, for Z to be a density process it must be a true martingale under P, i.e., $Z_t = \mathrm{E}_t[Z_T]$ for all t. However, in our case it is enough to show $\mathrm{E}[Z_T/Z_0] = 1$. This happens because Z is a local martingale (see equation 1.21) bounded below (by 0) and it is therefore a supermartingale. In turn, a supermartingale Z is a true martingale if and only if $\mathrm{E}[Z_T] = Z_0$.

[2] N.B. For Z to be a density process it must be a P-martingale and therefore also a P–local martingale. Therefore the right hand side of (1.21) must have zero P-drift.

1.12 Girsanov theorem for Itô processes

Theorem 1.17. *Suppose an Itô process Z is a density process associated with the change of measure $dP^\star/dP$. For any process X such that (X, Z) is jointly an Itô process one has*

$$b^X_{P^\star}(t) = b^X(t) + \frac{dX_t \frac{dZ_t}{Z_t}}{dt} = b^X(t) + \frac{dX_t d\ln Z_t}{dt},$$

that is by going from measure P to measure $P^\star$ the drift of Y increases by the rate of covariation between dX and dZ/Z (equivalently, between dX and $d\ln Z$).

Specifically if Z satisfies

$$\frac{dZ_t}{Z_t} = \eta_t(dX_t - b^X_t dt), \tag{1.24}$$

which is equivalent to

$$Z_t = Z_0 e^{\int_0^t \eta_s dX_s - \int_0^t (\eta_s b^X_s + \frac{1}{2}\eta_s c^X_s \eta_s^\top) ds},$$

then

$$b^X_{P^\star} = b^X + c^X \eta^\top.$$

Remark 1.18. *Girsanov theorem is easily derived from the Itô formula as a consequence of Proposition 1.16. Let $B^X_{P^\star}$ be the drift component of X under $P^\star$. We know that $B^X_{P^\star}$ is unique. By Proposition 1.16 $X - B^X_{P^\star}$ is a $P^\star$-local martingale if and only if $Z(X - B^X_{P^\star})$ is a P–local martingale. From the Itô formula*

$$\begin{aligned} d(Z_t(X_t - B^X_{P^\star}(t))) =& Z_t(dX_t - dB^X_{P^\star}(t)) + (X_t - B^X_{P^\star}(t))dZ_t + dZ_t dX_t \\ =& \underbrace{Z_t(dB^X(t) - dB^X_{P^\star}(t) + \frac{dZ_t}{Z_t}dX_t)}_{\text{predictable FV process}} + \underbrace{Z_t dM^X_t + (X_t - B^X_{P^\star}(t))dZ_t}_{P\text{–local martingale}}. \end{aligned}$$

The right-hand side is a P–local martingale (it has no drift under P) iff

$$dB^X(t) - dB^X_{P^\star}(t) + \frac{dZ_t}{Z_t}dX_t = 0,$$

which yields

$$dB^X_{P^\star}(t) = dB^X(t) + \frac{dZ_t}{Z_t}dX \tag{1.25}$$

Since $\frac{dZ_t}{Z_t} = d\ln Z_t +$ continuous FV process we obtain that dZ_t/Z_t can be replaced by $\ln Z_t$ in (1.25). Finally, in the special case (1.24), we have $dX_t dZ_t/Z_t = c^X\eta^\top dt$ and this concludes the proof.

1.13 Stochastic exponential and stochastic logarithm

In Finance one often encounters processes with increments of the form

$$\frac{dY_t}{Y_{t-}}.$$

This process has a name, it is called the *stochastic logarithm* of Y and it is denoted by $\mathcal{L}(Y)$

$$\begin{aligned} d\mathcal{L}(Y)_t =& \frac{dY_t}{Y_{t-}}, \\ \mathcal{L}(Y)_0 =& 0. \end{aligned}$$

The process is well defined if Y does not go to zero continuously.

Conversely, suppose we are given the SDE

$$\frac{dY_t}{Y_{t-}} = dX_t, \tag{1.26}$$

and we wish to express Y in terms of X. If $Y_0 = 1$ then the solution of (1.26) is called the *stochastic exponential* of X and it is denoted by $\mathscr{E}(X)$. Thus the stochastic exponential satisfies

$$\begin{aligned} d\mathscr{E}(X)_t =& \mathscr{E}(X)_{t-} dX_t, \\ \mathscr{E}(X)_0 =& 1. \end{aligned}$$

If you have ever used the change of measure in the Girsanov theorem, or worked with the Black–Scholes model, then you have come across both the stochastic logarithm and stochastic exponential, perhaps without realizing it.

The usefulness of stochastic logarithm and stochastic exponential stems from providing a very compact notation. Instead of saying “the process with increment dY/Y” we simply say $\mathcal{L}(Y)$. Instead of “capital gain rate process corresponding to the price process S” we say $\mathcal{L}(S)$. Since most of finance (not only in continuous time) is based on the interplay between the capital gain rate process $\mathcal{L}(S)$ and the log return process $\ln S$ it is very useful to have a simple notation for the former.

The stochastic exponential is similarly ubiquitous in finance. Instead of “fund value F generated by the cumulative rate of return process X” we can say $F_0\mathscr{E}(X)$.

1.14 Derivation of the Black–Scholes formula

Consider a stock without dividends and denote its price by S. Assume that the risk-free rate r and the dividend yield $\hat{\delta}$ are constant and that the capital gains rate process $\mathcal{L}(S)$ has constant volatility, $\sqrt{c^{\mathcal{L}(S)}} = \sigma$ (recall that $d\mathcal{L}(S) = \frac{dS}{S}$).

Denote the money market risk–neutral measure by P. The option pricing theory tells us that in the absence of arbitrage the expected rate of return equals the risk-free rate under P. That means

$$b^{\mathcal{L}(S)} = r - \hat{\delta}, \qquad c^{\mathcal{L}(S)} = \sigma^2.$$

Consequently $\mathcal{L}(S)$ is an Itô PII under P.

