An Introduction to Point Processes

from a

Martingale Point of View

Tomas Björk

KTH, 2011

Preliminary, incomplete, and probably with lots of typos

Contents

Part I

The Mathematics of Counting Processes

Chapter 1

Counting Processes

1.1 Generalities and the Poisson process

Good textbooks on point processes are [2] and [3]. The simplest type of a point process is a **counting process**, and the formal definition is as follows.

Definition 1.1.1 *A random process* $\{N_t;\ t \in R_+\}$ *is a* **counting process** *if it satisfies the following conditions.*

1. *The trajectories of* N *are, with probability one, right continuous and piecewise constant.*

2. *The process starts at zero, so*

$$N_0 = 0.$$

3. *For each* t

$$\Delta N_t = 0, \quad or \quad \Delta N_t = 1$$

with probability one. Here ΔN_t *denotes the jump size of* N *at time* t*, or more formally*

$$\Delta N_t = N_t - N_{t-}.$$

In more pedestrian terms, the process N starts at $N_0 = 0$ and stays at the level 0 until some random time T_1 when it jumps to $N_{T_1} = 1$. It then stays at level 1 until the another random time T_2 when it jumps to the value $N_{T_2} = 2$ etc. We will refer to the random times $\{T_n;\ n = 1, 2, \ldots\}$ as the **jump times** of N. Counting processes are often used to model situations where some sort of well specified **events** are occurring randomly in time. A typical example of an event could be the arrival of a new customer to a queue, an earthquake in a well specified geographical area, or a company going bankrupt. The interpretation is then that N_t denotes the number of events that has occurred in the time interval $[0, t]$. Thus N_t could be the number of customer which have arrived to a certain queue during the interval $[0, t]$ etc. With this interpretation, the jump times $\{T_n;\ n = 1, 2, \ldots\}$ are often also referred to as the **event times** of the process N.

Before we go on to the general theory of counting processes, we will study the **Poisson process** in some detail. The Poisson process is the single most important of all counting processes, and among counting processes it has very much the same position that the Wiener processes has among the diffusion processes. We start with some elementary facts concerning the Poisson distribution.

Definition 1.1.2 *A random variable X is said to have a* **Poisson distribution** *with parameter α if it takes values among the natural numbers, and the probability distribution has the form*

$$P(X = n) = e^{-\alpha}\frac{\alpha^n}{n!}, \quad n = 0, 1, 2, \ldots$$

We will often write this as $X \sim Po(\alpha)$.

We recall that, for any random variable X, its **characteristic function** φ_X is defined by

$$\varphi_X(u) = E\left[e^{iuX}\right], \quad u \in R,$$

where i is the imaginary unit. We also recall that the distribution of X is completely determined by φ_X. We will need the following well known result concerning the Poisson distribution.

Proposition 1.1.1 *Let X be $Po(\alpha)$. Then the characteristic function is given by*

$$\varphi_X(u) = e^{\alpha\left(e^{iu}-1\right)}$$

The mean and variance are given by

$$E[X] = \alpha, \quad Var(X) = \alpha.$$

Proof. This is left as an exercise. ∎

We now leave the Poisson distribution and go on to the Poisson process.

Definition 1.1.3 *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a given filtration $\mathbf{F} = \{\mathcal{F}_t\}_{t\geq 0}$, and let λ be a nonnegative real number. A counting process N is a* **Poisson process with intensity λ with respect to the filtration F** *if it satisfies the following conditions.*

1. *N is adapted to $\mathbf{F}$.*

2. *For all $s \leq t$ the random variable $N_t - N_s$ is independent of $\mathcal{F}_s$.*

3. *For all $s \leq t$, the conditional distribution of the increment $N_t - N_s$ is given by*

$$P\left(N_t - N_s = n \left| \mathcal{F}_s \right.\right) = e^{-\lambda(t-s)} \frac{\lambda^n (t-s)^n}{n!}, \quad n = 0, 1, 2, \ldots \tag{1.1}$$

In this definition we encounter the somewhat forbidding looking formula (1.1). As it turns out, there is another way of characterizing the Poisson process, which is much easier to handle than distributional specification above. This alternative characterization is done in terms of the "infinitesimal characteristics" of the process, and we now go on to discuss this.

1.2 Infinitesimal characteristics

One of the main ideas in modern process theory is that the "true nature" of a process is revealed by its "infinitesimal characteristics". For a diffusion process the infinitesimal characteristics are the drift and the diffusion terms. For a counting process, the natural infinitesimal object is the "predictable conditional jump probability per unit time", and informally we define this as

$$\frac{P\left(dN_t = 1 \left| \mathcal{F}_{t-} \right.\right)}{dt}.$$

The increment process dN is defined as

$$dN_t = N_t - N_{t-} = N_t - N_{t-dt},$$

and the sigma algebra $\mathcal{F}_{t-}$ is defined by

$$\mathcal{F}_{t-} = \bigvee_{0 \leq s < t} \mathcal{F}_s \tag{1.2}$$

The reason why we define dN_t as $N_t - N_{t-dt}$ instead of $N_t - N_{t+dt}$ is that we want the increment process dN to be adapted. The term "predictable" will be very important later on in the text, and will be given a precise mathematical definition. We also note that the increment dN_t only takes two possible values, namely $dN_t = 0$ or $dN_t = 1$ depending on whether or not an event has occurred at time t. We can thus write the conditional jump probability as an expected value, namely as

$$P\left(dN_t = 1 \left| \mathcal{F}_{t-} \right.\right) = E^P\left[dN_t \right| \mathcal{F}_{t-}].$$

Suppose now that N is a Poisson process with intensity λ, and that h is a small real number. According to the definition we then have

$$P\left(N_t - N_{t-h} = 1 \left| \mathcal{F}_{t-h} \right.\right) = e^{-\lambda h} \lambda h.$$

Expanding the exponential we thus have

$$P\left(N_t - N_{t-h} = 1 \left| \mathcal{F}_{t-h} \right.\right) = \lambda h \sum_{n=0}^{\infty} \frac{(\lambda h)^n}{n!}.$$

As h becomes "infinitesimally small" the higher order terms can be neglected and as a formal limit when $h \to dt$ we obtain

$$P\left(dN_t = 1 \left| \mathcal{F}_{t-}\right.\right) = \lambda dt, \tag{1.3}$$

or equivalently

$$E^P\left[\left. dN_t \right| \mathcal{F}_{t-}\right] = \lambda dt. \tag{1.4}$$

This entire discussion has obviously been very informal, but nevertheless the formula (1.4) has a great intuitive value. It says that we can interpret the parameter λ as the **conditional jump intensity**. In other words, λ is the (conditional) expected number of jumps per unit of time. The point of this is twofold.

- The concept of a conditional jump intensity is easy to interpret intuitively, and it can also easily be generalized to a large class of counting processes.
- As we will see below, the distribution of a counting process is completely determined by its conditional jump intensity, and equation (1.4) is **much** simpler than that equation (1.1).

The main project of this text is to develop a mathematically rigorous theory of counting processes, building on the intuitively appealing concept of a conditional jump intensity. As the archetypical example we will of course use the Poisson process, and to start with we need to reformulate the nice but very informal relation (1.4) to something more mathematically precise. To do this we start by noting (again informally) that if we subtract the conditional expected number of jumps λdt from the actual number of jumps dN_t then the result

$$dN_t - \lambda dt,$$

should have zero conditional mean. The implication of this is that we are led to conjecture that if we define the process M by

$$\begin{cases} dM_t &= dN_t - \lambda dt, \\ M_0 &= 0, \end{cases}$$

or, equivalently, on integrated form as

$$M_t = N_t - \lambda t,$$

then M should be a martingale. This conjecture is in fact true.

Proposition 1.2.1 *Assume that N is an $\mathbf{F}$- Poisson process with intensity λ. Then the process M, defined by*

$$M_t = N_t - \lambda t, \tag{1.5}$$

is an $\mathbf{F}$ martingale.

Proof. The proof is easy and left to the reader. ∎

This somewhat trivial result is much more important than it looks like at first sight. It is in fact the natural starting point of the "martingale approach" to counting processes. As we will see below, the martingale property of M above, is not only a **consequence** of the fact that N is a Poisson process but, in fact, the martingale property **characterizes** the Poisson process within the class of counting processes. More precisely, we will show below that if N is a arbitrary counting process and if the process M, defined as above is a martingale, then this **implies** that N must be Poisson with intensity λ. This is a huge technical step forward in the theory of counting processes, the reason being that it is often relatively easy to check the martingale property of M, whereas it is typically a very hard task to check that the conditional distribution of the increments of N is given by (1.1).

It also turns out that a very big class of counting processes can be characterized by a corresponding martingale property and this fact, coupled with a (very simple form of) stochastic differential calculus for counting processes, will provide us with a very powerful tool box for a fairly advanced study of counting processes on filtered probability spaces.

To develop this theory we need to carry out the following program.

1. Assuming that a process A is of bounded variation, we need to develop a theory of stochastic integrals of the form
$$\int_0^t h_s dA_s,$$
where the integrand h should be required have some nice measurability property.

2. In particular, if M is a martingale of bounded variation, we would like to under what conditions a process X of the form
$$X_t = \int_0^t h_s dM_s,$$
is a martingale. Is it for example enough that h is adapted? (Compare the Wiener case).

3. Develop a differential calculus for stochastic integrals of the type above. In particular we would like to derive an extension of the Itô formula to the counting process case.

4. Use the theory developed in the previous items to study general counting processes in terms of their martingale properties.

5. Given a Wiener process W, we recall that there exists a powerful martingale representation theorem which says that (for the internal filtration) **every** martingale X can be written as $X_t = X_0 + \int_o^t h_s dW_s$. Does there exist a corresponding theory for counting processes?

6. Study how the conditional jump intensity will change under an absolutely continuous change of measure. Does there exist a Girsanov theory for counting processes?

7. Finally we want to apply the theory above in order to study more concrete problems, like queuing theory, and arbitrage theory for economies where asset prices are driven by jump diffusions.

1.3 Exercises

Exercise 1.1 *Prove Proposition 1.1.1.*

Exercise 1.2 *Prove Proposition 1.2.1.*

Chapter 2

Stochastic Integrals and Differentials

2.1 Integrators of bounded variation

In this section,the main object is to develop a stochastic integration theory for integrals of the form

$$\int_0^t h_s dA_s,$$

where A is a process of bounded variation. In a typical application, the integrator A could for example be given by

$$A_t = N_t - \lambda t,$$

where N is a Poisson process with intensity λ, and in particular we will investigate under what conditions the process X defined by

$$X_t = \int_0^t h_s \left[dN_s - \lambda ds \right],$$

is a martingale. Apart from this, we also need to develop a stochastic differential calculus for processes of this kind, and to study stochastic differential equations, driven by counting processes.

Before we embark on this program,the following two points are worth mentioning.

- Compared to the definition of the usual Itô integral for Wiener processes, the integration theory for point processes is quite simple. Since all integrators will be of bounded variation, the integrals can be defined path wise, as opposed to the Itô integral which has to be defined as an L^2 limit.
- On the other hand, compare to the Itô integral, where the natural requirement is that the integrands are adapted, the point process integration theory requires much more delicate measurability properties of the

integrands. In particular we need to understand the fundamental concept of a **predictable process**.

In order to get a feeling for the predictability concept, and its relation to martingale theory, we will start by giving a brief recapitulation of discrete time stochastic integration theory.

2.2 Discrete time stochastic integrals

In this section we discuss briefly the simplest type of stochastic integration, namely integration of discrete time processes. This will thus serve as an introduction to the more complicated continuous time theory later on, and it is also important in its own right. We start by defining the discrete stochastic integral.

Definition 2.2.1 *Consider a probability space $(\Omega, \mathcal{F}, P)$, equipped with a discrete time, filtration $\mathbf{F} = \{\mathcal{F}_n\}_{n=0}^{\infty}$.*

- *For any random process X, the* **increment process** *ΔX is defined by*

$$(\Delta X)_n = X_n - X_{n-1}, \tag{2.1}$$

with the convention $X_{-1} = 0$. For simplicity of notation we will sometimes denote $(\Delta X)_n$ by ΔX_n.

- *For any two processes X and Y, the* **discrete stochastic integral** *process $X \star Y$ is defined by*

$$(X \star Y)_n = \sum_{k=0}^{n} X_k (\Delta Y)_k. \tag{2.2}$$

Instead of $(X \star Y)_n$ we will sometimes write $\int_0^n X_s dY_s$.

The reason why we define ΔX by "backward increments" above, is that in this way the process ΔX is adapted, whenever X is adapted.

From standard Itô integration theory we recall that if W is a Wiener process and if h is a square integrable adapted process the integral process Z, given by

$$Z_t = \int_0^t h_s dW_s$$

is a martingale. It is therefore natural to expect that a similar result holds for the discrete time integral, but this is not the case. As we will see below, the correct measurability concept is that of a **predictable process** rather than that of an adapted process.

Definition 2.2.2

- *A random process X is* **F-adapted** *if, for each n, X_n is $\mathcal{F}_n$ measurable.*

- *A random process X is* **F-predictable** *if, for each n, X_n is $\mathcal{F}_{n-1}$ measurable. Here we use the convention $\mathcal{F}_{-1} = \mathcal{F}_0$.*

We note that a predictable process is "known one step ahead in time".

The main result for stochastic integrals is that when you integrate a **predictable process** X w.r.t. a martingale M, then the result is a new martingale.

Proposition 2.2.1 *Assume that the space $(\Omega, \mathcal{F}, P, \mathbf{F})$ carries the processes X and M where X is predictable, M is a martingale, and $X_n(\Delta M)_n \in L^1$ for each n. Then the stochastic integral $X \star M$ is a martingale.*

Proof. We recall that in discrete time, a process Z is a martingale if and only if it satisfies the following condition.

$$E\left[\Delta Z_n \middle| \mathcal{F}_{n-1}\right] = 0, \quad n = 0, 1, \ldots$$

Defining Z as

$$Z_n = \sum_{k=0}^{n} X_k \Delta M_k.$$

it is clear that

$$\Delta Z_n = X_n \Delta M_n$$

and we obtain

$$E\left[\Delta Z_n \middle| \mathcal{F}_{n-1}\right] = E\left[X_n \Delta M_n \middle| \mathcal{F}_{n-1}\right] = X_n E\left[\Delta M_n \middle| \mathcal{F}_{n-1}\right] = 0.$$

In the second equality we used the fact that X is predictable, and in the third equality we used the martingale property of M. ∎

2.3 Stochastic integrals in continuous time

We now go back to continuous time and assume that we are given a filtered probability space $(\Omega, \mathcal{F}, P, \mathbf{F})$. Before going on to define the new stochastic integral we need to define a number of measurability properties for random processes, and in particular we need to define the continuous time version of the predictability concept.

Definition 2.3.1

- *A random process X is said to be* **cadlag** *(continu à droite, limites à gauche) if the trajectories are right continuous with left hand limits, with probability one.*

- *The class of adapted cadlag processes A with $A_0 = 0$, such that the trajectories of A are of finite variation on the interval $[0, T]$is denoted by $\mathcal{V}_T$. Such a process is said to be of* **finite variation on** $[0, T]$*, and will thus satisfy the condition*
$$\int_0^T |dA_t| < \infty.$$
- *We denote by $\mathcal{A}_T$ the class of processes in $\mathcal{V}_T$ such that*
$$E\left[\int_0^T |dA_t|\right] < \infty.$$
Such a process is said to be of **integrable variation on** $[0, T]$.
- *The class of processes belonging to $\mathcal{V}_T$ for all $T < 0$ and is denoted by $\mathcal{V}$. Such a process is said to be of* **finite variation**.
- *The class of processes belonging to $\mathcal{A}_T$ for all $T < 0$ and is denoted by $\mathcal{A}$. Such a process is said to be of* **integrable variation**.

Remark 2.3.1 *Note that the cadlag property, as well as the property of being adapted is built into the definition of $\mathcal{V}_T$ and $\mathcal{A}_T$.*

We now come to the two main measurability properties of random processes. before we go on to the definitions, we recall that a random process X on the time interval R_+ is a mapping
$$X : \Omega \times R_+ \to R,$$
where the value of X at time t, for the elementary outcome $\omega \in \Omega$ is denoted by either $X(t, \omega)$ or by $X_t(\omega)$.

Definition 2.3.2 *The* **optional** *σ-algebra on $R_+ \times \Omega$ is generated by all processes Y of the form*
$$Y_t(\omega) = Z(\omega) I\{r \le t < s\}, \tag{2.3}$$
where I is the indicator function, r and s are fixed real numbers, and Z is an $\mathcal{F}_s$ measurable random variable. A process X which, viewed as a mapping $X : \Omega \times R_+ \to R$, is measurable w.r.t the optional σ-algebra is said to be an **optional process**.

The definition above is perhaps somewhat forbidding when you meet it the first time. Note however, that every generator process Y above is adapted and cadlag, and we have in fact the following result, the proof of which is nontrivial and omitted.

Proposition 2.3.1 *The optional σ algebra is generated by the class of adapted cadlag processes.*

In particular it is clear that every process of finite variation, and every adapted process with continuous trajectories is optional. The optional measurability concept is in fact "the correct one" instead of the usual concept of a process being adapted. The difference between an adapted process and an optional one is that optionality for a process X implies a joint measurability property in (t, ω), whereas X being adapted only implies that the mapping $X_t : \Omega \to R$ is $\mathcal{F}_t$ measurable in ω for each fixed t. For "practical" purposes, the difference between an adapted process and an optional process is very small and the reader may, without great risk, interpret the term "optional" as "adapted". The main point of the optionality property is the following result, which shows that optionality is preserved under stochastic integration.

Proposition 2.3.2 *Assume that A is of finite variation and that h is an optional process satisfying the condition*

$$\int_0^t |h_s||dA_s| < \infty, \quad \textit{for all } t.$$

Then the following hold.

- *The process $X = h \star A$ defined, for each ω, by*

$$X_t(\omega) = \int_0^t h_s(\omega) dA_s(\omega),$$

  is well defined, for almost each ω, as a Lebesgue Stieltjes integral.

- *The process X is cadlag and optional, so in particular it is adapted.*

- *If h also satisfies the condition*

$$E\left[\int_0^t |h_s||dA_s|\right] < \infty, \quad \textit{for all } t.$$

  then X is of integrable variation.

Proof. The proposition is easy to prove if h is generator process of the form (2.3). The general case can then be proved by approximating h by a linear combination of generator processes. ∎

Remark 2.3.2 *Note again that since A is of finite variation it is, by definition, also optional. If we only require that h is adapted and A of finite variation (and thus adapted), then this would* **not** *guarantee that X is adapted.*

2.4 Stochastic integrals and martingales

Suppose that M is a martingale of integrable variation. We now turn to the question under which conditions on the integrand h, a stochastic process of the form

$$X_t = \int_0^t h_s dM_s,$$

is itself a martingale. With the Wiener theory in fresh memory, one is perhaps led to conjecture that it is enough to require that h (apart from obvious integrability properties) is adapted, or perhaps optional. This conjecture is, however, **not** correct and it is easy to construct a counter example.

