Jump Processes

Richard F. Bass

Contents

Chapter 1

Poisson processes

1.1 Definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A filtration is a collection of σ-fields $\mathcal{F}_t$ contained in $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s < t$. A filtration satisfies the usual conditions if it is complete: $N \in \mathcal{F}_t$ for all t whenever $\mathbb{P}(N) = 0$ and it is right continuous: $\mathcal{F}_{t+} = \mathcal{F}_t$ for all t, where $\mathcal{F}_{t+} = \cap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}$.

Definition 1.1 *Let $\{\mathcal{F}_t\}$ be a filtration, not necessarily satisfying the usual conditions. A* Poisson process *with parameter $\lambda > 0$ is a stochastic process X satisfying the following properties:*

(1) $X_0 = 0$, a.s.

(2) The paths of X_t are right continuous with left limits.

(3) If $s < t$, then $X_t - X_s$ is a Poisson random variable with parameter $\lambda(t - s)$.

(4) If $s < t$, then $X_t - X_s$ is independent of $\mathcal{F}_s$.

Define $X_{t-} = \lim_{s \to t, s < t} X_s$, the left hand limit at time t, and $\Delta X_t = X_t - X_{t-}$, the size of the jump at time t. We say a function f is increasing if $s < t$ implies $f(s) \leq f(t)$. We use 'strictly increasing' when $s < t$ implies $f(s) < f(t)$. We have the following proposition.

Proposition 1.2 *Let X be a Poisson process. With probability one, the paths of X_t are increasing and are constant except for jumps of size 1. There are only finitely many jumps in each finite time interval.*

Proof. For any fixed $s < t$, we have that $X_t - X_s$ has the distribution of a Poisson random variable with parameter $\lambda(t-s)$, hence is non-negative, a.s.; let $N_{s,t}$ be the null set of ω's where $X_t(\omega) < X_s(\omega)$. The set of pairs (s,t) with s and t rational is countable, and so $N = \cup_{s,t\in\mathbb{Q}_+} N_{s,t}$ is also a null set, where we write $\mathbb{Q}_+$ for the non-negative rationals. For $\omega \notin N$, $X_t \geq X_s$ whenever $s < t$ are rational. In view of the right continuity of the paths of X, this shows the paths of X are increasing with probability one.

Similarly, since Poisson random variables only take values in the non-negative integers, X_t is a non-negative integer, a.s. Using this fact for every t rational shows that with probability one, X_t takes values only in the non-negative integers when t is rational, and the right continuity of the paths implies this is also the case for all t. Since the paths have left limits, there can only be finitely many jumps in finite time.

It remains to prove that ΔX_t is either 0 or 1 for all t. Let $t_0 > 0$. If there were a jump of size 2 or larger at some time t strictly less than t_0, then for each n sufficiently large there exists $0 \leq k_n \leq 2^n$ such that $X_{(k_n+1)t_0/2^n} - X_{k_n t_0/2^n} \geq 2$. Therefore

$$\begin{aligned}
\mathbb{P}(\exists s < t_0 : \Delta X_s \geq 2) &\leq \mathbb{P}(\exists k \leq 2^n : X_{(k+1)t_0/2^n} - X_{kt_0/2^n} \geq 2) \\
&\leq 2^n \sup_{k\leq 2^n} \mathbb{P}(X_{(k+1)t_0/2^n} - X_{kt_0/2^n} \geq 2) \\
&= 2^n \mathbb{P}(X_{t_0/2^n} \geq 2^n) \\
&\leq 2^n (1 - \mathbb{P}(X_{t_0/2^n} = 0) - \mathbb{P}(X_{t_0/2^n} = 1)) \\
&= 2^n \Big(1 - e^{-\lambda t_0/2^n} - (\lambda t_0/2^n) e^{-\lambda t_0/2^n}\Big).
\end{aligned} \tag{1.1}$$

We used Definition 1.1(3) for the two equalities. By l'Hôpital's rules, $(1 - e^{-x} - xe^{-x})/x \to 0$ as $x \to 0$. We apply this with $x = \lambda t_0/2^n$, and see that the last line of (1.1) tends to 0 as $n \to \infty$. Since the left hand side of (1.1) does not depend on n, it must be 0. This holds for each t_0. □

1.2 Stopping times

Throughout this section we suppose we have a filtration $\{\mathcal{F}_t\}$ satisfying the usual conditions.

Definition 1.3 *A random variable $T : \Omega \to [0, \infty]$ is a* stopping time *if for all t, $(T < t) \in \mathcal{F}_t$. We say T is a finite stopping time if $T < \infty$, a.s. We say T is a bounded stopping time if there exists $K \in [0, \infty)$ such that $T \le K$, a.s.*

Note that T can take the value infinity. Stopping times are also known as *optional times*.

Given a stochastic process X, we define $X_T(\omega)$ to be equal to $X(T(\omega), \omega)$, that is, for each ω we evaluate $t = T(\omega)$ and then look at $X(\cdot, \omega)$ at this time.

Proposition 1.4 *Suppose $\mathcal{F}_t$ satisfies the usual conditions. Then*

(1) T is a stopping time if and only if $(T \le t) \in \mathcal{F}_t$ for all t.

(2) If $T = t$, a.s., then T is a stopping time.

(3) If S and T are stopping times, then so are $S \vee T$ and $S \wedge T$.

(4) If T_n, $n = 1, 2, \ldots$, are stopping times with $T_1 \le T_2 \le \cdots$, then so is $\sup_n T_n$.

(5) If T_n, $n = 1, 2, \ldots$, are stopping times with $T_1 \ge T_2 \ge \cdots$, then so is $\inf_n T_n$.

(6) If $s \ge 0$ and S is a stopping time, then so is $S + s$.

Proof. We will just prove part of (1), leaving the rest as an exercise. Note $(T \le t) = \cap_{n \ge N}(T < t + 1/n) \in \mathcal{F}_{t+1/N}$ for each N. Thus $(T \le t) \in \cap_N \mathcal{F}_{t+1/N} \subset \mathcal{F}_{t+} = \mathcal{F}_t$. $\square$

It is often useful to be able to approximate stopping times from the right. If T is a finite stopping time, that is, $T < \infty$, a.s., define

$$T_n(\omega) = (k+1)/2^n \qquad \text{if } k/2^n \le T(\omega) < (k+1)/2^n. \tag{1.2}$$

Define

$$\mathcal{F}_T = \{A \in \mathcal{F} : \text{ for each } t > 0,\ A \cap (T \le t) \in \mathcal{F}_t\}. \tag{1.3}$$

This definition of $\mathcal{F}_T$, which is supposed to be the collection of events that are "known" by time T, is not very intuitive. But it turns out that this definition works well in applications.

Proposition 1.5 *Suppose $\{\mathcal{F}_t\}$ is a filtration satisfying the usual conditions.*

(1) $\mathcal{F}_T$ is a σ-field.

(2) If $S \le T$, then $\mathcal{F}_S \subset \mathcal{F}_T$.

(3) If $\mathcal{F}_{T+} = \cap_{\varepsilon>0}\mathcal{F}_{T+\varepsilon}$, then $\mathcal{F}_{T+} = \mathcal{F}_T$.

(4) If X_t has right continuous paths, then X_T is $\mathcal{F}_T$-measurable.

Proof. If $A \in \mathcal{F}_T$, then $A^c \cap (T \le t) = (T \le t) \setminus [A \cap (T \le t)] \in \mathcal{F}_t$, so $A^c \in \mathcal{F}_T$. The rest of the proof of (1) is easy.

Suppose $A \in \mathcal{F}_S$ and $S \le T$. Then $A \cap (T \le t) = [A \cap (S \le t)] \cap (T \le t)$. We have $A \cap (S \le t) \in \mathcal{F}_t$ because $A \in \mathcal{F}_S$, while $(T \le t) \in \mathcal{F}_t$ because T is a stopping time. Therefore $A \cap (T \le t) \in \mathcal{F}_t$, which proves (2).

For (3), if $A \in \mathcal{F}_{T+}$, then $A \in \mathcal{F}_{T+\varepsilon}$ for every ε, and so $A \cap (T+\varepsilon \le t) \in \mathcal{F}_t$ for all t. Hence $A \cap (T \le t-\varepsilon) \in \mathcal{F}_t$ for all t, or equivalently $A \cap (T \le t) \in \mathcal{F}_{t+\varepsilon}$ for all t. This is true for all ε, so $A \cap (T \le t) \in \mathcal{F}_{t+} = \mathcal{F}_t$. This says $A \in \mathcal{F}_T$.

(4) Define T_n by (1.2). Note

$$(X_{T_n} \in B) \cap (T_n = k/2^n) = (X_{k/2^n} \in B) \cap (T_n = k/2^n) \in \mathcal{F}_{k/2^n}.$$

Since T_n only takes values in $\{k/2^n : k \ge 0\}$, we conclude $(X_{T_n} \in B) \cap (T_n \le t) \in \mathcal{F}_t$, and so $(X_{T_n} \in B) \in \mathcal{F}_{T_n} \subset \mathcal{F}_{T+1/2^n}$. Hence X_{T_n} is $\mathcal{F}_{T+1/2^n}$ measurable. If $n \ge m$, then X_{T_n} is measurable with respect to $\mathcal{F}_{T+1/2^n} \subset \mathcal{F}_{T+1/2^m}$. Since $X_{T_n} \to X_T$, then X_T is $\mathcal{F}_{T+1/2^m}$ measurable for each m. Therefore X_T is measurable with respect to $\mathcal{F}_{T+} = \mathcal{F}_T$. $\square$

1.3 Markov properties

Let us begin with the Markov property.

Theorem 1.6 *Let $\{\mathcal{F}_t\}$ be a filtration, not necessarily satisfying the usual conditions, and let P be a Poisson process with respect to $\{\mathcal{F}_t\}$. If u is a fixed time, then $Y_t = P_{t+u} - P_u$ is a Poisson process independent of $\mathcal{F}_u$.*

Proof. Let $\mathcal{G}_t = \mathcal{F}_{t+u}$. It is clear that Y has right continuous paths, is zero at time 0, has jumps of size one, and is adapted to $\{\mathcal{G}_t\}$. Since $Y_t - Y_s = P_{t+u} - P_{s+u}$, then $Y_t - Y_s$ is a Poisson random variable with mean $\lambda(t-s)$ that is independent of $\mathcal{F}_{s+u} = \mathcal{G}_s$. $\square$

The strong Markov property is the Markov property extended by replacing fixed times u by finite stopping times.

Theorem 1.7 *Let $\{\mathcal{F}_t\}$ be a filtration, not necessarily satisfying the usual conditions, and let P be a Poisson process adapted to $\{\mathcal{F}_t\}$. If T is a finite stopping time, then $Y_t = P_{T+t} - P_T$ is a Poisson process independent of $\mathcal{F}_T$.*

Proof. We will first show that whenever $m \geq 1$, $t_1 < \cdots < t_m$, f is a bounded continuous function on $\mathbb{R}^m$, and $A \in \mathcal{F}_T$, then

$$\mathbb{E}\left[f(Y_{t_1}, \ldots, Y_{t_m}); A\right] = \mathbb{E}\left[f(P_{t_1}, \ldots, P_{t_m})\right] \mathbb{P}(A). \tag{1.4}$$

Once we have done this, we will then show how (1.4) implies our theorem.

To prove (1.4), define T_n by (1.2). We have

$$\begin{aligned}
&\mathbb{E}\left[f(P_{T_n+t_1} - P_{T_n}, \ldots, P_{T_n+t_m} - P_{T_n}); A\right] \qquad (1.5)\\
&\quad = \sum_{k=1}^{\infty} \mathbb{E}\left[f(P_{T_n+t_1} - P_{T_n}, \ldots, P_{T_n+t_m} - P_{T_n}); A, T_n = k/2^n\right]\\
&\quad = \sum_{k=1}^{\infty} \mathbb{E}\left[f(P_{t_1+k/2^n} - P_{k/2^n}, \ldots, P_{t_m+k/2^n} - P_{k/2^n}); A, T_n = k/2^n\right].
\end{aligned}$$

Following the usual practice in probability that "," means "and," we use "$\mathbb{E}\left[\cdots; A, T_n = k/2^n\right]$" as an abbreviation for "$\mathbb{E}\left[\cdots; A \cap (T_n = k/2^n)\right]$." Since $A \in \mathcal{F}_T$, then $A \cap (T_n = k/2^n) = A \cap ((T < k/2^n) \setminus (T < (k-1)/2^n)) \in \mathcal{F}_{k/2^n}$. We use the independent increments property of Poisson process and the fact that $P_t - P_s$ has the same law as P_{t-s} to see that the sum in the last line of (1.5) is equal to

$$\begin{aligned}\sum_{k=1}^{\infty} &\mathbb{E}\left[f(P_{t_1+k/2^n} - P_{k/2^n}, \ldots, P_{t_m+k/2^n} - P_{k/2^n})\right]\mathbb{P}(A, T_n = k/2^n)\\ &= \sum_{k=1}^{\infty} \mathbb{E}\left[f(P_{t_1}, \ldots, P_{t_m})\right]\mathbb{P}(A, T_n = k/2^n)\\ &= \mathbb{E}\left[f(P_{t_1}, \ldots, P_{t_m})\right]\mathbb{P}(A),\end{aligned}$$

which is the right hand side of (1.4). Thus

$$\begin{aligned}\mathbb{E}\left[f(P_{T_n+t_1} - P_{T_n}, \ldots, P_{T_n+t_m} - P_{T_n}); A\right]&\\ = \mathbb{E}\left[f(P_{t_1}, \ldots, P_{t_m})\right]\mathbb{P}(A).&\end{aligned} \tag{1.6}$$

Now let $n \to \infty$. By the right continuity of the paths of P, the boundedness and continuity of f, and the dominated convergence theorem, the left hand side of (1.6) converges to the left hand side of (1.4).

If we take $A = \Omega$ in (1.4), we obtain

$$\mathbb{E}\left[f(Y_{t_1}, \ldots, Y_{t_m})\right] = \mathbb{E}\left[f(P_{t_1}, \ldots, P_{t_m})\right]$$

whenever $m \geq 1$, $t_1, \ldots, t_m \in [0, \infty)$, and f is a bounded continuous function on $\mathbb{R}^m$. This implies that the finite dimensional distributions of Y and P are the same. Since Y has right continuous paths, Y is a Poisson process.

Next take $A \in \mathcal{F}_T$. By using a limit argument, (1.4) holds whenever f is the indicator of a Borel subset B of $\mathbb{R}^d$, or in other words,

$$\mathbb{P}(Y \in B, A) = \mathbb{P}(Y \in B)\mathbb{P}(A) \tag{1.7}$$

whenever B is a cylindrical set. □

When we discuss the Skorokhod topology, we will be able be more precise for the independence argument.

Observe that what was needed for the above proof to work is not that P be a Poisson process, but that the process P have right continuous paths and that $P_t - P_s$ be independent of $\mathcal{F}_s$ and have the same distribution as P_{t-s}. We therefore have the following corollary.

Corollary 1.8 *Let $\{\mathcal{F}_t\}$ be a filtration, not necessarily satisfying the usual conditions, and let X be a process adapted to $\{\mathcal{F}_t\}$. Suppose X has paths that are right continuous with left limits and suppose $X_t - X_s$ is independent of $\mathcal{F}_s$ and has the same law as X_{t-s} whenever $s < t$. If T is a finite stopping time, then $Y_t = X_{T+t} - X_T$ is a process that is independent of $\mathcal{F}_T$ and X and Y have the same law.*

1.4 A characterization

Another characterization of the Poisson process is as follows. Let $T_1 = \inf\{t : \Delta X_t = 1\}$, the time of the first jump. Define $T_{i+1} = \inf\{t > T_i : \Delta X_t = 1\}$, so that T_i is the time of the i^{th} jump.

Proposition 1.9 *The random variables $T_1, T_2 - T_1, \ldots, T_{i+1} - T_i, \ldots$ are independent exponential random variables with parameter λ.*

Proof. In view of Corollary 1.8 it suffices to show that T_1 is an exponential random variable with parameter λ. If $T_1 > t$, then the first jump has not occurred by time t, so X_t is still zero. Hence

$$\mathbb{P}(T_1 > t) = \mathbb{P}(X_t = 0) = e^{-\lambda t},$$

using the fact that X_t is a Poisson random variable with parameter λt. □

We can reverse the characterization in Proposition 1.9 to construct a Poisson process. We do one step of the construction, leaving the rest as an exercise.

Let $U_1, U_2, \ldots$ be independent exponential random variables with parameter λ and let $T_j = \sum_{i=1}^{j} U_i$. Define

$$X_t(\omega) = k \qquad \text{if } T_k(\omega) \le t < T_{k+1}(\omega). \tag{1.8}$$

An examination of the densities shows that an exponential random variable has a gamma distribution with parameters λ and $r = 1$, so T_j is a gamma random variable with parameters λ and j. Thus

$$\mathbb{P}(X_t < k) = \mathbb{P}(T_k > t) = \int_t^\infty \frac{\lambda e^{-\lambda x} (\lambda x)^{k-1}}{\Gamma(k)}\, dx.$$

Performing the integration by parts repeatedly shows that

$$\mathbb{P}(X_t < k) = \sum_{i=0}^{k-1} e^{-\lambda t} \frac{(\lambda t)^i}{i!},$$

and so X_t is a Poisson random variable with parameter λt.

We will use the following proposition later.

Proposition 1.10 *Let $\{\mathcal{F}_t\}$ be a filtration satisfying the usual conditions. Suppose $X_0 = 0$, a.s., X has paths that are right continuous with left limits, $X_t - X_s$ is independent of $\mathcal{F}_s$ if $s < t$, and $X_t - X_s$ has the same law as X_{t-s} whenever $s < t$. If the paths of X are piecewise constant, increasing, all the jumps of X are of size 1, and X is not identically 0, then X is a Poisson process.*

Proof. Let $T_0 = 0$ and $T_{i+1} = \inf\{t > T_i : \Delta X_t = 1\}$, $i = 1, 2, \ldots$. We will show that if we set $U_i = T_i - T_{i-1}$, then the U_i's are i.i.d. exponential random variables.

By Corollary 1.8, the U_i's are independent and have the same law. Hence it suffices to show U_1 is an exponential random variable. We observe

$$\begin{aligned}
\mathbb{P}(U_1 > s + t) &= \mathbb{P}(X_{s+t} = 0) = \mathbb{P}(X_{s+t} - X_s = 0, X_s = 0) \\
&= \mathbb{P}(X_{t+s} - X_s = 0)\mathbb{P}(X_s = 0) = \mathbb{P}(X_t = 0)\mathbb{P}(X_s = 0) \\
&= \mathbb{P}(U_1 > t)\mathbb{P}(U_1 > s).
\end{aligned}$$

Setting $f(t) = \mathbb{P}(U_1 > t)$, we thus have $f(t + s) = f(t)f(s)$. Since $f(t)$ is decreasing and $0 < f(t) < 1$, we conclude $\mathbb{P}(U_1 > t) = f(t) = e^{-\lambda t}$ for some $\lambda > 0$, or U_1 is an exponential random variable. $\square$

1.5 Martingales

We define continuous time martingales. Let $\{\mathcal{F}_t\}$ be a filtration, not necessarily satisfying the usual conditions.

Definition 1.11 *M_t is a continuous time* martingale *with respect to the filtration $\{\mathcal{F}_t\}$ and the probability measure $\mathbb{P}$ if*

(1) $\mathbb{E}\,|M_t| < \infty$ for each t;

(2) M_t is $\mathcal{F}_t$ measurable for each t;

(3) $\mathbb{E}\,[M_t \mid \mathcal{F}_s] = M_s$, a.s., if $s < t$.

Part (2) of the definition can be rephrased as saying M_t is adapted to $\mathcal{F}_t$. If in part (3) "=" is replaced by "≥," then M_t is a *submartingale*, and if it is replaced by "≤," then we have a *supermartingale*.

Taking expectations in Definition 1.11(3), we see that if $s < t$, then $\mathbb{E}\,M_s \le \mathbb{E}\,M_t$ is M is a submartingale and $\mathbb{E}\,M_s \ge \mathbb{E}\,M_t$ if M is a supermartingale. Thus submartingales tend to increase, on average, and supermartingales tend to decrease, on average.

If P_t is a Poisson process with index λ, then $P_t - \lambda t$ is a continuous time martingale. To see this,

$$\begin{aligned}
\mathbb{E}\,[P_t - \lambda t \mid \mathcal{F}_s] &= \mathbb{E}\,[P_t - P_s \mid \mathcal{F}_s] - \lambda t + P_s \\
&= \mathbb{E}\,[P_t - P_s] - \lambda t + P_s \\
&= \lambda(t - s) = \lambda t + P_s \\
&= P_s - \lambda s.
\end{aligned}$$

We give another example of a martingale.

Example 1.12 Recall that given a filtration $\{\mathcal{F}_t\}$, each $\mathcal{F}_t$ is contained in $\mathcal{F}$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is our probability space. Let X be an integrable $\mathcal{F}$ measurable random variable, and let $M_t = \mathbb{E}\,[X \mid \mathcal{F}_t]$. Then

$$\mathbb{E}\,[M_t \mid \mathcal{F}_s] = \mathbb{E}\,[\mathbb{E}\,[X \mid \mathcal{F}_t] \mid \mathcal{F}_s] = \mathbb{E}\,[X \mid \mathcal{F}_s] = M_s,$$

and M is a martingale.

We derive the analogs of Doob's inequalities in the stochastic process context.

Theorem 1.13 *Suppose M_t is a martingale or non-negative submartingale with paths that are right continuous with left limits. Then*

(1)
$$\mathbb{P}(\sup_{s\le t}|M_s|\ge\lambda)\le\mathbb{E}\,|M_t|/\lambda.$$

(2) If $1<p<\infty$, *then*
$$\mathbb{E}\,[\sup_{s\le t}|M_s|]^p\le\Big(\frac{p}{p-1}\Big)^p\mathbb{E}\,|M_t|^p.$$

Proof. We will do the case where M_t is a martingale, the submartingale case being nearly identical. Let $\mathcal{D}_n=\{kt/2^n:0\le k\le 2^n\}$. If we set $N_k^{(n)}=M_{kt/2^n}$ and $\mathcal{G}_k^{(n)}=\mathcal{F}_{kt/2^n}$, it is clear that $\{N_k^{(n)}\}$ is a discrete time martingale with respect to $\{\mathcal{G}_k^{(n)}\}$. Let $A_n=\{\sup_{s\le t,s\in\mathcal{D}_n}|M_s|>\lambda\}$. By Doob's inequality for discrete time martingales,
$$\mathbb{P}(A_n)=\mathbb{P}(\max_{k\le 2^n}|N_k^{(n)}|>\lambda)\le\frac{\mathbb{E}\,|N_{2^n}^{(n)}|}{\lambda}=\frac{\mathbb{E}\,|M_t|}{\lambda}.$$

Note that the A_n are increasing, and since M_t is right continuous,
$$\cup_nA_n=\{\sup_{s\le t}|M_s|>\lambda\}.$$

Then
$$\mathbb{P}(\sup_{s\le t}|M_s|>\lambda)=\mathbb{P}(\cup_nA_n)=\lim_{n\to\infty}\mathbb{P}(A_n)\le\mathbb{E}\,|M_t|/\lambda.$$

If we apply this with λ replaced by $\lambda-\varepsilon$ and let $\varepsilon\to0$, we obtain (1).

The proof of (2) is similar. By Doob's inequality for discrete time martingales,
$$\mathbb{E}\,[\sup_{k\le 2^n}|N_k^{(n)}|^p]\le\Big(\frac{p}{p-1}\Big)^p\mathbb{E}\,|N_{2^n}^{(n)}|^p=\Big(\frac{p}{p-1}\Big)^p\mathbb{E}\,|M_t|^p.$$

Since $\sup_{k\le 2^n}|N_k^{(n)}|^p$ increases to $\sup_{s\le t}|M_s|^p$ by the right continuity of M, (2) follows by Fatou's lemma. □

Here is an example. If P_t is a Poisson process of index λ, then $P_t-\lambda t$ is a martingale. So $e^{a(P_t-\lambda t)}$ is a submartingale for any real number a. Then
$$\mathbb{P}(\sup_{s\le t}P_s-\lambda s\ge A)=\mathbb{P}(\sup_{s\le t}e^{a(P_s-\lambda s}>\ge e^{aA})\le e^{-aA}\mathbb{E}\,e^{aP_t}e^{-a\lambda t}.$$

We know

$$\mathbb{E}\, e^{aP_t} = \exp\Big((e^a - 1)\lambda t\Big).$$

We substitute this in the above and then optimize over a.

We will need Doob's optional stopping theorem for continuous time martingales.

Theorem 1.14 *Let $\{\mathcal{F}_t\}$ be a filtration satisfying the usual conditions. If M_t is a martingale or non-negative submartingale whose paths are right continuous, $\sup_{t\geq 0} \mathbb{E}\, M_t^2 < \infty$, and T is a finite stopping time, then $\mathbb{E}\, M_T \geq \mathbb{E}\, M_0$.*

Proof. We do the submartingale case, the martingale case being very similar. By Doob's inequality (Theorem 1.13(1)),

$$\mathbb{E}\,[\sup_{s\leq t} M_s^2] \leq 4\mathbb{E}\, M_t^2.$$

Letting $t \to \infty$, we have $\mathbb{E}\,[\sup_{t\geq 0} M_t^2] < \infty$ by Fatou's lemma.

Let us first suppose that $T < K$, a.s., for some real number K. Define T_n by (1.2). Let $N_k^{(n)} = M_{k/2^n}$, $\mathcal{G}_k^{(n)} = \mathcal{F}_{k/2^n}$, and $S_n = 2^n T_n$. By Doob's optional stopping theorem applied to the submartingale $N_k^{(n)}$, we have

$$\mathbb{E}\, M_0 = \mathbb{E}\, N_0^{(n)} \leq \mathbb{E}\, N_{S_n}^{(n)} = \mathbb{E}\, M_{T_n}.$$

Since M is right continuous, $M_{T_n} \to M_T$, a.s. The random variables $|M_{T_n}|$ are bounded by $1+\sup_{t\geq 0} M_t^2$, so by dominated convergence, $\mathbb{E}\, M_{T_n} \to \mathbb{E}\, M_T$.

We apply the above to the stopping time $T \wedge K$ to get $\mathbb{E}\, M_{T\wedge K} \geq \mathbb{E}\, M_0$. The random variables $M_{T\wedge K}$ are bounded by $1+\sup_{t\geq 0} M_t^2$, so by dominated convergence, we get $\mathbb{E}\, M_T \geq \mathbb{E}\, M_0$ when we let $K \to \infty$. □

We present the continuous time version of Doob's martingale convergence theorem. We will see that not only do we get limits as $t \to \infty$, but also a regularity result.

Let $\mathcal{D}_n = \{k/2^n : k \geq 0\}$, $\mathcal{D} = \cup_n \mathcal{D}_n$.

Theorem 1.15 *Let $\{M_t : t \in \mathcal{D}\}$ be either a martingale, a submartingale, or a supermartingale with respect to $\{\mathcal{F}_t : t \in \mathcal{D}\}$ and suppose $\sup_{t\in\mathcal{D}} \mathbb{E}\, |M_t| < \infty$. Then*

(1) $\lim_{t\to\infty} M_t$ *exists, a.s.*

(2) With probability one M_t *has left and right limits along* $\mathcal{D}$.

The second conclusion says that except for a null set, if $t_0 \in [0,\infty)$, then both $\lim_{t\in\mathcal{D},t\uparrow t_0} M_t$ and $\lim_{t\in\mathcal{D},t\downarrow t_0} M_t$ exist and are finite. The null set does not depend on t_0.

Proof. Martingales are also submartingales and if M_t is a supermartingale, then $-M_t$ is a submartingale, so we may without loss of generality restrict our attention to submartingales. By Doob's inequality,

$$\mathbb{P}(\sup_{t\in\mathcal{D}_n,t\le n} |M_t| > \lambda) \le \frac{1}{\lambda}\mathbb{E}\,|M_n|.$$

Letting $n\to\infty$ and using Fatou's lemma,

$$\mathbb{P}(\sup_{t\in\mathcal{D}} |M_t| > \lambda) \le \frac{1}{\lambda}\sup_t \mathbb{E}\,|M_t|.$$

This is true for all λ, so with probability one, $\{|M_t| : t\in\mathcal{D}\}$ is a bounded set.

Therefore the only way either (1) or (2) can fail is that if for some pair of rationals $a<b$ the number of upcrossings of $[a,b]$ by $\{M_t : t\in\mathcal{D}\}$ is infinite. Recall that we define upcrossings as follows.

Given an interval $[a,b]$ and a submartingale M, if $S_1 = \inf\{t : M_t \le a\}$, $T_i = \inf\{t > S_i : M_t \ge b\}$, and $S_{i+1} = \inf\{t > T_i : M_t \le a\}$, then the number of upcrossings up to time u is $\sup\{k : T_k \le u\}$.

Doob's upcrossing lemma tells us that if V_n is the number of upcrossings by $\{M_t : t \in \mathcal{D}_n \cap [0,n]\}$, then

$$\mathbb{E}\,V_n \le \frac{\mathbb{E}\,|M_n|}{b-a}.$$

Letting $n\to\infty$ and using Fatou's lemma, the number of upcrossings of $[a,b]$ by $\{M_t : t\in\mathcal{D}\}$ has finite expectation, hence is finite, a.s. If $N_{a,b}$ is the null set where the number of upcrossings of $[a,b]$ by $\{M_t : t\in\mathcal{D}\}$ is infinite and $N = \cup_{a<b,a,b\in\mathbb{Q}_+} N_{a,b}$, where $\mathbb{Q}_+$ is the collection of non-negative rationals, then $\mathbb{P}(N) = 0$. If $\omega \notin N$, then (1) and (2) hold. $\square$

As a corollary we have

Corollary 1.16 *Let $\{\mathcal{F}_t\}$ be a filtration satisfying the usual conditions, and let M_t be a martingale with respect to $\{\mathcal{F}_t\}$. Then M has a version that is also a martingale and that in addition has paths that are right continuous with left limits.*

Proof. Let $\mathcal{D}$ be as in the above proof. For each integer $N \geq 1$, $\mathbb{E}\,|M_t| \leq \mathbb{E}\,|M_N| < \infty$ for $t \leq N$ since $|M_t|$ is a submartingale by the conditional expectation form of Jensen's inequality. Therefore $M_{t\wedge N}$ has left and right limits when taking limits along $t \in \mathcal{D}$. Since N is arbitrary, M_t has left and right limits when taking limits along $t \in \mathcal{D}$, except for a set of ω's that form a null set. Let

$$\widetilde{M}_t = \lim_{u\in\mathcal{D}, u>t, u\to t} M_u.$$

It is clear that $\widetilde{M}$ has paths that are right continuous with left limits. Since $\mathcal{F}_{t+} = \mathcal{F}_t$ and $\widetilde{M}_t$ is $\mathcal{F}_{t+}$ measurable, then $\widetilde{M}_t$ is $\mathcal{F}_t$ measurable.

Let N be fixed. We will show $\{M_t; t \leq N\}$ is a uniformly integrable family of random variables. Let $\varepsilon > 0$. Since M_N is integrable, there exists δ such that if $\mathbb{P}(A) < \delta$, then $\mathbb{E}\,[|M_N|; A] < \varepsilon$. If L is large enough, $\mathbb{P}(|M_t| > L) \leq \mathbb{E}\,|M_t|/L \leq \mathbb{E}\,|M_N|/L < \delta$. Then

$$\mathbb{E}\,[|M_t|; |M_t| > L] \leq \mathbb{E}\,[|M_N|; |M_t| > L] < \varepsilon,$$

since $|M_t|$ is a submartingale and $(|M_t| > L) \in \mathcal{F}_t$. Uniform integrability is proved.

Now let $t < N$. If $B \in \mathcal{F}_t$,

$$\mathbb{E}\,[\widetilde{M}_t; B] = \lim_{u\in\mathcal{D}, u>t, u\to t} \mathbb{E}\,[M_u; B] = \mathbb{E}\,[M_t; B].$$

Here we used the Vitali convergence theorem and the fact that M_t is a martingale. Since $\widetilde{M}_t$ is $\mathcal{F}_t$ measurable, this proves that $\widetilde{M}_t = M_t$, a.s. Since N was arbitrary, we have this for all t. We thus have found a version of M that has paths that are right continuous with left limits. That $\widetilde{M}_t$ is a martingale is easy. □

The following technical result will be used in the next chapter. A function f is increasing if $s < t$ implies $f(s) \leq f(t)$. A process A_t has increasing paths if the function $t \to A_t(\omega)$ is increasing for almost every ω.

Proposition 1.17 *Suppose $\{\mathcal{F}_t\}$ is a filtration satisfying the usual conditions and suppose A_t is an adapted process with paths that are increasing, are right continuous with left limits, and $A_\infty = \lim_{t\to\infty} A_t$ exists, a.s. Suppose X is a non-negative integrable random variable, and M_t is a version of the martingale $\mathbb{E}\,[X \mid \mathcal{F}_t]$ which has paths that are right continuous with left limits. Suppose $\mathbb{E}\,[XA_\infty] < \infty$. Then*

$$\mathbb{E}\int_0^\infty X\,dA_s = \mathbb{E}\int_0^\infty M_s\,dA_s. \tag{1.9}$$

Proof. First suppose X and A are bounded. Let $n > 1$ and let us write $\mathbb{E}\int_0^\infty X\,dA_s$ as

$$\sum_{k=1}^\infty \mathbb{E}\,[X(A_{k/2^n} - A_{(k-1)/2^n})].$$

Conditioning the k^{th} summand on $\mathcal{F}_{k/2^n}$, this is equal to

$$\mathbb{E}\Big[\sum_{k=1}^\infty \mathbb{E}\,[X \mid \mathcal{F}_{k/2^n}](A_{k/2^n} - A_{(k-1)/2^n})\Big].$$

Given s and n, define s_n to be that value of $k/2^n$ such that $(k-1)/2^n < s \le k/2^n$. We then have

$$\mathbb{E}\int_0^\infty X\,dA_s = \mathbb{E}\int_0^\infty M_{s_n}\,dA_s. \tag{1.10}$$

For any value of s, $s_n \downarrow s$ as $n \to \infty$, and since M has right continuous paths, $M_{s_n} \to M_s$. Since X is bounded, so is M. By dominated convergence, the right hand side of (1.10) converges to

$$\mathbb{E}\int_0^\infty M_s\,dA_s.$$

This completes the proof when X and A are bounded. We apply this to $X \wedge N$ and $A \wedge N$, let $N \to \infty$, and use monotone convergence for the general case. $\square$

The only reason we assume X is non-negative is so that the integrals make sense. The equation (1.9) can be rewritten as

$$\mathbb{E}\int_0^\infty X\,dA_s = \mathbb{E}\int_0^\infty \mathbb{E}\,[X \mid \mathcal{F}_s]\,dA_s. \tag{1.11}$$

We also have

$$\mathbb{E} \int_0^t X \, dA_s = \mathbb{E} \int_0^t \mathbb{E} \left[X \mid \mathcal{F}_s \right] dA_s \tag{1.12}$$

for each t. This follows either by following the above proof or by applying Proposition 1.17 to $A_{s \wedge t}$.