1.14.1 Distribution of log returns

Consider a function f and a derivative with payoff $f(S_T)$. In our case

$$f(S_T) = (S_T - K)^+$$

corresponds to a European call option. The no-abitrage theory shows that the price of the derivative equals

$$C_0 = e^{-rT}\mathrm{E}\left[f(S_T)\right].$$

All we need to do is to work out the distribution of S_T. To this end, we apply the Itô formula to $\ln S$

$$d\ln S_t = \frac{dS_t}{S_t} - \frac{1}{2}\left(\frac{dS_t}{S_t}\right)^2 = d\mathcal{L}(S) - \frac{1}{2}c^{\mathcal{L}(S)}dt. \qquad (1.27)$$

This also implies

$$c^{\ln S} = c^{\mathcal{L}(S)},$$

meaning that capital gain rate and log return have the same volatility. Since $\mathcal{L}(S)$ is an Itô PII under P we conclude from (1.27) that $\ln S$ is also an Itô PII under P with

$$b^{\ln S} = b^{\mathcal{L}(S)} - c^{\mathcal{L}(S)}/2 = r - \hat{\delta} - \sigma^2/2,$$

and we can apply Proposition 1.9 to conclude

$$\ln S_T \overset{P}{\sim} N(\ln S_0 + (r - \hat{\delta} - \sigma^2/2)T, \sigma^2 T).$$

1.14.2 Pricing formula

We start by computing the expectation of a truncated lognormal variable. This is a standard result used ubiquitously in option pricing.

Lemma 1.19. *Let $Y \sim N(m, s^2)$ and $y_1 \leq y_2$. Then*

$$\begin{aligned}
\mathrm{E}\left[\mathrm{e}^{\alpha Y}\mathbf{1}_{y_1<Y<y_2}\right] =& \mathrm{e}^{\alpha m+\frac{1}{2}\alpha^2 s^2}\left(\Phi\left(\frac{y_2-m-\alpha s^2}{s}\right)-\Phi\left(\frac{y_1-m-\alpha s^2}{s}\right)\right)\\
=& \mathrm{e}^{\alpha m+\frac{1}{2}\alpha^2 s^2}\left(\Phi\left(\frac{m+\alpha s^2-y_1}{s}\right)-\Phi\left(\frac{m+\alpha s^2-y_2}{s}\right)\right).
\end{aligned}$$

In case $y_1 = -\infty$ or $y_2 = \infty$ use $\Phi(-\infty) = 0, \Phi(\infty) = 1$.

Proof. See for example Černý, 2009, Example B.25. □

It is now a simple matter to evaluate the option price by writing

$$\begin{aligned}
e^{rT}C_0 &= \mathrm{E}\left[(S_T-K)^+\right] = \mathrm{E}\left[(S_T-K)\mathbf{1}_{S_T>K}\right]\\
&= \mathrm{E}\left[(e^{\ln S_T}-K)\mathbf{1}_{\ln S_T>\ln K}\right]\\
&= \mathrm{E}\left[e^{\ln S_T}\mathbf{1}_{\ln S_T>\ln K}\right]-K\mathrm{E}\left[\mathbf{1}_{\ln S_T>\ln K}\right],
\end{aligned}$$

and employing Lemma 1.19 with $y_1 = \ln K, y_2 = \infty$ and $Y = \ln S_T$, implying

$$\begin{aligned}
m &= \ln S_0 + (b^{\mathcal{L}(S)} - c^{\mathcal{L}(S)}/2)T,\\
s^2 &= c^{\mathcal{L}(S)}T.
\end{aligned}$$

We set $\alpha = 1$ and 0 for the first and the second expectation, respectively. This yields

$$\begin{aligned}
C_0 =& S_0 e^{(b^{\mathcal{L}(S)}-r)T}\Phi\left(\frac{\ln\frac{S_0}{K}+(b^{\mathcal{L}(S)}+c^{\mathcal{L}(S)}/2)T}{\sqrt{c^{\mathcal{L}(S)}T}}\right)\\
&-Ke^{-rT}\Phi\left(\frac{\ln\frac{S_0}{K}+(b^{\mathcal{L}(S)}-c^{\mathcal{L}(S)}/2)T}{\sqrt{c^{\mathcal{L}(S)}T}}\right). \qquad (1.28)
\end{aligned}$$

1.14.3 Black–Scholes model with constant dividend yield

Recall that S denotes the ex-dividend share price and $\hat{\delta}$ is a constant dividend yield, meaning that the total dividend payment over time interval $[0, t]$ is $\int_0^t \hat{\delta} S_u du$. Financial theory tells us that the rate of return on the asset has to equal the risk-free rate under the money market measure P. Now the total rate of return arises from two sources -

capital gains rate dS_t/S_t and dividend yield $\hat{\delta}dt$

$$\frac{dS_t}{S_t} + \hat{\delta}dt = d\mathcal{L}(S)_t + \hat{\delta}dt.$$

Under the money market risk-neutral measure the expected rate of return is therefore

$$r = b^{\mathcal{L}(S)} + \hat{\delta},$$

implying we must have

$$b^{\mathcal{L}(S)} = r - \hat{\delta}.$$

1.15 Construction of PDEs in Finance

Partial differential equations used in the pricing of derivatives express a simple fact, namely that a certain process related to the price of the derivative, expressed as a function of some state variables is a martingale under a suitably chosen probability measure. One can vary the three main ingedients by changing i) the measure, ii) the target function and iii) the state variables, but the main principle is always the same — some quantity is a martingale.

Let us first examine the most natural setting dictated by financial rather than computational considerations. Here the target function will be the option price, the state variables will consist of time and stock price and the relevant measure will be the risk-neutral measure.