Example 2.4.1 *Let Z be a nontrivial random variable with*

$$E[Z] = 0, \quad E\left[Z^2\right] < \infty,$$

and define the process M by

$$M_t = \begin{cases} 0, & 0 \le t < 1, \\ Z, & t \ge 1. \end{cases}$$

If we define the filtration $\mathbf{F}$ by $\mathcal{F}_t = \sigma\{M_s;\ s \le t\}$, then it is easy to see that M is a martingale of integrable variation. In particular, M is optional, so let us define the integrand h as $h = M$. If we now define the process X by

$$X_t = \int_0^t h_s dM_s,$$

then it is clear that the integrator M has a point mass of size Z at $t = 1$. In particular we have $X_1 = h_1 \Delta M_1 = Z^2$, and we immediately obtain

$$X_t = \begin{cases} 0, & 0 \le t < 1, \\ Z^2, & t \ge 1. \end{cases}$$

From this it is clear that X is a non decreasing process, so in particular it is **not** *a martingale.*

Note, however, that if we define h as $h_t = M_{t-}$ then X will be a martingale (why?). As we will see, it is not a coincidence that this choice of h is **left** *continuous.*

It is clear from this example that we must demand more than mere optionality from the integrand h in order to ensure that that the stochastic integral $h \star M$ is a martingale. From the discrete time theory we recall that if M is a martingale and if h is **predictable**, then $h \star M$ is a martingale. We also recall that predictability of h in discrete time means that $h_n \in \mathcal{F}_{n-1}$ and the question is how to generalize this concept to continuous time.

The obvious idea is of course to say that a continuous time process h is predictable if $h_t \in \mathcal{F}_{t-}$ for all $t \in R_+$, and in order to see if this is a good idea we now give some informal and heuristic arguments. Let us thus assume that M is a martingale of bounded variation, that $h_t \in \mathcal{F}_{t-}$ for all $t \in R_+$, and that all necessary integrability conditions are satisfied. We define the process X by

$$X_t = \int_0^t h_s dM_s,$$

and we now want to check if X is a martingale. Loosely speaking, and comparing with discrete time theory, we expect the process X to be a martingale if and only if

$$E\left[\left. dX_t \right| \mathcal{F}_{t-}\right] = 0,$$

for all t. By definition we have

$$dX_t = h_t dM_t,$$

so we obtain

$$E\left[\left. dX_t \right| \mathcal{F}_{t-}\right] = E\left[\left. h_t dM_t \right| \mathcal{F}_{t-}\right].$$

Since $h_t \in \mathcal{F}_{t-}$, we can pull this term outside the expectation, and since M is a martingale we have $E\left[\left. dX_t \right| \mathcal{F}_{t-}\right] = 0$, so we obtain

$$E\left[\left. dX_t \right| \mathcal{F}_{t-}\right] = h_t E\left[\left. dM_t \right| \mathcal{F}_{t-}\right] = 0,$$

thus "proving" that X is a martingale.

This very informal argument is very encouraging, but it turns out that the requirement $h_t \in \mathcal{F}_{t-}$ is not quite good enough for our purposes, the main reason being that, for each fixed t, it is a measurability argument in the ω variable only. In particular the requirement $h_t \in \mathcal{F}_{t-}$ has the weakness that it does not guarantee that X is adapted. We thus need to refine the simple idea above, and it turns out that the following definition is exactly what we need.

Definition 2.4.1 *Given a filtered probability space* $(\Omega, \mathcal{F}, P, \mathbf{F})$*, we define the* **F-predictable** *σ-algebra Σ_P on $R_+ \times \Omega$ as the σ-algebra generated by all processes Y of the form*

$$Y_t(\omega) = Z(\omega) I\left\{r < t \leq s\right\}, \tag{2.4}$$

where r and s are real numbers and the random variable Z is $\mathcal{F}_r$- measurable. A process X which is measurable w.r.t. the predictable σ-algebra is said to be an **F-predictable process**.

This definition is the natural generalization of the predictability concept from discrete time theory, and it is extremely important to notice that all the generator processes Y above are **left** continuous and adapted. It is also possible to show the following result, the proof of which is omitted.

Proposition 2.4.1 *The predictable σ-algebra is also generated by the class of left continuous adapted processes.*

In particular, this result implies that every adapted left continuous process is predictable, and a very important special case of a predictable process is obtained if we start with an adapted cadlag process X and then define a new process Y by

$$Y_t = X_{t-}.$$

Since Y is left continuous and adapted it will certainly be predictable, and most of the predictable processes that we will meet in "practice" are in fact of this form.

Remark 2.4.1 *The working mathematician can, without great risk, interpret the term "predictable" as either "adapted and left continuous" or as "$\mathcal{F}_{t-}$-adapted".*

We can now state the main results of this section.

Proposition 2.4.2 *Assume that M is a martingale of bounded variation and that h is a predictable process satisfying the condition*

$$E\left[\int_0^t |h_s||dM_s| < \infty\right], \tag{2.5}$$

for all $t \geq 0$. Then the process X defined by

$$X_t = \int_0^t h_s dM_s,$$

is a martingale.

Proof. It is very easy to show that if h is a generator process of the form (2.4) then X is a martingale. The general result then follows by a (non trivial) approximation argument. ∎

We will also need the following result, which shows how the predictability property is inherited by stochastic integration.

Proposition 2.4.3 *Let A be a predictable process of bounded variation (so in particular A is cadlag) and let h be a predictable process satisfying*

$$E\left[\int_0^t |h_s||dA_s| < \infty\right], \tag{2.6}$$

for all $t \geq 0$. Then the integral process

$$X_t = \int_0^t h_s dA_s$$

is predictable.

Proof. The result is obvious when h is a generator process of the form (2.4). The general result follows by an approximation argument. ∎

We finish this section with a useful lemma.

Lemma 2.4.1 *Assume that X is optional and cadlag, and define the process Y by $Y_t = X_{t-}$. Then Y is predictable and for any optional process h we have*

$$\int_0^t h_s X_s ds = \int_0^t h_s Y_s ds,$$

for all $t \geq 0$.

Proof. *The predictability follows from the fact that Y is left continuous and adapted. Since X is cadlag, X and Y will (for a fixed trajectory) only differ on a finite number of points, and since we are integrating w.r.t. Lebesgue measure the integrals will coincide.* ∎

2.5 The Itô formula

Given the standard setting of a filtered probability space, let us consider an optional cadlag process X. If X can be represented on the form

$$X_t = X_0 + A_t + \int_0^t \sigma_s dW_s, \quad t \in R_+, \tag{2.7}$$

where the process, A, is of bounded variation, W is a Wiener process, and σ is an optional process, then we say that X has a **stochastic differential** and we write

$$dX_t = dA_t + \sigma_t dW_t. \tag{2.8}$$

In our applications, the process A will always be of the form

$$dA_t = \mu_t dt + h_t dN_t, \tag{2.9}$$

where μ and h are predictable and N is a counting process, but in principle we allow A to an arbitrary process of bounded variation (and thus cadlag and adapted). As in the pure Wiener case, it is important to note that the differential expression (2.8) is, by definition, nothing else than a shorthand notation for the integral expression (2.7).

The first question to ask is whether there exists an Itô formula for processes of this kind. In other words, let X have a stochastic differential of the form (2.8), let $F(t, x)$ be a given a smooth function, and define the process Z by

$$Z_t = F(t, X_t). \tag{2.10}$$

The question is now whether Z has a stochastic differential and, if so, what it looks like. This questions is answered within general semi martingale theory, but since that theory is outside the scope of the present text we will only discuss the simpler case when X has the differential

$$dX_t = \mu_t dt + \sigma_t dW_t + h_t dN_t. \tag{2.11}$$

Now, **between the jumps** of N the process X will have the dynamics

$$dX_t = \mu_t dt + \sigma_t dW_t,$$

and this is of course handled by the standard Itô formula

$$dZ_t = \left\{ \frac{\partial F}{\partial t}(t, X_t) + \mu_t \frac{\partial F}{\partial x}(t, X_t) + \frac{1}{2}\sigma_t^2 \frac{\partial^2 F}{\partial x^2}(t, X_t) \right\} dt + \sigma_t \frac{\partial F}{\partial x}(t, X_t) dW_t.$$

On the other hand, **at a jump time** t, the process N has a jump size of $\Delta N_t = N_t - N_{t-} = 1$ which implies that the process X will have a jump of size

$$\Delta X_t = h_t \Delta N_t = h_t.$$

Since $Z_t = F(t, X_t)$, the induced jump of Z is given by

$$\Delta Z_t = F(t, X_t) - F(t-, X_{t-}),$$

and since $X_t = X_{t-} + \Delta X_t = X_{t-} + h_t$ we obtain

$$\Delta Z_t = F(t, X_{t-} + h_t) - F(t-, X_{t-}),$$

and since F is assumed to be smooth we can also write this as

$$\Delta Z_t = F(t, X_{t-} + h_t) - F(t, X_{t-}).$$

If we note that $dN_t = 1$ at a jump time and that $dN_t = 0$ at times of no jumps, we can summarize our findings as follows, where the extension to a multi dimensional Wiener process is obvious.

Proposition 2.5.1 *Assume that X has dynamics of the form*

$$dX_t = \mu_t dt + \sigma_t dW_t + h_t dN_t, \tag{2.12}$$

where μ, σ, and h are predictable, and W is a Wiener process. Let F be a $C^{1,2}$ function. Then the following Itô formula holds.

$$\begin{aligned} dF(t, X_t) &= \left\{ \frac{\partial F}{\partial t}(t, X_t) + \mu_t \frac{\partial F}{\partial x}(t, X_t) + \frac{1}{2}\sigma_t^2 \frac{\partial^2 F}{\partial x^2}(t, X_t) \right\} dt \\ &+ \sigma_t \frac{\partial F}{\partial x}(t, X_t) dW_t \\ &+ \left\{ F(t, X_{t-} + h_t) - F(t, X_{t-}) \right\} dN_t. \end{aligned} \tag{2.13}$$

We see that this is just the standard Itô formula, with the added term

$$\{F(t, X_{t-} + h_t) - F(t, X_{t-})\}\, dN_t$$

If t is not a jump time of N, then $dN_t = 0$ so the jump term disappears. If, on the other hand, N has a jump at time t, then $dN_t = 1$ and the jump term $F(t, X_{t-} + h_t) - F(t, X_{t-})$ is added.

We can now compare this version of the Itô formula to what we get by doing a naive and straightforward Taylor expansion at $t-$. The first order terms are

$$\frac{\partial F}{\partial t}(t-, X_{t-})dt + \frac{\partial F}{\partial x}(t-, X_{t-})dX_t,$$

which by smoothness of F and Lemma 2.4.1can be written as

$$\frac{\partial F}{\partial t}(t, X_t)dt + \frac{\partial F}{\partial x}(t, X_{t-})dX_t.$$

By substituting (2.12) we obtain

$$\begin{aligned}
&\frac{\partial F}{\partial t}(t, X_t)dt + \frac{\partial F}{\partial x}(t, X_{t-})dX_t \\
&= \left\{ \frac{\partial F}{\partial t}(t, X_t) + \mu_t \frac{\partial F}{\partial x}(t, X_t) \right\} dt + \sigma_t \frac{\partial F}{\partial x}(t, X_t)dW_t + \frac{\partial F}{\partial x}(t, X_{t-})h_t dN_t \\
&= \left\{ \frac{\partial F}{\partial t}(t, X_t) + \mu_t \frac{\partial F}{\partial x}(t, X_t) \right\} dt + \sigma_t \frac{\partial F}{\partial x}(t, X_t)dW_t + \frac{\partial F}{\partial x}(t, X_{t-})\Delta X_t,
\end{aligned}$$

where we have used the fact that $h_t dN_t = \Delta X_t$. Comparing this expression to the Itô formula above, and writing $\{F(t, X_{t-} + h_t) - F(t, X_{t-})\}\, dN_t = \Delta F(t, X_t)$ we can write the Itô formula as

$$\begin{aligned}
dF(t, X_t) &= \frac{\partial F}{\partial t}(t, X_t)dt + \frac{\partial F}{\partial x}(t, X_{t-})dX_t + \frac{1}{2}\sigma_t^2 \frac{\partial^2 F}{\partial x^2}(t, X_t)dt \\
&+ \left\{ \Delta F(t, X_t) - \frac{\partial F}{\partial x}(t, X_{t-})\Delta X_t \right\}.
\end{aligned}$$

This result does in fact hold in great generality, and we formulate it as a proposition.

Theorem 2.5.1 *Assume that X has the dynamics*

$$dX_t = dA_t + \sigma_t dW_t + h_t dN_t, \tag{2.14}$$

where A is of bounded variation, and the other terms as above. Assume furthermore that F is a $C^{1,2}$ function. Then the following holds.

$$dF(t, X_t) = \frac{\partial F}{\partial t}(t, X_t)dt + \frac{\partial F}{\partial x}(t, X_{t-})dX_t + \frac{1}{2}\sigma_t^2 \frac{\partial^2 F}{\partial x^2}(t, X_t)dt \tag{2.15}$$

$$+ \left\{ \Delta F(t, X_t) - \frac{\partial F}{\partial x}(t, X_{t-})\Delta X_t \right\}. \tag{2.16}$$

Remark 2.5.1 *Note the evaluation of X at X_{t-} in the term $\frac{\partial F}{\partial x}(t, X_{t-})dX_t$. This implies that the process $\frac{\partial F}{\partial t}(t, X_{t-})$ is* **predictable**, *and thus that any martingale component in X will be integrated to a new martingale.*

Remark 2.5.2 *The Various forms of the Itô formula above generalize in the obvious way to the multi dimensional case.*

There is one important special case of the Itô formula for processes of bounded variation.

Proposition 2.5.2 *Assume that X and Y are processes of bounded variation (i.e. with no Wiener component). Then the following holds*

$$d\left(X_t Y_t\right) = X_{t-}dY_t + Y_{t-}dX_t + \Delta X_t \Delta Y_t. \tag{2.17}$$

Proof. The proof is left to the reader. ∎

Again the reason for the evaluation at $t-$ in $X_{t-}dY$ and $Y_{t-}dX_t$ is that this implies predictability of the integrands, and thus implies that martingale components of Y and X will be integrated to new martingales.

2.6 Stochastic differential equations

In this section we will apply the Itô formula in order to study stochastic differential equations driven by a counting process. This turns out to be a bit delicate, and there are some serious potential dangers, so let us start with a simple example without a driving Wiener process. Let us thus consider a counting process N, a real number x_0, and two real valued functions $\mu : R \to R$ and $\beta : R \to R$.

A first question is now to investigate under what conditions on μ and β the SDE

$$\left\{\begin{array}{rcl} dX_t &=& \mu(X_t)dt + \beta(X_t)dN_t, \\ X_0 &=& x_0, \end{array}\right. \tag{2.18}$$

has an adapted cadlag solution. A very natural, but naive, conjecture is that (2.18)-(**??**) will always possess a solution, as long as μ and β are n“nice enough”(such as for example Lipschitz and linear growth). This, however, is **wrong**, and it is very important to understand the following fact.

The SDE (2.18) is fundamentally ill posed.

To understand why this is so, let us consider the dynamics of X at a jump time t of the counting process N. Suppose therefore that N has a jump at time t. The X dynamics then says that

$$\Delta X_t = X_t - X_{t-} = \beta(X_t)dN_t = \beta(X_t), \tag{2.19}$$

which we can write as

$$X_t = X_{t-} + \beta(X_t). \tag{2.20}$$

The problem with this formula is that it describes a **non-causal** dynamic. The natural way of modeling the X dynamics is of course to model it as being generated "causally" by the N process, in the sense that at a jump time t, the jump size ΔX_t should be uniquely determined by X_{t-} and by dN_t. In (2.19) however, we see that if we are standing at $t-$, the jump size ΔX_t , is determined by X_t i.e. by the value of X **after** the jump. In particular we see that at a jump time t, the value of X_t (given X_{t-}) is being implicitly determined by the non linear equation (2.20).

By writing down a seemingly innocent expression like (2.18), one may in fact easily end up with completely nonsensical equations for which there is no solution. Consider for example the simple case when $\alpha \equiv 0$, $\beta(x) = x$, and $x_0 = 1$. We then have the SDE

$$\begin{cases} dX_t &= X_t dN_t, \\ X_0 &= 1. \end{cases}$$

This does not look particularly strange, but at a jump time t, equation (2.20) will now have the form

$$X_t = X_{t-} + X_t,$$

which implies that

$$X_{t-} = 0.$$

This however, is inconsistent with the initial condition $X_0 = 1$ (why?) so the SDE does not have a solution.

From the discussion above it should be clear that the correct way of writing an SDE driven by a counting process is to formulate it as

$$\begin{cases} dX_t &= \mu(X_{t-})dt + \beta(X_{t-})dN_t, \\ X_0 &= x_0, \end{cases}$$

where of course $\mu(X_{t-})$ can be replaced by $\mu(X_t)$ in the dt term. In fact, we have the following result.

Proposition 2.6.1 *Assume that the ODE*

$$\begin{cases} \frac{dX_t}{dt} &= \mu(X_t), \\ X_0 &= x_0, \end{cases}$$

has a unique global solution for every choice of x_0 and let $\beta : R \to R$ be an arbitrarily chosen function. Then the SDE

$$\begin{cases} dX_t &= \mu(X_{t-})dt + \beta(X_{t-})dN_t, \\ X_0 &= x_0, \end{cases}$$

has a unique global solution.

Proof. We have the following concrete algorithm.

1. Denote the jump times of N by $T_1, T_2, \ldots$.
2. For every fixed ω, solve the ODE

$$\begin{cases} \frac{dX_t}{dt} &= \mu(X_t), \\ X_0 &= x_0, \end{cases}$$

on the half open interval $[0, T_1)$. In particular we have now determined the value of X_{T_1-}.

3. Calculate the value of X_{T_1} by the formula

$$X_{T_1} = X_{T_1-} + \beta(X_{T_1-}).$$

4. Given X_{T_1} from the previous step, solve the ODE

$$\frac{dX_t}{dt} = \mu(X_t)$$

on the interval $[T_1, T_2)$. This will give us X_{T_2-}.

5. Compute X_{T_2} by the formula

$$X_{T_2} = X_{T_2-} + \beta(X_{T_2-}).$$

6. Continue by induction. ∎

We illustrate this methodology by solving a concrete SDE, namely the counting process analogue to geometrical Brownian motion

$$\begin{cases} dX_t &= \alpha X_{t-}dt + \beta X_{t-}dN_t, \\ X_0 &= x_0, \end{cases} \tag{2.21}$$

where α and β are real numbers.

