C8.2: Stochastic analysis and PDEs

Alison Etheridge
(with minor modifications by Ilya Chevyrev)

February 15, 2020

Contents

0 Introduction

0.1 Background

In this course we are going to explore topics that sit at the rich interface between probability and analysis and, especially, those that link the two areas through partial differential equations. To some extent, it is just a collection of things that I think you ought to know, ranging from powerful abstract results to tools that allow you to do hands on calculations,

We begin by recasting what you already know about continuous time Markov processes in the language of operator semigroups. We then develop the notion of martingale problem which provides us with a new way of characterising stochastic processes. This is particularly well suited to proving convergence theorems and we shall see how to use this framework to prove (weak) convergence of a sequence of stochastic processes. This is an important tool in modelling where it is often convenient to approximate a process taking discrete values by a continuous one. En route we shall encounter the method of duality, in which the distribution of one stochastic process is expressed in terms of that of another.

Each diffusion process can be associated with a second order differential operator, and through this operator the theory of diffusion processes is intricately connected with that of second order parabolic (and elliptic) differential equations. For example, if one can find the fundamental solution to the parabolic equation, then one can answer essentially any question about the finite-dimensional distributions of the stochastic process. However, in practice, it is rarely possible to find an explicit form for the fundamental solution. Nonetheless, in one spatial dimension a very large number of explicit calculations can be carried out through the theory of speed and scale. This theory uncovers and exploits the fundamental rôle played by Brownian motion as a building block from which to construct other diffusion processes.

Finally we shall return to general spatial dimensions and explore a menagerie of equations whose solutions can be expressed in terms of the law of a diffusion process which not only allows us to prove results about the diffusion, but also paves the way for proving results about the deterministic equations through their stochastic representation.

We will not work in the greatest possible generality, because it is too easy for simple ideas to be obscured by technical details. More general results can certainly be found in the references.

0.2 Recommended Reading

These notes and the accompanying problem sheets essentially define the course, but there are a huge number of books to which you can turn for supplementary reading. A few are listed below.

Important references:

(i) O. Kallenberg. *Foundations of Modern Probability.* Second Edition, Springer 2002. *This comprehensive text covers essentially the entire course, and much more, but should be supplemented with other references in order to develop experience of more examples.*

(ii) L.C.G Rogers & D. Williams. *Diffusions, Markov Processes and Martingales; Volume 1, Foundations* and *Volume 2, Itô calculus.* Cambridge University Press, 1987 and 1994. *These two volumes have a very different style to Kallenberg and complement it nicely. Again they cover much more material than this course.*

Supplementary reading:

(i) S.N. Ethier & T.G. Kurtz. *Markov Processes: characterization and convergence.* Wiley 1986. *It is not recommended to try to sit down and read this book cover to cover, but it is a treasure trove of powerful theory and elegant examples.*

(ii) S. Karlin & H.M. Taylor. *A second course in stochastic processes.* Academic Press 1981. *This classic text does not cover the material on semigroups and martingale problems that we shall develop, but it is a very accessible source of examples of diffusions and things one might calculate for them.*

Also highly recommended (in no particular order) are:

- D.W. Stroock & S.R.S. Varadhan, *Multidimensional diffusion processes,* Springer 1979. *This is an excellent place to learn about martingale problems - or more generally Stroock-Varadhan theory.*
- K. Yosida, *Functional Analysis,* Springer 1980.*A comprehensive source for material leading up to the Hille-Yosida Theorem.*
- K.J. Engel & R. Nagel, *A short course on operator semigroups*, Springer 2005. *A very accessible account of the Hille-Yosida Theorem. A longer book by the same authors is a good source of examples.*
- R.F. Bass, *Diffusions and elliptic operators,* Springer 1997.*Covers all the material on stochastic representation of solutions to differential equations which makes up the second half of the course. Also has a nice revision of stochastic differential equations if you need it.*
- D. Revuz & M. Yor. *Continuous martingales and Brownian motion.* Springer 1999. *An extremely thorough reference for the material on continuous parameter martingales.*
- T.M. Liggett, *Interacting Particle Systems,* Springer 1985. *We won't actually discuss interacting particle systems, but the very accessible introductions to the Hille-Yosida Theorem and to martingale problems form the basis of our approach to those topics in these the notes.*
- R. Durrett. *Stochastic calculus: a practical introduction.* CRC Press 1996. *A lot of relevant material, presented in a user-friendly way.*

If you feel that you need to revise the background functional analysis, then try

- S. Lang, *Real and Functional Analysis.* Springer 1978. *Offers a detailed treatment of calculus in Banach spaces.*
- E. Kreyszig, *Introductory functional analysis with applications.* Wiley 1978.

1 Review of some measure-theoretic probability

The purpose of this section is to refresh your memory and establish some notation. We will tend to suppress the measure-theoretic niceties in future sections. Those wanting more detail should consult Rogers & Williams or Revuz & Yor. I hope that you have seen most of this material before. If not, try reading the online notes for the Part B discrete parameter martingales course to get you started on the basic ideas and then look at, for example, Rogers & Williams to understand some of the subtleties when we move to continuous time.

This section of the notes draws heavily on Chapter II of Volume 1 of Rogers & Williams.

Definition 1.1 (Probability triple, random variable, law)**.** *(i) A* probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ *consists of a sample space Ω; a σ-algebra, $\mathcal{F}$, on that sample space; and a probability measure $\mathbb{P}$ on $\mathcal{F}$.*

(ii) Given a measure space $(E, \mathcal{E})$, an $(E, \mathcal{E})$-valued random variable *with carrier triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a measurable mapping X from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$.*

(iii) By the law μ_X of X, we mean the probability measure $\mu_X = \mathbb{P} \circ X^{-1}$ on $(E, \mathcal{E})$, so that

$$\mu_X(A) = \mathbb{P}[X \in A] = \mathbb{P}[\{\omega \in \Omega : X(\omega) \in A\}].$$

We shall often be interested in random variables taking values in Euclidean space, in which case $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ with $\mathcal{B}(\mathbb{R}^d)$ the Borel subsets of $\mathbb{R}^d$.

Definition 1.2 (σ-algebra generated by a random variable)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be an E-random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, that is a measurable function from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$. Then*

$$\begin{aligned} \sigma(X) &= \sigma\left(\{\{\omega \in \Omega : X(\omega) \in A\}; A \in \mathcal{E}\}\right) \\ &= \sigma\left(\left\{X^{-1}(A) : A \in \mathcal{E}\right\}\right). \end{aligned}$$

It is the smallest sub σ-algebra of $\mathcal{F}$ with respect to which X is a measurable function.

Definition 1.3 (Stochastic process)**.** *Let T be a set, $(E, \mathcal{E})$ a measure-space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability triple. A* stochastic process *with time-parameter set T, state space $(E, \mathcal{E})$ and carrier triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection $\{X_t\}_{t \in T}$ of $(E, \mathcal{E})$-valued random variables, carried by the triple $(\Omega, \mathcal{F}, \mathbb{P})$.*

Often we shall abuse notation and write X for $\{X_t\}_{t \in T}$.

Definition 1.4 (Sample path, realisation)**.** *Let X be a stochastic process. For $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ from T to E is called the* sample path *or* realisation *of $\{X_t\}_{t \in T}$ corresponding to ω.*

As things stand, we have made no regularity assumptions on our stochastic process, the mapping $X(\omega) : T \to E$ can be any function. For most of what follows, we shall take $(E, \mathcal{E})$ to be Euclidean space (or a subset thereof) and the time set T to be be $\mathbb{R}_+$, and we shall be primarily interested in stochastic processes for which each realisation is either a continuous function of time (we'll say the process has continuous paths) or a càdlàg function.

Definition 1.5 (Càdlàg function)**.** *A function $f : \mathbb{R}_+ \to \mathbb{R}^d$ is said to be* càdlàg *if it is 'right continuous with left limits', that is*

$$f(t) = f(t+) := \lim_{u \downarrow t} f(u)$$

and

$$f(t-) := \lim_{s \uparrow t} f(s) \text{ exists finitely for every } t > 0.$$

The name comes from the French *'continue à droite, limite à gauche'*.

We are not going to be able to guarantee regularity just through specifying the finite dimensional distributions of the stochastic process. It is not hard to cook up examples of processes with the same finite dimensional distributions, but different sample paths.

Definition 1.6 (Modification)**.** *A process X is called a* modification *of Y if X has the same state-space, (time) parameter set and carrier triple as Y and*

$$\mathbb{P}[X_t = Y_t] = 1 \quad \text{for every } t \in T.$$

We should like to know when we can find a *càdlàg* modification of a stochastic process with a particular law. For continuous parameter martingales (and supermartingales), Doob's Regularity Theorems tell us that the key is to consider martingales adapted to a 'big enough' *filtration*.

Definition 1.7 (Filtration, filtered space)**.** *By a* filtration *on* $(\Omega, \mathcal{F}, \mathbb{P})$*, we mean an increasing family* $\{\mathcal{F}_t\}_{t\geq 0}$ *of sub-*σ*-algebras of* $\mathcal{F}$ *such that for* $0 \leq s \leq t$

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}_\infty := \sigma\Big(\bigcup_{u\geq 0} \mathcal{F}_u\Big) \subseteq \mathcal{F}. \tag{1}$$

Then $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *is called a* filtered probability space.

Definition 1.8 (Adapted)**.** *The stochastic process* X *is* adapted *to the filtration* $\{\mathcal{F}_t\}_{t\geq 0}$ *if* X_t *is measurable with respect to* $\mathcal{F}_t$ *for all* t*.*

For discrete parameter stochastic processes, the most naïve example of filtration is adequate.

Definition 1.9 (Natural filtration)**.** *Let* $W = \{W_n\}_{n\geq 0}$ *be a discrete time stochastic process carried by the probability triple* $(\Omega, \mathcal{F}, \mathbb{P})$*. The* natural filtration $\{\mathcal{W}_n\}_{n\geq 0}$ *is defined to be the smallest filtration relative to which* W *is adapted. That is,*

$$\mathcal{W}_n = \sigma\left(W_0, W_1, \ldots, W_n\right).$$

The obvious analogue of this filtration for continuous martingales is not big enough for many purposes. However, suppose we take a supermartingale $\{Y_t\}_{t\geq 0}$ with respect to its 'natural' filtration, which we denote by $\{\mathcal{G}_t\}_{t\geq 0}$, then it turns out that for almost all ω, the limit through *rational* times

$$X_t(\omega) := \lim_{\mathbb{Q}\ni q\downarrow t} Y_q(\omega)$$

exists simultaneously for all t and defines a right-continuous supermartingale relative to the 'usual augmentation' $\{\mathcal{F}_t\}_{t\geq 0}$ of $\{\mathcal{G}_t\}_{t\geq 0}$ (see Definition 1.11). If the map $t \mapsto \mathbb{E}[Y_t]$ is continuous, then X is a modification of Y.

Definition 1.10 (The usual conditions)**.** *A filtered space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *is said to satisfy the* usual conditions *if in addition to (1) the following properties hold:*

(i) the σ*-algebra* $\mathcal{F}$ *is* $\mathbb{P}$*-complete (that is contains all subsets of all of the* $\mathbb{P}$*-null sets);*

(ii) $\mathcal{F}_0$ *contains all* $\mathbb{P}$*-null sets;*

(iii) $\{\mathcal{F}_t\}_{t\geq 0}$ *is right-continuous, that is*

$$\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{u>t} \mathcal{F}_u \qquad \text{for all } t \geq 0.$$

Definition 1.11 (Usual augmentation)**.** *The* usual augmentation *of the filtered space* $(\Omega, \mathcal{G}, \{\mathcal{G}_t\}_{t\geq 0}, \mathbb{P})$ *is the minimal enlargement* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *that satisfies the usual conditions.*

Here 'enlargement' just means that $\mathcal{G} \subseteq \mathcal{F}$ and $\mathcal{G}_t \subseteq \mathcal{F}_t$ for every t. Recall that the $\mathbb{P}$-completion of a σ-algebra is the smallest enlargement which is $\mathbb{P}$-complete.

Lemma 1.12 (See e.g. Rogers & Williams Lemma II.67.4)**.** *Writing* $\mathcal{N}$ *for the collection of* $\mathbb{P}$*-null sets, the usual enlargement is obtained by taking* $\mathcal{F}$ *to be the* $\mathbb{P}$*-completion of* $\mathcal{G}$ *and setting*

$$\mathcal{F}_t = \bigcap_{u>t} \sigma(\mathcal{G}_u, \mathcal{N}).$$

1.1 Continuous time Markov Processes

Informally, a Markov process models the (random) motion of a particle that moves around in a measurable space $(E, \mathcal{E})$ in a 'memoryless' way. More formally, we'll need a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0})$ on which we define an $\mathcal{F}_t$-measurable random variable X_t which gives the position of our particle at time t and we'll introduce a probability measure $\mathbb{P}^x$ for each point $x \in E$, which will determine the law of the process started from the point x. The Markov property will tie together the different laws $\mathbb{P}^x$.

Definition 1.13 (Markov process, transition function)**.**
A Markov process $X = (\Omega, \{\mathcal{F}_t\}_{t\geq 0}, \{X_t\}_{t\geq 0}, \{P_t\}_{t\geq 0}, \{\mathbb{P}^x\}_{x\in E})$ with state space $(E, \mathcal{E})$ is an E-valued stochastic process, adapted to $\{\mathcal{F}_t\}_{t\geq 0}$ such that for $0 \leq s \leq t$, and f a real-valued bounded measurable function on E and $x \in E$,

$$\mathbb{E}^x[f(X_{s+t})|\mathcal{F}_s] = \int_E f(y)P_t(X_s, dy) \qquad \mathbb{P}^x - a.s.$$

where $\{P_t\}_{t\geq 0}$ is a transition function on $(E, \mathcal{E})$; that is a family of kernels $P_t : E \times \mathcal{E} \to [0, 1]$ such that:

(i) for $t \geq 0$ and $x \in E$, $P_t(x, \cdot)$ is a measure on $\mathcal{E}$ with

$$P_t(x, E) \leq 1;$$

(ii) for $t \geq 0$ and $\Gamma \in \mathcal{E}$, $P_t(\cdot, \Gamma)$ is $\mathcal{E}$-measurable;

(iii) for $s, t \geq 0$, $x \in E$ and $\Gamma \in \mathcal{E}$,

$$P_{t+s}(x, \Gamma) = \int_E P_s(x, dy)P_t(y, \Gamma). \tag{2}$$

Remark 1.14. *(i) We shall sometimes write $P(t, x, \Gamma)$ instead of $P_t(x, \Gamma)$.*

(ii) Unless it leads to ambiguity, we will usually not specify the state space explicitly and, throughout, the filtration will be taken to be the usual enlargement of the natural filtration. Unless otherwise stated, we shall also assume that we are considering a càdlàg version (although sometimes we shall restrict to continuous processes).

(iii) Requiring only $P_t(x, E) \leq 1$ instead of $P_t(x, E) = 1$ allows for the possibility of killing *(in which our Markov process is typically thought of as being sent to a cemetery state).*

(iv) Equation (2) is called the Chapman-Kolmogorov equation*. It is this equation that captures the 'lack of memory' property, also known as the* Markov property*.*

Example 1.15. *(i)* (Pure jump processes.) *By a* pure jump process *we mean a model of a particle which, if it is currently at x, will wait an exponentially distributed amount of time with parameter $\alpha(x)$ before jumping to a new location determined by a probability measure $\mu(x, dy)$. That is, the probability that a particle currently at x does not jump in the next t units of time is $\exp(-\alpha(x)t)$ and, when it jumps, the probability that its new location is in the set B is $\mu(x, B) = \int_B \mu(x, dy)$.*

In the special case when α is constant, these processes are sometimes called pseudo-Poisson processes*.*

If the jumps of the process are homogeneous, in the sense that

$$\int_B \alpha(x)\mu(x, dy) = \int_{B-x} \alpha(0)\mu(0, dy),$$

then we call the process a compound Poisson process. *Such processes are the simplest examples of so-called* Lévy processes.

Of course μ does not have to have a density with respect to Lebesgue measure. For example, a continuous time random walk on the integers is a pure jump process for which μ is purely atomic.

(ii) (Brownian motion in $\mathbb{R}^d$.) *For d-dimensional Brownian motion, the transition functions are given by*

$$P_t(x, E) = \int_E \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x-y\|^2}{2t}\right) dy.$$

(iii) (Solutions to stochastic differential equations.) *The solutions to stochastic differential equations with coefficients depending on the current state are continuous time Markov processes. In this case we can specialise from càdlàg to continuous paths.*

By taking more general state spaces, we can construct Markov processes taking their values in, for example, spaces of functions, or measures, and we can model anything from spin systems in statistical physics to frequencies of different genetic types in a biological population. We can also sometimes gain insight into deterministic quantities by representing them in terms of expectations of functions of Markov processes.

Often we want to consider a stochastic process stopped at a random time. A particularly important class of random times is the stopping times.

Definition 1.16 (Stopping time)**.** *Suppose that the filtration $\{\mathcal{F}_t\}_{t\in T}$ satisfies the usual conditions. An* $\mathcal{F}_t$-stopping time *is a map $\tau : \Omega \to [0, \infty)$ such that $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for all $t \in [0, \infty)$.*

For these times we have the following important extension of the Markov property.

Definition 1.17 (Strong Markov property)**.** *Let X be a càdlàg E-valued Markov process, $\{\mathcal{F}_t\}_{t\geq 0}$ the usual enlargement of the natural filtration, and τ an $\mathcal{F}_t$-stopping time. Then X is* strong Markov at τ *if*

$$\mathbb{P}\left[X(t+\tau) \in \Gamma \middle| \mathcal{F}_\tau\right] = P(t, X(\tau), \Gamma), \quad \forall t \geq 0, \forall \Gamma \in \mathcal{E}.$$

Equivalently,

$$\mathbb{E}\left[f(X_{t+\tau}) \middle| \mathcal{F}_\tau\right] = \int f(y)P(t, X(\tau), dy).$$

We say that X is a strong Markov process *if it has the strong Markov property at all $\{\mathcal{F}_t\}$-stopping times.*

The strong Markov property was formulated by Doob in the 1940's and finally proved for Brownian motion by Hunt in 1956.

2 Feller Semigroups and the Hille Yosida Theorem

The starting point in discussing a Markov process is sometimes its sample paths. For example, last term you studied diffusion processes 'pathwise', as solutions of stochastic differential equations. There will only rarely be a closed form expression for the corresponding transition function. On the other

hand, sometimes we might be given the transition functions, and the challenge is to establish that they really do correspond to a well-defined Markov process. Theorem 2.12 below establishes very mild conditions under which we can guarantee that there is a Markov process corresponding to a collection of transition functions. However, it is rather rare to be provided with an explicit form for the transition functions. Instead, modelling considerations often lead to an 'infinitesimal' description of the process and then we must establish the existence of 'nice' transition functions. This will be the substance of the Hille-Yosida Theorem. Before stating it (a proof is beyond our scope here), we examine some analytic properties of transition functions.

2.1 Transition semigroups

We introduced the transition functions of a Markov process as a family of (sub) probability kernels, but we could equally think of them as a collection of positive bounded operators, of norm less than or equal to one, on the space of bounded measurable functions on E. To be precise, define

$$T_t f(x) := (T_t f)(x) = \int f(y) P_t(x, dy) = \mathbb{E}^x[f(X_t)].$$

Then, approximating f by simple functions, it is easy to see that $T_t f$ is again a bounded measurable function on E. Moreover, the operator T_t is a *positive contraction operator*, that is $0 \leq f \leq 1$ implies $0 \leq T_t f \leq 1$.

Lemma 2.1 (Transition kernels and operators)**.** *The probability kernels $\{P_t\}_{t\geq 0}$ satisfy the Chapman Kolmogorov relation (2) if and only if the corresponding transition operators $\{T_t\}_{t\geq 0}$ have the semigroup property:*

$$T_{s+t} = T_s T_t \qquad s, t \geq 0. \tag{3}$$

Example 2.2 (Brownian motion)**.** *For Brownian motion, the* heat semigroup *is the operator that maps a function f to the solution at time t to the heat equation with initial condition f; that is $T_t f = u$ where*

$$\frac{\partial u}{\partial t}(t, x) = \frac{1}{2}\Delta u(t, x), \qquad u(0, x) = f(x). \tag{4}$$

Of course we can write the solution as

$$u(t, x) = \int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{\|x - y\|^2}{2t}\right) f(y) dy.$$

The Chapman-Kolmogorov equation just tells us that to solve the equation at time $(t+s)$ is the same as solving at time t and then, starting from the resulting solution as initial condition, solving the equation at time s.

Remark 2.3. *The $1/2$ in front of the Laplacian in (4) is a probabilistic convention.*

2.2 Continuous time Markov chains with finite state space

To illustrate some of the ideas to follow, we work in the simple setting of a continuous time Markov chain on a finite state space.

Suppose that E is a finite set and $\mathcal{E}$ is the corresponding power set. We use the notation

$$p_{ij}(t) = P_t(i, \{j\}), \qquad P(t) = \{p_{ij}(t) : i, j \in E\},$$

where $\{P_t\}_{t\geq 0}$ is a transition function on $(E, \mathcal{E})$. We shall assume that $\{P_t\}_{t\geq 0}$ is *honest* in that $P_t(i, E) = 1$. (In other words there is no killing.)

In this setting, a transition semigroup is simply a semigroup of $|E| \times |E|$-matrices. If we want to construct a chain with a right-continuous version, then we are going to require some regularity of these matrices. For example, the chain which jumps immediately at time zero from its current state to a point picked according to some non-trivial distribution on E where it just stays, is perfectly well-defined, but it cannot have a right-continuous version. So we insist that

$$p_{ij}(t) \overset{t\downarrow 0}{\to} \delta_{ij}. \tag{5}$$

We shall see below (Proposition 2.14), that this condition is enough to guarantee the existence of a matrix $Q = \{q_{ij} : i, j \in E\}$, with

$$p'_{ij}(0) = q_{ij}.$$

Definition 2.4 (Q-matrix)**.** *The matrix Q is called the* Q-matrix*, or* infinitesimal generator*, of the Markov chain corresponding to $\{P_t\}$ and has the properties:*

(i) $q_{ij} \geq 0$ *for all* $i \neq j$*,*

(ii) $\sum_{k\in E} q_{ik} = 0$ *for all* $i \in E$*.*

Since $P'(0)$ exists and equals Q, it follows that for $t \geq 0$,

$$P'(t) = \lim_{\varepsilon\downarrow 0} \frac{P(t+\varepsilon) - P(t)}{\varepsilon} = \lim_{\varepsilon\downarrow 0} \frac{P(\varepsilon) - I}{\varepsilon} P(t) = QP(t). \tag{6}$$

We shall prove this in the more general setting of Proposition 2.14 below. Solving this equation, we find that

$$P(t) = \exp(tQ). \tag{7}$$

Indeed, we can start with any matrix satisfying the properties of Q given in Definition 2.4 and from it recover a transition semigroup using the recipe of (7) and the definition of the matrix exponential as

$$e^{tQ} = \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!}. \tag{8}$$

However, when we start to consider Markov processes on more general spaces, the analogue of the Q-matrix will, in general, be unbounded and then it is not clear how to make sense of the exponential formula, since (8) is only defined on $\bigcap_{n\geq 0} \mathcal{D}(Q^n)$, and for unbounded Q this could be a very small set of functions. To see how we might circumvent this problem, consider the real-valued function $p(t) \equiv \exp(ta)$. We can recover a from $p(t)$ either by differentiation or by integration:

$$\frac{1}{t}\left(p(t) - 1\right) \to a \quad \text{as } t \to 0,$$

or

$$\int_0^\infty e^{-\lambda t} p(t) dt = \frac{1}{\lambda - a}, \quad \text{for } \lambda > a. \tag{9}$$

If we replace $p(t)$ by the transition matrix of our Markov chain, then the first expression leads to the Q-matrix. The second expression also makes perfectly good sense (for any $\lambda > 0$) and leads to the so-called *resolvent*, $\{R_\lambda : \lambda > 0\}$ of the transition semigroup. In the case of the finite state space Markov

chains of this section, this is most easily understood as the (componentwise) Laplace transform of the semigroup. In other words,

$$(\lambda R_\lambda)_{ij} = \int_0^\infty \lambda e^{-\lambda t} p_{ij}(t) dt = \mathbb{P}[X_\tau = j | X_0 = i],$$

where τ, which is independent of X, is an exponentially distributed random variable with rate λ. It follows from (7) and (9) that

$$R_\lambda = (\lambda I - Q)^{-1}. \tag{10}$$

In what follows we shall see the same structure emerge in a much more general setting: subject to an analogue of (5), the infinitesimal generator of a transition semigroup will be its derivative at zero, the resolvent, which is just the Laplace transform of the semigroup, is defined by (10) and the semigroup is found by inverting the Laplace transform. In this way, sense can be made of the analogue of the exponential formula (7).

2.3 Poisson and pseudo-Poisson processes

Before moving onto some general theory, we work through two more important examples. The first is one of the simplest examples of a continuous time Markov chain with a countable state space and provides a fundamental building block from which we can construct a vast array of other models.

Example 2.5 (The Poisson process)**.** *The* Poisson process *is the continuous time Markov chain with states* $\{0, 1, 2, \ldots\}$ *and* Q*-matrix*

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

In this case it is not difficult to check that $Q^n = (q_{ij}^{(n)})_{i,j\geq 0}$ *with*

$$q_{ij}^{(n)} = \begin{cases} (\lambda)^n (-1)^{j-i} \binom{n}{j-i} & 0 \leq j - i \leq n \\ 0 & \textit{otherwise,} \end{cases}$$

and so

$$P_t = \sum_{n=0}^\infty \frac{(tQ)^n}{n!}$$

satisfies

$$(P_t)_{ij} = \begin{cases} \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t} & 0 \leq j - i < \infty \\ 0 & \textit{otherwise.} \end{cases}$$

In other words, the number of new arrivals in any time interval of length t *is Poisson with parameter* λt*. It is an easy exercise to check that* $P_t' = QP_t$*.*

Instead of this 'bare hands' approach, we could also have applied the resolvent method. To see how this works in this case, note that for $\rho > 0$*,*

$$(\rho I - Q) = \begin{pmatrix} \rho + \lambda & -\lambda & 0 & 0 & \cdots \\ 0 & \rho + \lambda & -\lambda & 0 & \cdots \\ 0 & 0 & \rho + \lambda & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

The coefficients r_{in} of the inverse matrix therefore satisfy

$$r_{00}(\rho+\lambda)=1, \quad \textit{and } r_{0,n-1}(-\lambda)+r_{0,n}(\rho+\lambda)=0, \quad \textit{for } n\geq 1.$$

Solving, this gives

$$r_{0,n}=\frac{\lambda}{\rho+\lambda}r_{0,n-1}=\frac{\lambda^n}{(\lambda+\rho)^{n+1}}.$$

Recall that

$$(R_\rho)_{ij}=\int_0^\infty e^{-\rho t}p_{ij}(t)dt,$$

so inverting the Laplace transform, we recover

$$(P_t)_{0,n}=\mathbb{P}\left[X_t=n|\, X_0=0\right]=\frac{(\lambda t)^n}{n!}e^{-\lambda t},$$

as before. To check this, take the Laplace transform of this expression and invoke uniqueness of Laplace transforms:

$$\int_0^\infty e^{-\rho t}\frac{(\lambda t)^n}{n!}e^{-\lambda t}dt=\frac{\lambda^n}{n!}\int_0^t e^{-(\lambda+\rho)t}t^n dt=\frac{\lambda^n}{(\lambda+\rho)^{n+1}}.$$

At the next level of complexity we have the pseudo-Poisson processes of Example 1.15 (i).