1.15.1 Black–Scholes PDE

As discussed above the input parameters governing the rate of return process are

$$b^{\mathcal{L}(S)} = r - \hat{\delta}, c^{\mathcal{L}(S)} = \sigma^2.$$

From the financial theory the price of the derivative equals

$$C_t = e^{-r(T-t)}\mathrm{E}_t\left[f(S_T)\right].$$

We have $dS = Sd\mathcal{L}(S)$ and consequently

$$b^S = Sb^{\mathcal{L}(S)}, c^S = S^2 c^{\mathcal{L}(S)}.$$

By Proposition 1.7 S is a Markov process under the risk-neutral measure P and there is a function g such that

$$C_t = g(t, S_t).$$

Furthermore the process $C_t e^{-rt} = \mathrm{E}_t\left[f(S_T)e^{-rT}\right]$ is a martingale by Proposition 1.14 and therefore it must have zero drift. From the Itô formula

$$\begin{aligned} d\left(C_t \mathrm{e}^{-rt}\right) &= \mathrm{e}^{-rt}(dC_t - rC_t dt) \\ dC_t &= dg(t,S_t) = \partial_1 g(t,S_t)dt + \partial_2 g(t,S_t)dS_t + \frac{1}{2}\partial_{22}g(t,S_t)c_t^S dt \end{aligned}$$

and the zero drift condition for Ce^{-rt} yields the Black–Scholes PDE

$$\begin{aligned} 0 &= \partial_1 g(t,S) + \partial_2 g(t,S)b^{\mathcal{L}(S)}S + \frac{1}{2}\partial_{22}g(t,S)S^2 c^{\mathcal{L}(S)} - rg(t,S). \\ g(T,S) &= (S-K)^+ \end{aligned}$$

1.15.2 Change of variables and change of target function

The previous derivation suggests a way to simplify the martingale PDE. Firstly we will write the pay-off of the derivative asset as a function of $\ln S_T$. For a call option with strike K this means taking $f(x) = (\mathrm{e}^x - K)^+$. We take as the target function the forward option price $Y_t = \mathrm{E}_t[f(\ln S_T)]$. Since $\ln S$ is a Markov process under P (with $b^{\ln S} = r - \hat{\delta} - \sigma^2/2$, $c^{\ln S} = \sigma^2$) we have that $Y_t = g(t, \ln S_t)$ for some function g. Moreover, by virtue of (1.16) Y is a P–martingale, and therefore $b^Y = 0$. If g is sufficiently smooth the Itô formula yields

$$\begin{aligned} 0 = b^Y &= \partial_1 g(t,x) + \partial_2 g(t,x)b^{\ln S} + \frac{1}{2}\partial_{22}g(t,x)c^{\ln S} \\ &= \partial_1 g(t,x) + \partial_2 g(t,x)(r - \hat{\delta} - \sigma^2/2) + \frac{1}{2}\partial_{22}g(t,x)\sigma^2, \end{aligned} \tag{1.29}$$

and the boundary condition reads

$$g(T,x) = (\mathrm{e}^x - K)^+ . \tag{1.30}$$

Once we solve for g the price of the derivative is given by $C_t = \mathrm{e}^{r(t-T)}g(t, \ln S_t)$.

1.15.3 Towards the heat equation

One can get rid of the drift term $\partial_2 g(t,x)b^{\ln S}$ in the previous PDE by using a state variable that is itself a martingale. One way to do this is to consider $X_t = \ln S_t + b^{\ln S}(T-t)$ under the money market risk-neutral measure P. Such a transformation works well for path-independent derivatives. For a call option with strike K this means solving

$$0 = \partial_1 g(t,x) + \frac{1}{2}\partial_{22}g(t,x)c^{\mathcal{L}(S)},$$

$$g(T,x) = (e^x - K)^+.$$

The option price at t is then given by

$$C_t = \mathrm{e}^{r(t-T)} g(t, \ln S_t + b^{\ln S}(T-t)).$$

1.15.4 Obtaining the heat equation by change of measure

Another useful way of transforming the pricing PDE is to keep $X = \ln S$, $f(x) = (\mathrm{e}^x - K)^+$ and consider a new measure $P^\star$ of the form

$$Z_t = \mathrm{e}^{\eta X_t + (\eta b^X + \frac{1}{2}\eta^2 c^X)(T-t)}, \tag{1.31}$$

$$\frac{dP^\star}{dP} = \frac{Z_T}{Z_0}, \tag{1.32}$$

for a constant η, as discussed in Section 1.11.2. Since we take P to be the money market risk-neutral measure, we have $b^X = b^{\ln S} = r - \hat{\delta} - \sigma^2/2$. In the Black–Scholes model $c^X = c^{\ln S} = \sigma^2$.

We know from Section 1.11.3 that $\mathrm{E}_t^{P^\star}[f(X_T)/Z_T]$ gives the forward price of $f(X_T)$ expressed in terms of the numeraire Z_t. The Girsanov theorem yields

$$b^X_{P^\star} = b^X + \eta c^X = r - \hat{\delta} - \frac{\sigma^2}{2} + \eta\sigma^2.$$

Consider the $P^\star$–martingale $\mathrm{E}_t^{P^\star}\left[f(X_T)e^{-\eta X_T}\right]$. Since X is Markov under $P^\star$ we conclude that there must be a function $g(t,x)$ such that

$$g(t, X_t) = \mathrm{E}_t^{P^\star}\left[f(X_T)e^{-\eta X_T}\right]. \tag{1.33}$$

Assuming that g is sufficiently differentiable, the Itô formula implies that g must satisfy the PDE

$$0 = \partial_1 g(t,x) + \partial_2 g(t,x) \underbrace{\left(r - \frac{\sigma^2}{2} + \eta\sigma^2\right)}_{b^X_{P^\star}} + \frac{1}{2}\partial_{22} g(t,x) \underbrace{\sigma^2}_{c^X}, \tag{1.34}$$

$$g(T,x) = \mathrm{e}^{-\eta x} f(x). \tag{1.35}$$

Recall that g gives the price of the derivative in terms of the numeraire Z. Once we solve for g, the spot price of the option at time t is given by

$$C_t = \mathrm{e}^{r(t-T)} Z_t g(t, X_t) \tag{1.36}$$

$$= \mathrm{e}^{(\eta(r-\sigma^2/2) + \frac{1}{2}\eta^2\sigma^2 - r)(T-t)} S_t^\eta g(t, \ln S_t). \tag{1.37}$$

We observe that the numeraire in this case is a fund whose value at time T equals S_T^η, a power contract. Two values of η are worth pointing out. Firstly,

$$\eta = -b^{\ln S}/c^{\ln S} = 1/2 + (\hat{\delta} - r)/\sigma^2$$

will give $b_{P^\star}^{\ln S} = 0$. Secondly, $\eta = 1$ will turn the payoff in the numeraire units, $f(X_T)\mathrm{e}^{-\eta X_T} = \left(1 - K\mathrm{e}^{-X_T}\right)^+$, into a function bounded between 0 and 1. The value $\eta = 1$ corresponds to the T–forward stock risk-neutral measure because the numeraire is one share sold forward to be delivered at T.