Chapter 2

Lévy processes

A *Lévy process* is a process with stationary and independent increments whose paths are right continuous with left limits. Having *stationary increments* means that the law of $X_t - X_s$ is the same as the law of $X_{t-s} - X_0$ whenever $s < t$. Saying that X has *independent increments* means that $X_t - X_s$ is independent of $\sigma(X_r; r \leq s)$ whenever $s < t$.

We want to examine the structure of Lévy processes. We know three examples: the Poisson process, Brownian motion, and the deterministic process $X_t = t$. It turns out all Lévy processes can be built up out of these as building blocks. We will show how to construct Lévy processes and we will give a representation of an arbitrary Lévy process.

Recall that we use $X_{t-} = \lim_{s<t, s\to t} X_s$ and $\Delta X_t = X_t - X_{t-}$.

2.1 Examples

Let us begin at looking at some simple Lévy processes. Let P_t^j, $j = 1, \ldots, J$, be a sequence of independent Poisson processes with parameters λ_j, resp. Each P_t^j is a Lévy process and the formula for the characteristic function of a Poisson random variable shows that the characteristic function of P_t^j is

$$\mathbb{E}\, e^{iuP_t^j} = \exp(t\lambda_j(e^{iu} - 1)).$$

Therefore the characteristic function of $a_jP_t^j$ is

$$\mathbb{E}\, e^{iua_jP_t^j} = \exp(t\lambda_j(e^{iua_j} - 1))$$

and the characteristic function of $a_jP_t^j - a_j\lambda_jt$ is

$$\mathbb{E}\, e^{iua_jP_j^t-a)j\lambda_jt} = \exp(t\lambda_j(e^{iua_j} - 1 - iua_j)).$$

If we let m_j be the measure on $\mathbb{R}$ defined by $m_j(dx) = \lambda_j\delta_{a_j}(dx)$, where $\delta_{a_j}(dx)$ is point mass at a_j, then the characteristic function for $a_jP_t^j$ can be written as

$$\exp\Big(t\int_{\mathbb{R}}[e^{iux} - 1]\, m_j(dx)\Big) \tag{2.1}$$

and the one for $a_jP_t^j - a_j\lambda_jt$ as

$$\exp\Big(t\int_{\mathbb{R}}[e^{iux} - 1 - iux]\, m_j(dx)\Big). \tag{2.2}$$

Now let

$$X_t = \sum_{j=1}^{J} a_jP_t^j.$$

It is clear that the paths of X_t are right continuous with left limits, and the fact that X has stationary and independent increments follows from the corresponding property of the P^j's. Moreover the characteristic function of a sum of independent random variables is the product of the characteristic functions, so the characteristic function of X_t is given by

$$\mathbb{E}\, e^{iuX_t} = \exp\Big(t\int_{\mathbb{R}}[e^{iux} - 1]\, m(dx)\Big) \tag{2.3}$$

with $m(dx) = \sum_{j=1}^{J}\lambda_j\delta_{a_j}(dx)$.

The process $Y_t = X_t - t\sum_{j=1}^{J} a_j\lambda_j$ is also a Lévy process and its characteristic function is

$$\mathbb{E}\, e^{iuY_t} = \exp\Big(t\int_{\mathbb{R}}[e^{iux} - 1 - iux]\, m(dx),\Big) \tag{2.4}$$

again with $m(dx) = \sum_{j=1}^{J}\lambda_j\delta_{a_j}(dx)$.

Remark 2.1 Recall that if φ is the characteristic function of a random variable Z, then $\varphi'(0) = i\mathbb{E}\, Z$ and $\varphi''(0) = -\mathbb{E}\, Z^2$. If Y_t is as in the paragraph above, then clearly $\mathbb{E}\, Y_t = 0$, and calculating the second derivative of $\mathbb{E}\, e^{iuY_t}$ at 0, we obtain

$$\mathbb{E}\, Y_t^2 = t \int x^2\, m(dx).$$

The following lemma is a restatement of Corollary 1.8.

Lemma 2.2 *If X_t is a Lévy process and T is a finite stopping time, then $X_{T+t} - X_T$ is a Lévy process with the same law as $X_t - X_0$ and independent of $\mathcal{F}_T$.*

We will need the following lemma:

Lemma 2.3 *Suppose $X_1, \ldots, X_n$ are independent exponential random variables with parameters $a_1, \ldots, a_n$, resp.*
(1) Then $\min(X_1, \ldots, X_n)$ is an exponential random variable with parameter $a_1 + \cdots + a_n$.
(2) The probability that X_i is the smallest of the n exponentials is

$$\frac{a_i}{a_1 + \cdots + a_n}.$$

Proof. (1) Write

$$\begin{aligned} \mathbb{P}(\min(X_1, \ldots, X_n) > t) &= \mathbb{P}(X_1 > t, \ldots, X_n > t) = \mathbb{P}(X_1 > t) \cdots \mathbb{P}(X_n > t) \\ &= e^{-a_1 t} \cdots e^{-a_n t} = e^{-(a_1 + \cdots + a_n)t}. \end{aligned}$$

(2) Without loss of generality we may suppose $i = 1$. Let's first do the case $n = 2$. The joint density of (X_1, X_2) is $a_1 e^{-a_1 x} a_2 e^{-a_2 y}$ and we want to integrate this over $x < y$. Doing the integration yields $a_1/(a_1 + a_2)$. For the general case of $n > 2$, apply the above to X_1 and $\min(X_2, \ldots, X_n)$. $\square$

If $P_1, \ldots, P_n$ are independent Poisson processes with parameters $\lambda_1, \ldots, \lambda_n$ resp., let $X_t = \sum_{i=1}^n b_i P_i(t)$. By the above lemma, the times between jumps of X are independent exponentials with parameter $\lambda_1 + \cdots + \lambda_n$. At each jump, X jumps b_i with probability $\lambda_i/(\lambda_1 + \cdots + \lambda_n)$.

Thus another way to construct X is to let $U_1, U_2, \ldots$ be independent exponentials with parameter $\lambda_1 + \ldots + \lambda_n$ and let $Y_1, Y_2, \ldots$ be a sequence of i.i.d. random variables independent of the U_i's such that $\mathbb{P}(Y_k = b_j) = \lambda_j/(\lambda_1 + \ldots + \lambda_n)$. We then let X_0 be 0, let X_t be piecewise constant, and at time $\sum_{i=1}^m U_i$ we let X jump by the amount Y_m.

2.2 Construction of Lévy processes

A process X has *bounded jumps* if there exists a real number $K > 0$ such that $\sup_t |\Delta X_t| \le K$, a.s.

Lemma 2.4 *If X_t is a Lévy process with bounded jumps and with $X_0 = 0$, then X_t has moments of all orders, that is, $\mathbb{E}\,|X_t|^p < \infty$ for all positive integers p.*

Proof. Suppose the jumps of X_t are bounded in absolute value by K. Since X_t is right continuous with left limits, there exists $M > K$ such that $\mathbb{P}(\sup_{s\le t} |X_s| \ge M) \le 1/2$.

Let $T_1 = \inf\{t : |X_t| \ge M\}$ and $T_{i+1} = \inf\{t > T_i : |X_t - X_{T_i}| > M\}$. For $s < T_1$, $|X_s| \le M$, and then $|X_{T_1}| \le |X_{T_1-}| + |\Delta X_{T_1}| \le M + K \le 2M$. We have

$$
\begin{aligned}
\mathbb{P}(T_{i+1} < t) &\le \mathbb{P}(T_i < t, T_{i+1} - T_i < t)\\
&= \mathbb{P}(T_{i+1} - T_i < t)\mathbb{P}(T_i < t)\\
&= \mathbb{P}(T_1 < t)\mathbb{P}(T_i < t),
\end{aligned}
$$

using Lemma 2.2. Now

$$
\mathbb{P}(T_1 < t) \le \mathbb{P}(\sup_{s\le t} |X_s| \ge M) \le \tfrac{1}{2},
$$

so $\mathbb{P}(T_{i+1} < t) \le \frac{1}{2}\mathbb{P}(T_i < t)$, and then by induction, $\mathbb{P}(T_i < t) \le 2^{-i}$. Therefore

$$
\mathbb{P}(\sup_{s\le t} |X_s| \ge 2(i+1)M) \le \mathbb{P}(T_i < t) \le 2^{-i}
$$

and the lemma now follows immediately. $\square$

A key lemma is the following.

Lemma 2.5 *Suppose I is a finite interval of the form (a, b), $[a, b)$, $(a, b]$, or $[a, b]$ with $a > 0$ and m is a finite measure on $\mathbb{R}$ giving no mass to I^c. Then there exists a Lévy process X_t satisfying* (2.3)

Proof. First let us consider the case where $I = [a, b)$. We approximate m be a discrete measure. If $n \geq 1$, let $z_j = a + j(b-a)/2^n$, $j = 0, \ldots, 2^n - 1$, and let

$$m_n(dx) = \sum_{j=0}^{2^n-1} m([z_j, z_{j+1}))\delta_{z_j}(dx),$$

where δ_{z_j} is point mass at z_j. The measures m_n converge weakly to m as $n \to \infty$ in the sense that

$$\int f(x)\, m_n(dx) \to \int f(x)\, dx$$

whenever f is a bounded continuous function on $\mathbb{R}$.

We let $U_1, U_2, \ldots$ be independent exponential random variables with parameter $m(I)$. Let $Y_1, Y_2, \ldots$ be i.i.d. random variables independent of the U_i's with $\mathbb{P}(Y_i \in dx) = m(dx)/m(I)$. We let X_t start at 0 and be piecewise constant with jumps of size Y_m at times $\sum_{i=1}^m U_i$.

If we define X_t^n is the exact same way, except that we replace m by m_n and we let $Y_i^n = z_j$ if $Y_i \in [z_j, z_{j+1})$, then we know from the previous section that X_t^n is a Lévy process with Lévy measure m_n. Moreover each Y_i^n differs from Y_i by at most $(b-a)2^{-n}$, so

$$\sup_{s \leq t} |X_s^n - X_s| \leq (b-a)2^n N,$$

where N is the number of jumps of these processes before time t. N is a Poisson random variable with parameter $m(I)$, so has moments of all orders. It follows that X_t^n converges uniformly to X_t almost surely on each finite interval, and the difference goes to 0 in L^p for each p.

We conclude that the law of $X_t - X_s$ is independent of $\mathcal{F}_s$ and has the same law as that of X_{t-s} because these hold for each X^n.

Since $x \to e^{iux}$ is a bounded continuous function and m_n converges weakly to m, starting with

$$\mathbb{E} \exp(iuX_t^n) = \exp\Big(t \int [e^{iux} - 1]\, m_n(dx)\Big),$$

and passing to the limit, we obtain that the characteristic function of X under $\mathbb{P}$ is given by (2.3).

If now the interval I contains the point b, we follow the above proof, except we let $P_t^{2^n-1}$ be a Poisson random variable with parameter $m([z_{n-1}, b])$. Similarly, if I does not contain the point a, we change P_t^0 to be a Poisson random variable with parameter $m((a, z_1))$. With these changes, the proof works for intervals I, whether or not they contain either of their endpoints. $\square$

Remark 2.6 If X is the Lévy process constructed in Lemma 2.5, then $Y_t = X_t - \mathbb{E}\, X_t$ will be a Lévy process satisfying (2.4).

Here is the main theorem of this section.

Theorem 2.7 *Suppose m is a measure on $\mathbb{R}$ with $m(\{0\}) = 0$ and*

$$\int (1 \wedge x^2) m(dx) < \infty.$$

Suppose $b \in \mathbb{R}$ and $\sigma \geq 0$. There exists a Lévy process X_t such that

$$\mathbb{E}\, e^{iuX_t} = \exp\Big(t\Big\{iub - \sigma^2 u^2/2 + \int_{\mathbb{R}} [e^{iux} - 1 - iux 1_{(|x| \leq 1)}] m(dx)\Big\}\Big). \quad (2.5)$$

The above equation is called the *Lévy-Khintchine formula*. The measure m is called the *Lévy measure*. If we let

$$m(dx) = \frac{1 + x^2}{x^2} m'(dx)$$

and

$$b = b' + \int_{(|x|\le 1)} \frac{x^3}{1+x^2} m(dx) - \int_{(|x|>1)} \frac{x}{1+x^2} m(dx),$$

then we can also write

$$\mathbb{E}\, e^{iuX_t} = \exp\Big(t\Big\{iub' - \sigma^2 u^2/2 + \int_{\mathbb{R}} \Big[e^{iux} - 1 - \frac{iux}{1+x^2}\Big] \frac{1+x^2}{x^2} m'(dx)\Big\}\Big).$$

Both expressions for the Lévy-Khintchine formula are in common use.

Proof. Let $m(dx)$ be a measure supported on $(0,1]$ with $\int x^2\, m(dx) < \infty$. Let $m_n(dx)$ be the measure m restricted to $(2^{-n}, 2^{-n+1}]$. Let Y_t^n be independent Lévy processes whose characteristic functions are given by (2.4) with m replaced by m_n; see Remark 2.6. Note $\mathbb{E}\, Y_t^n = 0$ for all n by Remark 2.1. By the independence of the Y^n's, if $M < N$,

$$\mathbb{E}\Big(\sum_{n=M}^N Y_t^n\Big)^2 = \sum_{n=M}^N \mathbb{E}\,(Y_t^n)^2 = \sum_{n=M}^N t \int x^2\, m_n(dx) = t\int_{2^{-N}}^{2^{-M}} x^2\, m(dx).$$

By our assumption on m, this goes to 0 as $M, N \to \infty$, and we conclude that $\sum_{n=0}^N Y_t^n$ converges in L^2 for each t. Call the limit Y_t. It is routine to check that Y_t has independent and stationary increments. Each Y_t^n has independent increments and is mean 0, so

$$\mathbb{E}\,[Y_t^n - Y_s^n \mid \mathcal{F}_s] = \mathbb{E}\,[Y_t^n - Y_s^n] = 0,$$

or Y^n is a martingale. By Doob's inequalities and the L^2 convergence,

$$\mathbb{E}\, \sup_{s\le t} \Big| \sum_{n=M}^N Y_s^n\Big|^2 \to 0$$

as $M, N \to \infty$, and hence there exists a subsequence M_k such that $\sum_{n=1}^{M_k} Y_s^n$ converges uniformly over $s \le t$, a.s. Therefore the limit Y_t will have paths that are right continuous with left limits.

If m is a measure supported in $(1,\infty)$ with $m(\mathbb{R}) < \infty$, we do a similar procedure starting with Lévy processes whose characteristic functions are of the form (2.3). We let $m_n(dx)$ be the restriction of m to $(2^n, 2^{n+1}]$, let X_t^n be independent Lévy processes corresponding to m_n, and form $X_t = \sum_{n=0}^\infty X_t^n$.

Since $m(\mathbb{R}) < \infty$, for each t_0, the number of times t less than t_0 at which any one of the X_t^n jumps is finite. This shows X_t has paths that are right continuous with left limits, and it is easy to then see that X_t is a Lévy process.

Finally, suppose $\int x^2 \wedge 1\, m(dx) < \infty$. Let X_t^1, X_t^2 be Lévy processes with characteristic functions given by (2.3) with m replaced by the restriction of m to $(1, \infty)$ and $(-\infty, -1)$, resp., let X_t^3, X_t^4 be Lévy processes with characteristic functions given by (2.4) with m replaced by the restriction of m to $(0, 1]$ and $[-1, 0)$, resp., let $X_t^5 = bt$, and let X_t^6 be σ times a Brownian motion. Suppose the X^i's are all independent. Then their sum will be a Lévy process whose characteristic function is given by (2.5). □

A key step in the construction was the centering of the Poisson processes to get Lévy processes with characteristic functions given by (2.4). Without the centering one is forced to work only with characteristic functions given by (2.3).

2.3 Representation of Lévy processes

We now work towards showing that every Lévy process has a characteristic function of the form given by (2.5).

Lemma 2.8 *If X_t is a Lévy process and A is a Borel subset of $\mathbb{R}$ that is a positive distance from 0, then*

$$N_t(A) = \sum_{s \le t} 1_A(\Delta X_s)$$

is a Poisson process.

Saying that A is a positive distance from 0 means that $\inf\{|x| : x \in A\} > 0$.

Proof. Since X_t has paths that are right continuous with left limits and A is a positive distance from 0, then there can only be finitely many jumps of X that lie in A in any finite time interval, and so $N_t(A)$ is finite and has paths that are right continuous with left limits. It follows from the fact that X_t has stationary and independent increments that $N_t(A)$ also has stationary and independent increments. We now apply Proposition 1.10. □

Our main result is that $N_t(A)$ and $N_t(B)$ are independent if A and B are disjoint.

Theorem 2.9 *Let $\{\mathcal{F}_t\}$ be a filtration satisfying the usual conditions. Suppose that $N_t(A)$ is a Poisson point process with respect to the measure λ. If $A_1, \ldots, A_n$ are pairwise disjoint measurable subsets of $\mathbb{R}$ with $\mathbb{E}\, N_t(A_k) < \infty$ for $k = 1, \ldots, n$, then the processes $N_t(A_1), \ldots, N_t(A_n)$ are mutually independent.*

Proof. Define $\lambda(A) = \mathbb{E}\, N_1(A)$. The previous lemma shows that if $\lambda(A) < \infty$, then $N_t(A)$ is a Poisson process, and clearly its parameter is $\lambda(A)$. We first make the observation that because $A_1, A_2, \ldots, A_n$ are disjoint, then no two of the $N_t(A_k)$ have jumps at the same time.

To prove the theorem, it suffices to let $0 = r_0 < r_1 < \cdots < r_m$ and show that the random variables

$$\{N_{r_j}(A_k) - N_{r_{j-1}}(A_k) : 1 \le j \le m, 1 \le k \le n\}$$

are independent. Since for each j and each k, $N_{r_j}(A_k) - N_{r_{j-1}}(A_k)$ is independent of $\mathcal{F}_{r_{j-1}}$, it suffices to show that for each $j \le m$, the random variables

$$\{N_{r_j}(A_k) - N_{r_{j-1}}(A_k) : 1 \le k \le n\}$$

are independent. We will do the case $j = m = 1$ and write r for r_j for simplicity; the case when $j, m > 1$ differs only in notation.

We will prove this using induction. We start with the case $n = 2$ and show the independence of $N_r(A_1)$ and $N_r(A_2)$. Each $N_t(A_k)$ is a Poisson process, and so $N_t(A_k)$ has moments of all orders. Let $u_1, u_2 \in \mathbb{R}$ and set

$$\phi_k = \lambda(A_k)(e^{iu_k} - 1), \qquad k = 1, 2.$$

Let

$$M_t^k = e^{iu_k N_t(A_k) - t\phi_k}.$$

We see that M_t^k is a martingale because $\mathbb{E}\, e^{iu_k N_t(A_k)} = e^{t\phi_k}$, and therefore

$$\begin{aligned} \mathbb{E}\,[M_t^k \mid \mathcal{F}_s] &= M_s^k \mathbb{E}\,[e^{iu(N_t(A_k) - N_s(A_k))) - (t-s)\phi_k} \mid \mathcal{F}_s] \\ &= M_s^k e^{-(t-s)\phi_k} \mathbb{E}\,[e^{iu(N_t(A_k) - N_s(A_k))}] = M_s^k, \end{aligned}$$

using the independence and stationarity of the increments of a Poisson process.

Now we can write

$$\begin{aligned}\mathbb{E}\,[M_t^1 M_t^2] &= \mathbb{E}\,[M_t^1] + \mathbb{E}\int_0^t M_t^1\,dM_s^2 \\ &= 1 + \mathbb{E}\int_0^t M_s^1\,dM_s^2,\end{aligned}$$

using that $M_0^2 = 1$, M^1 is a martingale, and Proposition 1.17. (Here M^2 is the difference of two increasing processes; the adjustments needed are easy.)

Since we have argued that no two of the $N_t(A_k)$ jump at the same time, the same is true for the M_t^k and so the above is equal to

$$1 + \mathbb{E}\int_0^t M_{s-}^1\,dM_s^2.$$

It therefore remains to prove that the above integral is equal to 0.

If H_s is a process of the form

$$H_s(\omega) = K(\omega)1_{(a,b]}(s)$$

where K is $\mathcal{F}_a$ measurable, then

$$\int_0^t H_s\,dM_s^2 = K(M_{t\wedge b}^2 - M_{t\wedge a}^2),$$

and conditioning on $\mathcal{F}_a$, the expectation is zero:

$$\mathbb{E}\,[K(M_{t\wedge b}^2 - M_{t\wedge a}^2)] = \mathbb{E}\,[K\mathbb{E}\,[M_{t\wedge b}^2 - M_{t\wedge a}^2 \mid \mathcal{F}_a]\,] = 0,$$

using that M^2 is a martingale. We are doing Lebesgue-Stieltjes integrals here, but the argument is similar to one used with stochastic integrals. The expectation is also 0 for linear combinations of such H's. Since M_{s-}^1 is left continuous, we can approximate it by such H's, and therefore the integral is 0 as required.

We thus have

$$\mathbb{E}\,M_r^1 M_r^2 = 1.$$

This implies

$$\mathbb{E}\left[e^{i(u_1 N_r(A_1)+u_2 N_r(A_2))}\right] = e^{r\phi_1}e^{r\phi_2} = \mathbb{E}\left[e^{iu_1 N_r(A_1)}\right]\mathbb{E}\left[e^{iu_2 N_r(A_2)}\right].$$

Since this holds for all u_1, u_2, then $N_r(A_1)$ and $N_r(A_2)$ are independent. We conclude that the processes $N_t(A_1)$ and $N_t(A_2)$ are independent.

To handle the case $n = 3$, we first show that $M_t^1 M_t^2$ is a martingale. We write

$$\begin{aligned}
&\mathbb{E}\,[M_t^1 M_t^2 \mid \mathcal{F}_s] \\
&\qquad = M_s^1 M_s^2 e^{-(t-s)(\phi_1+\phi_2)}\mathbb{E}\,[e^{i(u_1(N_t(A_1)-N_s(A_1))+u_2(N_t(A_2)-N_s(A_2)))} \mid \mathcal{F}_s] \\
&\qquad = M_s^1 M_s^2 e^{-(t-s)(\phi_1+\phi_2)}\mathbb{E}\,[e^{i(u_1(N_t(A_1)-N_s(A_1))+u_2(N_t(A_2)-N_s(A_2)))}] \\
&\qquad = M_s^1 M_s^2,
\end{aligned}$$

using the fact that $N_t(A_1)$ and $N_t(A_2)$ are independent of each other and each have stationary and independent increments.

Note that $M_t^3 = e^{iu_3 N_t(A_3) - t\phi_3}$ has no jumps in common with M_t^1 or M_t^2. Therefore if $\overline{M}_t^3 = M_{t\wedge r}^3$, then

$$\mathbb{E}\,[\overline{M}_\infty^3(\overline{M}_\infty^1 \overline{M}_\infty^2)] = 0,$$

and as before this leads to

$$\mathbb{E}\,[M_r^3(M_r^1 M_r^2)] = 1.$$

As above this implies that $N_r(A_1), N_r(A_2)$, and $N_r(A_3)$ are independent. To prove the general induction step is similar. □

We will also need the following corollary.

Corollary 2.10 *Let $\mathcal{F}_t$ and $N_t(A_k)$ be as in Theorem 4.2. Suppose Y_t is a process with paths that are right continuous with left limits such that $Y_t - Y_s$ is independent of $\mathcal{F}_s$ whenever $s < t$ and $Y_t - Y_s$ has the same law as Y_{t-s} for each $s < t$. Suppose moreover that Y has no jumps in common with any of the $N_t(A_k)$. Then the processes $N_t(A_1), \ldots, N_t(A_n)$, and Y_t are independent.*

Proof. The law of Y_0 is the same as that of $Y_t - Y_t$, so $Y_0 = 0$, a.s. By the fact that Y has stationary and independent increments,

$$\mathbb{E}\, e^{iuY_{s+t}} = \mathbb{E}\, e^{iuY_s} \mathbb{E}\, e^{iu(Y_{s+t}-Y_s)} = \mathbb{E}\, e^{iuY_s} \mathbb{E}\, e^{iuY_t},$$

which implies that the characteristic function of Y is of the form $\mathbb{E}\, e^{iuY_t} = e^{t\psi(u)}$ for some function $\psi(u)$.

We fix $u \in \mathbb{R}$ and define

$$M_t^Y = e^{iuY_t - t\psi(u)}.$$

As in the proof of Theorem 4.2, we see that M_t^Y is a martingale. Since M^Y has no jumps in common with any of the M_t^k, if $\overline{M}_t^Y = M_{t\wedge r}^Y$, we see as above that

$$\mathbb{E}\, [\overline{M}_\infty^Y (\overline{M}_\infty^1 \cdots \overline{M}_\infty^n)] = 1,$$

or

$$\mathbb{E}\, [M_r^Y M_r^1 \cdots M_r^n] = 1.$$

This leads as above to the independence of Y from all the $N_t(A_k)$'s. □

Here is the representation theorem for Lévy processes.

Theorem 2.11 *Suppose X_t is a Lévy process with $X_0 = 0$. Then there exists a measure m on $\mathbb{R} - \{0\}$ with*

$$\int (1 \wedge x^2)\, m(dx) < \infty$$

and real numbers b and σ such that the characteristic function of X_t is given by (2.5)*.*

Proof. Define $m(A) = \mathbb{E}\, N_1(A)$ if A is a bounded Borel subset of $(0, \infty)$ that is a positive distance from 0. Since $N_1(\cup_{k=1}^\infty A_k) = \sum_{k=1}^\infty N_1(A_k)$ if the A_k are pairwise disjoint and each is a positive distance from 0, we see that m is a measure on $[a, b]$ for each $0 < a < b < \infty$, and m extends uniquely to a measure on $(0, \infty)$.

First we want to show that $\sum_{s\le t}\Delta X_s 1_{(\Delta X_s>1)}$ is a Lévy process with characteristic function

$$\exp\Big(t\int_1^\infty [e^{iux}-1]\,m(dx)\Big).$$

Since the characteristic function of the sum of independent random variables is equal to the product of the characteristic functions, it suffices to suppose $0<a<b$ and to show that

$$\mathbb{E}\,e^{iuZ_t}=\exp\Big(t\int_{(a,b]}[e^{iux}-1]\,m(dx)\Big),$$

where

$$Z_t=\sum_{s\le t}\Delta X_s 1_{(a,b]}(\Delta X_s).$$

Let $n>1$ and $z_j=a+j(b-a)/n$. By Lemma 2.8, $N_t((z_j,z_{j+1}])$ is a Poisson process with parameter

$$\ell_j=\mathbb{E}\,N_1((z_{j-1},z_j])=m((z_j,z_{j+1}]).$$

Thus $\sum_{j=0}^{n-1} z_j N_t((z_j,z_{j+1}])$ has characteristic function

$$\prod_{j=0}^{n-1}\exp(t\ell_j(e^{iuz_j}-1))=\exp\Big(t\sum_{j=0}^{n-1}(e^{iuz_j}-1)\ell_j\Big),$$

which is equal to

$$\exp\Big(t\int(e^{iux}-1)\,m_n(dx)\Big), \tag{2.6}$$

where $m_n(dx)=\sum_{j=0}^{n-1}\ell_j\delta_{z_j}(dx)$. Since Z_t^n converges to Z_t as $n\to\infty$, passing to the limit shows that Z_t has a characteristic function of the form (2.5).

Next we show that $m(1,\infty)<\infty$. (We write $m(1,\infty)$ instead of $m((1,\infty))$ for esthetic reasons.) If not, $m(1,K)\to\infty$ as $K\to\infty$. Then for each fixed L and each fixed t,

$$\limsup_{K\to\infty}\mathbb{P}(N_t(1,K)\le L)=\limsup_{K\to\infty}\sum_{j=0}^{L}e^{-tm(1,K)}\frac{m(1,K)^j}{j!}=0.$$

This implies that $N_t(1,\infty)=\infty$ for each t. However, this contradicts the fact that X_t has paths that are right continuous with left limits.

We define m on $(-\infty, 0)$ similarly.

We now look at

$$Y_t = X_t - \sum_{s \le t} \Delta X_s 1_{(|\Delta X_s| > 1)}.$$

This is again a Lévy process, and we need to examine its structure. This process has bounded jumps, hence has moments of all orders. By subtracting $c_1 t$ for an appropriate constant c_1, we may suppose Y_t has mean 0. Let $I_1, I_2, \ldots$ be an ordering of the intervals $\{[2^{-(m+1)}, 2^{-m}), (-2^{-m}, -2^{-(m+1)}] : m \ge 0\}$. Let

$$\widetilde{X}_t^k = \sum_{s \le t} \Delta X_s 1_{(\Delta X_s \in I_k)}$$

and let $X_t^k = \widetilde{X}_t^k - \mathbb{E}\, \widetilde{X}_t^k$. By the fact that all the X^k have mean 0 and are independent,

$$\sum_{k=1}^\infty \mathbb{E}\,(X_t^k)^2 \le \mathbb{E}\Big[\Big(Y_t - \sum_{k=1}^\infty X_t^k\Big)^2\Big] + \mathbb{E}\Big[\Big(\sum_{k=1}^\infty X_t^k\Big)^2\Big] = \mathbb{E}\,(Y_t)^2 < \infty.$$

Hence

$$\mathbb{E}\Big[\sum_{k=M}^N X_t^k\Big]^2 = \sum_{k=M}^N \mathbb{E}\,(X_t^k)^2$$

tends to 0 as $M, N \to \infty$, and thus $X_t - \sum_{k=1}^N X_t^k$ converges in L^2. The limit, X_t^c, say, will be a Lévy process independent of all the X_t^k. Moreover, X^c has no jumps, i.e., it is continuous. Since all the X^k have mean 0, then $\mathbb{E}\, X_t^c = 0$. By the independence of the increments,

$$\mathbb{E}\,[X_t^c - X_s^c \mid \mathcal{F}_s] = \mathbb{E}\,[X_t^c - X_s^c] = 0,$$

and we see X^c is a continuous martingale. Using the stationarity and independence of the increments,

$$\begin{aligned}\mathbb{E}\,[(X_{s+t}^c)^2] &= \mathbb{E}\,[(X_s^c)^2] + 2\mathbb{E}\,[X_s^c(X_{s+t}^c - X_s^c)] + \mathbb{E}\,[(X_{s+t}^c - X_s^c)^2]\\ &= \mathbb{E}\,[(X_s^c)^2] + \mathbb{E}\,[(X_t^c)^2],\end{aligned}$$

which implies that there exists a constant c_2 such that $\mathbb{E}\,(X_t^c)^2 = c_2 t$. We then have

$$\begin{aligned}\mathbb{E}\,[(X_t^c)^2 - c_2 t \mid \mathcal{F}_s] &= (X_s^c)^2 - c_2 s + \mathbb{E}\,[(X_t^c - X_s^c)^2 \mid \mathcal{F}_s] - c_2(t-s)\\ &= (X_s^c)^2 - c_2 s + \mathbb{E}\,[(X_t^c - X_s^c)^2] - c_2(t-s)\\ &= (X_s^c)^2 - c_2 s.\end{aligned}$$

The quadratic variation process of X^c is therefore $c_2 t$, and by Lévy's theorem, $X_t^c/\sqrt{c_2}$ is a constant multiple of Brownian motion.

To complete the proof, it remains to show that $\int_{-1}^1 x^2\, m(dx) < \infty$. But by Remark 2.1,

$$\int_{I_k} x^2\, m(dx) = \mathbb{E}\,(X_1^k)^2,$$

and we have seen that

$$\sum_k \mathbb{E}\,(X_1^k)^2 \le \mathbb{E}\, Y_1^2 < \infty.$$

Combining gives the finiteness of $\int_{-1}^1 x^2\, m(dx)$. $\square$

2.4 Symmetric stable processes

Let $\alpha \in (0,2)$. If

$$m(dx) = \frac{c}{|x|^{1+\alpha}}\, dx,$$

we have what is called a symmetric stable process of index α. We see that $\int 1 \wedge x^2\, m(dx)$ is finite.

Because $|x|^{-1-\alpha}$ is symmetric, in the Lévy-Khintchine formula we can take $iux\, 1_{(|x|<a)}$ for any a instead of $iux\, 1_{(|x|<1)}$. Then

$$\begin{aligned}\int \Big[e^{iux} - 1 - iux\, 1_{(|x|<1)}\Big] \frac{c}{|x|^{1+\alpha}}\, dx &= \int \Big[e^{iux} - 1 - iux\, 1_{(|x|<1/|u|)}\Big] \frac{c}{|x|^{1+\alpha}}\, dx\\ &= \int \Big[e^{iy} - 1 - iy\, 1_{(|y|<1)}\Big] |u|^{1+\alpha} \frac{dy}{|u|}\\ &= -c'|u|^\alpha.\end{aligned}$$

In the last line we have the negative sign because the imaginary part of $e^{iy} - 1 - iy\,1_{(|y|<1)}$ is zero and the real part is negative since $|\cos y| \le 1$. Therefore is X_t is a symmetric stable process of index α,

$$\mathbb{E}\, e^{iuX_t} = e^{-c't|u|^\alpha}.$$

An exercise is to show that if $a > 0$ and X_t is a symmetric stable process of index α, then X_{at} and $a^{1/\alpha}X_t$ have the same law.