To solve this SDE we note that up to the first jump time T_1 we have the ODE

$$\begin{cases} \frac{dX_t}{dt} &= \alpha X_t, \\ X_0 &= x_0, \end{cases}$$

with the exponential solution

$$X_t = e^{\alpha t} x_0,$$

so in particular we have $X_{T_1-} = e^{\alpha T_1} x_0$. The jump size at T_1 is given by

$$\Delta X_{T_1} = \beta X_{T_1-},$$

so we have

$$X_{T_1} = X_{T_1-} + \beta X_{T_1-} = (1+\beta)X_{T_1-} = (1+\beta)e^{\alpha T_1}x_0.$$

We now solve the ODE

$$\begin{cases} \frac{dX_t}{dt} &= \alpha X_t, \\ X_{T_1} &= (1+\beta)e^{\alpha T_1}x_0, \end{cases}$$

on the interval $[T_1, T_2)$ to obtain

$$X_t = e^{\alpha(t-T_1)}(1+\beta)e^{\alpha T_1}x_0 = e^{\alpha t}(1+\beta)x_0$$

and in particular

$$X_{T_2-} = (1+\beta)e^{\alpha T_2}x_0.$$

As before, the the jump condition gives us

$$X_{T_2} = X_{T_2-} + \beta X_{T_2-} = (1+\beta)X_{T_2-} = (1+\beta)^2 e^{\alpha T_2}x_0.$$

Continuing in this way we see that the solution is given by the formula

$$X_t = x_0(1+\beta)^{N_t}e^{\alpha t}.$$

We may in fact generalize this result as follows.

Proposition 2.6.2 *Assume that X satisfies the SDE*

$$\begin{cases} dX_t &= \alpha_t X_{t-}dt + \beta_t X_{t-}dN_t, \\ X_0 &= x_0, \end{cases}$$

where α and β are predictable processes. Then X can be represented as

$$X_t = x_0 e^{\int_0^t \alpha_s ds} \prod_{T_n \le t} (1+\beta_{T_n}),$$

or equivalently as

$$X_t = x_0 e^{\int_o^t \alpha_s ds + \int_0^t \ln(1+\beta_s)dN_s}.$$

2.7 The Watanabe Theorem

The object of this section is to prove the Watanabe Characterization Theorem for the Poisson process. Before we do this, we make a slight extension of the definition of a Poisson process.

Definition 2.7.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a given filtration $\mathbf{F} = \{\mathcal{F}_t\}_{t\ge 0}$, and let $t \to \lambda_t$ be a deterministic function of time. A counting process N is a* **Poisson process with intensity function λ with respect to the filtration F** *if it satisfies the following conditions.*

1. *N is adapted to $\mathbf{F}$.*
2. *For all $s \leq t$ the random variable $N_t - N_s$ is independent of $\mathcal{F}_s$.*
3. *For all $s \leq t$, the conditional distribution of the increment $N_t - N_s$ is given by*

$$P\left(N_t - N_s = n \left| \mathcal{F}_s \right.\right) = e^{-\Lambda_{s,t}} \frac{(\Lambda_{s,t})^n}{n!}, \quad n = 0, 1, 2, \dots \tag{2.22}$$

where

$$\Lambda_{s,t} = \int_s^t \lambda_u du. \tag{2.23}$$

We have the following easy result concerning the characteristic function for a Poisson process.

Lemma 2.7.1 *For a Poisson process as above, the following hold for all $s < t$*

$$E\left[e^{iu(N_t - N_s} \middle| \mathcal{F}_s \right] = e^{\Lambda_{s,t}\left(e^{iu}-1\right)} \tag{2.24}$$

With the definition above it is easy to see that the process X defined by

$$N_t - \int_0^T \lambda_s ds,$$

is an $\mathbf{F}$ martingale. The Watanabe Theorem says that this martingale property of X is not only a consequence of, but in fact characterizes the Poisson process among the class of counting processes.

Theorem 2.7.1 (The Watanabe Characterization) *Assume that N is a counting process and that $t \to \lambda_t$ is a deterministic function. Assume furthermore that the process M, defined by*

$$M_t = N_t - \int_0^t \lambda_s ds, \tag{2.25}$$

is an $\mathbf{F}$ martingale. Then N is Poisson w.r.t. $\mathbf{F}$ with intensity function λ.

Proof. Using a slight extension of Proposition 1.1.1, it is enough to show that

$$E\left[e^{iu(N_t - N_s)} \middle| \mathcal{F}_s \right] = \exp\left\{ \Lambda_{s,t}\left(e^{iu} - 1\right) \right\}, \tag{2.26}$$

with Λ as in (2.23). We start by proving this for the simpler case when $s = 0$. We thus want to show that

$$E\left[e^{iuN_t} \right] = \exp\left\{ \Lambda_{0,t}\left(e^{iu} - 1\right) \right\},$$

It is now natural to define the process Z by

$$Z_t = e^{iuN_t},$$

and an application of the Itô formula of Proposition 2.5.1 immediately gives us

$$dZ_t = \left\{ e^{iu(N_{t-}+1)} - e^{iuN_{t-}} \right\} dN_t = e^{iuN_{t-}} \left\{ e^{iu} - 1 \right\} dN_t.$$

We now use the relation

$$dN_t = \lambda_t dt + dM_t,$$

where the martingale M is defined by (2.25) to obtain

$$dZ_t = Z_{t-} \left\{ e^{iu} - 1 \right\} \lambda_t dt + Z_{t-} \left\{ e^{iu} - 1 \right\} dM_t.$$

Integrating this over $[0, t]$ we obtain (using the fact that $Z_0 = 1$)

$$Z_t = 1 + \left\{ e^{iu} - 1 \right\} \int_0^t Z_{s-} \lambda_s ds + \left\{ e^{iu} - 1 \right\} \int_0^t Z_{s-} \lambda_s dM_s.$$

Since M is a martingale and the integrand Z_{s-} is predictable (left continuity!) the dM integral is also a martingale so, after taking expectations, we obtain

$$E\left[Z_t\right] = 1 + \left\{ e^{iu} - 1 \right\} \int_0^t E\left[Z_{s-}\right] \lambda_s ds.$$

Let us now, for a fixed u, define the deterministic function y by

$$y_t = E\left[e^{iuN_t}\right] = E\left[Z_t\right].$$

We thus have

$$y_t = 1 + \left\{ e^{iu} - 1 \right\} \int_0^t y_{s-} \lambda_s ds,$$

and since we are integrating over Lebesgue measure we can (why?) write this as

$$y_t = 1 + \left\{ e^{iu} - 1 \right\} \int_0^t y_s \lambda_s ds.$$

Taking the derivative w.r.t t we obtain the ODE

$$\begin{cases} \dfrac{dy_t}{dt} &= y_t \left(e^{iu} - 1 \right) \lambda_t, \\ y_0 &= 1, \end{cases}$$

with the solution

$$y_t = e^{\left(e^{iu}-1\right) \int_0^t \lambda_s ds}.$$

This proves (2.26) for the special case when $s = 0$. For the general case, it is clearly enough to show (why?) that

$$E\left[I_A e^{iu(N_t - N_s)}\right] = E\left[I_A\right] \exp\left\{ \Lambda_{s,t} \left(e^{iu} - 1 \right) \right\},$$

for every event $A \in \mathcal{F}_s$. To do this we now define, for fixed s, u and A, the process Z on the time interval $[s, \infty)$ by

$$Z_t = I_A e^{iu(N_t - N_s)},$$

and basically copy the argument above. ∎

2.8 Exercises

Exercise 2.1 *Show that the SDE*

$$\begin{cases} dX_t &= aX_t dt + \beta dN_t \\ X_0 &= x_0 \end{cases}$$

where a, β, and x_0 are real numbers, and N is a counting process, has the solution

$$X_t = e^{at}x_0 + \beta \int_0^t e^{a(t-s)} dN_s$$

Exercise 2.2 *Consider the SDE of the previous exercise, and assume that N is Poisson with constant intensity λ. Compute $E[X_t]$.*

Exercise 2.3 *Consider the following SDEs, where N^x and N^y are counting processes without common jumps, and where the parameters $\alpha_X, \alpha_Y, \beta_X, \beta_Y$ are known constants.*

$$\begin{aligned} dX_t &= \alpha_X X_t dt + \beta_X X_{t-} dN_t^x, \\ dY_t &= \alpha_Y Y_t dt + \beta_Y Y_{t-} dN_t^y, \end{aligned}$$

Define the process Z by $Z_t = X_t Y_t$. Then Z will satisfy an SDE. Find this SDE, and compute $E[Z_t]$ in the case when N^x and N^y are Poisson with intensities λ_x and λ_y.

Exercise 2.4 *Consider the SDEs of the previous exercise. Define the process Z by $Z_t = X_t/Y_t$. Then Z will satisfy an SDE. Find this SDE, and compute $E[Z_t]$ in the case when N^x and N^y are Poisson with intensities λ_x and λ_y.*

Exercise 2.5 *Consider two discrete time processes X and Y. Prove the product formula*

$$\Delta(XY)_n = X_{n-1}\Delta Y_n + Y_{n-1}\Delta X_n + \Delta X_n \Delta Y_n.$$

Exercise 2.6 *Consider two continuous time processes X and Y which are both of bounded variation (i.e. they have no driving Wiener process). Use the Itô formula to prove the product formula*

$$d(XY)_t = X_{t-} dY_t + Y_{t-} dX_t + \Delta X_t \Delta Y_t.$$

As usual, $\Delta X_n = X_n - X_{n-1}$ etc.

Chapter 3

Counting Processes with Stochastic Intensities

In this section we will generalize the concept of an intensity from the Poisson case to the case of a fairly general counting process. We consider a filtered probability space $(\Omega, \mathcal{F}, P, \mathbf{F})$ carrying an optional counting process N.

3.1 Definition of stochastic intensity

Definition 3.1.1 *Consider an optional counting process N on the filtered space $(\Omega, \mathcal{F}, P, \mathbf{F})$. Let λ be a non negative optional random process such that*

$$\int_0^t \lambda_s ds < \infty, \quad \textit{for all } t \geq 0. \tag{3.1}$$

If the condition

$$E\left[\int_0^\infty h_t \lambda_t dt\right] = E\left[\int_0^\infty h_t dN_t\right] \tag{3.2}$$

for every non negative predictable process h, then we say that N has the **F-intensity** λ.

At first sight, this definition may look rather forbidding, but the intuitive interpretation is that, modulo integrability, it says that the difference $dN_t - \lambda_t dt$ is a martingale increment. This is clear from the following result.

Proposition 3.1.1 *Assume that N has the* **F** *intensity λ and that N is* **integrable**, *in the sense that*

$$E[N_t] < \infty, \quad \textit{for all } t \geq 0. \tag{3.3}$$

Then the process M defined by

$$M_t = N_t - \int_0^t \lambda_s ds$$

is an **F** *martingale.*

Proof. Fix s and t with $s < t$, and choose an arbitrary event $A \in \mathcal{F}_s$. If we now define the process h by

$$h_u(\omega) = I_A(\omega) I\left\{s < u \leq t\right\},$$

then h is non negative and predictable (why?). With this choice of h, the relation (3.2) becomes

$$E\left[I_A \int_s^t \lambda_u du\right] = E\left[I_A\left(N_t - N_s\right)\right].$$

Because of (3.3) we may now subtract the left hand side from the right hand side without any risk of expressions of the type $+\infty - \infty$. The result is

$$E\left[I_A\left(M_t - M_s\right)\right] = 0,$$

which shows that M is a martingale. ∎

Remark 3.1.1 *The reason why we define the intensity concept by the condition (3.2), rather than by the martingale property of M above, is that (3.2) also covers the case when $E\left[N_t\right] = \infty$.*

We now have a number of obvious questions to answer.

- Does every counting process have an intensity?
- Is the intensity unique?
- How does the intensity depend on the filtration **F**?
- What is the intuitive interpretation of λ?

3.2 Existence

The existence of an intensity is a technically non trivial problem which is outside the scope of this text. Roughly speaking, the story is as follows.

For every counting process N there will always exist an increasing predictable process A, called the **compensator** process with the property that the process M defined by

$$M_t = N_t - A_t,$$

is a martingale. This is in fact a special case of a very general results known as the "Doob-Meyer decomposition of a submartingale of class D".

If we **assume** that the compensator A is absolutely continuous w.r.t. Lebesgue measure, then we can write A as

$$A_t = \int_0^y \lambda_s ds,$$

for some process λ, and this λ is of course our intensity. We thus see that only those counting processes for which the compensator is absolutely continuous will possess an intensity. Furthermore one can show that if a counting process has an intensity, then the distribution of every jump time will have a density w.r.t Lebesgue measure. This implies that if we restrict ourselves (as we will do for the rest of the text) to counting processes with intensities then we are basically excluding counting processes with jumps at predetermined points in time.

3.3 Uniqueness

From Definition 3.1.1 is should be clear that we can **not** expect the intensity process λ to be unique. Suppose for example that λ is an intensity for N and that λ is cadlag. If we now define the process μ by

$$\mu_t = \lambda_{t-},$$

then it is clear that

$$E\left[\int_0^\infty h_t \lambda_t dt\right] = E\left[\int_0^\infty h_t \mu_t dt\right]$$

so μ is also an intensity. If, however, we require **predictability**, then we have uniqueness.

Proposition 3.3.1 *Assume that N has an* **F** *intensity λ^*. The N will also possess an* **F predictable** *intensity λ. Furthermore, λ is unique in the sense that if μ is another predictable intensity, then we have*

$$\mu_t(\omega) = \lambda_t(\omega), \quad dPdN_t - a.e.$$

Proof. The formal proof is rather technical and left out. The intuitive idea behind the proof is however very easy to understand. We simply define the process λ by the prescription

$$\lambda_t = E\left[\lambda_t^* \middle| \mathcal{F}_{t-}\right],$$

and since λ_t is clearly $\mathcal{F}_{t-}$-measurable for each t, we see that λ is predictable, thus proving the existence of a predictable intensity.

This is in fact, where the formal proof gets technical since λ defined above is not really defined as a bona fide random process. Instead we have defined λ_t as an equivalence class of random variables for each t, and the problem is to show that we can choose one member of each equivalence class and "glue" these together in such a way as to obtain a predictable process.

To show uniqueness, it enough to show that for every predictable non negative process h we have

$$E\left[\int_0^\infty h_t \lambda_t dN_t\right] = E\left[\int_0^\infty h_t \mu_t dN_t\right].$$

If, in the left hand side, we use the assumption that N has the intensity μ and on the right hand side use the fact that N has the intensity λ we see that both sides equal

$$E\left[\int_0^\infty h_t \mu_t \lambda dt\right]. \quad \blacksquare$$

3.4 Interpretation

We now go on the intuitive interpretation of the intensity concept. Let us thus assume that N has the predictable intensity process λ. Modulo integrability, this implies that

$$dN_t - \lambda dt$$

is a martingale increment, and heuristically we will thus have

$$E\left[\left. dN_t - \lambda dt \right| \mathcal{F}_{t-}\right] = 0.$$

Since λ is predictable we have $\lambda_t \in \mathcal{F}_{t-}$ so we can move $\lambda_t dt$ outside the expectation and obtain

$$E\left[\left. dN_t \right| \mathcal{F}_{t-}\right] = \lambda_t dt.$$

We thus see that the predictable intensity λ has the interpretation that λ_t is the conditionally expected number of jumps per unit of time. Since we know that the predictable intensity is unique, we can summarize the moral of this section so far in the following slogan:

> The natural description of the dynamics for a counting process N is in terms of its **predictable intensity** λ, with the interpretation
>
> $$E\left[\left. dN_t \right| \mathcal{F}_{t-}\right] = \lambda_t dt. \tag{3.4}$$

3.5 Dependence on the filtration

It is important to note that the intensity concept is tied to a particular choice of filtration. If we have two different filtrations $\mathbf{F}$ and $\mathbf{G}$, and a counting process which is optional w.r.t. to both $\mathbf{F}$ and $\mathbf{G}$, then there is no reason to believe that the $\mathbf{F}$ intensity $\lambda^{\mathbf{F}}$ will coincide with the $\mathbf{G}$ intensity $\lambda^{\mathbf{G}}$. In the general case there are no interesting relations between $\lambda^{\mathbf{F}}$ and $\lambda^{\mathbf{G}}$, but in the special case when $\mathbf{G}$ is a sub filtration of $\mathbf{F}$, we have a very precise result.

Proposition 3.5.1 *Assume that N has the predictable $\mathbf{F}$ intensity $\lambda^{\mathbf{F}}$, and assume that we are given a filtration $\mathbf{G}$ such that*

$$\mathcal{G}_t \subseteq \mathcal{F}_t, \quad \textit{for all } t \geq 0.$$

Then there exists a predictable $\mathbf{G}$ intensity $\lambda^{\mathbf{G}}$ with the property that

$$\lambda_t^{\mathbf{G}} = E\left[\left. \lambda_t^{\mathbf{F}} \right| \mathcal{G}_{t-}\right].$$

Proof. Using the intuitive interpretation (3.4) the result follows at once from the calculation

$$\lambda_t^{\mathbf{G}} = E\left[dN_t \middle| \mathcal{G}_{t-}\right] = E\left[E\left[dN_t \middle| \mathcal{F}_{t-}\right] \middle| \mathcal{G}_{t-}\right] = E\left[\lambda_t^{\mathbf{F}} \middle| \mathcal{G}_{t-}\right].$$

A more formal proof is as follows. Let h be an arbitrary non negative $\mathbf{G}$ predictable process. Then h will also be $\mathbf{F}$ predictable (why?) and we have

$$\begin{aligned} E\left[\int_0^\infty h_t dN_t\right] &= E\left[\int_0^\infty h_t \lambda_t^{\mathbf{F}} dt\right] = E\left[\int_0^\infty E\left[h_t \lambda_t^{\mathbf{F}} \middle| \mathcal{G}_{t-}\right] dt\right] \\ &= E\left[\int_0^\infty h_t E\left[\lambda_t^{\mathbf{F}} \middle| \mathcal{G}_{t-}\right] dt\right] = E\left[\int_0^\infty h_t \lambda_t^{\mathbf{G}} dt\right], \end{aligned}$$

which shows that $\lambda^{\mathbf{G}}$ is the predictable $\mathbf{G}$ intensity of N. ▮

Chapter 4

Martingale Representation

In the next two chapters we present the two main theoretical workhorses for counting process theory: the Martingale Representation Theorem and the Girsanov Theorem. These results will be used over and over again in connection with general counting process theory, and they are fundamental for the analysis of arbitrage free capital markets.

4.1 The Martingale Representation Theorem

Assume that we are given a filtered space $(\Omega, \mathcal{F}, P, \mathbf{F})$ carrying an integrable adapted point process N with $\mathbf{F}$-intensity λ. From Propositions 2.4.2 and 3.1.1 we know that, for every choice of a predictable (and sufficiently integrable) process h the process X defined by

$$X_t = \int_0^T h_s \left[dN_s - \lambda_s ds \right] \tag{4.1}$$

will be an $\mathbf{F}$-martingale. An interesting question is now to ask whether also the converse statement also is true, i.e. to ask if **every** $\mathbf{F}$-martingale X can be represented on the form (4.1). That this cannot possible be the case is clear from the following counter example.

Assume for simplicity that N is Poisson with constant intensity, and assume that the space also carries an independent $\mathbf{F}$-Wiener process W. Then, setting $X = W$, it is clear that X is an $\mathbf{F}$-martingale, but it is also clear that X cannot have the representation (4.1). The reason is of course that X has continuous trajectories, whereas a stochastic integral w.r.t. the compensated N process, will have trajectories with jumps. The more informal reason is of course that the Wiener process W "has nothing at all to do with the point process N". In order to have any chance of obtaining a positive result we therefore have to guarantee that the space carries "nothing else than the process N itself". The natural condition is given in the following fundamental result, which is the point process analogue of the corresponding martingale representation result for Wiener processes.