Remark 1.20. *Wilmott, Howison, and Dewynne (1995) use $\eta = -b^{\ln S}/c^{\ln S}$ and assume zero dividend yield, $\hat{\delta} = 0$, which leads to $\eta = 1/2 - r/\sigma^2$ leading to $b_{P^\star}^{\ln S} = 0$. They use slightly different state variables: $X = \ln S - \ln K$ instead of $\ln S$ and so-called "operational time to maturity" $\tau = \frac{\sigma^2}{2}(T-t)$ instead of t. The target function is $u(\tau, x) = g(t,x)/K$. On account of $\frac{\partial}{\partial t} = \frac{d\tau}{dt}\frac{\partial}{\partial \tau}$ equations (1.34, 1.35) yield*

$$\begin{aligned} 0 &= -\frac{\sigma^2}{2}\partial_1 u(\tau,x) + \frac{\sigma^2}{2}\partial_{22}u(\tau,x), \\ u(0,x) &= \mathrm{e}^{-\eta x}\left(e^x - 1\right)^+, \\ C_t &= K\mathrm{e}^{-\eta \ln K}\mathrm{e}^{\eta \ln S_t + \left(\eta(r-\sigma^2/2) - r + \frac{1}{2}\eta^2\sigma^2\right)(T-t)}u\left(\frac{\sigma^2}{2}(T-t), \ln\frac{S_t}{K}\right). \end{aligned} \tag{1.38}$$

On rearranging the PDE turns into a heat equation

$$\partial_1 u(\tau,x) = \partial_{22}u(\tau,x).$$

References

Černý, A. (2009). *Mathematical Techniques in Finance: Tools for Incomplete Markets*. 2nd. Princeton University Press, Princeton, NJ.

Černý, A. and J. Ruf (2019). "Finance without Brownian motions: An introduction to simplified stochastic calculus". Available from ssrn.com/abstract=3500384.

Dellacherie, C. and P.-A. Meyer (1978). *Probabilities and potential*. Vol. 29. North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York, pp. viii+189. ISBN: 0-7204-0701-X.

Grigoriu, M. (2002). *Stochastic Calculus: Applications in Science and Engineering*. Springer Science + Business Media, New York.

Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes*. 2nd. Vol. 288. Comprehensive Studies in Mathematics. Springer-Verlag, Berlin, pp. xx+661. ISBN: 3-540-43932-3.

Kunita, H. and S. Watanabe (1967). "On square integrable martingales". In: *Nagoya Math. J.* 30, pp. 209–245. ISSN: 0027-7630.

Larsson, M. and J. Ruf (2014). "Convergence of Local Supermartingales and Novikov-Kazamaki Type Conditions for Processes With Jumps". Available from `arxiv.org/abs/1411.6229`.

McKean Jr., H. P. (1969). *Stochastic Integrals*. Probability and Mathematical Statistics, No. 5. Academic Press, New York-London, pp. xiii+140.

Meyer, P.-A. (1967). "Intégrales stochastiques. I, II, III, IV". In: *Séminaire de Probabilités I, Strasbourg*. Lecture Notes in Math. Springer, Berlin, pp. 72–94, 95–117, 118–141, 142–162.

— (1976). "Un cours sur les intégrales stochastiques". In: *Séminaire de Probabilités X, Strasbourg*. Vol. 511. Lecture Notes in Math. Springer, Berlin, pp. 245–400.

Wilmott, P., S. Howison, and J. Dewynne (1995). *The Mathematics of Financial Derivatives*. Cambridge University Press.

Chapter 2

Principles of asset pricing

2.1 Brute force derivation of the Black–Scholes formula

2.1.1 Black–Scholes model with constant dividend yield

Consider a stock with dividends, denoting the ex-dividend share price by S. Assume that the risk-free rate r and the dividend yield $\hat{\delta}$ are constant. The dividends are payed continuously to give the total dividend payment over time interval $[0, t]$ of $\int_0^t \hat{\delta} S_u du$. We further assume, in line with the Black–Scholes model, that the capital gains rate process $\mathcal{L}(S)$ has constant volatility σ, meaning $c^{\mathcal{L}(S)} = \sigma^2$ (recall that $d\mathcal{L}(S)_t = \frac{dS_t}{S_t}$).

The total rate of return arises from two sources - capital gains rate dS_t/S_t and dividend yield $\hat{\delta}dt$

$$\frac{dS_t}{S_t} + \hat{\delta}dt = d\mathcal{L}(S)_t + \hat{\delta}dt.$$

Under the money market risk-neutral measure the expected rate of return on the stock investment must equal risk-free return,

$$r = b^{\mathcal{L}(S)} + \hat{\delta}.$$

Denote the money market risk–neutral measure by P. We have

$$b^{\mathcal{L}(S)} = r - \hat{\delta}, \qquad c^{\mathcal{L}(S)} = \sigma^2.$$

Consequently $\mathcal{L}(S)$ is an Itô PII under P. The Itô formula

$$d \ln S_t = \frac{dS_t}{S_t} - \frac{1}{2}\left(\frac{dS_t}{S_t}\right)^2$$

then yields

$$b^{\ln S} = r - \hat{\delta} - \frac{1}{2}\sigma^2, \qquad c^{\ln S} = \sigma^2.$$

Therefore $\ln S$, too, is an Itô PII under P.