By Exercise 6.7.4 of Chung's book,

$$\mathbb{P}(X_1 > A) \sim cA^{-\alpha}, \qquad A \to \infty, \tag{2.7}$$

where $f \sim g$ means the ratio of the two sides goes to 0.

Since $e^{-c't|u|^\alpha}$ is integrable, X_t has a continuous density function $p_t(x)$. We have

$$p_t(0) = \frac{1}{2\pi}\int e^{-c't|u|^\alpha}\,du, \tag{2.8}$$

and by a change of variables,

$$p_t(0) = ct^{-1/\alpha}. \tag{2.9}$$

If $x \neq 0$, then

$$p_t(x) = \frac{1}{2\pi}\int e^{-iux}e^{-c't|u|^\alpha}\,du = \frac{1}{2\pi}\int (\cos ux - i\sin ux)e^{-c't|u|^\alpha}\,du.$$

Since $\sin ux$ is an odd function of u, the imaginary term is 0. Since $\cos ux \le 1$ and in fact is strictly less than 1 except at countably many values of u, we see that

$$p_t(x) < p_t(0). \tag{2.10}$$

If $\beta < 1$, we can take $m(dx) = c/|x|^{1+\beta}$ for $x > 0$ and 0 for $x < 0$. We can also take the Lévy-Khintchine exponent to be just $[e^{iux} - 1]$ if we take the drift term to cancel out the $\int iux\,1_{(|x|<1)}$ term. This reflects that here we do not need to subtract the mean to get convergence of the compound Poisson processes. In this case we get the one-sided stable processes of index β. The paths of such a process only increase.

There is a notion of subordination which is very curious. Suppose that T_t is a one-sided stable process of index β with $\beta < 1$. Let W_t be a Brownian motion independent of T_t. Then $Y_t = W_{T_t}$ is a symmetric stable process of index 2β. Let's see why that is so.

That Y is a Lévy process is not hard to see. We must therefore calculate the Lévy measure m. If P_t is a Poisson process with parameter λ, then

$$\mathbb{E}\, e^{uP_t} = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} e^{uk} = e^{-\lambda t} e^{u\lambda t} = e^{\lambda t(e^u - 1)}.$$

Using that the moment generating function of independent random variables is the product of the moment generating functions and taking limits, we see that

$$\mathbb{E}\, e^{-uT_t} = e^{-cu^\beta}.$$

Then

$$\begin{aligned}
\mathbb{E}\, e^{iuY_t} &= \mathbb{E} \int e^{iuW_s} \mathbb{P}(T_t \in ds) \\
&= \int e^{-u^2 s/2} \mathbb{P}(T_t \in ds) \\
&= \mathbb{E}\, e^{-u^2 T_t/2} \\
&= e^{-ct(u^2/2)^\beta} \\
&= e^{-c'|u|^{2\beta}}.
\end{aligned}$$

Chapter 3

Stochastic calculus

In this chapter we investigate the stochastic calculus for processes which may have jumps as well as a continuous component. If X is not a continuous process, it is no longer true that $X_{t\wedge T_N}$ is a bounded process when $T_N = \inf\{t : |X_t| \geq N\}$, since there could be a large jump at time T_N. We investigate stochastic integrals with respect to square integrable (not necessarily continuous) martingales, Itô's formula, and the Girsanov transformation. We prove the reduction theorem that allows us to look at semimartingales that are not necessarily bounded.

We will need the Doob-Meyer decomposition, which can be found in Chapter 16 of Bass, *Stochastic Processes*. That in turn depends on the debut and section theorems. A simpler proof than the standard one for the debut and section theorems can be found in the Arxiv:

`http://arxiv.org/abs/1001.3619`.

3.1 Decomposition of martingales

We assume throughout this chapter that $\{\mathcal{F}_t\}$ is a filtration satisfying the usual conditions. This means that each $\mathcal{F}_t$ contains every $\mathbb{P}$-null set and $\cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon} = \mathcal{F}_t$ for each t.

Let us with a few definitions and facts. The *predictable σ-field* is the σ-field of subsets of $[0,\infty)\times\Omega$ generated by the collection of bounded, left continuous processes that are adapted to $\{\mathcal{F}_t\}$. A stopping time T is *predictable* and predicted by the sequence of stopping times T_n if $T_n \uparrow T$, and $T_n < T$ on the event $(T > 0)$. A stopping time T is *totally inaccessible* if $\mathbb{P}(T = S) = 0$ for every predictable stopping time S. The graph of a stopping time T is $[T, T] = \{(t, \omega) : t = T(\omega) < \infty\}$. If X_t is a process that is right continuous with left limits, we set $X_{t-} = \lim_{s \to t, s<t} X_s$ and $\Delta X_t = X_t - X_{t-}$. Thus ΔX_t is the size of the jump of X_t at time t.

Let's look at some examples. If W_t is a Brownian motion and $T = \inf\{t : W_t = 1\}$, then $T_n = \inf\{t : W_t = 1 - (1/n)\}$ are stopping times that predict T.

On the other hand, if P_t is a Poisson process (with parameter 1, say, for convenience), then we claim that $T = \inf\{t : P_t = 1\}$ is totally inaccessible. To show this, suppose S is a stopping time and $S_n \uparrow S$ are stopping times such that $S_n < S$ on $(S > 0)$. We will show that $\mathbb{P}(S = T) = 0$. To do that, it suffices to show that $\mathbb{P}(S \wedge N = T) = 0$ for each positive integer N. Since $P_t - t$ is a martingale, $\mathbb{E}\, P_{S_n \wedge N} = \mathbb{E}\,(S_n \wedge N)$. Letting $n \to \infty$, we obtain (by monotone convergence) that $\mathbb{E}\, P_{(S \wedge N)-} = \mathbb{E}\,(S \wedge N)$. We also know that $\mathbb{E}\, P_{S \wedge N} = \mathbb{E}\,(S \wedge N)$. Therefore $\mathbb{E}\, P_{(S \wedge N)-} = \mathbb{E}\, P_{S \wedge N}$. Since P has increasing paths, this implies that $P_{(S \wedge N)-} = P_{S \wedge N}$, and we conclude $\mathbb{P}(S \wedge N = T) = 0$.

In this chapter we will assume througout for simplicity that every jump time of whichever process we are considering is totally inaccessible. The general case is not much harder, but the differences are only technical.

A supermartingale Z is of class D if the family of random variables:

$$\{Z_T : T \text{ a finite stopping time}\}$$

is uniformly integrable.

Theorem 3.1 *(Doob-Meyer decomposition) Let $\{\mathcal{F}_t\}$ be a filtration satisfying the usual conditions and let Z be a supermartingale of class D whose paths are right continuous with left limits. Then Z can be written $Z_t = M_t - A_t$ in one and only one way, where M and A are adapted processes whose paths are right continuous with left limits, A has continuous increasing paths and $A_\infty = \lim_{t \to \infty} A_t$ is integrable, and M is a uniformly integrable martingale*

Suppose A_t is a bounded increasing process whose paths are right continuous with left limits. Recall that a function f is increasing if $s < t$ implies

$f(s) \le f(t)$. Then trivially A_t is a submartingale, and by the Doob-Meyer decomposition, there exists a continuous increasing process $\widetilde{A}_t$ such that $A_t - \widetilde{A}_t$ is a martingale. We call $\widetilde{A}_t$ the *compensator* of A_t.

If $A_t = B_t - C_t$ is the difference of two increasing processes B_t and C_t, then we can use linearity to define $\widetilde{A}_t$ as $\widetilde{B}_t - \widetilde{C}_t$. We can even extend the notion of compensator to the case where A_t is complex valued and has paths that are locally of bounded variation by looking at the real and imaginary parts.

We will use the following lemma. For any increasing process A we let $A_\infty = \lim_{t\to\infty} A_t$.

Lemma 3.2 *Suppose A_t has increasing paths that are right continuous with left limits, $A_t \le K$ a.s. for each t, and let B_t be its compensator. Then $\mathbb{E}\, B_\infty^2 \le 2K^2$.*

Proof. If $M_t = A_t - B_t$, then M_t is a martingale, and then

$$\mathbb{E}\,[M_\infty - M_t \mid \mathcal{F}_t] = 0.$$

We then write

$$\begin{aligned}\mathbb{E}\, B_\infty^2 &= 2\mathbb{E} \int_0^\infty (B_\infty - B_t)\, dB_t = 2\mathbb{E} \int_0^\infty \mathbb{E}\,[B_\infty - B_t \mid \mathcal{F}_t]\, dB_t \\ &= 2\mathbb{E} \int_0^\infty \mathbb{E}\,[A_\infty - A_t \mid \mathcal{F}_t]\, dB_t \le 2K\mathbb{E} \int_0^\infty dB_t \\ &= 2K\mathbb{E}\, B_\infty = 2K\mathbb{E}\, A_\infty \le 2K^2.\end{aligned}$$

□

From the lemma we get the following corollary.

Corollary 3.3 *If $A_t = B_t - C_t$, where B_t and C_t are increasing right continuous processes with $B_0 = C_0 = 0$, a.s., and in addition B and C are bounded, then*

$$\mathbb{E} \sup_{t\ge 0} \widetilde{A}_t^2 < \infty.$$

Proof. By a proposition, $\mathbb{E}\,\widetilde{B}_\infty^2 < \infty$ and $\mathbb{E}\,\widetilde{C}_\infty^2 < \infty$, and so

$$\mathbb{E}\,\sup_{t\ge 0} \widetilde{A}_t^2 \le \mathbb{E}\,[2\sup_{t\ge 0} \widetilde{B}_t^2 + 2\sup_{t\ge 0} \widetilde{C}_t^2] \le 2\mathbb{E}\,\widetilde{B}_\infty^2 + 2\mathbb{E}\,\widetilde{C}_\infty^2 < \infty.$$

We are done. □

A key result is the following *orthogonality lemma.*

Lemma 3.4 *Suppose A_t is a bounded increasing right continuous process with $A_0 = 0$, a.s., $\widetilde{A}_t$ is the compensator of A, and $M_t = A_t - \widetilde{A}_t$. Suppose N_t is a right continuous square integrable martingale such that $(\Delta N_t)(\Delta M_t) = 0$ for all t. Then $\mathbb{E}\,M_\infty N_\infty = 0$.*

Proof. By Lemma 3.3, M is square integrable. Suppose

$$H(s,\omega) = K(\omega)1_{(a,b]}(s)$$

with K being $\mathcal{F}_a$ measurable. Since M_t is of bounded variation, we have (this is a Lebesgue-Stieltjes integral here)

$$\mathbb{E}\int_0^\infty H_s\,dM_s = \mathbb{E}\,[K(M_b - M_a)] = \mathbb{E}\,[K\mathbb{E}\,[M_b - M_a \mid \mathcal{F}_a]\,] = 0.$$

We see that linear combinations of such H's generate the predictable σ-field. Thus by linearity and taking limits, $\mathbb{E}\int_0^\infty H_s\,dM_s = 0$ if H_s is a predictable process such that $\mathbb{E}\int_0^\infty |H_s|\,|dM_s| < \infty$. In particular, since N_{s-} is left continuous and hence predictable, $\mathbb{E}\int_0^\infty N_{s-}\,dM_s = 0$, provided we check integrability:

$$\begin{aligned}\mathbb{E}\Big|\int_0^\infty |N_{s-}|\,|dM_s|\Big| &\le \mathbb{E}\int_0^\infty (\sup_r |N_r|)\,|dM_s| \\ &= \mathbb{E}\,[(\sup_r |N_r|)\,(A_\infty + \widetilde{A}_\infty)] < \infty\end{aligned}$$

by the Cauchy-Schwarz inequality.

By hypothesis, $\mathbb{E}\int_0^\infty \Delta N_s\,dM_s = 0$, so $\mathbb{E}\int_0^\infty N_s\,dM_s = 0$. On the other hand, using Proposition 1.17, we see

$$\mathbb{E}\,M_\infty N_\infty = \mathbb{E}\int_0^\infty N_\infty\,dM_s = \mathbb{E}\int_0^\infty N_s\,dM_s = 0.$$

The proof is complete. □

If we apply the above to $N_{t\wedge T}$, we have $\mathbb{E}\, M_\infty N_T = 0$. If we then condition on $\mathcal{F}_T$,

$$\mathbb{E}\left[M_T N_T\right] = \mathbb{E}\left[N_T \mathbb{E}\left[M_\infty \mid \mathcal{F}_T\right]\right] = \mathbb{E}\left[N_T M_\infty\right] = 0. \tag{3.1}$$

The reason for the name "orthogonality lemma" is that by (3.1), $M_t N_t$ is a martingale. This implies that $\langle M, N\rangle_t$ (which we will define soon, and is defined similarly to the case of continuous martingales) is identically equal to 0.

Let M_t be a square integrable martingale with paths that are right continuous and left limits, so that $\mathbb{E}\, M_\infty^2 < \infty$. For each $i \in \mathbb{Z}$, let $T_{i1} = \inf\{t : |\Delta M_t| \in [2^i, 2^{i+1})\}$, $T_{i2} = \inf\{t > T_{i1} : |\Delta M_t| \in [2^i, 2^{i+1})\}$, and so on; i can be both positive and negative. Since M_t is right continuous with left limits, for each i, $T_{ij} \to \infty$ as $j \to \infty$. We conclude that M_t has at most countably many jumps. We relabel the jump times as $S_1, S_2, \ldots$ so that each S_k is totally inaccessible, the graphs of the S_k are disjoint, M has a jump at each time S_k and only at these times, and $|\Delta M_{S_k}|$ is bounded for each k. We do not assume that $S_{k_1} \le S_{k_2}$ if $k_1 \le k_2$, and in general it would not be possible to arrange this.

If S_i is a totally inaccessible stopping time, let

$$A_i(t) = \Delta M_{S_i} 1_{(t \ge S_i)} \tag{3.2}$$

and

$$M_i(t) = A_i(t) - \widetilde{A}_i(t), \tag{3.3}$$

where $\widetilde{A}_i$ is the compensator of A_i. $A_i(t)$ is the process that is 0 up to time S_i and then jumps an amount ΔM_{S_i}; thereafter it is constant. We know that $\widetilde{A}$ is continuous. Note that $M - M_i$ has no jump at time S_i.

Theorem 3.5 *Suppose M is a square integrable martingale and we define M_i as in* (3.3).

(1) Each M_i is square integrable.

(2) $\sum_{i=1}^\infty M_i(\infty)$ converges in L^2.

(3) If $M_t^c = M_t - \sum_{i=1}^\infty M_i(t)$, then M^c is square integrable and we can find a version that has continuous paths.

(4) For each i and each stopping time T, $\mathbb{E}\,[M_T^c M_i(T)] = 0$.

Proof. (1) If S_i is a totally inaccessible stopping time and we let $B_t = (\Delta M_{S_i})^+ 1_{(t \ge S_i)}$ and $C_t = (\Delta M_{S_i})^- 1_{(t\ge S_i)}$, then (1) follows by Corollary 3.3.

(2) Let $V_n(t) = \sum_{i=1}^n M_i(t)$. By the orthogonality lemma (Lemma 3.4), $\mathbb{E}\,[M_i(\infty)M_j(\infty)] = 0$ if $i \neq j$ and $\mathbb{E}\,[M_i(\infty)(M_\infty - V_n(\infty)] = 0$ if $i \le n$. We thus have

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}\, M_i(\infty)^2 &= \mathbb{E}\, V_n(\infty)^2 \\ &\le \mathbb{E}\,\Big[M_\infty - V_n(\infty)\Big]^2 + \mathbb{E}\, V_n(\infty)^2 \\ &= \mathbb{E}\,\Big[M_\infty - V_n(\infty) + V_n(\infty)\Big]^2 \\ &= \mathbb{E}\, M_\infty^2 < \infty.\end{aligned}$$

Therefore the series $\mathbb{E}\,\sum_{i=1}^n M_i(\infty)^2$ converges. If $n > m$,

$$\mathbb{E}\,[(V_n(\infty) - V_m(\infty)]^2 = \mathbb{E}\,\Big[\sum_{i=m+1}^n M_i(\infty)\Big]^2 = \sum_{i=m+1}^n \mathbb{E}\, M_i(\infty)^2.$$

This tends to 0 as $n, m \to \infty$, so $V_n(\infty)$ is a Cauchy sequence in L^2, and hence converges.

(3) From (2), Doob's inequalities, and the completeness of L^2, the random variables $\sup_{t\ge 0}[M_t - V_n(t)]$ converge in L^2 as $n \to \infty$. Let $M_t^c = \lim_{n\to\infty}[M_t - V_n(t)]$. There is a sequence n_k such that

$$\sup_{t\ge 0} |(M_t - V_{n_k}(t)) - M_t^c| \to 0, \qquad \text{a.s.}$$

We conclude that the paths of M_t^c are right continuous with left limits. By the construction of the M_i, $M - V_{n_k}$ has jumps only at times S_i for $i > n_k$. We therefore see that M^c has no jumps, i.e., it is continuous.

(4) By the orthogonality lemma and (3.1),

$$\mathbb{E}\,[M_i(T)(M_T - V_n(T)] = 0$$

if T is a stopping time and $i \le n$. Letting n tend to infinity proves (4). $\square$

3.2 Stochastic integrals

If M_t is a square integrable martingale, then M_t^2 is a submartingale by Jensen's inequality for conditional expectations. Just as in the case of continuous martingales, we can use the Doob-Meyer decomposition to find a predictable increasing process starting at 0, denoted $\langle M \rangle_t$, such that $M_t^2 - \langle M \rangle_t$ is a martingale.

Let us define

$$[M]_t = \langle M^c \rangle_t + \sum_{s \le t} |\Delta M_s|^2. \tag{3.4}$$

Here M^c is the continuous part of the martingale M as defined in Theorem 3.5. As an example, if $M_t = P_t - t$, where P_t is a Poisson process with parameter 1, then $M_t^c = 0$ and

$$[M]_t = \sum_{s \le t} \Delta P_s^2 = \sum_{s \le t} \Delta P_s = P_t,$$

because all the jumps of P_t are of size one. In this case $\langle M \rangle_t = t$; this follows from Proposition 3.6 below.

In defining stochastic integrals, one could work with $\langle M \rangle_t$, but the process $[M]_t$ is the one that shows up naturally in many formulas, such as the product formula.

Proposition 3.6 *$M_t^2 - [M]_t$ is a martingale.*

Proof. By the orthogonality lemma and (3.1) it is easy to see that

$$\langle M \rangle_t = \langle M^c \rangle_t + \sum_i \langle M_i \rangle_t.$$

Since $M_t^2 - \langle M \rangle_t$ is a martingale, we need only show $[M]_t - \langle M \rangle_t$ is a martingale. Since

$$[M]_t - \langle M \rangle_t = \Big(\langle M^c \rangle_t + \sum_{s \le t} |\Delta M_s|^2 \Big) - \Big(\langle M^c \rangle_t + \sum_i \langle M_i \rangle_t \Big),$$

it suffices to show that $\sum_i \langle M_i \rangle_t - \sum_i \sum_{s \le t} |\Delta M_i(s)|^2$ is a martingale.

By an exercise

$$M_i(t)^2 = 2\int_0^t M_i(s-)\,dM_i(s) + \sum_{s\le t} |\Delta M_i(s)|^2, \tag{3.5}$$

where the first term on the right hand side is a Lebesgue-Stieltjes integral. If we approximate this integral by a Riemann sum and use the fact that M_i is a martingale, we see that the first term on the right in (3.5) is a martingale. Thus $M_i^2(t) - \sum_{s\le t} |\Delta M_i(s)|^2$ is a martingale. Since $M_i^2(t) - \langle M_i \rangle_t$ is a martingale, summing over i completes the proof. □

If H_s is of the form

$$H_s(\omega) = \sum_{i=1}^n K_i(\omega) 1_{(a_i,b_i]}(s), \tag{3.6}$$

where each K_i is bounded and $\mathcal{F}_{a_i}$ measurable, define the stochastic integral by

$$N_t = \int_0^t H_s\,dM_s = \sum_{i=1}^n K_i[M_{b_i\wedge t} - M_{a_i\wedge t}].$$

Very similar proofs to those in the Brownian motion case (see Chapter 10 of Bass, *Stochastic Processes*), show that the left hand side will be a martingale and (with $[\cdot]$ instead of $\langle\cdot\rangle$), $N_t^2 - [N]_t$ is a martingale.

If H is $\mathcal{P}$-measurable and $\mathbb{E}\int_0^\infty H_s^2\,d[M]_s < \infty$, approximate H by integrands H_s^n of the form (3.6) so that

$$\mathbb{E}\int_0^\infty (H_s - H_s^n)^2\,d[M]_s \to 0$$

and define N_t^n as the stochastic integral of H^n with respect to M_t. By almost the same proof as that of the construction of stochastic integrals with respect to Brownian motion, the martingales N_t^n converge in L^2. We call the limit $N_t = \int_0^t H_s\,dM_s$ the *stochastic integral* of H with respect to M. A subsequence of the N^n converges uniformly over $t \ge 0$, a.s., and therefore the limit has paths that are right continuous with left limits. The same arguments as those for Brownian motion apply to prove that the stochastic integral is a martingale and

$$[N]_t = \int_0^t H_s^2\,d[M]_s.$$

A consequence of this last equation is that

$$\mathbb{E}\Big(\int_0^t H_s\,dM_s\Big)^2 = \mathbb{E}\int_0^t H_s^2\,d[M]_s. \tag{3.7}$$

3.3 Itô's formula

We will first prove Itô's formula for a special case, namely, we suppose $X_t = M_t + A_t$, where M_t is a square integrable martingale and A_t is a process of bounded variation whose total variation is integrable. The extension to semimartingales without the integrability conditions will be done later in the chapter (in Section 3.5) and is easy. Define $\langle X^c\rangle_t$ to be $\langle M^c\rangle_t$.

Theorem 3.7 *Suppose $X_t = M_t + A_t$, where M_t is a square integrable martingale and A_t is a process with paths of bounded variation whose total variation is integrable. Suppose f is C_b^2 on $\mathbb{R}$. Then*

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_0^t f'(X_{s-})\,dX_s + \tfrac12\int_0^t f''(X_{s-})\,d\langle X^c\rangle_s \\ &+ \sum_{s\le t}[f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s]. \end{aligned} \tag{3.8}$$

Proof. The proof will be given in several steps. Set

$$S(t) = \int_0^t f'(X_{s-})\,dX_s, \qquad Q(t) = \tfrac12\int_0^t f''(X_{s-})\,d\langle X^c\rangle_s,$$

and

$$J(t) = \sum_{s\le t}[f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s].$$

We use these letters as mnemonics for "stochastic integral term," "quadratic variation term," and "jump term," resp.

Step 1: Suppose X_t has a single jump at time T which is a totally inaccessible stopping time and there exists $N > 0$ such that $|\Delta M_T| + |\Delta A_T| \le N$ a.s.

Let $C_t = \Delta M_T 1_{(t \geq T)}$ and let $\widetilde{C}_t$ be the compensator. If we replace M_t by $M_t - C_t + \widetilde{C}_t$ and A_t by $A_t + C_t - \widehat{C}_t$, we may assume that M_t is continuous.

Let $B_t = \Delta X_T 1_{(t \geq T)}$. Set $\widehat{X}_t = X_t - B_t$ and $\widehat{A}_t = A_t - B_t$. Then $\widehat{X}_t = M_t + \widehat{A}_t$ and $\widehat{X}_t$ is a continuous process that agrees with X_t up to but not including time T. We have $\widehat{X}_{s-} = \widehat{X}_s$ and $\Delta \widehat{X}_s = 0$ if $s \leq T$. By Ito's formula for continuous processes,

$$\begin{aligned} f(\widehat{X}_t) &= f(\widehat{X}_0) + \int_0^t f'(\widehat{X}_s)\, d\widehat{X}_s + \tfrac{1}{2}\int_0^t f''(\widehat{X}_s)\, d\langle M\rangle_s \\ &= f(\widehat{X}_0) + \int_0^t f'(\widehat{X}_{s-})\, d\widehat{X}_s + \tfrac{1}{2}\int_0^t f''(\widehat{X}_{s-})\, d\langle \widetilde{X}^c\rangle_s \\ &\qquad + \sum_{s\leq t} [f(\widehat{X}_s) - f(\widehat{X}_{s-}) - f'(\widehat{X}_{s-})\Delta \widehat{X}_s], \end{aligned}$$

since the sum on the last line is zero. For $t < T$, $\widehat{X}_t$ agrees with X_t. At time T, $f(X_t)$ has a jump of size $f(X_T) - f(X_{T-})$. The integral with respect to $\widehat{X}$, $S(t)$, will jump $f'(X_{T-})\Delta X_T$, $Q(t)$ does not jump at all, and $J(t)$ jumps $f(X_T) - f(X_{T-}) - f'(X_{T-})\Delta X_T$. Therefore both sides of (3.8) jump the same amount at time T, and hence in this case we have (3.8) holding for $t \leq T$.

Step 2: Suppose there exist times $T_1 < T_2 < \cdots$ with $T_n \to \infty$, each T_i is a totally inaccessible stopping time stopping time, for each i, there exists $N_i > 0$ such that $|\Delta M_{T_i}|$ and $|\Delta A_{T_i}|$ are bounded by N_i, and X_t is continuous except at the times $T_1, T_2, \ldots$. Let $T_0 = 0$.

Fix i for the moment. Define $X'_t = X_{(t-T_i)^+}$, define A'_t and M'_t similarly, and apply Step 1 to X' at time $T_i + t$. We have for $T_i \leq t \leq T_{i+1}$

$$\begin{aligned} f(X_t) &= f(X_{T_i}) + \int_{T_i}^t f'(X_{s-})\, dX_s + \tfrac{1}{2}\int_{T_i}^t f''(X_{s-})\, d\langle X^c\rangle_s \\ &\qquad + \sum_{T_i < s \leq t} [f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s]. \end{aligned}$$

Thus for any t we have

$$\begin{aligned} f(X_{T_{i+1}\wedge t}) &= f(X_{T_i \wedge t}) + \int_{T_i\wedge t}^{T_{i+1}\wedge t} f'(X_{s-})\, dX_s + \tfrac{1}{2}\int_{T_i\wedge t}^{T_{i+1}\wedge t} f''(X_{s-})\, d\langle X^c\rangle_s \\ &\qquad + \sum_{T_i\wedge t < s \leq T_{i+1}\wedge t} [f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s]. \end{aligned}$$

Summing over i, we have (3.8) for each t.

Step 3: We now do the general case. As in the paragraphs preceding Theorem 3.5, we can find stopping times $S_1, S_2, \ldots$ such that each jump of X occurs at one of the times S_i and so that for each i, there exists $N_i > 0$ such that $|\Delta M_{S_i}| + |\Delta A_{S_i}| \le N_i$. Moreover each S_i is a totally inaccessible stopping time. Let M be decomposed into M^c and M_i as in Theorem 3.5 and let

$$A_t^c = A_t - \sum_{i=1}^{\infty} \Delta A_{S_i} 1_{(t \ge S_i)}.$$

Since A_t is of bounded variation, then A^c will be finite and continuous. Define

$$M_t^n = M_t^c + \sum_{i=1}^{n} M_i(t)$$

and

$$A_t^n = A_t^c + \sum_{i=1}^{n} \Delta A_{S_i} 1_{(t \ge S_i)},$$

and let $X_t^n = M_t^n + A_t^n$. We already know that M^n converges uniformly over $t \ge 0$ to M in L^2. If we let $B_t^n = \sum_{i=1}^n (\Delta A_{S_i})^+ 1_{(t \ge S_i)}$ and $C_t^n = \sum_{i=1}^n (\Delta A_{S_i})^- 1_{(t \ge S_i)}$ and let $B_t = \sup_n B_t^n$, $C_t = \sup_n C_t^n$, then the fact that A has paths of bounded variation implies that with probability one, $B_t^n \to B_t$ and $C_t^n \to C_t$ uniformly over $t \ge 0$ and $A_t = B_t - C_t$. In particular, we have convergence in total variation norm:

$$\mathbb{E} \int_0^\infty |d(A_t^n) - A_t)| \to 0.$$

We define $S^n(t)$, $Q^n(t)$, and $J^n(t)$ analogously to $S(t)$, $Q(t)$, and $J(t)$, resp. By applying Step 2 to X^n, we have

$$f(X_t^n) = f(X_0^n) + S^n(t) + Q^n(t) + J^n(t),$$

and we need to show convergence of each term. We now examine the various terms.

Uniformly in t, X_t^n converges to X_t in probability, that is,

$$\mathbb{P}(\sup_{t \ge 0} |X_t^n - X_t| > \varepsilon) \to 0$$

as $n \to \infty$ for each $\varepsilon > 0$. Since $\int_0^t d\langle M^c \rangle_s < \infty$, by dominated convergence

$$\int_0^t f''(X^n_{s-})\, d\langle M^c \rangle_s \to \int_0^t f''(X_{s-})\, d\langle M^c \rangle_s$$

in probability. Therefore $Q^n(t) \to Q(t)$ in probability. Also, $f(X^n_t) \to f(X_t)$ and $f(X_0) \to f(X_0)$, both in probability.

We now show $S^n(t) \to S(t)$. Write

$$\begin{aligned}
\int_0^t f'(X^n_{s-})\, dA^n_s &- \int_0^t f'(X_{s-})\, dA_s \\
&= \Big[\int_0^t f'(X^n_{s-})\, dA^n_s - \int_0^t f'(X^n_{s-})\, dA_s \Big] \\
&\qquad + \Big[\int_0^t f'(X^n_{s-})\, dA_s - \int_0^t f'(X_{s-})\, dA_s \Big] \\
&= I^n_1 + I^n_2.
\end{aligned}$$

We see that

$$|I^n_1| \le \|f'\|_\infty \int_0^t |dA^n_s - dA_s| \to 0$$

as $n \to \infty$, while by dominated convergence, $|I^n_2|$ also tends to 0.

We next look at the stochastic integral part of $S^n(t)$.

$$\begin{aligned}
\int_0^t f'(X^n_{s-})\, dM^n_s &- \int_0^t f'(X_{s-})\, dM_s \\
&= \Big[\int_0^t f'(X^n_{s-})\, dM^n_s - \int_0^t f'(X_{s-})\, dM^n_s \Big] \\
&\qquad + \Big[\int_0^t f'(X_{s-})\, dM^n_s - \int_0^t f'(X_{s-})\, dM_s \Big] \\
&= I^n_3 + I^n_4.
\end{aligned}$$

The L^2 norm of I^n_3 is bounded by

$$\mathbb{E} \int_0^t |f'(X^n_{s-}) - f'(X_{s-})|^2\, d[M^n]_s \le \mathbb{E} \int_0^t |f'(X^n_{s-}) - f'(X_{s-})|^2\, d[M]_s,$$

which goes to 0 by dominated convergence. Also

$$I^n_4 = \int_0^t f'(X_{s-}) \sum_{i=n+1}^{\infty} dM_i(s),$$

so using the orthogonality lemma (Lemma 3.4), the L^2 norm of I_4^n is less than

$$\|f'\|_\infty^2 \sum_{i=n+1}^\infty \mathbb{E}\,[M_i]_\infty \le \|f'\|_\infty^2 \sum_{i=n+1}^\infty \mathbb{E}\,M_i(\infty)^2,$$

which goes to 0 as $n \to \infty$.

Finally, we look at the convergence of J^n. The idea here is to break both $J(t)$ and $J^n(t)$ into two parts, the jumps that might be relatively large (jumps at times S_i for $i \le N$ where N will be chosen appropriately) and the remaining jumps. Let $N > 1$ be chosen later.

$$\begin{aligned}
J(t) - J^n(t) &= \sum_{s\le t}[f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s] \\
&\qquad - \sum_{s\le t}[f(X_s^n) - f(X_{s-}^n) - f'(X_{s-}^n)\Delta X_s^n] \\
&= \sum_{\{i:S_i\le t\}} [f(X_{S_i}) - f(X_{S_i-}) - f'(X_{S_i-})\Delta X_{S_i}] \\
&\qquad - \sum_{\{i:S_i\le t\}} [f(X_{S_i}^n) - f(X_{S_i-}^n) - f'(X_{S_i-}^n)\Delta X_{S_i}^n] \\
&= \sum_{\{i>N:S_i\le t\}} [f(X_{S_i}) - f(X_{S_i-}) - f'(X_{S_i-})\Delta X_{S_i}] \\
&\qquad - \sum_{\{i>N:S_i\le t\}} [f(X_{S_i}^n) - f(X_{S_i-}^n) - f'(X_{S_i-}^n)\Delta X_{S_i}^n] \\
&\qquad + \sum_{\{i\le N,S_i\le t\}} \Big\{[f(X_{S_i}) - f(X_{S_i-}) - f'(X_{S_i-})\Delta X_{S_i}] \\
&\qquad\qquad - [f(X_{S_i}^n) - f(X_{S_i-}^n) - f'(X_{S_i-}^n)\Delta X_{S_i}^n]\Big\} \\
&= I_5^N - I_6^{n,N} + I_7^{n,N}.
\end{aligned}$$

By the fact that M and A are right continuous with left limits, $|\Delta M_{S_i}| \le 1/2$ and $|\Delta A_{S_i}| \le 1/2$ if i is large enough (depending on ω), and then $|\Delta X_{S_i}| \le 1$, and also

$$\begin{aligned}
|\Delta X_{S_i}|^2 &\le 2|\Delta M_{S_i}|^2 + 2|\Delta A_{S_i}|^2 \\
&\le 2|\Delta M_{S_i}|^2 + |\Delta A_{S_i}|.
\end{aligned}$$

We have

$$|I_5^N| \le \|f''\|_\infty \sum_{i>N, S_i\le t} (\Delta X_{S_i})^2$$

and

$$|I_6^{n,N}| \le \|f''\|_\infty \sum_{n\ge i>N, S_i\le t} (\Delta X_{S_i})^2.$$

Since $\sum_{i=1}^\infty |\Delta M_{S_i}|^2 \le [M]_\infty < \infty$ and $\sum_{i=1}^\infty |\Delta A_{S_i}| < \infty$, then given $\varepsilon > 0$, we can choose N large such that

$$\mathbb{P}(|I_5^N| + |I_6^{n,N}| > \varepsilon) < \varepsilon.$$

Once we choose N, we then see that $I_7^{n,N}$ tends to 0 in probability as $n \to \infty$, since X_t^n converges in probability to X_t uniformly over $t \ge 0$. We conclude that $J^n(t)$ converges to $J(t)$ in probability as $n \to \infty$.