Theorem 4.1.1 *Assume that N is an integrable point process with intensity λ, and that the filtration is the* **internal** *one, generated by N, i.e.*

$$\mathcal{F}_t = \mathcal{F}_t^N. \tag{4.2}$$

Then, for every $\mathbf{F}$-martingale X there will exist a predictable process h such that

$$X_t = X_0 + \int_0^T h_s \left[dN_s - \lambda_s ds \right]. \tag{4.3}$$

Furthermore, the process h is unique $dP(\omega)dN_t(\omega) -$ a.e..

Proof. This is a deep and difficult result and the reader is referred to the specialist literature for a proof. ∎

Remark 4.1.1 *We remark that this is an abstract existence result. There is generally no concrete characterization of the integrand h.*

The result above generalizes immediately to the multi dimensional setting, and we can also include a finite number of driving Wiener processes.

Theorem 4.1.2 *Let $(\Omega, \mathcal{F}, P, \mathbf{F})$ be a filtered probability space carrying k counting processes $N^1, \ldots, N^k$, as well as a standard d-dimensional Wiener process $W^1, \ldots, W^d$. Assume that the filtration $\mathbf{F}$ is the internal one, i.e.*

$$\mathcal{F}_t = \sigma \left\{ N_s^i, W_s^j;\ i = 1, \ldots, k,\ j = 1, \ldots, \ldots, d;\ s \le t \right\} \tag{4.4}$$

Assume furthermore that N^i has the predictable intensity λ^i for $i = 1, \ldots, k$. Then, for every F-martingale X, there will exist predictable processes $h^1, \ldots, h^k$ and $g^1, \ldots, g^d$ such that

$$X_t = X_0 + \sum_{i=1}^k h_s^i \left[dN_s^i - \lambda_s^i ds \right] + \sum_{j=1}^d \int_0^t g_s^j dW_s^j. \tag{4.5}$$

Chapter 5

Girsanov Transformations

In this chapter we will study how the intensity λ of a counting process N changes when we transform the original measure P to a new measure $Q << P$. As a result we will obtain a counting process version of the standard Girsanov Theorem for Wiener processes. At the end of the chapter we apply the theory to maximum likelihood estimation and we also define and prove the existence of Cox processes.

5.1 The Girsanov Theorem

To set the scene we consider a given filtered space $(\Omega, \mathcal{F}, P, \mathbf{F})$ carrying an adapted counting process N with $\mathbf{F}$-predictable P-intensity λ. We study the process N on a fixed time interval $[0, T]$.

Let us now assume that we change measure form P to Q, where $Q << P$ on $\mathcal{F}_T$, and let L be the induced likelihood process, given by

$$L_t = \frac{dQ}{dP}, \quad \text{on } \mathcal{F}_t, \quad 0 \le t \le T. \tag{5.1}$$

Under the new measure Q, the counting process N will have a predictable intensity $\lambda^Q \neq \lambda$ and our task is to derive an expression for the Q-intensity λ^Q.

In order to understand more clearly what is going on, we start with some heuristics. To this end we recall that the intuitive interpretation of λ^Q is given by the relation

$$\lambda_t^Q dt = E^Q \left[dN_t | \mathcal{F}_{t-} \right],$$

and in order to compute the expected value the obvious tool to use is the abstract Bayes formula. Using this and the fact that the likelihood process L is a P martingale we have

$$\lambda_t^Q dt \quad = \quad E^Q \left[dN_t | \mathcal{F}_{t-} \right] = \frac{E^P \left[L_t dN_t | \mathcal{F}_{t-} \right]}{E^P \left[L_t | \mathcal{F}_{t-} \right]}$$

$$
\begin{aligned}
&= \frac{E^P\left[\left(L_{t-}+dL_t\right)dN_t\right|\mathcal{F}_{t-}]}{L_{t-}} \\
&= E^P\left[dN_t\right|\mathcal{F}_{t-}] + \frac{E^P\left[dL_t dN_t\right|\mathcal{F}_{t-}]}{L_{t-}}.
\end{aligned}
$$

Recalling that $\lambda_t dt = E^P\left[dN_t\right|\mathcal{F}_{t-}]$ we thus obtain

$$
\lambda_t^Q dt = \lambda_t dt + \frac{E^P\left[dL_t dN_t\right|\mathcal{F}_{t-}]}{L_{t-}}. \tag{5.2}
$$

At this degree of generality we are not able to go further but, as in the Wiener case, the above expression will simplify considerably if we make some further assumptions about L. Since we know that L is a P martingale, and since also the compensated process

$$
dN_t - \lambda_t dt
$$

is a martingale increment under P, it natural to investigate what will happen if we assume that L has the particular structure

$$
dL_t = g_t\left[dN_t - \lambda_t dt\right],
$$

where g is a predictable process. Using the facts that $dN_t dt = 0$, $\left(dN_t\right)^2 = dN_t$, and that $g_t \in \mathcal{F}_{t-}$ (why?), we get

$$
\begin{aligned}
\frac{E^P\left[dL_t dN_t\right|\mathcal{F}_{t-}]}{L_{t-}} &= \frac{E^P\left[g_t\left\{dN_t - \lambda_t dt\right\}dN_t\right|\mathcal{F}_{t-}]}{L_{t-}} \\
&= g_t\frac{E^P\left[\left.\left(dN_t\right)^2\right|\mathcal{F}_{t-}\right]}{L_{t-}} = g_t\frac{E^P\left[dN_t\right|\mathcal{F}_{t-}]}{L_{t-}} \\
&= g_t\frac{\lambda_t dt}{L_{t-}}.
\end{aligned}
$$

From this we see that if we now define the predictable process h by

$$
h_t = \frac{g_t}{L_{t-}},
$$

the above expression simplifies to

$$
\frac{E^P\left[dL_t dN_t\right|\mathcal{F}_{t-}]}{L_{t-}} = h_t \lambda_t dt,
$$

and if we plug this into (5.2) we obtain

$$
\lambda_t^Q dt = \lambda_t dt + h_t \lambda_t dt = \lambda_t\left(1 + h_t\right)dt,
$$

or

$$
\lambda_t^Q = \lambda_t\left(1 + h_t\right).
$$

The point of these rather informal calculations is that we are able to guess what the Girsanov Theorem will look like for a counting process. We can now state and prove the formal result.

Theorem 5.1.1 (The Girsanov Theorem) *Let N be an optional counting process on the filtered space $(\Omega, \mathcal{F}, P, \mathbf{F})$ and assume that N has the predictable intensity λ. Let h be a predictable process such that*

$$h_t \geq -1, \quad P - a.s. \tag{5.3}$$

and define the process L by

$$\begin{cases} dL_t &= L_{t-} h_t \{dN_t - \lambda_t dt\}, \\ L_0 &= 1, \end{cases} \tag{5.4}$$

on the interval $[0, T]$. Assume furthermore that

$$E^P[L_T] = 1. \tag{5.5}$$

Now define a new probability measure $Q << P$ on $\mathcal{F}_T$ by

$$dQ = L_T dP. \tag{5.6}$$

Then N has the Q intensity λ^Q , given by

$$\lambda_t^Q = \lambda_t (1 + h_t).$$

Proof. We need to show that, for every non negative predictable process g, we have

$$E^Q\left[\int_0^T g_t dN_t\right] = E^Q\left[\int_0^T g_t \lambda_t (1 + h_t) dt\right]. \tag{5.7}$$

We start with the right hand side to obtain

$$\begin{aligned} E^Q\left[\int_0^T g_t \lambda_t (1 + h_t) dt\right] &= \int_0^T E^Q[g_t \lambda_t (1 + h_t)] dt \\ = \int_0^T E^P[L_t g_t \lambda_t (1 + h_t)] dt &= E^P\left[\int_0^T L_t g_t \lambda_t (1 + h_t) dt\right] \\ = E^P\left[\int_0^T L_{t-} g_t \lambda_t (1 + h_t) dt\right] & \end{aligned} \tag{5.8}$$

Turning to the left hand side of (5.7) we obtain

$$E^Q\left[\int_0^T g_t dN_t\right] = E^P\left[L_T \int_0^T g_t dN_t\right],$$

and it is therefore natural to study the process Z, defined by

$$Z_t = L_t Y_t = L_t \int_0^t g_s dN_s.$$

It is clear that Z is the product of two processes of bounded variation, so from the product rule (Proposition 2.5.2) we have

$$\begin{aligned} dZ_t &= L_{t-}dY_t + Y_{t-}dL_t + \Delta L_t \Delta Y_t \\ &= L_{t-}g_t dN_t + Y_{t-}dL_t + L_{t-}g_t h_t dN_t \\ &= L_{t-}g_t\left(1+h_t\right)dN_t + Y_{t-}dL_t. \end{aligned}$$

Integrating this, and recalling that λ is the P intensity of N, gives us

$$\begin{aligned} E^Q\left[\int_0^T g_t dN_t\right] &= E^P\left[L_T \int_0^T g_t dN_t\right] \\ &= E^P\left[\int_0^T L_{t-}g_t\left(1+h_t\right)dN_t\right] + E^P\left[\int_0^T Y_{t-}dL_t\right] \\ &= E^P\left[\int_0^T L_{t-}g_t\left(1+h_t\right)\lambda_t dt\right], \end{aligned} \tag{5.9}$$

where we have also used the fact that, since L is a P martingale and the process $t \to Y_{t-}$ is predictable, the dL integral has zero expected value. The equality (5.7) now follows from (5.8) and (5.9). ∎

This result generalizes easily to the multi dimensional case, and we can also include a finite number of driving Wiener processes in the obvious way.

Theorem 5.1.2 (The Girsanov Theorem) *Consider the filtered probability space $(\Omega, \mathcal{F}, P, \mathbf{F})$ and assume that $N^1, \ldots, N^k$ are optional counting processes with predictable intensities $\lambda^1, \ldots, \lambda^k$. Assume furthermore that $W^1, \ldots, W^d$ are standard independent $(\mathbf{F}, P)$-Wiener processes. Let $h^1, \ldots, h^k$ be predictable processes with*

$$h_t^i < -1, \quad i = 1, \ldots, k, \quad P - a.s,$$

and let $g^1, \ldots, g^d$ be optional processes. Define the process L on $[0, T]$ by

$$\begin{cases} dL_t = L_t \sum_{i=1}^d g_t^i dW_t^i + L_{t-} \sum_{j=1}^k h_t \left\{dN_t^i - \lambda_t^i dt\right\}, \\ L_0 = 1, \end{cases} \tag{5.10}$$

and assume that

$$E^P\left[L_T\right] = 1.$$

Define the measure Q on $\mathcal{F}_T$ by $dQ = L_T dP$. Then the following hold

- *We can write*

  $$dW_t^i = g_t^i dt + dW_t^{Q,i}, \quad i = 1, \ldots, d,$$

  where $W^{Q,1}, \ldots, W^{Q,d}$ are Q Wiener processes.

- *The Q intensities of $N^1, \ldots N^k$ are given by*

  $$\lambda_t^{Q,i} = \lambda_t^i\left(1 + h_t^i\right), \quad i = 1, \ldots, k.$$

5.2 The converse of the Girsanov Theorem

If we start with a measure P and perform a Girsanov transformation according to (5.4)-(5.6) in order to define a new measure Q, then we know that $Q << P$. A natural question to ask is whether **all** measures $Q << P$ are obtained by the procedure (5.4)-(5.6).

In the general case it is obvious that the answer is no. Consider for example a case where the stochastic basis carries a Poisson process N with constant intensity λ as well as $N[0,1]$ distributed random variable Z, which is independent of N. Suppose furthermore that we change P to Q by changing the distribution of Z from $N[0,1]$ to $N[5,1]$, while keeping the distribution fixed for N. It is then obvious that $Q \sim P$, but since the Girsanov transformation (5.4)-(5.6) is completely determined by N, and not in any way involving Z, it is intuitively obvious that the change form P to Q can not be achieved by a Girsanov transformation of the type (5.4)-(5.6).

From this discussion it is reasonable to assume that if we restrict ourselves to the case when the filtration $\mathbf{F}$ is the **internal** one, generated by the counting process N, then we may hope for a converse of the Girsanov Theorem.

Proposition 5.2.1 (The Converse of the Girsanov Theorem) *Let N be a counting process on $(\Omega, \mathcal{F}, P, \mathbf{F})$ with intensity process λ, and assume that the filtration $\mathbf{F}$ is the* **internal** *one, i.e. that*

$$\mathcal{F}_t = \mathcal{F}_t^N, \quad t \geq 0. \tag{5.11}$$

Assume furthermore that there exists a measure Q such that for a fixed T, we have $Q << P$ on $\mathcal{F}_T$, and let L denote the corresponding likelihood process, i.e.

$$L_t = \frac{dQ}{dP}, \quad on\ \mathcal{F}_t, \quad 0 \leq t \leq T$$

Then there exists a predictable process h such that L has the dynamics

$$\begin{cases} dL_t &= L_{t-} h_t \left\{ dN_t - \lambda_t dt \right\}, \\ L_0 &= 1, \end{cases} \tag{5.12}$$

and the Q intensity is given by

$$\lambda_t^Q = \lambda_t (1 + h_t).$$

Proof. Defining L as a above, we know from general theory that L is a P martingale. Since we have the internal filtration, the Martingale Representation Theorem 4.1.1 guarantees the existence of a predictable process g such that

$$dL_t = g_t \left\{ dN_t - \lambda_t dt \right\},$$

and if we define h by

$$h_t = \frac{g_t}{L_{t-}},$$

we are done. There is a potential problem when $L_{t-} = 0$, but also this can be handled. ∎

We can of course extend this result to the case of a multidimensional counting process and a multi dimensional Wiener process. The proof is almost identical.

Proposition 5.2.2 (The Converse of the Girsanov Theorem) *Consider a filtered space $(\Omega, \mathcal{F}, P, \mathbf{F})$ carrying the counting processes $N^1, \ldots, N^k$ with predictable intensities $\lambda^1, \ldots, \lambda^k$, as well as the standard independent Wiener processes $W^1, \ldots, W^d$. We assume that the filtration is the* **internal** *one , generated by N and W, i.e.*

$$\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{F}_t^W.$$

Assume furthermore that there exists a measure Q such that for a fixed T, we have $Q << P$ on $\mathcal{F}_T$, and let L denote the corresponding likelihood process, i.e.

$$L_t = \frac{dQ}{dP}, \quad on\ \mathcal{F}_t, \quad 0 \le t \le T$$

Then there exists a k-dimensional predictable process h, and a d-dimensional predictable process g such that L has the dynamics

$$\left\{ \begin{array}{rcl} dL_t & = & L_{t-} \sum_{i=1}^k h_s^i \left[dN_s^i - \lambda_s^i ds \right] + L_t \sum_{j=1}^d \int_0^t g_s^j dW_s^j \\ L_0 & = & 1, \end{array} \right. \tag{5.13}$$

5.3 Maximum Likelihood Estimation

In this section we give a brief introduction to maximum likelihood (ML) estimation for counting processes.

We need the concept of a statistical model.

Definition 5.3.1 *A dynamic* **statistical model** *over a finite time interval $[0, T]$ consists of the following objects.*

- *A measurable space $(\Omega, \mathcal{F})$.*
- *A filtration $\mathbf{F}$.*
- *An indexed family of probability measures $\{P_\alpha;\ \alpha \in A\}$, defined on the space $(\Omega, \mathcal{F})$, where A is some index set and where all measures are assumed to be absolutely continuous on $\mathcal{F}_T$ w.r.t. some base measure P_{α_0}, i.e.*

$$P_\alpha << P_{\alpha_0}, \quad for\ all\ \alpha \in A$$

In most concrete applications (see examples below) the parameter α will be a real number or a finite dimensional vector, i.e. A will be the real line or some finite dimensional Euclidean space. The filtration will typically be generated by some observation process X.

The interpretation of all this is that the probability distribution is governed by some measure P_α, but we do no know which. We do have, however, access to a flow of information over time, and this is formalized by the filtration above, so at time t we have the information contained in $\mathcal{F}_t$. Our problem is to try to estimate α given this flow of observations, or more precisely: for every t we want an estimate α_t of α, based upon the information contained in $\mathcal{F}_t$, i.e. based on the observations over the time interval $[0, t]$. The last requirement is formalized by requiring that the estimation process should be adapted to $\mathbf{F}$, i.e. that $\alpha_t \in \mathcal{F}_t$.

One of the most common techniques used in this context is that of finding, for each t, the **maximum likelihood** estimate of α. Formally the procedure works as follows.

- Compute, for each α the corresponding Likelihood process L^α (where α only has the role of being an index, and not a power) is defined by

$$L_t^\alpha = \frac{dP_\alpha}{dP_{\alpha_0}}, \quad \text{on } \mathcal{F}_t.$$

- For each fixed t, find the value of α which maximizes the likelihood ratio L_t^α.

- The optimal α is denoted by $\widehat{\alpha}_t$ and is called the **maximum likelihood estimate** of α based on the information gathered over $[0, t]$.

As the simplest possible example let us consider the problem of estimating the constant but unknown intensity of a scalar Poisson process.

In this example we do in fact have an obvious candidate for the intensity estimate. Indeed, if N is Poisson with intensity λ, then λ is the mean number of jumps per unit time, so the natural estimate is given by

$$\widehat{\lambda}_t = \frac{N_t}{t}, \quad t > 0.$$

To formalize our problem within the more abstract framework above, we need to build a statistical model. To this end we consider a filtered space $(\Omega, \mathcal{F}, P, \mathbf{F})$ carrying a Poisson process N with **unit intensity** under P. The filtration is assumed to be the internal one, i.e. $\mathcal{F}_t = \mathcal{F}_t^N$. For any non negative real number λ we can now define the measure P_λ by the Girsanov transformation

$$\left\{ \begin{array}{rcl} dL_t^\lambda & = & L_t^\lambda (\lambda - 1) \{dN_t - dt\}, \\ L_0^\lambda & = & 1. \end{array} \right. \tag{5.14}$$

From the Girsanov Theorem, and the Watanabe Characterization Theorem it is clear that under P_λ, the process N will be Poisson with the constant intensity λ. The SDE (5.14) can easily be solved, for example by using Proposition 2.6.2, and we obtain

$$L_t^\lambda = e^{N_t \ln(\lambda) - t(\lambda - 1)}.$$

We thus have to maximize the expression

$$N_t \ln(\lambda) - t(\lambda - 1)$$

over $\lambda > 0$ and we immediately obtain the ML estimate as

$$\widehat{\lambda}_t = \frac{N_t}{t}. \tag{5.15}$$

We see that in this example the ML estimator actually coincides with our naive guess above. The point of using the ML technique is of course that in a more complicated situation (see the exercises) we may have no naive candidate, whereas the ML technique in principle is always applicable.

5.4 Cox Processes

In point process theory and its applications, such as for example in credit risk theory, a very important role is played by a particular class of counting processes known as "Cox processes", or "doubly stochastic Poisson processes". In this section we will define the Cox process concept and then use the Girsanov theory developed above to prove the existence of Cox processes.