Proposition 2.6. *Let T_t be the transition semigroup associated with a pseudo-Poisson process X in E. Then $T_t=e^{tA}$ for all $t\geq 0$, where for any bounded measurable function $f:E\to\mathbb{R}$,*

$$(Af)(x)=\alpha\int_{y\in E}(f(y)-f(x))\,\mu(x,dy), \qquad x\in E.$$

Proof. The pseudo-Poisson process can be written as $X=Y\circ N$ where N is a Poisson process with parameter α and Y is a *discrete time* Markov chain, independent of N, with jumps determined by μ. That is, the number of jumps that the process has made by time t is Poisson with parameter αt and the distribution of each jump is determined by the probability kernel μ. Now write G for the transition operator associated with μ. That is G, which is the analogue of the transition matrix for a finite state space Markov chain, governs a single jump of the process, so that

$$Gf(x)=\int f(y)\mu(x,dy).$$

Then for any $t\geq 0$ and bounded measurable f,

$$\begin{aligned}
T_tf(x) &= \mathbb{E}^x[f(X_t)]=\sum_{n=0}^\infty \mathbb{E}^x[f(Y_n);N_t=n]\\
&= \sum_{n=0}^\infty \mathbb{P}[N_t=n]\mathbb{E}^x[f(Y_n)]\\
&= \sum_{n=0}^\infty e^{-\alpha t}\frac{(\alpha t)^n}{n!}G^nf(x)\\
&= e^{\alpha t(G-I)}f(x).
\end{aligned}$$

□

Once again (c.f. (6)), since $Af(x)=\alpha(G-I)f(x)$, we see that T_tf is the solution to Cauchy's equation:

$$(T_tf)'=AT_tf, \qquad T_0f=f.$$

2.4 A little bit of functional analysis

Let $(E, \mathcal{E})$ be complete (contains all its limit points) compact and separable (contains a countable dense set). Recall that we can always make E compact by adding a 'point at infinity'. We write $C(E)$ for the space of continuous real-valued functions on E equipped with the *sup* norm, that is $\|f\|_\infty = \sup_{x\in E}\{|f(x)|\}$. This norm makes $C(E)$ into a *Banach space* (that is a complete normed vector space).

Definition 2.7 (Operator norm)**.** *Given a linear operator $T : C(E) \to C(E)$, we define its* operator norm *by*

$$\|T\| = \sup\{\|Tf\|_\infty : f \in C(E), \|f\|_\infty = 1\}.$$

Definition 2.8 (Closed operator)**.** *A linear operator A with domain $\mathcal{D}(A)$ is* closed *if its graph is closed.*

If A is closed, then if x_n is a Cauchy sequence in $\mathcal{D}(A)$, converging to x and Ax_n converges to y, then $x \in \mathcal{D}(A)$ and $Ax = y$.

To make formulae easier to read, we sometimes write $T(t)$ in place of T_t.

Definition 2.9 (Strongly continuous contraction semigroup)**.** *A family of bounded operators $\{T(t), 0 \le t \le \infty\}$ on a Banach space C is called a strongly continuous contraction semigroup if*

(i) $T(0) = I$;

(ii) $T(s)T(t) = T(s+t), \forall s, t \in \mathbb{R}_+$;

(iii) $\|T(t)\| \le 1$ *for all* $t \ge 0$.

(iv) for any $z \in C$, $t \mapsto T(t)z$ is continuous.

When our Banach space is $C(E)$, we shall use f rather than z to denote a typical element and we shall abuse notation and use 1 to denote the function $f(x) \equiv 1$.

Definition 2.10 (Feller semigroup)**.** *A strongly continuous contraction semigroup on $C(E)$ with the additional properties*

(i) $T(t)1 = 1$ *and*

(ii) $T(t)f \ge 0$ *for all non-negative* $f \in C(E)$,

is called a Feller semigroup.

Remark 2.11. *In fact there is some inconsistency in the literature over the definition of Feller semigroup. For example, if one only assumes that E is locally compact (rather than compact), then it is usual to replace $C(E)$ by $C_0(E)$, the continuous functions on E which tend to zero at infinity. Such functions are continuous in the one-point compactification of E.*

Theorem 2.12. *Suppose that $T(t)$ is a Feller semigroup on $C(E)$. Then there exists a unique Markov process, X, taking values in càdlàg paths on E, with law $\{\mathbb{P}^\eta, \eta \in E\}$ such that $\mathbb{P}^\eta[X_0 = \eta] = 1$ and*

$$T(t)f(\eta) = \mathbb{E}^\eta\left[f(X_t)\right]$$

for all $f \in C(E)$, $\eta \in E$ and $t \ge 0$.

Remark 2.13. *Notice that our assumptions include the statement that the mapping $f \mapsto g$, where $g(\eta) = \mathbb{E}^\eta[f(X_t)]$, maps $C(E)$ to $C(E)$. We then say that the Markov process X is a* Feller process *or has the* Feller property*.*

We don't prove Theorem 2.12. For the construction of the process from the semigroup we refer to Chapter 4 of Ethier & Kurtz (1986).

Proposition 2.14 (Infinitesimal generator)**.** *Let $T(t)$ be a strongly continuous contraction semigroup on a Banach space C and define*

$$A_t = \frac{1}{t}\left(T(t) - I\right), \quad \forall t > 0.$$

Set $Az = \lim_{t\downarrow 0} A_t z$ where $\mathcal{D}(A) = \{z : \lim_{t\downarrow 0} A_t z \text{ exists}\}$. Then A is a densely defined closed operator and is called the infinitesimal generator *of $T(t)$. Moreover, for any $z \in \mathcal{D}(A)$ and $t \geq 0$, we have $T(t)z \in \mathcal{D}(A)$ and*

$$\frac{dT(t)z}{dt} = AT(t)z = T(t)Az. \tag{11}$$

Remark 2.15. *Derivatives $\frac{d}{dt}f(t)$ (and integrals $\int_0^t f(s)ds$ in the following proof) for Banach space-valued functions f may appear unfamiliar, but these objects are defined similarly as for $\mathbb{R}$-valued functions: $\frac{d}{dt}f(t) = \lim_{h\to 0} h^{-1}(f(t+h) - f(t))$, while integrals are first defined for simple functions and then extended by continuity (often called the Bochner integral). The question of which functions are integrable can be subtle, but we will only ever look at $\int_0^t f(s)ds$ for continuous $f\colon [0,\infty) \to C$, in which case it is clear that f can be uniformly approximated on every compact interval by simple functions. All familiar properties of integrals (triangle inequality, fundamental theorem of calculus, Fubini's theorem, etc.) hold in this setting. See Parts 3 and 4 of Lang (1993) for a detailed treatment.*

Proof. For any $z \in C$ and $s > 0$ define the mean value

$$z_s = \frac{1}{s}\int_0^s T(u)z\,du.$$

Then for $t < s$,

$$\begin{aligned}
A_t z_s &= \frac{1}{st}\int_0^s [T(t+u) - T(u)]z\,du \qquad \text{(semigroup property)} \\
&= \frac{1}{st}\int_t^{t+s} T(u)z\,du - \frac{1}{st}\int_0^s T(u)z\,du \\
&= \frac{1}{st}\int_s^{t+s} T(u)z\,du - \frac{1}{st}\int_0^t T(u)z\,du \qquad \text{(since } t < s\text{, so } \textstyle\int_t^s \text{ cancels)} \\
&= \frac{1}{s}\left(T(s) - I\right) z_t \to A_s z, \quad \text{as } t \downarrow 0 \qquad \text{(since } t \mapsto T(t)z \text{ is continuous).}
\end{aligned}$$

This shows that $z_s \in \mathcal{D}(A)$ and so, since $z_s \to z$ as $s \to 0$, $\mathcal{D}(A)$ is a dense set.

To verify (11), first observe that by definition of A_s, we have

$$A_s T(t)z = \frac{1}{s}\left(T(t+s) - T(t)\right) z = T(t)A_s z.$$

If $z \in \mathcal{D}(A)$, then we can take the limit as $s \downarrow 0$ on the right hand side. But then this shows that $T(t)z \in \mathcal{D}(A)$ and so $T(t) : \mathcal{D}(A) \to \mathcal{D}(A)$. Moreover,

$$\frac{d^+T(t)z}{dt} = \lim_{s\downarrow 0} \frac{T(t+s)z - T(t)z}{s} = T(t)Az = AT(t)z.$$

To complete the proof of (11), we must check that for $t > 0$, $\frac{d^- T(t)z}{dt} = \frac{d^+ T(t)z}{dt}$. For any $t > \delta > 0$,

$$\begin{aligned}\left\| \frac{T(t)z - T(t-\delta)z}{\delta} - T(t)Az \right\| &\leq \left\| T(t-\delta)\left(\frac{T(\delta)z - z}{\delta} - Az \right) \right\| + \left\| T(t-\delta)\left(Az - T(\delta)Az\right) \right\| \\ &\leq \left\| T(t-\delta) \right\| \left\| \frac{T(\delta)z - z}{\delta} - Az \right\| + \left\| T(t-\delta) \right\| \left\| Az - T(\delta)Az \right\| .\end{aligned}$$

Since $\|T(t-\delta)\| \leq 1$, letting $\delta \downarrow 0$, we have that

$$\frac{d^- T(t)z}{dt} = T(t)Az = \frac{d^+ T(t)z}{dt}$$

as required.

Finally, we check that A is closed. Suppose that $z_n \in \mathcal{D}(A)$ is such that $z_n \to z$ and $y_n = Az_n \to y$. Then

$$\begin{aligned}\lim_{t\downarrow 0} A_t z &= \lim_{t\downarrow 0} \lim_{n\to\infty} A_t z_n \\ &= \lim_{t\downarrow 0} \lim_{n\to\infty} \frac{1}{t}\left(T(t) - I\right) z_n \\ &= \lim_{t\downarrow 0} \lim_{n\to\infty} \frac{1}{t} \int_0^t T(u) A z_n du \\ &= \lim_{t\downarrow 0} \frac{1}{t} \int_0^t T(u) y du = y.\end{aligned}$$

Thus $z \in \mathcal{D}(A)$ and $y = Az$ as required. □

Definition 2.16. *Let $T(t)$ be a strongly continuous contraction semigroup on a Banach space C. The resolvent $(R_\lambda)_{\lambda>0}$ is a family of bounded linear operators on C defined through the Laplace transform of $T(t)$, i.e.*

$$R_\lambda z = \int_0^\infty e^{-\lambda t} T(t) z dt \ , \quad \forall z \in C \ .$$

The following corollary of Proposition 2.14 extends formula (9) to the general setting.

Corollary 2.17. *Let $\lambda > 0$ and $T(t)$ be a strongly continuous contraction semigroup on a Banach space C. The operator $\lambda - A \colon \mathcal{D}(A) \to C$ is invertible with inverse*

$$(\lambda - A)^{-1} = R_\lambda \ .$$

Proof. Define $S(t) = e^{-\lambda t} T(t)$, which is again a strongly continuous contraction semigroup on C. The infinitesimal generator of $S(t)$ is easily seen to be $A - \lambda$ with the same domain $\mathcal{D}(A - \lambda) = \mathcal{D}(A)$. It follows from (11) that, for all $z \in \mathcal{D}(A)$,

$$\int_0^\infty e^{-\lambda t} T(t)(A - \lambda) z dt = \int_0^\infty \frac{d}{dt} S(t) z dt = -z \ ,$$

from which we see that R_λ is the left inverse of $(\lambda - A)$ on $\mathcal{D}(A)$. Using the fact that $T(t) \colon \mathcal{D}(A) \to \mathcal{D}(A)$, the same argument shows that R_λ is the right inverse of $(\lambda - A)$ on $\mathcal{D}(A)$.

Consider now $g \in C$. By density of $\mathcal{D}(A)$ in C, there exists a sequence $g_n \in \mathcal{D}(A)$ such that $g_n \to g$. Defining $f_n = R_\lambda g_n$, by boundedness of R_λ, we have $(f_n, g_n) \to (f, g)$ where $f = R_\lambda g$. Since $\lambda - A$ is closed and $g_n = (\lambda - A) f_n$, we have $f \in \mathcal{D}(A)$ with $(\lambda - A) f = g$, from which it follows that $R_\lambda \colon C \to \mathcal{D}(A)$ is indeed the inverse of $\lambda - A$. □

2.5 The Hille-Yosida Theorem

So far we have proved that to every strongly continuous contraction semigroup, there corresponds an infinitesimal generator. Our next aim is to go the other way: starting from an infinitesimal generator, we'd like to recover the transition semigroup of a Markov process. This section follows Liggett (1985) closely.

First we need to know how to recognise an infinitesimal generator when we see one.

Definition 2.18 (Markov pregenerator)**.** *A (usually unbounded) linear operator A on $C(E)$ with domain $\mathcal{D}(A)$ is said to be a* Markov pregenerator *if it satisfies the following conditions:*

(i) $1 \in \mathcal{D}(A)$ *and* $A1 = 0$*;*

(ii) $\mathcal{D}(A)$ *is dense in* $C(E)$*;*

(iii) If $f \in \mathcal{D}(A)$*,* $\lambda \geq 0$ *and* $f - \lambda Af = g$*, then*

$$\min_{\zeta \in E} f(\zeta) \geq \min_{\zeta \in E} g(\zeta).$$

Exercise 2.19. *Show that a Markov pregenerator has the property that if $f \in \mathcal{D}(A)$, $\lambda \geq 0$ and $f - \lambda Af = g$, then $\|f\| \leq \|g\|$. Deduce that, in particular, g determines f uniquely.*

To verify (iii) one usually uses the following result.

Proposition 2.20. *Suppose that the linear operator A on $C(E)$ satisfies the following property: if $f \in \mathcal{D}(A)$ and $f(\eta) = \min_{\zeta \in E} f(\zeta)$, then $Af(\eta) \geq 0$. Then A satisfies property (iii) of Definition 2.18.*

Proof. Suppose $f \in \mathcal{D}(A)$, $\lambda \geq 0$ and $f - \lambda Af = g$. Let η be any point at which f attains its minimum. Such a point exists by the compactness of E and the continuity of f. Then

$$\min_{\zeta \in E} f(\zeta) = f(\eta) \geq f(\eta) - \lambda Af(\eta) = g(\eta) \geq \min_{\zeta \in E} g(\zeta).$$

□

A Markov pregenerator necessarily satisfies a *maximum principle*. Recall from elementary calculus that a function $f \in C(a, b)$ taking a local maximum at x satisfies $f'(x) = 0$ and $f''(x) \leq 0$. This is often rephrased to say that if $f \in C^2([a, b])$ satisfies $f'' \geq 0$, and f attains its maximum in (a, b), then f is constant. This can be extended to *elliptic operators* on $\mathbb{R}^d$.

Definition 2.21 (Elliptic operator)**.** *A second order differential operator*

$$L(u) = \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x) \frac{\partial u}{\partial x_i} - c(x)u \tag{12}$$

is called elliptic *at a point x if the matrix $A(x) = (a_{ij}(x))$ is* positive definite*, that is satisfies*

$$z^T A(x) z \geq \mu(x) \|z\|^2, \qquad \forall z \in \mathbb{R}^d,$$

for some positive μ, and is called uniformly elliptic *if $\mu > 0$ can be taken independent of x.*

Lemma 2.22 (Maximum principle for elliptic operators)**.** *Let L be an elliptic operator as in (12) with $c(x) \geq 0$, and suppose that $Lu \geq 0$ in some domain $E \subseteq \mathbb{R}^d$. If u takes a non-negative maximum in the interior of E, then u is constant.*

This same *positive maximum principle* holds for a Feller semigroup:

Lemma 2.23 (Positive maximum principle for Feller semigroups)**.** *Let A be the generator of a Feller semigroup acting on functions on $E \subseteq \mathbb{R}^d$. If f is such that $Af \geq 0$ in E and f takes a non-negative maximum inside E, then f is constant.*

Remark 2.24. *The fact that elliptic operators and Feller semigroups satisfy the same maximum principle reflects a much deeper connection between the two classes. Write C_c^∞ for the class of infinitely differentiable functions with compact support on $\mathbb{R}^d$. If X is a continuous Feller process on $[0,T]$ with infinitesimal generator A and $C_c^\infty \subseteq \mathcal{D}(A)$, then A is an elliptic operator.*

Example 2.25. *It can be easily checked using Proposition 2.20 that the following are Markov pregenerators:*

(i) $A = G - I$ *where G is a positive operator (maps non-negative functions to non-negative functions) defined on all of $C(E)$ such that $G1 = 1$.*

(ii) $E = [0,1]$ *and* $Af(\eta) = \frac{1}{2}f''(\eta)$ *with*

$$\mathcal{D}(A) = \{f \in C(E) : f'' \in C(E), f'(0) = 0 = f'(1)\}.$$

(iii) $E = [0,1]$ *and* $Af(\eta) = \frac{1}{2}f''(\eta)$ *with*

$$\mathcal{D}(A) = \{f \in C(E) : f'' \in C(E), f''(0) = 0 = f''(1)\}.$$

Exercise 2.26. *Calculate the infinitesimal generator of the pure jump process of Example 1.15 (i) and check that it is of the form of Example 2.25 (i) above.*

Definition 2.27 (Closure of a linear operator)**.** *Let A be a linear operator on $C(E)$. A linear operator $\bar{A}$ is called the closure of A if it is the smallest closed extension of A.*

Not every linear operator has a closure, the difficulty being that the closure of the graph of a linear operator may not correspond to the graph of a linear operator (see the problem sheet for an example), but instead to a 'multivalued' operator. Happily, this won't affect us:

Proposition 2.28. *Suppose that A is a Markov pregenerator. Then A has a closure $\bar{A}$ which is again a Markov pregenerator.*

Proof. The first step is to check that the closure of the graph of A is the graph of a single-valued linear operator $\bar{A}$. To this end, suppose that $f_n \in \mathcal{D}(A)$, $f_n \to 0$ and $Af_n \to h$. Choose any $g \in \mathcal{D}(A)$. Then, as we already observed (Exercise 2.19), for $\lambda \geq 0$,

$$\|(I - \lambda A)(f_n + \lambda g)\| \geq \|f_n + \lambda g\|,$$

and so letting $n \to \infty$

$$\| - \lambda h + \lambda(g - \lambda A g)\| = \|\lambda g - \lambda h - \lambda^2 A g\| \geq \|\lambda g\|.$$

Dividing by λ and letting $\lambda \downarrow 0$ gives $\|g - h\| \geq \|g\|$. Since g was an arbitrary element of the dense set $\mathcal{D}(A)$ we deduce that $h = 0$. This implies that the closure of the graph of A is indeed the graph of a single-valued linear operator.

It remains to check that $\bar{A}$ is also a Markov pregenerator, for which we must check (iii) of Definition 2.18. Take $f \in \mathcal{D}(\bar{A})$, $\lambda \geq 0$ and $g = f - \lambda \bar{A} f$. By definition, there is a sequence $f_n \in \mathcal{D}(A)$

such that $f_n \to f$, and $Af_n \to \bar{A}f$. Setting $g_n = f_n - \lambda A f_n$, and using the fact that A is a Markov pregenerator,

$$\min_{\zeta \in E} f_n(\zeta) \geq \min_{\zeta \in E} g_n(\zeta).$$

Since $g_n \to g$ as $n \to \infty$, passing to the limit we obtain property (iii) of Definition 2.18 as required. □

Proposition 2.29. *Suppose that A is a closed Markov pregenerator. Then for each $\lambda > 0$, the range of $I - \lambda A$, which we denote $\mathcal{R}(I - \lambda A)$ is a closed subset of $C(E)$.*

Proof. Suppose that $g_n \in \mathcal{R}(I - \lambda A)$ and $g_n \to g$ as $n \to \infty$. Define f_n by $f_n - \lambda A f_n = g_n$. Then by linearity,

$$(f_n - f_m) - \lambda A(f_n - f_m) = g_n - g_m$$

so that by Exercise 2.19, $\|f_n - f_m\| \leq \|g_n - g_m\|$ and, since g_n is a Cauchy sequence, the same must be true of f_n. Let $f = \lim_{n\to\infty} f_n$, then

$$\lim_{n\to\infty} Af_n = \frac{1}{\lambda} \lim_{n\to\infty} (f_n - g_n) = \frac{1}{\lambda}(f - g)$$

and so, since A is closed (and rearranging),

$$f - \lambda A f = g.$$

We have shown that $g \in \mathcal{R}(I - \lambda A)$, as required. □

Definition 2.30 (Markov Generator)**.** *A* Markov generator *is a closed Markov pregenerator A for which $\mathcal{R}(I - \lambda A) = C(E)$ for all $\lambda \geq 0$.*

Remark 2.31. *On the problem sheet, you are asked to prove the following claims:*

(i) For a closed Markov pregenerator, if $\mathcal{R}(I - \lambda A) = C(E)$ for all sufficiently small positive λ, then the same will follow for all non-negative λ.

(ii) If a Markov pregenerator is everywhere defined and is a bounded *operator, then it is automatically a Markov generator.*

We are finally in a position to state the main result of this section.

Theorem 2.32 (Hille-Yosida Theorem)**.** *There is a one-to-one correspondence between Markov generators on $C(E)$ and Feller semigroups on $C(E)$, given as follows:*

(i)

$$\mathcal{D}(A) = \left\{ f \in C(E) : \lim_{t \downarrow 0} \frac{T(t)f - f}{t} \text{ exists} \right\},$$

and

$$Af = \lim_{t \downarrow 0} \frac{T(t)f - f}{t} \text{ for } f \in \mathcal{D}(A).$$

(ii)

$$T(t)f = \lim_{n\to\infty} \left(I - \frac{t}{n} A \right)^{-n} f \quad \text{for } f \in C(E) \text{ and } t \geq 0.$$

Furthermore,

(iii) if $f \in \mathcal{D}(A)$, it follows that $T(t)f \in \mathcal{D}(A)$ and

$$\frac{d}{dt}T(t)f = AT(t)f = T(t)Af,$$

and

(iv) for $g \in C(E)$ and $\lambda \geq 0$, the solution to $f - \lambda Af = g$ is given by

$$f = \int_0^\infty e^{-t}T(\lambda t)g dt = \frac{1}{\lambda}R_{1/\lambda}g.$$

Remark 2.33. *Parts (iii) and (iv) of Theorem 2.32 are simply restatements of Proposition 2.14 and Corollary 2.17 for general strongly continuous contraction semigroups.*

Usually, in order to show existence of a process, first we provide an infinitesimal description and use this to write down a Markov pregenerator, A. The next, and most difficult, step is to show that $\mathcal{R}(I - \lambda A)$ is dense in $C(E)$ (for sufficiently small positive λ). Then the closure of A exists and is a Markov generator, which the Hille-Yosida Theorem tells us generates a Feller semigroup.

Example 2.34. *On the problem sheet you are asked to check that each of the pregenerators in Example 2.25 is actually a Markov generator.*

The Markov processes to which these examples correspond are

(i) The pure jump process of Example 1.15 (i);

(ii) Brownian motion on $[0, 1]$ with reflecting barriers at 0 and 1;

(iii) Brownian motion on $[0, 1]$ with absorbing barriers at 0 and 1.

Notice that the intersection of the domains of the generators in (ii) and (iii) is a dense subset of $C([0, 1])$ and the generators agree on that set, even though the processes generated are quite different. This tells us that a Markov generator is not automatically determined by its values on a dense subset of $C(E)$. For this reason we make the following definition.

Definition 2.35 (Core of a generator)**.** *Suppose that A is a Markov generator on $C(E)$. A linear subspace $\mathcal{D}$ of $\mathcal{D}(A)$ is said to be a* core *for A if A is the closure of its restriction to $\mathcal{D}$.*

Trivially, A is uniquely determined by its values on a core. If we follow the procedure described immediately after Theorem 2.32 to write down a generator, then the domain of the original pregenerator is automatically a core for the generator. In fact, in most cases we won't know the full domain of a generator explicitly, but knowing a core is just as good.

Example 2.36 (Wright-Fisher diffusion)**.** *This diffusion process is of fundamental importance in population genetics, where it is used to model the evolution of gene frequencies, and we shall return to it many times in the rest of the course. It takes its values in $[0, 1]$ and the infinitesimal generator takes the form*

$$Af(x) = \frac{1}{2}x(1 - x)f''(x),$$

for an appropriate subset of $C^2([0, 1])$ (the twice continuously differentiable functions on $[0, 1]$). The issue of exactly which functions are in $\mathcal{D}(A)$ is not clear, but to construct the process from the Hille-Yosida Theorem, we define Af for a reasonable class of functions f and then check that $\mathcal{R}(I - \lambda A)$

is dense in $C([0,1])$. To do so, take a polynomial $g(x) = \sum_{k=0}^n a_k x^k$, and try to solve the equation $f - \lambda Af = g$. Let us write $f(x) = \sum_{k=0}^n c_k x^k$. This leads to a recursion for the coefficients c_k:

$$c_k - \frac{\lambda}{2}\left[k(k+1)c_{k+1} - (k-1)kc_k\right] = a_k,$$

which can be solved. It follows that if we define A on the set of polynomials then $\mathcal{R}(I - \lambda A)$ is dense in $C([0,1])$ as required. Taking the closure, it follows that the domain of A is

$$\{f : \text{ there exist polynomials } p_n \text{ such that } p_n \to f \text{ uniformly and } x(1-x)p_n'' \text{ has a uniform limit}\}.$$

In particular, we see that every $f \in \mathcal{D}(A)$ satisfies $Af(0) = 0 = Af(1)$. This corresponds to the fact that if gene frequencies hit zero or one, then they stay there.

Remark 2.37. *The Wright-Fisher diffusion solves the stochastic differential equation*

$$dX_t = \sqrt{X_t(1-X_t)}dB_t,$$

where B_t is Brownian motion. In particular, notice that the coefficient in front of the noise is not uniformly Lipschitz, but only locally Lipschitz continuous.

Often we obtain stochastic processes of interest by an approximation procedure. In this context the following result is helpful.

Theorem 2.38 (Trotter-Kurtz)**.** *Suppose that A_n and A are the generators of Feller semigroups $T_n(t)$ and $T(t)$ respectively. If there is a core $\mathcal{D}$ for A such that $\mathcal{D} \subseteq \mathcal{D}(A_n)$ for all n and $A_n f \to Af$ for all $f \in \mathcal{D}$, then*

$$T_n(t)f \to T(t)f$$

for all $f \in C(E)$ uniformly for t in compact sets.