2.1.2 Distribution of log returns

Consider a function f and a derivative with payoff $f(S_T)$. In our case

$$f(S_T) = (S_T - K)^+$$

corresponds to a European call option. The no-abitrage theory shows that the price of the derivative equals

$$C_0 = e^{-rT}\mathrm{E}\left[f(S_T)\right].$$

All we need to do is to work out the distribution of S_T. To this end, we apply the Itô formula to $\ln S$

$$d\ln S_t = \frac{dS_t}{S_t} - \frac{1}{2}\left(\frac{dS_t}{S_t}\right)^2 = d\mathcal{L}(S) - \frac{1}{2}c^{\mathcal{L}(S)}dt. \tag{2.1}$$

This also implies

$$c^{\ln S} = c^{\mathcal{L}(S)},$$

meaning that capital gain rate and log return have the same volatility. Since $\mathcal{L}(S)$ is an Itô PII under P we conclude from (2.1) that $\ln S$ is also an Itô PII under P with

$$b^{\ln S} = b^{\mathcal{L}(S)} - \frac{1}{2}c^{\mathcal{L}(S)} = r - \hat{\delta} - \frac{1}{2}\sigma^2,$$

and we can apply Proposition 1.9 to conclude

$$\ln S_T \overset{P}{\sim} N(\ln S_0 + (r - \hat{\delta} - \sigma^2/2)T, \sigma^2 T).$$

2.1.3 Pricing formula

We start by computing the expectation of a truncated lognormal variable. This is a standard result used ubiquitously in option pricing.

Lemma 2.1. *Let $Y \sim N(m, s^2)$ and $y_1 \leq y_2$. Then*

$$\begin{aligned}\mathrm{E}\left[\mathrm{e}^{\alpha Y}\mathbf{1}_{y_1<Y<y_2}\right] &= \mathrm{e}^{\alpha m + \frac{1}{2}\alpha^2 s^2}\left(\Phi\left(\frac{y_2 - m - \alpha s^2}{s}\right) - \Phi\left(\frac{y_1 - m - \alpha s^2}{s}\right)\right)\\ &= \mathrm{e}^{\alpha m + \frac{1}{2}\alpha^2 s^2}\left(\Phi\left(\frac{m + \alpha s^2 - y_1}{s}\right) - \Phi\left(\frac{m + \alpha s^2 - y_2}{s}\right)\right).\end{aligned}$$

In case $y_1 = -\infty$ or $y_2 = \infty$ use $\Phi(-\infty) = 0, \Phi(\infty) = 1$.

Proof. See for example Černý (2009), Example B.25. □

It is now a simple matter to evaluate the option price by writing

$$e^{rT}C_0 = \mathrm{E}\left[(S_T - K)^+\right] = \mathrm{E}\left[(S_T - K)\mathbf{1}_{S_T>K}\right]$$

$$
\begin{aligned}
&= \mathrm{E}\left[\left(e^{\ln S_T} - K\right)\mathbf{1}_{\ln S_T > \ln K}\right] \\
&= \mathrm{E}\left[e^{\ln S_T}\mathbf{1}_{\ln S_T > \ln K}\right] - K\mathrm{E}\left[\mathbf{1}_{\ln S_T > \ln K}\right],
\end{aligned}
$$

and employing Lemma 2.1 with $y_1 = \ln K, y_2 = \infty$, and $Y = \ln S_T$. This means we must take

$$
\begin{aligned}
m &= \ln S_0 + b^{\ln S} T, \\
s^2 &= c^{\ln S} T.
\end{aligned}
$$

We set $\alpha = 1$ for the first and $\alpha = 0$ for the second expectation, respectively. This yields

$$
\begin{aligned}
C_0 = S_0 e^{(b^{\ln S} + c^{\ln S}/2 - r)T} \Phi\left(\frac{\ln \frac{S_0}{K} + (b^{\ln S} + c^{\ln S}/2)T}{\sqrt{c^{\ln S} T}}\right) \\
- K e^{-rT} \Phi\left(\frac{\ln \frac{S_0}{K} + b^{\ln S} T}{\sqrt{c^{\ln S} T}}\right).
\end{aligned} \tag{2.2}
$$

2.2 Wealth going forward in time

2.2.1 Stock fund

Consider a stock with a continuous ex-dividend price S and cumulative dividend $\int_0^t \hat{\delta} S_u du$. Let us now build an investment fund with initial value F_0 that invests all its assets in this particular stock. We have

$$dF_t = \frac{F_t}{S_t}(dS_t + \hat{\delta} S_t dt).$$

Use the Itô formula to pass from dF_t to $d \ln F_t$,

$$d \ln F_t = d \ln S_t + \hat{\delta} dt,$$

then integrate and exponentiate to obtain

$$F_t = F_0 \frac{S_t}{S_0} \mathrm{e}^{\hat{\delta} t}.$$

If we start with one share, $F_0 = S_0$ then our stock fund will have the value

$$F_t = S_t \mathrm{e}^{\hat{\delta} t}.$$

Observe that

- one share at time 0 is worth more than one share at time t because of the dividend stream;
- the spot value of *one share delivered at time* T is precisely

$$S_T \frac{F_t}{F_T} = S_t \mathrm{e}^{-\hat{\delta}(T-t)}.$$

- we have made no assumptions about the volatility of S;
- we have not assumed existence of a risk-free money market account either.

2.2.2 Money market fund

Imagine now a stock with a constant ex-dividend value $S = 1$ and stochastic dividend yield r_t. The same derivation as above gives

$$F_t = F_0 \mathrm{e}^{\int_0^t r_u du}.$$

Now F represents the money market fund, which we can think of as the balance of a bank account where the interest is continuously compounded and deposited back into the account continuously. If the interest rates are deterministic then the spot value of \$1 delivered at time T is

$$\frac{F_t}{F_T} = \mathrm{e}^{-\int_t^T r_u du} = P(t, T),$$

i.e., the same as the price of time-T zero coupon discount bond.