This completes the proof. □

3.4 The reduction theorem

Let M be a process adapted to $\{\mathcal{F}_t\}$. If there exist stopping times T_n increasing to ∞ such that each process $M_{t\wedge T_n}$ is a uniformly integrable martingale, we say M is a *local martingale*. If each $M_{t\wedge T_n}$ is a square integrable martingale, we say M is a *locally square integrable martingale*. We say a stopping time T *reduces* a process M if $M_{t\wedge T}$ is a uniformly integrable martingale.

Lemma 3.8 *(1) The sum of two local martingales is a local martingale.*

(2) If S and T both reduce M, then so does $S \vee T$.

(3) If there exist times $T_n \to \infty$ such that $M_{t\wedge T_n}$ is a local martingale for each n, then M is a local martingale.

Proof. (1) If the sequence S_n reduces M and the sequence T_n reduces N, then $S_n \wedge T_n$ will reduce $M + N$.

(2) $M_{t\wedge(S\vee T)}$ is bounded in absolute value by $|M_{t\wedge T}| + |M_{t\wedge S}|$. Both $\{|M_{t\wedge T}|\}$ and $\{|M_{t\wedge S}|\}$ are uniformly integrable families of random variables.

Now use that the sum of two uniformly integrable families is uniformly integrable.

(3) Let S_{nm} be a family of stopping times reducing $M_{t\wedge T_n}$ and let $S'_{nm} = S_{nm} \wedge T_n$. Renumber the stopping times into a single sequence $R_1, R_2, \ldots$ and let $H_k = R_1 \vee \cdots \vee R_k$. Note $H_k \uparrow \infty$. To show that H_k reduces M, we need to show that R_i reduces M and use (2). But $R_i = S'_{nm}$ for some m, n, so $M_{t\wedge R_i} = M_{t\wedge S_{nm}\wedge T_n}$ is a uniformly integrable martingale. □

Let M be a local martingale with $M_0 = 0$. We say that a stopping time T *strongly reduces* M if T reduces M and the martingale $\mathbb{E}[\,|M_T| \mid \mathcal{F}_s]$ is bounded on $[0, T)$, that is, there exists $K > 0$ such that

$$\sup_{0\le s<T} \mathbb{E}[\,|M_T| \mid \mathcal{F}_s] \le K, \qquad \text{a.s.}$$

Lemma 3.9 *(1) If T strongly reduces M and $S \le T$, then S strongly reduces M.*

(2) If S and T strongly reduce M, then so does $S \vee T$.

(3) If Y_∞ is integrable, then $\mathbb{E}[\mathbb{E}[Y_\infty \mid \mathcal{F}_T] \mid \mathcal{F}_S] = \mathbb{E}[Y_\infty \mid \mathcal{F}_{S\wedge T}]$.

Proof. (1) Note $\mathbb{E}[\,|M_S| \mid \mathcal{F}_s] \le \mathbb{E}[\,|M_T| \mid \mathcal{F}_s]$ by Jensen's inequality, hence S strongly reduces M.

(2) It suffices to show that $\mathbb{E}[|M_{S\vee T}| \mid \mathcal{F}_t]$ is bounded for $t < T$, since by symmetry the same will hold for $t < S$. For $t < T$ this expression is bounded by

$$\mathbb{E}[\,|M_T| \mid \mathcal{F}_t] + \mathbb{E}[\,|M_S| 1_{(S>T)} \mid \mathcal{F}_t].$$

The first term is bounded since T strongly reduces M. For the second term, if $t < T$,

$$\begin{aligned} 1_{(t<T)}\mathbb{E}[\,|M_S|1_{(S>T)} \mid \mathcal{F}_t] &= \mathbb{E}[\,|M_S|1_{(S>T)}1_{(t<T)} \mid \mathcal{F}_t] \\ &\le \mathbb{E}[\,|M_S|1_{(t<S)} \mid \mathcal{F}_t] \\ &= \mathbb{E}[\,|M_S| \mid \mathcal{F}_t]1_{(t<S)}, \end{aligned}$$

which in turn is bounded since S strongly reduces M.

(3) Let Y_t be the right continuous version of $\mathbb{E}[X \mid \mathcal{F}_t]$. We thus need to show that $\mathbb{E}[Y_S \mid \mathcal{F}_T] = Y_{S\wedge T}$. The right hand side is $\mathcal{F}_{S\wedge T}$ measurable and $\mathcal{F}_{S\wedge T} \subset \mathcal{F}_T$. We thus need to show that if $A \in \mathcal{F}_T$, then

$$\mathbb{E}[Y_S; A] = \mathbb{E}[Y_{S\wedge T}; A].$$

Let $B = (S \le T)$. We will show

$$\mathbb{E}[Y_S; A\cap B] = \mathbb{E}[Y_{S\wedge T}; A\cap B] \tag{3.9}$$

and

$$\mathbb{E}[Y_S; A\cap B^c] = \mathbb{E}[Y_{S\wedge T}; A\cap B^c] \tag{3.10}$$

Adding (3.9) and (3.10) will achieve our goal.

Since $Y_S = Y_{S\wedge T}$ on B, the right hand side of (3.9) is equal to $\mathbb{E}[Y_S; A\cap B]$ as required.

For (3.10), $S > T$ on B^c, so $S = S \vee T$ on B^c. Also $A \cap B^c \in \mathcal{F}_T \subset \mathcal{F}_{S\vee T}$. Since Y is a martingale,

$$\mathbb{E}[Y_S; A\cap B^c] = \mathbb{E}[Y_{S\vee T}; A\cap B^c] = \mathbb{E}[Y_T; A\cap B^c] = \mathbb{E}[Y_{S\wedge T}; A\cap B^c],$$

which is (3.10)

□

Lemma 3.10 *If M is a local martingale with $M_0 = 0$, then there exist stopping times $T_n \uparrow \infty$ that strongly reduce M.*

Proof. Let $R_n \uparrow \infty$ be a sequence reducing M. Let

$$S_{nm} = R_n \wedge \inf\{t : \mathbb{E}[\,|M_{R_n}| \mid \mathcal{F}_t] \ge m\}.$$

Arrange the stopping times S_{nm} into a single sequence $\{U_n\}$ and let $T_n = U_1 \vee \cdots \vee U_n$. In view of the preceding lemmas, we need to show U_i strongly reduces M, which will follow if S_{nm} does for each n and m.

Let $Y_t = \mathbb{E}\,[\,|M_{R_n}|\mid \mathcal{F}_t]$, where we take a version whose paths are right continuous with left limits. Y is bounded by m on $[0, S_{nm})$. By Jensen's inequality for conditional expectations and Lemma 3.9

$$\begin{aligned}
\mathbb{E}\,[\,|M_{S_{nm}}|1_{(t<S_{nm})}\mid \mathcal{F}_t] &\le \mathbb{E}\,[\,|\mathbb{E}\,[\,|M_{R_n}|\mid \mathcal{F}_{S_{nm}}]|1_{(t<S_{nm})}\mid \mathcal{F}_t]\\
&= \mathbb{E}\,[\,\mathbb{E}\,[\,|M_{R_n}|1_{(t<S_{nm})}\mid \mathcal{F}_{S_{nm}}]\mid \mathcal{F}_t]\\
&= \mathbb{E}\,[\,|M_{R_n}|1_{(t<S_{nm})}\mid \mathcal{F}_{S_{nm}\wedge t}]\\
&= Y_{S_{nm}\wedge t}1_{(t<S_{nm})}\\
&= Y_t 1_{(t<S_{nm})} \le m.
\end{aligned}$$

We used that $1_{(t<S_{nm})}$ is $\mathcal{F}_{S_{nm}\wedge t}$ measurable; to see that we have by Lemma 3.9(3) that

$$\begin{aligned}
\mathbb{E}\,[1_{(t<S_{nm})}\mid \mathcal{F}_{S_{nm}\wedge t}] &= \mathbb{E}\,[\mathbb{E}\,[1_{(t<S_{nm})}\mid \mathcal{F}_{S_{nm}}]\;\mid \mathcal{F}_t] = \mathbb{E}\,[1_{(t<S_{nm})}\mid \mathcal{F}_t]\\
&= 1_{(t<S_{nm})}.
\end{aligned}$$

We are done. □

Our main theorem of this section is the following.

Theorem 3.11 *Suppose M is a local martingale. Then there exist stopping times $T_n \uparrow \infty$ such that $M_{t\wedge T_n} = U^n_t + V^n_t$, where each U^n is a square integrable martingale and each V^n is a martingale whose paths are of bounded variation and such that the total variation of the paths of V_n is integrable. Moreover, $U^n_t = U^n_{T_n}$ and $V^n_t = V^n_{T_n}$ for $t \ge T_n$.*

The last sentence of the statement of the theorem says that U^n and V^n are both constant from time T_n on.

Proof. It suffices to prove that if M is a local martingale with $M_0 = 0$ and T strongly reduces M, then $M_{t\wedge T}$ can be written as $U + V$ with U and V of the described form. Thus we may assume $M_t = M_T$ for $t \ge T$, $|M_T|$ is integrable, and $\mathbb{E}\,[\,|M_T|\mid \mathcal{F}_t]$ is bounded, say by K, on $[0, T)$.

Let $A_t = M_T 1_{(t\ge T)} = M_t 1_{(t\ge T)}$, let $\widetilde{A}$ be the compensator of A, let $V = A - \widetilde{A}$, and let $U = M - A + \widetilde{A}$. Then V is a martingale of bounded variation. We compute the expectation of the total variation of V. Let $B_t = M_T^+ 1_{(t\ge T)}$ and $C_t = M_T^- 1_{(t\ge T)}$. Then the expectation of the total variation of A is bounded by $\mathbb{E}\,|M_T| < \infty$ and the expectation of the total variation of $\widetilde{A}$ is bounded by

$$\mathbb{E}\,\widetilde{B}_\infty + \mathbb{E}\,\widetilde{C}_\infty = \mathbb{E}\,B_\infty + \mathbb{E}\,C_\infty \le \mathbb{E}\,|M_T| < \infty.$$

We need to show U is square integrable. Note

$$\begin{aligned}|M_t - A_t| &= |M_t| 1_{(t<T)} = |\mathbb{E}\,[M_\infty \mid \mathcal{F}_t]|\, 1_{(t<T)} \\ &= |\mathbb{E}\,[\mathbb{E}\,[M_\infty \mid \mathcal{F}_{T\vee t}] \mid \mathcal{F}_t]|\, 1_{(t<T)} = |\mathbb{E}\,[M_{T\vee t} \mid \mathcal{F}_t]|\, 1_{(t<T)} \\ &= |\mathbb{E}\,[M_T \mid \mathcal{F}_t]|\, 1_{(t<T)} \le \mathbb{E}\,[\,|M_T| \mid \mathcal{F}_t] 1_{(t<T)} \le K.\end{aligned}$$

Therefore it suffices to show $\widetilde{A}$ is square integrable.

Our hypotheses imply that $\mathbb{E}\,[M_T^+ \mid \mathcal{F}_t]$ is bounded by K on $[0, T)$, hence $\mathbb{E}\,[B_\infty - B_t \mid \mathcal{F}_t]$ is bounded, and so $\mathbb{E}\,\widetilde{B}_\infty^2 < \infty$. Similarly, $\mathbb{E}\,\widetilde{C}_\infty^2 < \infty$. Since $A = B - C$, then $\widetilde{A} = \widetilde{B} - \widetilde{C}$, and it follows that $\sup_{t\ge 0} \widetilde{A}_t$ is square integrable. □

3.5 Semimartingales

We define a *semimartingale* to be a process of the form $X_t = X_0 + M_t + A_t$, where X_0 is finite, a.s., and is $\mathcal{F}_0$ measurable, M_t is a local martingale, and A_t is a process whose paths have bounded variation on $[0, t]$ for each t.

If M_t is a local martingale, let T_n be a sequence of stopping times as in Theorem 3.11. We set $M^c_{t\wedge T_n} = (U^n)^c_t$ for each n and

$$[M]_{t\wedge T_n} = \langle M^c \rangle_{t\wedge T_n} + \sum_{s\le t\wedge T_n} \Delta M_s^2.$$

It is easy to see that these definitions are independent of how we decompose of M into $U^n + V^n$ and of which sequence of stopping times T_n strongly reducing M we choose. We define $\langle X^c \rangle_t = \langle M^c \rangle_t$ and define

$$[X]_t = \langle X^c \rangle_t + \sum_{s\le t} \Delta X_s^2.$$

We say an adapted process H is *locally bounded* if there exist stopping times $S_n \uparrow \infty$ and constants K_n such that on $[0, S_n]$ the process H is bounded by K_n. If X_t is a semimartingale and H is a locally bounded predictable process, define $\int_0^t H_s\,dX_s$ as follows. Let $X_t = X_0 + M_t + A_t$. If $R_n = T_n \wedge S_n$, where the T_n are as in Theorem 3.11 and the S_n are the stopping times used in the definition of locally bounded, set $\int_0^{t\wedge R_n} H_s\,dM_s$ to be the stochastic integral as defined in Section 3.2. Define $\int_0^{t\wedge R_n} H_s\,dA_s$ to be the usual Lebesgue-Stieltjes integral. Define the stochastic integral with respect to X as the sum of these two. Since $R_n \uparrow \infty$, this defines $\int_0^t H_s\,dX_s$ for all t. One needs to check that the definition does not depend on the decomposition of X into M and A nor on the choice of stopping times R_n.

We now state the general Itô formula.

Theorem 3.12 *Suppose X is a semimartingale and f is C^2. Then*

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})\,dX_s + \tfrac{1}{2}\int_0^t f''(X_{s-})\,d\langle X^c\rangle_s + \sum_{s\le t}[f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s].$$

Proof. First suppose f has bounded first and second derivatives. Let T_n be stopping times strongly reducing M_t, let $S_n = \inf\{t : \int_0^t |dA_s| \ge n\}$, let $R_n = T_n \wedge S_n$, and let $X_t^n = X_{t\wedge R_n} - \Delta A_{R_n}$. Since the total variation of A_t is bounded on $[0, R_n)$, it follows that X^n is a semimartingale which is the sum of a square integrable martingale and a process whose total variation is integrable. We apply Theorem 3.7 to this process. X_t^n agrees with X_t on $[0, R_n)$. As in the proof of Theorem 3.7, by looking at the jump at time R_n, both sides of Itô's formula jump the same amount at time R_n, and so Itô's formula holds for X_t on $[0, R_n]$. If we now only assume that f is C^2, we approximate f by a sequence f_m of functions that are C^2 and whose first and second derivatives are bounded, and then let $m \to \infty$; we leave the details to the reader. Thus Itô's formula holds for t in the interval $[0, R_n]$ and for f without the assumption of bounded derivatives. Finally, we observe that $R_n \to \infty$, so except for a null set, Itô's formula holds for each t. □

The proof of the following corollary is similar to the proof of Itô's formula.

Corollary 3.13 *If $X_t = (X^1_t, \ldots, X^d_t)$ is a process taking values in $\mathbb{R}^d$ such that each component is a semimartingale, and f is a C^2 function on $\mathbb{R}^d$, then*

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_0^t \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_{s-})\, dX^i_s \\ &+ \tfrac{1}{2} \int_0^t \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{s-})\, d\langle (X^i)^c, (X^j)^c \rangle_s \\ &+ \sum_{s \le t} \Big[f(X_s) - f(X_{s-}) - \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_{s-}) \Delta X^i_s \Big], \end{aligned}$$

where $\langle Y, Z \rangle_t = \frac{1}{2}[\langle Y+Z \rangle_t - \langle Y \rangle_t - \langle Z \rangle_t]$.

If X and Y are real-valued semimartingales, define

$$[X,Y]_t = \tfrac{1}{2}([X+Y]_t - [X]_t - [Y]_t). \tag{3.11}$$

The following corollary is the product formula for semimartingales with jumps.

Corollary 3.14 *If X and Y are semimartingales of the above form,*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_{s-}\, dY_s + \int_0^t Y_{s-}\, dX_s + [X,Y]_t.$$

Proof. Apply Theorem 3.12 with $f(x) = x^2$. Since in this case

$$f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s = \Delta X^2_s,$$

we obtain

$$X^2_t = X^2_0 + 2 \int_0^t X_{s-}\, dX_s + [X]_t. \tag{3.12}$$

Applying (3.12) with X replaced by Y and by $X+Y$ and using

$$X_t Y_t = \tfrac{1}{2}[(X_t + Y_t)^2 - X^2_t - Y^2_t]$$

gives our result. $\square$

3.6 The Girsanov theorem

Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probability measures, that is, $\mathbb{P}$ and $\mathbb{Q}$ are mutually absolutely continuous. Let M_∞ be the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ and let $M_t = \mathbb{E}[M_\infty \mid \mathcal{F}_t]$. The martingale M_t is uniformly integrable since $M_\infty \in L^1(\mathbb{P})$. Once a non-negative martingale hits zero, it is easy to see that it must be zero from then on. Since $\mathbb{Q}$ and $\mathbb{P}$ are equivalent, then $M_\infty > 0$, a.s., and so M_t never equals zero, a.s. Observe that M_T is the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ on $\mathcal{F}_T$.

If $A \in \mathcal{F}_t$, we have

$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}[M_\infty; A] = \mathbb{E}_{\mathbb{P}}[M_t; A],$$

using that M is a martingale.

Theorem 3.15 *Suppose X is a local martingale with respect to $\mathbb{P}$. Then $X_t - D_t$ is a local martingale with respect to $\mathbb{Q}$, where*

$$D_t = \int_0^t \frac{1}{M_s}\, d[X, M]_s.$$

Note that in the formula for D, we are using a Lebesgue-Stieltjes integral.

Proof. Since $\mathbb{E}_{\mathbb{Q}}[X_t - D_t; A] = \mathbb{E}_{\mathbb{P}}[M_t(X_t - D_t); A]$ if $A \in \mathcal{F}_t$ and the same with t replaced by s, it suffices to show that $M_t(X_t - D_t)$ is a local martingale with respect to $\mathbb{P}$. By Corollary 3.14,

$$\begin{aligned} d(M(X-D))_t &= (X-D)_{t-}\, dM_t + M_{t-}\, dX_t - M_{t-}\, dD_t \\ &\quad + d[M, X-D]_t. \end{aligned}$$

The first two terms on the right are local martingales with respect to $\mathbb{P}$. Since D is of bounded variation, the continuous part of D is zero, hence

$$[M, D]_t = \sum_{s \le t} \Delta M_s \Delta D_s = \int_0^t \Delta M_s\, dD_s.$$

Thus

$$M_t(X_t - D_t) = \text{ local martingale } + [M, X]_t - \int_0^t M_s\, dD_s.$$

Using the definition of D shows that $M_t(X_t - D_t)$ is a local martingale. $\square$

Chapter 4

Stochastic differential equations

4.1 Poisson point processes

Poisson point processes are random measures that are related to Poisson processes. Poisson point processes are also useful in the study of excursions, even excursions of a continuous process such as Brownian motion, and they arise when studying stochastic differential equations with jumps.

Let $\mathcal{S}$ be a metric space, $\mathcal{G}$ the collection of Borel subsets of $\mathcal{S}$, and λ a measure on $(\mathcal{S}, \mathcal{G})$.

Definition 4.1 *We say a map*

$$N : \Omega \times [0, \infty) \times \mathcal{G} \to \{0, 1, 2, \ldots\}$$

(writing $N_t(A)$ for $N(\omega, t, A)$) is a Poisson point process *if*

(1) for each Borel subset A of $\mathcal{S}$ with $\lambda(A) < \infty$, the process $N_t(A)$ is a Poisson process with parameter $\lambda(A)$, and

(2) for each t and ω, $N(t, \cdot)$ is a measure on $\mathcal{G}$.

A model to keep in mind is where $\mathcal{S} = \mathbb{R}$ and λ is Lebesgue measure. For each ω there is a collection of points $\{(s, z)\}$ (where the collection depends on ω). The number of points in this collection with $s \leq t$ and z in a subset A is $N_t(A)(\omega)$. Since $\lambda(\mathbb{R}) = \infty$, there are infinitely many points in every time interval.

Another example is to let X be a Lévy process and let $N_t(A)$ be the number of jumps of size A before time t. A consequence of the definition is that since $\lambda(\emptyset) = 0$, then $N_t(\emptyset)$ is a Poisson process with parameter 0; in other words, $N_t(\emptyset)$ is identically zero.

Our main result is that $N_t(A)$ and $N_t(B)$ are independent if A and B are disjoint.

Theorem 4.2 *Let $\{\mathcal{F}_t\}$ be a filtration satisfying the usual conditions. Let $\mathcal{S}$ be a metric space furnished with a positive measure λ. Suppose that $N_t(A)$ is a Poisson point process with respect to the measure λ. If $A_1, \ldots, A_n$ are pairwise disjoint measurable subsets of $\mathcal{S}$ with $\lambda(A_k) < \infty$ for $k = 1, \ldots, n$, then the processes $N_t(A_1), \ldots, N_t(A_n)$ are mutually independent.*

This is proved exactly the same way we proved that $N_t(A)$ and $N_t(B)$ are independent in the study of Lévy processes.

We now turn to stochastic integrals with respect to Poisson point processes. In the same way that a non-decreasing function on the reals gives rise to a measure, so $N_t(A)$ gives rise to a random measure $\mu(dt, dz)$ on the product σ-field $\mathcal{B}[0,\infty) \times \mathcal{G}$, where $\mathcal{B}[0,\infty)$ is the Borel σ-field on $[0,\infty)$; μ is determined by

$$\mu([0,t] \times A)(\omega) = N_t(A)(\omega);$$

Define a non-random measure ν on $\mathcal{B}[0,\infty) \times \mathcal{G}$ by $\nu([0,t] \times A) = t\lambda(A)$ for $A \in \mathcal{G}$. If $\lambda(A) < \infty$, then $\mu([0,t] \times A) - \nu([0,t] \times A)$ is the same as a Poisson process minus its mean, hence is locally a square integrable martingale.

We can define a stochastic integral with respect to the random measure $\mu - \nu$ as follows. Suppose $H(\omega, s, z)$ is of the form

$$H(\omega, s, z) = \sum_{i=1}^n K_i(\omega) 1_{(a_i, b_i]}(s) 1_{A_i}(z), \tag{4.1}$$

where for each i the random variable K_i is bounded and $\mathcal{F}_{a_i}$ measurable and $A_i \in \mathcal{G}$ with $\lambda(A_i) < \infty$. For such H we define

$$\begin{aligned} N_t &= \int_0^t \int H(\omega, s, z)\, d(\mu - \nu)(ds, dz) \qquad (4.2) \\ &= \sum_{i=1}^n K_i (\mu - \nu)(((a_i, b_i] \cap [0,t]) \times A_i). \end{aligned}$$

Let us assume without loss of generality that the A_i are disjoint. It is not hard to see that N_t is a martingale, that $N^c = 0$, and that

$$[N]_t = \int_0^t \int H(\omega, s, z)^2 \, \mu(ds, dz). \tag{4.3}$$

Since $\langle N \rangle_t$ must be predictable and all the jumps of N are totally inaccessible, it follows that $\langle N \rangle_t$ is continuous. Since $[N]_t - \langle N \rangle_t$ is a martingale, we conclude

$$\langle N \rangle_t = \int_0^t \int H(\omega, s, z)^2 \, \nu(ds, dz). \tag{4.4}$$

Suppose $H(s, z)$ is predictable process in the following sense: H is measurable with respect to the σ-field generated by all processes of the form (4.1). Suppose also that

$$\mathbb{E} \int_0^\infty \int_{\mathcal{S}} H(s, z)^2 \, \nu(ds, dz) < \infty.$$

Take processes H^n of the form (4.1) converging to H in the space L^2 with norm $(\mathbb{E} \int_0^\infty \int_{\mathcal{S}} H^2 \, d\nu)^{1/2}$. The corresponding $N_t^n = \int_0^t H^n(s, z) \, d(\mu - \nu)$ are easily seen to be a Cauchy sequence in L^2, and the limit N_t we call the *stochastic integral of H with respect to $\mu - \nu$*. As in the continuous case, we may prove that $\mathbb{E}\, N_t^2 = \mathbb{E}\, [N]_t = \mathbb{E}\, \langle N \rangle_t$, and it follows from this, (4.3), and (4.4) that

$$[N]_t = \int_0^t \int_{\mathcal{S}} H(s, z)^2 \, \mu(ds, dz), \qquad \langle N \rangle_t = \int_0^t \int_{\mathcal{S}} H(s, z)^2 \, \nu(ds, dz). \tag{4.5}$$

One may think of the stochastic integral as follows: if μ gives unit mass to a point at time t with value z, then N_t jumps at this time t and the size of the jump is $H(t, z)$.

4.2 The Lipschitz case

We consider the the stochastic differential equation (SDE)

$$X_t = x_0 + \int_0^t \int F(X_{s-}, z)[\mu(dz, ds) - \nu(dz, ds)]. \tag{4.6}$$

Theorem 4.3 *Suppose there is a constant c such that*

$$\int \sup_x F(x,z)^2\,\lambda(dz) < \infty$$

and

$$\int |F(x,z) - F(y,z)|^2\,\lambda(dz) \le c|x-y|^2.$$

Then there exists a solution to (4.6) *and the solution is pathwise unique.*

Proof. Let $X_t^0 = x_0$ and define

$$X_t^{n+1} = x_0 + \int_0^t \int F(X_{s-}^n, z)[\mu(dz,ds) - \nu(dz,ds)].$$

X^n is a martingale, and by Doob's inequality,

$$\begin{aligned}\mathbb{E}\sup_{s\le t}|X_s^{n+1} - X_s^n|^2 &\le 4\mathbb{E}\,|X_t^{n+1} - X_t^n|^2\\ &= 4\mathbb{E}\int_0^t \int [F(X_{s-}^n, z) - F(X_{s-}^{n-1}, z)|^2\,\lambda(dz)\,ds\\ &\le 4c\int_0^t |X_{s-}^n - X_{s-}^{n-1}|^2\,ds.\end{aligned}$$

If we let

$$g_n(t) = \mathbb{E}\sup_{s\le t}|X_s^n - X_s^{n-1}|^2,$$

we then have

$$g_n(t) \le A\int_0^t g_{n-1}(s)\,ds.$$

Since g_1 is easily seen to be bounded, say, by B, we have by induction that $g_2(t) \le ABt$, $g_3(t) \le A^2Bt^2/2$, and so on, and therefore

$$\sum_{n=1}^\infty g_n(t)^{1/2} < \infty.$$

It follows that $\sum \sup_{s\le t}|X_s^n - X_s^{n-1}|$ converges in L^2, and it is routine to see that therefore $X_t = \lim X_t^n$ exists and is the solution to (4.6).

If X and Y are two solutions and $g(t) = \mathbb{E} \sup_{s \le t} |X_s - Y_s|^2$, we obtain similarly that

$$g(t) \le A \int_0^t g(s)\, ds.$$

We may also assume that g is bounded by B for $t \le t_0$. We then obtain $g(t) \le ABt$, $g(t) \le A^2 B t^2/2!$, and so on, and therefore g must be identically zero, or we have pathwise uniqueness. □

4.3 Analogue of Yamada-Watanabe theorem

Our main result in this section is the analogue of the Yamada-Watanabe condition for diffusions. We suppose X is a symmetric stable process of index $\alpha \in (1,2)$ and look at the equation

$$dV_t = F(V_{t-})\, dX_t. \tag{4.7}$$

Theorem 4.4 *Suppose $\alpha \in (1,2)$, suppose F is bounded and continuous, and suppose ρ is a nondecreasing continuous function on $[0,\infty)$ with $\rho(0) = 0$ and $|F(x) - F(y)| \le \rho(|x-y|)$ for all $x, y \in \mathbb{R}$. If*

$$\int_{0+} \frac{1}{\rho(x)^\alpha} dx = \infty, \tag{4.8}$$

then the solution to the SDE (4.7) is pathwise unique.

We normalize our symmetric stable processes so that $\sum_{s \le t} 1_{\{|\Delta X_s| \in A\}}$ is a Poisson process with parameter $\int_A |y|^{-1-\alpha} dy$.

Recall (2.8), (2.9), and (2.10).

Suppose X_t is a symmetric stable process of index $\alpha \in (1,2)$. We define the Poisson point process μ by

$$\mu(A \times [0,t]) = \sum_{s \le t} 1_A(\Delta X_s),$$

the number of times before time t that X_t has jumps whose size lies in the set A. We define the compensating measure ν by

$$\nu(A) = \mathbb{E}\,\mu(A \times [0,1]) = \int_A \frac{1}{|x|^{1+\alpha}} dx.$$

Set

$$\mathcal{L}f(x) = \int [f(x+w) - f(x) - f'(x)w]\,|w|^{-1-\alpha} dw \tag{4.9}$$

for C_b^2 functions f, where C_b^2 is the set of C^2 functions f such that f, f, and f'' are bounded. There is convergence of the integral for large w since $\alpha > 1$. There is convergence for small w by using Taylor's theorem and the fact that $\alpha < 2$.

For C_b^2 functions $\mathcal{L}$ coincides with the infinitesimal generator of X. Let us explain this further.

If X_t is a Lévy process with Lévy measure m, then

$$\mathbb{E}\,e^{iu(X_t+x)} - e^{iux} = e^{iux}\Big(\mathbb{E}\,e^{iuX_t} - 1\Big) = e^{iux}\Big(e^{t\int [e^{iuh}-1-iuh1_{(|h|\le 1)}]\,m(dh)} - 1\Big).$$

Dividing by t and letting $t \to 0$,

$$\lim_{t\to 0} \frac{\mathbb{E}\,e^{iu(X_t+x)} - e^{iux}}{t} = e^{iux}\int [e^{iuh} - 1 - iuh1_{(|h|\le 1)}]\,m(dh).$$

Replacing u by $-u$, multiplyng by $\frac{1}{2\pi}\widehat{f}(u)$, and integrating u over $\mathbb{R}$, we get

$$\lim_{t\to 0} \frac{\mathbb{E}\,f(X_t+x) - f(x)}{t} = \int [f(x+h) - f(x) - hf'(x)1_{(|h|\le 1)}]\,m(dh) = \mathcal{L}f(x),$$

or

$$\lim_{t\to 0} \frac{P_t f(x) - f(x)}{t} = \mathcal{L}f(x),$$

provided $f \in C_b^2$, the collection of C^2 functions such that f, f', and f'' are bounded and provided one shows that it is valid to interchange the limit with the integral in two places (it is).

There is a slight discrepancy here in the definition of $\mathcal{L}$. We use that for $\alpha \in (1,2)$ and $m(dw) = |w|^{-(1+\alpha}\,dw$, we have that $\int w1_{(|w|>1)}\,m(dw)$ is integrable and equals 0.

Let

$$G_\lambda f(x) = \int_0^\infty e^{-\lambda s} P_s f(x)\, ds$$

where $P_s f(x) = \mathbb{E}\, f(X_s + x)$. We have

$$\begin{aligned} P_t G_\lambda f(x) &= \int_0^\infty e^{-\lambda s} P_{s+t} f(x)\, ds = e^{\lambda t} \int_0^\infty e^{-\lambda(s+t)} P_{s+t} f(x)\, ds \\ &= e^{\lambda t} \int_t^\infty e^{-\lambda s} P_s f(x)\, ds \\ &= (e^{\lambda t} - 1) \int_t^\infty e^{-\lambda s} P_s f(x)\, ds + \int_t^\infty e^{-\lambda s} P_s f(x)\, ds. \end{aligned}$$

So

$$\frac{P_t G_\lambda f(x) - G_\lambda f(x)}{t} = \frac{e^{\lambda t} - 1}{t} \int_t^\infty e^{-\lambda s} P_s f(x)\, ds - \int_0^t e^{-\lambda s} P_s f(x)\, ds.$$

Since $P_s f(x) = \mathbb{E}\, f(X_s + x) \to f(x)$ as $s \to 0$, we obtain

$$\mathcal{L} G_\lambda f(x) = \lambda G_\lambda f(x) - f(x).$$

Proposition 4.5 *Suppose $\alpha \in (1,2)$, f is in C_b^2, and*

$$Z_t = \int_0^t H_s\, dX_s,$$

where H_t is a bounded predictable process. Then

$$f(Z_t) = f(Z_0) + M_t + \int_0^t |H_s|^\alpha \mathcal{L} f(Z_{s-}) ds, \tag{4.10}$$

where M_t is a martingale.

Proof. Let $X_t^n = \sum_{s \le t} \Delta X_s 1_{(|\Delta X_s| \le n)}$ and $Y_t^n = X_t - X_t^n$. Then X_t^n is a Lévy process with symmetric Lévy measure which is equal to ν on $[-n, n]$ and 0 outside this interval. Hence X_t^n is a square integrable martingale and so $\int_0^t H_s dX_s^n$ is also a square integrable martingale since H is bounded. On the other hand

$$\mathbb{E} \Big| \int_0^t H_s dY_s^n \Big| \le \|H\|_\infty \mathbb{E} \sum_{s \le t} |\Delta X_s| 1_{(|\Delta X_s| > n)} < \infty$$

because $\alpha \in (1,2)$. The right hand side tends to 0 as $n \to \infty$ by dominated convergence. Therefore Z_t is the L^1 limit of the square integrable martingales $\int_0^t H_s\, dX_s^n$, and it follows that Z_t is a martingale.