Part ((iii)) of the Hille-Yosida Theorem tells us that $F(t) = T(t)f$ gives a solution to the Cauchy problem:

$$\frac{d}{dt}F(t) = AF(t), \qquad F(0) = f, \qquad \text{for } f \in \mathcal{D}(A).$$

Sometimes one would like to know that the semigroup actually gives the *unique* solution to a problem of this type:

Theorem 2.39. *Suppose that A is the generator of a Feller semigroup. Further, suppose that $F(t)$ and $G(t)$ are functions on $[0,\infty)$ with values in $C(E)$ which satisfy*

(i) $F(t) \in \mathcal{D}(A)$ for each $t \geq 0$;

(ii) $G(t)$ is continuous on $[0,\infty)$; and

(iii)

$$\frac{d}{dt}F(t) = AF(t) + G(t), \qquad \text{for } t \geq 0.$$

Then

$$F(t) = T(t)F(0) + \int_0^t T(t-s)G(s)ds. \tag{13}$$

Proof. We establish the derivative of $T(t-s)F(s)$ with respect to s. First observe that

$$\frac{T(t-s-h)F(s+h)-T(t-s)F(s)}{h} = T(t-s)\left[\frac{F(s+h)-F(s)}{h}\right]+\left[\frac{T(t-s-h)-T(t-s)}{h}\right]F(s)$$
$$+\left[T(t-s-h)-T(t-s)\right]F'(s)+\left[T(t-s-h)-T(t-s)\right]\left[\frac{F(s+h)-F(s)}{h}-F'(s)\right].$$

Since $T(t-s)$ is a bounded operator, the first term on the right hand side converges to $T(t-s)F'(s)$ as $h \to 0$. Using ((iii)) of Theorem 2.32 (and the first assumption), the second term converges to $-T(t-s)AF(s)$. The third term tends to zero since for each $f \in C(E)$ the mapping $t \mapsto T(t)f$ is uniformly continuous; and the fourth tends to zero since $T(t-s)$ and $T(t-s-h)$ are both contractions. Combining these observations, we have that, for $0 < s < t$,

$$\frac{d}{ds}\left(T(t-s)F(s)\right) = T(t-s)F'(s) - T(t-s)AF(s),$$

which, by the third assumption, becomes

$$\frac{d}{ds}T(t-s)F(s) = T(t-s)G(s). \tag{14}$$

Finally, by the second assumption (and continuity of the mapping $t \mapsto T(t)f$ for $f \in C(E)$), $T(t-s)G(s)$ is a continuous function of s. Integrating (14) then gives

$$F(t) - T(t)F(0) = \int_0^t T(t-s)G(s)ds,$$

as required. □

Remark 2.40. *Evidently we can specialise equation (13). For example, if $A = \frac{1}{2}\Delta$, then we see that we can represent the solution to*

$$\frac{\partial f}{\partial t}(t,x) = \frac{1}{2}\Delta f(t,x) + g(t,x)$$

as

$$f(t,x) = \mathbb{E}_x\left[f(0,B_t) + \int_0^t g(s,B_{t-s})ds\right],$$

where B_t is a Brownian motion and the subscript 'x' in the expectation indicates that it starts at $B_0 = x$. This gives us a nontrivial example of a stochastic representation of a solution to a deterministic PDE.

If we consider Brownian motion on $[0,1]$ with absorbing boundary (or some higher-dimensional analogue), then for example setting $g \equiv 1$ gives the expected time for the process to leave the domain. We return to this later.

3 Martingale problems

We now turn to the martingale approach to stochastic processes. Originally developed by Stroock and Varadhan for the construction and analysis of diffusion processes with continuous coefficients, it is particularly well adapted to weak convergence arguments. A thorough exposition in the diffusion context can be found in Stroock & Varadhan (1979). Connections to weak convergence are covered in Ethier & Kurtz (1986).

3.1 Martingale problems and Markov (pre-)generators

Throughout this subsection we let $(E, \mathcal{E})$ be a measurable space in which E is a metric pace.

Definition 3.1 (The space D and its associated filtration)**.** *We shall write $D[0,\infty)$, or sometimes just D, for the space of càdlàg functions from $[0,\infty)$ to E. This is the canonical path space for a Markov process with state space E. For $s \in [0,\infty)$, the* evaluation mapping $\pi_s : D[0,\infty) \to E$ *is defined by $\pi_s(\eta_\cdot) = \eta_s$. Let $\mathcal{F}$ be the smallest σ-algebra with respect to which all the mappings π_s are measurable and for $t \in [0,\infty)$, let $\mathcal{F}_t$ be the smallest σ-algebra on $D[0,\infty)$ relative to which all the mappings π_s for $0 \le s \le t$ are measurable.*

Definition 3.2 (Martingale problem)**.** *Suppose that A is a Markov pregenerator and $\mu \in \mathcal{P}(E)$, where $\mathcal{P}(E)$ is the set of all probability measures on $(E, \mathcal{E})$. A probability measure $\mathbb{P}$ on $D[0,\infty)$ is said to* solve the martingale problem (A, μ) *if*

(i) $\mathbb{P}[\{\zeta \in D : \zeta_0 \in \Gamma\}] = \mu(\Gamma)$ *for all* $\Gamma \in \mathcal{E}$ *and*

(ii)

$$f(\eta_t) - \int_0^t Af(\eta_s)ds$$

is a (local) martingale relative to $\mathbb{P}$ and the σ-algebras $\{\mathcal{F}_t, t \ge 0\}$ for all $f \in \mathcal{D}(A)$.

If $\mu = \delta_\eta$, the Dirac delta at $\eta \in E$, then the first condition is equivalent to $\mathbb{P}[\{\zeta \in D : \zeta_0 = \eta\}] = 1$, and we simply say that $\mathbb{P}$ solves the martingale problem for A with initial point η.

Notice that in contrast to the Hille-Yosida Theorem, which required A to be a Markov generator, here we only require that it be a Markov *pre*generator. Existence of a solution to the martingale problem often requires only rather minimal conditions on the infinitesimal parameters of the process. Uniqueness is typically much more difficult to establish. We shall present a powerful approach to uniqueness in §5, but first we explore the connection between Feller processes and martingale problems in a little more detail.

Theorem 3.3. *Suppose that E is compact and separable and that A is a Markov pregenerator for which the closure $\bar{A}$ is a Markov generator. Let $\{\mathbb{P}^x, x \in E\}$ be the unique Feller process that corresponds to $\bar{A}$ (through Theorem 2.12 and Theorem 2.32). Then for each $x \in E$, $\mathbb{P}^x$ is the unique solution to the martingale problem for A with initial point x.*

Proof. In order to check that $\mathbb{P}^x$ is a solution to the martingale problem, fix $f \in \mathcal{D}(A)$ and, for $0 < r < t$, write

$$\begin{aligned}
\mathbb{E}^x\left[\int_r^t Af(\eta_s)ds \middle| \mathcal{F}_r\right] &= \mathbb{E}^{\eta_r}\left[\int_0^{t-r} Af(\eta_s)ds\right] \\
&= \int_0^{t-r} \mathbb{E}^{\eta_r}[Af(\eta_s)]ds \\
&= \int_0^{t-r} T(s)Af(\eta_r)ds \\
&= \int_0^{t-r} \bar{A}T(s)f(\eta_r)ds \\
&= T(t-r)f(\eta_r) - f(\eta_r).
\end{aligned}$$

The last two equalities follow from ((iii)) of Theorem 2.32. Using this we now compute that for $r < t$

$$\begin{aligned}\mathbb{E}^x\left[f(\eta_t) - \int_0^t Af(\eta_s)ds\middle|\mathcal{F}_r\right] &= \mathbb{E}^x\left[f(\eta_t)|\,\mathcal{F}_r\right] - \int_0^r Af(\eta_s)ds - \mathbb{E}^x\left[\int_r^t Af(\eta_s)ds\middle|\mathcal{F}_r\right] \\ &= \mathbb{E}^{\eta_r}[f(\eta_{t-r})] - \int_0^r Af(\eta_s)ds - T(t-r)f(\eta_r) + f(\eta_r) \\ &= f(\eta_r) - \int_0^r Af(\eta_s)ds.\end{aligned}$$

Therefore $\mathbb{P}^x$ is a solution to the martingale problem for A.

We now turn to uniqueness. Fix $x \in E$ and let $\mathbb{P}$ be any solution to the martingale problem for A (and hence for $\bar{A}$) with initial point x. We are going to check that $\mathbb{P}$ has the same finite-dimensional distributions as $\mathbb{P}^x$. Given $g \in C(E)$ and $\lambda > 0$, there is $f \in \mathcal{D}(\bar{A})$ for which

$$(\lambda - \bar{A})f = g. \tag{15}$$

Since $\mathbb{P}$ solves the martingale problem, for $r < t$ we have

$$\mathbb{E}\left[f(\eta_t) - \int_r^t \bar{A}f(\eta_s)ds\middle|\mathcal{F}_r\right] = f(\eta_r),$$

where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$. Multiplying by $\lambda e^{-\lambda t}$ and integrating from r to infinity, we obtain

$$\begin{aligned}\mathbb{E}\left[\int_r^\infty \lambda e^{-\lambda t}f(\eta_t)dt - \int_r^\infty\int_r^t \lambda e^{-\lambda t}\bar{A}f(\eta_s)dsdt\middle|\mathcal{F}_r\right] &= \mathbb{E}\left[\int_r^\infty \lambda e^{-\lambda t}f(\eta_t)dt - \int_r^\infty e^{-\lambda s}\bar{A}f(\eta_s)ds\middle|\mathcal{F}_r\right] \\ &= e^{-\lambda r}f(\eta_r),\end{aligned}$$

which, by (15), can be written

$$\mathbb{E}\left[\int_r^\infty e^{-\lambda t}g(\eta_t)dt\middle|\mathcal{F}_r\right] = e^{-\lambda r}f(\eta_r). \tag{16}$$

Setting $r = 0$ in (16), and using the uniqueness theorem for Laplace transforms, we see that the one-dimensional distributions of $\mathbb{P}$ are the same as those for $\mathbb{P}^x$, since $\mathbb{P}^x$ is another solution to the martingale problem. We shall use induction to extend this to higher dimensional distributions. Take $0 < s_1 < s_2 < \cdots < s_n$, $\lambda_i > 0$ and $h_i \in C(E)$. Setting $r = s_n$ in (16), multiplying by

$$\exp\left(-\sum_{i=1}^n \lambda_i s_i\right)\prod_{i=1}^n h_i(\eta_{s_i}),$$

taking expectations of both sides and integrating we obtain

$$\begin{aligned}&\int\cdots\int_{0<s_1<\cdots<s_n<t} \exp\left(-\lambda t - \sum_{i=1}^n \lambda_i s_i\right)\mathbb{E}\left[g(\eta_t)\prod_{i=1}^n h_i(\eta_{s_i})\right]ds_1\cdots ds_n dt \\ &\qquad = \int\cdots\int_{0<s_1<\cdots<s_n} \exp\left(-\lambda s_n - \sum_{i=1}^n \lambda_i s_i\right)\mathbb{E}\left[f(\eta_{s_n})\prod_{i=1}^n h_i(\eta_{s_i})\right]ds_1\cdots ds_n.\end{aligned} \tag{17}$$

Applying uniqueness for multidimensional Laplace transforms to (17) (and the corresponding expression with $\mathbb{E}$ replaced by $\mathbb{E}^x$), we see that the equality of the n-dimensional distributions of $\mathbb{P}$ and $\mathbb{P}^x$ implies

that of the $(n+1)$-dimensional distributions. By induction, $\mathbb{P}$ and $\mathbb{P}^x$ have the same finite-dimensional distributions and so $\mathbb{P} = \mathbb{P}^x$ as required. □

Notice that we didn't use the Hille-Yosida Theorem. The key point was to be able to solve the resolvent equation (15).

Definition 3.4 (Well-posed martingale problem)**.** *Suppose that A is a Markov pregenerator. The martingale problem for A is said to be* well-posed *if, for each $x \in E$, the martingale problem for A with initial point x has a unique solution.*

Martingale problems provide a very powerful way to characterise the Markov processes associated with a given generator. Liggett (1985), which we have been following thus far, moves on to discuss martingale problems associated with interacting particle systems. We now depart from Liggett and turn to the setting in which martingale problems were first developed.

The rest of this section and the next owes a lot to lecture notes on stochastic calculus written by Nathanaël Berestycki of the University of Cambridge, which in turn follow Durrett (1996) and Stroock & Varadhan (1979).

3.2 Recap on Stochastic Differential Equations

Last term, you saw how to make sense of the notion of solution to a *stochastic differential equation* (SDE).

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t,$$

which is really shorthand for

$$X_t = X_0 + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dB_s \ , \tag{18}$$

where $\{B_t\}_{t\geq 0}$ is Brownian motion and the last term on the right is an Itô integral.

More precisely, let B be a Brownian motion in $\mathbb{R}^m$ with $m \geq 1$. Let $d \geq 1$ and suppose

$$\sigma(x) = \big(\sigma_{ij}(x)\big)_{\substack{1\leq i\leq d \\ 1\leq j\leq m}} : \mathbb{R}^d \to \mathbb{R}^{d\times m}$$

and

$$b(x) = \big(b_i(x)\big)_{1\leq i\leq d} : \mathbb{R}^d \to \mathbb{R}^d$$

are given Borel functions, bounded on compact sets. Consider the equation in $\mathbb{R}^d$:

$$dX_t = \sigma(X_t)\, dB_t + b(X_t)\, dt \ , \tag{19}$$

which may be written componentwise as

$$dX_t^i = \sum_{j=1}^m \sigma_{ij}(X_t)\, dB_t^j + b_i(X_t)\, dt \ , \quad 1 \leq i \leq d \ . \tag{20}$$

This general SDE will be called $E(\sigma, b)$.

Definition 3.5 (Solutions of stochastic differential equations)**.** *A* solution *to $E(\sigma, b)$ in (19) consists of*

(i) a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\big)$ satisfying the usual conditions;

(ii) an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $B = (B^1, \ldots, B^m)$ taking values in $\mathbb{R}^m$;

(iii) an $(\mathcal{F}_t)_{t\geq 0}$-adapted continuous process $X = (X^1, \ldots, X^d) \in \mathbb{R}^d$ such that

$$X_t = X_0 + \int_0^t \sigma(X_s)\, dB_s + \int_0^t b(X_s)\, ds.$$

When, in addition, $X_0 = x \in \mathbb{R}^d$, we say that X is a solution started from x.

There are several different notions of existence and uniqueness for SDE's.

Definition 3.6 (Uniqueness in law, pathwise uniqueness, strong solution)**.** *Let $E(\sigma, b)$ be the SDE in (19).*

(i) We say that $E(\sigma, b)$ has a solution *if for all $x \in \mathbb{R}^d$, there exists a solution to the SDE started from x.*

(ii) There is uniqueness in law *if all solutions to $E(\sigma, b)$ started from x have the same distribution.*

(iii) There is pathwise uniqueness *if, when we fix $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ and B then any two solutions X and X' satisfying $X_0 = X'_0$ a.s. are indistinguishable from one another (that is their trajectories coincide a.s.).*

(iv) We say that a solution X of $E(\sigma, b)$ started from x is a strong solution *if X is adapted to the natural filtration of B.*

Remark 3.7 (Weak solution)**.** *In general, $\sigma(B_s, s \leq t) \subseteq \mathcal{F}_t$ and a solution might not be measurable with respect to the Brownian motion B. A strong solution depends only on $x \in \mathbb{R}^d$ and the Brownian motion B, and is moreover non-anticipative: if the path of B is known up to time t, then so is the path of X up to time t. We will term* weak *any solution that is not strong.*

Since we are primarily interested in relating SDEs to martingale problems, we shall be concerned with weak solutions and uniqueness in law, but for completeness we make some remarks on the relationship between the various notions of existence and uniqueness.

Remark 3.8. *If every solution is strong, then pathwise uniqueness holds. Indeed, any solution must then be a certain measurable functional of the path B. If two functionals F_1 and F_2 of B gave two solutions to the SDE, then we would construct a third one by tossing a coin and choosing X_1 or X_2. This third solution would then not be adapted to $\mathcal{F}^B$.*

It is possible to have existence of a weak solution and uniqueness in law, without pathwise uniqueness (see the second problem sheet).

Theorem 3.9 (Yamada-Watanabe Theorem)**.** *Let σ, b be measurable functions. If pathwise uniqueness holds for $E(\sigma, b)$ and there exist solutions, then there is also uniqueness in law. In this case, for every filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and every $\mathcal{F}_t$-Brownian motion $B = (B_t, t \geq 0)$, and for every $x \in \mathbb{R}^d$, the unique solution X to $E_x(\sigma, b)$ is strong.*

In particular, provided that there exist solutions, pathwise uniqueness is stronger than uniqueness in law.

Lipschitz coefficients

Last term you were primarily concerned with SDEs with *Lipschitz* coefficients.

Definition 3.10 (Lipschitz continuous function)**.** *For $U \subseteq \mathbb{R}^d$ and $f : U \to \mathbb{R}^d$, we say that f is* Lipschitz *with* Lipschitz constant $K < \infty$ *if*

$$\big|f(x) - f(y)\big| \leq K|x - y| \quad \text{for all } x, y \in U \ , \tag{21}$$

where $|.|$ denotes the Euclidean norm on $\mathbb{R}^d$. (If $f : U \to \mathbb{R}^{d\times m}$ then the left-hand side is the Euclidean norm in $\mathbb{R}^{d\times m}$).

The key result which you proved last term is that SDE's with Lipschitz coefficients have pathwise unique solutions which are furthermore always strong.

Finally we note that weak uniqueness implies the strong Markov property.

3.3 Diffusion processes and martingale problems

Lévy (1948) characterised standard Brownian motion as the unique continuous process B such that $B(t)$ and $B^2(t) - t$ are both martingales. Our next aim is to find an analogous characterisation of the diffusion processes which arise as solutions of stochastic differential equations and relate it to the notion of martingale problem.

We shall sometimes be sloppy and say that a process solves a martingale problem, when really we mean that its law does. We shall also sometimes fail to specify the initial value in a martingale problem, in which case saying that a solution exists means that a solution exists for any choice of initial value.

Let $\sigma_{i,j}(x)_{1\leq i,j\leq d}$ and $(b_i(x))_{1\leq i\leq d}$ be a family of measurable functions with values in $\mathbb{R}$. Let $a(x) = \sigma(x)\sigma^T(x)$. (Here we assume for simplicity that $m = d$).

Definition 3.11 (Martingale problem $\mathbf{M}(a, b)$)**.** *We say that a process $X = (X_t, t \geq 0)$ with values in $\mathbb{R}^d$, together with a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, solves the* martingale problem $\mathbf{M}(a, b)$ *if for all $1 \leq i, j \leq d$,*

$$M^i = \left(X^i_t - \int_0^t b_i(X_s)ds; t \geq 0\right)$$

and

$$\left(M^i_t M^j_t - \int_0^t a_{ij}(X_s)ds; t \geq 0\right)$$

are local martingales.

In particular, when $d = 1$ the statements become

$$M_t = X_t - \int_0^t b(X_s)ds$$

and

$$M^2_t - \int_0^t \sigma^2(X_s)ds$$

are local martingales.

If σ, b are Lipschitz continuous, then the solution to the SDE (18) exists, that is, there is a process X with

$$X_t - X_0 = \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s.$$

The stochastic integral is a local martingale and so

$$M_t = X_t - X_0 - \int_0^t b(X_s)ds$$

is a mean zero local martingale. Squaring and using the Itô isometry for stochastic integrals, yields

$$M_t^2 - \int_0^t \sigma^2(X_s)ds$$

is a local martingale. In other words, X then solves the martingale problem $\mathbf{M}(a, b)$. Rather remarkably, as the next theorem shows, these are the only solutions.

Theorem 3.12. *Let $a = \sigma\sigma^T$ and let X and $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a solution to $\mathbf{M}(a, b)$. Then there exists an $(\mathcal{F}_t)$-Brownian motion $(B_t, t \geq 0)$ in $\mathbb{R}^d$ defined on an enlarged probability space, such that (X, B) solves $E(\sigma, b)$.*

Sketch of proof: We satisfy ourselves with a sketch of the proof in $d = 1$. For simplicity, we suppose that $\sigma(x) > \delta > 0$ for all $x \in \mathbb{R}$. We have that

$$M_t = X_t - X_0 - \int_0^t b(X_s)ds$$

is a local martingale, as is

$$M_t^2 - \int_0^t \sigma^2(X_s)ds.$$

Since $\sigma(X_s)$ is bounded away from zero, we can define

$$B_t = \int_0^t \frac{1}{\sigma(X_s)} dM_s$$

and since $d[B_t] = dt$, by Lévy's characterisation of Brownian motion, B is a Brownian motion on $\mathbb{R}$. Moreover,

$$\int_0^t \sigma(X_s)dB_s = M_t - M_0 = X_t - \int_0^t b(X_s)ds. \tag{22}$$

But (22) is simply the statement that (X, B) solves $E(\sigma, b)$. □

Theorem 3.12 shows that there is a one-to-one correspondence between solutions to the stochastic differential equation $E(\sigma, b)$ and the martingale problem $\mathbf{M}(a, b)$. In particular, there is *uniqueness in distribution* to the solutions of $E(\sigma, b)$, if and only if the solutions to the martingale problem $\mathbf{M}(a, b)$ are unique, where uniqueness means that all solutions to $\mathbf{M}(a, b)$ with identical starting points have the same law.

Of course we should connect this version of the martingale problem to the one that we introduced in Definition 3.2. It is clear from Itô's formula that any solution to the SDE (18) will provide a solution to the martingale problem corresponding to the infinitesimal generator A given by

$$Af(x) = \frac{1}{2}\sum_{ij} a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_i b_i(x)\frac{\partial f}{\partial x_i}(x). \tag{23}$$

In one dimension this takes the form

$$Af(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x).$$

We shall call b and a the *infinitesimal drift* and *infinitesimal variance* of the diffusion. That a solution to the martingale problem for A is also a solution to the martingale problem $\mathbf{M}(a,b)$ is the content of the next lemma.

Lemma 3.13. *A solution (in the sense of Definition 3.2) to the martingale problem for A given by (23) is also a solution to the martingale problem* $\mathbf{M}(a,b)$.

Proof. For simplicity we provide a sketch of the proof in one dimension only.

Taking $f(x) \equiv x$, it is immediate that

$$M_t = X_t - \int_0^t b(X_s)ds$$

is a local martingale. Now consider M_t^2.

$$\begin{aligned} M_t^2 &= \left(X_t - \int_0^t b(X_s)ds\right)^2 \\ &= X_t^2 - 2X_t\int_0^t b(X_s)ds + \left(\int_0^t b(X_s)ds\right)^2 \\ &= X_t^2 - 2X_t\int_0^t b(X_s)ds + \int_0^t 2b(X_s)\int_s^t b(X_u)duds \\ &= X_t^2 - 2\int_0^t b(X_s)\left(X_t - \int_s^t b(X_u)du\right)ds. \end{aligned}$$

Now setting $f(x) = x^2$ we have that

$$X_t^2 - \int_0^t \{2X_s b(X_s) + a(X_s)\}\, ds$$

is a local martingale, so we will be able to deduce that

$$M_t^2 - \int_0^t a(X_s)ds$$

is a local martingale if we can show that

$$2\int_0^t b(X_s)\left(X_t - \int_s^t b(X_u)du\right)ds - 2\int_0^t X_s b(X_s)ds = 2\int_0^t b(X_s)\left(X_t - X_s - \int_s^t b(X_u)du\right)ds$$

is a local martingale. To see this, we use the tower property of conditional expectations. Suppose that $0 < r < t$. First observe that

$$\begin{aligned} \mathbb{E}\left[\int_r^t b(X_s)\left(X_t - X_s - \int_s^t b(X_u)du\right)ds\middle|\mathcal{F}_r\right] &= \mathbb{E}\left[\int_r^t \mathbb{E}\left[b(X_s)\left(X_t - X_s - \int_s^t b(X_u)du\right)\middle|\mathcal{F}_s\right]ds\middle|\mathcal{F}_r\right] \\ &= \mathbb{E}\left[\int_r^t b(X_s)\mathbb{E}\left[\left(X_t - X_s - \int_s^t b(X_u)du\right)\middle|\mathcal{F}_s\right]ds\middle|\mathcal{F}_r\right] \\ &= 0, \end{aligned}$$

since M_t is a local martingale. Now note that

$$\begin{aligned} \mathbb{E}\left[\int_0^r b(X_s)\left(X_t - X_s - \int_s^t b(X_u)du\right)ds\middle|\mathcal{F}_r\right] &= \int_0^r b(X_s)\left\{\mathbb{E}\left[\left(X_t - X_r - \int_r^t b(X_u)du\right)\middle|\mathcal{F}_r\right]\right. \\ &\qquad \left. + X_r - X_s - \int_s^r b(X_u)du\right\}ds \\ &= \int_0^r b(X_s)\left(X_r - X_s - \int_s^r b(X_u)du\right)ds, \end{aligned}$$

as required (where again we have used that M_t is a local martingale).

Remark 3.14. *The infinitesimal generators corresponding to solutions to $E(\sigma, b)$ are the second order differential operators given in (23) where we assume that σ and b are Lipschitz. This condition is stronger than we need for the process to exist, as we have already seen in Example 2.36, in which $\sigma(x) = \sqrt{x(1-x)}$ is certainly not uniformly Lipschitz. In that example, we also allowed $\sigma^2(x)$ to vanish on the boundary. This is crucial in many applications, but to avoid pathologies we shall* only *allow σ to vanish on the boundary, and never on the interior, of the domain.*

4 Stroock-Varadhan theory of diffusion approximation

For section C students: proofs of results in this section are *not* examinable. For the exam, I expect you to be able to state and apply the results of Section 4.2.

4.1 Notions of weak convergence of processes

In this section we describe (without proof) some basic results in the theory of weak convergence of processes. We will use these in the next section to derive conditions under which Markov chains (suitably rescaled) converge to (weak) solutions of stochastic differential equations.

This subsection is challenging. Try to get a feel for the results, but for the rest of the course we shall concentrate on the case where the limiting process has continuous paths, and for that we really only need Definition 4.10 and Theorem 4.11.

We view a process as a random variable with values in a space D of trajectories (paths) and so it is convenient to deal with weak convergence in a general metric space.

Definition 4.1 (Complete, separable, metric space, Borel σ-algebra)**.** *A metric space (S, d) consists of a nonempty set S and a map $d : S \times S \to \mathbb{R}$ such that:*

(i) $d(x, y) = 0$ *if and only if* $x = y$*;*

(ii) $d(x, y) \geq 0$*;*

(iii) $d(x, z) \leq d(x, y) + d(y, z)$*.*

The open ball $B(x, r)$ is the set $\{y \in S : d(x, y) < r\}$. The Borel σ-algebra *is the σ-algebra generated by all open sets.*

We say that (S, d) is separable *if it contains a countable dense subset and* complete *if every Cauchy sequence in (S, d) converges to a limit in S.*

When we talk about measurability, we shall implicitly assume that we are using the Borel σ-algebra. The notion of convergence in distribution is defined in terms of integration against test functions.