2.2.3 Fixed proportions investment fund

Let us now condider a fund investing fixed proportion α in the stock and the rest in the money market. The value of such fund must satisfy

$$dF_t = F_t \left((1-\alpha) r_t dt + \alpha \left(\frac{dS_t}{S_t} + \hat{\delta} dt \right) \right).$$

Once again, pass to $\ln F_t$ and $\ln S_t$

$$d \ln F_t = (1-\alpha) r_t dt + \alpha \hat{\delta} dt + \alpha d \ln S_t + \frac{1}{2}(\alpha - \alpha^2) dC_t^{\ln S}.$$

Now integrate and exponentiate

$$F_t = F_0 \frac{S_t^\alpha}{S_0^\alpha} e^{(1-\alpha)\int_0^t r_u du + \alpha\hat{\delta}t + \frac{1}{2}(\alpha-\alpha^2)C_t^{\ln S}}. \tag{2.3}$$

As long as r and $C^{\ln S}$ are deterministic, we are able to determine the spot price of one *α–power contract* delivered at time T, with payoff S_T^α. This is done by choosing

$$F_0 = \frac{S_0^\alpha}{e^{(1-\alpha)\int_0^T r_u du + \alpha\hat{\delta}T + \frac{1}{2}(\alpha-\alpha^2)C_T^{\ln S}}} \tag{2.4}$$

in (2.3).

- With $\alpha = 0$ we get back to the zero-coupon discount bond.
- With $\alpha = 1$ we obtain the stock fund that delivers one share at time T.

In these two cases no assumption on $C^{\ln S}$ is required.

2.3 Wealth going back in time

In the previous section, we have specified some simple investment strategies and computed their value at a terminal date T. Now we will go in the opposite direction, first specify the value of a derivative asset at T and then construct an investment fund that delivers this value. We assume throughout the risk-free rate r, the divident yield $\hat{\delta}$, and the log stock return volatility $\sqrt{c^{\ln S}} = \sigma$ are constant. We will denote the spot price of the derivative by F_t, and the spot value of the numeraire fund by either $\widetilde{F}$ or by $\overline{F}$.

2.3.1 Cash-or-nothing binary

The payoff of a cash-or-nothing binary call is $F_T = \mathbf{1}_{S_T > K}$. We will take $\overline{F}$ to be the fund with value of \$1 at T, i.e.,

$$\overline{F}_t = e^{-r(T-t)}$$

and $\overline{P}$ to be the corresponding risk-neutral measure which is commonly called the T-forward measure. Observe that $\overline{F}$ is just a fixed multiple of the money market fund of Subsection 2.2.2, therefore $\overline{P}$ coincides with the money market risk-neutral measure. The T–forward price equals

$$\overline{E}_t\left[\frac{F_T}{\overline{F}_T}\right] = \overline{E}_t[\mathbf{1}_{S_T>K}] = \overline{P}(S_T > K|\mathscr{F}_t).$$

To obtain the spot price, multiply the T–fwd price by the value of T–zero coupon discount bond,

$$\begin{aligned} F_t = \quad \bar{F}_t \quad &\times \quad \bar{\mathrm{E}}_t\left[\frac{F_T}{\bar{F}_T}\right] \\ = \mathrm{e}^{-r(T-t)}&\Phi\left(\frac{\ln\frac{S_t}{K} + (r - \hat{\delta} - \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned}$$

2.3.2 Asset-or-nothing binary

The payoff of an asset-or-nothing binary call is $F_T = S_T \mathbf{1}_{S_T>K}$. As a numeraire we will use the fund that contains one share at time T,

$$\widetilde{F}_t = S_t \mathrm{e}^{-\hat{\delta}(T-t)}.$$

We will call the corresponding risk-neutral measure the "stock–T–fwd" measure. Observe that $\widetilde{F}$ is a fixed multiple of the stock fund in Subsection 2.2.1, therefore the stock–T–fwd measure coincides with the stock fund risk-neutral measure in this setting. The stock–T–fwd price of the asset-or-nothing binary equals

$$\widetilde{\mathrm{E}}_t\left[\frac{F_T}{\widetilde{F}_T}\right] = \widetilde{\mathrm{E}}_t[\mathbf{1}_{S_T>K}] = \widetilde{P}(S_T > K|\mathscr{F}_t).$$

To obtain the spot price, multiply the stock–T–fwd price by the spot value of one share delivered at T,

$$\begin{aligned} F_t = \quad \widetilde{F}_t \quad &\times \quad \widetilde{\mathrm{E}}_t\left[\frac{F_T}{\widetilde{F}_T}\right] \\ = S_t \mathrm{e}^{-\hat{\delta}(T-t)}&\Phi\left(\frac{\ln\frac{S_t}{K} + (r - \hat{\delta} + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned}$$

Exercise 2.1. *Compute the price of the derivative with the payoff $S_T^\alpha \mathbf{1}_{S_T>K}$ for arbitrary $\alpha \in \mathbb{R}$.*

2.3.3 Black–Scholes formula revisited

We can now see that the Black–Scholes formula has a much nicer financial interpretation as a mix of asset-or-nothing and cash-or-nothing binary call options,

$$C_t = \widetilde{F}_t \times \widetilde{P}(S_T > K|\mathscr{F}_t) - K\bar{F}_t \times \bar{P}(S_T > K|\mathscr{F}_t).$$

2.3.4 Log contract

We will price $F_T = \ln S_T$ using the T–forward measure. We know that $\ln S$ is a continuous PII under this measure. The conditional mean of $\ln S_T - \ln S_t$ is $B_T^{\ln S} - B_t^{\ln S} = (r - \hat{\delta} - \sigma^2/2)(T-t)$. We therefore have

$$F_t = \mathrm{e}^{-r(T-t)}\left(\ln S_t + \left(r - \hat{\delta} - \frac{\sigma^2}{2}\right)\right)(T-t).$$

2.3.5 Power contract

We will price $F_T = S_T^\alpha$ using the T–forward measure. We know that

$$\ln S_T \sim N\left(\ln S_t + \left(r - \hat{\delta} - \sigma^2/2\right)(T-t), \sigma^2(T-t)\right),$$

hence

$$\alpha \ln S_T \sim N\left(\alpha \ln S_t + \alpha\left(r - \hat{\delta} - \sigma^2/2\right)(T-t), \alpha^2\sigma^2(T-t)\right).$$

Use the moment generating function of the normal distribution to obtain the forward price

$$\frac{F_t}{\overline{F}_t} = \mathrm{E}_t\left[S_T^\alpha\right] = \mathrm{E}_t\left[\mathrm{e}^{\alpha \ln S_T}\right] = S_t^\alpha \mathrm{e}^{\alpha\left(r-\hat{\delta}-\frac{\sigma^2}{2}\right)(T-t)+\frac{1}{2}\alpha^2\sigma^2(T-t)}.$$

Compare this formula for F_t with the constant proportions fund in (2.3) with initial value (2.4).