The intuitive idea of a Cox process is very simple and goes roughly as follows.

1. Consider a fixed random process λ on some probability space Ω.
2. Fix one particular trajectory of λ, say the one corresponding to the outcome $\omega \in \Omega$.
3. For this fixed ω, the mapping $t \to \lambda_t(\omega)$ is a deterministic function of time.
4. Construct, again for this fixed ω, a counting process N which is Poisson with intensity function $\lambda_.(\omega)$.
5. Repeat this procedure for all choices of $\omega \in \Omega$.

If this informal procedure can be carried out (this is not at all clear), then it seems that it would produce a counting process N with the following property.

> Conditional on the entire λ-trajectory, the process N is Poisson with that particular λ-trajectory as intensity function. This is, intuitively, the definition of a Cox process.

The construction above is of course not very precise from a mathematical point of view, and this also holds for the statement concerning the properties of N. For example, what **exactly** do we mean by the sentence "Conditional on the entire λ-trajectory, the process N is Poisson with that particular λ-trajectory as intensity function"? We now have a small research program consisting of the following items.

- Define, in a mathematically precise way, the concept of a Cox process.
- Prove that, given an intensity process λ, the corresponding Cox processes exist.

The formal definition is as follows.

Definition 5.4.1 *Consider a probability space $(\Omega, \mathcal{F}, P)$, carrying a counting process N as well as a non negative process λ. We say that N is a* **Cox process** *with intensity process λ if the relation*

$$E\left[e^{iu(N_t-N_s)}\middle|\mathcal{F}_s^N \vee \mathcal{F}_\infty^\lambda\right] = e^{\Lambda_{s,t}\left(e^{iu}-1\right)} \tag{5.16}$$

holds for all $s < t$, where

$$\Lambda_{s,t} = \int_s^t \lambda_u du.$$

If we compare this with Definition 2.7.1 and Lemma 2.7.1 we see that the interpretation is indeed that, "conditional on the λ-trajectory the process N is Poisson with that particular λ-trajectory as intensity function".

We now go on to show that, for any given process λ, there actually exists a Cox process with λ as the intensity. The formal statement is given below, and the proof is a nice example of Girsanov technique.

Proposition 5.4.1 *Consider a probability space $(\Omega, \mathcal{F}, P_0)$, carrying a non negative random process λ. Assume that the space also carries a Poisson process N, with unit intensity, and that N is independent of λ. Then there exists a probability measure $P \sim P_0$ with the following properties.*

- *The distribution of λ under P is the same as the distribution under P_0.*
- *Under P, the counting process N is a Cox process with intensity λ.*

Proof. We start by defining the filtration $\mathbf{F}$ by

$$\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{F}_\infty^\lambda,$$

and note that this implies that $\mathcal{F}_\infty^\lambda \subseteq \mathcal{F}_0$. Next we define a likelihood process L in the obvious way by

$$\begin{cases} dL_t &= L_{t-}(\lambda_t - 1)\{dN_t - dt\}, \\ L_0 &= 1, \end{cases} \tag{5.17}$$

and define the new measure P by

$$L_t = \frac{dP}{dP_0} \quad \text{on } \mathcal{F}_t.$$

From the Girsanov Theorem it is clear that N has the intensity λ under P but this is not enough. Denoting, by $\mathcal{L}^Q(X)$ the distribution of a random variable or process X under a measure Q we have to show that

(a) $\mathcal{L}^P(\lambda) = \mathcal{L}^{P_0}(\lambda)$.

(b) Under P, the process N is Cox with intensity λ.

Item (a) is easy. Since $L_0 = 1$, the measures P_0 and P coincide on $\mathcal{F}_0$, and since $\sigma\{\lambda\} = \mathcal{F}_\infty^\lambda$ is included in $\mathcal{F}_0$, we conclude that $\mathcal{L}^P(\lambda) = \mathcal{L}^{P_0}(\lambda)$.

Item (b) requires a little bit of work. We need to show that

$$E^P\left[e^{iu(N_t - N_s)} \middle| \mathcal{F}_s \right] = e^{\Lambda_{s,t}\left(e^{iu}-1\right)}, \tag{5.18}$$

and the obvious idea is of course to use the Bayes formula. We then have

$$E^P\left[e^{iu(N_t - N_s)} \middle| \mathcal{F}_s \right] = \frac{E^{P_0}\left[e^{iu(N_t - N_s)} L_t \middle| \mathcal{F}_s \right]}{L_s}. \tag{5.19}$$

where we have used the fact that $E^{P_0}\left[L_t \middle| \mathcal{F}_s \right] = L_s$. In order to compute the last conditional expectation, we choose a fixed s, and consider the process Z on $[s, \infty)$ defined by

$$Z_t = e^{iu(N_t - N_s)} L_t.$$

Defining the process Y by

$$Y_t = e^{iu(N_t - N_s)},$$

we can write

$$Z_t = Y_t L_t,$$

and since both processes are of bounded variation we can use the product rule to obtain

$$dZ_t = Y_{t-} dL_t + L_{t-} dY_t + \Delta Y_t \Delta L_t.$$

The differential dL_t is given by (5.17), and for Y we easily obtain

$$dY_t = \left(e^{iu(N_{t-}+1)} - e^{iuN_{t-}} \right) e^{-iuN_s} dN_t = Y_{t-}\left(e^{iu} - 1\right) dN_t.$$

From this expression and from (5.17) we have

$$\Delta Y_t \Delta L_t = Y_{t-}\left(e^{iu} - 1\right) L_{t-}(\lambda_t - 1)\, dN_t = Z_{t-}\left(e^{iu} - 1\right)(\lambda_t - 1)\, dN_t.$$

Denoting the P_0 martingale $N_t - t$ by M_t we thus obtain

$$\begin{aligned} dZ_t &= Y_{t-} L_{t-}(\lambda_t - 1)\, dM_t + L_{t-} Y_{t-}\left(e^{iu} - 1\right) dN_t + Z_{t-}\left(e^{iu} - 1\right)(\lambda_t - 1)\, dN_t \\ &= Z_{t-}\lambda_t\left(e^{iu} - 1\right) dN_t + Z_{t-}(\lambda_t - 1)\, dM \\ &= Z_{t-}\lambda_t\left(e^{iu} - 1\right) dt + Z_{t-}\left(\lambda_t e^{iu} - 1\right) dM_t \end{aligned}$$

Integrating this we obtain

$$Z_t = Z_s + \int_s^t Z_u \lambda_u \left(e^{iu} - 1\right) du + \int_s^t Z_{u-}\left(\lambda_u e^{iu} - 1\right) dM_u.$$

We now note that, since $\mathcal{F}_\infty^\lambda \subseteq \mathcal{F}_0$, we always have $\lambda_t \in \mathcal{F}_0 \subseteq \mathcal{F}_{t-}$, so the process λ is in fact $\mathbf{F}$ predictable, implying that the dM integral is a P_0 martingale, Let $E_s^0[\cdot]$ denote the conditional expectation $E^{P_0}[\cdot| \mathcal{F}_s]$, and take E_s^0 expectations. Using the martingale property of the dM integral and the fact that $Z_s = L_s$ gives us

$$E_s^0[Z_t] = E_s^0[L_s] + \int_s^t E_s^0[Z_u \lambda_u]\left(e^{iu} - 1\right) du$$

Since $\mathcal{F}_\infty^\lambda \subseteq \mathcal{F}_0$ we have $E_s^0[Z_u \lambda_u] = \lambda_u E_s^0[Z_u]$ giving us

$$E_s^0[Z_t] = L_s + \int_s^t \lambda_u \left(e^{iu} - 1\right) E_s^0[Z_u] du$$

We now denote $E_s^0[Z_t]$ by x_t (suppressing the fixed s) we have the integral equation

$$x_t = L_s + \int_s^t \lambda_u \left(e^{iu} - 1\right) x_u du,$$

which is the integral form of the ODE

$$\begin{cases} \dot{x}_t &= x_t \lambda_u \left(e^{iu} - 1\right), \\ x_s &= L_s, \end{cases}$$

with the solution

$$x_t = L_s e^{\Lambda_{s,t}\left(e^{iu}-1\right)}.$$

We have thus shown that

$$E^0\left[e^{iu(N_t - N_s)} L_t \middle| \mathcal{F}_s\right] = L_s e^{\Lambda_{s,t}\left(e^{iu}-1\right)}$$

and inserting this into (5.19) gives us (5.18). ∎

5.5 Exercises

Exercise 5.1 *Assume that the counting process Nhas the $\mathcal{F}_t^N$ predictable intensity*

$$\lambda_t = \alpha g(N_{t-})$$

where g is a non negative known deterministic function, and $\alpha$is an unknown parameter. Show that the ML estimate of α is given by

$$\widehat{\alpha}_t = \frac{N_t}{\int_0^t g(N_s) ds}.$$

Exercise 5.2 *Assume that we can observe a process X with P dynamics given by*

$$dX_t = \mu dt + \sigma dW_t + dN_t$$

where σ is a known constant, μ is an unknown parameter, N is Poisson with unknown intensity λ, and W is standard Wiener. Our task is to estimate μ and λ, given observations of X.

(a) *Define the process X under a base measure P_0 according to the dynamics*

$$dX_t = \sigma dW_t^0 + dN_t$$

where W^0 is Wiener and N is Poisson with unit intensity under P_0. Convince yourself that N and W^0 are observable, given the X observations, in the sense that

$$\mathcal{F}_t^N \subseteq \mathcal{F}_t^X, \quad \mathcal{F}_t^{W^0} \subseteq \mathcal{F}_t^X.$$

(b) *Perform a Girsanov transformation from P_0 to P, such that X has the P dynamics $dX_t = \mu dt + \sigma dW_t + dN_t$. Write down an explicit expression for the likelihood process L.*

(c) *Maximize L w.r.t. μ and λ and show that the ML estimates are given by*

$$\widehat{\lambda}_t = \frac{N_t}{t}, \quad \widehat{\mu}_t = \frac{X_t - N_t}{t}$$

Chapter 6

Connections to PIDEs

In this chapter we will study how stochastic differential equations (SDEs) driven by counting- and Wiener processes are connected to certain types of partial integro-differential equations (PIDEs).

6.1 SDEs and Markov processes

On a filtered space $(\Omega, \mathcal{F}, P, \mathbf{F})$ we consider a scalar SDE of the form

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t + \beta(t-, X_{t-})dN_t, \tag{6.1}$$

$$X_0 = x_0. \tag{6.2}$$

Here we assume that $\mu(t,x)$, $\sigma(t,x)$ and $\beta(t,x)$ are given deterministic functions, and that x_0 is a given real number. The Wiener process W and the counting process N are allowed to be multi dimensional, in which case σ and β are row vectors of the proper dimensions and the products σdW and βdN are interpreted as inner products. We need one more important assumption.

Assumption 6.1.1 *We assume that the counting process N has a predictable intensity λ of the form*

$$\lambda_t = \lambda(t-, X_{t-}). \tag{6.3}$$

Here, with slight abuse of notation the λ in the right hand side denotes a smooth deterministic function of (t,x).

It is reasonable to expect that under these assumptions, the process X is Markov, and this is in fact true.

Proposition 6.1.1 *Under the assumptions above, the process X will be a Markov process.*

Proof. The proof is rather technical and therefore omitted. ∎

6.2 The infinitesimal generator

To every SDE of the form (6.1), and in fact to every Markov process, one can associate a natural operator, the "infinitesimal generator" $\mathcal{A}$ of the process. The infinitesimal generator is an operator on a function space, and in order to define it, let us consider a function $f : R_+ \times R \rightarrow R$. We now fix (t, x) and consider the difference quotient

$$\frac{1}{h} E_{t,x} \left[f(t+h, X_{t+h}) - f(t,x) \right].$$

The limit of this (if it exists) as $h \rightarrow 0$ would have the interpretation of a "mean derivative" of the composite process $t \longmapsto f(t, X_t)$, and this leads us to the following definition.

Definition 6.2.1 *Let us by C_b denote the space of bounded continuous mappings $f : R_+ \times R \rightarrow R$. The* **infinitesimal generator** *$\mathcal{A} : \mathcal{D} \rightarrow C_b$ is defined by*

$$\left[\mathcal{A} f \right](t,x) = \lim_{h \downarrow 0} \frac{1}{h} E_{t,x} \left[f(t+h, X_{t+h}) - f(t,x) \right], \tag{6.4}$$

where $\mathcal{D}$ is the subspace of C_b for which the limits exists for all (t, x).

Remark 6.2.1 *The* **domain** *$\mathcal{D}$ is obviously important in the definition, but in the sequel we will be rather imprecise about the exact description of $\mathcal{D}$, and we will also apply $\mathcal{A}$ to unbounded functions. The boundedness requirement above only serves to guarantee that the expected values are finite.*

Perhaps somewhat surprisingly, it turns out that the infinitesimal generator provides a huge amount of information about the underlying process X. We have for example the following central result.

Proposition 6.2.1 *The distribution of a Markov process X is uniquely determined by the infinitesimal generator $\mathcal{A}$*

We will not be able to prove this result for a general Markov process, but we will at least make it believable for the case of a SDE like in (6.1).

We thus consider the SDE (6.1) and go on to determine the shape of the infinitesimal generator. Fort simplicity we consider only the case when W and N are scalar. If $f \in C^{1,2}$ we have, from the Itô formula,

$$\begin{aligned} df(t, X_t) &= \left\{ \frac{\partial f}{\partial t}(t, X_t) + \mu(t, X_t) \frac{\partial f}{\partial t}(t, X_t) + \frac{1}{2} \sigma^2(t, X_t) \frac{\partial^2 f}{\partial x^2}(t, X_t) \right\} dt \\ &+ \sigma(t, X_t) \frac{\partial f}{\partial x}(t, X_t) dW_t + f_\beta(t-, X_{t-}) dN_t, \end{aligned}$$

where f_β is defined by

$$f_\beta(t,x) = f(t, x + \beta(t,x)) - f(t,x). \tag{6.5}$$

We now compensate the counting process by adding and subtracting the term $\lambda(T-, X_{t-})dt$ to dN_t, We then have

$$df(t, X_t) = \mathcal{A}f(t, X_t)dt + \sigma(t, X_t)\frac{\partial f}{\partial x}(t, X_t)dW_t + f_\beta(t-, X_{t-})d\bar{N}_t, \tag{6.6}$$

where the operator $\mathcal{A}$ is defined by

$$\mathcal{A}f(t, x) = \frac{\partial f}{\partial t}(t, x) + \mu(t, x)\frac{\partial f}{\partial t}(t, x) + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2 f}{\partial x^2}(t, x) + f_\beta(t, x)\lambda(t, x),$$

and where the compensated increment $d\bar{N}$ is defined by

$$d\bar{N}_t = dN_t - \lambda(t-, X_{t-})dt,$$

Remark 6.2.2 *We note that the last term of $\mathcal{A}$ in (6.6) should formally be written as*

$$f_\beta(t-, X_{t-})\lambda(t-, X_{t-}).$$

However, because of the assumed continuity of f, β, λ, and the fact that we are integrating w.r.t. Lebesgue measure dt, we are allowed to evaluate this term at (t, X_t).

The point of writing df as in (6.6) is that we have decomposed df in a **drift term** given by the dt term, and a **martingale term**, given by the sum of the dW and $d\bar{N}$ terms. Let us now fix (t, x) as initial conditions for X. We may then integrate (6.6) to obtain

$$\begin{aligned} f(t+h, X_{t+h}) &= f(t, x) + \int_t^{t+h} \mathcal{A}f(s, X_s)ds \\ &+ \int_t^{t+h} \sigma(s, X_s)\frac{\partial f}{\partial x}(s, X_s)dW_s \\ &+ \int_t^{t+h} f_\beta(s-, X_{s-})d\bar{N}_s. \end{aligned}$$

Here the dW integral is obviously a martingale and since $\bar{N}$ is a martingale and the f_β term is predictable (why?) we see that also the $d\bar{N}$ integral is a martingale. Taking expectations we thus obtain

$$\frac{1}{h}E_{t,x}\left[f(t+h, X_{t+h}) - f(t, x)\right] = E_{t,x}\left[\frac{1}{h}\int_t^{t+h} \mathcal{A}f(s, X_s)ds\right]$$

Letting $h \to 0$ and using the fundamental theorem of calculus, we obtain

$$\frac{1}{h}E_{t,x}\left[f(t+h, X_{t+h}) - f(t, x)\right] = \mathcal{A}f(t, x)$$

with $\mathcal{A}$ defined as above. We have thus proved the following main result.

Proposition 6.2.2 *Assume that X has the dynamics*

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t + \beta(t-, X_{t-})dN_t, \tag{6.7}$$

and that the intensity of N given by $\lambda(t-, X_{t-})$. Then the following hold.

- *The infinitesimal generator of X is given by*

$$\mathcal{A}f = \frac{\partial f}{\partial t}(t, x) + \mu(t, x)\frac{\partial f}{\partial t}(t, x) + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2 f}{\partial x^2} + f_\beta(t, x)\lambda(t, x), \tag{6.8}$$

  where

$$f_\beta(t, x) = f(t, x + \beta(t, x)) - f(t, x). \tag{6.9}$$

- *The process $f(t, X_t)$ is a (local) martingale if and only if it satisfies the equation*

$$\mathcal{A}f(t, x) = 0, \quad (t, x) \in R_+ \times R \tag{6.10}$$

Remark 6.2.3 *We note that the equation (6.10) contains a number of partial derivatives and a (degenerate) integral term (the f_β term). It is thus a partial integro-differential equation (PIDE).*

Remark 6.2.4 *The result extends, in the obvious way, to the case of a multi dimensional Wiener process and a multi dimensional counting process.*

We end this section by noting that the second item above can be generalized to any Markov process. We have in fact the following general result, which in fact holds for a Markov processes on a very general state space.

Proposition 6.2.3 (Dynkin's Formula) *Assume that X is a Markov process with infinitesimal generator $\mathcal{A}$. Then, for every f in the domain of $\mathcal{A}$, the process*

$$f(t, X_t) - \int_0^t \mathcal{A}f(s, X_s)ds$$

is a martingale. Furthermore, the process $f(t, X_t)$ is a martingale if and only if

$$\mathcal{A}f(t, x) = 0, \quad (t, x) \in R_+ \times R$$

Proof. The proof in the general case is a quite technical so we omit it. From an intuitive point of view the result is, however, more or less obvious, and we give an informal argument. Letting $h \to dt$ in the definition of $\mathcal{A}$ we obtain

$$E\left[df(t, X_t)\middle| \mathcal{F}_t\right] = \mathcal{A}f(t, X_t)dt$$

with the interpretation $df(t, X_t) = f(t + dt, X_{t+dt}) - f(t, X_t)$. From this it is clear that the "conditionally detrended" difference

$$df(t, X_t) - \mathcal{A}f(t, X_t)dt$$

should be a martingale increment, and this is precisely the content of the Dynkin formula. The second statement follows directly from the Dynkin formula. ∎

6.3 The Kolmogorov backward equation

We continue to study the SDE

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t + \beta(t-, X_{t-})dN_t, \tag{6.11}$$

where N has the intensity function $\lambda(t, x)$. Let us now consider a fixed point in time T and a real valued function Φ. The object of this section is to understand how one can compute the expectation $E[\Phi(X_T)]$. In order to do this we consider the process Z defined by

$$Z_t = E\left[\Phi(X_T) | \mathcal{F}_t^X\right]. \tag{6.12}$$

We first note that, since X is Markov, we can write Z as

$$Z_t = E\left[\Phi(X_T) | X_t\right],$$

so in fact we have

$$Z_t = f(t, X_t),$$

where the deterministic function f is defined by $f(t, x) = E[\Phi(X_T)| X_t = x]$ or, equivalently by

$$f(t, x) = E_{t,x}[\Phi(X_T)].$$

Secondly we note that since Z is given by the conditional expectation (6.12), the process Z is a martingale. From Proposition 6.2.2 we thus see that f must satisfy

$$\mathcal{A}f(t, x) = 0,$$

with the obvious boundary condition $f(T, x) = \Phi(x)$. We have thus proved the following result.