Write $K(s,y)$ for $[f(Z_{s-}+H_s y)-f(Z_{s-})-f'(Z_{s-})H_s y]$. Note that $\Delta Z_s = H_s \Delta X_s$. Note also that $|K(s,y)|$ is bounded by a constant times $(|y| \wedge y^2)$. If $f \in C_b^2$, we have by Ito's formula that

$$\begin{aligned} f(Z_t) &= f(Z_0) + \int_0^t f'(Z_{s-})dZ_s + \sum_{s \le t} [f(Z_s) - f(Z_{s-}) - f'(Z_{s-})\Delta Z_s] \\ &= f(Z_0) + \int_0^t f'(Z_{s-})dZ_s + \int_0^t \int K(s,y)\mu(dy,ds) \\ &= f(Z_0) + M_t + \int_0^t \int K(s,y)\nu(dy)ds, \end{aligned}$$

where

$$M_t = \int_0^t f'(Z_{s-})dZ_s + \int_0^t \int K(s,y)(\mu(dy,ds) - \nu(dy)ds).$$

The first term on the right is a martingale by the argument of the first paragraph of this proof. For each m we have then that $\int_{|y|\le m} K(s,y)^2 \nu(dy)$ is bounded, and so for each m

$$W_t^m = \int_0^t \int_{|y|\le m} K(s,y)(\mu(dy,ds) - \nu(dy)ds)$$

is a martingale. Since $W_t^k - W_t^m$ is a martingale for each k, then

$$\begin{aligned} \mathbb{E} \int_0^t \int_{m<|y|\le k} |K(s,y)|(\mu(dy,ds) + \nu(dy)ds) &\le c_1 \int_0^t \int_{m<|y|\le k} |y|\nu(dy)ds \\ &\le c_2 m^{1-\alpha}, \end{aligned}$$

where c_1 and c_2 are positive finite constants not depending on m or k. Letting $k \to \infty$, we see that

$$\mathbb{E} \int_0^t \int_{m<|y|} |K(s,y)|(\mu(dy,ds) + \nu(dy)ds) \le c_2 m^{1-\alpha}.$$

Therefore M_t is the limit in L^1 of the martingales $\int_0^t f(Z_{s-})dZ_s + W_t^m$, and hence is itself a martingale.

We make the change of variable $w = H_s y$. Since $y \to H_s y$ is monotone if $H_s \neq 0$ we have that the integral with respect to $\nu(dy)$ is

$$\begin{aligned}\int [f(Z_{s-} + H_s y) - f(Z_{s-}) - f'(Z_{s-}) H_s y] \frac{dy}{|y|^{1+\alpha}} \\ &= \int [f(Z_{s-} + w) - f(Z_{s-}) - f'(Z_{s-}) w] |H_s|^{\alpha} |w|^{-1-\alpha} dw, \\ &= |H_s|^{\alpha} \mathcal{L} f(Z_{s-})\end{aligned}$$

if $H_s \neq 0$. This equality clearly also holds when $H_s = 0$. We therefore arrive at (4.10). □

We note for future reference that we have shown (take $H_s = 1$ a.s.) that $f(X_t) - f(X_0) - \int_0^t \mathcal{L} f(X_s)\, ds$ is a martingale if $f \in C_b^2$.

We now prove Theorem 4.4.

Proof of Theorem 4.4. Let Y^1 and Y^2 be any two solutions to (4.7), let $Z_t = Y_t^1 - Y_t^2$, and let $H_t = F(Y_{t-}^1) - F(Y_{t-}^2)$. Then $Z_t = \int_0^t H_s dX_s$.

Let a_n be numbers decreasing to 0 so that $\int_{a_{n+1}}^{a_n} \rho(x)^{-\alpha} dx = n$. For each n let h_n be a nonnegative C^2 function with support in $[a_{n+1}, a_n]$ whose integral is 1, and with $h_n(x) \leq 2/(n\rho(x)^{\alpha})$. This is possible since

$$\int_{a_{n+1}}^{a_n} 1/(n\rho(x)^{\alpha}) dx = 1.$$

Fix $\lambda > 0$, let $g_\lambda(x) = \int_0^\infty e^{-\lambda t} p_t(x, 0) dt$, where $p_t(x, y)$ is the transition density for X_t, and let $G_\lambda f(x) = \int f(y) g_\lambda(x-y) dy$. We have shown that that $g_\lambda(x)$ is bounded, and is continuous in x. Furthermore, $g_\lambda(x) < g_\lambda(0)$ if $x \neq 0$. Let $f_n(x) = G_\lambda h_n(x)$. By interchanging differentiation and integration and using translation invariance, f_n is in C_b^2 since h_n is C_b^2.

Define $A_t = \int_0^t |H_s|^{\alpha} ds$. By Ito's product formula,

$$\begin{aligned}\mathbb{E}\, e^{-\lambda A_t} f_n(Z_t) - f_n(0) \\ &= \mathbb{E} \int_0^t e^{-\lambda A_s} d[f_n(Z_s)] - \mathbb{E} \int_0^t e^{-\lambda A_s} \lambda |H_s|^{\alpha} f_n(Z_{s-}) ds \\ &= \mathbb{E} \int_0^t e^{-\lambda A_s} |H_s|^{\alpha} \mathcal{L} f_n(Z_{s-}) ds - \mathbb{E} \int_0^t e^{-\lambda A_s} \lambda |H_s|^{\alpha} f_n(Z_{s-}) ds.\end{aligned}$$

Since $\mathcal{L}f_n = \mathcal{L}G_\lambda h_n = \lambda G_\lambda h_n - h_n = \lambda f_n - h_n$, we have

$$f_n(0) - \mathbb{E}\, e^{-\lambda A_t} f_n(Z_t) = \mathbb{E} \int_0^t e^{-\lambda A_s} |H_s|^\alpha h_n(Z_{s-}) ds.$$

Note $|H_s| \le \rho(|Z_{s-}|)$, so using our bound for h_n, the right hand side is less than $2t/n$ in absolute value, which tends to 0 as $n \to \infty$. The measures $h_n(y)dy$ all have mass 1 and they tend weakly to point mass at 0. Since g_λ is continuous in x, then $f_n(x) \to g_\lambda(x)$ as $n \to \infty$. We conclude

$$g_\lambda(0) - \mathbb{E}\, e^{-\lambda A_t} g_\lambda(Z_t) = 0.$$

We noted above that $g_\lambda(x) < g_\lambda(0)$ if $x \ne 0$, while clearly $A_t < \infty$ since F is bounded. We deduce $\mathbb{P}(Z_t = 0) = 1$. This holds for each t, and we conclude that Z is identically 0. $\square$

Remark 4.6 The above proof breaks down for $\alpha = 1$ since g_λ is no longer a bounded function.

Chapter 5

The space $D[0,1]$

5.1 Convergence of probability measures

We suppose we have a sequence of probabilities on a metric space $\mathcal{S}$ and we want to define what it means for the sequence to converge weakly. Alternately, we may have a sequence of random variables and want to say what it means for the random variables to converge weakly.

For now our state space is assumed to be an arbitrary metric space, although we will soon add additional assumptions on $\mathcal{S}$. We use the Borel σ-field on $\mathcal{S}$, which is the σ-field generated by the open sets in $\mathcal{S}$. We write $A^0, \overline{A}$, and ∂A for the interior, closure, and boundary of A, resp.

5.2 The portmanteau theorem

Clearly the definition of weak convergence of real-valued random variables in terms of distribution functions has no obvious analog. The appropriate generalization is the following.

Definition 5.1 *A sequence of probabilities* $\{\mathbb{P}_n\}$ *on a metric space* $\mathcal{S}$ *furnished with the Borel* σ*-field is said to* converge weakly *to* $\mathbb{P}$ *if* $\int f\,d\mathbb{P}_n \to \int f\,d\mathbb{P}$ *for every bounded and continuous function* f *on* $\mathcal{S}$*. A sequence of random variables* $\{X_n\}$ *taking values in* $\mathcal{S}$ converges weakly *to a random* *variable X taking values in $\mathcal{S}$ if $\mathbb{E}\, f(X_n) \to \mathbb{E}\, f(X)$ whenever f is a bounded and continuous function.*

Saying X_n converges weakly to X is the same as saying that the laws of X_n converge weakly to the law of X. To see this, if $\mathbb{P}_n$ is the law of X_n, that is, $\mathbb{P}_n(A) = \mathbb{P}(X_n \in A)$ for each Borel subset A of $\mathcal{S}$, then $\mathbb{E}\, f(X_n) = \int f\, d\mathbb{P}_n$ and $\mathbb{E}\, f(X) = \int f\, d\mathbb{P}$. (This holds when f is an indicator by the definition of the law of X_n and X, then for simple functions by linearity, then for non-negative measurable functions by monotone convergence, and then for arbitrary bounded and Borel measurable f by linearity.)

The following theorem, known as the portmanteau theorem, gives some other characterizations of weak convergence. For this chapter we let

$$F_\delta = \{x : d(x, F) < \delta\} \tag{5.1}$$

for closed sets F, the set of points within δ of F, where $d(x, F) = \inf\{d(x, y) : y \in F\}$.

Theorem 5.2 *Suppose $\{\mathbb{P}_n, n = 1, 2, \ldots\}$ and $\mathbb{P}$ are probabilities on a metric space. The following are equivalent.*

(1) $\mathbb{P}_n$ converges weakly to $\mathbb{P}$.

(2) $\limsup_n \mathbb{P}_n(F) \le \mathbb{P}(F)$ for all closed sets F.

(3) $\liminf_n \mathbb{P}_n(G) \ge \mathbb{P}(G)$ for all open sets G.

(4) $\lim_n \mathbb{P}_n(A) = \mathbb{P}(A)$ for all Borel sets A such that $\mathbb{P}(\partial A) = 0$.

Proof. The equivalence of (2) and (3) is easy because if F is closed, then $G = F^c$ is open and $\mathbb{P}_n(G) = 1 - \mathbb{P}_n(F)$.

To see that (2) and (3) imply (4), suppose $\mathbb{P}(\partial A) = 0$. Then

$$\begin{aligned}\limsup_n \mathbb{P}_n(A) &\le \limsup_n \mathbb{P}_n(\overline{A}) \le \mathbb{P}(\overline{A})\\ &= \mathbb{P}(A^0) \le \liminf \mathbb{P}_n(A^0) \le \liminf \mathbb{P}_n(A).\end{aligned}$$

Next, let us show (4) implies (2). Let F be closed. If $y \in \partial F_\delta$, then $d(y, F) = \delta$. The sets ∂F_δ are disjoint for different δ. At most countably many of them can have positive $\mathbb{P}$-measure, hence there exists a sequence $\delta_k \downarrow 0$ such that $\mathbb{P}(\partial F_{\delta_k}) = 0$ for each k. Then

$$\limsup_n \mathbb{P}_n(F) \le \limsup_n \mathbb{P}_n(\overline{F_{\delta_k}}) = \mathbb{P}(\overline{F_{\delta_k}}) = \mathbb{P}(F_{\delta_k})$$

for each k. Since $\mathbb{P}(F_{\delta_k}) \downarrow \mathbb{P}(F)$ as $\delta_k \to 0$, this gives (2).

We show now that (1) implies (2). Suppose F is closed. Let $\varepsilon > 0$. Take $\delta > 0$ small enough so that $\mathbb{P}(\overline{F_\delta}) - \mathbb{P}(F) < \varepsilon$. Then take f continuous, to be equal to 1 on F, to have support in $\overline{F_\delta}$, and to be bounded between 0 and 1. For example, $f(x) = 1 - (1 \wedge \delta^{-1} d(x, F))$ would do. Then

$$\begin{aligned}\limsup_n \mathbb{P}_n(F) &\le \limsup_n \int f d\mathbb{P}_n = \int f d\mathbb{P} \\ &\le \mathbb{P}(\overline{F_\delta}) \le \mathbb{P}(F) + \varepsilon.\end{aligned}$$

Since this is true for all ε, (2) follows.

Finally, let us show (2) implies (1). Let f be bounded and continuous. If we show

$$\limsup_n \int f d\mathbb{P}_n \le \int f d\mathbb{P}, \tag{5.2}$$

for every such f, then applying this inequality to both f and $-f$ will give (1). By adding a sufficiently large positive constant to f and then multiplying by a suitable constant, without loss of generality we may assume f is bounded and takes values in $(0, 1)$. Let $F_i = \{x : f(x) \ge i/k\}$, which is closed.

$$\begin{aligned}\int f d\mathbb{P}_n &\le \sum_{i=1}^k \frac{i}{k} \mathbb{P}_n\Big(\frac{i-1}{k} \le f(x) < \frac{i}{k}\Big) \\ &= \sum_{i=1}^k \frac{i}{k} [\mathbb{P}_n(F_{i-1}) - \mathbb{P}_n(F_i)] \\ &= \sum_{i=0}^{k-1} \frac{i+1}{k} \mathbb{P}_n(F_i) - \sum_{i=1}^k \frac{i}{n} \mathbb{P}_n(F_i) \\ &\le \frac{1}{k} + \frac{1}{k} \sum_{i=1}^k \mathbb{P}_n(F_i).\end{aligned}$$

Similarly,

$$\int f d\mathbb{P} \geq \frac{1}{k} \sum_{i=1}^{k} \mathbb{P}(F_i).$$

Then

$$\begin{aligned}\limsup_n \int f d\mathbb{P}_n &\leq \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \limsup_n \mathbb{P}_n(F_i) \\ &\leq \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \mathbb{P}(F_i) \leq \frac{1}{k} + \int f d\mathbb{P}.\end{aligned}$$

Since k is arbitrary, this gives (5.2). □

If $x_n \to x$, $\mathbb{P}_n = \delta_{x_n}$, and $\mathbb{P} = \delta_x$, it is easy to see $\mathbb{P}_n$ converges weakly to $\mathbb{P}$. Letting $A = \{x\}$ shows that one cannot, in general, have $\lim_n \mathbb{P}_n(F) = \mathbb{P}(F)$ for all closed sets F.

5.3 The Prohorov theorem

It turns out there is a simple condition that ensures that a sequence of probability measures has a weakly convergent subsequence.

Definition 5.3 *A sequence of probabilities $\mathbb{P}_n$ on a metric space $\mathcal{S}$ is* tight *if for every ε there exists a compact set K (depending on ε) such that $\sup_n \mathbb{P}_n(K^c) \leq \varepsilon$.*

The important result here is Prohorov's theorem.

Theorem 5.4 *If a sequence of probability measures on a metric space $\mathcal{S}$ is tight, there is a subsequence that converges weakly to a probability measure on $\mathcal{S}$.*

Proof. Suppose first that the metric space $\mathcal{S}$ is compact. Then $C(\mathcal{S})$, the collection of continuous functions on $\mathcal{S}$, is a separable metric space when furnished with the supremum norm. Let $\{f_i\}$ be a countable collection of non-negative elements of $C(\mathcal{S})$ whose linear span is dense in $C(\mathcal{S})$. For each i, $\int f_i\, d\mathbb{P}_n$ is a bounded sequence, so we have a convergent subsequence. By a diagonalization procedure, we can find a subsequence n' such that $\int f_i\, d\mathbb{P}_{n'}$ converges for all i. By the term "diagonalization procedure," we are referring to the well known method of proof of the Ascoli-Arzelà theorem; see any book on real analysis for a detailed explanation. Call the limit Lf_i. Clearly $0 \leq Lf_i \leq \|f_i\|_\infty$, L is linear, and so we can extend L to a bounded linear functional on $\mathcal{S}$. By the Riesz representation theorem, there exists a measure $\mathbb{P}$ such that $Lf = \int f d\mathbb{P}$. Since $\int f_i\, d\mathbb{P}_{n'} \to \int f_i\, d\mathbb{P}$ for all f_i, it is not hard to see, since each $\mathbb{P}_{n'}$ has total mass 1, that $\int f\, d\mathbb{P}_{n'} \to \int f\, d\mathbb{P}$ for all $f \in C(\mathcal{S})$. Therefore $\mathbb{P}_{n'}$ converges weakly to $\mathbb{P}$. Since $Lf \geq 0$ if $f \geq 0$, then $\mathbb{P}$ is a positive measure. The function that is identically equal to 1 is bounded and continuous, so $1 = \mathbb{P}_{n'}(\mathcal{S}) = \int 1\, d\mathbb{P}_{n'} \to \int 1\, d\mathbb{P}$, or $\mathbb{P}(\mathcal{S}) = 1$.

Next suppose that $\mathcal{S}$ is a Borel subset of a compact metric space $\mathcal{S}'$. Extend each $\mathbb{P}_n$, initially defined on $\mathcal{S}$, to $\mathcal{S}'$ by setting $\mathbb{P}_n(\mathcal{S}' \setminus \mathcal{S}) = 0$. By the first paragraph of the proof, there is a subsequence $\mathbb{P}_{n'}$ that converges weakly to a probability $\mathbb{P}$ on $\mathcal{S}'$ (the definition of weak convergence here is relative to the topology on $\mathcal{S}'$). Since the $\mathbb{P}_n$ are tight, there exist compact subsets K_m of $\mathcal{S}$ such that $\mathbb{P}_n(K_m) \geq 1 - 1/m$ for all n. The K_m will also be compact relative to the topology on $\mathcal{S}'$, so by Theorem 5.2,

$$\mathbb{P}(K_m) \geq \limsup_{n'} \mathbb{P}_{n'}(K_m) \geq 1 - 1/m.$$

Since $\cup_m K_m \subset \mathcal{S}$, we conclude $\mathbb{P}(\mathcal{S}) = 1$.

If G is open in $\mathcal{S}$, then $G = H \cap \mathcal{S}$ for some H open in $\mathcal{S}'$. Then

$$\liminf_{n'} \mathbb{P}_{n'}(G) = \liminf_{n'} \mathbb{P}_{n'}(H) \geq \mathbb{P}(H) = \mathbb{P}(H \cap \mathcal{S}) = \mathbb{P}(G).$$

Thus by Theorem 5.2, $\mathbb{P}_{n'}$ converges weakly to $\mathbb{P}$ relative to the topology on $\mathcal{S}$.

Now let $\mathcal{S}$ be an arbitrary metric space. Since all the $\mathbb{P}_n$'s are supported on $\cup_m K_m$, we can replace $\mathcal{S}$ by $\cup_m K_m$, or we may as well assume that $\mathcal{S}$ is σ-compact, and hence separable. It remains to embed the separable metric space $\mathcal{S}$ into a compact metric space $\mathcal{S}'$. If d is the metric on $\mathcal{S}$, $d \wedge 1$ will also be an equivalent metric, that is, one that generates the same collection of open sets, so we may assume d is bounded by 1. Now $\mathcal{S}$ can be embedded in $\mathcal{S}' = [0,1]^{\mathbb{N}}$ as follows. We define a metric on $\mathcal{S}'$ by

$$d'(a,b) = \sum_{i=1}^{\infty} 2^{-i}(|a^i - b^i| \wedge 1), \qquad a = (a^1, a^2, \ldots), b = (b^1, b^2, \ldots). \quad (5.3)$$

Being the product of compact spaces, $\mathcal{S}'$ is itself compact. If $\{z_j\}$ is a countable dense subset of $\mathcal{S}$, let $I : \mathcal{S} \to [0,1]^{\mathbb{N}}$ be defined by

$$I(x) = (d(x, z_1), d(x, z_2), \ldots).$$

We leave it to the reader to check that I is a one-to-one continuous open map of $\mathcal{S}$ to a subset of $\mathcal{S}'$. Since $\mathcal{S}$ is σ-compact, and the continuous image of compact sets is compact, then $I(\mathcal{S})$ is a Borel set. □

5.4 Metrics for $D[0,1]$

We define the space $D[0,1]$ to be the collection of real-valued functions on $[0,1]$ which are right continuous with left limits. We will introduce a topology on $D = D[0,1]$, the Skorokhod topology, which makes D into a complete separable metric space.

We write $f(t-)$ for $\lim_{s<t, s\to t} f(s)$. We will need the following observation. If f is in D and $\varepsilon > 0$, let $t_0 = 0$, and for $i > 0$ let $t_{i+1} = \inf\{t > t_i : |f(t) - f(t_i)| > \varepsilon\} \wedge 1$. Because f is right continuous with left limits, then from some i on, t_i must be equal to 1.

Our first try at a metric, ρ, makes D into a separable metric space, but one that is not complete. Let's start with ρ anyway, since we need it on the way to the metric d we end up with.

Let Λ be the set of functions λ from $[0,1]$ to $[0,1]$ that are continuous, strictly increasing, and such that $\lambda(0) = 0$, $\lambda(1) = 1$. Define

$$\begin{aligned}\rho(f,g) = \inf\{\varepsilon > 0 : \exists \lambda \in \Lambda \text{ such that } &\sup_{t\in[0,1]} |\lambda(t) - t| < \varepsilon,\\ &\sup_{t\in[0,1]} |f(t) - g(\lambda(t))| < \varepsilon\}.\end{aligned}$$

Since the function $\lambda(t) = t$ is in Λ, then $\rho(f,g)$ is finite if $f, g \in D$. Clearly $\rho(f,g) \geq 0$. If $\rho(f,g) = 0$, then either $f(t) = g(t)$ or else $f(t) = g(t-)$ for each t; since elements of D are right continuous with left limits, it follows that $f = g$. If $\lambda \in \Lambda$, then so is λ^{-1} and we have, setting $s = \lambda^{-1}(t)$ and noting both s and t range over $[0,1]$,

$$\sup_{t\in[0,1]} |\lambda^{-1}(t) - t| = \sup_{s\in[0,1]} |s - \lambda(s)|$$

and

$$\sup_{t\in[0,1]} |f(\lambda^{-1}(t)) - g(t)| = \sup_{s\in[0,1]} |f(s) - g(\lambda(s))|,$$

and we conclude $\rho(f,g) = \rho(g,f)$. The triangle inequality follows from

$$\sup_{t\in[0,1]} |\lambda_2 \circ \lambda_1(t) - t| \le \sup_{t\in[0,1]} |\lambda_1(t) - t| + \sup_{s\in[0,1]} |\lambda_2(s) - s|$$

and

$$\sup_{t\in[0,1]} |f(t) - h(\lambda_2 \circ \lambda_1(t))| \le \sup_{t\in[0,1]} |f(t) - g(\lambda_1(t))| + \sup_{s\in[0,1]} |g(s) - h(\lambda_2(s))|.$$

Look at the set of f in D for which there exists an integer k such that f is constant and equal to a rational on each interval $[(i-1)/k, i/k)$. It is not hard to check that the collection of such f's is dense in D with respect to ρ, which shows (D,ρ) is separable.

The space D with the metric ρ is not, however, complete; one can show that $f_n = 1_{[1/2,1/2+1/n]}$ is a Cauchy sequence which does not converge. We therefore introduce a slightly different metric d. Define

$$\|\lambda\| = \sup_{s\ne t, s,t\in[0,1]} \Big| \log \frac{\lambda(t) - \lambda(s)}{t-s} \Big|$$

and let

$$d(f,g) = \inf\{\varepsilon > 0 : \exists \lambda \in \Lambda \text{ such that } \|\lambda\| \le \varepsilon, \sup_{t\in[0,1]} |f(t) - g(\lambda(t))| \le \varepsilon.\}$$

Note $\|\lambda^{-1}\| = \|\lambda\|$ and $\|\lambda_2 \circ \lambda_1\| \le \|\lambda_1\| + \|\lambda_2\|$. The symmetry of d and the triangle inequality follow easily from this, and we conclude d is a metric.

Lemma 5.5 *There exists ε_0 such that*

$$\rho(f,g) \le 2d(f,g)$$

if $d(f,g) < \varepsilon_0$.

(It turns out $\varepsilon_0 = 1/4$ will do.)

Proof. Since $\log(1+2x)/(2x) \to 1$ as $x \to 0$, we have

$$\log(1-2\varepsilon) < -\varepsilon < \varepsilon < \log(1+2\varepsilon)$$

if ε is small enough. Suppose $d(f,g) < \varepsilon$ and λ is the element of Λ such that $d(f,g) < \|\lambda\| < \varepsilon$ and $\sup_{t\in[0,1]} |f(t) - g(\lambda(t))| < \varepsilon$. Since $\lambda(0) = 0$, we have

$$\log(1-2\varepsilon) < -\varepsilon < \log \frac{\lambda(t)}{t} < \varepsilon < \log(1+2\varepsilon), \tag{5.4}$$

or

$$1 - 2\varepsilon < \frac{\lambda(t)}{t} < 1 + 2\varepsilon, \tag{5.5}$$

which implies $|\lambda(t) - t| < 2\varepsilon$, and hence $\rho(f,g) \le 2d(f,g)$. □

We define the analog ξ_f of the modulus of continuity for a function in D as follows. Define $\theta_f[a,b) = \sup_{s,t\in[a,b)} |f(t) - f(s)|$ and

$$\xi_f(\delta) = \inf\{\max_{1\le i\le n} \theta_f[t_{i-1}, t_i) : \exists n \ge 1, 0 = t_0 < t_1 < \cdots < t_n = 1$$
$$\text{such that } t_i - t_{i-1} > \delta \text{ for all } i \le n\}.$$

Observe that if $f \in D$, then $\xi_f(\delta) \downarrow 0$ as $\delta \downarrow 0$.

Lemma 5.6 *Suppose $\delta < 1/4$. Let $f \in D$. If $\rho(f,g) < \delta^2$, then $d(f,g) \le 4\delta + \xi_f(\delta)$.*

Proof. Choose t_i's such that $t_i - t_{i-1} > \delta$ and $\theta_f[t_{i-1}, t_i) < \xi_f(\delta) + \delta$ for each i. Pick $\mu \in \Lambda$ such that $\sup_t |f(t) - g(\mu(t))| < \delta^2$ and $\sup_t |\mu(t) - t| < \delta^2$. Then $\sup_t |f(\mu^{-1}(t)) - g(t)| < \delta^2$. Set $\lambda(t_i) = \mu(t_i)$ and let λ be linear in between. Since $\mu^{-1}(\lambda(t_i)) = t_i$ for all i, then t and $\mu^{-1}\circ\lambda(t)$ always lie in the same subinterval $[t_{i-1}, t_i)$. Consequently

$$\begin{aligned}|f(t)-g(\lambda(t))| &\le |f(t)-f(\mu^{-1}(\lambda(t)))| + |f(\mu^{-1}(\lambda(t)))-g(\lambda(t))|\\ &\le \xi_f(\delta)+\delta+\delta^2 < \xi_f(\delta)+4\delta.\end{aligned}$$

We have

$$\begin{aligned}|\lambda(t_i)-\lambda(t_{i-1})-(t_i-t_{i-1})| &= |\mu(t_i)-\mu(t_{i-1})-(t_i-t_{i-1})|\\ &\le 2\delta^2 < 2\delta(t_i-t_{i-1}).\end{aligned}$$

Since λ is defined by linear interpolation,

$$|\lambda(t)-\lambda(s))-(t-s)| \le 2\delta|t-s|, \qquad s,t\in[0,1],$$

which leads to

$$\Big|\frac{\lambda(t)-\lambda(s)}{t-s}-1\Big| \le 2\delta,$$

or

$$\log(1-2\delta) \le \log\Big(\frac{\lambda(t)-\lambda(s)}{t-s}\Big) \le \log(1+2\delta).$$

Since $\delta < \frac{1}{4}$, we have $\|\lambda\| \le 4\delta$. □

Proposition 5.7 *The metrics d and ρ are equivalent, i.e., they generate the same topology.*

In particular, (D,d) is separable.

Proof. Let $B_\rho(f,r)$ denote the ball with center f and radius r with respect to the metric ρ and define $B_d(f,r)$ analogously. Let $\varepsilon > 0$ and let $f\in D$. If $d(f,g) < \varepsilon/2$ and ε is small enough, then $\rho(f,g) \le 2d(f,g) < \varepsilon$, and so $B_d(f,\varepsilon/2)\subset B_\rho(f,\varepsilon)$.

To go the other direction, what we must show is that given f and ε, there exists δ such that $B_\rho(f,\delta)\subset B_d(f,\varepsilon)$. δ may depend on f; in fact, it has to in general, for otherwise a Cauchy sequence with respect to d would be a Cauchy sequence with respect to ρ, and vice versa. Choose δ small enough that $4\delta^{1/2}+\xi_f(\delta^{1/2}) < \varepsilon$. By Lemma 5.6, if $\rho(f,g) < \delta$, then $d(f,g) < \varepsilon$, which is what we want.

Finally, suppose G is open with respect to the topology generated by ρ. For each $f \in G$, let r_f be chosen so that $B_\rho(f, r_f) \subset G$. Hence $G = \cup_{f\in G} B_\rho(f, r_f)$. Let s_f be chosen so that $B_d(f, s_f) \subset B_\rho(f, r_f)$. Then $\cup_{f\in G} B_d(f, s_f) \subset G$, and in fact the sets are equal because if $f \in G$, then $f \in B_d(f, s_f)$. Since G can be written as the union of balls which are open with respect to d, then G is open with respect to d. The same argument with d and ρ interchanged shows that a set that is open with respect to d is open with respect to ρ. □

5.5 Compactness and completeness

We now show completeness for (D, d).

Theorem 5.8 *The space D with the metric d is complete.*

Proof. Let f_n be a Cauchy sequence with respect to the metric d. If we can find a subsequence n_j such that f_{n_j} converges, say, to f, then it is standard that the whole sequence converges to f. Choose n_j such that $d(f_{n_j}, f_{n_{j+1}}) < 2^{-j}$. For each j there exists λ_j such that

$$\sup_t |f_{n_j}(t) - f_{n_{j+1}}(\lambda_j(t))| \le 2^{-j}, \qquad \|\lambda_j\| \le 2^{-j}.$$

As in (5.4) and (5.5),

$$|\lambda_j(t) - t| \le 2^{-j+1}.$$

Then

$$\begin{aligned}\sup_t |\lambda_{n+m+1} \circ \lambda_{m+n} \circ \cdots \circ \lambda_n(t) - \lambda_{n+m} \circ \cdots \circ \lambda_n(t)| &\\ = \sup_s |\lambda_{n+m+1}(s) - s| &\\ \le 2^{-(n+m)} &\end{aligned}$$

for each n. Hence for each n, the sequence $\lambda_{m+n} \circ \cdots \circ \lambda_n$ (indexed by m) is a Cauchy sequence of functions on $[0,1]$ with respect to the supremum norm on $[0,1]$. Let ν_n be the limit. Clearly $\nu_n(0)=0$, $\nu_n(1)=1$, ν_n is continuous, and nondecreasing. We also have

$$\begin{aligned}
\Big|\log&\frac{\lambda_{n+m}\circ\cdots\circ\lambda_n(t)-\lambda_{n+m}\circ\cdots\circ\lambda_n(s)}{t-s}\Big| \\
&\le \|\lambda_{n+m}\circ\cdots\circ\lambda_n\| \\
&\le \|\lambda_{n+m}\|+\cdots+\|\lambda_n\| \\
&\le \frac{1}{2^{n-1}}.
\end{aligned}$$

If we then let $m\to\infty$, we obtain

$$\Big|\log\frac{\nu_n(t)-\nu_n(s)}{t-s}\Big|\le\frac{1}{2^{n-1}},$$

which implies $\nu_n\in\Lambda$ with $\|\nu_n\|\le 2^{1-n}$.

We see that $\nu_n=\nu_{n+1}\circ\lambda_n$. Consequently

$$\sup_t|f_{n_j}(\nu_j^{-1}(t))-f_{n_{j+1}}(\nu_{j+1}^{-1}(t))|=\sup_s|f_{n_j}(s)-f_{n_{j+1}}(\lambda_j(s))|\le 2^{-j}.$$

Therefore $f_{n_j}\circ\nu_j^{-1}$ is a Cauchy sequence on $[0,1]$ with respect to the supremum norm. Let f be the limit. Since

$$\sup_t|f_{n_j}(\nu_j^{-1}(t))-f(t)|\to 0$$

and $\|\nu_j\|\to 0$ as $j\to\infty$, then $d(f_{n_j},f)\to 0$. □

We turn to compactness.

Theorem 5.9 *A set A has compact closure in $D[0,1]$ if*

$$\sup_{f\in A}\sup_t|f(t)|<\infty$$

and

$$\lim_{\delta\to 0}\sup_{f\in A}\xi_f(\delta)=0.$$

The converse of this theorem is also true, but we won't need this.

Proof. A complete and totally bounded set in a metric space is compact, and $D[0,1]$ is a complete metric space. Hence it suffices to show that A is totally bounded: for each $\varepsilon > 0$ there exist finitely many balls of radius ε that cover A.

Let $\eta > 0$ and choose k large such that $1/k < \eta$ and $\xi_f(1/k) < \eta$ for each $f \in A$. Let $M = \sup_{f\in A} \sup_t |f(t)|$ and let $H = \{-M + j/k : j \le 2kM\}$, so that H is an η-net for $[-M, M]$. Let B be the set of functions $f \in D[0,1]$ that are constant on each interval $[(i-1)/k, i/k)$ and that take values only in the set H. In particular, $f(1) \in H$.

We first prove that B is a 2η-net for A with respect to ρ. If $f \in A$, there exist $t_0, \ldots, t_n$ such that $t_0 = 0$, $t_n = 1$, $t_i - t_{i-1} > 1/k$ for each i, and $\theta_f[t_{i-1}, t_i) < \eta$ for each i. Note we must have $n \le k$. For each i choose integers j_i such that $j_i/k \le t_i < (j_i + 1)/k$. The j_i are distinct since the t_i are at least $1/k$ apart. Define λ so that $\lambda(j_i/k) = t_i$ and λ is linear on each interval $[j_i/k, j_{i+1}/k]$. Choose $g \in B$ such that $|g(m/k) - f(\lambda(m/k))| < \eta$ for each $m \le k$. Observe that each $[m/k, (m+1)/k)$ lies inside some interval of the form $[j_i/k, j_{i+1}/k)$. Since λ is increasing, $[\lambda(m/k), \lambda((m+1)/k))$ is contained in $[\lambda(j_i/k), \lambda(j_{i+1}/k)) = [t_i, t_{i+1})$. The function f does not vary more than η over each interval $[t_i, t_{i+1})$, so $f(\lambda(t))$ does not vary more than η over each interval $[m/k, (m+1)/k)$. g is constant on each such interval, and hence

$$\sup_t |g(t) - f(\lambda(t))| < 2\eta.$$

We have

$$|\lambda(j_i/k) - j_i/k| = |t_i - j_i/k| < 1/k < \eta$$

for each i. By the piecewise linearity of λ, $\sup_t |\lambda(t) - t| < \eta$. Thus $\rho(f, g) < 2\eta$. We have proved that given $f \in A$, there exists $g \in B$ such that $\rho(f, g) < 2\eta$, or B is a 2η-net for A with respect to ρ.