2.3.6 Average log contract

The payoff at T is

$$F_T = \frac{1}{T-T_0}\int_{T_0}^{T} \ln S_t dt.$$

Here $T > T_0 \geq 0$ is the forward start date of the contract. For details see handwritten notes.

2.4 Model-independent pricing

Some calculations above are universal, i.e., model-independent. For example, we have several times converted rate of return to log return. When X is continuous the universal formula for this reads

$$\mathscr{E}(X)_t = \mathrm{e}^{X_t - X_0 + \frac{1}{2}[X,X]_t}.$$

There is also a completely universal formula (works with jumps)

$$\mathscr{E}(X)_t \mathscr{E}(Y)_t = \mathscr{E}(X + Y + [X, Y])_t.$$

In particular, when X is continuous and Y has finite variation then $[X, Y] = 0$ and we obtain the useful simplification

$$\mathscr{E}(X + Y)_t = \mathscr{E}(X)_t \mathscr{E}(Y)_t.$$

These properties can be used to derive (2.3) quickly and easily.

2.5 Universal pricing formula

The common principle behind all the results we have seen so far is

$$\text{price of total return} = 1.$$

More mathematically,

$$t\text{–price of } \frac{F_T}{F_t} = 1.$$

Even more mathematically, fixing a numeraire fund $\overline{F}$ and the corresponding risk-neutral measure $\overline{P}$ one obtains

$$\overline{E}_t \left[\frac{F_T/F_t}{\overline{F}_T/\overline{F}_t} \right] = 1. \tag{2.5}$$

One immediate consequence of (2.5) is that if we consider another numeraire fund, say $\widetilde{F}$, then the two risk-neutral probabilities $\widetilde{P}$ and $\overline{P}$ must be related by the formula

$$\frac{d\widetilde{P}_t}{d\overline{P}_t} = \frac{\widetilde{F}_t/\widetilde{F}_0}{\overline{F}_t/\overline{F}_0},$$

meaning that

- the change between two risk-neutral measures equals the ratio of the corresponding numeraire fund returns;
- the two risk-neutral measures will be the same if and only if the two numeraire funds involved are a fixed multiple of each other;
- the ratio $F/\overline{F}$ is a $\overline{P}$–martingale for any F and any matching pair $(\overline{F}, \overline{P})$.

2.6 Construction of PDEs in Finance

Partial differential equations used in the pricing of derivatives express a simple fact, namely that a certain process related to the price of the derivative, expressed as a function of some state variables is a martingale under a suitably chosen probability measure. One can vary the three main ingredients by changing i) the measure, ii) the target function and iii) the state variables, but the main principle is always the same — some quantity is a martingale.

Let us first examine the most natural setting dictated by financial rather than computational considerations. Here the target function will be the option price, the state variables will consist of time and stock price and the relevant measure will be the risk-neutral measure.

2.6.1 Black–Scholes PDE

As discussed above the input parameters governing the rate of return process are

$$b^{\mathcal{L}(S)} = r - \hat{\delta}, c^{\mathcal{L}(S)} = \sigma^2.$$

From the financial theory the price of the derivative equals

$$C_t = e^{-r(T-t)} \mathrm{E}_t \left[f(S_T) \right].$$

We have $dS = Sd\mathcal{L}(S)$ and consequently

$$b^S = Sb^{\mathcal{L}(S)}, c^S = S^2 c^{\mathcal{L}(S)}.$$

By Proposition 1.7 S is a Markov process under the risk-neutral measure P and there is a function g such that

$$C_t = g(t, S_t).$$

Furthermore the process $C_t e^{-rt} = \mathrm{E}_t \left[f(S_T) e^{-rT} \right]$ is a martingale by Proposition 1.14 and therefore it must have zero drift. From the Itô formula

$$\begin{aligned} d\left(C_t \mathrm{e}^{-rt}\right) &= \mathrm{e}^{-rt}(dC_t - rC_t dt) \\ dC_t &= dg(t, S_t) = \partial_1 g(t, S_t) dt + \partial_2 g(t, S_t) dS_t + \frac{1}{2} \partial_{22} g(t, S_t) c_t^S dt \end{aligned}$$

and the zero drift condition for Ce^{-rt} yields the Black–Scholes PDE

$$\begin{aligned} 0 &= \partial_1 g(t, S) + \partial_2 g(t, S) b^{\mathcal{L}(S)} S + \frac{1}{2} \partial_{22} g(t, S) S^2 c^{\mathcal{L}(S)} - rg(t, S). \\ g(T, S) &= (S - K)^+ \end{aligned}$$

2.6.2 Change of variables and change of target function

The previous derivation suggests a way to simplify the martingale PDE. Firstly we will write the pay-off of the derivative asset as a function of $\ln S_T$. For a call option with strike K this means taking $f(x) = (\mathrm{e}^x - K)^+$. We take as the target function the forward option price $Y_t = \mathrm{E}_t[f(\ln S_T)]$. Since $\ln S$ is a Markov process under P (with $b^{\ln S} = r - \hat{\delta} - \sigma^2/2$, $c^{\ln S} = \sigma^2$) we have that $Y_t = g(t, \ln S_t)$ for some function g. Moreover, by virtue of (1.16) Y is a P–martingale, and therefore $b^Y = 0$. If g is sufficiently smooth the Itô formula yields

$$\begin{aligned}0 = b^Y &= \partial_1 g(t,x) + \partial_2 g(t,x) b^{\ln S} + \frac{1}{2}\partial_{22} g(t,x) c^{\ln S} \\ &= \partial_1 g(t,x) + \partial_2 g(t,x)(r - \hat{\delta} - \sigma^2/2) + \frac{1}{2}\partial_{22} g(t,x)\sigma^2,\end{aligned} \tag{2.6}$$

and the boundary condition reads

$$g(T,x) = (\mathrm{e}^x - K)^+ . \tag{2.7}$$

Once we solve for g the price of the derivative is given by $C_t = \mathrm{e}^{r(t-T)} g(t, \ln S_t)$.