Definition 4.2 (Weak convergence, convergence in distribution/law)**.** *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability distributions on S. We say that $\mu_n \to \mu$* weakly *as $n \to \infty$, if $\int_S f d\mu_n \to \int_S f d\mu$ as $n \to \infty$ for all bounded continuous functions f. If μ_n is the law of a random variable $X^{(n)}$ and μ that of a random variable X, we say that $X^{(n)} \to X$* in distribution *(or* in law*).*

Remark: What a probabilist calls weak convergence is what an analyst calls weak* convergence.

There are a number of ways one can reformulate the notion of weak convergence in terms of the mass assigned to events that are either closed or open.

Theorem 4.3 (Portmanteau Theorem). *Let $(X^{(n)})_{n\geq 1}$ be a sequence of random variables taking values in S. The following are equivalent.*

(i) *$X^{(n)} \to X$ in distribution.*

(ii) *For any closed set $K \subseteq S$, $\limsup_{n\to\infty} \mathbb{P}(X^{(n)} \in K) \leq \mathbb{P}(X \in K)$.*

(iii) *For any open set $O \subseteq S$, $\liminf_{n\to\infty} \mathbb{P}(X^{(n)} \in O) \geq \mathbb{P}(X \in O)$.*

(iv) *For all Borel sets A such that $\mathbb{P}[X \in \partial A] = 0$, $\lim_{n\to\infty} \mathbb{P}[X^{(n)} \in A] = \mathbb{P}[X \in A]$.*

(v) *For any bounded function f, denote by D_f the set of discontinuities of f. Then for any f such that $\mathbb{P}[X \in D_f] = 0$, $\mathbb{E}[f(X^{(n)})] \to \mathbb{E}[f(X)]$ as $n \to \infty$.*

The notion of weak convergence makes perfectly good sense even when the random variables $X^{(n)}$ are not related to one another in any particular way. Indeed they don't even have to be defined on the same probability space. However, it turns out that provided the space (S, d) is sufficiently nice, one can find a single probability space and a collection of random variables $Y^{(n)}$ on that space with the property that, for each n, $Y^{(n)}$ has the same law as $X^{(n)}$ and the sequence $Y^{(n)}$ converges *almost surely*. This is the content of the following theorem (which we include for completeness).

Theorem 4.4 (Skorokhod Representation Theorem). *Suppose that (S, d) is a complete, separable metric space. If $\mu_n \to \mu$ weakly then there exist random variables $Y^{(n)}$ defined on $[0, 1]$ equipped with the Lebesgue measure $\mathbb{P}$, such that $Y^{(n)} \stackrel{d}{=} \mu_n$ for all $n \geq 1$, and $\lim_{n\to\infty} Y^{(n)} = Y$, $\mathbb{P}$-almost surely, where $Y \stackrel{d}{=} \mu$.*

A standard strategy for proving that a sequence x_n converges to a limit x is to prove that (a) the sequence takes its values in a compact set and (b) there is only one possible subsequential limit. In the context of stochastic processes, the analogous strategy is to first check that the laws of the processes are *tight* and then to prove uniqueness of the limit point.

Definition 4.5 (Tightness). *Let (S, d) be a metric space. A family of probability measures $M \subseteq \mathcal{P}(S)$ is said to be* tight *if for each $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subseteq S$ such that for all $\mu \in M$, $\mu(K_\varepsilon) > 1 - \varepsilon$.*

By Prokhorov's theorem, tightness implies that the closure of M is sequentially compact in the topology of weak convergence, and is equivalent to relative compactness if (S, d) is complete and separable. Therefore, granted tightness, any sequence in M will contain a convergent subsequence.

In order to apply this notion to the càdlàg stochastic processes, we are required to turn $D([0, \infty), E)$, the space of càdlàg paths with values in E, into a metric space.

For any complete separable metric space (E, d), one can define the Skorokhod metric on the space $D([0, \infty), E)$, and this space will again be complete and separable.

Definition 4.6 (The Skorokhod topology). *Let (E, d) be a metric space, and let $\{\eta_n\}_{n\geq 1}$ be a sequence and η a point in $D([0, \infty), E)$. Then $\eta_n \to \eta$ as $n \to \infty$ in the* Skorokhod topology *if and only if, whenever t_n is a sequence of points in $[0, \infty)$ converging to $t \in [0, \infty)$, the following conditions are satisfied.*

(i)

$$\lim_{n\to\infty} \min\{d(\eta_n(t_n), \eta(t)), d(\eta_n(t_n), \eta(t-))\} = 0$$

(ii) If $\lim_{n\to\infty} d(\eta_n(t_n), \eta(t)) = 0$*, then for any sequence of times* s_n *with* $s_n \geq t_n$ *and* $\lim_{n\to\infty} s_n = t$*,*

$$\lim_{n\to\infty} d(\eta_n(s_n), \eta(t)) = 0.$$

(iii) If $\lim_{n\to\infty} d(\eta_n(t_n), \eta(t-)) = 0$*, then for any sequence of times* s_n *with* $s_n \leq t_n$ *and* $\lim_{n\to\infty} s_n = t$*,*

$$\lim_{n\to\infty} d(\eta_n(s_n), \eta(t-)) = 0.$$

This is not the usual definition of the Skorokhod topology, but it is equivalent (see Ethier & Kurtz, Proposition 3.6.5). This form has the advantage of (relative) transparency: the first condition ensures that there are at most two limit points of the sequence $\{\eta_n(t_n)\}$, the second condition guarantees right continuity of the limiting path η at t, the third left limits.

The following result, Theorem 3.7.2 of Ethier & Kurtz, demonstrates the nature of the hurdles to be overcome in proving tightness of a sequence of processes with càdlàg sample paths. It requires some notation. Let us suppose that (E, d) is complete and separable. For a path $x \in D([0,\infty), E)$, define

$$w'(x, \delta, T) = \inf_{\{t_i\}} \max_i \sup_{s,t\in[t_{i-1},t_i)} d(x(s), x(t)),$$

where $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \ldots < t_m = T$ with $m \geq 1$ and $\min_{1\leq i\leq m}(t_i - t_{i-1}) > \delta$. The rôle of w' will become clear soon.

For a subset $K \subseteq E$, let $K^\varepsilon = \{x \in E : \inf_{y\in K} d(x, y) < \varepsilon\}$ denote the ε-expansion of K.

Theorem 4.7 (Ethier & Kurtz)**.** *Let* $\{X^{(n)}\}_{n\geq 1}$ *be a sequence of processes taking values in* $D([0,\infty), E)$*. Then* $\{X^{(n)}\}_{n\geq 1}$ *is relatively compact if and only if the following two conditions hold.*

(i) For every $\varepsilon > 0$ *and every (rational)* $t \geq 0$*, there exists a compact set* $\gamma_{\varepsilon,t} \subseteq E$*, such that*

$$\liminf_{n\to\infty} \mathbb{P}\left[X^{(n)}(t) \in \gamma^\varepsilon_{\varepsilon,t}\right] \geq 1 - \varepsilon \tag{24}$$

(ii) For every $\varepsilon > 0$ *and* $T > 0$*, there exists* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} \mathbb{P}\left[w'(X^{(n)}, \delta, T) \geq \varepsilon\right] \leq \varepsilon \tag{25}$$

The first condition could instead be replaced by the requirement of relative compactness of the one-dimensional distributions at each fixed time. The second condition prevents large jumps from being too close together. That is, for any $\varepsilon > 0$, there won't be a limit point of jumps of size greater than ε. (If the jumps are sufficiently 'spread out' then we can choose the partition in such a way that each jump point corresponding to a jump of size more than ε is a point t_i. If they are not, then no matter how small δ, we cannot arrange for $w'(x, \delta, T)$ to be less than ε.)

This condition is rather unwieldy to work with. If we are dealing with real-valued processes, then the following criterion due to Aldous, is extremely helpful.

Theorem 4.8 (Aldous (1978))**.** *Let* $Y^{(n)}$ *be a sequence of real valued processes with càdlàg paths. Suppose that the following conditions are satisfied.*

(i) For each fixed t*,* $Y^{(n)}_t$ *are tight.*

(ii) Given a sequence of stopping times τ_n*, bounded by* T*, for each* $\varepsilon > 0$ *there exists* $\delta > 0$ *and* n_0 *such that*

$$\sup_{n\geq n_0} \sup_{\theta\in[0,\delta]} \mathbb{P}\left[\left|Y^{(n)}(\tau_n + \theta) - Y^{(n)}(\tau_n)\right| > \varepsilon\right] \leq \varepsilon. \tag{26}$$

Then the $Y^{(n)}$ are tight.

If the $Y^{(n)}$ are semimartingales, then we write $V^{(n)}$ for the corresponding predictable finite variation process and $[M^{(n)}]$ for the quadratic variation of the martingale part. The following version of Theorem 4.8 can be found in Rebolledo (1980).

Theorem 4.9 (The Aldous-Rebolledo Criterion)**.** *Let $Y^{(n)}$ be a sequence of real valued semimartingales with càdlàg paths. Suppose that the following conditions are satisfied.*

(i) For each fixed t, $Y_t^{(n)}$ are tight.

(ii) Given a sequence of stopping times τ_n, bounded by T, for each $\varepsilon > 0$ there exists $\delta > 0$ and n_0 such that

$$\sup_{n \geq n_0} \sup_{\theta \in [0,\delta]} \mathbb{P}\left[\left|V^{(n)}(\tau_n + \theta) - V^{(n)}(\tau_n)\right| > \varepsilon\right] \leq \varepsilon, \tag{27}$$

and

$$\sup_{n \geq n_0} \sup_{\theta \in [0,\delta]} \mathbb{P}\left[\left|[M^{(n)}]_{\tau_n+\theta} - [M^{(n)}]_{\tau_n}\right| > \varepsilon\right] \leq \varepsilon. \tag{28}$$

Then the $Y^{(n)}$ are tight.

This is more generally applicable than it looks. For example, provided that one can prove a compact containment condition, tightness of measure-valued processes is implied by that of the real-valued processes obtained by integrating against test functions from a large enough class.

So far in this section we have considered processes with càdlàg paths. If we specialise to the case where the random variables $X^{(n)}$ take values in the space C of continuous paths over a compact time interval, $[0, 1]$ for example, then things are simpler. For definiteness (and since it is all we use in what follows), let us also specialise to $E \subseteq \mathbb{R}^d$. We equip C with the distance induced by the sup-norm:

$$d(f, g) = \|f - g\|_\infty = \sup_{t \in [0,1]} |f(t) - g(t)|. \tag{29}$$

This turns C into a complete, separable metric space, which is continuously embedded in the space D. Whereas in the case of càdlàg paths, we used the quantity $w'(x, \delta, T)$ to control the accumulation of 'large jumps', for continuous paths, we can work with a much simpler quantity:

Definition 4.10 (Modulus of continuity at precision δ, $\mathrm{osc}_\delta(\omega)$)**.** *For a continuous path $\omega(t), t \in [0, 1]$, let*

$$\mathrm{osc}_\delta(\omega) = \sup\{|\omega(s) - \omega(t)| : s, t \in [0, 1], |s - t| \leq \delta\}.$$

osc_δ is simply the modulus of continuity of the path ω, at precision δ.

Theorem 4.11. *Suppose that $\{X^{(n)}\}_{n\geq 1}$ is a sequence of processes with values in C. Then $\{X^{(n)}\}_{n\geq 1}$ is tight, if and only if for each $\varepsilon > 0$, there exist $n_0, M \geq 1$ and $\delta > 0$ such that:*

(i) $\mathbb{P}[|X^{(n)}(0)| > M] \leq \varepsilon$ *for all* $n \geq n_0$.

(ii) $\mathbb{P}^{(n)}[\mathrm{osc}_\delta(X^{(n)}) > \varepsilon] \leq \varepsilon$ *for all* $n \geq n_0$.

Thus, to show that a sequence converges weakly in C, it suffices to prove that the conditions of Theorem 4.11 are satisfied and that there is a unique weak subsequential limit. This uniqueness is for instance the case if one has already established convergence of the finite-dimensional distributions, i.e., convergence of the k-dimensional vector $(X_{t_1}^{(n)}, \ldots, X_{t_k}^{(n)})$ towards $(X_{t_1}, \ldots, X_{t_k})$, for any $k \geq 1$ and

any choice of 'test times' $t_1, \ldots, t_k$. The first condition of Theorem 4.11 says that the initial value of the process, $X^{(n)}(0)$, takes values in a compact set with arbitrarily high probability. This is of course trivial in the case when the starting point of a process is a deterministic point.

In the next section, we will prove weak convergence of certain rescaled Markov chains towards diffusion processes. For this, we will usually use the fact that any weak subsequential limit must satisfy the associated martingale problem $\mathbf{M}(a, b)$ for which sufficient smoothness of the coefficients proves uniqueness in distribution. Although the limiting processes will be continuous, the approximating Markov chains are not. We could circumvent this by linear interpolation, but then we lose the Markov property, and so we prefer to work in the space D. Evidently if we can check the second condition of Theorem 4.11 for these *càdlàg* processes, then the second condition of Theorem 4.7 follows *a fortiori*. Moreover, combining this estimate on the 'oscillations' of the process with the first condition of Theorem 4.11 is enough to guarantee that the first condition of Theorem 4.7 is also satisfied. Thus it will suffice to check the conditions of Theorem 4.11 for our càdlàg processes and, in this case, any subsequential limit will actually be a continuous process (in the sense that $\mathbb{P}[X \in C] = 1$). Furthermore, weak convergence with respect to the Skorokhod topology towards a continuous process X, implies weak convergence in C of the linear interpolations. We shall also use the fact that if $x_n \to x$ in the Skorokhod topology, then $x_n(t) \to x(t)$ for all $t \geq 0$.

4.2 Markov chains and diffusions

The result which we now discuss is due to Stroock and Varadhan (1969). It can be found in Chapter 11 of Stroock & Varadhan (1979). It shows how to obtain diffusion processes as limits of sequences of Markov chains and is applicable in a remarkably wide variety of contexts. Our treatment follows rather closely the book of Durrett.

We take a sequence of Markov chains Y^h indexed by a parameter $h > 0$. We shall focus on discrete time chains, but see Remark 4.13 below.

The chain Y^h takes its values in a certain set $E_h \subseteq \mathbb{R}^d$. The transition probabilities of Y^h are given by a transition kernel Π_h,

$$\mathbb{P}[Y^h_{n+1} \in A | Y^h_n = x] = \Pi_h(x, A).$$

We define the random process X^h on $[0, 1]$ by

$$X^h_t = Y^h_{\lfloor t/h \rfloor}, \qquad t \in [0, 1],$$

so that X^h is almost surely right-continuous and is constant between successive jumps, which occur every h units of time. We write K_h for the rescaled transition kernel:

$$K_h(x, dy) = \frac{1}{h}\Pi_h(x, dy).$$

For $1 \leq i, j \leq d$, define:

$$a^h_{ij}(x) = \int_{|y-x|\leq 1} (y_i - x_i)(y_j - x_j) K_h(x, dy), \tag{30}$$

$$b^h_i(x) = \int_{|y-x|\leq 1} (y_i - x_i) K_h(x, dy), \tag{31}$$

and

$$\Delta^h_\varepsilon(x) = K_h(x, B(x, \varepsilon)^c).$$

Suppose that the coefficients a_{ij} and b_i are continuous functions on $\mathbb{R}^d$ for which the martingale problem $\mathbf{M}(a,b)$ is *well posed*, i.e. for each $x \in \mathbb{R}^d$ there is a unique in distribution process $(X_t, 0 \le t \le 1)$ such that $X_0 = x$ almost surely, and

$$M_t^i = X_t^i - \int_0^t b_i(X_s)ds \quad \text{and} \quad M_t^i M_t^j - \int_0^t a_{ij}(X_s)ds$$

are all local martingales.

Theorem 4.12. *Under the assumptions above, suppose further that for each $1 \le i,j \le d$, and for every $R > 0$, $\varepsilon > 0$,*

(i) $\lim_{h\to 0} \sup_{|x|\le R} |a_{ij}^h(x) - a_{ij}(x)| = 0.$

(ii) $\lim_{h\to 0} \sup_{|x|\le R} |b_i^h(x) - b_i(x)| = 0.$

(iii) $\lim_{h\to 0} \sup_{|x|\le R} \Delta_\varepsilon^h(x) = 0.$

Then if $X_0^h = x_h \to x_0$, we have $(X_t^h, 0 \le t \le 1) \to (X_t, 0 \le t \le 1)$ weakly in D, where X_t solves the martingale problem $\mathbf{M}(a,b)$, and in particular, the linear interpolations of Y^h converge weakly in C.

Remark 4.13 (Convergence of continuous time chains)**.** *We have stated the result for discrete time Markov chains. In fact, Durrett treats discrete and continuous time chains simultaneously. Suppose that we have continuous time chains $\{X_t^h\}_{t\ge 0}$ taking values in $S_h \subseteq \mathbb{R}$. In place of the sequence of transition probabilities Π_h for the discrete time chain, we have a sequence of transition rates:*

$$\frac{d}{dt}\mathbb{P}[X_t^h \in A | X_0^h = x]\Big|_{t=0} = Q_h(x,A), \qquad \text{for } x \in S_h, A \subseteq \mathbb{R}^d, x \notin A.$$

We assume that for any compact set K,

$$\sup_{x\in K} Q_h(x, \mathbb{R}^d \backslash \{x\}) < \infty. \tag{32}$$

Let us write

$$K_h(x, dy) = Q_h(x, dy).$$

Then the statement of Theorem 4.12 remains valid.

The rest of this subsection is devoted to the proof of Theorem 4.12. The idea is simple. The first two conditions ensure convergence of the infinitesimal mean and variance to those of the diffusion process. The third guarantees that there are no macroscopic jumps in the limit. By localization, see Durrett (1996), Section 8.7(c) p.304, we can, and do, replace (i), (ii) and (iii) by the following stronger conditions:

(i) $\lim_{h\to 0} \sup_{x\in\mathbb{R}^d} |a_{ij}^h(x) - a_{ij}(x)| = 0.$

(ii) $\lim_{h\to 0} \sup_{x\in\mathbb{R}^d} |b_i^h(x) - b_i(x)| = 0.$

(iii) $\lim_{h\to 0} \sup_{x\in\mathbb{R}^d} \Delta_\varepsilon^h(x) = 0.$

(iv) Moreover, a_{ij}^h, b_i^h, Δ_ε^h are uniformly bounded in h and x.

Step 1. Tightness

Let f be a bounded and measurable function. Define the operator L^h by

$$L^h f(x) = \int K_h(x, dy)(f(y) - f(x)). \tag{33}$$

This plays the rôle of the infinitesimal generator for our discrete time process. In particular, the process

$$f(Y_k^h) - \sum_{j=0}^{k-1} hL^h f(Y_j^h), \qquad k = 1, 2, \ldots \tag{34}$$

is a (discrete-time) martingale. For our proof of tightness we are going to need an estimate on the time needed by the chain to make a deviation of size roughly $\varepsilon > 0$, when it starts at position $y \in \mathbb{R}^d$. To do this, we introduce a function $g \in \mathcal{C}^2(\mathbb{R})$ (the twice continuously differentiable functions on $\mathbb{R}$) with $0 \le g \le 1$, $g(0) = 1$ and $g(x) = 0$ for $x \ge 1$. For $x \in \mathbb{R}^d$, set $f_\varepsilon(x) = g(|x|^2/\varepsilon^2)$ which is also $\mathcal{C}^2$, and vanishes for $|x| \ge \varepsilon$. Finally, for $a \in \mathbb{R}^d$, let $f_{a,\varepsilon}(x) = f_\varepsilon(x - a)$.

Here, and elsewhere, we shall use $D_i f$ to denote $\frac{\partial f}{\partial x_i}$ and $D_{ij} f$ for $\frac{\partial^2 f}{\partial x_i \partial x_j}$.

Lemma 4.14. *There exists $C_\varepsilon < \infty$, independent of h, such that $|L^h f_{a,\varepsilon}(x)| \le C_\varepsilon$ for all $a, x \in \mathbb{R}^d$.*

Proof. This is simply an application of Taylor's theorem. For $x, y \in \mathbb{R}^d$, there exists $c_{xy} \in [0, 1]$ such that

$$f_{a,\varepsilon}(y) - f_{a,\varepsilon}(x) = \sum_{i=1}^d (y_i - x_i) D_i f_{a,\varepsilon}(x) + \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) D_{ij} f_{a,\varepsilon}(z_{xy}),$$

where $z_{xy} = x + c_{xy}(y - x) \in [x, y]$ (the straight line segment joining x and y).

To obtain $L^h f_{a,\varepsilon}(x)$, we integrate with respect to $K_h(x, dy)$ which yields the bound

$$\begin{aligned}
\left| L^h f_{a,\varepsilon}(x) \right| &= \left| \int_{y \in \mathbb{R}^d} K_h(x, dy)(f_{a,\varepsilon}(y) - f_{a,\varepsilon}(x)) \right| \\
&\le \left| \nabla f_{a,\varepsilon}(x) \cdot \int_{|y-x| \le 1} (y - x) K_h(x, dy) \right| \\
&\quad + \left| \int_{|y-x| \le 1} \sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f_{a,\varepsilon}(z_{xy}) K_h(x, dy) \right| \\
&\quad + 2\|f_{a,\varepsilon}\|_\infty K_h(x, B(x, 1)^c).
\end{aligned}$$

Let $A_\varepsilon = \sup_x |\nabla f_{a,\varepsilon}(x)|$ and $B_\varepsilon = \sup_z \|D f_{a,\varepsilon}(z)\|$, where $Df = (D_{ij} f)_{1 \le i,j \le d}$ is the Hessian matrix of f and for a $d \times d$ matrix M, the matrix norm is defined by

$$\|M\| := \sup_{u \in \mathbb{R}^d : |u| = 1} |\langle u, Mu \rangle|.$$

In this notation

$$\left| \sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f_{a,\varepsilon}(z_{xy}) \right| \le \|y - x\|^2 B_\varepsilon.$$

Hence

$$L^h f_{a,\varepsilon}(x) \le A_\varepsilon |b^h(x)| + B_\varepsilon \int_{|y-x| \le 1} |y - x|^2 K_h(x, dy) + 2K_h(x, B(x, 1)^c).$$

Since $\int_{|y-x|\le 1} |y-x|^2 K_h(x,dy) = \sum_i a_{ii}^h(x)$, using the uniform boundedness assumptions of (iv), the lemma is proved. □

To estimate $\mathrm{osc}_\delta(X^h)$, we first define a sequence of stopping times $\{\tau_n\}_{n\ge 0}$. Set $\tau_0 = 0$ and then, inductively,

$$\tau_n = \inf\{t \ge \tau_{n-1} : |X_t^h - X_{\tau_{n-1}}^h| \ge \varepsilon\}.$$

Now set

$$N = \min\{n : \tau_n > 1\},$$
$$\sigma = \min\{\tau_n - \tau_{n-1} : 1 \le n \le N\}$$

and, finally

$$\theta = \max\{|X^h(t) - X^h(t-)| : 0 < t \le 1\}.$$

The relationship between these random variables and tightness is provided by the following lemma:

Lemma 4.15. *Assume that $\sigma > \delta$ and that $\theta < \varepsilon$. Then $\mathrm{osc}_\delta(X^h) \le 4\varepsilon$.*

Proof. The proof is straightforward. We want to show that for all $s, t \in [0,1]$ with $|s-t| \le \delta$, $|X^h(s) - X^h(t)| \le 4\varepsilon$. The point is that since $|s-t| \le \delta < \sigma$, s and t can only span at most one of the intervals $[\tau_{n-1}, \tau_n]$, and the definition of the stopping times τ_n ensures that X^h does not vary by more than 2ε on such an interval. More explicitly, if $\tau_{n-1} \le s < t < \tau_n$, then $|X^h(s) - X^h(t)| \le 2\varepsilon$. If on the other hand, $\tau_{n-1} \le s < \tau_n \le t$, then

$$\begin{aligned} |X^h(s) - X^h(t)| &\le |X^h(s) - X^h(\tau_{n-1})| + |X^h(t) - X^h(\tau_n)| \\ &\quad + |X^h(\tau_n) - X^h(\tau_n-)| + |X^h(\tau_n-) - X^h(\tau_{n-1})| \\ &\le 4\varepsilon. \end{aligned}$$

□

We now check the conditions of Theorem 4.11. Since it is assumed that the starting point $X_0^h = x_0^h$ is nonrandom and converges towards a fixed x_0, it suffices to check the second condition: for all $\varepsilon > 0$, there exists $\delta > 0$ and h_0 such that for $h \le h_0$,

$$\mathbb{P}[\mathrm{osc}_\delta(X^h) \ge \varepsilon] \le \varepsilon.$$

From Lemma 4.15, this will follow if we show that for all δ sufficiently small and all $x \in \mathbb{R}^d$,

$$\mathbb{P}_x[\theta > \varepsilon/4] < \varepsilon/2 \quad \text{and} \quad \mathbb{P}_x[\sigma < \delta] < \varepsilon/2 \text{ as } h \to 0.$$

The first statement is very simple: since there are at most $1/h$ time steps in the interval $[0,1]$, a simple union bound yields

$$\mathbb{P}_x[\theta > \varepsilon] \le \frac{1}{h} \sup_y \Pi_h(y, B(y,\varepsilon)^c) \le \sup_y \Delta_\varepsilon^h(y) \to 0$$

by (iii), so certainly for h sufficiently small $\mathbb{P}_x[\theta > \varepsilon/4] < \varepsilon/2$. The second claim requires more work. The first step is to estimate $\mathbb{P}_x[\tau_1 \le u]$ for small u. Note that by Lemma 4.14,

$$\mathbb{E}[f_{x,\varepsilon}(Y_{k+1}^h) - f_{x,\varepsilon}(Y_k^h) + hC_\varepsilon] \ge 0,$$

from which it follows that the process

$$f_{x,\varepsilon}(Y_k^h) + C_\varepsilon hk, \quad k = 0, 1, \ldots$$

is a submartingale. We set $\tau = \inf\{k \geq 1 : |Y_k^h - x| > \varepsilon\}$, so that $\tau_1 = h\tau$. Using the Optional Stopping Theorem at $\tau \wedge u'$ with $u' = u/h$,

$$\mathbb{E}_x \left[f_{x,\varepsilon}(Y^h_{\tau \wedge u'}) + hC_\varepsilon \, \tau \wedge u' \right] \geq \mathbb{E}_x[f_{x,\varepsilon}(Y_0^h)] = 1. \tag{35}$$

On the event $\{\tau \leq u'\}$, we have that $|Y^h_{\tau\wedge u'} - x| \geq \varepsilon$, so that $f_{x,\varepsilon}(Y^h_{\tau\wedge u'}) = 0$, and evidently $\tau \wedge u' \leq u'$, so we have:

$$\mathbb{P}_x[\tau_1 \leq u] = \mathbb{P}[\tau \leq u'] \leq \mathbb{E}_x \left[1 - f_{x,\varepsilon}(Y^h_{\tau\wedge u'}) \right] \leq hC_\varepsilon u' = C_\varepsilon u. \tag{36}$$

(To see this, observe that the expression under the expectation is non-negative and equals one if $\tau \leq u'$ and rearrange (35).) Consequently, for all $u > 0$, letting $p = \mathbb{P}_x[\tau_1 \leq u]$:

$$\begin{aligned} \mathbb{E}_x[e^{-\tau_1}] &\leq \mathbb{P}_x[\tau_1 \leq u] + e^{-u}\mathbb{P}_x[\tau_1 \geq u] \\ &= p + e^{-u}(1-p) = e^{-u} + p(1 - e^{-u}) \\ &\leq e^{-u} + pu \leq 1 - u + C_\varepsilon u^2. \end{aligned}$$

Thus, by choosing u small enough, we can find $\lambda < 1$, independent of x or δ (depending solely on ε through C_ε), such that $\mathbb{E}_x[e^{-\tau_1}] \leq \lambda$. Now, iterating and using the strong Markov property at times $\tau_1, \ldots, \tau_n$, since λ does not depend on x, we have

$$\mathbb{E}_x[e^{-\tau_n}] \leq \lambda^n,$$

and thus, by Markov's inequality,

$$\begin{aligned} \mathbb{P}_x[N > n] = \mathbb{P}_x[\tau_n < 1] &\leq \mathbb{P}_x[e^{-\tau_n} \geq e^{-1}] \\ &\leq e\mathbb{E}_x[e^{-\tau_n}] \leq e\lambda^n. \end{aligned}$$

Now observe that for any $k \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}_x[\sigma \leq \delta] &\leq k \sup_y \mathbb{P}_y[\tau_1 \leq \delta] + \mathbb{P}_x[N > k] \\ &\leq k \, C_\varepsilon \delta + e\lambda^k, \end{aligned}$$

where we used (36) to estimate $\mathbb{P}[\tau_1 \leq \delta]$. Thus taking k large enough that $e\lambda^k < \varepsilon/4$ and then δ small enough that $C_\varepsilon k\delta < \varepsilon/4$, we have that $\mathbb{P}[\sigma \leq \delta] < \varepsilon/2$ and tightness is proved.