Now let $\varepsilon > 0$ and choose $\delta > 0$ small so that $4\delta + \xi_f(\delta) < \varepsilon$ for each $f \in A$. Set $\eta = \delta^2/4$. Choose B as above to be a 2η-net for A with respect to ρ. By Lemma 5.6, if $\rho(f, g) < 2\eta < \delta^2$, then $d(f, g) \le 4\delta + \xi_f(\delta) < \varepsilon$. Therefore B is an ε-net for A with respect to d. □

The following corollary is proved exactly similarly to the continuous case.

Corollary 5.10 *Suppose X_n are processes whose paths are right continuous with left limits. Suppose for each ε and η there exists n_0, R, and δ such that*

$$\mathbb{P}(\xi_{X_n}(\delta) \geq \varepsilon) \leq \eta \tag{5.6}$$

and

$$\mathbb{P}(\sup_{t\in[0,1]} |X_n(t)| \geq R) \leq \eta. \tag{5.7}$$

Then the X_n are tight with respect to the topology of $D[0,1]$.

Proof. Since each X_i is in $D[0,1]$, then for each i, $\mathbb{P}(\xi_{X_i}(\delta) \geq \varepsilon) \to 0$ as $\delta \to 0$ by dominated convergence. Hence, given ε and η we can, by taking δ smaller if necessary, assume that (5.6) holds for all n.

Choose $\varepsilon_m = \eta_m = 2^{-m}$ and consider the δ_m and A_m so that

$$\sup_n \mathbb{P}(\xi_{X_n}(\delta_m) \geq 2^{-m}) \leq 2^{-m}$$

and

$$\sup_n \mathbb{P}(\sup_t |X_n(t)| \geq A_m) \leq 2^{-m}.$$

Let

$$K_{m_0} = \{f \in D[0,1] : \xi_f(\delta_m) \leq 2^{-m} \text{ for all } m \geq m_0, \\ \sup_t |f(t)| \leq A_{m_0}\}.$$

Each K_{m_0} is a compact subset of $D[0,1]$. We have

$$\begin{aligned} \mathbb{P}(X_n \notin K_{m_0}) &\leq \mathbb{P}(\sup_t |X_n(t)| \geq A_{m_0}) + \sum_{m=m_0}^{\infty} \mathbb{P}(\xi_{X_n}(\delta_m) \geq \varepsilon_m) \\ &\leq 2^{-m_0} + \sum_{m=m_0}^{\infty} 2^{-m} = 3 \cdot 2^{-m_0}. \end{aligned}$$

This proves tightness. □

We show that if $f_n \to f$ with respect to d and $f \in C[0,1]$, the convergence is in fact uniform.

Proposition 5.11 *Suppose $f_n \to f$ in the topology of $D[0,1]$ with respect to d and $f \in C[0,1]$. Then $\sup_{t\in[0,1]} |f_n(t) - f(t)| \to 0$.*

Proof. Let $\varepsilon > 0$. Since f is uniformly continuous on $[0,1]$, there exists δ such that $|f(t) - f(s)| < \varepsilon/2$ if $|t - s| < \delta$. For n sufficiently large there exists $\lambda_n \in \Lambda$ such that $\sup_t |f_n(t) - f(\lambda_n(t))| < \varepsilon/2$ and $\sup_t |\lambda_n(t) - t| < \delta$. Therefore $|f(\lambda_n(t)) - f(t)| < \varepsilon/2$, and so $|f_n(t) - f(t)| < \varepsilon$. □

5.6 The Aldous criterion

A very useful criterion for tightness is the following one due to Aldous.

Theorem 5.12 *Let $\{X_n\}$ be a sequence in $D[0,1]$. Suppose*

$$\lim_{R\to\infty} \sup_n \mathbb{P}(|X_n(t)| \geq R) = 0 \tag{5.8}$$

for each $t \in [0,1]$ and that whenever τ_n are stopping times for X_n and $\delta_n \to 0$ are reals,

$$|X_n(\tau_n + \delta_n) - X_n(\tau_n)| \tag{5.9}$$

converges to 0 in probability as $n \to \infty$.

Proof. We will set $X_n(t) = X_n(1)$ for $t \in [1,2]$ to simplify notation. The proof of this theorem comprises four steps.

Step 1. We claim that (5.9) implies the following: given ε there exist n_0 and δ such that

$$\mathbb{P}(|X_n(\tau_n + s) - X_n(\tau_n)| \geq \varepsilon) \leq \varepsilon \tag{5.10}$$

for each $n \geq n_0$, $s \leq 2\delta$, and τ_n a stopping time for X_n. For if not, we choose an increasing subsequence n_k, stopping times τ_{n_k}, and $s_{n_k} \leq 1/k$ for which (5.10) does not hold. Taking $\delta_{n_k} = s_{n_k}$ gives a contradiction to (5.9).

Step 2. Let $\varepsilon > 0$, fix $n \geq n_0$, and let $T \leq U \leq 1$ be two stopping times for X_n. We will prove

$$\mathbb{P}(U \leq T + \delta, |X_n(U) - X_n(T)| \geq 2\varepsilon) \leq 16\varepsilon. \tag{5.11}$$

To prove this, we start by letting λ be Lebesgue measure. If

$$A_T = \{(\omega, s) \in \Omega \times [0, 2\delta] : |X_n(T+s) - X_n(T)| \geq \varepsilon\},$$

then for each $s \leq 2\delta$ we have $\mathbb{P}(\omega : (\omega, s) \in A_T) \leq \varepsilon$ by (5.10) with τ_n replaced by T. Writing $\mathbb{P} \times \lambda$ for the product measure, we then have

$$\mathbb{P} \times \lambda(A_T) \leq 2\delta\varepsilon. \tag{5.12}$$

Set $B_T(\omega) = \{s : (\omega, s) \in A_T\}$ and $C_T = \{\omega : \lambda(B_T(\omega)) \geq \frac{1}{4}\delta\}$. From (5.12) and the Fubini theorem,

$$\int \lambda(B_T(\omega))\, \mathbb{P}(d\omega) \leq 2\delta\varepsilon,$$

so

$$\mathbb{P}(C_T) \leq 8\varepsilon.$$

We similarly define B_U and C_U, and obtain $\mathbb{P}(C_T \cup C_U) \leq 16\varepsilon$.

If $\omega \notin C_T \cup C_U$, then $\lambda(B_T(\omega)) \leq \frac{1}{4}\delta$ and $\lambda(B_U(\omega)) \leq \frac{1}{4}\delta$. Suppose $U \leq T + \delta$. Then

$$\lambda\{t \in [T, T+2\delta] : |X_n(t) - X_n(T)| \geq \varepsilon\} \leq \tfrac{1}{4}\delta,$$

and

$$\lambda\{t \in [U, U+\delta] : |X_n(t) - X_n(U)| \geq \varepsilon\} \leq \tfrac{1}{4}\delta.$$

Hence there exists $t \in [T, T+2\delta] \cap [U, U+\delta]$ such that $|X_n(t) - X_n(T)| < \varepsilon$ and $|X_n(t) - X_n(U)| < \varepsilon$; this implies $|X_n(U) - X_n(T)| < 2\varepsilon$, which proves (5.11).

Step 3. We obtain a bound on ξ_{X_n}. Let $T_{n0} = 0$ and

$$T_{n,i+1} = \inf\{t > T_{ni} : |X_n(t) - X_n(T_{ni})| \geq 2\varepsilon\} \wedge 2.$$

Note we have $|X_n(T_{n,i+1}) - X_n(T_{ni})| \geq 2\varepsilon$ if $T_{ni} < 2$. We choose n_0, δ as in Step 1. By Step 2 with $T = T_{ni}$ and $U = T_{n,i+1}$,

$$\mathbb{P}(T_{n,i+1} - T_{ni} < \delta, T_{ni} < 2) \leq 16\varepsilon. \tag{5.13}$$

Let $K = [2/\delta] + 1$ and apply (5.10) with ε replaced by ε/K to see that there exist $n_1 \geq n_0$ and $\zeta \leq \delta \wedge \varepsilon$ such that if $n \geq n_1$, $s \leq 2\zeta$, and τ_n is a stopping time, then

$$\mathbb{P}(|X_n(\tau_n + s) - X_n(\tau_n)| > \varepsilon/K) \leq \varepsilon/K. \tag{5.14}$$

By (5.11) with $T = T_{ni}$ and $U = T_{n,i+1}$ and δ replaced by ζ,

$$\mathbb{P}(T_{n,i+1} \le T_{ni} + \zeta) \le 16\varepsilon/K \tag{5.15}$$

for each i and hence

$$\mathbb{P}(\exists i \le K : T_{n,i+1} \le T_{ni} + \zeta) \le 16\varepsilon. \tag{5.16}$$

We have

$$\begin{aligned}\mathbb{E}\,[T_{ni} - T_{n,i-1}; T_{nK} < 1] &\ge \delta\mathbb{P}(T_{ni} - T_{n,i-1} \ge \delta, T_{nK} < 1)\\ &\ge \delta[\mathbb{P}(T_{nK} < 1) - \mathbb{P}(T_{ni} - T_{n,i-1} < \delta, T_{nK} < 1)]\\ &\ge \delta[\mathbb{P}(T_{nK} < 1) - 16\varepsilon],\end{aligned}$$

where we used (5.13) in the last step. Summing over i from 1 to K,

$$\begin{aligned}\mathbb{P}(T_{nK} < 1) \ge \mathbb{E}\,[T_{nK}; T_{nK} < 1] &= \sum_{i=1}^{K} \mathbb{E}\,[T_{ni} - T_{n,i-1}; T_{nK} < 1]\\ &\ge K\delta[\mathbb{P}(T_{nK} < 1) - 16\varepsilon] \ge 2[\mathbb{P}(T_{nK} < 1) - 16\varepsilon],\end{aligned}$$

or $\mathbb{P}(T_{nK} < 1) \le 32\varepsilon$. Hence except for an event of probability at most 32ε, we have $\xi_{X_n}(\zeta) \le 4\varepsilon$.

Step 4. The last step is to obtain a bound on $\sup_t |X_n(t)|$. Let $\varepsilon > 0$ and choose δ and n_0 as in Step 1. Define

$$D_{Rn} = \{(\omega, s) \in \Omega \times [0,1] : |X_n(s)(\omega)| > R\}$$

for $R > 0$. The measurability of D_{Rn} with respect to the product σ-field $\mathcal{F} \times \mathcal{B}[0,1]$ where $\mathcal{B}[0,1]$ is the Borel σ-field on $[0,1]$ follows by the fact that X_n is right continuous with left limits. Let

$$G(R, s) = \sup_n \mathbb{P}(|X_n(s)| > R).$$

By (5.8), $G(R, s) \to 0$ as $R \to \infty$ for each s. Pick R large so that

$$\lambda(\{s : G(R, s) > \varepsilon\delta\}) < \varepsilon\delta.$$

Then

$$\int 1_{D_{Rn}}(\omega, s)\,\mathbb{P}(d\omega) = \mathbb{P}(|X_n(s)| > R) \le \begin{cases} 1, & G(r, s) > \varepsilon\delta,\\ \varepsilon\delta, & \text{otherwise.}\end{cases}$$

Integrating over $s \in [0,1]$,

$$\mathbb{P} \times \lambda(D_{Rn}) < 2\varepsilon\delta.$$

If $E_{Rn}(\omega) = \{s : (\omega, s) \in D_{Rn}\}$ and $F_{Rn} = \{\omega : \lambda(E_{Rn}) > \delta/4\}$, we have

$$\tfrac{1}{4}\delta\mathbb{P}(F_{Rn}) = \int_{F_{Rn}} \tfrac{1}{4}\delta\,\mathbb{P}(d\omega) \le \int\int_0^1 1_{D_{Rn}}(\omega,s)\,\lambda(ds)\,\mathbb{P}(d\omega) \le 2\varepsilon\delta,$$

so $\mathbb{P}(F_{Rn}) \le 8\varepsilon$.

Define $T = \inf\{t : |X_n(t)| \ge R + 2\varepsilon\} \wedge 2$ and define A_T, B_T, and C_T as in Step 2. We have

$$\mathbb{P}(C_T \cup F_{Rn}) \le 16\varepsilon.$$

If $\omega \notin C_T \cup F_{Rn}$ and $T < 2$, then $\lambda(E_{Rn}(\omega)) \le \delta/4$. Hence there exists $t \in [T, T + 2\delta]$ such that $|X_n(t)| \le R$ and $|X_n(t) - X_n(T)| \le \varepsilon$. Therefore $|X_n(T)| \le R + \varepsilon$, which contradicts the definition of T. We conclude that T must equal 2 on the complement of $C_T \cup F_{Rn}$, or in other words, except for an event of probability at most 16ε, we have $\sup_t |X_n(t)| \le R + 2\varepsilon$, provided, of course, that $n \ge n_0$.

An application of Corollary 5.10 completes the proof. □

Chapter 6

Markov processes

6.1 Introduction

It is not uncommon for a Markov process to be defined as a sextuple $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \theta_t, \mathbb{P}^x)$, and for additional notation (e.g., $\zeta, \Delta, \mathcal{S}, P_t, R_\lambda$, etc.) to be introduced rather rapidly. This can be intimidating for the beginner. We will explain this notation in as gentle a manner as possible. We will consider a Markov process to be a pair $(X_t, \mathbb{P}^x)$ (rather than a sextuple), where X_t is a single stochastic process and $\{\mathbb{P}^x\}$ is a family of probability measures, one probability measure $\mathbb{P}^x$ corresponding to each element x of the state space.

The idea that a Markov process consists of one process and many probabilities is one that takes some getting used to. To explain this, let us first look at an example. Suppose $X_1, X_2, \ldots$ is a Markov chain with stationary transition probabilities with 5 states: $1, 2, \ldots, 5$. Everything we want to know about X can be determined if we know $p(i,j) = \mathbb{P}(X_1 = j \mid X_0 = i)$ for each i and j and $\mu(i) = \mathbb{P}(X_0 = i)$ for each i. We sometimes think of having a different Markov chain for every choice of starting distribution $\mu = (\mu(1), \ldots, \mu(5))$. But instead let us define a new probability space by taking Ω' to be the collection of all sequences $\omega = (\omega_0, \omega_1, \ldots)$ such that each ω_n takes one of the values $1, \ldots, 5$. Define $X_n(\omega) = \omega_n$. Define $\mathcal{F}_n$ to be the σ-field generated by $X_0, \ldots, X_n$; this is the same as the σ-field generated by sets of the form $\{\omega : \omega_0 = a_0, \ldots, \omega_n = a_n\}$, where $a_0, \ldots, a_n \in \{1, 2, \ldots, 5\}$.

For each $x = 1, 2, \dots, 5$, define a probability measure $\mathbb{P}^x$ on Ω' by

$$\begin{aligned}\mathbb{P}^x(X_0 = x_0, X_1 = x_1, \dots X_n = x_n) \\ = 1_{\{x\}}(x_0)p(x_0, x_1) \cdots p(x_{n-1}, x_n).\end{aligned} \tag{6.1}$$

We have 5 different probability measures, one for each of $x = 1, 2, \dots, 5$, and we can start with an arbitrary probability distribution μ if we define $\mathbb{P}^\mu(A) = \sum_{i=1}^5 \mathbb{P}^i(A)\mu(i)$. We have lost no information by this redefinition, and it turns out this works much better when doing technical details.

The value of $X_0(\omega) = \omega_0$ can be any of $1, 2, \dots, 5$; the notion of starting at x is captured by $\mathbb{P}^x$, not by X_0. The probability measure $\mathbb{P}^x$ is concentrated on those ω's for which $\omega_0 = x$ and $\mathbb{P}^x$ gives no mass to any other ω.

Let us now look at a Lévy process, and see how this framework plays out there. Let $\mathbb{P}$ be a probability measure and let Z_t be a Lévy process with respect to $\mathbb{P}$ started at 0. Then $Z_t^x = x + Z_t$ is a Lévy process started at x. Let Ω' be the set of right continuous left limit functions from $[0, \infty)$ to $\mathbb{R}$, so that each element ω in Ω' is a right continuous left limit function. (We do not require that $\omega(0) = 0$ or that $\omega(0)$ take any particular value of x.) Define

$$X_t(\omega) = \omega(t). \tag{6.2}$$

This will be our process. Let $\mathcal{F}$ be the σ-field on Ω', the right continuous left limit functions, generated by the cylindrical subsets. Now define $\mathbb{P}^x$ to be the law of Z^x. This means that $\mathbb{P}^x$ is the probability measure on $(\Omega', \mathcal{F})$ defined by

$$\mathbb{P}^x(X \in A) = \mathbb{P}(Z^x \in A), \qquad x \in \mathbb{R}, A \in \mathcal{F}. \tag{6.3}$$

The probability measure $\mathbb{P}^x$ is determined by the fact that if $n \geq 1$, $t_1 \leq \cdots \leq t_n$, and $B_1, \dots, B_n$ are Borel subsets of $\mathbb{R}$, then

$$\mathbb{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) = \mathbb{P}(Z_{t_1}^x \in B_1, \dots, Z_{t_n}^x \in B_n).$$

6.2 Definition of a Markov process

We want to allow our Markov processes to take values in spaces other than the Euclidean ones. For now, we take our state space $\mathcal{S}$ to be a separable metric space, furnished with the Borel σ-field. For the first time around, just think of $\mathbb{R}$ in place of $\mathcal{S}$.

To define a Markov process, we start with a measurable space $(\Omega, \mathcal{F})$ and we suppose we have a filtration $\{\mathcal{F}_t\}$ (not necessarily satisfying the usual conditions).

Definition 6.1 *A* Markov process $(X_t, \mathbb{P}^x)$ *is a stochastic process*

$$X : [0, \infty) \times \Omega \to \mathcal{S}$$

and a family of probability measures $\{\mathbb{P}^x : x \in \mathcal{S}\}$ *on* $(\Omega, \mathcal{F})$ *satisfying the following.*

(1) For each t, X_t *is* $\mathcal{F}_t$ *measurable.*

(2) For each t *and each Borel subset* A *of* $\mathcal{S}$, *the map* $x \to \mathbb{P}^x(X_t \in A)$ *is Borel measurable.*

(3) For each $s, t \geq 0$, *each Borel subset* A *of* $\mathcal{S}$, *and each* $x \in \mathcal{S}$, *we have*

$$\mathbb{P}^x(X_{s+t} \in A \mid \mathcal{F}_s) = \mathbb{P}^{X_s}(X_t \in A), \qquad \mathbb{P}^x - \text{a.s.} \tag{6.4}$$

Some explanation is definitely in order. Let

$$\varphi(x) = \mathbb{P}^x(X_t \in A), \tag{6.5}$$

so that φ is a function mapping $\mathcal{S}$ to $\mathbb{R}$. Part of the definition of filtration is that each $\mathcal{F}_t \subset \mathcal{F}$. Since we are requiring X_t to be $\mathcal{F}_t$ measurable, that means that $(X_t \in A)$ is in $\mathcal{F}$ and it makes sense to talk about $\mathbb{P}^x(X_t \in A)$. Definition 6.1(2) says that the function φ is Borel measurable. This is a very mild assumption, and will be satisfied in the examples we look at.

The expression $\mathbb{P}^{X_s}(X_t \in A)$ on the right hand side of (6.4) is a random variable and its value at $\omega \in \Omega$ is defined to be $\varphi(X_s(\omega))$, with φ given by (6.5). Note that the randomness in $\mathbb{P}^{X_s}(X_t \in A)$ is thus all due to the X_s term and not the X_t term. Definition 6.1(3) can be rephrased as saying that for each s, t, each A, and each x, there is a set $N_{s,t,x,A} \subset \Omega$ that is a null set with respect to $\mathbb{P}^x$ and for $\omega \notin N_{s,t,x,A}$, the conditional expectation $\mathbb{P}^x(X_{s+t} \in A \mid \mathcal{F}_s)$ is equal to $\varphi(X_s)$.

We have now explained all the terms in the sextuple $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \theta_t, \mathbb{P}^x)$ except for θ_t. These are called shift operators and are maps from $\Omega \to \Omega$ such that $X_s \circ \theta_t = X_{s+t}$. We defer the precise meaning of the θ_t and the rationale for them until Section 6.4, where they will appear in a natural way.

In the remainder of the section and in Section 6.3 we define some of the additional notation commonly used for Markov processes. The first one is almost self-explanatory. We use $\mathbb{E}^x$ for expectation with respect to $\mathbb{P}^x$. As with $\mathbb{P}^{X_s}(X_t \in A)$, the notation $\mathbb{E}^{X_s} f(X_t)$, where f is bounded and Borel measurable, is to be taken to mean $\psi(X_s)$ with $\psi(y) = \mathbb{E}^y f(X_t)$.

If we want to talk about our Markov process started with distribution μ, we define

$$\mathbb{P}^\mu(B) = \int \mathbb{P}^x(B)\,\mu(dx),$$

and similarly for $\mathbb{E}^\mu$; here μ is a probability on $\mathcal{S}$.

6.3 Transition probabilities

If $\mathcal{B}$ is the Borel σ-field on a metric space $\mathcal{S}$, a *kernel* $Q(x, A)$ on $\mathcal{S}$ is a map from $\mathcal{S} \times \mathcal{B} \to \mathbb{R}$ satisfying the following.

(1) For each $x \in \mathcal{S}$, $Q(x, \cdot)$ is a measure on $(\mathcal{S}, \mathcal{B})$.

(2) For each $A \in \mathcal{B}$, the function $x \to Q(x, A)$ is Borel measurable.

The definition of Markov transition probabilities or simply transition probabilities is the following.

Definition 6.2 *A collection of kernels* $\{P_t(x, A); t \geq 0\}$ *are* Markov transition probabilities *for a Markov process* $(X_t, \mathbb{P}^x)$ *if*

(1) $P_t(x, \mathcal{S}) = 1$ *for each* $t \geq 0$ *and each* $x \in \mathcal{S}$.

(2) For each $x \in \mathcal{S}$, *each Borel subset* A *of* $\mathcal{S}$, *and each* $s, t \geq 0$,

$$P_{t+s}(x, A) = \int_{\mathcal{S}} P_t(y, A) P_s(x, dy). \tag{6.6}$$

(3) For each $x \in \mathcal{S}$, *each Borel subset* A *of* $\mathcal{S}$, *and each* $t \geq 0$,

$$P_t(x, A) = \mathbb{P}^x(X_t \in A). \tag{6.7}$$

Definition 6.2(3) can be rephrased as saying that for each x, the measures $P_t(x, dy)$ and $\mathbb{P}^x(X_t \in dy)$ are the same. We define

$$P_t f(x) = \int f(y) P_t(x, dy) \tag{6.8}$$

when $f : \mathcal{S} \to \mathbb{R}$ is Borel measurable and either bounded or non-negative.

The equations (6.6) are known as the *Chapman-Kolmogorov equations*. They can be rephrased in terms of equality of measures: for each x

$$P_{s+t}(x, dz) = \int_{y \in \mathcal{S}} P_t(y, dz) P_s(x, dy). \tag{6.9}$$

Multiplying (6.9) by a bounded Borel measurable function $f(z)$ and integrating gives

$$P_{s+t} f(x) = \int P_t f(y) P_s(x, dy). \tag{6.10}$$

The right hand side is the same as $P_s(P_t f)(x)$, so we have

$$P_{s+t} f(x) = P_s P_t f(x), \tag{6.11}$$

i.e., the functions $P_{s+t} f$ and $P_s P_t f$ are the same. The equation (6.11) is known as the *semigroup property*.

P_t is a linear operator on the space of bounded Borel measurable functions on $\mathcal{S}$. We can then rephrase (6.11) simply as

$$P_{s+t} = P_s P_t. \tag{6.12}$$

Operators satisfying (6.12) are called a *semigroup*, and are much studied in functional analysis.

One more observation about semigroups: if we take expectations in (6.4), we obtain

$$\mathbb{P}^x(X_{s+t} \in A) = \mathbb{E}^x \Big[\mathbb{P}^{X_s}(X_t \in A) \Big].$$

The left hand side is $P_{s+t} 1_A(x)$ and the right hand side is

$$\mathbb{E}^x [P_t 1_A(X_s)] = P_s P_t 1_A(x),$$

and so (6.4) encodes the semigroup property.

The *resolvent* or *λ-potential* of a semigroup P_t is defined by

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} P_t f(x)\, dt, \qquad \lambda \geq 0, \quad x \in \mathcal{S}.$$

This can be recognized as the Laplace transform of P_t. By the Fubini theorem, we see that

$$R_\lambda f(x) = \mathbb{E}^x \int_0^\infty e^{-\lambda t} f(X_t)\,dt.$$

Resolvents are useful because they are typically easier to work with than semigroups.

When practitioners of stochastic calculus tire of a martingale, they 'stop' it. Markov process theorists are a harsher lot and they 'kill' their processes. To be precise, attach an isolated point Δ to $\mathcal{S}$. Thus one looks at $\widehat{\mathcal{S}} = \mathcal{S} \cup \Delta$, and the topology on $\widehat{\mathcal{S}}$ is the one generated by the open sets of $\mathcal{S}$ and $\{\Delta\}$. Δ is called the *cemetery point*. All functions on $\mathcal{S}$ are extended to $\widehat{\mathcal{S}}$ by defining them to be 0 at Δ. At some random time ζ the Markov process is killed, which means that $X_t = \Delta$ for all $t \geq \zeta$. The time ζ is called the *lifetime* of the Markov process.

6.4 The canonical process and shift operators

Suppose we have a Markov process $(X_t, \mathbb{P}^x)$ where $\mathcal{F}_t = \sigma(X_s; s \leq t)$. Suppose that X_t has right continuous left limit paths. For this to even make sense, we need the set $\{t \to X_t \text{ is not right continuous left limit}\}$ to be in $\mathcal{F}$, and then we require this event to be $\mathbb{P}^x$-null for each x. Define $\widetilde{\Omega}$ to be the set of right continuous left limit functions on $[0, \infty)$. If $\widetilde{\omega} \in \widetilde{\Omega}$, set $\widetilde{X}_t = \widetilde{\omega}(t)$. Define $\widetilde{\mathcal{F}}_t = \sigma(\widetilde{X}_s; s \leq t)$ and $\widetilde{\mathcal{F}}_\infty = \vee_{t \geq 0} \widetilde{\mathcal{F}}_t$. Finally define $\widetilde{\mathbb{P}}^x$ on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}_\infty)$ by $\widetilde{\mathbb{P}}^x(\widetilde{X} \in \cdot) = \mathbb{P}^x(X \in \cdot)$. Thus $\widetilde{\mathbb{P}}^x$ is specified uniquely by

$$\widetilde{\mathbb{P}}^x(\widetilde{X}_{t_1} \in A_1, \ldots, \widetilde{X}_{t_n} \in A_n) = \mathbb{P}^x(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n)$$

for $n \geq 1$, $A_1, \ldots, A_n$ Borel subsets of $\mathcal{S}$, and $t_1 < \cdots < t_n$. Clearly there is so far no loss (nor gain) by looking at the Markov process $(\widetilde{X}_t, \widetilde{\mathbb{P}}^x)$, which is called the *canonical process*.

Let us now suppose we are working with the canonical process, and we drop the tildes everywhere. We define the *shift operators* $\theta_t : \Omega \to \Omega$ as follows. $\theta_t(\omega)$ will be an element of Ω and therefore is a continuous function from $[0, \infty)$ to $\mathcal{S}$. Define

$$\theta_t(\omega)(s) = \omega(t + s).$$

Then

$$X_s \circ \theta_t(\omega) = X_s(\theta_t(\omega)) = \theta_t(\omega)(s) = \omega(t+s) = X_{t+s}(\omega).$$

The shift operator θ_t takes the path of X and chops off and discards the part of the path before time t.

We will use expressions like $f(X_s) \circ \theta_t$. If we apply this to $\omega \in \Omega$, then

$$(f(X_s) \circ \theta_t)(\omega) = f(X_s(\theta_t(\omega))) = f(X_{s+t}(\omega)),$$

or $f(X_s) \circ \theta_t = f(X_{s+t})$.

Even if we are not in this canonical setup, from now on we will suppose there exist shift operators mapping Ω into itself so that

$$X_s \circ \theta_t = X_{s+t}.$$

6.5 Enlarging the filtration

Throughout the remainder of this chapter we assume that X has paths that are right continuous with left limits. To be more precise, if

$$N = \{\omega : \text{ the function } t \to X_t(\omega) \text{ is not right continuous with left limits}\},$$

then we assume $N \in \mathcal{F}$ and N is $\mathbb{P}^x$-null for every $x \in \mathcal{S}$.

Let us first introduce some notation. Define

$$\mathcal{F}_t^{00} = \sigma(X_s; s \le t), \qquad t \ge 0. \tag{6.13}$$

This is the smallest σ-field with respect to which each X_s is measurable for $s \le t$. We let $\mathcal{F}_t^0$ be the completion of $\mathcal{F}_t^{00}$, but we need to be careful what we mean by completion here, because we have more than one probability measure present. Let $\mathcal{N}$ be the collection of sets that are $\mathbb{P}^x$-null for every $x \in \mathcal{S}$. Thus $N \in \mathcal{N}$ if $(\mathbb{P}^x)^*(N) = 0$ for each $x \in \mathcal{S}$, where $(\mathbb{P}^x)^*$ is the outer probability corresponding to $\mathbb{P}^x$. The outer probability $(\mathbb{P}^x)^*$ is defined by

$$(\mathbb{P}^x)^*(S) = \inf\{\mathbb{P}^x(B) : A \subset B, B \in \mathcal{F}\}.$$

Let

$$\mathcal{F}_t^0 = \sigma(\mathcal{F}_t^{00} \cup \mathcal{N}). \tag{6.14}$$

Finally, let

$$\mathcal{F}_t = \mathcal{F}^0_{t+} = \cap_{\varepsilon>0}\mathcal{F}^0_{t+\varepsilon}. \tag{6.15}$$

We call $\{\mathcal{F}_t\}$ the *minimal augmented filtration* generated by X. The reason for worrying about which filtrations to use is that $\{\mathcal{F}^{00}_t\}$ is too small to include many interesting sets (such as those arising in the law of the iterated logarithm, for example), while if the filtration is too large, the Markov property will not hold for that filtration.

The filtration matters when defining a Markov process; see Definition 6.1(3).

We will make the following assumption.

Assumption 6.3 *Suppose $P_t f$ is continuous on $\mathcal{S}$ whenever f is bounded and continuous on $\mathcal{S}$.*

Markov processes satisfying Assumption 6.3 are called *Feller processes* or *weak Feller processes*. If $P_t f$ is continuous whenever f is bounded and Borel measurable, then the Markov process is said to be a *strong Feller process*.

One can show that under Assumption 6.3 we have

$$\mathbb{P}^x(X_{s+t}\in A \mid \mathcal{F}_s) = \mathbb{P}^{X_s}(X_t\in A), \qquad \mathbb{P}^x - \text{a.s.}$$

6.6 The Markov property

We start with the Markov property:

$$\mathbb{E}^{\,x}[f(X_{s+t}) \mid \mathcal{F}_s] = \mathbb{E}^{\,X_s}[f(X_t)], \qquad \mathbb{P}^x - \text{a.s.} \tag{6.16}$$

Since $f(X_{s+t}) = f(X_t)\circ\theta_s$, if we write Y for the random variable $f(X_t)$, we have

$$\mathbb{E}^{\,x}[Y\circ\theta_s \mid \mathcal{F}_s] = \mathbb{E}^{\,X_s}Y, \qquad \mathbb{P}^x - \text{a.s.} \tag{6.17}$$

We wish to generalize this to other random variables Y.

Proposition 6.4 *Let $(X_t,\mathbb{P}^x)$ be a Markov process and suppose* (6.16) *holds. Suppose $Y=\prod_{i=1}^n f_i(X_{t_i-s})$, where the f_i are bounded, Borel measurable, and $s\le t_1\le\ldots\le t_n$. Then* (6.17) *holds.*

Proof. We will prove this by induction on n. The case $n = 1$ is (6.16), so we suppose the equality holds for n and prove it for $n + 1$.

Let $V = \prod_{j=2}^{n+1} f_j(X_{t_j - t_1})$ and $h(y) = \mathbb{E}^y V$. By the induction hypothesis,

$$\begin{aligned}
\mathbb{E}^x\Big[\prod_{j=1}^{n+1} f_j(X_{t_j})|\mathcal{F}_s\Big] &= \mathbb{E}^x\Big[\mathbb{E}^x[V \circ \theta_{t_1}|\mathcal{F}_{t_1}]f_1(X_{t_1})|\mathcal{F}_s\Big]\\
&= \mathbb{E}^x\Big[(\mathbb{E}^{X_{t_1}}V)f_1(X_{t_1})|\mathcal{F}_s\Big]\\
&= \mathbb{E}^x[(hf_1)(X_{t_1})|\mathcal{F}_s].
\end{aligned}$$

By (6.16) this is $\mathbb{E}^{X_s}[(hf_1)(X_{t_1-s})]$. For any y,

$$\begin{aligned}
\mathbb{E}^y[(hf_1)(X_{t_1-s})] &= \mathbb{E}^y[(\mathbb{E}^{X_{t_1-s}}V)f_1(X_{t_1-s})]\\
&= \mathbb{E}^y\Big[\mathbb{E}^y[V \circ \theta_{t_1-s}|\mathcal{F}_{t_1-s}]f_1(X_{t_1-s})\Big]\\
&= \mathbb{E}^y[(V \circ \theta_{t_1-s})f_1(X_{t_1-s})].
\end{aligned}$$

If we replace V by its definition, replace y by X_s, and use the definition of θ_{t_1-s}, we get the desired equality for $n+1$ and hence the induction step. □

We now come to the general version of the Markov property. As usual, $\mathcal{F}_\infty = \vee_{t\geq 0}\mathcal{F}_t$. The expression $Y \circ \theta_t$ for general Y may seem puzzling at first.

Theorem 6.5 *Let* $(X_t, \mathbb{P}^x)$ *be a Markov process and suppose* (6.16) *holds. Suppose* Y *is bounded and measurable with respect to* $\mathcal{F}_\infty$. *Then*

$$\mathbb{E}^x[Y \circ \theta_s \mid \mathcal{F}_s] = \mathbb{E}^{X_s}Y, \qquad \mathbb{P}^x - \text{a.s.} \tag{6.18}$$

The proof follows from the previous proposition by a monotone class argument.

