1 Counting Processes

In this chapter we introduce the concept of a counting process. For references to the literature on counting processes and more general point processes, see the Notes at the end of the chapter.

1.1 Generalities and the Poisson Process

Our main objective is to study point processes on the real positive line, and the simplest type of a point process is a *counting process*. The formal definition is as follows.

Definition 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
$$\Delta N_t = 0, \quad \text{or} \quad \Delta N_t = 1.$$
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 at 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 have occurred in the time interval $[0, t]$. Thus N_t could be the number of customers who have arrived at 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 occupies very much the same place that the Wiener process does among diffusion processes. We start with some elementary facts concerning the Poisson distribution.

Definition 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 $X \sim \text{Po}(\alpha)$.

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 1.3 *Let X be* $\text{Po}(\alpha)$. *Then the characteristic function is given by*

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

The mean and variance are given by

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

Proof This is left as an exercise. □

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

Definition 1.4 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 real number with $\lambda > 0$. 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

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

In concrete terms this says that the increment $N_t - N_s$ is Poisson with parameter $\lambda(T-s)$ and independent of $\mathcal{F}_s$. In the definition above 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

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

The increment process dN is informally interpreted as

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

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 $N_{t+dt} - N_t$ is that we want the increment process dN to be adapted. The term "predictable" will be very important later on, 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

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

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

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

Expanding the exponential we thus have

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

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

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

or equivalently

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

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.
- 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 difference

$$dN_t - \lambda dt,$$

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 1.5 *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 We have to show that $E\left[N_t - N_s | \mathcal{F}_s\right] = \lambda(t-s)$. This however follows directly from the fact that the conditional distribution of $N_t - N_s$, given $\mathcal{F}_s$, is Poisson with parameter $\lambda(t-s)$. □

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. Indeed, 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 an arbitrary counting process and if the process M, defined 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).

Furthermore, it 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 toolbox 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 to have some nice measurability property.
2. In particular, if M is a martingale of bounded variation, we would like to know 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 with 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 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 optimal control, and arbitrage theory for economies where asset prices are driven by jump diffusions.

1.3 Notes

The textbook Brémaud (1981) is a classic in the field. The monograph Last & Brandt (1995) contains an almost encyclopedic study of (marked) point processes. In Cont & Tankov (2003) the reader will find an in depth study of Lévy processes and their applications to finance. For general semimartingale theory see Cohen & Elliott (2015), Jacod & Shiryaev (1987), or Protter (2004).

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, derive the relevant Itô formula, 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 pathwise, as opposed to the Itô integral which has to be defined as an L^2 limit.
- On the other hand, compared 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 briefly discuss the simplest type of stochastic integration, namely integration of discrete-time processes. This will 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 2.1 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 Y, the **increment process** ΔY is defined by
$$(\Delta Y)_n = Y_n - Y_{n-1}, \tag{2.1}$$
with the convention $Y_{-1} = 0$. For simplicity of notation we will sometimes denote $(\Delta Y)_n$ by ΔY_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 ΔY by "backward increments" as $(\Delta Y)_n = Y_n - Y_{n-1}$, instead of "forward increments" $(\Delta Y)_n = Y_{n+1} - Y_n$, is that by using backwards increments the process ΔY is adapted whenever Y 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, then 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 would hold for the discrete-time integral, but this is not the case. Indeed, 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

- 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 with respect to a martingale M, then the result is a new martingale.

Proposition 2.3 *Assume that the filtered probability 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

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

Thus, defining Z as

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

it follows that

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

and we obtain

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

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 discrete-time version of the predictability concept.

Definition 2.4

- 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

$$\int_0^T |dA_t| < \infty \quad P\text{-a.s.}$$

- 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 $T < \infty$ 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 $T < \infty$ is denoted by $\mathcal{A}$. Such a process is said to be of **integrable variation**.

Remark Note that both the cadlag property, and the property of being adapted, are parts of 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 either by $X(t, \omega)$ or by $X_t(\omega)$.

Definition 2.5 The **optional** σ-algebra on $R_+ \times \Omega$ is generated by all processes Y of the form

$$Y_t(\omega) = Z(\omega) I\{r \leq t < s\}, \tag{2.3}$$

where I is the indicator function, r and s are fixed real numbers, and Z is an $\mathcal{F}_r$ measurable random variable. A process X which, viewed as a mapping $X : \Omega \times R_+ \to R$, is measurable with respect to 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 2.6 *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 2.7 *Assume that A is of finite variation and that h is an optional process satisfying the condition*

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

Then the following assertions 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, or by using a monotone class argument. □

Remark Note again that since A is of finite variation it is, by definition, 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 fresh in the 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 counterexample.

Example Let Z be a non-trivial 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

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

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