Step 2. Uniqueness of the weak subsequential limits.
Since we have assumed that the martingale problem $\mathbf{M}(a, b)$ is well posed, it suffices to show that the limit of any weakly convergent subsequence solves the martingale problem $\mathbf{M}(a, b)$. The first step is to show that the operator L^h defined in (33) converges in a suitable sense to the infinitesimal generator L of the diffusion:

$$Lf(x) = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{d} b_i(x) \frac{\partial f}{\partial x_i}(x).$$

Lemma 4.16. *Let $f \in \mathcal{C}^2_K$ (the space of twice continuously differentiable functions with compact support). Then $L^h f(x) \to Lf(x)$ uniformly in $x \in \mathbb{R}^d$ as $h \to 0$.*

Proof. As in Lemma 4.14, we apply a Taylor expansion to $L^h f(x)$, and, recalling the definition of $b_i^h(x)$, we write

$$L^h f(x) = \sum_{i=1}^d b_i^h(x) D_i f(x) + \frac{1}{2} \int_{|y-x|\le 1} \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) D_{ij} f(z_{xy}) K_h(x, dy) + \int_{|y-x|>1} [f(y) - f(x)] K_h(x, dy)$$

The final term on the right-hand side converges to 0, uniformly in x, by assumption (iii) with $\varepsilon = 1$. To deal with the first term, note that

$$\left| \sum_{i=1}^d b_i^h(x) D_i f(x) - \sum_{i=1}^d b_i(x) D_i f(x) \right| \le \sup_{1\le i\le d} |b_i^h(x) - b_i(x)| \sum_{i=1}^d \|D_i f\|_\infty,$$

which converges to 0 uniformly in x by assumption (ii) (since $f \in \mathcal{C}_K^2$). It remains to control the middle term. Recalling the definition of $a_{ij}^h(x)$, we obtain:

$$\begin{aligned} &\left| \int_{|y-x|\le 1} \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) D_{ij} f(z_{xy}) K_h(x, dy) - \sum_{i,j=1}^d a_{ij}(x) D_{ij} f(x) \right| \\ &\qquad \le \left| \sum_{i,j=1}^d a_{ij}^h(x) D_{ij} f(x) - \sum_{i,j=1}^d a_{ij}(x) D_{ij} f(x) \right| \\ &\qquad\quad + \left| \int_{|y-x|\le 1} \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) [D_{ij} f(z_{xy}) - D_{ij} f(x)] K_h(x, dy) \right|. \end{aligned}$$

The first term converges to 0 uniformly in x by (i) and the fact that the derivatives of f are uniformly bounded. The second term can be split into integrals over $|y - x| > \varepsilon$ and $|y - x| \le \varepsilon$. The first of these converges to 0 uniformly in $x \in \mathbb{R}^d$ thanks to (iii) and the fact that the integrand is bounded. For the other term, let

$$\Gamma(\varepsilon) = \sup_{1\le i,j\le d} \sup_{|y-x|\le\varepsilon} |D_{ij} f(z_{xy}) - D_{ij} f(x)|,$$

then since z_{xy} lies on the segment between x and y, and since $D_{ij} f$ is continuous on the compact set K (and hence uniformly continuous), $\Gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$. On the other hand, by the Cauchy-Schwarz inequality,

$$\begin{aligned} &\left| \int_{|y-x|\le\varepsilon} \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) [D_{ij} f(z_{xy}) - D_{ij} f(x)] K_h(x, dy) \right| \\ &\qquad\qquad \le \Gamma(\varepsilon) \int_{|y-x|\le\varepsilon} d|y - x|^2 K_h(x, dy), \end{aligned}$$

and the proof of the lemma is complete. □

To complete the proof of Theorem 4.12, fix a sequence $h_n \to 0$ for which $X^{h_n} \to X$ weakly (in D) as $n \to \infty$. Recall from (34) that

$$f(X_{kh_n}^{h_n}) - \sum_{j=0}^{k-1} h_n L^{h_n} f(X_{jh_n}^{h_n}), \quad k = 0, 1, \ldots$$

is a (discrete parameter) martingale. In particular, taking $k_n = \lceil s/h_n \rceil$, and $\ell_n = \lceil t/h_n \rceil$, we obtain that for any $\mathcal{F}_s$-measurable $F : D \to \mathbb{R}$,

$$\mathbb{E}_x \left[F(X^{h_n}) \left\{ f(X^{h_n}_{\ell_n h_n}) - f(X^{h_n}_{k_n h_n}) - \sum_{j=k_n}^{\ell_n - 1} h_n L^{h_n} f(X^{h_n}_{j h_n}) \right\} \right] = 0. \tag{37}$$

Using the Skorokhod Representation Theorem, we can find Y^n such that $Y^n \stackrel{d}{=} X^{h_n}$ and $Y^n \to Y$ almost surely, where $Y \stackrel{d}{=} X$. We recognize a Riemann sum in the expectation on the left hand side of (37). Since almost sure convergence in D implies almost sure convergence of the marginals, we deduce that

$$\mathbb{E}_x \left[F(X) \left\{ f(X_t) - f(X_s) - \int_s^t Lf(X_u) du \right\} \right] = 0.$$

Since F is an arbitrary, it follows that

$$f(X_t) - \int_0^t Lf(X_u) du, \quad t \geq 0$$

is a martingale for all $f \in \mathcal{C}^2_K$. Since the martingale problem has a unique solution, we have proved uniqueness of limits and the proof of Theorem 4.12 is complete. □

Theorem 4.12 is a remarkably powerful result, not only as a means of constructing diffusions, but also for identifying an appropriate diffusion process with which to approximate a Markov chain. Often the limiting diffusion is simpler to study than the Markov chain that it approximates. Moreover, our models are only ever a caricature of reality. The fact that a diffusion approximates so many different local structures suggests that predictions made from the diffusion approximation are 'robust' to changes in the (often unreliable) local details of our modelling.

Here we present one simple example of a diffusion approximation. There are more on the problem sheet.

Example 4.17 (The Ehrenfest chain)**.** *The Ehrenfest chain was proposed by Tatania and Paul Ehrenfest in 1907 to explain the second law of thermodynamics. We suppose that $2n$ molecules are placed in two chambers. In the original model, molecules independently change container at rate $1/2n$. (In other words, at the times of a rate one Poisson process, a molecule is picked independently at random to change container.) We shall work with the discrete time version, in which at each time step a molecule is picked at random to change container (the limiting diffusion is the same as for the original model). Let Z^n_t denote the number of molecules in the first contained at time t. The transition probabilities of this chain are, in an obvious notation,*

$$q_{i,i-1} = \frac{i}{2n} \qquad q_{i,i+1} = \frac{2n-i}{2n}.$$

The chain has an equilibrium distribution which we denote by π_i. Mark Kac proved in 1947 that if the initial system is not in equilibrium, then entropy*, given by*

$$H(t) = -\sum_i \mathbb{P}[Z_t = i] \log \big(\mathbb{P}[Z_t = i]/\pi_i \big),$$

increases.

Define the centred and normalized process $X^n_t = (Z^n_{\lfloor tn \rfloor} - n)/\sqrt{n}$, and suppose (for definiteness) that $Z^n_0 = n$.

The process $(X_t^n, 0 \le t \le 1)$ converges weakly to an Ornstein-Uhlenbeck diffusion $(X_t, 0 \le t \le 1)$ with unit viscosity, that is the pathwise unique solution to

$$dX_t = -X_t dt + dB_t, X_0 = 0.$$

We prove this using Theorem 4.12 with $Y_t^n = (Z_t^n - n)/\sqrt{n}$ so that X_t^n is just a timechange of Y_t^n. Here $h_n = 1/n$ and we have abused notation by writing X^n for X^{h_n}. The state space of X^n is $E_n = \{k/\sqrt{n} : -n \le k \le n\}$ and (again abusing notation) $K^n(x, dy) = n\Pi^n(x, dy)$, where

$$\Pi_n(x, x + n^{-1/2}) = \frac{n - x\sqrt{n}}{2n}, \qquad \Pi_n(x, x - n^{-1/2}) = \frac{n + x\sqrt{n}}{2n}.$$

To see this, if $Y_t^n = x$, then the number of particles in the first container is $n + x\sqrt{n}$ and if one of those is chosen (which happens with probability $(n + x\sqrt{n})/2n$) then the number of molecules in the container will go down by one.

Since the chain can only make jumps of size $1/\sqrt{n}$, condition (iii) of Theorem 4.12 holds trivially, and for all $n \ge 1$,

$$b^n(x) = \int (y - x) K_n(x, dy) = n \left\{ n^{-1/2} \frac{n - x\sqrt{n}}{2n} - n^{-1/2} \frac{n + x\sqrt{n}}{2n} \right\} = -x,$$

while

$$a^n(x) = \int (y - x)^2 K_n(x, dy) = n \left\{ n^{-1} \frac{n - x\sqrt{n}}{2n} + n^{-1} \frac{n + x\sqrt{n}}{2n} \right\} = 1.$$

Since the coefficients of the Ornstein-Uhlenbeck diffusion are Lipschitz, there is pathwise uniqueness for the associated SDE and thus uniqueness in distribution. Therefore, $(X_t^n, 0 \le t \le 1)$ converges weakly to $(X_t, 0 \le t \le 1)$, by Theorem 4.12.

5 The method of duality

Often, proving that a sequence of processes is tight, and that any limit point solves a martingale problem, is relatively straightforward, but uniqueness of the solution to the martingale problem is much harder. Sometimes in this context, the method of duality can be helpful. It has also found widespread use as a means of characterising the distribution of a complex stochastic process in terms of another, simpler, process (or even a deterministic equation).

Theorem 5.1 (The method of duality)**.** *Let E_1, E_2 be metric spaces and suppose that $\mathbb{P}$ and $\mathbb{Q}$ are probability distributions on the space of càdlàg functions from $[0, \infty)$ to E_1, E_2 respectively. (These spaces are endowed with the natural σ-fields.) Take two bounded functions f, g on $E_1 \times E_2$ for which the following are true:*

(i) For each $y \in E_2$, $f(\cdot, y)$ and $g(\cdot, y)$ are continuous functions on E_1.

(ii) For each $x \in E_1$, $f(x, \cdot)$ and $g(x, \cdot)$ are continuous functions on E_2.

(iii) For $y \in E_2$,

$$f(X(t), y) - \int_0^t g(X(s), y) ds$$

is a $\mathbb{P}$-martingale.

(iv) For $x \in E_1$,

$$f(x, Y(t)) - \int_0^t g(x, Y(s))ds$$

is a $\mathbb{Q}$-martingale.

Then

$$\mathbb{E}^{\mathbb{P}}_{X(0)}\left[f(X(t), Y(0))\right] = \mathbb{E}^{\mathbb{Q}}_{Y(0)}\left[f(X(0), Y(t))\right].$$

Proof. Taking expectations and differentiating gives that for $s \geq 0$,

$$\frac{d}{ds}\mathbb{E}^{\mathbb{P}}\left[f\left(X(s), y\right)\right] = \mathbb{E}^{\mathbb{P}}\left[g\left(X(s), y\right)\right],$$

and

$$\frac{d}{ds}\mathbb{E}^{\mathbb{Q}}\left[f\left(x, Y(s)\right)\right] = \mathbb{E}^{\mathbb{Q}}\left[g\left(x, Y(s)\right)\right].$$

(The interchange of differentiation and integration is valid, since f and g are bounded.) Thus for each fixed t and $0 \leq s \leq t$,

$$\frac{d}{ds}\mathbb{E}^{\mathbb{P}\times\mathbb{Q}}\left[f\left(X(s), Y(t-s)\right)\right] = 0.$$

Integrating s over $[0, t]$ yields the result. □

(There is no ambiguity if we drop the superscripts $\mathbb{P}$, $\mathbb{Q}$, from the expectations. From now on, we do so.)

The application that we have in mind is the following. If we can find a sufficiently large class of functions $f(\cdot, y)$ for which the conditions of Theorem 5.1 are satisfied, then we can use Y to characterise the distribution of X (or vice versa). In particular, *existence* of a dual process Y (for a sufficiently rich class of functions f) is sufficient to guarantee *uniqueness* of X. To see why, suppose that we have two solutions to the martingale problem, X, $\tilde{X}$. Then, for a dual process Y,

$$\mathbb{E}[f(X(t), Y(0))] = \mathbb{E}[f(X(0), Y(t))] = \mathbb{E}[f(\tilde{X}(0), Y(t))] = \mathbb{E}[f(\tilde{X}(t), Y(0))].$$

If the class of f is wide enough, it follows that the one dimensional distributions of X and $\tilde{X}$ coincide. But an inductive argument similar to the one we saw in the proof of Theorem 3.3 shows that all finite dimensional distributions coincide in this case, i.e., the solution to the martingale problem is unique. Specifically, one can apply the following result.

Lemma 5.2 (Ethier & Kurtz (1986) Chapter 4 Theorem 4.2 (a))**.** *Let A be a Markov pregenerator on a compact separable space E. Suppose that for all $\mu \in \mathcal{P}(E)$, if X, Y solve the martingale problem (A, μ), then X, Y have the same one dimensional distributions. Then, for all $\mu \in \mathcal{P}(E)$, any two solutions of the martingale problem (A, μ) have the same finite dimensional distributions.*

In general, it is far simpler to prove existence than uniqueness and so this observation has found widespread application.

The method of duality can be generalised in several ways. For example, in the form above we insisted that if we write A_X for the generator of the process X on E_1 and A_Y for the generator of the process Y on E_2, then $A_X f(\cdot, y)(x) = A_Y f(x, \cdot)(y)$. If instead we have

$$A_X f(\cdot, y)(x) + \alpha(x)f(x, y) = A_Y f(x, \cdot)(y) + \beta(y)f(x, y),$$

then at least provided that $\int_0^t |\alpha(X(s))|ds < \infty$ a.s. and $\int_0^t |\beta(Y(s))|ds < \infty$ a.s. and we have the additional integrability conditions

$$\mathbb{E}\left[\left|f(X(t),Y(0))\exp\left(\int_0^t \alpha(X(s))ds\right)\right|\right] < \infty$$

and

$$\mathbb{E}\left[\left|f(X(0),Y(t))\exp\left(\int_0^t \beta(Y(s))ds\right)\right|\right] < \infty,$$

the duality formula can be modified to

$$\mathbb{E}\left[f(X(t),Y(0))\exp\left(\int_0^t \alpha(X(s))ds\right)\right] = \mathbb{E}\left[f(X(0),Y(t))\exp\left(\int_0^t \beta(Y(s))ds\right)\right].$$

A proof of this more general form of the method of duality (and some examples of its ramifications) can be found in Ethier & Kurtz, p.188ff.

Remark 5.3 (Approximate duality)**.** *A very useful extension of the method of duality is the following observation of Mytnik (1996). The purpose of the dual process Y is to provide an expression for the one-dimensional distributions of a solution to the martingale problem for X. The same objective is achieved if one can find an 'approximate dual'. That is, if, for each y, we can find a sequence of processes Y_n such that*

$$\mathbb{E}\left[f(X(t),y)\right] = \lim_{n\to\infty} \mathbb{E}\left[f(X(0),Y_n(t))\right]. \tag{38}$$

Uniqueness of the solution to the martingale problem follows as before.

The function f of Theorem 5.1 is often referred to as a 'duality function'. In general, finding a duality function is something of a black art. However, for diffusion processes arising in genetics, one can often use a so-called *moment dual*.

Example 5.4 (The Wright-Fisher diffusion and Kingman's coalescent)**.** *Suppose that*

$$dX_t = \sqrt{X_t(1-X_t)}dB_t.$$

Then

$$\mathbb{E}\left[X_t^{N_0}\right] = \mathbb{E}\left[X_0^{N_t}\right],$$

where N_t is a pure death process in which $N_t \mapsto N_t - 1$ at rate $\binom{N_t}{2}$.

To check this, we take $f(x,n) = x^n$ as the duality function in Theorem 5.1. Then writing $\mathcal{L}_{WF}$ for the infinitesimal generator of the Wright-Fisher diffusion,

$$\mathcal{L}_{WF}f(\cdot,n)(x) = \binom{n}{2}x^{n-2}x(1-x) = \binom{n}{2}(x^{n-1}-x^n).$$

Similarly, writing $\mathcal{L}_K$ for the infinitesimal generator of the pure death process,

$$\mathcal{L}_K f(x,\cdot)(n) = \binom{n}{2}(f(x,n-1)-f(x,n)) = \binom{n}{2}(x^{n-1}-x^n).$$

We take $g(x,n) = \binom{n}{2}(x^{n-1}-x^n)$ in Theorem 5.1. Since $0 < x < 1$ and n only ever decreases through positive integer values, boundedness is immediate and so the result follows.

There is actually a much stronger relationship between the pure death process we have just described and the Wright-Fisher diffusion. If we consider the ancestors of a sample taken from the population at the current time, then as we trace backwards in time, the death process counts the number of individuals ancestral to the sample. A death event, corresponds to two ancestral lineages having a common parent. The process that describes the whole tree of ancestral lineages is known as Kingman's coalescent and has proved to be an immensely powerful tool in population genetics.

6 The theory of speed and scale

With apologies, in this section we shall change our notation. Diffusions will have infinitesimal drift μ and infinitesimal variance σ^2 and will be defined on the interval (a, b).

6.1 The speed measure and the scale function

Evidently diffusion processes are intricately connected with second order parabolic equations. Indeed, if we have an expression for the fundamental solution to the corresponding equation, then we have the transition density of the diffusion and so can calculate essentially any quantity of interest. However, in general, it is certainly not going to be possible to find an explicit expression for the transition density. Nonetheless, a nice feature of one dimensional diffusions is that many quantities can be calculated explicitly. This is because (except at certain singular points which, under our conditions, will only ever be at the endpoints of the interval in which our diffusion lives) all one-dimensional diffusions can be transformed into Brownian motion, first by a change of space variable (through the so-called scale function) and then a timechange (through what is known as the speed measure).

To see how this works, we first investigate what happens to a diffusion when we change the timescale. Suppose that a diffusion $\{Z_t\}_{t\geq 0}$ has generator $\mathcal{L}_Z$, with infinitesimal drift $\mu_Z(x)$ and infinitesimal variance $\sigma_Z^2(x)$. We use Δ_h to denote the operator that reports the increment of a process over the next time step of length h (where we are thinking of h as being infinitesimally small). Thus, for example,

$$\mathbb{E}\left[\Delta_h Z(0)\middle|\, Z_0 = y\right] = \mu_Z(y)h + o(h),$$

and

$$\mathbb{E}\left[(\Delta_h Z(0))^2\middle|\, Z_0 = y\right] = \sigma_Z^2(y)h + o(h).$$

We define a new process $\{Y_t\}_{t\geq 0}$ by $Y_t = Z_{\tau(t)}$ where

$$\tau(t) = \int_0^t \beta(Y_s)ds,$$

for some function $\beta(x)$ which we assume to be bounded, continuous and strictly positive. So if $Y_0 = Z_0$, then the increment of Y_t over an infinitesimal time interval $(0, dt)$ is that of Z_t over the interval $(0, d\tau(t)) = (0, \beta(Y_0)dt)$. In our previous notation,

$$\mathbb{E}[\Delta_h Y(0)|Y_0 = y] = \beta(Y_0)h\mu_Z(Z_0) + o(h) = \beta(y)\mu_Z(y)h + o(h),$$

and

$$\mathbb{E}[(\Delta_h Y(0))^2|Y_0 = y] = \beta(Y_0)h\sigma_Z^2(Z_0) + o(h) = \beta(y)\sigma_Z^2(y)h + o(h).$$

In other words,

$$\mathcal{L}_Y f(x) = \beta(x)\mathcal{L}_Z f(x).$$

In the simplest example, β is a constant and we are simply changing our time units in a spatially homogeneous way. In general, the rate of our 'clock' depends upon where we are in space. We are now in a position to understand speed and scale. Let $\{X_t\}_{t\geq 0}$ be governed by the infinitesimal generator

$$\mathcal{L}f(x) = \frac{1}{2}\sigma^2(x)\frac{d^2 f}{dx^2} + \mu(x)\frac{df}{dx}, \tag{39}$$

for f twice continuously differentiable on (a, b). We assume that $\mu(x)$ and $\sigma(x)$ are bounded and locally Lipschitz on (a, b) with $\sigma^2(x) > 0$ on (a, b) (although it can vanish on $\{a, b\}$). As we know

from Example 2.25, in general, we are going to need to be more specific about which functions lie in the domain of our generator, but we defer that to §6.2. Suppose now that $S(x)$ is a strictly increasing function on (a,b) and consider the new process $Z_t = S(X_t)$. Then the generator $\mathcal{L}_Z$ of Z can be calculated as

$$\begin{aligned}
\mathcal{L}_Z f(x) &= \frac{d}{dt}\mathbb{E}\left[f(Z_t)\middle| Z_0 = x\right]\Big|_{t=0} \\
&= \frac{d}{dt}\mathbb{E}\left[f(S(X_t))\middle| S(X_0) = x\right]\Big|_{t=0} \\
&= \mathcal{L}_X(f\circ S)(S^{-1}(x)) \\
&= \frac{1}{2}\sigma^2(S^{-1}(x))\frac{d^2}{dx^2}(f\circ S)(S^{-1}(x)) + \mu(S^{-1}(x))\frac{d}{dx}(f\circ S)(S^{-1}(x)) \\
&= \frac{1}{2}\sigma^2(S^{-1}(x))\left\{(S'(S^{-1}(x)))^2\frac{d^2f}{dx^2}(x) + S''(S^{-1}(x))\frac{df}{dx}(x)\right\} \\
&\quad +\mu(S^{-1}(x))S'(S^{-1}(x))\frac{df}{dx}(x) \\
&= \frac{1}{2}\sigma^2(S^{-1}(x))S'(S^{-1}(x))^2\frac{d^2f}{dx^2}(x) + \mathcal{L}S(S^{-1}(x))\frac{df}{dx}(x). \qquad (40)
\end{aligned}$$

Now if we can find a strictly increasing function S that satisfies $\mathcal{L}S \equiv 0$, then the drift term in (40) will vanish and so Z_t will just be a time change of Brownian motion on the interval $(S(a), S(b))$. Such an S is provided by the scale function of the diffusion.

Definition 6.1 (Scale function)**.** *For a diffusion X_t on (a,b) with drift μ and variance σ^2, the* scale function *is defined by*

$$S(x) = \int_{x_0}^{x}\exp\left(-\int_{\eta}^{y}\frac{2\mu(z)}{\sigma^2(z)}dz\right)dy,$$

where x_0, η are points fixed (arbitrarily) in (a,b).

Definition 6.2 (Natural scale)**.** *We shall say that a diffusion is in* natural scale *if $S(x)$ can be taken to be linear.*

The scale change $X_t \mapsto S(X_t)$ resulted in a time changed Brownian motion on $(S(a), S(b))$. The change of time required to transform this into standard Brownian motion is dictated by the speed measure.

Definition 6.3 (Speed measure)**.** *The function $m(\xi) = \frac{1}{\sigma^2(\xi)S'(\xi)}$ is the* density of the speed measure *or just the* speed density *of the process X_t. We write*

$$M(x) = \int_{x_0}^{x} m(\xi)d\xi.$$

Remark 6.4. *The function m plays the rôle of β before. Naively, looking at equation (40), we might expect to timechange via $\beta(\xi) = 1/(\sigma^2(\xi)S'(\xi)^2)$. However, notice that*

$$\int_{x_0}^{x} m(\xi)d\xi = \int_{S(x_0)}^{S(x)} m(S^{-1}(y))\frac{1}{S'(S^{-1}(y))}dy = \int_{S(x_0)}^{S(x)}\frac{1}{\sigma^2(S^{-1}(y))\big(S'(S^{-1}(y))\big)^2}dy.$$

The additional $S'(y)$ in the generator (40) has been absorbed in the change of coordinates since our time change is applied to $S(X_t)$ on $(S(a), S(b))$, not to X_t itself.

In summary, we have the following.

Lemma 6.5. *Denoting the scale function and the speed measure by S and M respectively we have*

$$\mathcal{L}f = \frac{1}{2}\frac{1}{dM/dS}\frac{d^2f}{dS^2} = \frac{1}{2}\frac{d}{dM}\left(\frac{df}{dS}\right).$$

Proof

$$\begin{aligned}
\frac{1}{2}\frac{d}{dM}\left(\frac{df}{dS}\right) &= \frac{1}{2}\frac{1}{dM/dx}\frac{d}{dx}\left(\frac{1}{dS/dx}\frac{df}{dx}\right) \\
&= \frac{1}{2}\sigma^2(x)S'(x)\frac{d}{dx}\left(\frac{1}{S'(x)}\frac{df}{dx}\right) \\
&= \frac{1}{2}\sigma^2(x)\frac{d^2f}{dx^2} - \frac{1}{2}\sigma^2(x)S'(x)\frac{S''(x)}{(S'(x))^2}\frac{df}{dx} \\
&= \frac{1}{2}\sigma^2(x)\frac{d^2f}{dx^2} + \mu(x)\frac{df}{dx}
\end{aligned} \tag{41}$$

(since S solves $\mathcal{L}S = 0$) as required. □

6.2 Hitting probabilities and Feller's boundary classification

Before going further, let's see how we might apply this. Recall that the Wright-Fisher diffusion on $(0, 1)$ is used to model evolution of gene frequencies. Suppose that X_t represents the frequency of a-alleles in a population and that zero and one are traps for the process. (The different forms a gene can take are called alleles.) One question that we should like to answer is "What is the probability that the a-allele is eventually lost from the population?" In other words, what is the probability that the diffusion hits zero before one? To prove a general result we need first to be able to answer this question for Brownian motion.