2.6.3 Towards the heat equation

One can get rid of the drift term $\partial_2 g(t,x) b^{\ln S}$ in the previous PDE by using a state variable that is itself a martingale. One way to do this is to consider $X_t = \ln S_t + b^{\ln S}(T-t)$ under the money market risk-neutral measure P. Such a transformation works well for path-independent derivatives. For a call option with strike K this means solving

$$\begin{aligned}0 &= \partial_1 g(t,x) + \frac{1}{2}\partial_{22} g(t,x) c^{\mathcal{L}(S)}, \\ g(T,x) &= (e^x - K)^+.\end{aligned}$$

The option price at t is then given by

$$C_t = \mathrm{e}^{r(t-T)} g(t, \ln S_t + b^{\ln S}(T-t)).$$

2.6.4 Obtaining the heat equation by change of measure

Another useful way of transforming the pricing PDE is to keep $X = \ln S$, $f(x) = (\mathrm{e}^x - K)^+$ and consider a new measure $P^\star$ of the form

$$Z_t = \mathrm{e}^{\eta X_t + (\eta b^X + \frac{1}{2}\eta^2 c^X)(T-t)}, \tag{2.8}$$

$$\frac{dP^\star}{dP} = \frac{Z_T}{Z_0}, \tag{2.9}$$

for a constant η, as discussed in Section 1.11.2. Since we take P to be the money market risk-neutral measure, we have $b^X = b^{\ln S} = r - \hat{\delta} - \sigma^2/2$. In the Black–Scholes model $c^X = c^{\ln S} = \sigma^2$.

We know from Section 1.11.3 that $\mathrm{E}_t^{P^\star}[f(X_T)/Z_T]$ gives the forward price of $f(X_T)$ expressed in terms of the numeraire Z_t. The Girsanov theorem yields

$$b^X_{P^\star} = b^X + \eta c^X = r - \hat{\delta} - \frac{\sigma^2}{2} + \eta\sigma^2.$$

Consider the $P^\star$–martingale $\mathrm{E}_t^{P^\star}\left[f(X_T)e^{-\eta X_T}\right]$. Since X is Markov under $P^\star$ we conclude that there must be a function $g(t,x)$ such that

$$g(t, X_t) = \mathrm{E}_t^{P^\star}\left[f(X_T)e^{-\eta X_T}\right]. \tag{2.10}$$

Assuming that g is sufficiently differentiable, the Itô formula implies that g must satisfy the PDE

$$0 = \partial_1 g(t,x) + \partial_2 g(t,x)\underbrace{\left(r - \frac{\sigma^2}{2} + \eta\sigma^2\right)}_{b^X_{P^\star}} + \frac{1}{2}\partial_{22} g(t,x)\underbrace{\sigma^2}_{c^X}, \tag{2.11}$$

$$g(T,x) = \mathrm{e}^{-\eta x} f(x). \tag{2.12}$$

Recall that g gives the price of the derivative in terms of the numeraire Z. Once we solve for g, the spot price of the option at time t is given by

$$C_t = \mathrm{e}^{r(t-T)} Z_t g(t, X_t) \tag{2.13}$$

$$= \mathrm{e}^{(\eta(r-\sigma^2/2)+\frac{1}{2}\eta^2\sigma^2 - r)(T-t)} S_t^\eta g(t, \ln S_t). \tag{2.14}$$

We observe that the numeraire in this case is a fund whose value at time T equals S_T^η, a power contract. Two values of η are worth pointing out. Firstly,

$$\eta = -b^{\ln S}/c^{\ln S} = 1/2 + (\hat{\delta} - r)/\sigma^2$$

will give $b^{\ln S}_{P^\star} = 0$. Secondly, $\eta = 1$ will turn the payoff in the numeraire units, $f(X_T)\mathrm{e}^{-\eta X_T} = \left(1 - K\mathrm{e}^{-X_T}\right)^+$, into a function bounded between 0 and 1. The value $\eta = 1$ corresponds to the T–forward stock risk-neutral measure because the numeraire is one share sold forward to be delivered at T.

Remark 2.2. *Wilmott, Howison, and Dewynne (1995) use $\eta = -b^{\ln S}/c^{\ln S}$ and assume zero dividend yield, $\hat{\delta} = 0$, which leads to $\eta = 1/2 - r/\sigma^2$ leading to $b^{\ln S}_{P^\star} = 0$. They use slightly different state variables: $X = \ln S - \ln K$ instead of $\ln S$ and so-called "operational time to maturity" $\tau = \frac{\sigma^2}{2}(T-t)$ instead of t. The target function is $u(\tau, x) = g(t,x)/K$. On* *account of* $\frac{\partial}{\partial t} = \frac{d\tau}{dt}\frac{\partial}{\partial \tau}$ *equations (2.11, 2.12) yield*

$$\begin{aligned} 0 &= -\frac{\sigma^2}{2}\partial_1 u(\tau,x) + \frac{\sigma^2}{2}\partial_{22} u(\tau,x), \\ u(0,x) &= \mathrm{e}^{-\eta x}\left(e^x - 1\right)^+, \\ C_t &= K\mathrm{e}^{-\eta \ln K}\mathrm{e}^{\eta \ln S_t + \left(\eta(r-\sigma^2/2) - r + \frac{1}{2}\eta^2\sigma^2\right)(T-t)} u\left(\frac{\sigma^2}{2}(T-t), \ln\frac{S_t}{K}\right). \end{aligned} \tag{2.15}$$

On rearranging the PDE turns into a heat equation

$$\partial_1 u(\tau,x) = \partial_{22} u(\tau,x).$$

References

Černý, A. (2009). *Mathematical Techniques in Finance: Tools for Incomplete Markets*. 2nd. Princeton University Press, Princeton, NJ.

Wilmott, P., S. Howison, and J. Dewynne (1995). *The Mathematics of Financial Derivatives*. Cambridge University Press.