Proposition 6.3.1 (The Kolmogorov Backward Equation) *Let X be the solution of (6.7), T a fixed point in time, and Φ any function such that $\Phi(X_T) \in L^1$. Define the function f by*

$$f(t, x) = E_{t,x}[\Phi(X_T)].$$

Then f satisfies the Kolmogorov backward equation

$$\begin{cases} \mathcal{A}f(t, x) &= 0, \quad (t, x) \in R_+ \times R \\ f(T, x) &= \Phi(x), \quad x \in R \end{cases} \tag{6.13}$$

where $\mathcal{A}$ is given by (6.8).

Remark 6.3.1 *It is clear from the discussion around the Dynkin formula that the Kolmogorov backward equation is valid, not only for the solution to an SDE of the form (6.7), but in fact for a general Markov process.*

In particular we may choose $\Phi(x) = I_A(x)$ where $A \subseteq R$ is a Borel set in R, and I denotes the indicator function. In this case we see that

$$f(t,x) = P\left(X_T \in A \,|X_t = x\right).$$

and we see that these **transition probabilities** must satisfy the backward equation with the boundary condition

$$f(T,x) = I_A(x).$$

Even more in particular, if we assume that X has transition **densities** $p(t,x;T,z)$ with the interpretation

$$p(t,x;T,z)dz = P\left(X_T \in dz \,|X_t = x\right)$$

then also these transition densities must satisfy the Kolmogorov equation in the (t,x) variables, with boundary condition

$$p(T,x;T,z) = \delta_z(x),$$

where δ_z is the Dirac measure at z. We thus see that the transition probabilities, and thus the entire distribution, of the process X are completely determined by the infinitesimal generator $\mathcal{A}$.

We can also turn the Kolmogorov equation around. Instead of starting with an expected value and deriving a PIDE, we may start with the PIDE and derive an expected value.

Proposition 6.3.2 (Feynman-Kac) *Assume that X is as above, and assume that a function f solves the PIDE*

$$\left\{\begin{array}{rcll} \mathcal{A}f(t,x) & = & 0, & (t,x) \in R_+ \times R \\ f(T,x) & = & \Phi(x), & x \in R \end{array}\right. ,$$

where $\mathcal{A}$ is given by (6.8). Then we have

$$f(t,x) = E_{t,x}\left[\Phi(X_T)\right].$$

Proof. Assume that f satisfies the backward equation. If we consider the process $f(t,X_t)$ then, since $\mathcal{A}f = 0$, it is clear from the Dynkin formula that $f(t,X_t)$ is a martingale. Using the Markov property and the boundary condition we then obtain

$$f(t,X_t) = E\left[\left. f(T,X_T)\right| \mathcal{F}_t\right] = E\left[\left.\Phi(X_T)\right| X_t\right]. \quad \blacksquare$$

In many finance applications it is natural to consider problems with discounting. A small variation of the arguments above gives us the following result.

Proposition 6.3.3 (Feynman-Kac) *Assume that X is as above, and assume that a function f solves the PIDE*

$$\left\{ \begin{array}{rcl} \mathcal{A}f(t,x) - rf(t,x) & = & 0, \quad (t,x) \in R_+ \times R \\ f(T,x) & = & \Phi(x), \quad x \in R \end{array} \right. ,$$

where $\mathcal{A}$ is given by (6.8), and r is a real number Then we have

$$f(t,x) = e^{-r(T-t)} E_{t,x}\left[\Phi(X_T)\right].$$

Part II

Arbitrage Theory

Chapter 7

Portfolio Dynamics and Martingale Measures

7.1 Portfolios

We now turn to the problem of pricing financial derivatives in models which are driven, not only by a finite number of Wiener processes, but also by a number of counting processes. In this chapter we recall some central concepts and result from general arbitrage theory. For details the reader is referred to [1] or any other standard textbook on the subject.

We consider a market model consisting of $N+1$ financial assets (without dividends). As usual we assume that the market is perfectly liquid, that there is no credit risk, no bid-ask spread, and that prices are not affected by our portfolios.

We are given a filtered probability space $(\Omega, \mathcal{F}, P, \mathbf{F})$, and by S_t^i we denote the price at time t of one unit of asset No. i, for $i = 0, \ldots, N$. We let S denote the corresponding N dimensional column vector process, and all asset price processes are assumed to be optional. The asset S^0 will play a special role below. In general the price processes are allowed to be semi martingales but the reader can, without any loss of good ideas, think of the simpler case when the prices are driven by finite number of Wiener and counting processes.

We now go on to define the concept of a "self financing portfolio". Intuitively this is a portfolio strategy whee there is no external withdrawal from, or infusion of money to, the portfolio. It is far from trivial how this should be formalized in continuous time, but a careful discretization argument leads to the following formal definition.

Definition 7.1.1 *A* **portfolio strategy** *is an $N+1$ dimensional predictable (row vector) process $h = (h^1, \ldots, h^N)$. For a given strategy h, the corresponding*

value process V^h *is defined by*

$$V_t^h = \sum_{i=o}^{N} h_t^i S_t^i = h_t S_t. \tag{7.1}$$

or equivalently

$$V_t^h = h_t S_t. \tag{7.2}$$

The strategy is said to be **self financing** *if*

$$dV_t^h = \sum_{i=o}^{N} h_t^i dS_t^i, \tag{7.3}$$

or equivalently

$$V_t^h = h_t dS_t. \tag{7.4}$$

For a given strategy h*, the corresponding* **relative portfolio** $u = (u^1, \ldots, u^N)$ *is defined by*

$$u_t^i = \frac{h_t^i S_{t-}^i}{V_{t-}^h}, \quad i = 0, \ldots, N, \tag{7.5}$$

and we will obviously have

$$\sum_{i=0}^{N} u_t^i = 1.$$

It is important to note the requirement of predictability for h in the definition above. Informally this means that at $t - dt$ we decide on our portfolio h_t, and we then hold this portfolio over the infinitesimal interval $[t - dt, t]$.

We should, in all honesty, also require some minimal integrability properties for our admissible portfolios, but we will suppress these and some other technical conditions. The reader is referred to the specialist literature for details.

As in the Wiener case, it is often easier to work with the relative portfolio u than with the portfolio h. We immediately have the following obvious result.

Proposition 7.1.1 *If* u *is the relative portfolio corresponding to a self financing portfolio* h*, then we have*

$$dV_t^h = V_{t-}^h \sum_{i=o}^{N} u_t^i \frac{dS_t^i}{S_{t-}^i}. \tag{7.6}$$

In most market models we have a (locally) risk free asset, and the formal definition is as follows.

Definition 7.1.2 *Suppose that one of the asset price processes, henceforth denoted by* B*, has dynamics of the form*

$$dB_t = r_t B_{t-} dt, \tag{7.7}$$

where r *is some predictable random process. In such a case we say that the asset* B *is (locally)* **risk free***, and we refer to* B *as the* **bank account***. The process* r *is referred to as the corresponding* **short rate***.*

The term "locally risk free" is more or less obvious. If we are standing at time $t - dt$ then, because of predictability, we know the value of r_t. We also know B_{t-}, which implies that already at time $t - dt$ we know the value B_t of B at time t. The asset B is thus risk free on the local (infinitesimal) time scale, even if the short rate r is random. The interpretation is the usual, i.e. we can think of B as the value of a bank account where we have the short rate r. Typically we will choose B as the asset S^0.

7.2 Arbitrage

The definition of arbitrage is standard.

Definition 7.2.1 *A portfolio strategy h is an* **arbitrage** *strategy on the time interval $[0, T]$ if the following conditions are satisfied.*

1. *The strategy h is self financing*
2. *The initial cost of h is zero, i.e.*

$$V_0^h = 0.$$

3. *At time T it holds that*

$$\begin{aligned} P\left(V_T^h \geq 0\right) &= 1, \\ P\left(V_T^h > 0\right) &> 0. \end{aligned}$$

An arbitrage strategy is thus a money making machine which produces positive amounts of money out of nothing. The economic interpretation is that the existence of an arbitrage opportunity signifies a serious case of mispricing in the market, and a minimal requirement of market efficiency is that there are no arbitrage opportunities. The single most important result in mathematical finance is the "first fundamental theorem" which connects absence of arbitrage to the existence of a martingale measure.

Definition 7.2.2 *Consider a market model consisting of $N+1$ assets $S^0, \ldots, S^N$, and assume that the* **numeraire asset** *S^0 has the property that $S_t^0 > 0$ with probability one for all t. An equivalent* **martingale measure** *is a probability measure Q with the properties that*

1. *Q is equivalent ot P, i.e. $Q \sim P$.*
2. *The normalized price processes $Z_t^0, \ldots Z_t^N$, defined by*

$$Z_t^i = \frac{S_t^i}{S_t^0}, \quad i = 0, \ldots N,$$

are (local) martingales under Q.

We can now sate the main abstract result.

Theorem 7.2.1 (The First Fundamental Theorem) *The market model is free of arbitrage possibilities if and only if there exists a martingale measure Q.*

Proof. This is a very deep result, and the reader is referred to the literature for a proof. ∎

We note that if there exists a martingale measure Q, then it will depend upon the choice of the numeraire asset S^0, so we should really index Q as Q^0. In most cases the numeraire asset will be the bank account B, and in this case the measure Q, which more precisely should be denoted by Q^B, is known as the **risk neutral martingale measure**.

The First Fundamental Theorem above is very powerful and general result. In some more restricted cases, especially when we use "the classical delta hedging approach" to arbitrage free pricing below, we will not need the full force of the First Fundamental Theorem. In these cases, the following, very easy, result will be enough for our purposes.

Proposition 7.2.1 *Assume that the self financing portfolio strategy h is such that the corresponding value process V^h has dynamics of the form*

$$dV_t^h = V_t k_t dt,$$

where k is some adapted process. Assume furthermore that the market contains also a bank account with short rate process r. Then we must have

$$k_t = r_t, \quad P - a.s. \quad \forall t \geq 0,$$

otherwise there will exist an arbitrage opportunity.

Proof. The point of the result is that if the V process has dynamics as above (with no driving noise process), then V represents a (locally) **risk free** investment opportunity, and in order to avoid arbitrage between the portfolio and the bank account we must have $k = r$. If for example $r_t < k_t$ then we borrow in the bank and invest in the portfolio, and vice versa if $r_t > k_t$. ∎

7.3 Martingale Pricing

We now study the possibility of pricing contingent claims. The formal definition of a claim is as follows.

Definition 7.3.1 *Given a a stochastic basis $(\Omega, \mathcal{F}, P, \mathbf{F})$ and a specified point in time T, often referred to as "the exercise date") a* **contingent T-claim** *is a random variable $X \in \mathcal{F}_T$.*

The interpretation is that the holder of the claim will obtain the random amount X of money at time T. We now consider the "primary" or "underlying" market $S^0, S^1, \ldots, S^N$ as given *a priori*, and we fix a T-claim X. Our task is that of determining a "reasonable" price process $\Pi(t; X)$ for X, and we assume that the primary market is arbitrage free. A main idea is the following.

> The derivative should be priced in a way that is **consistent** with the prices of the underlying assets. More precisely we should demand that the extended market $\Pi(\ ; X), S^0, S^1, \cdots, S^N$ is free of arbitrage possibilities.

In this approach we thus demand that there should exist a martingale measure Q for the extended market $\Pi(X), S^0, S^1, \cdots, S^N$. Letting Q denote such a measure, assuming enough integrability, and applying the definition of a martingale measure we obtain

$$\frac{\Pi(t; X)}{S_t^0} = E^Q\left[\left.\frac{\Pi(T; X)}{S_T^0}\right| \mathcal{F}_t\right] = E^Q\left[\left.\frac{X}{S_T^0}\right| \mathcal{F}_t\right] \tag{7.8}$$

where we have used the fact that, in order to avoid arbitrage at time T we must have $\Pi(T; X) = X$. We thus have the following result.

Theorem 7.3.1 (General Pricing Formula) *The arbitrage free price process for the T-claim X is given by*

$$\Pi(t; X) = S_t^0 E^Q\left[\left.\frac{X}{S_T^0}\right| \mathcal{F}_t\right], \tag{7.9}$$

where Q is the (not necessarily unique) martingale measure for the a priori given market $S^0, S^1, \cdots, S^N$, with S^0 as the numeraire .

Note that different choices of Q will generically give rise to different price processes. In particular we note that if we assume that if S^0 is the money account

$$S_t^0 = S_0^0 \cdot e^{\int_0^t r(s)ds},$$

where r is the short rate, then (7.9) reduced to the familiar "risk neutral valuation formula".

Theorem 7.3.2 (Risk Neutral Valuation Formula)
Assuming the existence of a short rate, the pricing formula takes the form

$$\Pi(t; X) = E^Q\left[\left. e^{-\int_t^T r(s)ds} X\right| \mathcal{F}_t\right]. \tag{7.10}$$

where Q is a (not necessarily unique) martingale measure with the money account as the numeraire.

The pricing formulas (7.9) and (7.10) are very nice, but it is clear that if there exists more than one martingale measure (for the chosen numeraire), then the formulas do not provide a unique arbitrage free price for a given claim X. It is thus natural to ask under what conditions the martingale measure is unique, and this turns out to be closely linked to the possibility of hedging contingent claims.

7.4 Hedging

Consider a market model $S^0, \ldots, S^N$ and a contingent T-claim X.

Definition 7.4.1 *If there exists a self financing portfolio h such that the corresponding value process V^h satisfies the condition*

$$V_T^h = X, \quad P - a.s. \tag{7.11}$$

then we say that h **replicates** *X, that h is a* **hedge** *against X, or that X is* **attained** *by h. If, for every T, all T-claims can be replicated, then we say that the market is* **complete**.

Given the hedging concept, we now have a second approach to pricing. Let let us assume that X can be replicated by h. Since the holding of the derivative contract and the holding of the replicating portfolio are equivalent from a financial point of view, we see that price of the derivative must be given by the formula

$$\Pi(t; X) = V_t^h, \tag{7.12}$$

since otherwise there would be an arbitrage possibility (why?).

We now have two obvious problems.

- What will happen in a case when X can be replicated by two different portfolios g and h?
- How is the formula (7.12) connected to the previous pricing formula (7.9)?

To answer these question, let us assume that the market is free of arbitrage, and let us also assume at the T claim X is replicated by the portfolios g and h. We choose the bank account B as the numeraire and consider a fixed martingale measure Q. Since Q is a martingale measure for the underlying market $S^0, \ldots, S^N$, it is easy to see that this implies that Q is also a martingale measure for V^g and V^h in the sense that V^h/B and V^g/B are Q martingales. Using this we obtain

$$\frac{V_t^h}{B_t} = E^Q\left[\left.\frac{V_T^h}{B_T}\right| \mathcal{F}_T\right]$$

and similarly for V^g. Since, by assumption, we have $V_T^h = X$ we thus have

$$V_t^h = E^Q\left[\left. X\frac{B_t}{B_T}\right| \mathcal{F}_T\right],$$

which will hold for any replicating portfolio and for any martingale measure Q. Assuming absence of arbitrage we have thus proved the following.

- If X is replicated by g and h, then

$$V_t^h = V_t^g, \quad t \geq 0.$$

- For an attainable claim, the value of the replicating portfolio coincides with the risk neutral valuation formula, i.e.

$$V_t^h = E^Q \left[e^{-\int_t^T r_s ds X} \middle| \mathcal{F}_T \right].$$

From (7.10) it is obvious that every claim X will have a unique price if and only if the martingale measure Q is unique. On the other hand, it follows from the alternative pricing formula (7.12) that there will exist a unique price for every claim if every claim can be replicated. The following result is therefore not surprising.

Theorem 7.4.1 (Second Fundamental Theorem) *Given a fixed numeraire S^0, the corresponding martingale measure Q^0 is unique if and only if the market is complete.*

Proof. We have already seen above that if the market is complete, then the martingale measure is unique. The other implication is a very deep result, and the reader is referred to the literature. ∎

7.5 Heuristic results

In this section we will provide a very useful and general rule of thumb which can be used to determine whether a certain model is complete and/or free of arbitrage. The arguments will be purely heuristic.

Let us consider a model with N traded underlying assets **plus** the risk free asset (i.e. totally $N + 1$ assets). We assume that the price processes of the underlying assets are driven by R "random sources". We cannot give a precise definition of what constitutes a "random source" here, but the following infromal rules will be enough for our purposes.

- Every independent Wiener process counts as one source of randomness. Thus, if we have five independent Wiener processes, then $R = 5$.

- Every independent Poisson process counts as one source of randomness. Thus, if we have five independent Wiener processes,and three independent Poisson processes, then $R = 5 + 3 = 8$.

- If we have a driving point process with random jump size, then **every possible jump size** counts as one source of randomness. Thus, if we have a compound Poisson process with three possible jump sizes at each jump time then $R = 3$. If the jump size has a probability distribution allowing a density w.r.t. Lebesgue measure, then $R = \infty$.

When discussing completeness and absence of arbitrage it is important to realize that these concepts work in opposite directions. Let the number of random sources R be fixed. Then every new underlying asset added to the model (without increasing R) will of course give us a potential opportunity of creating an arbitrage portfolio, so in order to have an arbitrage free market the number M of underlying assets must be small in comparison to the number of random sources R.

On the other hand we see that every new underlying asset added to the model gives us new possibilities of replicating a given contingent claim, so completeness requires M to be great in comparison to R.

We cannot formulate and prove a precise result here, but the following rule of thumb, or "meta-theorem", is nevertheless extremely useful. In concrete cases it can in fact be given a precise formulation and a precise proof.

Meta-Theorem 7.5.1 *Let M denote the number of underlying* **traded** *assets in the model* **excluding** *the risk free asset, and let R denote the number of random sources. Generically we then have the following relations.*

1. *The model is arbitrage free if and only if $M \leq R$.*
2. *The model is complete if and only if $M \geq R$.*
3. *The model is complete and arbitrage free if and only if $M = R$.*

As an example we take the Black–Scholes model, where we have one underlying asset S plus the risk free asset so $M = 1$. We have one driving Wiener process, giving us $R = 1$, so in fact $M = R$. Using the meta-theorem above we thus expect the Black–Scholes model to be arbitrage free as well as complete and this is indeed the case.