Lemma 6.6. *Let $\{B_t\}_{t\geq 0}$ be standard Brownian motion on the line. For each $y \in \mathbb{R}$, let T_y denote the random time at which it hits y for the* first *time. Then for $a < x < b$,*

$$\mathbb{P}[T_a < T_b |\, B_0 = x] = \frac{b-x}{b-a}.$$

Sketch of Proof

Let $u(x) = \mathbb{P}[T_a < T_b |\, B_0 = x]$ and assume that $\mathbb{P}[T_a \wedge T_b < h | B_0 = x] = o(h)$ as $h \to 0$. If we suppose that u is sufficiently smooth, then, using the Markov property,

$$\begin{aligned}
u(x) &= \mathbb{E}\left[u(B_h) |\, B_0 = x\right] + o(h) \\
&= \mathbb{E}\left[u(x) + (B_h - x)u'(x) + \frac{1}{2}(B_h - x)^2 u''(x)\right] + o(h) \\
&= u(x) + \frac{1}{2}hu''(x) + o(h).
\end{aligned}$$

Subtracting $u(x)$ from each side, dividing by h and letting h tend to zero, we obtain $u''(x) = 0$. We also have the boundary conditions $u(a) = 1$ and $u(b) = 0$. This is easily solved to give

$$u(x) = \frac{b-x}{b-a},$$

as required. □

Of course this reflects the corresponding result for simple random walk that you prove in elementary probability courses. In general we can reduce the corresponding question for $\{X_t\}_{t\geq 0}$ to solution of the equation $\mathcal{L}u(x) = 0$ with $u(a) = 1$ and $u(b) = 0$, but in fact we have already done all the work we need. We have the following result.

Lemma 6.7 (Hitting probabilities)**.** *Let $\{X_t\}_{t\geq 0}$ be a one-dimensional diffusion on (a, b) with infinitesimal drift $\mu(x)$ and variance $\sigma^2(x)$ satisfying the conditions above. If $a < a_0 < x < b_0 < b$ then writing T_y for the first time at which $X_t = y$,*

$$\mathbb{P}[T_{a_0} < T_{b_0} | X_0 = x] = \frac{S(b_0) - S(x)}{S(b_0) - S(a_0)}, \tag{42}$$

where S is the scale function for the diffusion.

Remark 6.8. *Our definition of the scale function, S, depended upon arbitrary choices of η and x_0, but η cancels in the ratio and x_0 in the difference, so that the expression on the right hand side of (42) is well-defined.*

Proof of Lemma 6.7

Evidently it is enough to consider the corresponding hitting probabilities for the process $Z_t = S(X_t)$, where S is the scale function. The process $\{Z_t\}_{t\geq 0}$ is a time changed Brownian motion, but since we only care about *where* not *when* the process exits the interval $(S(a_0), S(b_0))$, then we need only determine the hitting probabilities for Brownian motion and the result follows immediately from Lemma 6.6. □

Before continuing to calculate quantities of interest, we fill in a gap left earlier, when we failed to completely specify the domain of the generators of our one-dimensional diffusions. Whether or not functions in the domain must satisfy boundary conditions at a and b is determined by the nature of those boundaries from the perspective of the diffusion. More precisely, we have the following classification.

Definition 6.9 (Feller's boundary classification)**.** *For a one-dimensional diffusion on the interval with endpoints a, b (with $a < b$), define*

$$u(x) = \int_{x_0}^{x} M dS, \qquad v(x) = \int_{x_0}^{x} S dM,$$

where S is the scale function of Definition 6.1 and M the speed measure of Definition 6.3. The boundary b is said to be

regular	*if*	$u(b) < \infty$	*and*	$v(b) < \infty$
exit	*if*	$u(b) < \infty$	*and*	$v(b) = \infty$
entrance	*if*	$u(b) = \infty$	*and*	$v(b) < \infty$
natural	*if*	$u(b) = \infty$	*and*	$v(b) = \infty$

with symmetric definitions at a.

Regular and exit boundaries are said to be accessible *while entrance and natural boundaries are called* inaccessible*.*

Theorem 6.10. *If neither a nor b is regular, the domain of the generator (39) is the continuous functions f on $[a, b]$ which are twice continuously differentiable on the interior and for which*

(i) if a and b are inaccessible there are no further conditions,

(ii) if b (resp. a) is an exit boundary, then

$$\lim_{x\to b} \mathcal{L}f(x) = 0$$

$$\left(\textit{resp. } \lim_{x\to a} \mathcal{L}f(x) = 0\right).$$

If b (resp. a) is a regular boundary, then for each fixed $q \in [0,1]$ there is a Feller semigroup corresponding to the generator (39) with domain as above plus the additional condition

$$q \lim_{x\to b} \mathcal{L}f(x) = -(1-q) \lim_{x\to b} \frac{1}{S'(x)} f'(x) \tag{43}$$

$$\left(\textit{resp. } q \lim_{x\to a} \mathcal{L}f(x) = (1-q) \lim_{x\to a} \frac{1}{S'(x)} f'(x)\right).$$

For a more careful discussion see Ethier & Kurtz (1986), Chapter 8.

6.3 Green's functions

Lemma 6.7 tells us the probability that we exit (a,b) for the first time through a, but can we glean some information about how long we must wait for $\{X_t\}_{t\geq 0}$ to exit the interval (a,b) (either through a or b) or, more generally, writing T^* for the first exit time of (a,b), can we say anything about $\mathbb{E}[\int_0^{T^*} g(X_s)ds|X_0 = x]$? (Putting $g = 1$ gives the mean exit time.) Let us write

$$w(x) = \mathbb{E}\left[\int_0^{T^*} g(X_s)ds|X_0 = x\right]$$

and we'll derive the differential equation satisfied by w.

Suppose for simplicity that g is Lipschitz continuous on (a,b) with Lipschitz constant K. First note that $w(a) = w(b) = 0$. Now consider a small interval of time of length h. We're going to split the integral into the contribution up to time h and after time h. Because $\{X_t\}_{t\geq 0}$ is a Markov process,

$$\mathbb{E}\left[\int_h^{T^*} g(X_s)ds|X_h = z\right] = \mathbb{E}\left[\int_0^{T^*} g(X_s)ds|X_0 = z\right] = w(z)$$

and so for $a < x < b$

$$w(x) \approx \mathbb{E}\left[\int_0^h g(X_s)ds|X_0 = x\right] + \mathbb{E}\left[w(X_h)|X_0 = x\right]. \tag{44}$$

The '$\approx$' here is because we have ignored the possibility that $h > T^*$. Since g is Lipschitz continuous, we have the approximation

$$\begin{aligned}\left|\mathbb{E}\left[\int_0^h g(X_s)ds\middle| X_0 = x\right] - h\,g(x)\right| &= \mathbb{E}\left[\left|\int_0^h g(X_s)ds - hg(x)\right|\middle| X_0 = x\right] \\ &\leq \mathbb{E}\left[\int_0^h K|X_s - x|ds\middle| X_0 = x\right] \leq K\int_0^h \sqrt{\mathbb{E}\left[|X_s - x|^2|\, X_0 = x\right]}ds = \mathcal{O}(h^{3/2}).\end{aligned}$$

Now substitute this estimate in (44), subtract $w(x)$ from both sides, divide by h and let $h \downarrow 0$ to obtain

$$\mu(x)w'(x) + \frac{1}{2}\sigma^2(x)w''(x) = -g(x), \quad w(a) = 0 = w(b). \tag{45}$$

Let us now turn to solving this equation. Using (41) from the proof of Lemma 6.5 with $w = f$,

$$\mathcal{L}w(x) = \frac{1}{2}\frac{1}{m(x)}\frac{d}{dx}\left(\frac{1}{S'(x)}w'(x)\right)$$

and so we have

$$\frac{d}{dx}\left(\frac{1}{S'(x)}w'(x)\right) = -2g(x)m(x),$$

whence

$$\frac{1}{S'(x)}w'(x) = -2\int_a^x g(\xi)m(\xi)d\xi + \beta$$

where β is a constant. Multiplying by $S'(x)$ and integrating gives

$$w(x) = -2\int_a^x S'(\xi)\int_a^\xi g(\eta)m(\eta)d\eta d\xi + \beta(S(x) - S(a)) + \alpha$$

for constants α, β. Since $w(a) = 0$, we immediately have that $\alpha = 0$. Reversing the order of integration,

$$\begin{aligned} w(x) &= -2\int_a^x \int_\eta^x S'(\xi)d\xi g(\eta)m(\eta)d\eta + \beta(S(x) - S(a)) \\ &= -2\int_a^x (S(x) - S(\eta))g(\eta)m(\eta)d\eta + \beta(S(x) - S(a)) \end{aligned}$$

and $w(b) = 0$ now gives

$$\beta = \frac{2}{S(b) - S(a)}\int_a^b (S(b) - S(\eta))g(\eta)m(\eta)d\eta.$$

Finally then

$$\begin{aligned} w(x) &= \frac{2}{S(b) - S(a)}\Big\{(S(x) - S(a))\int_a^b (S(b) - S(\eta))g(\eta)m(\eta)d\eta \\ &\qquad - (S(b) - S(a))\int_a^x (S(x) - S(\eta))g(\eta)m(\eta)d\eta\Big\} \\ &= \frac{2}{S(b) - S(a)}\Big\{(S(x) - S(a))\int_x^b (S(b) - S(\eta))g(\eta)m(\eta)d\eta \\ &\qquad + (S(b) - S(x))\int_a^x (S(\eta) - S(a))g(\eta)m(\eta)d\eta\Big\} \end{aligned}$$

where the last line is obtained by splitting the first integral into $\int_a^b = \int_x^b + \int_a^x$.

Theorem 6.11. *For a continuous function g,*

$$\mathbb{E}\left[\int_0^{T*} g(X_s)ds | X_0 = x\right] = \int_a^b G(x,\xi)g(\xi)d\xi,$$

where for $a < x < b$ we have

$$G(x,\xi) = \begin{cases} 2\frac{(S(x)-S(a))}{(S(b)-S(a))}(S(b) - S(\xi))m(\xi), & \text{for } x < \xi < b \\ 2\frac{(S(b)-S(x))}{(S(b)-S(a))}(S(\xi) - S(a))m(\xi), & \text{for } a < \xi < x, \end{cases}$$

with S the scale function given in Definition 6.1 and $m(\xi) = \frac{1}{\sigma^2(\xi)S'(\xi)}$, the density of the speed measure.

Definition 6.12. *The function $G(x,\xi)$ is called the* Green's function *of the process $\{X_t\}_{t\geq 0}$.*

Of course, the Green's function is familiar from the elementary theory of differential equations, but now we have a probabilistic interpretation. By taking g to approximate $\mathbf{1}_{(x_1,x_2)}$ we see that $\int_{x_1}^{x_2} G(x,\xi)d\xi$ is the mean time spent by the process in (x_1,x_2) before exiting (a,b) if initially $X_0 = x$. Sometimes, the Green's function is called the *sojourn density*.

Example 6.13. *Consider the Wright-Fisher diffusion with generator*

$$\mathcal{L}f(p) = \frac{1}{2}p(1-p)f''(p).$$

Notice that since it has no drift term ($\mu = 0$) it is already in natural scale, $S(p) = p$ (up to an arbitrary additive constant). What about $\mathbb{E}[T^*]$?

Using Theorem 6.11 with $g = 1$ we have

$$\begin{aligned}
\mathbb{E}_p[T^*] &= \mathbb{E}\left[\int_0^{T^*} 1ds | X_0 = p\right] = \int_0^1 G(p,\xi)d\xi \\
&= 2\int_p^1 p(1-\xi)\frac{1}{\xi(1-\xi)}d\xi + 2\int_0^p (1-p)\xi\frac{1}{\xi(1-\xi)}d\xi \\
&= 2p\int_p^1 \frac{1}{\xi}d\xi + 2(1-p)\int_0^p \frac{1}{1-\xi}d\xi \\
&= -2\left\{p\log p + (1-p)\log(1-p)\right\}.
\end{aligned}$$

On the problem sheet you show that the Wright-Fisher diffusion arises as a diffusion approximation both to the (continuous time) Moran model and the (discrete time) Wright-Fisher model. Example 6.13 suggests that in our Moran model, at least if the population is large, if the current proportion of a-alleles is p, the time until either the a-allele or the A-allele is fixed in the population should have mean approximately

$$-2\left\{p\log p + (1-p)\log(1-p)\right\}. \tag{46}$$

In fact by conditioning on whether the proportion of a-alleles increases or decreases at the first reproduction event, one obtains a recurrence relation for the *number of jumps* until the Moran process first hits either zero or one. This recurrence relation can be solved explicitly and since jumps occur at independent exponentially distributed times with mean $1/\binom{N}{2}$, it is easy to verify that (46) is indeed a good approximation. For the Wright-Fisher model, in its original timescale, there is no explicit expression for the expected time to fixation, $t(p)$. However, since changes in p over a single generation are typically small, one can expand $t(p)$ in a Taylor series, and thus verify that for a large population,

$$p(1-p)t''(p) = -2N, \quad t(0) = 0 = t(1).$$

This is readily solved to give

$$t(p) = -2N\left\{p\log p + (1-p)\log(1-p)\right\},$$

as predicted by our diffusion approximation. (The Moran model is already in the diffusive timescale, whereas the Wright-Fisher model is not, accounting for the extra factor of N.)

6.4 Stationary distributions and reversibility

Before moving on to stochastic representations of solutions to various PDEs, we consider one last quantity for our one-dimensional diffusions. First a general definition.

Definition 6.14 (Stationary distribution)**.** *Let $\{X_t\}_{t\geq 0}$ be a Markov process on the space E. A* stationary distribution *for $\{X_t\}_{t\geq 0}$ is a probability distribution ψ on E such that if X_0 has distribution ψ, then X_t has distribution ψ for all $t \geq 0$.*

In particular this definition tells us that if ψ is a stationary distribution for $\{X_t\}_{t\geq 0}$, then

$$\frac{d}{dt}\mathbb{E}\left[f(X_t)\middle| X_0 \sim \psi\right] = 0,$$

where we have used $X_0 \sim \psi$ to indicate that X_0 is distributed according to ψ. In other words

$$\frac{d}{dt}\int_E \mathbb{E}\left[f(X_t)\middle| X_0 = x\right]\psi(dx) = 0.$$

Evaluating the time derivative at $t = 0$ gives

$$\int_E \mathcal{L}f(x)\psi(dx) = 0.$$

Sometimes this allows us to find an explicit expression for $\psi(dx)$. Let $\{X_t\}_{t\geq 0}$ be a one-dimensional diffusion on (a, b) with generator given by (39). We're going to suppose that there is a stationary distribution which is absolutely continuous with respect to Lebesgue measure. Let us abuse notation a little by using $\psi(x)$ to denote the density of $\psi(dx)$ on (a, b). Then, integrating by parts, we have that for all $f \in \mathcal{D}(\mathcal{L})$,

$$\begin{aligned} 0 &= \int_a^b \left\{\frac{1}{2}\sigma^2(x)\frac{d^2f}{dx^2}(x) + \mu(x)\frac{df}{dx}(x)\right\}\psi(x)dx \\ &= \int_a^b f(x)\left\{\frac{1}{2}\frac{d^2}{dx^2}\left(\sigma^2(x)\psi(x)\right) - \frac{d}{dx}\left(\mu(x)\psi(x)\right)\right\}dx + \text{boundary terms.}\end{aligned}$$

This equation must hold for all f in the domain of $\mathcal{L}$ and so, in particular, choosing f and f' to vanish on the boundary,

$$\frac{1}{2}\frac{d^2}{dx^2}\left(\sigma^2(x)\psi(x)\right) - \frac{d}{dx}\left(\mu(x)\psi(x)\right) = 0 \quad \text{for } x \in (a, b). \tag{47}$$

Integrating once gives

$$\frac{1}{2}\frac{d}{dx}\left(\sigma^2(x)\psi(x)\right) - \mu(x)\psi(x) = C_1,$$

for some constant C_1 and then using $S'(x)$ as an integrating factor we obtain

$$\frac{d}{dy}\left(S'(y)\sigma^2(y)\psi(y)\right) = C_1S'(y),$$

from which

$$\psi(x) = C_1\frac{S(x)}{S'(x)\sigma^2(x)} + C_2\frac{1}{S'(x)\sigma^2(x)} = m(x)\left[C_1S(x) + C_2\right].$$

If we can arrange constants so that $\psi \geq 0$ and

$$\int_a^b \psi(\xi)d\xi = 1$$

then the stationary distribution exists and has density ψ. In particular, if $\int_a^b m(y)dy < \infty$, then taking $C_1 = 0$,

$$\psi(x) = \frac{m(x)}{\int_a^b m(y)dy} \tag{48}$$

is the density of a stationary distribution for the diffusion.

We know from the theory of Markov chains that uniqueness of the stationary measure of a chain requires irreducibility. The corresponding condition here is regularity.

Definition 6.15. *For a one dimensional diffusion process on the interval I, let us write*

$$H_y = \inf\{t > 0 : X_t = y\}.$$

The diffusion is said to be regular *if for all $x \in I^0$ (the interior of I) and all $y \in I$ (including finite endpoints) $\mathbb{P}_x[H_y < \infty] > 0$.*

Theorem 6.16 (Watanabe & Motoo 1958)**.** *A regular diffusion in natural scale with no absorbing boundary points has a stationary distribution if and only if the speed measure is finite and then it is given by (48).*

Under these conditions there is also an ergodic theorem.

Example 6.17. *Recall the generator of the Wright-Fisher diffusion with mutation from the problem sheet,*

$$\mathcal{L}f(p) = \frac{1}{2}p(1-p)\frac{d^2f}{dp^2} + \Big(\nu_2(1-p) - \nu_1 p\Big)\frac{df}{dp}.$$

What is the stationary distribution?

For this diffusion

$$\begin{aligned} S'(p) &= \exp\left(-\int_{p_0}^{p} \frac{2\mu(z)}{\sigma^2(z)}dz\right) \\ &= \exp\left(-\int_{p_0}^{p} \frac{2\nu_2(1-z) - 2\nu_1 z}{z(1-z)}dz\right) \\ &= C\exp\left(-2\nu_2 \log p - 2\nu_1 \log(1-p)\right) \\ &= Cp^{-2\nu_2}(1-p)^{-2\nu_1}, \end{aligned}$$

where the value of the constant C depends on p_0. In this case we have

$$m(p) = \frac{1}{\sigma^2(p)S'(p)} = Cp^{2\nu_2-1}(1-p)^{2\nu_1-1}.$$

Now

$$\int_0^1 m(p)dp = \int_0^1 Cp^{2\nu_2-1}(1-p)^{2\nu_1-1}dp = C\frac{\Gamma(2\nu_1)\Gamma(2\nu_2)}{\Gamma(2(\nu_1+\nu_2))}$$

(where Γ is Euler's Gamma function) and so the stationary distribution is just

$$\psi(p) = \frac{\Gamma(2(\nu_1+\nu_2))}{\Gamma(2\nu_1)\Gamma(2\nu_2)}p^{2\nu_2-1}(1-p)^{2\nu_1-1}. \tag{49}$$

Ethier & Kurtz (1986), Chapter 10, Lemma 2.1 gives a direct proof of uniqueness of this stationary distribution. □

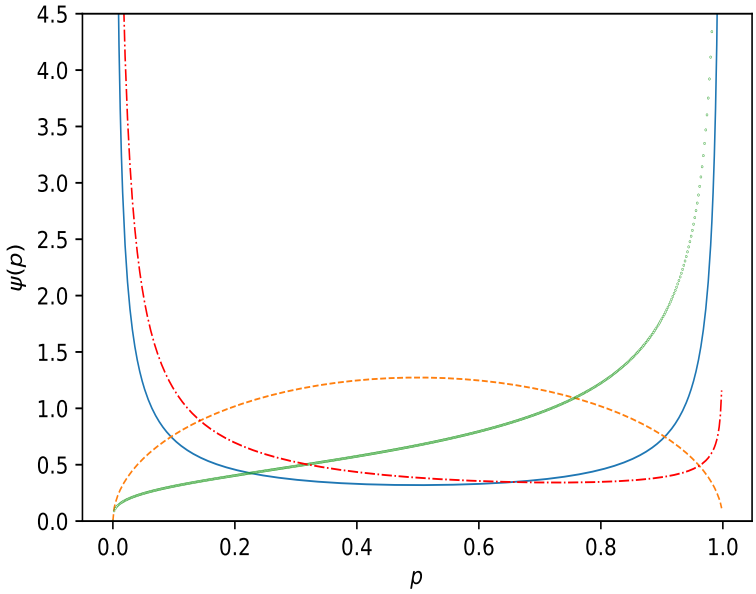

Figure 1: **Stationary distribution of the Wright-Fisher diffusion.** The graphs plot the density ψ, given by equation (49) for: $2\nu_1 = 2\nu_2 = 0.2$ (blue solid line), $2\nu_1 = 2\nu_2 = 1.5$ (orange dashed line), $2\nu_1 = 0.5, 2\nu_2 = 1.3$ (green dotted line) and $2\nu_1 = 0.7, 2\nu_2 = 0.2$ (red alternating dashes and dots).

The stationary distribution gives us some understanding of the longterm balance between the competing forces of mutation (which maintains genetic diversity) and the stochastic term (confusingly usually referred to as the genetic drift), which removes variation from the population. Figure 1 shows the density of the stationary distribution of the Wright-Fisher diffusion with mutation for a variety of parameters. When $2\nu_1$ and $2\nu_2$ are both bigger than 1, the stationary distribution is peaked around its mean, but when they are both less than one it has singularities at $\{0, 1\}$. Of course, if there is no mutation, then the process eventually becomes trapped in $\{0, 1\}$.

Finally let us demonstrate one very powerful technique that is often applied in settings where the speed measure is a stationary distribution. The idea is familiar from the study of discrete time and space Markov chains.

Definition 6.18. *A discrete time and space Markov chain with transition probabilities $p(i, j)$ is said to be* reversible *with respect to the stationary measure π if it satisfies the* detailed balance *equation:*

$$\pi(i)p(i, j) = \pi(j)p(j, i)$$

for all i and j in the state space.

For such chains we can say things about events backwards in time by considering the forwards in time transition probabilities. The analogue of the detailed balance equation for a one-dimensional diffusion is

$$\psi(x)p(t, x, y) = \psi(y)p(t, y, x) \quad \text{for all } x, y, t.$$

Now multiplying by arbitrary functions $f(x)$ and $g(y)$ in the domain of the generator of the diffusion

we obtain

$$\int \psi(x)f(x)\left(\int p(t,x,y)g(y)dy\right)dx = \int \psi(y)g(y)\left(\int p(t,y,x)f(x)dx\right)dy.$$

Now observe that the inner integrals are

$$\mathbb{E}[g(X_t)|X_0 = x] \quad \text{and} \quad \mathbb{E}[f(X_t)|X_0 = y]$$

and differentiate with respect to t at $t = 0$ to obtain

$$\int f(x)\mathcal{L}g(x)\psi(x)dx = \int \mathcal{L}f(y)g(y)\psi(y)dy. \tag{50}$$

Definition 6.19. *If the identity (50) is satisfied for all f and g, then ψ is called a* reversible stationary distribution *and we say that the diffusion is* reversible *with respect to ψ.*

Now suppose that the stationary distribution of the diffusion is given by $\psi(x) = m(x)/\int m(y)dy$. Then choosing f and g to vanish at the boundary of the domain to force the boundary terms to vanish when we integrate by parts (twice), we obtain

$$\begin{aligned}
\int_a^b f(x)\mathcal{L}g(x)m(x)dx &= \frac{1}{2}\int_a^b f(x)\frac{1}{m(x)}\frac{d}{dx}\left(\frac{1}{S'(x)}\frac{dg}{dx}\right)m(x)dx \\
&= \frac{1}{2}\int_a^b \frac{d}{dx}\left(\frac{1}{S'(x)}\frac{df}{dx}\right)g(x)dx \\
&= \frac{1}{2}\int_a^b \frac{1}{m(x)}\frac{d}{dx}\left(\frac{1}{S'(x)}\frac{df}{dx}\right)g(x)m(x)dx \\
&= \int_a^b \mathcal{L}f(x)g(x)m(x)dx,
\end{aligned}$$

so this is indeed a *reversible* stationary distribution.

7 PDEs that can be solved by running a Brownian motion (or a diffusion)

We now turn to higher dimensions and investigate some of the classes of equation that can be 'solved' by running a Brownian motion. We shall take our underlying diffusion to be Brownian motion, but by replacing it with a diffusion with generator $\mathcal{L}$, we obtain probabilistic representations of the corresponding equations with $\frac{1}{2}\Delta$ replaced by $\mathcal{L}$. Here we shall use the representations to deduce properties of Brownian motion. In general, where explicit solutions of the pde are not available, one can also use the diffusion to deduce properties of the solution to the pde. This section owes a great deal to lecture notes of Peter Mörters.

7.1 The Dirichlet Problem and exit times

Definition 7.1. *Let U be an open, bounded domain in $\mathbb{R}^d$ and let ∂U be its boundary. Suppose $\phi : \partial U \to \mathbb{R}$ is a continuous function on its boundary. A function $u : \overline{U} \to \mathbb{R}$, which is twice continuously differentiable on U and continuous on the closure $\overline{U}$ is a* solution to the Dirichlet problem *with boundary value ϕ, if*

$$\Delta u(x) = 0 \text{ for all } x \in U$$

and $u(x) = \phi(x)$ for $x \in \partial U$. Here Δ is the Laplace operator $\Delta = \sum_{i=1}^d D_{ii}$ and a function with $\Delta u(x) = 0$ for all $x \in U$ is said to be harmonic on U.

This problem was posed by Gauss in 1840. In fact Gauss thought he showed that there is always a solution, but his reasoning was wrong and Zaremba in 1909 and Lebesgue in 1913 gave counterexamples. However, if the domain is sufficiently nice there is a solution and in this case the solution can be represented and simulated using Brownian motion. To understand the connection between Brownian motion and harmonic functions, we have a closer look at Itô's formula.

Let $B = \{B(t) = (B^1(t), \dots, B^d(t)) : t \geq 0\}$ be a *d-dimensional Brownian motion* started in $x = (x^1, \dots, x^d)$, i.e. $B^1, \dots, B^d$ are independent standard Brownian motions started in $x^1, \dots, x^d$. Recall that

$$\langle B^k \rangle_t = t \text{ and } \langle B^k, B^l \rangle_t = 0 \text{ for all } l \neq k.$$

With this information Itô's formula reads

$$f(B(t)) - f(x) = \int_0^t \sum_{i=1}^d D_i f(B(s))\, dB_s^i + \frac{1}{2} \int_0^t \sum_{i=1}^d D_{ii} f(B(s))\, ds,$$

for $f : \mathbb{R}^d \to \mathbb{R}$ twice continuously differentiable. Noting that the last integrand is just $\Delta f(B(s))$ we obtain the fundamental relation.