6.7 Strong Markov property

Given a stopping time T, recall that the σ-field of events known up to time T is defined to be

$$\mathcal{F}_T = \big\{A \in \mathcal{F}_\infty : A \cap (T \leq t) \in \mathcal{F}_t \text{ for all } t > 0\big\}$$

We define θ_T by $\theta_T(\omega)(t) = \omega(T(\omega) + t)$. Thus, for example, $X_t \circ \theta_T(\omega) = X_{T(\omega)+t}(\omega)$ and $X_T(\omega) = X_{T(\omega)}(\omega)$.

Now we can state the strong Markov property.

Suppose $(X_t, \mathbb{P}^x)$ is a Markov process with respect to $\{\mathcal{F}_t\}$. The strong Markov property is said to hold if whenever T is a finite stopping time and Y is bounded and measurable with respect to $\mathcal{F}_\infty$, then

$$\mathbb{E}^x[Y \circ \theta_T | \mathcal{F}_T] = \mathbb{E}^{X_T} Y, \qquad \mathbb{P}^x - \text{a.s.}$$

Recall that we are restricting our attention to Markov processes whose paths are right continuous with left limits. If we have a Markov process $(X_t, \mathbb{P}^x)$ whose paths are right continuous with left limits, which has shift operators $\{\theta_t\}$, and which satisfies the strong Markov property, whether or not Assumption 6.3 holds, then we say that $(X_t, \mathbb{P}^x)$ is a *strong Markov process*. A strong Markov process is said to be *quasi-left continuous* if $X_{T_n} \to X_T$, a.s., on $\{T < \infty\}$ whenever T_n are stopping times increasing up to T. Unlike in the definition of predictable stopping times, we are not requiring the T_n to be strictly less than T. A *Hunt process* is a strong Markov process that is quasi-left continuous. Quasi-left continuity does not imply left continuity; consider the Poisson process.

Chapter 7

Stable-like processes

7.1 Martingale problems

We saw in the chapter on SDEs that if X_t is a Lévy process, then

$$f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)\,ds$$

is a martingale, where $f \in C_b^2$, the C^2 functions such that f, f', f'' are bounded, and

$$\mathcal{L}f(x) = \int [f(x+h) - f(x) - f'(x)h1_{(|h|\le 1)}]\,m(dh).$$

There is something analogous for all Markov processes. Given a Markov process $(X_t, \mathbb{P}^x)$, we say $\mathcal{L}$ is the weak infinitesimal generator if

$$\frac{P_t f(x) - f(x)}{t} \to \mathcal{L}f(x)$$

boundedly and pointwise as $t \to 0$ for all f in some domain. This is different from the usual definition of infinitesimal generator in functional analysis, as there the convergence has to be in norm.

Proposition 7.1 *If f is in the domain of the weak infinitesimal generator, then $M_t = f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)\,ds$ is a martingale.*

Proof. Note

$$\frac{P_{s+h}f(x) - P_s f(x)}{h} = P_s\Big(\frac{P_h f - f}{h}\Big)(x).$$

By dominated convergence, the right hand side converges to $P_s\mathcal{L}f(x)$ as $h \to 0$. Therefore the derivative of $P_s f(x)$ is $P_s\mathcal{L}f(x)$.

Since $P_s\mathcal{L}f$ is bounded,

$$P_t f(x) - f(x) = \int_0^t P_s\mathcal{L}f(x)\,ds,$$

or

$$\mathbb{E}^x f(X_t) - \mathbb{E}^x f(X_0) = \mathbb{E}^x \int_0^t \mathcal{L}f(X_s)\,ds.$$

By the Markov property,

$$\mathbb{E}^x[M_t - M_s \mid \mathcal{F}_s] = \mathbb{E}^{X_s} f(X_{t-s}) - f(X_s) - \mathbb{E}^{X_s}\int_0^{t-s} \mathcal{L}f(X_r)\,dr = 0.$$

This is what we want. □

Here is some terminology. Let $\mathcal{L}$ be an operator, x_0 a point in the state space. We say a probability measure $\mathbb{P}$ is a solution to the martingale problem for $\mathcal{L}$ started at x_0 if
(1) $\mathbb{P}(X_0 = x_0) = 1$ a.s.
(2) for all f in the domain of $\mathcal{L}$, $f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)\,ds$ is a martingale.

The martingale problem is well posed if there is a solution to the martingale problem started at x_0 for each x_0 in the state space and the solution is unique.

Here is a fact that we will not prove.

Theorem 7.2 *Suppose the martingale problem is well posed. If $\mathbb{P}^x$ is the solution to the martingale problem started at x_0, then $(X_t, \mathbb{P}^x)$ is a strong Markov process.*

7.2 Stable-like processes

The term stable-like process refers to several types of processes. Here is one of them. We have seen that if we let $\mathcal{L}f$ be defined by

$$\mathcal{L}f(x) = \int [f(x+h) - f(x) - f'(x)h1_{(|h|\le 1)}]\, m(dh)$$

for $f \in C_b^2$, then $\mathbb{P}$ is a solution to the martingale problem for the Lévy process started at 0. In particular, if we have a symmetric stable process of index α, then $m(dh)$ gets replaced by $c/|h|^{1+\alpha}\, dh$.

Let us suppose we have a strong Markov process such that for every x, $\mathbb{P}^x$ is the unique solution to the martingale problem for $\mathcal{L}$ started at x, where we define

$$\mathcal{L}f(x) = \int [f(x+h) - f(x) - f'(x)h1_{(|h|\le 1)}]\, \frac{A(x,h)}{|h|^{1+\alpha}}\, dh,$$

for $f \in C_b^2$. The way to think about this is that it is a lot like a stable process, but its intensity varies from point to point, and the intensity also varies with the size of the jump. We suppose there exist constants c_1, c_2 such that

$$0 < c_1 \le A(x,h) \le c_2 < \infty$$

for all x and h.

We also want to go to higher dimensions, so we replace $|h|^{1+\alpha}$ by $|h|^{d+\alpha}$ and $f'(x)h$ by $\nabla f(x)\cdot h$. Now C_b^2 refers to C^2 functions f such that f and all its first and second partial derivatives are bounded.

7.3 Some properties

Let us begin by describing more carefully the processes we wish to consider. A probability measure $\mathbb{P}$ on the space $D[0,\infty)$ is a solution to the martingale problem for $\mathcal{L}$ started at x if $X_t(\omega) = \omega(t)$ are the coordinate maps, $\mathcal{F}_t$ is the σ-field generated by the cylindrical sets, and
(1) we have $\mathbb{P}(X_0 = x) = 1$, and
(2) for each $f \in C_b^2$ we have that

$$f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)ds \tag{7.1}$$

is a $\mathbb{P}$-martingale, where

$$\mathcal{L}f(x) = \int_{\mathbb{R}^d - \{0\}} [f(x+h) - f(x) - \nabla f(x) \cdot h 1_{(|h| \le 1)}] n(x,h) dh.$$

The symmetry assumption we will impose on n will make the presence of the ∇f term have no effect; moreover we could replace the $1_{(|h|\le 1)}$ term by $1_{(|h|\le M)}$ with any $M > 0$ whatsoever.

We assume that $(\mathbb{P}^x, X_t)$ is a strong Markov process with state space $\mathbb{R}^d$ such that for each x the probability measure $\mathbb{P}^x$ is a solution to the martingale problem for $\mathcal{L}$ started at x.

Throughout this chapter we make the following assumption.

Assumption 7.3 (a) For all x and h we have $n(x,-h) = n(x,h)$.
(b) There exist constants $\kappa \in (0,1)$ and $\alpha \in (0,2)$ such that for all x and h we have

$$\frac{\kappa}{|h|^{d+\alpha}} \le n(x,h) \le \frac{\kappa^{-1}}{|h|^{d+\alpha}}. \tag{7.2}$$

The proof of the following scaling property is an easy change of variables argument.

Proposition 7.4 *Suppose $(\mathbb{P}^x, X_t)$ is as above, $a > 0$, and $Y_t = aX_{a^{-\alpha}t}$. Define $\mathbb{Q}^x = \mathbb{P}^{x/a}$. Then $(\mathbb{Q}^x, Y_t)$ is a strong Markov process. We have $\mathbb{Q}^x(Y_0 = x) = 1$ and if $f \in C^2$, then $f(Y_t) - f(Y_0) - \int_0^t \widetilde{\mathcal{L}} f(Y_s) ds$ is a $\mathbb{Q}^x$-martingale, where $\widetilde{\mathcal{L}} f(x) = \int [f(x+h) - f(x) - \nabla f(x) \cdot h 1_{(|h|\le 1)}] \widetilde{n}(x,h) dh$ and $\widetilde{n}$ satisfies (7.2) with the same values of κ and α.*

Proof. Because $(\mathbb{P}^x, X_t)$ is strong Markov and Y_t is a constant multiple of a time change of X_t, then $(\mathbb{Q}^x, Y_t)$ is strong Markov. That $\mathbb{Q}^x(Y_0 = 1) = 1$ is clear. Let

$$\widetilde{n}(y,k) = a^{-(d+\alpha)} n(a^{-1}y, a^{-1}k)$$

and

$$\widetilde{\mathcal{L}} f(y) = \int_{\mathbb{R}^d - \{0\}} [f(y+k) - f(y) - \nabla f(y) \cdot k 1_{(|k|\le 1)}] \widetilde{n}(y,k) dk.$$

Clearly $\widetilde{n}$ satisfies (2.3) with the same values of κ and α. Let $f \in C^2$ and set $g(x) = f(ax)$. Then

$$g(X_{a^{-\alpha}t}) - g(X_0) - \int_0^{a^{-\alpha}t} \mathcal{L}g(X_s)ds$$

is a martingale, hence so is

$$g(X_{a^{-\alpha}t}) - g(X_0) - \int_0^t a^{-\alpha}\mathcal{L}g(X_{a^{-\alpha}s})ds.$$

Consequently

$$f(Y_t) - f(Y_0) - \int_0^t a^{-\alpha}\mathcal{L}g(a^{-1}Y_s)ds$$

is also a martingale.

It remains to check that $a^{-\alpha}\mathcal{L}g(a^{-1}y) = \widetilde{\mathcal{L}}f(y)$. This follows because, omitting the gradient term for simplicity,

$$\begin{aligned}
a^{-\alpha}\mathcal{L}g(a^{-1}y) &= a^{-\alpha}\int [g(a^{-1}y+k) - g(a^{-1}y)]n(a^{-1}y,k)dk \\
&= a^{-\alpha}\int [f(y+ak) - f(y)]n(a^{-1}y,k)dk \\
&= a^{d}\int [f(y+ak) - f(y)]\widetilde{n}(y,ak)dk \\
&= \int [f(y+h) - f(y)]\widetilde{n}(y,h)dh \\
&= \widetilde{\mathcal{L}}f(y).
\end{aligned}$$

□

We will also need the following fact, known as the Lévy system formula.

Proposition 7.5 *Suppose A and B are Borel sets that are a positive distance from each other. Then*

$$\sum_{s\le t} 1_{(X_{s-}\in A, X_s\in B)} - \int_0^t 1_A(X_s)\int_B n(X_s, u-X_s)du\,ds$$

is a $\mathbb{P}^x$-martingale for each x.

Proof. Let $f \in C^2$ with $f = 0$ on A and $f = 1$ on B. Let M_t^f denote the martingale in (7.1). Then $\int_0^t 1_A(X_{s-})dM_t^f$ is also a martingale under $\mathbb{P}^x$, since the stochastic integral with respect to a martingale is a martingale. Since $f(X_t) - f(X_0) = \sum_{s\le t}[f(X_s) - f(X_{s-})]$, this says that

$$\sum_{s\le t}[1_A(X_{s-})(f(X_s) - f(X_{s-}))] - \int_0^t 1_A(X_{s-})\mathcal{L}f(X_s)ds$$

is a martingale. Since $X_{s-} \neq X_s$ for only countably many values of s, then

$$\sum_{s\le t}[1_A(X_{s-})(f(X_s) - f(X_{s-}))] - \int_0^t 1_A(X_s)\mathcal{L}f(X_s)ds \tag{7.3}$$

is also a martingale. Now if $x \in A$, then $f(x)$ and $\nabla f(x)$ are both equal to 0, and so

$$\mathcal{L}f(x) = \int_{\mathbb{R}^d-\{0\}} f(x+h)n(x,h)dh = \int_{\mathbb{R}^d-\{0\}} f(u)n(x,u-x)du\,.$$

Note $n(x,h)$ is integrable over h in the complement of any neighborhood of the origin. Because A and B are a positive distance from each other, the sum on the left of (7.3) is actually a finite sum. With these facts we can pass to a limit to see that

$$\sum_{s\le t}[1_A(X_{s-})(1_B(X_s) - 1_B(X_{s-})] - \int_0^t 1_A(X_s)\int_B n(X_s,u-X_s)du\,ds$$

is a martingale, which is equivalent to what we wanted to prove. □

By taking limits, it is not necessary to assume that A and B are a positive distance apart, but only that they are disjoint.

We let $B(x,r)$ denote the ball of radius r centered at x. We use $|A|$ to denote the Lebesgue measure of A. Set

$$\tau_A = \inf\{t>0 : X_t \notin A\}, \qquad T_A = \inf\{t>0 : X_t \in A\}.$$

7.4 Harnack inequality

We begin this section by proving a tightness result.

Proposition 7.6 *There exists c_1 depending only on κ and not x such that*

$$\mathbb{P}^x(\sup_{s\le t}|X_s - X_0| > 1) \le c_1 t.$$

Proof. Let f be a C^2 function taking values in $[0,1]$ such that $f(0)=0$ and $f(y)=1$ if $|y|\ge 1$. Let $f_x(y)=f(y-x)$. By the Taylor expansion of f_x,

$$|(f_x(z+h)-f_x(z))+(f_x(z-h)-f_x(z))| \le c_2|h|^2. \tag{7.4}$$

Since n is symmetric, this and and our assumptions imply

$$\begin{aligned}
|\mathcal{L}f_x(z)| &\le \Big|\int_{|h|\le 1}[f_x(z+h)-f_x(z)]n(z,h)dh\Big| \\
&\qquad + \Big|\int_{|h|>1}[f_x(z+h)-f_x(z)]n(z,h)dh\Big| \\
&\le c_3\int_{|h|\le 1}|h|^2 n(z,h)dh + c_4\int_{|h|>1} n(z,h)dh \\
&\le c_5.
\end{aligned}$$

We now use (7.1) to write

$$\mathbb{E}^x f_x(X_{\tau_{B(x,1)}\wedge t}) - f_x(x) = \mathbb{E}^x\int_0^{\tau_{B(x,1)}\wedge t}\mathcal{L}f_x(X_s)ds \le c_5 t.$$

If X_t exits $B(x,1)$ before time t then $f_x(X_{\tau_{B(x,1)}\wedge t})=1$, and so the left hand side is greater than $\mathbb{P}^x(\tau_{B(x,1)}\le t)$. □

Lemma 7.7 *Let $\varepsilon>0$. There exists c_1 depending only on ε such that if $x\in\mathbb{R}^d$ and $r>0$, then*

$$\inf_{z\in B(x,(1-\varepsilon)r)}\mathbb{E}^z\tau_{B(x,r)} \ge c_1 r^\alpha.$$

Proof. By scaling we may assume $r = 1$. By the previous proposition and scaling, if $z \in B(x, 1-\varepsilon)$

$$\mathbb{P}^z(\tau_{B(x,1)} \le \varepsilon^\alpha t) \le \mathbb{P}^z(\sup_{s\le\varepsilon^\alpha t} |X_s - X_0| \ge \varepsilon) \le c_2 t.$$

Thus

$$\mathbb{E}^z \tau_{B(x,1)} \ge \varepsilon^\alpha t \mathbb{P}^z(\tau_{B(x,1)} \ge \varepsilon^\alpha t) \ge \varepsilon^\alpha t(1 - c_2 t)\,.$$

Taking $t = 1/(2c_2)$ yields a uniform lower bound. □

Lemma 7.8 *There exists* c_1 *such that* $\sup_z \mathbb{E}^z \tau_{B(x,r)} \le c_1 r^\alpha$.

Proof. By scaling, we may suppose $r = 1$. Let S be the time of the first jump larger than 2. We want to show there exists $c_2 \in (0, \frac{1}{2})$ such that $\mathbb{P}^z(S \le 1) > c_2$ for all z. For z such that $\mathbb{P}^z(S \le 1) \ge \frac{1}{2}$, there is nothing to show. So suppose z is such that $\mathbb{P}^z(S \le 1) < \frac{1}{2}$. By an argument similar to that in Proposition 7.5,

$$\sum_{s\le t} 1_{(|X_s - X_{s-}|>2)} - \int_0^t \int_{(|h|>2)} n(X_s, h)dh$$

is a martingale. Then by optional stopping and by the lower bounds on n

$$\begin{aligned}\mathbb{P}^z(S \le 1) &= \mathbb{E}^z \sum_{s\le S\wedge 1} 1_{(|X_s - X_{s-}|>2)} \\ &= \mathbb{E}^z \int_0^{S\wedge 1} \int_{(|h|>2)} n(X_s, h)dh\, ds \\ &\ge c_3 \mathbb{E}^z(S \wedge 1) \ge c_3 \mathbb{P}^z(S > 1) \ge c_3/2.\end{aligned}$$

Letting $c_2 = (1 \wedge c_3)/2$, we have $\mathbb{P}^z(S \le 1) \ge c_2$.

If there is a jump larger than 2 before time 1, then $\tau_{B(x,1)} \le 1$. So

$$\sup_z \mathbb{P}^z(\tau_{B(x,1)} > 1) \le 1 - c_2.$$

Let θ_t be the usual shift operator for Markov processes. By the Markov property,

$$\begin{aligned}\mathbb{P}^z(\tau_{B(x,1)} > m+1) &\le \mathbb{P}^z(\tau_{B(x,1)} > m, \tau_{B(x,1)} \circ \theta_m > 1)\\ &= \mathbb{E}^z\Big[\mathbb{P}^{X_m}(\tau_{B(x,1)} > 1); \tau_{B(x,1)} > m\Big]\\ &\le (1-c_2)\mathbb{P}^z(\tau_{B(x,1)} > m).\end{aligned}$$

By induction, $\mathbb{P}^z(\tau_{B(x,1)} > m) \le (1-c_2)^m$, which implies that $\tau_{B(x,1)}$ has moments of all orders. □

Next we show X_t will hit sets of positive Lebesgue measure with positive probability.

Proposition 7.9 *Suppose $A \subset B(x,1)$. There exists c_1 not depending on x or A such that*

$$\mathbb{P}^y(T_A < \tau_{B(x,3)}) \ge c_1|A|, \qquad y \in B(x,2).$$

Proof. Fix $y \in B(x,2)$. Write τ for $\tau_{B(x,3)}$. If X_t is in A for some t less than time τ with probability larger than $1/4$, we are done, so assume $\mathbb{P}^y(T_A < \tau) \le 1/4$. Using a previous proposition and scaling, choose t_0 small enough so that the probability that τ occurs before time t_0 is also less than $1/4$. Note that T_A cannot equal τ because $A \subset B(x,1)$. For $|h| \le 4$, $n(X_s,h)$ is bounded below by our assumptions. Hence for $X_s \in B(x,3)$ and $u \in A \subset B(x,1)$, we have $|X_s - u| \le 4$, and consequently $n(X_s, u - X_s)$ is bounded below. So

$$\begin{aligned}\mathbb{P}^y(T_A < \tau) &\ge \mathbb{E}^y \sum_{s \le T_A \wedge \tau \wedge t_0} 1_{(X_{s-} \ne X_s, X_s \in A)}\\ &= \mathbb{E}^y \int_0^{T_A \wedge \tau \wedge t_0} \int_A n(X_s, u - X_s)du\,ds\\ &\ge c_2|A|\mathbb{E}^y(T_A \wedge \tau \wedge t_0).\end{aligned}$$

Now

$$\begin{aligned}\mathbb{E}^y(T_A \wedge \tau \wedge t_0) &\ge \mathbb{E}^y(t_0; T_A \ge \tau \ge t_0)\\ &= t_0\mathbb{P}^y(T_A \ge \tau \ge t_0)\\ &\ge t_0[1 - \mathbb{P}^y(T_A < \tau) - \mathbb{P}^y(\tau < t_0)] \ge t_0/2.\end{aligned}$$

Combining this with the above,

$$\mathbb{P}^y(T_A < \tau) \geq c_2 |A| t_0/2.$$

$\square$

Proposition 7.10 *There exist c_1 and c_2 such that if $x \in \mathbb{R}^d, r > 0, z \in B(x,r)$, and H is a bounded nonnegative function supported in $B(x,2r)^c$, then*

$$\mathbb{E}^z H(X_{\tau_{B(x,r)}}) \leq c_1 \Big(\mathbb{E}^z \tau_{B(x,r)}\Big) \int \frac{H(y)}{|y-x|^{d+\alpha}} dy$$

and

$$\mathbb{E}^z H(X_{\tau_{B(x,r)}}) \geq c_2 \Big(\mathbb{E}^z \tau_{B(x,r)}\Big) \int \frac{H(y)}{|y-x|^{d+\alpha}} dy.$$

Proof. Note $H(w) = 0$ if $w \in B(x,r)$ and $H(X_{\tau_{B(x,r)}}) > 0$ only if there is a jump from $B(x,r)$ to $B(x,2r)^c$. By optional stopping, if $B \subset B(x,2r)^c$

$$\begin{aligned}
\mathbb{E}^z 1_{(X_{t\wedge\tau(B(x,r))} \in B)} &= \mathbb{E}^z \int_0^{t\wedge\tau(B(x,r))} \int_B n(X_s, u - X_s) du\, ds \\
&\leq \mathbb{E}^z \int_0^{t\wedge\tau(B(x,r))} \int_B \frac{c_3}{|u - X_s|^{d+\alpha}} du\, ds \\
&\leq c_4 \mathbb{E}^z (t \wedge \tau_{B(x,r)}) \int_B \frac{dy}{|y-x|^{d+\alpha}}.
\end{aligned}$$

Letting $t \to \infty$, using monotone convergence on the right and dominated convergence on the left, we have

$$\mathbb{E}^z 1_B(X_{\tau_{B(x,r)}}) \leq c_4 \Big(\mathbb{E}^z \tau_{B(x,r)}\Big) \int \frac{1_B(y)}{|y-x|^{d+\alpha}} dy.$$

Using linearity we have the above when 1_B is replaced by a simple function; approximating H by simple functions and taking limits, we have the first inequality in the statement of the proposition.

The proof of the second inequality is exactly similar, using the lower bound for n instead of the upper bound. $\square$

We say a bounded function $h : \mathbb{R}^d \to \mathbb{R}$ is $\mathcal{L}$-harmonic in a domain D if $h(X_{t\wedge\tau_D})$ is a $\mathbb{P}^x$-martingale for all x. It is easy to see that if h is C^2 in D, and $\mathcal{L}h(x) = 0$ for $x \in D$, then h will be $\mathcal{L}$-harmonic.

Theorem 7.11 *There exists c_1 such that if h is nonnegative and bounded on $\mathbb{R}^d$ and $\mathcal{L}$-harmonic in $B(x_0, 16)$, then*

$$h(x) \le c_1 h(y), \qquad x, y \in B(x_0, 1).$$

Proof. By looking at a constant multiple of h, we may assume $\inf_{B(x_0,1)} h = \frac{1}{2}$. Choose $z_0 \in B(x_0, 1)$ such that $h(z_0) \le 1$. We want to show that h is bounded above in $B(x_0, 1)$ by a constant not depending on h. We will establish this by contradiction: if there exists a point $x \in B(x_0, 1)$ with $h(x) = K$ where K is too large, we can obtain a sequence of points in $B(x_0, 2)$ on which h is unbounded.

Let $\varepsilon < \frac{1}{3}$ be chosen so that $|B(0, 1-\varepsilon)|/|B(0, 1)| \ge \frac{3}{4}$. Using our lemmas and propositions, there exists c_2 such that if $x \in \mathbb{R}^d, r > 0$, and H is a nonnegative function supported on $B(x, 2r)^c$, then for $y, z \in B(x, (1-\varepsilon)r)$,

$$\mathbb{E}^{\,z} H(X_{\tau(B(x,r))}) \le c_2 \mathbb{E}^{\,y} H(X_{\tau(B(x,r))}). \tag{7.5}$$

By a proposition there exists c_3 such that if $A \subset B(x_0, 4)$,

$$\mathbb{P}^y(T_A < \tau_{B(x_0,16)}) \ge c_3 |A|, \qquad y \in B(x_0, 8). \tag{7.6}$$

Also there exists $c_4 \le 1$ such that if $x \in \mathbb{R}^d, r > 0$, and $C \subset B(x, r/3)$ with $|C|/|B(x, r/3)| \ge \frac{1}{3}$, then

$$\mathbb{P}^x(T_C < \tau_{B(x,r)}) \ge c_4. \tag{7.7}$$

Let

$$\eta = \frac{c_4}{3}, \qquad \zeta = \frac{1}{3} \wedge (c_2^{-1}\eta). \tag{7.8}$$

Now suppose there exists $x \in B(x_0, 2)$ with $h(x) = K$ for some $K > 2$. Let r be chosen so that

$$|B(x, r/3)| = 2/(c_3 \zeta K). \tag{7.9}$$

Note this implies

$$r = c_5 K^{-1/d}. \tag{7.10}$$

Let us write B_r for $B(x,r)$, τ_r for $\tau_{B(x,r)}$ and similarly B_{2r} and τ_{2r}. Let A be a compact set contained in

$$A' = \{w \in B(x, r/3) : h(w) \geq \zeta K\}.$$

By (7.6) and optional stopping,

$$\begin{aligned} 1 \geq h(z_0) &\geq \mathbb{E}^{z_0}[h(X_{T_A \wedge \tau_{B(x_0,16)}}); T_A < \tau_{B(x_0,16)}] \\ &\geq \zeta K \mathbb{P}^{z_0}(T_A < \tau_{B(x_0,16)}) \\ &\geq c_3 \zeta K |A|, \end{aligned}$$

hence

$$\frac{|A|}{|B(x,r/3)|} \leq \frac{1}{c_3 \zeta K |B(x,r/3)|} \leq \frac{1}{2}.$$

This implies $|A'|/|B(x,r/3)| \leq \frac{1}{2}$. Let C be a compact set contained in $B(x,r/3) - A'$ such that

$$\frac{|C|}{|B(x,r/3)|} \geq \frac{1}{3}. \tag{7.11}$$

Let $H = h 1_{B_{2r}^c}$. We claim

$$\mathbb{E}^x[h(X_{\tau_r}); X_{\tau_r} \notin B_{2r}] \leq \eta K$$

If not

$$\mathbb{E}^x H(X_{\tau_r}) > \eta K,$$

and by (7.5), for all $y \in B(x, r/3)$,

$$\begin{aligned} h(y) &\geq \mathbb{E}^y h(X_{\tau_r}) \geq \mathbb{E}^y[h(X_{\tau_r}); X_{\tau_r} \notin B_{2r}] \\ &\geq c_2^{-1} \mathbb{E}^x H(X_{\tau_r}) \geq c_2^{-1} \eta K \\ &\geq \zeta K, \end{aligned}$$

contradicting (7.11) and the definition of A', noting that $|C|/|B(x,r/3)| \geq \frac{1}{3}$ and so A' is a proper subset of $B(x,r/3)$.

Let $M = \sup_{B_{2r}} h(z)$. We then have

$$\begin{aligned} K = h(x) &= \mathbb{E}^x[h(X_{T_C}); T_C < \tau_r] + \mathbb{E}^x[h(X_{\tau_r}); \tau_r < T_C, X_{\tau_r} \in B_{2r}] \\ &\quad + \mathbb{E}^x[h(X_{\tau_r}); \tau_r < T_C, X_{\tau_r} \notin B_{2r}] \\ &\le \zeta K \mathbb{P}^x(T_C < \tau_r) + M\mathbb{P}^x(\tau_r < T_C) + \eta K \\ &= \zeta K \mathbb{P}^x(T_C < \tau_r) + M(1 - \mathbb{P}^x(T_C < \tau_r)) + \eta K, \end{aligned}$$

or

$$\frac{M}{K} \ge \frac{1 - \eta - \zeta \mathbb{P}^x(T_C < \tau_r)}{1 - \mathbb{P}^x(T_C < \tau_r)}.$$

Since $\zeta < \frac{1}{3}$, then $c_4(1-\zeta) > c_4/3 = \eta$, and then

$$\mathbb{P}^x(T_C < \tau_r)(1-\zeta) \ge c_4(1-\zeta) > \eta,$$

hence

$$\mathbb{P}^x(T_C < \tau_r) > \eta + \zeta \mathbb{P}^x(T_C < \tau_r).$$

This implies

$$1 - \mathbb{P}^x(T_C < \tau_r) < 1 - \eta - \zeta \mathbb{P}^x(T_C < \tau_r),$$

and therefore $M/K > 1$.

Using (7.7) and (7.8) there exists $\beta > 0$ such that $M \ge K(1+2\beta)$. Therefore there exists $x' \in B(x, 2r)$ with $h(x') \ge K(1+\beta)$.

Now suppose there exists $x_1 \in B(x_0, 1)$ with $h(x_1) = K_1$. Define r_1 in terms of K_1 analogously to (7.9). Using the above argument (with x_1 replacing x and x_2 replacing x'), there exists $x_2 \in B(x_1, 2r_1)$ with $h(x_2) = K_2 \ge (1+\beta)K_1$. We continue and obtain r_2 and then x_3, K_3, r_3, etc. Note $x_{i+1} \in B(x_i, 2r_i)$ and $K_i \ge (1+\beta)^{i-1}K_1$. In view of (7.10),

$$|x_{i+1} - x_i| \le 2r_i \le c/K_i^{1/d} \le \frac{c}{K_1^{1/d}(1+\beta)^{i/d}},$$

which is summable, hence

$$\sum_i |x_{i+1} - x_i| \le c_6 K_1^{-1/d}.$$

So if $K_1 > c_6^d$, then we have a sequence $x_1, x_2, \ldots$ contained in $B(x_0, 2)$ with $h(x_i) \ge (1+\beta)^{i-1}K_1 \to \infty$, a contradiction to h being bounded on $\mathbb{R}^d$. Therefore we cannot take K_1 larger than $c_1 = c_6^d$, and thus $\sup_{B(x_0,1)} h(y) \le c_1$, which is what we wanted to prove. □

Corollary 7.12 *Suppose D is a bounded connected domain and $r > 0$. There exists c_1 depending only on D and r such that if h is nonnegative and bounded in $\mathbb{R}^d$ and $\mathcal{L}$-harmonic in D, then $h(x) \le c_1 h(y)$ if $x, y \in D$ and* $\mathrm{dist}\,(x, \partial D)$ *and* $\mathrm{dist}\,(y, \partial D)$ *are both greater than r.*

Proof. We form a sequence $x = y_0, y_1, y_2, \ldots, y_m = y$ such that $|y_{i+1} - y_i| < (a_{i+1} \wedge a_i)/32$, where $a_i = \mathrm{dist}\,(y_i, \partial D)$ and each $a_i < r$. By compactness we can choose M depending only on r so that no more than M points y_i are needed. By scaling and and the previous theorem, $h(y_i) \le c_2 h(y_{i+1})$ with $c_2 > 1$. So

$$h(x) = h(y_0) \le c_2 h(y_1) \le \cdots \le c_2^m h(y_m) = c_2^m h(y) \le c_2^M h(y).$$

□

7.5 Regularity

In this section we obtain some estimates on equicontinuity of resolvents.

Theorem 7.13 *If h is bounded on $\mathbb{R}^d$ and $\mathcal{L}$-harmonic in a ball $B(x_0, 2)$, then h is Hölder continuous in $B(x_0, 1)$: there exist c_1 and $\beta > 0$ such that*

$$|h(x) - h(y)| \le c_1 \|h\|_\infty |x - y|^\beta, \qquad x, y \in B(x_0, 1).$$

Proof. By a proposition there exists c_2 such that if $x \in \mathbb{R}^d, r > 0$, and $A \subset B(x, r/3)$ with $|A|/|B(x, r/3)| \ge \frac{1}{3}$, then

$$\mathbb{P}^x(T_A < \tau_{B(x,r)}) \ge c_2. \tag{7.12}$$

By our propositions and lemmas with $H = 1_{B(x,s)^c}$, there exists c_3 such that if $s \ge 2r$, then

$$\mathbb{P}^x(X_{\tau_{B(x,r)}} \notin B(x, s)) \le c_3 r^\alpha / s^\alpha. \tag{7.13}$$

Let

$$\gamma = \Big(1 - \frac{c_2}{4}\Big)^{1/2}, \qquad \rho = \frac{1}{3} \wedge \Big(\frac{\gamma}{2}\Big)^{1/\alpha} \wedge \Big(\frac{c_2 \gamma^2}{8 c_3}\Big)^{1/\alpha}.$$

By linearity and scaling it suffices to suppose $0 \le h \le M$ on $\mathbb{R}^d$ and h is $\mathcal{L}$-harmonic on $B(x,1)$. We will show

$$\sup_{B(x,\rho^k)} h - \inf_{B(x,\rho^k)} h \le M\gamma^k \tag{7.14}$$

for all k.

We write B_i for $B(x,\rho^i)$ and τ_i for $\tau_{B(x,\rho^i)}$. Let

$$a_i = \inf_{B_i} h, \qquad b_i = \sup_{B_i} h.$$

Suppose $b_i - a_i \le M\gamma^i$ for all $i \le k$; we want to show

$$b_{k+1} - a_{k+1} \le M\gamma^{k+1}. \tag{7.15}$$

We have $a_k \le h \le b_k$ on B_{k+1}. Let

$$A' = \{z \in B_{k+1} : h(z) \le (a_k + b_k)/2\}.$$

We may suppose $|A'|/|B_{k+1}| \ge \frac{1}{2}$, for if not we look at $M - h$ instead of h. Let A be a compact set contained in A' with $|A|/|B_{k+1}| \ge \frac{1}{3}$. Let $\varepsilon > 0$, pick $y \in B_{k+1}$ with $h(y) \ge b_{k+1} - \varepsilon$, and pick $z \in B_{k+1}$ with $h(z) \le a_{k+1} + \varepsilon$.