Chapter 8

Poisson Driven Stock Prices

8.1 Introduction

In this chapter we will study arbitrage pricing in a concrete model. The model below is extremely simple and very unrealistic from an economic point of view. We use it mostly as a small laboratory model, and it it is nevertheless instructive to analyze it.

The model is very similar to the standard Black-Scholes model, the only difference being that while the stock price in the Black-Scholes model is driven by a Wiener process, the stock price in our model will be driven by a Poisson process. The Wiener driven and the Poisson driven models are structurally very close, so the reader will hopefully recognize concepts and techniques from the Wiener case. Exactly as in the Wiener case, we have three different methods for pricing financial derivatives.

1. Construction of locally risk free portfolios.
2. Construction of replicating portfolios.
3. Construction of equivalent martingale measures.

As one may expect, the martingale approach is the most general one, but it is still very instructive (and a good exercise) to see how far it is possible to go using the techniques 1-2 above.

As usual we consider a filtered space $(\Omega, \mathcal{F}, P, \mathbf{F})$ and we assume that the space carries a Poisson process N with constant intensity λ. The filtration is the internal one generated by N. The market we will study is a very simple one. It consists of two assets, namely a risky asset with price process S and a the usual bank account B. The dynamics are as follows, where α, β and r are known constants.

$$dS_t = \alpha S_{t-}dt + \beta S_{t-}dN_t, \quad (8.1)$$
$$dB_t = rB_t dt. \quad (8.2)$$

In order to have an economic interpretation of the stock price dynamics (8.1) we note that **between jumps** the price evolves according to the deterministic ODE

$$\frac{dS_t}{dt} = \alpha S_t,$$

so between jumps the stock price grows exponentially with the factor α. We thus see that the constant α is the local mean rate of return of the stock **between jumps**.

The question is now to get a grip on the overall mean rate of return, and in order to do this we recall that if we define the process M by

$$M_t = N_t - \lambda t, \tag{8.3}$$

then M is a martingale, and we will write (8.3) as

$$dN_t = \lambda dt + dM_t. \tag{8.4}$$

If we now plug (8.4) into (8.1) we obtain, after some reshuffling,

$$dS_t = S_{t-} (\alpha + \beta\lambda)\, dt + \beta S_{t-} dM_t, \tag{8.5}$$

and we see that the overall mean rate of return of the stock, **including jumps** is given by $\alpha + \beta\lambda$.

We see that we have two equivalent ways of viewing the stock price dynamics, since can write the dynamics either as (8.1) or as (8.5). From a probabilistic point of view, (8.5) is the most natural one, since it decomposes the dynamics into a predictable **drift** part (the dt term), and a **martingale** part (the dM term). This is known as "the semimartingale decomposition" of the S dynamics. The representation (8.1), on the other hand, is often easier to use when we want to apply the Ito formula. As we will see below we will often switch between the two representations.

If we now move to the dN term in (8.1) we see that if N has a jump at time t, then the induced jump size of S is given by

$$\Delta S_t = \beta S_{t-},$$

so β is the **relative jump size** of the stock price, and we will sometimes refer to β as the "jump volatility" of S. We also see that the sign of β determines the sign of the jump: Assuming $S_t > 0$, if $\beta > 0$ then all jumps are upwards whereas if $\beta < 0$ all jump are downwards. In particular we see that if $\beta = -1$ then, if there is a jump at t, we obtain $\Delta S_t = -S_{t-}$, i.e.

$$S_t = S_{t-} + \Delta S_t = S_{t-} - S_{t-} = 0.$$

In other words, if $\beta = -1$, then the stock price will jump to zero at the first jump of N (and the stock price will stay forever at the value zero). This indicates clearly that we may use counting processes in order to model bankruptcy phenomena.

We can now collect our findings so far.

Proposition 8.1.1 *If S is given by (8.1) we have the following interpretation.*

- *The constant α is the local mean rate of return of the stock* **between jumps**.
- *Denoting the overall mean rate of return of the stock,* **including jumps** *under the measure P by μ^P, we have*

$$\mu^P = \alpha + \beta\lambda. \tag{8.6}$$

- *The relative jump size is given by β.*

Before we go on to pricing in this simple model, let us informally discuss conditions of no arbitrage.

Let us first assume that $\beta > 0$, and that $S_0 > 0$. Then all jumps are positive and it is clear that a necessary condition for no arbitrage is that $r > \alpha$, since otherwise the stock return would dominate the return of the bank between jumps, and dominate even more at a jump time (because of the positive jumps). In other words, if $r < \alpha$ then we would have an arbitrage by borrowing in the bank and investing in the stock.

If, on the other hand, $\beta < 0$, then all jumps are negative, and a necessary condition for no arbitrage is that $\alpha > r$, since otherwise the bank would dominate the stock between jumps and even more so at a jump time. We can summarize the findings as follows.

Proposition 8.1.2 *A necessary condition for absence of arbitrage is given by*

$$\frac{r-\alpha}{\beta} > 0. \tag{8.7}$$

8.2 The classical approach to pricing

We now turn the the problem of pricing derivatives in the model above. In this section we will basically follow the "classical" Black-Scholes delta hedging methodology and to this end we consider a contingent T-claim X of the form

$$X = \Phi(S_T),$$

where Φ is some given contract function. A typical example would be a European call option with exercise date T and strike price K, in which case Φ would have the form

$$\Phi(s) = \max\left[s - K, 0\right].$$

We now assume that the derivative is traded on a liquid market, and that the price $\Pi(t; X)$ is of the form

$$\Pi(t; X) = F(t, S_t),$$

for some smooth function $F(t, s)$. Our job is to find out what the pricing function F must look like in order to avoid arbitrage on the extended market (S, B, F). To this end we carry out the following program.

1. Form a self financing portfolio based on the stock S and the derivative F and denote the corresponding value process of the portfolio by V
2. Choose the portfolio weights such that the dN terms in the V dynamics cancel.
3. The V dynamics will then be of the form
$$dV_t = V_t k_t dt,$$
for some random process k.
4. Thus V is a **risk free** portfolio and in order to avoid arbitrage possibilities between V and the bank account we must, according to Proposition 7.2.1 have the equation
$$k_t = r, \quad t \geq 0.$$
5. The equation above turns out to be a PIDE for the determination for the pricing function F.

We now go on to carry out this small program. Denoting the relative weights on the underlying stock and the derivative by u^S and u^F respectively we have the portfolio dynamics

$$dV_t = dV_{t-} \left\{ u_t^S \frac{dS_t}{S_{t-}} + u_t^F \frac{dF(t, S_t)}{F(t, S_{t-})} \right\},$$

which we write more compactly as

$$dV = V^- \left\{ u^S \frac{dS}{S^-} + u^F \frac{dF}{F^-} \right\},$$

where S^- is shorthand for S_{t-} etc. From the Itô formula we immediately have

$$dF(t, S_t) = \left\{ \frac{\partial F}{\partial t}(t, S_t) + \alpha S_t \frac{\partial F}{\partial s}(t, S_t) \right\} dt + F_\beta(t, S_{t-}) dN_t,$$

where

$$F_\beta(t, s) = F(t, s + \beta s) - F(t, s),$$

and we can rewrite this in shorthand as

$$dF = \alpha_F F dt + \beta_F^- F^- dN_t,$$

where

$$\alpha_F(t, s) = \frac{\frac{\partial F}{\partial t}(t, s) + \alpha s \frac{\partial F}{\partial s}(t, s)}{F(t, s)} \qquad \beta_F(t, s) = \frac{F_\beta(t, s)}{F(t, s)}.$$

With this notation the portfolio dynamics takes the form

$$dV = V^- \left\{ u^S(\alpha dt + \beta dN) + u^F(\alpha_F dt + \beta_F^- dN) \right\},$$

or alternatively

$$dV = V^{-}\left\{u^S\alpha + \alpha_F\right\}dt + V^{-}\left\{u^S\beta + u^F\beta_F^{-}\right\}dN$$

We thus see that if we choose the relative portfolio u such that

$$u^S\beta + u^F\beta_F^{-} = 0,$$

then the driving Poisson noise in the portfolio dynamics will vanish. Recalling the the portfolio weights must sum to unity, we thus define the relative portfolio by the system

$$\begin{aligned} u^S\beta + u^F\beta_F &= 0 \\ u^S + u^F &= 1. \end{aligned}$$

This is a simple 2×2 system of linear equations with the solution

$$\begin{aligned} u^S &= \frac{\beta_F^{-}}{\beta_F^{-} - \beta}, \\ u^F &= -\frac{\beta}{\beta_F^{-} - \beta}. \end{aligned}$$

Using this portfolio, the V dynamics takes the form

$$dV = V^{-}\left\{\frac{\alpha\beta_F}{\beta_F - \beta} - \frac{\alpha_F\beta}{\beta_F - \beta}\right\}dt$$

which represents the dynamics of a risk free asset. From Proposition 7.2.1 we thus see that, in order to avoid arbitrage, the condition

$$\frac{\alpha\beta_F}{\beta_F - \beta} - \frac{\alpha_F\beta}{\beta_F - \beta} = r,$$

must be satisfied $P - a.s.$. Substituting the definitions for α_F and β_F and reshuffling this equation will finally give us the equation

$$\frac{\partial F}{\partial t} + \alpha s\frac{\partial F}{\partial s} + \frac{r - \alpha}{\beta}F_\beta - rF = 0.$$

This is the required no arbitrage condition for the pricing function F. Recalling that we have the obvious (why?) boundary condition $F(T, s) = \Phi(s)$, we have our first main pricing result.

Proposition 8.2.1 *Consider the pure Poisson model (8.1)-(8.2) and a T claim X of the form $X = \Phi(S_T)$. Assume that the price process $\Pi(t; X)$ is of the form $\Pi(t; X) = F(T, S_t)$. Then, in order to avoid arbitrage, the pricing function F must satisfy the following PIDE on the time interval $[0, T]$.*

$$\left\{\begin{aligned} \frac{\partial F}{\partial t}(t,s) + \alpha s\frac{\partial F}{\partial s}(t,s) + \frac{r - \alpha}{\beta}F_\beta(t,s) - rF(t,s) &= 0, \\ F(T,s) &= \Phi(s), \end{aligned}\right. \tag{8.8}$$

where, F_β is defined by

$$F_\beta(t,s) = F(t, s+\beta s) - F(t,s). \tag{8.9}$$

Comparing the PIDE above with our results in Section 6.3 we see (with great satisfaction) that it is precisely of the form which allows for a Feynman-Kac representation. In fact, using Propositions 6.2.2 and 6.3.3 we can write the solution to the PIDE (8.9) as

$$F(t,s) = e^{-r(T-t)} E^Q_{t,s}\left[\Phi(S_T)\right],$$

where S has dynamics

$$dS_t = \alpha S_t dt + \beta S_t dN_t, \tag{8.10}$$

and N is Poisson with intensity $(r-\alpha)/\beta$ under the measure Q. This looks quite nice, but at this point we have to be a bit careful, since in order to apply the relevant Feynman-Kac Theorem we need to assume that the condition

$$\frac{r-\alpha}{\beta} > 0, \tag{8.11}$$

is satisfied, otherwise we are dealing with a Poisson process with negative intensity, and such animals do not exist. This condition is, however, exactly the necessary condition for absence of arbitrage that we encountered in Proposition 8.1.2, and we can summarize our finding as follows.

Proposition 8.2.2 *Consider the pure Poisson model (8.1)-(8.2), where N is Poisson with constant intensity λ under the objective measure P, and where we assume that the no arbitrage condition (8.11) is satisfied. Consider a T claim X of the form $X = \Phi(S_T)$ and assume that the price process $\Pi(t;X)$ is of the form $\Pi(t;X) = F(T, S_t)$. Absence of arbitrage will then imply that F has the representation*

$$F(t,s) = e^{-r(T-t)} E^Q_{t,s}\left[\Phi(S_T)\right], \tag{8.12}$$

where the S dynamics are given by (8.1.2), but where the process N under the measure Q is Poisson with intensity

$$\lambda^Q = \frac{r-\alpha}{\beta}.$$

An explicit formula for F is given by

$$F(t,s) = e^{-r(T-t)} \sum_{n=0}^{\infty} \Phi\left(s(1+\beta)^n e^{\alpha(T-t)}\right) \frac{(r-\alpha)^n (T-t)^n}{\beta^n n!} e^{-\frac{r-\alpha}{\beta}(T-t)}. \tag{8.13}$$

We end this section by discussing how F depends on the various model parameters. The most striking fact, which we see from Propositions 8.2.1 and 8.2.2 is that while F depends on the parameters α, β and r, it does **not** depend on the parameter λ, which is the Poisson intensity λ under the objective measure P. This is the point process version of the fact that in a Winer driven model, the pricing function does not depend on the local mean rate of return.

We also see that the dynamics of S is given by

$$dS_t = \alpha S_t dt + \beta S_{t-} dN_t,$$

under the objective measure P as well as under the martingale measure Q. The difference between P and Q is that while N is Poisson with intensity λ under P, it is Poisson with intensity $\lambda^Q = (r - \alpha)/\beta$ under Q.

We can also easily compute the local rate of return of S under Q. We can write

$$dN_t = \left(\frac{r-\alpha}{\beta}\right) dt + dM_t^Q,$$

where M^Q is a Q martingale. Inserting this into the S dynamics above gives us

$$dS_t = r S_t dt + \beta S_{t-} dM_t^Q,$$

which shows that the local mean rate of return under Q equals the short rate r.

This shows, as was expected, that Q is a risk neutral martingale measure for S, i.e. that the process S/B is a Q martingale. It is also easy to see that F/S is a Q martingale.

8.3 The martingale approach to pricing

In this section we will study the simple Poisson market described above in terms of martingale measures. We recall the P dynamics of the stock price

$$dS_t = \alpha S_t dt + \beta S_{t-} dN_t, \tag{8.14}$$

and our first task is to find the relevant no arbitrage conditions. By elementary arguments we have already derived the condition (8.11), but the logical status of this condition is that it is only a necessary condition. We now want to find necessary **and** sufficient conditions, and to this end we now determine the class of equivalent martingale measures (with the bank account as numeraire) for our market model on the compact interval $[0, T]$.

Since the filtration is the internal one, we know from the converse of the Girsanov Theorem that every measure $Q \sim P$, regardless of whether Q is a martingale measure or not, is obtained by a Girsanov transformation of the form

$$L_t = \frac{dQ}{dP}, \quad \text{on } \mathcal{F}_t, \quad 0 \le t \le T,$$

where the likelihood process L has the dynamics

$$\begin{cases} dL_t &= L_{t-} h_t \left\{ dN_t - \lambda_t dt \right\}, \\ L_0 &= 1, \end{cases}$$

for some predictable process H with $h > -1$. From the Girsanov Theorem we know that the Q intensity of N is given by

$$\lambda_t^Q = (1 + h_t)\lambda,$$

so we can write

$$dN_t = (1 + h_t)\lambda dt + dM_t^Q,$$

where M^Q is a Q martingale. Subsituting this into the S dynamics gives us the semimartingale mdecomposition of S under Q as

$$dS_t = S_t \left\{ \alpha + \beta(1 + h_t)\lambda \right\} dt + \beta S_{t-} dM_t^Q,$$

so the local mean rate of return under Q is given by

$$\mu_t^Q = \alpha + \beta(1 + h_t)\lambda.$$

A martingale measure with B as numeraire is characterized by the fact that $\mu_t^Q = r$, so Q is a risk neutral martingale measure if and only if

$$\alpha + \beta(1 + h_t)\lambda = r. \tag{8.15}$$

First Fundamental Theorem now says that the market model is free of arbitrage if and only if there exists an equivalent martingale measure, so we see that we have absence of arbitrage if and only if eqn (8.15) has a solution h such that $h_t > -1$. Since (8.15) has the simple solution

$$h_t = \frac{r - \alpha}{\lambda\beta} - 1,$$

and $\lambda > 0$ we have more or less proved the following

Proposition 8.3.1 *The pure Poisson market model (8.1)-(8.2) is free of arbitrage if and only if the condition*

$$\frac{r - \alpha}{\beta} > 0, \tag{8.16}$$

is satisfied. If the condition is satisfied,then the market is also complete, and the process N is Poisson under Q with intensity given by

$$\lambda_t^Q = \frac{r - \alpha}{\beta}. \tag{8.17}$$

Given this result, pricing of derivatives is very , and we have the following result.

Proposition 8.3.2 *Assume that the no arbitrage condition (8.16) is satisfied, and consider any contingent T claim X. Then X will have a unique arbitrage free price process $\Pi(t; X)$ given by*

$$\Pi(t; X) = e^{-r(T-t)} E^Q \left[X \middle| \mathcal{F}_t \right]. \tag{8.18}$$

Furthermore, if X is of the form $X = \Phi(S_T)$ for some deterministic contract function Φ, then we have

$$\Pi(t; X) = F(t, S_t),$$

where the pricing function F satisfies the PIDE

$$\left\{ \begin{array}{rcl} \frac{\partial F}{\partial t}(t,s) + \alpha s \frac{\partial F}{\partial s}(t,s) + \frac{r-\alpha}{\beta} F_\beta(t,s) - rF(t,s) & = & 0, \\ F(T,s) & = & \Phi(s), \end{array} \right. \tag{8.19}$$

Proof. The risk neutral valuation formula (8.18) is just a special case of Proposition 7.3.2. If X is of the form $X = \Phi(S_T)$ then, since S is Markov under Q, we can write

$$e^{-r(T-t)} E^Q \left[X | \mathcal{F}_t \right] = e^{-r(T-t)} E^Q \left[X | S_t \right] = F(t, S_t),$$

and the PIDE (8.19) is the Kolmogorov backward equation. ∎

8.4 The martingale approach to hedging

As we saw in the previous section, the martingale measure is unique, so the Second Fundamental Theorem guarantees that the Poisson model is complete. In this simple case we can in fact also provide a self contained proof of market completeness. Let us thus consider a T-claim X.

Our formal job is to construct three processes, V, u^B and u^S such that the following hold.

- The (prospective) weights u^B and u^S are predictable and sum to unity, i.e.

$$u_t^B + u_t^S = 1.$$

- The V dynamics have the form

$$dV_t = V_{t-} \left\{ u_t^S \frac{dS_t}{S_{t-}} + u_t^B \frac{dB_t}{B_{t-}} \right\}. \tag{8.20}$$

- V replicates X at maturity, i.e.

$$V_T = X.$$

In order to do this we consider the associated the price process

$$\Pi(t; X) = e^{-r(T-t)} E^Q \left[X | \mathcal{F}_t \right]. \tag{8.21}$$

Furthermore, we hope that for the replicating relative portfolio u we will have

$$V_t^u = \Pi(t; X).$$

The idea is now to use a martingale representation result to obtain the dynamics for $\Pi(t;X)$ from (8.21), compare these dynamics with (8.20) and thus identify the portfolio weights.