Theorem 7.2. *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is harmonic on U and let $T = \inf\{t \geq 0 : B(t) \notin U\}$. Then $\{f(B(t)) : t \geq 0\}$ is a local martingale on $[0, T)$. More precisely, for all $t < T$,*

$$f(B(t)) = f(x) + \int_0^t \sum_{i=1}^d D_i f(B(s))\, dB_s^i.$$

Next we prove the probabilistic representation of the solutions of the Dirichlet problem. This theorem will be the key to several path properties of Brownian motion, which we derive in the sequel.

Theorem 7.3. *Suppose U is an open bounded set and suppose that u is a solution of the Dirichlet problem on U with boundary value ϕ. Define the stopping time*

$$T = \inf\{t \geq 0 : B(t) \notin U\}.$$

Then, for every $x \in U$,

$$u(x) = \mathbb{E}_x \left[\phi(B(T))\right], \tag{51}$$

where $\mathbb{E}_x$ refers to the expectation with respect to the Brownian motion started in x. In particular, the solution is uniquely determined by the boundary value.

Remark 7.4. *(i) We already saw a special case of this result in Lemma 6.6.*

(ii) We have assumed differentiability of u in the statement of the result. It is then not hard to check that u solves the Dirichlet problem. To circumvent this assumption, one can use an alternative characterisation of harmonic functions, as those functions whose value at each point $x \in U$ is equal to the average of their value over the surface of any sphere centred on x and completely contained in U. That this is true for the function u defined by (51) is an elementary application of the strong Markov property and symmetry of Brownian motion.

(iii) The solution $u(x)$ can be simulated using the formula (51), by running many independent Brownian motions, starting in $x \in U$, until they hit the boundary of U and letting $u(x)$ be the average of the values of ϕ on the hitting points.

(iv) It is a straightforward consequence of this representation that the solutions u to the Dirichlet problem always attain their maximum (and minimum) on the boundary of the domain.

To prove Theorem 7.3 we first show that the stopping time T is almost surely finite (and even has moments of all orders).

Lemma 7.5. *For all $0 < p < \infty$ we have $\sup_{x\in U} \mathbb{E}_x[T^p] < \infty$.*

Proof. We start by expressing the moment as a Lebesgue integral and using a change of variable,

$$\mathbb{E}_x[T^p] = \int_0^\infty \mathbb{P}_x[T^p \geq t]\, dt \quad = \quad \int_0^\infty ps^{p-1}\, \mathbb{P}_x[T \geq s]\, ds.$$

It thus suffices to show that, for some $q < 1$, $\mathbb{P}_x[T \geq k] \leq q^k$. To prove this, let $K = \sup\{|x - y| : x, y \in U\}$ be the diameter of U. If $x \in U$ and $|B_1 - x| > K$, then $B_1 \notin U$ and hence $T < 1$. Thus

$$\mathbb{P}_x[T < 1] \geq \mathbb{P}_x[|B_1 - x| > K] = \mathbb{P}_0[|B_1| > K] =: \tilde{p} > 0.$$

Letting $q = 1 - \tilde{p}$ we have shown $\mathbb{P}_x[T \geq 1] \leq q$ for all x, which is the start of an induction argument. Now we can use the Markov property and the inductive hypothesis to infer

$$\mathbb{P}_x[T \geq k] \leq \frac{1}{(2\pi)^{d/2}} \int_U e^{-|x-y|^2/2} \mathbb{P}_y[T \geq k-1]\, dy \leq q^{k-1} \mathbb{P}_x[B_1 \in U] \leq q^k.$$

This is what we had to show. □

Proof of Theorem 7.3. As $\Delta u(x) = 0$ for all $x \in U$ we see from Itô's formula that $u(B(s))$ is a continuous local martingale on the random interval $[0, T)$. Now (see the Appendix or recall B8.2/C8.1) there exists a time change $\gamma : [0, \infty) \to [0, T)$ such that $X = \{X_t = u(B(\gamma(t))) : t \geq 0\}$ is a bounded martingale with respect to the new filtration defined by $\mathcal{G}(t) := \mathcal{F}(\gamma(t))$. Being a bounded martingale, X converges, by the Martingale Convergence Theorem, almost surely and in L^1 to a limit X_∞ with $X_t = \mathbb{E}_x[X_\infty|\mathcal{G}(t)]$. Because T is almost surely finite and u is continuous on the closure of U, we have $X_\infty = u(B(T))$. Hence we infer

$$u(x) = X_0 = \mathbb{E}_x[X_0] = \mathbb{E}_x[X_\infty] = \mathbb{E}_x[u(B(T))].$$

As $B(T) \in \partial U$ we have that $u(B(T)) = \phi(B(T))$ and we are done. □

We now show a simple application of this result to the problem of *recurrence* and *transience* of Brownian motion in various dimensions.

Definition 7.6 (Recurrence, neighbourhood recurrence, transience)**.** *We say that a (Markov) process X with values in $\mathbb{R}^d$ is*

(i) recurrent*, if for every $x \in \mathbb{R}^d$ there is a (random) sequence $t_n \uparrow \infty$ such that $X(t_n) = x$. We say that x is* visited infinitely often*,*

(ii) neighbourhood recurrent*, if, for every $x \in \mathbb{R}^d$ and $\varepsilon > 0$, the ball $B(x, \varepsilon)$ around x of radius ε is visited infinitely often. Equivalently, every open set is visited infinitely often.*

(iii) transient*, if it converges to infinity almost surely.*

Theorem 7.7. *Brownian motion is*

(i) recurrent *in dimension $d = 1$,*

(ii) neighbourhood recurrent*, but not recurrent, in* $d = 2$*,*

(iii) transient *in dimension* $d \geq 3$*.*

We prove this through a series of lemmas. We begin with $d = 1$. Recall from the proof of Lemma 6.6 that in $d = 1$ and for $a < x < b$, if $T := \inf\{t \geq 0 : B_t \notin (a, b)\}$, then,

$$\mathbb{P}_x[B_T = a] = \frac{b-x}{b-a} \text{ and } \mathbb{P}_x[B_T = b] = \frac{x-a}{b-a}.$$

Lemma 7.8. *Let* $T_x := \inf\{t > 0 : B_t = x\}$*. Then* $\mathbb{P}_y[T_x < \infty] = 1$*.*

Proof. We may assume that $x = 0$ and, by reflection, $y > 0$. Then, by Lemma 6.6 and using that the exit time from any bounded interval is finite, e.g. by Lemma 7.5,

$$\mathbb{P}_y[T_0 < \infty] \geq \lim_{M\to\infty} \mathbb{P}_y[T_0 < T_{M-y}] = \lim_{M\to\infty} \frac{M-y}{M} = 1.$$

□

By Lemma 7.8, one dimensional Brownian motion eventually visits every point x. As T_x is a stopping time, by the strong Markov property, $\{B(T_x + t) : t \geq 0\}$ is again a Brownian motion, which visits every point. We wait until the new motion visits a point $y \neq x$, say at time T_y. Then $\{B(T_x + T_y + t) : t \geq 0\}$ is a Brownian motion started in y, which visits x again, and so forth. With a fixed positive probability it takes at least, say, one time unit before the motion started in y visits x. Because we have infinitely many independent trials for this experiment, there are infinitely many successes (by the Borel-Cantelli lemma). This proves that we visit x infinitely often, which means that the process is recurrent in $d = 1$ and (i) is proved. (Note that the last step was required because our definition of recurrence requires visits to take place at arbitrarily large times.)

Let us now move to dimensions $d \geq 2$. Start the motion at a point x contained in some annulus

$$x \in A := \{x \in \mathbb{R}^d : r \leq |x| \leq R\} \text{ for } 0 < r < R < \infty.$$

What is the probability that the Brownian motion hits the inner ring before it hits the outer ring? In order to copy the proof of the one-dimensional exit problem, we have to find harmonic functions u (that is functions with $\Delta u = 0$) on the annulus A.

By symmetry, such functions will be radially symmetric and so setting $u(x) = \psi(|x|)$ and writing the Laplacian in polar coordinates, we find

$$\psi''(\rho) + \frac{d-1}{\rho}\psi'(\rho) = 0.$$

This yields, for $\{|x| \neq 0\}$, general solutions of the form

$$u(x) = \psi(|x|) = \begin{cases} c_1|x| + c_2 & \text{if } d = 1, \\ c_1 \log|x| + c_2 & \text{if } d = 2, \\ c_1|x|^{2-d} + c_2 & \text{if } d \geq 3. \end{cases} \tag{52}$$

Now we can use Theorem 7.3. Define stopping times

$$T_r = \inf\{t > 0 : |B_t| = r\} \text{ for } r > 0.$$

Letting $T = T_r \wedge T_R$ be the first exit time of the Brownian motion B from A, we have

$$u(x) = \mathbb{E}_x[u(B_T)] = \psi(r)\mathbb{P}_x[T_r < T_R] + \psi(R)(1 - \mathbb{P}_x[T_r < T_R]).$$

Rearranging,

$$\mathbb{P}_x[T_r < T_R] = \frac{\psi(R) - \psi(|x|)}{\psi(R) - \psi(r)}.$$

Remark 7.9. *Notice by Lemma 6.7 that this is $\mathbb{P}[T_r < T_R]$ for the one-dimensional diffusion on (r, R) with generator*

$$\frac{1}{2}\frac{\partial^2 f}{\partial x^2} + \frac{d-1}{2x}\frac{\partial f}{\partial x}.$$

This is the Bessel process of dimension d, which is the modulus of d-dimensional Brownian motion. It is a (convenient) special property of Brownian motion that its modulus is once again a diffusion.

The scale function for this diffusion is as in (52) (with $c_1 = 1$, $c_2 = 0$).

Summarizing we have:

Lemma 7.10. *Suppose B is a Brownian motion in dimension $d \geq 2$ started at a point x in the annulus*

$$A := \{x \in \mathbb{R}^d \,:\, r \leq |x| \leq R\},$$

where $0 < r < R < \infty$. Then, if $d = 2$,

$$\mathbb{P}_x\{T_r < T_R\} = \frac{\log R - \log|x|}{\log R - \log r}.$$

In dimension $d \geq 3$,

$$\mathbb{P}_x\{T_r < T_R\} = \frac{R^{2-d} - |x|^{2-d}}{R^{2-d} - r^{2-d}}.$$

Now consider dimension $d = 2$, fix $r = \varepsilon$ and let $R \uparrow \infty$ in the previous formula. Then we obtain, for arbitrary $x \notin B(0, \varepsilon)$,

$$\mathbb{P}_x[T_\varepsilon < \infty] = \mathbb{P}_x\Big[\bigcup_{R>0}\{T_\varepsilon < T_R\}\Big] = \lim_{R\to\infty}\mathbb{P}_x[T_\varepsilon < T_R] = \lim_{R\to\infty}\frac{\log R - \log|x|}{\log R - \log\varepsilon} = 1.$$

Hence the motion eventually hits every small ball around 0. Then, of course, it must hit every small ball eventually and, as before (using the strong Markov property) it hits every small ball infinitely often. This proves that in $d = 2$ Brownian motion is *neighbourhood recurrent*. □

A compact set $A \subseteq \mathbb{R}^d$ with the property that

$$\mathbb{P}_x[B_t \in A \text{ for some } t > 0] = 0 \text{ for all } x \in \mathbb{R}^d$$

is called a *polar set*. We show that points are polar sets for Brownian motion in $d \geq 2$, which proves that in these dimensions Brownian motion cannot be recurrent.

Lemma 7.11. *For Brownian motion in $d \geq 2$ points are polar sets.*

Proof. It suffices to consider the point 0. Define

$$S_0 = \inf\{t > 0 \,:\, B_t = 0\}.$$

Fix $R > 0$ and let $r \downarrow 0$. First let $x \neq 0$. Then

$$\mathbb{P}_x[S_0 < T_R] \leq \lim_{r\to 0}\mathbb{P}_x[T_r < T_R] = 0.$$

As this holds for all R and $T_R \to \infty$ as $R \to \infty$, by continuity of Brownian motion, we have $\mathbb{P}_x[S_0 < \infty] = 0$ for all $x \neq 0$. To extend this to $x = 0$ we observe that the Markov property implies

$$\mathbb{P}_0[B_t = 0 \text{ for some } t \geq \varepsilon] = \mathbb{E}_0\Big[\mathbb{P}_{B_\varepsilon}[T_0 < \infty]\Big] = 0\,,$$

noting that $B_\varepsilon \neq 0$ almost surely. Hence

$$\mathbb{P}_0[S_0 < \infty] = \lim_{\varepsilon\downarrow 0}\mathbb{P}_0[B_t = 0 \text{ for some } t \geq \varepsilon] = 0.$$

□

Remark 7.12. *(i) It is clear that for all x and all $t > 0$ we have $\mathbb{P}[B_t \neq x] = 1$. Our statement however is that, for all x, $\mathbb{P}[B_t \neq x \text{ for all } t] = 1$. This is much harder, because uncountable unions of nullsets usually fail to be nullsets (to see this try to take the quantifier 'for all x' inside the probability).*

(ii) The proof in the case $d = 2$ also shows that in $d \geq 3$ line segments are polar sets. This does not hold in dimension $d = 2$.

To complete the picture we have to show that Brownian motion is transient in dimensions $d \geq 3$. First observe that the proof of neighbourhood recurrence in dimension $d = 2$ does not work here, because in $d \geq 3$,

$$\mathbb{P}_x[T_r < \infty] = \lim_{R\to\infty} \mathbb{P}_x[T_r < T_R] = \frac{r^{d-2}}{|x|^{d-2}} < 1$$

for all $|x| > r$. However, we can use this formula to show transience.

Lemma 7.13. *In $d \geq 3$, $\lim_{t\to\infty} B_t = \infty$.*

Proof. Consider the event

$$A_n := \Big\{ |B_t| > \sqrt{n} \text{ for all } t \geq T_n \Big\},$$

and recall that $T_n < \infty$ almost surely by Lemma 7.5. By the strong Markov property we have

$$\mathbb{P}_x[A_n^c] = \mathbb{E}_x\Big[\mathbb{P}_{B(T_n)}[T_{\sqrt{n}} < \infty]\Big] = \mathbb{E}_x\Big[\lim_{R\to\infty} \mathbb{P}_{B(T_n)}[T_{\sqrt{n}} < T_R]\Big] = \Big(\frac{1}{\sqrt{n}}\Big)^{d-2} \longrightarrow 0.$$

Now let A be the event that infinitely many of the events A_n occur. We have

$$\mathbb{P}_x[A] = \mathbb{P}_x\Big[\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\Big] = \limsup_{n\to\infty} \mathbb{P}_x\Big[\bigcup_{k=n}^{\infty} A_k\Big] \geq \limsup_{n\to\infty} \mathbb{P}_x[A_n] = 1\,,$$

hence A holds almost surely, which means that, for infinitely many n, the path eventually does not return inside the ball of radius $\sqrt{n}$, so it must go to infinity almost surely. This concludes the proof of Theorem 7.7. □

For Brownian motion, exit problems from domains are closely linked to the behaviour of harmonic functions (and the Laplace operator) on the domain. We conclude this section with one more problem from this realm.

Suppose we start a d-dimensional Brownian motion at some point x inside an open bounded domain U. Let

$$T = \inf\{t \geq 0 \,:\, B(t) \notin U\}.$$

We ask for the distribution of the point $B(T)$ where the Brownian motion first leaves U. This distribution, on the boundary ∂U, is usually called the *harmonic measure on ∂U*. Of course it depends on the starting point x, but by a famous theorem of Harnack all these measures are absolutely continuous with respect to each other.

We will concentrate on the case of exit distributions from the unit ball

$$U = \{x \in \mathbb{R}^d \,:\, |x| < 1\}.$$

If $x = 0$ the distribution of $B(T)$ is (by symmetry) the uniform distribution π, but if x is another point it is an interesting problem to determine this distribution in terms of a probability density. Since the solution to the Dirichlet problem can be written as $\mathbb{E}[g(B_T)]$, it is not surprising that the harmonic measure is intricately connected with the Dirichlet problem.

Theorem 7.14 (Poisson's formula)**.** *Suppose that $A \subseteq \partial U$ is a Borel subset of the unit sphere $\partial U \subseteq \mathbb{R}^d$ for $d \geq 2$. Then, for all $x \in U$,*

$$\mathbb{P}_x[B(T) \in A] = \int_A \frac{1-|x|^2}{|x-y|^d}\, d\pi(y),$$

where π denotes the uniform distribution on the unit sphere.

Remark 7.15. *The density is of course the Poisson kernel often appearing in courses on complex analysis and appears frequently in potential theory.*

Proof. To prove the theorem we actually show that for every bounded measurable $f : \mathbb{R}^d \to \mathbb{R}$ we have

$$\mathbb{E}_x[f(B(T))] = \int_{\partial U} \frac{1-|x|^2}{|x-y|^d} f(y)\, d\pi(y). \tag{53}$$

It suffices to consider C^∞-functions. By Theorem 7.3, to prove (53) we just have to show that the right hand side, as a function of $x \in U$, defines a solution to the Dirichlet problem on U with boundary value f.

To check this, one first checks that $\frac{1-|x|^2}{|x-y|^d}$ is harmonic on U, which is just a calculation, and then argues that it is allowed to differentiate twice under the integral sign. We omit the details, referring readers to Durrett. To check the boundary condition first look at the case $f \equiv 1$. Then we have to show that, for all $x \in U$,

$$I(x) := \int_{\partial U} \frac{1-|x|^2}{|x-y|^d}\, \pi(dy) \equiv 1.$$

This can be computed as well, but we argue mathematically. Observe that $I(0) = 1$, I is invariant under rotation and $\Delta I = 0$ on U. Now let $x \in U$ with $|x| = r < 1$ and let $\tau := \inf\{t : |B_t| > r\}$. By Theorem 7.3

$$I(0) = \mathbb{E}_0[I(B_\tau)] = I(x)\,,$$

using rotational invariance in the second step. Hence $I \equiv 1$. Now we show that the right hand side of (53) can be extended continuously to all points $y \in \partial U$ by $f(y)$. We write D_0 for ∂U with a δ-neighbourhood $U(y,\delta)$ removed and $D_1 = \partial U \setminus D_0$. We have, using that $I \equiv 1$, for all $x \in U(y,\delta/2) \cap U$,

$$\begin{aligned}
&\Big| f(y) - \int_{\partial U} \frac{1-|x|^2}{|x-z|^d} f(z)\, d\pi(z) \Big| \\
&= \Big| \int_{\partial U} \frac{1-|x|^2}{|x-z|^d} (f(y) - f(z))\, d\pi(z) \Big| \\
&\leq 2\|f\|_\infty \int_{D_0} \frac{1-|x|^2}{|x-z|^d}\, d\pi(z) + \sup_{z \in D_1} |f(y) - f(z)|.
\end{aligned}$$

For fixed $\delta > 0$ the first term goes to 0 as $x \to y$ by the Dominated Convergence Theorem, whereas the second can be made arbitrarily small by choice of δ (since f is continuous). □

The Poisson integral formula of (53) is, of course, very special, whereas our representation of solutions to the Poisson equation is very general. In fact it extends to highly irregular boundaries where proofs of existence and uniqueness of solutions to the Dirichlet problem can depend on fine properties of the Brownian motion.

7.2 The Poisson problem and occupation times

We now turn to a multidimensional version of the results of Section 6.3 and ask how much time does a d-dimensional Brownian motion spend in a domain U before it leaves the domain? As before, the answer can be expressed in terms of solutions to the *Poisson problem*.

Definition 7.16 (Poisson problem)**.** *Let U be an open bounded domain and $u : \overline{U} \to \mathbb{R}$ be a continuous function, which is twice continuously differentiable on U. Let $g : U \to \mathbb{R}$ be continuous. Then u is said to be the* solution of Poisson's problem for g *if $u(x) = 0$ for all $x \in \partial U$ and*

$$\frac{1}{2}\Delta u(x) = -g(x) \text{ for all } x \in U.$$

The $1/2$ in front of the Laplacian is, of course, a probabilistic convention.

Theorem 7.17. *Suppose g is bounded and u a bounded solution of Poisson's problem for g. Then this solution has the form*

$$u(x) = \mathbb{E}_x\Big[\int_0^T g(B_t)\,dt\Big] \text{ for } x \in U\,,$$

where $T := \inf\{t > 0 : B(t) \notin U\}$. In particular, the solution, if it exists, is always uniquely determined.

Just as in $d = 1$, if $g \equiv 1$, then $u(x) = \mathbb{E}_x[T]$.

Rather than repeating our one-dimensional argument, we mimic our approach to the Dirichlet problem and use Itô's formula to find a local martingale.

Lemma 7.18. *Let U be a bounded open domain and $T := \inf\{t > 0 : B(t) \notin U\}$. If u is a solution to Poisson's problem for g, then $M = \{M_t : t \geq 0\}$ defined by*

$$M_t = u(B(t)) + \int_0^t g(B(s))\,ds$$

is a local martingale on $[0, T)$.

Proof. Applying Poisson's equation and Itô's formula gives, for all $t < T$,

$$\begin{aligned} u(B(t)) + \int_0^t g(B(s))\,ds &= u(B(t)) - \frac{1}{2}\int_0^t \Delta u(B(s))\,ds \\ &= u(B(0)) + \int_0^t \sum_{i=1}^d D_i u(B(s))\,dB^i(s), \end{aligned}$$

which is a local martingale on $[0, T)$. □

Proof of Theorem 7.17: M is a local martingale on $[0, T)$ and (c.f. the proof of Theorem 7.3) we let $\gamma : [0, \infty) \to [0, T)$ be the time change such that $\{M(\gamma(t)) : t \geq 0\}$ is a martingale. As u and g are bounded,

$$\sup_{t\geq 0} |M(\gamma(t))| \leq \|u\|_\infty + T\|g\|_\infty.$$

The right hand side is L^2-integrable by Lemma 7.5 and hence we have that $\{M(\gamma(t)) : t \geq 0\}$ is a uniformly integrable martingale. The Martingale Convergence Theorem tells us that $\lim_{t\to\infty} M(\gamma(t)) =: M_\infty$ exists almost surely and $\mathbb{E}_x[M_0] = \mathbb{E}_x[M_\infty]$. We can now use continuity of u and g to obtain

$$u(x) = \mathbb{E}_x[M_0] = \mathbb{E}_x[M_\infty] = \mathbb{E}_x\Big[\lim_{t\uparrow T}\big\{u(B_t) + \int_0^t g(B_s)\,ds\big\}\Big] = \mathbb{E}_x\Big[\int_0^T g(B_s)\,ds\Big],$$

as required. □

We now address the following question: given a bounded open domain $U \subseteq \mathbb{R}^d$, does Brownian motion spend an infinite or a finite amount of time in U? It is not surprising that the answer depends on the dimension; more interestingly, it depends neither on U, nor on the starting point.

Theorem 7.19. *Let $U \subseteq \mathbb{R}^d$ be a bounded open domain and $x \in \mathbb{R}^d$ arbitrary. Then, if $d \leq 2$,*

$$\int_0^\infty \mathbf{1}_U(B_t)\,dt = \infty \qquad \mathbb{P}_x\text{-almost surely,}$$

and, if $d \geq 3$,

$$\mathbb{E}_x \int_0^\infty \mathbf{1}_U(B_t)\,dt < \infty.$$

Proof. As U is contained in a ball, and contains a ball, it suffices to show this for balls. By shifting, we can even restrict to balls $U = B(0,r)$ centred in the origin. Let us start with the first claim. We let $d \leq 2$ and let $G = B(0,2r)$. Let $T_0 = 0$ and, for all $k \geq 1$, let

$$S_k = \inf\{t > T_{k-1} \,:\, B_t \in U\} \text{ and } T_k = \inf\{t > S_k \,:\, B_t \notin G\}.$$

For $d \leq 2$, there is neighbourhood recurrence and these form an infinite sequence of stopping times. From the strong Markov property and rotational invariance we infer, for $k \geq 1$,

$$\begin{aligned}\mathbb{P}_x\Big[\int_{S_k}^{T_k} \mathbf{1}_U(B_t)\,dt \geq s \,\Big|\, \mathcal{F}(S_k)\Big] &= \mathbb{P}_{B(S_k)}\Big[\int_0^{T_1} \mathbf{1}_U(B_t)\,dt \geq s\Big] \\ &= \mathbb{E}_x\Big[\mathbb{P}_{B(S_k)}\Big[\int_0^{T_1} \mathbf{1}_U(B_t)\,dt \geq s\Big]\Big] = \mathbb{P}_x\Big[\int_{S_k}^{T_k} \mathbf{1}_U(B_t)\,dt \geq s\Big].\end{aligned}$$

The second expression does not depend on k, so that the random variables

$$\int_{S_k}^{T_k} \mathbf{1}_U(B_t)\,dt$$

are independent and identically distributed. As they are not identically zero, but non-negative, they have positive expected value and, by the strong law of large numbers,

$$\int_0^\infty \mathbf{1}_U(B_t)\,dt = \lim_{n\to\infty} \sum_{k=1}^{n} \int_{S_k}^{T_k} \mathbf{1}_U(B_t)\,dt = \infty,$$

which proves the first claim.