By optional stopping

$$\begin{aligned} h(y) &- h(z) \\ &= \mathbb{E}^y[h(X_{T_A}) - h(z); T_A < \tau_k] \\ &\qquad + \mathbb{E}^y[h(X_{\tau_k}) - h(z); \tau_k < T_A, X_{\tau_k} \in B_{k-1}] \\ &\qquad + \sum_{i=1}^{\infty} \mathbb{E}^y[h(X_{\tau_k}) - h(z); \tau_k < T_A, X_{\tau_k} \in B_{k-i-1} - B_{k-i}]. \end{aligned}$$

The first term on the right is bounded by

$$\Big(\frac{a_k + b_k}{2} - a_k\Big)\mathbb{P}^y(T_A < \tau_k).$$

The second term is bounded by

$$(b_{k-1} - a_{k-1})\mathbb{P}^y(\tau_k < T_A) = (b_{k-1} - a_{k-1})(1 - \mathbb{P}^y(T_A < \tau_k)).$$

Using (7.13) the infinite sum is bounded by

$$\begin{aligned}\sum_{i=1}^{\infty}&(b_{k-i-1}-a_{k-i-1})\mathbb{P}^y(X_{\tau_k}\notin B_{k-i})\\ &\leq \sum_{i=1}^{\infty} c_3 M\gamma^{k-i-1}(\rho^k)^\alpha/(\rho^{k-i})^\alpha\\ &= c_3M\gamma^{k-1}\sum_{i=1}^{\infty}(\rho^\alpha/\gamma)^i\\ &\leq 2c_3M\gamma^{k-2}\rho^\alpha\\ &\leq \frac{c_2}{4}M\gamma^k.\end{aligned}$$

Therefore

$$\begin{aligned}&h(y)-h(z)\\ &\quad\leq \frac{1}{2}(b_k-a_k)\mathbb{P}^y(T_A<\tau_k)+(b_{k-1}-a_{k-1})(1-\mathbb{P}^y(T_A<\tau_k))+c_2M\gamma^k/4\\ &\quad\leq M\gamma^k\Big(\frac{1}{\gamma}-\Big(\frac{1}{\gamma}-\frac{1}{2}\Big)\mathbb{P}^y(T_A<\tau_k)\Big)+c_2M\gamma^k/4\\ &\quad\leq M\gamma^k\Big(\frac{1}{\gamma}-\Big(\frac{1}{\gamma}-\frac{1}{2}\Big)c_2\Big)+c_2M\gamma^k/4.\end{aligned}$$

Since $\gamma<1$ and

$$1-c_2+\frac{3c_2\gamma}{4}\leq 1-\frac{c_2}{4}=\gamma^2,$$

then

$$\frac{1}{\gamma}-\Big(\frac{1}{\gamma}-\frac{1}{2}\Big)c_2+\frac{c_2}{4}\leq\gamma,$$

and so

$$h(y)-h(z)\leq M\gamma^{k+1}.$$

We conclude that

$$b_{k+1}-a_{k+1}\leq M\gamma^{k+1}+2\varepsilon.$$

Since ε is arbitrary, this proves (7.15) and hence (7.14).

If $x,y\in B(x_0,1)$, let k be the smallest integer such that $|x-y|<\rho^k$. Then $\log|x-y|\geq(k+1)\log\rho$, $y\in B(x,\rho^k)$, and

$$\begin{aligned}|h(y)-h(x)|&\leq M\gamma^k=Me^{k\log\gamma}\\ &\qquad\leq c_4Me^{\log|x-y|(\log\gamma/\log\rho)}=c_4M|x-y|^{\log\gamma/\log\rho}.\end{aligned}$$

□

Define

$$S_\lambda g(x) = \mathbb{E}^x \int_0^\infty e^{-\lambda t} g(X_t)dt.$$

Proposition 7.14 *Suppose g is bounded and has compact support. There exists $c_1 > 2$ and $\beta \in (0,1)$ such that*

$$|S_0 g(x) - S_0 g(y)| \le c_1(\|S_0 g\|_\infty + \|g\|_\infty)(|x-y| \wedge 1)^\beta.$$

Proof. Suppose $|x-y| \le 1$, for otherwise there is nothing to prove. We write

$$S_0 g(x) = \mathbb{E}^x \int_0^{\tau_{B(x,r)}} g(X_s)ds + \mathbb{E}^x S_0 g(X_{\tau_{B(x,r)}})$$

and

$$S_0 g(y) = \mathbb{E}^y \int_0^{\tau_{B(x,r)}} g(X_s)ds + \mathbb{E}^y S_0 g(X_{\tau_{B(x,r)}}).$$

Taking the difference,

$$|S_0 g(x) - S_0 g(y)| \le 2\|g\|_\infty \sup_z \mathbb{E}^z \tau_{B(x,r)} + c_2 \|S_0 g\|_\infty \Big(\frac{|x-y|}{r}\Big)^\beta,$$

using the previous theorem, scaling, and the fact that $z \to \mathbb{E}^z S_0 g(X_{\tau_{B(x,r)}})$ is $\mathcal{L}$-harmonic inside $B(x,r)$. Taking $r = |x-y|^{1/2}$ and using a lemma, we obtain our result. □

Theorem 7.15 *Suppose g is bounded and $\lambda > 0$. There exists $c_1 > 0$ and $\beta \in (0,1)$ such that*

$$|S_\lambda g(x) - S_\lambda g(y)| \le c_1 \|g\|_\infty (|x-y| \wedge 1)^\beta.$$

Proof. Without loss of generality assume $g \ge 0$. Temporarily assume g has compact support. Let $h = g - \lambda S_\lambda g$. Note $S_0 h \le S_0 g + \lambda S_\lambda S_0 g$, so h is bounded. We have $S_\lambda g = S_0 h$ by the resolvent equation. Since $\|S_\lambda g\|_\infty \le c_2 \|g\|_\infty$, then $\|S_0 h\|_\infty + \|h\|_\infty \le c_3 \|g\|_\infty$. Our result now follows by a proposition if g has compact support. Taking limits allows us to remove this restriction. $\square$

The solution to the integral equation

$$\mathcal{L}u(x) - \lambda u(x) = -g(x)$$

is given by $u(x) = S_\lambda g(x)$. So our theorem provides a regularity result for the solutions of such integral equations.

Chapter 8

Symmetric jump processes

8.1 Dirichlet forms

Let us now suppose $\mathcal{S}$ is a locally compact separable metric space together with a σ-finite measure m defined on the Borel subsets of $\mathcal{S}$. We want to give a definition of Dirichlet form in this more general context. We suppose there exists a dense subset $\mathcal{D} = \mathcal{D}(\mathcal{E})$ of $L^2(S, m)$ and a non-negative bilinear symmetric form $\mathcal{E}$ defined on $\mathcal{D} \times \mathcal{D}$, which means

$$\begin{aligned} \mathcal{E}(f,g) &= \mathcal{E}(g,f), \quad \mathcal{E}(f+g,h) = \mathcal{E}(f,h) + \mathcal{E}(g,h) \\ \mathcal{E}(af,g) &= a\mathcal{E}(f,g), \quad \mathcal{E}(f,f) \ge 0 \end{aligned}$$

for $f, g, h \in \mathcal{D}$, $a \in \mathbb{R}$.

We will frequently write $\langle f, g\rangle$ for $\int f(x)g(x)\, m(dx)$. For $a > 0$ define

$$\mathcal{E}_a(f,f) = \mathcal{E}(f,f) + a\langle f, f\rangle.$$

We can define a norm on $\mathcal{D}$ using the inner product $\mathcal{E}_a$: the norm of f equals $(\mathcal{E}_a(f,f))^{1/2}$; we call this the norm induced by $\mathcal{E}_a$. Since $a\langle f, f\rangle \le \mathcal{E}_a(f,f)$, then

$$\begin{aligned} \mathcal{E}_a(f,f) \le \mathcal{E}_b(f,f) &= \mathcal{E}_a(f,f) + (b-a)\langle f, f\rangle \\ &\le \Big(1 + \frac{b-a}{a}\Big)\mathcal{E}_a(f,f) \end{aligned}$$

if $a < b$, so the norms induced by different a's are all equivalent. We say $\mathcal{E}$ is *closed* if $\mathcal{D}$ is complete with respect to the norm induced by $\mathcal{E}_a$ for some a. Equivalently, $\mathcal{E}$ is closed if whenever $u_n \in \mathcal{D}$ satisfies $\mathcal{E}_1(u_n - u_m, u_n - u_m) \to 0$ as $n, m \to \infty$, then there exists $u \in \mathcal{D}$ such that $\mathcal{E}(u_n - u, u_n - u) \to 0$ as $n \to \infty$.

We say $\mathcal{E}$ is *Markovian* if whenever $u \in \mathcal{D}$, then $v = 0 \vee (u \wedge 1) \in \mathcal{D}$ and $\mathcal{E}(v, v) \le \mathcal{E}(u, u)$. (A slightly weaker definition of Markovian is sometimes used.) A *Dirichlet form* is a non-negative bilinear symmetric form that is closed and Markovian.

Absorbing Brownian motion on $[0, \infty)$ is a symmetric process. The corresponding Dirichlet form is

$$\mathcal{E}(f, f) = \tfrac{1}{2} \int_0^\infty |f'(x)|^2 \, dx,$$

and the appropriate domain turns out to be the completion of the set of C^1 functions with compact support contained in $(0, \infty)$ with respect to the norm induced by $\mathcal{E}_1$. In particular, any function with compact support contained in $(0, \infty)$ will be zero in a neighborhood of 0. In a domain D in higher dimensions, the Dirichlet form for absorbing Brownian motion becomes

$$\mathcal{E}(f, f) = \tfrac{1}{2} \int |\nabla f(x)|^2 \, dx, \tag{8.1}$$

with the domain of $\mathcal{E}$ being the completion with respect to $\mathcal{E}_1$ of the C^1 functions whose support is contained in the interior of D.

Reflecting Brownian motion is also a symmetric process. For a domain D, the Dirichlet form is given by (8.1) and the domain $\mathcal{D}(\mathcal{E})$ of the form is given by the completion with respect to the norm induced by $\mathcal{E}_1$ of the C^1 functions on $\overline{D}$ with compact support, where $\overline{D}$ is the closure of D. One might expect there to be some restriction on the normal derivative $\partial f/\partial n$ on the boundary of D, but in fact there is no such restriction. To examine this further, consider the case of $D = (0, \infty)$. If one takes the class of functions f which are C^1 with compact support and with $f'(0) = 0$ and takes the closure with respect to the norm induced by $\mathcal{E}_1$, one can show that one gets the same class as $\mathcal{D}(\mathcal{E})$.

One nice consequence of the fact that one doesn't need to impose a restriction on the normal derivative in the domain of $\mathcal{E}$ for reflecting Brownian motion is that this allows us to define reflecting Brownian motion in any domain, even when the boundary is not smooth enough for the notion of normal derivative to be defined.

8.2 Construction of the semigroup

We now want to construct the resolvent corresponding to a Dirichlet form. We are going to arrange things so that

$$\mathcal{E}_a(R_a f, g) = \langle f, g \rangle \tag{8.2}$$

for all $a > 0$ and all f, g such that $R_a f, g \in \mathcal{D}$. Our Banach space $\mathcal{B}$ will be $L^2(\mathcal{S}, m)$. An operator T is symmetric if $\langle Tf, g\rangle = \langle f, Tg\rangle$.

Recall the Hille-Yosida theorem says that if $\mathcal{L}$ is a densely defined unbounded operator such that $R_\lambda = (\lambda I - \mathcal{L})^{-1}$ exists for all real $\lambda > 0$ and $\|R_\lambda\| \le 1/\lambda$, then $\mathcal{L}$ is the infinitesimal generator of a strongly continuous semigroup of contractions whose resolvents are R_λ. Strongly continuous means that $P_t f \to f$ in norm as $t \to 0$ for every f in the Banach space, contraction means $\|P_t\| \le 1$, and saying R_λ is the resolvent means that $R_\lambda f = \int_0^\infty e^{-\lambda t} P_t f\, dt$ for all f.

An alternate phrasing is that if R_λ, $\lambda > 0$ is a collection of bounded operators such that the resolvent identity holds, that is, $R_a - R_b = (b-a)R_a R_b$, we have $\|R_\lambda\| \le 1/\lambda$, and the range of R_λ is dense in the Banach space, then there exists a strongly continuous semigroup of contractions whose resolvents are R_λ.

Theorem 8.1 *If $\mathcal{E}$ is a Dirichlet form, there exists a family of resolvent operators $\{R_\lambda\}$ such that*
(1) the R_λ satisfy the resolvent equation,
(2) $\|\lambda R_\lambda\| \le 1$ for all $\lambda > 0$,
(3) $\lambda R_\lambda f \to f$ as $\lambda \to \infty$,
(4) $\mathcal{E}_a(R_a f, g) = \langle f, g\rangle$ if $a > 0$, $R_a f, g \in \mathcal{D}$,
(5) R_a is a symmetric operator if $a > 0$,
(6) every function in the domain of the infinitesimal generator $\mathcal{L}$ is in $R_\lambda(\mathcal{B})$, where $\mathcal{B}$ is the Banach space,
(7) $\langle \mathcal{L} f, g\rangle = -\mathcal{E}(f, g)$ if f is in the domain of $\mathcal{L}$ and $g \in \mathcal{D}(\mathcal{E})$.

Proof. Fix $f \in \mathcal{B}$ and define a linear functional on $\mathcal{B}$ by $I(g) = \langle f, g\rangle$. This functional is also a bounded linear functional on $\mathcal{D}$ with respect to the norm induced by $\mathcal{E}_a$, that is, there exists c such that $|I(g)| \le c\mathcal{E}_a(g,g)^{1/2}$. This follows because

$$|I(g)| = \Big|\int fg\Big| \le \langle f,f\rangle^{1/2}\langle g,g\rangle^{1/2} \le \langle f,f\rangle^{1/2}(\tfrac{1}{a}\mathcal{E}_a(g,g))^{1/2}$$

by the Cauchy-Schwarz inequality. Since $\mathcal{E}$ is closed, $\mathcal{D}$ is a Hilbert space with respect to the norm induced by $\mathcal{E}_a$. By the Riesz representation theorem for Hilbert spaces, there exists a unique element $u \in \mathcal{D}$ such that $I(g) = \mathcal{E}_a(u,g)$ for all $g \in \mathcal{D}$. We set $R_a f = u$. In particular, (8.2) holds, and $R_a f \in \mathcal{D}$.

We show the resolvent equation holds. If $g \in \mathcal{D}$,

$$\begin{aligned}
\mathcal{E}_a(R_a f - R_b f, g) &= \mathcal{E}_a(R_a f, g) - \mathcal{E}(R_b f, g) - a\langle R_b f, g\rangle \\
&= \langle f,g\rangle - \mathcal{E}(R_b f, g) - b\langle R_b f, g\rangle + (b-a)\langle R_b f, g\rangle \\
&= \langle f,g\rangle - \mathcal{E}_b(R_b f, g) + (b-a)\langle R_b f, g\rangle \\
&= (b-a)\langle R_b f, g\rangle \\
&= \mathcal{E}_a((b-a)R_a R_b f, g).
\end{aligned}$$

Since this holds for all $g \in \mathcal{D}$ and $\mathcal{D}$ is dense in $\mathcal{B}$, then $R_a f - R_b f = (b-a)R_a R_b f$.

Next we show that $\|aR_a f\| \le \|f\|$, or equivalently,

$$\langle aR_a f, aR_a f\rangle \le \langle f, f\rangle. \tag{8.3}$$

If $\langle R_a f, R_a f\rangle$ is zero, then (8.3) trivially holds, so suppose it is positive. We have

$$a\langle R_a f, R_a f\rangle \le \mathcal{E}_a(R_a f, R_a f) = \langle f, R_a f\rangle \le \langle f,f\rangle^{1/2}\langle R_a f, R_a f\rangle^{1/2}$$

by (8.2) and the Cauchy-Schwarz inequality. If we now divide both sides by $\langle R_a f, R_a f\rangle^{1/2}$ and then square both sides, we obtain (8.3).

We show that $bR_b f \to f$ as $b \to \infty$ when $f \in \mathcal{B}$. If $f \in \mathcal{D}$, then by the Cauchy-Schwarz inequality and (8.3)

$$\begin{aligned}
\langle bR_b f, f\rangle &\le \langle bR_b f, bR_b f\rangle^{1/2}\langle f,f\rangle^{1/2} \\
&\le \langle f, f\rangle.
\end{aligned}$$

Using this,

$$\begin{aligned}
b\langle bR_bf - f, bR_bf - f\rangle &\le \mathcal{E}_b(bR_bf - f, bR_bf - f)\\
&= b^2\mathcal{E}_b(R_bf, R_bf) - 2b\mathcal{E}_b(R_bf, f) + \mathcal{E}_b(f,f)\\
&= b^2\langle R_bf, f\rangle - 2b\langle f, f\rangle + \mathcal{E}(f,f) + b\langle f, f\rangle\\
&\le \mathcal{E}(f,f).
\end{aligned}$$

Now divide both sides by b to get $\|bR_bf - f\|^2 \le \mathcal{E}(f,f)/b \to 0$ as $b \to \infty$. Since $\mathcal{D}$ is dense in $\mathcal{B}$ and $\|bR_b\| \le 1$ for all b, we conclude $bR_bf \to f$ for all $f \in \mathcal{B}$.

To see the symmetry, we have

$$\langle f, R_ag\rangle = \mathcal{E}_a(R_af, R_ag) = \mathcal{E}_a(R_ag, R_af) = \langle g, R_af\rangle.$$

If f is in the domain of the infinitesimal generator, then $h = \mathcal{L}f$ is in the Banach space, and hence $\lambda f - h = (\lambda - \mathcal{L})f$ is in the Banach space $\mathcal{B}$, which implies $h = R_\lambda(\lambda f - h)$. Thus every function in the domain of the infinitesimal generator is in $R_\lambda(\mathcal{B})$.

If $f = R_\lambda h$, then $\mathcal{L}f = \lambda R_\lambda h - h$, and then

$$\langle \mathcal{L}f, g\rangle = \lambda\langle R_\lambda h, g\rangle - \langle h, g\rangle.$$

We have

$$\langle h, g\rangle = \mathcal{E}_\lambda(R_\lambda h, g) = \mathcal{E}(R_\lambda h, g) + \lambda\langle R_\lambda h, g\rangle.$$

Solving,

$$\mathcal{E}(f,g) = \mathcal{E}(R_\lambda h, g) = \langle h, g\rangle - \lambda\langle R_\lambda h, g\rangle.$$

□

Theorem 8.2 *If $f \in \mathcal{B}$ satisfies $0 \le f(x) \le 1$, m-a.e., then for all $a > 0$*

$$0 \le aR_af \le 1, \qquad m - a.e. \tag{8.4}$$

Proof. Fix $f \in \mathcal{B}$ with $0 \le f \le 1$, m-a.e., and let $a > 0$. Define a functional ψ on $\mathcal{D}$ by

$$\psi(v) = \mathcal{E}(v,v) + a\Big\langle v - \frac{f}{a}, v - \frac{f}{a}\Big\rangle.$$

We claim

$$\psi(R_a f) + \mathcal{E}_a(R_a f - v, R_a f - v) = \psi(v), \qquad v \in \mathcal{D}. \tag{8.5}$$

To see this, start with the left hand side, which is equal to

$$\begin{aligned}
\mathcal{E}(R_a f, R_a f) &+ a\Big\langle R_a f - \frac{1}{a} f, R_a f - \frac{1}{a} f\Big\rangle + \mathcal{E}_a(R_a f - v, R_a f - v)\\
&= \mathcal{E}_a(R_a f, R_a f) - 2\langle R_a f, f\rangle + \frac{1}{a}\langle f, f\rangle + \mathcal{E}_a(R_a f, R_a f)\\
&\qquad - 2\mathcal{E}_a(R_a f, v) + \mathcal{E}_a(v, v)\\
&= \frac{1}{a}\langle f, f\rangle - 2\langle f, v\rangle + \mathcal{E}(v, v) + a\langle v, v\rangle\\
&= \psi(v).
\end{aligned}$$

If follows from (8.5) and the fact that $\mathcal{E}_a(g, g)$ is non-negative for any $g \in \mathcal{D}$ that $R_a f$ is the function that minimizes ψ.

Set $\phi(x) = 0 \vee (x \wedge (1/a))$ and let $w = \phi(R_a f)$. Observe that $|\phi(t) - s| \leq |t - s|$ for $t \in \mathbb{R}$ and $s \in [0, 1/a]$, so

$$\Big|w(x) - \frac{f(x)}{a}\Big| \leq \Big|R_a f(x) - \frac{f(x)}{a}\Big|,$$

and therefore

$$\Big\langle w - \frac{f}{a}, w - \frac{f}{a}\Big\rangle \leq \Big\langle R_a f - \frac{f}{a}, R_a f - \frac{f}{a}\Big\rangle. \tag{8.6}$$

Since $\mathcal{E}$ is Markovian, then $aw = 0 \vee ((aR_a f) \wedge 1)$, which leads to

$$\mathcal{E}(w, w) \leq \frac{1}{a^2}\mathcal{E}(aR_a f, aR_a f) = \mathcal{E}(R_a f, R_a f). \tag{8.7}$$

Adding (8.6) and (8.7), we conclude $\psi(w) \leq \psi(R_a f)$. Since $R_a f$ is the minimizer for ψ, then $w = R_a f$, m-a.e. But $0 \leq w \leq 1/a$, and hence $aR_a f$ takes values in $[0, 1]$, m-a.e. □

Corollary 8.3 *(1) If $0 \leq f \leq 1$, m-a.e., then $0 \leq P_t f \leq 1$, m-a.e.*
(2) P_t is a symmetric operator.

Proof. If $0 \le f \le 1$, m-a.e., then $0 \le bR_b f \le 1$, m-a.e, by Theorem 8.1, and iterating, $0 \le (bR_b)^i f \le 1$, m-a.e., for every i. Using the proof of the Hille-Yosida theorem,

$$Q_t^b f(x) = e^{-bt} \sum_{i=0}^{\infty} (bt)^i (bR_b)^i f(x)/i!,$$

which will be non-negative, m-a.e., and bounded by $e^{-bt} \sum_{i=0}^{\infty} (bt)^i / i!$, m- a.e. Passing to the limit as $b \to \infty$, we see that $P_t f$ takes values in $[0,1]$, m-a.e.

The proof of the symmetry of P_t is similar. □

When it comes to using the semigroup P_t derived from a Dirichlet form to construct a Markov process X, there is a difficulty that we did not have before. Since P_t is constructed using an L^2 procedure, $P_t f$ is defined only up to almost everywhere equivalence. Without some continuity properties of $P_t f$ for enough f's, we must neglect some null sets. If the only null sets we could work with were sets of m-measure 0, we would be in trouble. For example, when $\mathcal{S}$ is the plane and m is two-dimensional Lebesgue measure, the x axis has measure zero, but a continuous process will (in general) hit the x axis. Fortunately there is a notion of sets of capacity zero, which are null sets that are smaller than sets of measure zero. It is possible to construct a process X starting from all points x in $\mathcal{S}$ except for those in a set $\mathcal{N}$ of capacity zero and to show that starting from any point not in $\mathcal{N}$, the process never hits $\mathcal{N}$.

There is another difficulty when working with Dirichlet forms. In general, one must look at $\widetilde{\mathcal{S}}$, a certain compactification of $\mathcal{S}$, which is a compact set containing $\mathcal{S}$. Even when our state space is a domain in $\mathbb{R}^d$, $\widetilde{\mathcal{S}}$ is not necessarily equal to $\overline{\mathcal{S}}$, the Euclidean closure of $\mathcal{S}$, and one must work with $\widetilde{\mathcal{S}}$ instead of $\overline{\mathcal{S}}$. It can be shown that this problem will not occur if the Dirichlet form is regular. Let C_K be the set of continuous functions with compact support. A Dirichlet form $\mathcal{E}$ is *regular* if $\mathcal{D} \cap C_K$ is dense in $\mathcal{D}$ with respect to the norm induced by $\mathcal{E}_1$ and $\mathcal{D} \cap C_K$ also is dense in C_K with respect to the supremum norm.

8.3 Symmetric jump processes

We are going to define

$$\mathcal{E}(f,g) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{[f(y)-f(x)]\,[g(y)-g(x)]}{|x-y|^{d+\alpha}} A(x,y)\, dy\, dx,$$

where A is symmetric and bounded above and below by positive constants, but first we need to specify the domain.

Define

$$\nu(f) = \|f\|_2 + \Big(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{[f(y)-f(x)]^2}{|x-y|^{d+\alpha}}\, dy\, dBig)^{1/2}.$$

Let us show that if $f \in C^2$ with compact support, say in $B(0,M)$ with $M \ge 2$, then $\nu(f)$ is finite. Let $|A|$ be the Lebesgue measure of a set A.

Fix x and first suppose $|x| \ge 3M$. The numerator of the second integral in the definition of $\nu(f)$ is 0 unless $y \in B(0,M)$, and in that case $|x-y|^{d+\alpha} \le c|x|^{d+\alpha}$. So for $|x| \ge 3M$ we can bound the inside integral by

$$c^{-1} \int_{|y|\le M} \frac{|f(y)|}{|x|^{d+\alpha}}\, dy \le c' \|f\|_\infty |B(0,M)| |x|^{-d-\alpha}.$$

This is integrable over $|x| \ge 3M$.

Now suppose $|x| \le 3M$. Since $f \in C^2$,

$$\begin{aligned}\int_{\mathbb{R}^d} \frac{|f(y)-f(x)|^2}{|x-y|^{d+\alpha}}\, dy \le \|f'\|_\infty^2 &\int_{|x-y|\le 1} \frac{|x-y|^2}{|x-y|^{d+\alpha}}\, dy \\ &+ 2\|f\|_\infty^2 \int_{|x-y|>1} \frac{1}{|x-y|^{d+\alpha}}\, dy,\end{aligned}$$

which is bounded by a constant depending on f. So the integral over $B(0,3M)$ is finite. Therefore $\nu(f)$ is finite.

We now let D be the completion of C^2_K, the C^2 functions with compact support, with respect to the norm $\nu(f)$.

We suppose $A(x,y) = A(y,x)$ for all x and y and

$$0 < c_1 \le A(x,y) \le c_2 < \infty$$

for constants c_1, c_2. That $\mathcal{E}$ is bilinear, symmetric, and $\mathcal{E}(f,f) \geq 0$ is obvious. That $\mathcal{E}$ is closed follows easily because $\mathcal{E}_\alpha(f,f)$ is comparable to $\nu(f)$.

It remains to show $\mathcal{E}$ is Markovian. If $g = (f \wedge 1) \vee 0$, note that

$$|g(y) - g(x)| \leq |f(y) - f(x)|.$$

It then follows easily that $\mathcal{E}(g,g) \leq \mathcal{E}(f,f)$.

We note that $\mathcal{E}$ is regular by the way the domain of $\mathcal{E}$ was constructed.

8.4 The Poincaré and Nash inequalities

Let $Q = Q_h = [-h/2, h/2]^d$. Define

$$f_Q = \frac{1}{|Q|} \int_Q f(y)\,dy.$$

Theorem 8.4 *(Poincaré inequality) There exists a constant c such that*

$$\int_Q |f(y) - f_Q|^2\,dy \leq ch^\alpha \int_Q \int_Q \frac{(f(y) - f(x))^2}{|x-y|^{d+\alpha}}\,dy\,dx.$$

Proof. Let's first do the case $h = 1$. We write

$$\begin{aligned}\int_Q \int_Q [f(y) - f(x)]^2\,dy\,dx &= \int_Q \int_Q [f(y)^2 - 2f(x)f(y) + f(x)^2]\,dy\,dx \\ &= 2\int_Q f(x)^2\,dx - 2f_Q f_Q = 2\int_Q (f - f_Q)^2.\end{aligned}$$

Since $|x-y|^{d+\alpha} \geq c$ on $Q \times Q$, the left hand side is less than or equal to

$$c \int_Q \int_Q \frac{(f(y) - f(x))^2}{|x-y|^{d+\alpha}}\,dy\,dx.$$

The case of general h is done by a scaling argument (apply the above to $g(x) = f(x/h)$ and use a change of variables). $\square$

Before proving the Nash inequality, observe that

$$|f_Q| = \frac{1}{|Q|}\Big|\int_Q f\Big| \le h^{-d}\int_Q |f|.$$

We will also use the inequality that if all the $a_i \ge 0$, then

$$\sum a_i^2 \le \Big(\sum a_i\Big)^2.$$

Theorem 8.5 *(Nash inequality)*

$$\|f\|_2^{2(1+\alpha/d)} \le c\mathcal{E}(f,f)\|f\|_1^{2\alpha/d}.$$

Proof. Break $\mathbb{R}^d$ into cubes $Q_1, Q_2, \ldots$ whose union is $\mathbb{R}^d$ and whose interiors are pairwise disjoint. Note

$$\frac{1}{|Q_i|}\int_{Q_i}(f(x)-f_{Q_i})^2 = \frac{1}{|Q_i|}\int_{Q_i} f(x)^2\,dx - (f_{Q_i})^2.$$

Then

$$\begin{aligned}
\|f\|_2^2 &= \int f(x)^2\,dx = \sum_i \int_{Q_i} f(x)^2\\
&= \sum h^d \frac{1}{|Q_i|}\int_{Q_i} f(x)^2\,dx\\
&= \sum h^d \frac{1}{|Q_i|}\int_{Q_i} |f(x)-f_{Q_i}|^2\,dx + \sum h^d f_{Q_i}^2\\
&\le \sum_i h^\alpha \int_{Q_i}\int_{Q_i} \frac{f(y)-f(x))^2}{|x-y|^{d+\alpha}}\,dy\,dx + h^d\Big(\sum |f|_{Q_i}\Big)^2\\
&\le ch^\alpha \mathcal{E}(f,f) + h^{-d}\|f\|_1^2.
\end{aligned}$$

Now choose h so that the two terms on the last line are equal, namely,

$$h = \Big(\frac{\|f\|_1^2}{\mathcal{E}(f,f)}\Big)^{1/(\alpha+d)},$$

and we obtain

$$\|f\|_2^2 \le c\|f\|_1^{2\alpha/(\alpha+d)}\mathcal{E}(f,f)^{d/(\alpha+d)}.$$

Taking both sides to the power $(\alpha+d)/d$ gives the inequality. $\square$

8.5 Upper bounds on the transition densities

Notice

$$|P_t f(x)| = \Big|\int f(y) P_t(x, dy)\Big| \le \int |f(y)| P_t(x, dy) = P_t|f|(x).$$

Since P_t is symmetric, we have

$$\|P_t f\|_1 = \langle |P_t f|, 1\rangle \le \langle P_t|f|, 1\rangle = \langle |f|, P_t 1\rangle = \langle |f|, 1\rangle = \|f\|_1.$$

Theorem 8.6 *There exists a function $p(t,x,y)$ such that*

$$P_t f(x) = \int p(t,x,y) f(y)\, dy$$

(for almost every x) and such that $p(t,x,y) \le ct^{-d/\alpha}$.

Proof. We will show

$$\|P_t f\|_2 \le ct^{-d/2\alpha}\|f\|_1$$

for f in the domain of the infinitesimal generator. By taking limits, this holds for all $f \in L^1 \cap L^2$. Taking $f = 1_A$ where $|A| = 0$, we get that $\|P_t f\|_2 = 0$, or $P_t f = 0$ a.e. So

$$\int_A P_t(x, dy) = 0,$$

or $P_t(x, dy)$ is absolutely continuous with respect to Lebesgue measure. Thus $p(t,x,y)$, the Radon-Nikodym derivative exists (for almost every x).

Also, if $g \in L^1$,

$$|\langle P_t f, g\rangle| = |\langle f, P_t g\rangle| \le \|f\|_2 \|P_t g\|_2 \le ct^{-d/2\alpha}\|f\|_2\|g\|_1.$$

Taking the supremum over the set of $g \in L^1 \cap L^2$ with $\|g\|_1 \le 1$, we obtain

$$\|P_t f\|_\infty \le ct^{-d/2\alpha}\|f\|_2.$$

Then

$$\|P_t f\|_\infty = \|P_{t/2}(P_{t/2} f)\|_\infty \le ct^{-d/2\alpha}\|P_{t/2} f\|_2 \le ct^{-d/\alpha}\|f\|_1.$$

Letting $f = 1_A$, we have

$$\int_A p(t,x,y)\,dy \le ct^{-d/\alpha}|A|,$$

and the bound on $p(t,x,y)$ follows.

Now let $f \in C_K^2$. f is in the domain of the infinitesimal generator, so $P_t f$ is also. Suppose $\|f\|_1 = 1$, so that $\|P_t f\|_1 \le 1$. Let

$$E(t) = \int P_t f(x)^2\,dx.$$

Using the Nash inequality and the fact that $\mathcal{E}(f,f)$, which has an $A(x,y)$ term, is comparable to

$$\int\int \frac{(f(y)-f(x))^2}{|x-y|^{d+\alpha}}\,dy\,dx,$$

we have

$$\begin{aligned} E'(t) &= \int 2P_t f(x)\frac{\partial}{\partial t}P_t f(x)\,dx = 2\int P_t f(x) L P_t f(x)\,dx \\ &= -2\mathcal{E}(P_t f, P_t f) \le -c\int\int \frac{(P_t f(y)-P_t f(x))^2}{|x-y|^{d+\alpha}}\,dy\,dx \\ &\le -2c\|P_t f\|_2^{2(1+\alpha/d)} = -cE(t)^{1+\alpha/d}. \end{aligned}$$

So

$$\frac{E'(t)}{E(t)^{1+\alpha/d}} \le -c,$$

hence

$$-E(t)^{-\alpha/d} \le -E(t)^{-\alpha/d} + E(0)^{-\alpha/d} \le -ct,$$

and therefore

$$E(t) \le ct^{-d/\alpha}.$$

By linearity,

$$\|P_t f\|_2^2 \le ct^{-d/\alpha}\|f\|_1^2.$$

□