We start by noting that $\Pi(t;X)$, henceforth abbreviated as π_t, can be written as

$$\pi_t = e^{-r(T-t)}Y_t,$$

where Y is defined by

$$E^Q[X|\mathcal{F}_t].$$

It is clear that Y is a Q martingale, so by a slight variation of the Martingale Representation Theorem 4.1.1 we deduce the existence of a predictable process g such that

$$dY_t = g_t Y_{t-}\left\{dN_t - \lambda^Q dt\right\}.$$

where

$$\lambda^Q = \frac{r-\alpha}{\beta}.$$

From the product formula we then obtain, after some reshuffling of terms,

$$d\pi_t = \pi_{t-}\left\{r - g_t\lambda^Q\right\}dt + g_t\pi_{t-}dN_t. \tag{8.22}$$

The V dynamics above can be written in more detail as

$$dV_t = V_{t-}\left\{u_t^S\alpha + u_t^B r\right\}dt + V_{t-}u_t^S\beta dN_t, \tag{8.23}$$

and, comparing (8.22) with (8.23), we can now identify u^S from the dN term. More formally, let us define u^S by

$$u_t^S = \frac{g_t}{\beta}.$$

We can then write (8.22) as

$$d\pi_t = \pi_{t-}\left\{r - g_t\frac{r-\alpha}{\beta}\right\}dt + \pi_{t-}u_t^S\beta dN_t.$$

i.e.

$$d\pi_t = \pi_{t-}\left\{u_t^S\alpha + (1-u_t^S)r\right\}dt + \pi_{t-}u_t^S\beta dN_t.$$

and we see that if we define u^B by

$$u_t^B = 1 - u_t^S,$$

and define the process V by $V_t = \pi_t$ we obtain

$$dV_t = V_{t-}\left\{u_t^S\alpha + u_t^B r\right\}dt + V_{t-}u_t^S\beta dN_t.$$

which are the dynamics of a self financing portfolio with weights u^B and u^S (summing to unity), and we obviously have $V_T = X$.

Chapter 9

Jump Diffusion Models

9.1 Introduction

The model in the previous chapter was an extremely unrealistic one, and we now go on to present a more realistic model driven by a Wiener process as well as by an independent Poisson process.

Formally we consider a stochastic basis $(\Omega, \mathcal{F}, P, \mathbf{F})$ carrying a standard Wiener process W as well as an independent Poisson Process N with constant P-intensity λ. The market consists of a risky asset S and a bank account B with constant short rate. The asset dynamics are given by

$$dS_t = \alpha S_t dt + \sigma S_t dW_t + \beta S_{t-} dN_t, \tag{9.1}$$

$$dB_t = rB_t dt, \tag{9.2}$$

where α, σ, β, and the short rate r are known constants.

Exactly like in the previous section we can interpret α as the local mean rate of return **between jumps**, and β as the relative jump size. In particular this implies that if S is the price of a common stock and thus non negative, then we must have $\beta \geq -1$. To obtain the rate or return including jumps we write

$$dN_t = \lambda dt + dM_t,$$

where M is a P martingale, and substitute this into (9.1) to obtain

$$dS_t = \{\alpha + \beta\lambda\} S_t dt + \sigma S_t dW_t + \beta S_{t-} dM_t.$$

Since the dW and the dM terms are martingales this implies that the local mean rate of return **including jumps** under P is given by

$$\mu^P = \alpha + \beta\lambda.$$

Before we go on to perform and concrete calculations, let us informally discuss what we may expect from the model above. Referring to the Metatheorem

7.5.1 we see that we have one risky asset S, so $N = 1$, and two sources of randomness W and N, so $R = 2$. From the Metatheorem we thus conclude that we may expect the model to be arbitrage free but **not complete**. In particular we should that the martingale measure is not unique, that there will not be unique arbitrage free prices for financial derivatives, and that it will not be possible to form a riks free portfolio based on a derivative and the underlying asset.

9.2 Classical technique

Vi now go on to study derivatives pricing using the methodology of risk free portfolios developed in Section 8.2. To this end we consider a T-claim X of the form

$$X = \Phi(S_T),$$

and we assume that this derivative asset is traded on a liquid market with a price process of the form

$$\Pi(t; X) = F(t, S_t).$$

Our job is to see what we can say about the pricing function F, given the requirement of an arbitrage free market. In the first round we just try to copy the arguments from Section 8.2 so we try to form a risk free portfolio based on S and F. To do this we need the price dynamics of the derivative, and from the Itô formula we have

$$dF = \alpha_F F dt + \sigma_F dW + \beta_F^- F^- dN_t, \tag{9.3}$$

where upper case index, like in F^-, denotes evaluation at $(t-, S_{t-})$. The coefficients are given by

$$\alpha_F(t,s) = \frac{F_t(t,s) + \alpha s F_s(t,s) + \frac{1}{2}\sigma^2 s^2 F_{ss}(t,s)}{F(t,s)}, \tag{9.4}$$

$$\sigma_F(t,s) = \frac{\sigma s F_s(t,s)}{F(t,s)}, \tag{9.5}$$

$$\beta_F(t,s) = \frac{F_\beta(t,s)}{F(t,s)}. \tag{9.6}$$

Here we have used the notation $F_t = \frac{\partial F}{\partial t}$ and similarly for F_s and F_{ss}. The function F_β is given by

$$F_\beta(t,s) = F(t, s + \beta s) - F(T, s). \tag{9.7}$$

If we now form a self financing portfolio based on S and F, we obtain the following dynamics of the value process V.

$$dV = V^- \left\{ u^S \frac{dS}{S^-} + u^F \frac{dF}{F^-} \right\}.$$

Inserting the expression for dF above and collecting terms we obtain

$$\begin{aligned} dV &= V^{-}\left\{u^S\alpha + u^F\alpha_F^{-}\right\}dt + V^{-}\left\{u^S\sigma + u^F\sigma_F^{-}\right\}dW \\ &+ V^{-}\left\{u^S\beta + u^F\beta_F^{-}\right\}dN. \end{aligned}$$

We now want to balance the portfolio in such a way that it becomes locally risk free, i.e. we want to choose the portfolio weights u^S and u^F such that the dW and the dN terms vanish. Recalling that the weights must sum to unity we then have the following system of equations

$$\begin{aligned} u^S\sigma + u^F\sigma_F^{-} &= 0, \\ u^S\beta + u^F\beta_F^{-} &= 0, \\ u^S + u^F &= 1. \end{aligned}$$

This system is, however, overdetermined since we have two unknowns and three equations, so in general it will not have a solution.

The economic reason for this is clear. If we want to hedge the derivative by using the underlying asset, then we have only one instrument (the stock) to hedge two sources of randomness (W and N).

We can summarize the situation as follows.

- The price of a particular derivative Φ will **not** be completely determined by the specification of the S-dynamics and the requirement that the market (B, S, F) is free of arbitrage.
- The reason for this fact is that arbitrage pricing is always a case of pricing a derivative **in terms of** the price of some underlying assets. In our market we do not have sufficiently many underlying assets.

Thus we will not obtain a unique price of a particular derivative. This fact does not mean, however, that prices of various derivatives can take any form whatsoever. From the discussion above we see that the reason for the incompleteness is that we do not have enough underlying assets, so if we adjoin one more asset to the market, without introducing any new Wiener or Poisson processes, then we expect the market to be complete. This idea can be expressed in the following ways.

- We **cannot** say anything about the price of any **particular** derivative.
- The requirement of an arbitrage free derivative market implies that **prices of different derivatives** (i.e. claims with different contract functions or different times of expiration) will have to satisfy certain **internal consistency relations** in order to avoid arbitrage possibilities on the derivatives market.

We now go on to investigate these internal consistency relations so we assume that, apart form the claim $X = \Phi(S_T)$ there is also another T-claim $Y = \Gamma(S_T)$ traded on the market, and we assume that the price of Y is of the form

$$\Pi(t; Y) = G(t, S_t),$$

for some pricing function G. We assume that the market (B, S, F, G) is free of arbitrage and we now form a portfolio based on these assets, with predictable weights u^B, u^S, u^F, u^G. Since the weights must sum to unity we can write

$$u^B = 1 - u^s - u^F - u^G,$$

where u^S, u^F, u^G can be choose without constraints. The corresponding value dynamics are then given by

$$dV = V^- \left\{ (1 - u^s - u^F - u^G)\frac{dB}{B} + u^S \frac{dS}{S^-} + u^F \frac{dF}{F^-} + u^G \frac{dG}{G^-} \right\}$$

The differential dF is already given by (9.3)-(9.6), and we will of course have exactly the same structure for the differential dG. Collecting the various terms, we obtain

$$\begin{aligned} dV &= V^- \left\{ r + (\alpha - r)u^S + (\alpha_F^- - r)u^F + (\alpha_G^- - r)u^G \right\} dt \\ &+ V^- \left\{ \sigma u^S + \sigma_F^- u^F + \sigma_G^- u^G \right\} dW_t \\ &+ V^- \left\{ \beta u^S + \beta_F^- u^F + \beta_G^- u^G \right\} dN_t \end{aligned}$$

If we now choose the weights such that the dW and dN terms vanish we obtain the system

$$\begin{aligned} \sigma u^S + \sigma_F^- u^F + \sigma_G^- u^G &= 0, \\ \beta u^S + \beta_F^- u^F + \beta_G^- u^G &= 0 \end{aligned}$$

With such a a choice of weights, the portfolio becomes locally risk free, and absence of arbitrage now implies that we must also have

$$r + (\alpha - r)u^S + (\alpha_F^- - r)u^F + (\alpha_G^- - r)u^G = r,$$

or equivalently

$$(\alpha - r)u^S + (\alpha_F^- - r)u^F + (\alpha_G^- - r)u^G = 0.$$

The result of all this is that absence of arbitrage on the derivatives market implies that the system

$$\begin{aligned} (\alpha - r)u^S + (\alpha_F^- - r)u^F + (\alpha_G^- - r)u^G &= 0, \\ \sigma u^S + \sigma_F^- u^F + \sigma_G^- u^G &= 0, \\ \beta u^S + \beta_F^- u^F + \beta_G^- u^G &= 0 \end{aligned}$$

admits a **non trival** solution. (The trivial solution $u^S = u^F = u^G = 0$ corresponds to putting all the money in the bank). This is equivalent to saying that the coefficient matrix

$$\begin{bmatrix} \alpha - r & \alpha_F - r & \alpha_G - r \\ \sigma & \sigma_F & \sigma_G \\ \beta & \beta_F & \beta_G \end{bmatrix}$$

is singular. This, in turn, implies that the rows must be linearly dependent, so there must exists functions $\gamma_0(t,s)$ (the lower case index will soon disappear) and $\varphi(t,s)$ such that

$$\begin{aligned} \alpha - r &= \varphi\sigma + \gamma_0\beta, && (9.8) \\ \alpha_F - r &= \varphi\sigma_F + \gamma_0\beta_F, && (9.9) \\ \alpha_G - r &= \varphi\sigma_G + \gamma_0\beta_G. && (9.10) \end{aligned}$$

This system allows a natural economic interpretation, and to see this we recall that α is the local mean rate of return for the stock **excluding jumps**. The local mean rate of return for the stock **including jumps** is given by

$$\mu^P = \alpha + \beta\lambda,$$

and in the same way the local mean rates of return for the F and G contracts are given by μ_F^P and μ_G^P, where

$$\begin{aligned} \mu_F^P &= \alpha_F + \beta_F\lambda, && (9.11) \\ \mu_G^P &= \alpha_G + \beta_G\lambda. && (9.12) \end{aligned}$$

Defining the function h by

$$\gamma(t,s) = \gamma_0(t,s) - \lambda$$

we can write (9.8)-(9.10) as

$$\begin{aligned} \mu^P - r &= \varphi\sigma + \gamma\beta, && (9.13) \\ \mu_F^P - r &= \varphi\sigma_F + \gamma\beta_F, && (9.14) \\ \mu_G^P - r &= \varphi\sigma_G + \gamma\beta_G. && (9.15) \end{aligned}$$

These equations allow a very natural economic interpretation. On the left hand side we have the **risk premium** for the assets S, F, and G. On the right hand side we have a sum of the diffusion and jump volatilities multiplied by the coefficients φ and γ respectively. The main point to observe is that whereas the risk premium, the diffusion volatility, and the jump volatility vary from asset to asset, the coefficients φ and γ are **the same for all assets**. We can thus interpret φ as "the risk premium per unit of diffusion volatility", and γ as "the risk premium per unit of jump volatility". A less precise, but very common, way of referring to φ and γ is that

$$\begin{aligned} \varphi &= \text{"the market price of diffusion risk"}, \\ \gamma &= \text{"the market price of jump risk"}. \end{aligned}$$

We have now proved out first result.

Proposition 9.2.1 *Consider the model (9.1)-(9.2) and assume absence of arbitrage on the derivatives market. The there will exist functions $\varphi(t,s)$ and*

$\gamma(t,s)$ such that, for any claim of the form $\Phi(S_T)$ with pricing function $F(t,s)$ the following condition will hold

$$\mu_F^P - r = \varphi\sigma_F + \gamma\beta_F, \tag{9.16}$$

where the local mean rate of return μ_F^P, the diffusion volatility σ_F, and the jump volatility β_F are defined by (9.4)-(9.6), and (9.16). The functions φ and γ are **universal** *in the sense that they do not depend on the particular choice of the derivative. In particular we have*

$$\alpha + \lambda\beta - r = \varphi\sigma + \gamma\beta. \tag{9.17}$$

We may also use the relations above to obtain an equation for the F function. Plugging the definitions of μ_F^P, α_F, σ_F and β_F into (9.16) and using (9.17) gives us the following pricing result.

Proposition 9.2.2 *The pricing function F for the contract $\Phi(S_T)$ will satisfy the PIDE*

$$\begin{cases} F_t + \{r - \beta(\lambda-\gamma)\}\, sF_s + \dfrac{1}{2}\sigma^2 s^2 F_{ss} + (\lambda-\gamma)F_\beta - rF &= 0, \\ F(T,s) &= \Phi(s), \end{cases} \tag{9.18}$$

where $F_t = \frac{\partial F}{\partial t}$ etc.

This is the pricing PIDE for F, but in order to provide an explicit solution we need to know the market price of jump risk γ, and this object is not given *a prior* neither is it determined within the model. The reason for this si of course that our model is incomplete, so there are infinitely many market prices of risk which are consistent with no arbitrage. Thus there are also potentially infinitely many arbitrage free price processes for the contract Φ. In a concrete market, exactly one of these price processes will be chosen by the market, and this process will be determined, not only on the requirement of absence of arbitrage, but also by the preferences towards risk on the market. These preferences are then codified in the the market choice of the market price of risk γ. Obviously the same argument goes for the market price of diffusion risk φ, but the (9.17) allows us to express φ in terms of γ. We could of course also express γ in terms of φ, and we would then obtain a pricing equation involving φ but not γ. A more precise statement is therefore that the market chooses φ and γ subject to the constraint (9.17).

We can of course also apply the Feynman-Kac representation formula to (9.18). This gives us the following risk neutral pricing formula.

Proposition 9.2.3 *With notations as above, the pricing function F can also be represented by the following risk neutral valuation formula.*

$$F(t,s) = e^{-r(T-t)} E_{t,s}^{Q}\left[\Phi(S_T)\right] \tag{9.19}$$

The dynamics of S under Q are given by

$$dS_t = (\alpha - \varphi\sigma - \gamma\beta)\, S_t dt + S_t \sigma dW_t^Q + \beta S_{t-} dN_t, \tag{9.20}$$

where W^Q is Q-Wiener and N is Poisson under Q with intensity of $\lambda - \gamma$.

Remark 9.2.1 *By using (9.17) we can of course also write the dt term of S as $\{r - \beta(\lambda - \gamma)\}$, but the notation $(\alpha - \varphi\sigma - \gamma\beta)$ emphasizes the dependence on both φ and γ more clearly.*

The next proposition shows that, as expected, the measure Q above is indeed the martingale measure with B as numeraire.

Proposition 9.2.4 *The measure Q above has the property that the processes S_t/B_t as well as $F(t, S_t)/B_t$ are martingales under Q.*

Proof. Exercise for the reader. ∎

9.3 Martingale analysis

We now go on to study the jump diffusion model

$$\begin{aligned} dS_t &= \alpha S_t dt + \sigma S_t dW_t + \beta S_{t-} dN_t, & (9.21)\\ dB_t &= r B_t dt, & (9.22) \end{aligned}$$

from the point of view of martingale measures. This turns out to be very easy, and as usual we start looking for a potential martingale measure by applying the Girsanov Theorem. We thus choose two predictable processes g and h with $h > -1$, define the likelihood process L by

$$\begin{cases} dL_t = L_t g_t dW_t + L_{t-} h_t \{dN_T - \lambda dt\}, \\ L_0 = 1 \end{cases}, \tag{9.23}$$

and define a new measure Q by

$$L_t = \frac{dQ}{dP}, \quad \text{on } \mathcal{F}_t,\ 0 \le t \le T.$$

From the Girsanov Theorem we know that N has Q intensity

$$\lambda_t^Q = (1 + h_t)\lambda, \tag{9.24}$$

and that we can write

$$dW_t = g_t dt + dW_t^Q, \tag{9.25}$$

where W^Q is Q-Wiener. Plugging (9.25) into the S dynamics (9.21) and compensating N under Q gives us the Q dynamics of S as

$$\begin{aligned} dS_t &= \{\alpha + g_t\sigma + (1+h_t)\beta\lambda\} S_t dt + \sigma S_t dW_t^Q && (9.26) \\ &+ \beta S_- \{dN_t - (1+h_t)\lambda dt\} && (9.27) \end{aligned}$$

From this we see that Q is a martingale measure if and only if the relation

$$\alpha + g_t\sigma + (1+h_t)\beta\lambda = r, \tag{9.28}$$

is satisfied, and we can write this as

$$\alpha + \beta\lambda - r = -g_t\sigma - h_t\lambda\beta. \tag{9.29}$$

This equation is of course the same equation as (9.17) and we have the following result.

Proposition 9.3.1 *The measure Q above is a martingale measure if and only if the following conditions are satisfied.*

$$\begin{aligned} h_t &> -1, && (9.30) \\ \alpha + \beta\lambda - r &= -g_t\sigma - h_t\lambda\beta, && (9.31) \\ E^P[L_T] &= 1. && (9.32) \end{aligned}$$

Furthermore, the Girsanov kernels g and h are related to the market price of diffusion risk φ and the market price of jump risk γ by

$$\begin{aligned} g_t &= -\varphi(t-, S_{t-}), && (9.33) \\ h_t &= -\frac{\gamma(t-, S_{t-})}{\lambda}. && (9.34) \end{aligned}$$

Chapter 10

List of topics to be added

The following topics will be added later.

- Diversification and Arbitrage Pricing Theory (APT)
- Marked point processes
- Optimal control theory with finance applications
- Interest rate theory in the presence of jumps
- Credit risk theory
- Nonlinear filtering
- Queuing theory and Jackson networks

Bibliography

[1] Björk, T. *Arbitrage Theory in Continuous Time*, 2 ed. Oxford University Press, 2004.

[2] Brémaud, P. *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, 1981.

[3] Cont, R., and Tankov, P. *Finanical Modelling with Jump Processes*. Chapman and Hall /CRC, 2003.