For the second claim, first let f be non-negative and measurable. Fubini's Theorem implies

$$\begin{aligned}\mathbb{E}_x \int_0^\infty f(B_t)\,dt &= \int_0^\infty \mathbb{E}_x f(B_t)\,dt = \int_0^\infty \int p_t(x,y) f(y)\,dy\,dt \\ &= \int \int_0^\infty p_t(x,y)\,dt\, f(y)\,dy,\end{aligned}$$

where $p_t(x,y) = (2\pi t)^{-d/2} \exp(-|x-y|^2/2t)$ is the transition density of Brownian motion. Note that for fixed x, y, for large t we have $p_t(x,y) \sim (2\pi t)^{-d/2}$, hence $\int_0^\infty p_t(x,y)\,dt = \infty$ if $d \leq 2$. In $d \geq 3$, however, we can define, for $x \neq y$,

$$G(x,y) := \int_0^\infty p_t(x,y)\,dt < \infty\,,$$

a quantity called *the potential kernel*. It can be calculated explicitly, through the change of variables $s = |x-y|^2/2t$:

$$\begin{aligned} G(x,y) &= \int_0^\infty \frac{1}{(2\pi t)^{-d/2}} e^{-|x-y|^2/2t}\, dt = \int_\infty^0 \Big(\frac{s}{\pi|x-y|^2}\Big)^{d/2} e^{-s}\Big(-\frac{|x-y|^2}{2s^2}\Big)\, ds \\ &= \frac{|x-y|^{2-d}}{2\pi^{d/2}} \int_0^\infty s^{(d/2)-2} e^{-s}\, ds = \frac{\Gamma(d/2-1)}{2\pi^{d/2}} |x-y|^{2-d}, \end{aligned}$$

where $\Gamma(x) = \int_0^\infty s^{x-1}e^{-s}\, ds$ is the gamma function. We denote the constant in the last term by $c(d)$. To summarise what we have done so far,

$$\mathbb{E}_x \int_0^\infty f(B_t)\, dt = \int G(x,y) f(y)\, dy = c(d) \int \frac{f(y)}{|x-y|^{d-2}}\, dy. \tag{54}$$

To complete the proof, let $f = \mathbf{1}_{B(0,r)}$. If $x = 0$ we may change to polar coordinates and then, writing $C(d)$ for a constant depending only on d,

$$\mathbb{E}_0 \int_0^\infty \mathbf{1}_{B(0,r)}(B_s)\, ds = \int_{B(0,r)} G(0,y)\, dy = C(d) \int_0^r s^{d-1} s^{2-d}\, ds = (C(d)/2) r^2 < \infty.$$

To start in an arbitrary $x \neq 0$ we consider a Brownian motion started in 0 and a stopping time T, which is the first hitting time of the sphere $\partial B(0,|x|)$. Using spherical symmetry and the strong Markov property we obtain

$$\mathbb{E}_x \int_0^\infty \mathbf{1}_{B(0,r)}(B_s)\, ds = \mathbb{E}_0 \int_T^\infty \mathbf{1}_{B(0,r)}(B_s)\, ds \leq \mathbb{E}_0 \int_0^\infty \mathbf{1}_{B(0,r)}(B_s)\, ds < \infty.$$

□

We shall now study the *expected occupation measures* of a Brownian motion in a Borel subset A of some bounded open domain U. These are defined as

$$\mathbb{E}_x \int_0^T \mathbf{1}_A(B_s)\, ds,$$

in other words the expected time Brownian motion spends in A before leaving the domain U.

Theorem 7.20. *Let $d \geq 3$ and $U \subseteq \mathbb{R}^d$ a bounded domain, let T be the first exit time from the domain. Define the potential kernel $G(x,y) = c(d)|x-y|^{2-d}$ where $c(d) = \frac{\Gamma(d/2-1)}{2\pi^{d/2}}$, for $x \in U$ recall the definition of the harmonic measure*

$$\mu(x, dz) = \mathbb{P}_x\{B_T \in dz\}$$

and define the Green's function *of the domain U as*

$$G_U(x,y) = G(x,y) - \int_{\partial U} G(z,y)\, \mu(x,dz) \text{ for } x, y \in U \text{ with } x \neq y.$$

Then we have, for all $x \in U$ and $A \subseteq U$ Borel,

$$\mathbb{E}_x\Big[\int_0^T \mathbf{1}_A(B_s)\, ds\Big] = \int_A G_U(x,y)\, dy.$$

Remark 7.21. *(i) Probabilistically the Green's function $G_U(x,y)$ is the density of the expected occupation measure of a Brownian motion started in x and stopped upon leaving U.*

(ii) Note that the theorem also tells us that the bounded solution of the Poisson problem for g is $\int G_U(x,y)g(y)\,dy$. This is, of course, the PDE interpretation of the Green's function.

(iii) Letting $g(y)\,dy$ in Theorem 7.17 converge to the Dirac measure δ_y we recover the physical interpretation of the Green's function $G_U(x,y)$ as the electrostatic potential of a unit mass at y with ∂U grounded.

(iv) Although in $d=1,2$ it is not possible to put $G(x,y)=\int_0^\infty p_t(x,y)\,dt$, it is also possible to define a potential kernel G such that the Theorem 7.20 holds true. As in $d=3$ these potential kernels are just constant multiples of $u(x-y)$ for the harmonic functions u on the punctured disc, see (52).

Proof of Theorem 7.20.

Suppose that g is a non-negative, bounded function with compact support. Then, by Theorem 7.19, we can define

$$w(x) := \mathbb{E}_x \int_0^\infty g(B_s)\,ds < \infty.$$

By the strong Markov property,

$$w(x) = \mathbb{E}_x \int_0^T g(B_s)\,ds + \mathbb{E}_x w(B_T).$$

We already know that

$$w(x) \;=\; \int_0^\infty \mathbb{E}_x[g(B_s)]\,ds = \int_0^\infty \int p_s(x,y)g(y)\,dy\,ds = \int G(x,y)\,g(y)\,dy.$$

Substituting (and partitioning over the possible values of B_T) this gives

$$\mathbb{E}_x\Big[\int_0^T g(B_s)\,ds\Big] = \int G(x,y)\,g(y)\,dy - \mathbb{E}_x \int G(B_T,y)\,g(y)\,dy = \int G_U(x,y)g(y)\,dy,$$

which is the required formula. □

In the case $U=B(0,1)$ this can be calculated explicitly, using Theorem 7.17. We have

$$\begin{aligned} G_U(x,y) &= G(x,y) - \int_{\partial U} \frac{1-|x|^2}{|x-z|^d}\,G(z,y)\,\pi(dz) \\ &= \frac{\Gamma(d/2-1)}{(2\pi)^{d/2}} \int_{\partial U} \frac{1-|x|^2}{|x-z|^d}\Big(|x-y|^{2-d} - |z-y|^{2-d}\Big)\pi(dz). \end{aligned}$$

The latter integral can be evaluated as follows.

Lemma 7.22. *Let $d\geq 3$ and $U\subseteq\mathbb{R}^d$ the unit disc and let T be the first exit time from U. Then,*

$$\int_{\partial U} \frac{1-|x|^2}{|x-z|^d}\,G(z,y)\,\pi(dz) = c(d)\frac{|y|^{d-2}}{|\,x|y|^2 - y\,|^{d-2}} \text{ for all } x,y\in U\ .$$

Proof. We already know from the proof of Poisson's formula and the definition of G_U that, for every $y\in U$,

$$u(x) := G(x,y) - G_U(x,y) = \int_{\partial U} \frac{1-|x|^2}{|x-z|^d} G(z,y)\pi(dz)$$

can be extended continuously to the boundary of U to give the solution to the Dirichlet problem with boundary function $\phi(x) = G(x, y)$. By uniqueness it suffices to check that

$$v(x) := c(d)\,\frac{|y|^{d-2}}{|\,x|y|^2 - y\,|^{d-2}}$$

also has this property, so that $v = u$ and we are done. To see this note that

$$v(x) = c(d)|y|^{2-d}u(x - y/|y|^2),$$

where $u(x-y)$ is defined in (52). Since u is harmonic on any punctured disc, and $y/|y|^2 \notin U$, we deduce that v is harmonic on U. Clearly, v is continuous on $\overline{U}$. To determine the value on the boundary observe that, if $|x| = 1$,

$$|\,x|y| - y/|y|\,|^2 = |x|^2|y|^2 - 2x \cdot y + 1 = |y|^2 - 2x \cdot y + |x|^2 = |x - y|^2.$$

Therefore, on $|x| = 1$

$$v(x) = c(d)\frac{1}{|x|y| - y/|y||^{d-2}} = \frac{c(d)}{|x - y|^{d-2}} = G(x, y),$$

as required. □

7.3 The Feynman-Kac formula

For our final example of equations that can be solved by running a Brownian motion, we consider a parabolic equation.

Definition 7.23 (Heat equation with dissipation)**.** *A continuous function $u : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$, which is twice continuously differentiable, is said to satisfy the heat equation with heat dissipation rate $c : (0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ if we have*

(i) u is continuous at each point of $\{0\} \times \mathbb{R}^d$ and $u(0, x) = f(x)$,

(ii) $\dfrac{\partial u}{\partial t}(t, x) = \dfrac{1}{2}\Delta u(t, x) - c(t, x)u(t, x)$ *on* $(0, \infty) \times \mathbb{R}^d$.

If $c(t, x) \geq 0$ then $u(t, x)$ describes the temperature at time t at x for a heat flow with cooling. Here the initial temperature distribution is given by f and $c(t, x)$ describes the heat dissipation rate at point x at time t. As usual, the first step is to find a local martingale.

Lemma 7.24. *If u is a solution to the heat equation with dissipation rate c, then $M = \{M_s : s \geq 0\}$ with*

$$M_s := u(t - s, B_s) \exp\Big(- \int_0^s c(t - r, B_r)\, dr\Big)$$

is a local martingale on $[0, t)$.

Proof. Let us write $c_s^t = -\int_0^s c(t - r, B_r)\, dr$. We apply Itô's formula (with the semimartingales $X_s^0 = t - s$, $X_s^i = B_s^i$ for $1 \leq i \leq d$ and $X_s^{d+1} = c_s^t$).

$$\begin{aligned} &u(t - s, B_s)\exp(c_s^t) - u(t, B_0) \\ &= \int_0^s -\frac{\partial u}{\partial t}(t - r, B_r)\exp(c_r^t)\, dr + \sum_{j=1}^{d}\int_0^s \exp(c_r^t)\frac{\partial u}{\partial x_j}(t - r, B_r)\, dB_r^j \\ &\quad + \int_0^s u(t - r, B_r)\exp(c_r^t)\, dc_r^t + \frac{1}{2}\int_0^s \Delta u(t - r, B_r)\exp(c_r^t)\, dr, \end{aligned}$$

since we have

$$\langle X^i, X^j\rangle_t = \begin{cases} t & \text{if } 1 \le i = j \le d \\ 0 & \text{otherwise.} \end{cases}$$

Using $dc_r^t = -c(t-r, B_r)\,dr$ and rearranging, the right hand side is

$$= \int_0^s \Big(-\frac{\partial u}{\partial t} - cu + \frac{1}{2}\Delta u\Big)(t-r, B_r)\exp(c_r^t)\,dr + \sum_{j=1}^d \int_0^s \exp(c_r^t)\frac{\partial u}{\partial x_j}(t-r, B_r)\,dB_r^j.$$

This proves the claim, because $-\frac{\partial u}{\partial t} - cu + \frac{1}{2}\Delta u = 0$ and the second term is a local martingale. □

The resulting representation theorem for the solutions of our heat equation with dissipation term is called the Feynman-Kac formula.

Theorem 7.25 (Feynman-Kac formula)**.** *Suppose that the dissipation rate c is bounded and u is a solution of the heat equation with dissipation rate c, which is bounded on every set $[0,t]\times\mathbb{R}^d$. Then u is uniquely determined and satisfies*

$$u(t,x) = \mathbb{E}_x\Big[f(B_t)\exp\Big(-\int_0^t c(t-r, B_r)\,dr\Big)\Big].$$

Proof. Under our assumptions on c and u, M is a bounded martingale on $[0,t)$ and $M_t = \lim_{s\uparrow t} M_s = f(B_t)\exp(c_t^t)$. Since M is uniformly integrable we deduce that

$$\mathbb{E}_x\Big[f(B_t)\exp\Big(-\int_0^t c(t-r, B_r)\,dr\Big)\Big] = \mathbb{E}_x[M_t] = \mathbb{E}_x[M_0] = u(t,x).$$

□

Example 7.26 (Brownian motion in a soft potential)**.** *Imagine that the bounded function $c:\mathbb{R}^d\to[0,\infty)$ defines a potential landscape, so that it is hard for a particle to go through the hills of c and easier to go through the valleys. To model this, suppose for example, that for a particle following the path $\{B(t) : t\ge 0\}$, the probability of survival up to time T is*

$$\exp\Big(-\int_0^T c(B_r)\,dr\Big).$$

It is possible to construct such a process $\{X(t) : t\ge 0\}$ with values in $\mathbb{R}^d\cup\{\dagger\}$, called a killed Brownian motion*. If $u_1:[0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ is a solution of our problem with dissipation rate c and initial value $f\equiv 1$, which is bounded on the set $[0,T]\times\mathbb{R}^d$, then the probability that a path $\{X(t) : t\ge 0\}$ started in x survives up to time T is*

$$u_1(T,x) := \mathbb{E}_x\Big[\exp\Big(-\int_0^T c(B_r)\,dr\Big)\Big].$$

Suppose now f is a bounded function and we are interested in the expected value of $f(X(T))$ for the killed Brownian motion X conditioned on survival up to time T. First note that we do not expect the result to be the same as with ordinary Brownian motion, as killed Brownian motion conditioned on survival up to time T is intuitively more likely to go through the valleys of c than over the hills. Suppose that $u_2:[0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ is a solution of our problem with initial value f, which is bounded on $[0,T]\times\mathbb{R}^d$, then

$$\mathbb{E}_x[f(X(T))\,|\,X(T)\ne\dagger] = \frac{\mathbb{E}_x\big[f(B_T)\exp\big(-\int_0^T c(B_r)\,dr\big)\big]}{\mathbb{E}_x\big[\exp\big(-\int_0^T c(B_r)\,dr\big)\big]} = \frac{u_2(T,x)}{u_1(T,x)}.$$

The process X conditioned on survival up to time T, i.e. on $\{X(T) \neq \dagger\}$, is called Brownian motion in the soft potential c and, if we denote it by $Y = \{Y_t : t \in [0,T]\}$, its marginal distributions at the endpoint are given by the Feynman-Kac formula as

$$\mathbb{E}[f(Y(T))] = \frac{u_2(T,x)}{u_1(T,x)}.$$

Often, especially in mathematical finance, the Feynman-Kac formula is expressed in 'backwards time'. It can be thought of as providing a solution to a PDE subject to a terminal condition. To alleviate notation we present this in one dimension and set $c = 0$, but this time we replace Brownian motion by a more general diffusion.

Theorem 7.27 (Another Feynman-Kac stochastic representation)**.** *Assume that the function F solves the boundary value problem*

$$\begin{aligned} \frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x) = 0 \qquad 0 \leq t \leq T, \\ F(T,x) = \Phi(x). \end{aligned} \tag{55}$$

Define $\{X_t\}_{0\leq t\leq T}$ to be the solution of the stochastic differential equation

$$dX_t = \mu(t,X_t)dt + \sigma(t,X_t)dW_t, \qquad 0 \leq t \leq T,$$

where $\{W_t\}_{t\geq 0}$ is standard Brownian motion under the measure $\mathbb{P}$. If

$$\int_0^T \mathbb{E}\left[\left(\sigma(t,X_t)\frac{\partial F}{\partial x}(t,X_t)\right)^2\right] dt < \infty, \tag{56}$$

then

$$F(t,x) = \mathbb{E}^{\mathbb{P}}\left[\Phi(X_T)\middle|\, X_t = x\right].$$

Proof: We apply Itô's formula to $\{F(s,X_s)\}_{t\leq s\leq T}$.

$$\begin{aligned} F(T,X_T) &= F(t,X_t) + \int_t^T \left\{\frac{\partial F}{\partial s}(s,X_s) + \mu(s,X_s)\frac{\partial F}{\partial x}(s,X_s) + \frac{1}{2}\sigma^2(s,X_s)\frac{\partial^2 F}{\partial x^2}(s,X_s)\right\} ds \\ &\qquad + \int_t^T \sigma(s,X_s)\frac{\partial F}{\partial x}(s,X_s)dW_s. \end{aligned} \tag{57}$$

Now using assumption (56), which is enough to ensure that the Itô integral exists and defines a mean zero local martingale,

$$\mathbb{E}\left[\int_t^T \sigma(s,X_s)\frac{\partial F}{\partial x}(s,X_s)dW_s \middle| X_t = x\right] = 0.$$

Moreover, since F satisfies (55), the deterministic integral on the right hand side of (57) vanishes, so, taking expectations,

$$\mathbb{E}\left[F(T,X_T)\middle|\, X_t = x\right] = F(t,x)$$

and substituting $F(T,X_T) = \Phi(X_T)$ gives the required result. □

Exercise 7.28. *Convince yourself that the solution to the boundary value problem*

$$\frac{\partial F}{\partial t}(t,x)+\mu(t,x)\frac{\partial F}{\partial x}(t,x)+\frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x)-V(x,t)F(x,t)+f(x,t)=0 \qquad 0\le t\le T,$$
$$F(T,x)=\Phi(x).$$

can be written as

$$F(t,x)=\mathbb{E}^{\mathbb{P}}\left[\int_t^T e^{-\int_t^r V(X_u,u)du}f(X_r,r)dr+e^{-\int_t^T V(X_u,u)du}\Phi(X_T)\middle| X_t=x\right].$$

Example 7.29. *Solve*

$$\begin{aligned}\frac{\partial F}{\partial t}+\frac{1}{2}\frac{\partial^2 F}{\partial x^2}&=0,\\ F(T,x)&=\Phi(x).\end{aligned} \tag{58}$$

Solution: The corresponding stochastic differential equation is

$$dX_t=dW_t$$

so, by the Feynman-Kac representation,

$$F(t,x)=\mathbb{E}\left[\Phi(W_T)|\,W_t=x\right].$$

In fact we knew this already. The transition density of Brownian motion is

$$p(t,x,y)=\frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{(x-y)^2}{2t}\right). \tag{59}$$

This gives

$$\mathbb{E}\left[\Phi(W_T)|\,W_t=x\right]=\int p\left(T-t,x,y\right)\Phi(y)dy.$$

To check that this really is the solution, differentiate and use the fact that $p(t,x,y)$ given by (59) is the fundamental solution to the equation

$$\frac{\partial u}{\partial t}=\frac{1}{2}\frac{\partial^2 u}{\partial x^2},$$

to obtain (58). □

7.4 Semilinear equations and branching processes

All the examples we have seen so far have been stochastic representations of solutions to linear partial differential equations. The theory does not end there. In this short section, we see a beautiful representation of solutions to certain semilinear equations in terms of an object called *branching Brownian motion.*

Branching Brownian motion is most easily described as a simple model of an evolving population. As the name suggests, during their lifetimes, individuals are assumed to follow Brownian motion in space. They reproduce according to a continuous time Galton Watson branching process. More formally, we have the following definition.

Definition 7.30. Branching Brownian motion *has three ingredients:*

(i) **The spatial motion:** *During her lifetime, each individual in the population moves around in $\mathbb{R}^d$ (independently of all other individuals) according to a Brownian motion.*

(ii) **The branching rate, V:** *Each individual has an exponentially distributed lifetime with parameter V. That is, given that she is alive at time t, the probability that she dies in the time interval $[t, t+\delta t)$ is $V\delta t + o(\delta t)$.*

(iii) **The branching mechanism, Φ:** *When she dies, an individual leaves behind (at the location where she died) a random number of offspring with probability generating function $\Phi(s) = \sum_{k=0}^{\infty} p_k s^k$. Conditional on their time and place of birth, offspring evolve independently of each other (in the same way as their parent).*

Remark 7.31. *There is nothing special about Brownian motion on $\mathbb{R}^d$ here. It could be replaced by a Markov process on any Polish space without extra work. Of course, the lifetimes of individuals must be exponentially distributed if we want the branching Brownian motion to be a Markov process.*

It is often convenient to think of the branching Brownian motion as a measure-valued process. It will take values among purely atomic measures on $\mathbb{R}^d$. We represent an individual at the point $x \in \mathbb{R}^d$ by δ_x, a unit point mass at x. In this section, we write Y_t^i for the position of the ith member of the population at time t. The symbol ξ_t will denote the measure representing the whole population at time t. That is

$$\xi_t = \sum \delta_{Y_t^i},$$

where the sum is over all individuals alive at time t.

If the initial population is the (purely atomic) measure ν, then we write $\mathbb{P}_\nu(\cdot)$ for the distribution of the process and $P_t(\cdot, \nu)$ for the corresponding transition probability.

Two immediate consequences of the definition will be of particular importance to us in what follows:

(i) **The Markov property.** Since the exponential lifetime of each individual and the Brownian motion that determines her spatial movement are both Markov processes, the branching Brownian motion inherits this property.

(ii) **The branching property.** Since the evolution of the descendants of different individuals in the population are independent,

$$P_t(\cdot, \nu_1 + \nu_2) = P_t(\cdot, \nu_1) * P_t(\cdot, \nu_2),$$

where $*$ denotes convolution. In words, the distribution of the process with initial value $\nu_1 + \nu_2$ is equal to that of the sum of two independent copies of the process with initial values ν_1, ν_2 respectively.

We exploit these properties in characterising the distribution of branching Brownian motion in terms of a semilinear heat equation. This result is usually attributed to McKean (1975), but is also implicit in the work of Skorokhod (1964).

Before stating the result, we need some notation. We use $\langle \phi, \mu \rangle$ to denote the integral $\int \phi d\mu$.

Theorem 7.32 (Characterisation via a pde)**.** *The distribution of branching Brownian motion can be characterised as follows: For $\psi \in C_b^+(R^d) \cap \mathcal{D}(\Delta)$ with $0 \le \psi(x) \le 1$ for $x \in \mathbb{R}^d$,*

$$\mathbb{E}_{\delta_x}\left[\prod \psi(Y_t^i)\right] = v(t,x), \tag{60}$$

where the product is over all individuals alive at time t, and v solves

$$\begin{cases} \frac{\partial v}{\partial t} = \frac{1}{2}\Delta v + V\left(\Phi(v) - v\right), \\ \\ \qquad v(0,x) = \psi(x). \end{cases} \tag{61}$$

Equivalently, if the initial state of the population is represented by the (purely atomic) measure ν, and its state at time t is the measure ξ_t,

$$\mathbb{E}_\nu\left[\exp\left(\langle \log\psi(x), \xi_t(dx)\rangle\right)\right] = \exp\left(\langle \log v(t,x), \nu(dx)\rangle\right). \tag{62}$$

Remark 7.33. *The function of the assumption that $0 \leq \psi \leq 1$ is to ensure that the function v, defined by equation (60), is finite. The representation is actually somewhat more general than this.*

Proof.

If we define v to be the expression on the left hand side of equation (60), then it is evident that $v(0,x) = \psi(x)$. We must evaluate the time derivative of v, to check that it solves the differential equation.

We assume that at time t, v is twice continuously differentiable with respect to the space variable, x. If we can show that, under this assumption, the function satisfies the differential equation, then standard regularity theory for the heat semigroup tells us that, in fact, the function v is *smooth* as a function of x at all later times. (More pedantic readers might prefer to work with the integrated form of the equation.)

Suppose then that we wish to evaluate $v(t+\delta t, x)$. The idea is to condition on the behaviour of the original ancestor in the first δt of time. One of two things can have happened to her: either she died before time δt, leaving behind a random number of offspring, or she is still alive at time δt.

We need some notation. For each non-negative integer k, we write A_k for the event that the original ancestor dies in the first δt of time leaving behind exactly k offspring, and we denote by A^c the complementary event (that she survived until time δt). Then evidently

$$\mathbb{E}_{\delta_x}\left[\prod \psi(Y^i_{t+\delta t})\right] = \sum\nolimits_{k=0}^{\infty} \mathbb{E}_{\delta_x}\left[\prod \psi(Y^i_{t+\delta t})|A_k\right]\mathbb{P}[A_k] + \mathbb{E}_{\delta_x}\left[\prod \psi(Y^i_{t+\delta t})|A^c\right]\mathbb{P}[A^c].$$

We now estimate each term on the right hand side of this equation. Since we are interested in calculating the first (time) derivative, we only care about terms up to order δt.

Consider first the event A_k. In the notation of Definition 7.30, the probability of A_k is $p_k V\delta t + o(\delta t)$. Conditional on A_k, our original ancestor left behind exactly k offspring at the location where she died. The probability that any of these descendants have, themselves, died by time δt is of order δt and so we may ignore that possibility. Moreover, the branching property tells us that conditional on their location at time δt, the evolution of the descendants of each of these k first generation individuals over the time interval $[\delta t, t+\delta t]$ is independent of that of the descendants of her siblings.

Consider then the family tree of descendants of a single individual who is located at the point y at time δt. By the Markov property of branching Brownian motion, the distribution of this subpopulation at time $t+\delta t$ is exactly the same as if we had started from a single individual at y at time *zero*, and allowed the population to evolve until time t.

Evidently at time δt, the k first generation individuals have travelled no further than if they had followed Brownian motions from time zero, and so combining the observations above with continuity of the function $v(t,\cdot)$, we have

$$\mathbb{E}_{\delta_x}\left[\prod \psi(Y^i_{t+\delta t})|A_k\right] = v(t,x)^k + o(1).$$

Under A^c, we condition further on the position, $B_{\delta t}$, of the ancestor at time δt to obtain (in the notation of equation (60)),

$$v(t+\delta t, x) = V\delta t \sum_{k=0}^{\infty} p_k v(t,x)^k + E_x\left[v\left(t, B_{\delta t}\right)\right](1 - V\delta t) + o(\delta t),$$

where E_x denotes expectation for a *Brownian motion* started at the point x at time zero. Finally, subtracting $v(t,x)$ from both sides, dividing by δt, and letting $\delta t \downarrow 0$, we see that, since the infinitesimal generator of Brownian motion is $\frac{1}{2}\Delta$, v defined by (60) does, indeed, satisfy (61). □

McKean was interested in the special case in which $\Phi(v) = v^2$ (so that when an individual dies, she leaves behind exactly two offspring). Equation (61) is then the celebrated Fisher KPP equation. (KPP are Kolmogorov, Petrovskii and Piscunov.) This equation arises in population genetics, and McKean used the representation of solutions in terms of branching Brownian motion to study travelling wave solutions. A considerable industry has grown out of extending McKean's work to other semilinear partial differential equations.

A Definition and examples of local martingales

This appendix is copied almost verbatim from lecture notes of Peter Mörters. It is intended to provide an easy reference for the relationship between local martingales and martingales. Local martingales are an extension of the idea of martingales. The essential idea is that we require that certain properties of a process (like the martingale property) need only hold locally.

Definition A.1 (Stopped process, local martingale, reducing sequence)**.** *Let $\{X(t) : t \geq 0\}$ be an adapted process and T a stopping time with respect to the filtration $\{\mathcal{F}(t) : t \geq 0\}$. Define the* stopped process *$\{X^T(t) : t \geq 0\}$ by $X^T(t) = X(T \wedge t)$. The process $\{X(t) : t \geq 0\}$ is called a* local martingale *with respect to $\{\mathcal{F}(t) : t \geq 0\}$ if there exists a sequence*

$$0 = T_0 \leq T_1 \leq \ldots \leq T_n \uparrow \infty$$

of stopping times such that $\{X^{T_n}(t) : t \geq 0\}$ is a martingale. We say that $\{T_n\}$ is a reducing sequence of stopping times.

Remark A.2. *Every martingale is a local martingale, because every sequence of stopping times, which increases to infinity is reducing. In the definition we can equivalently require that $\{X^{T_n}(t) : t \geq 0\}$ is a martingale with respect to $\{\mathcal{F}(t \wedge T_n) : t \geq 0\}$ instead of $\{\mathcal{F}(t) : t \geq 0\}$. Show this as an exercise.*

Example A.3. *We construct a local martingale $\{X(t) : t \geq 0\}$, which is not a martingale. Although we will not prove all the details, I hope you get the flavour.*

We let a particle perform a symmetric random walk, but in continuous time and the waiting times between the jumps are random times. Fix a probability distribution on the integers by denoting the weight at the integer n by $p_n > 0$. We will assume that

$$\sum_{n=-\infty}^{\infty} n^2 p_n < \infty\,,$$

which means that the p_n are decreasing rather rapidly at both ends of the sequence. Now the process starts in 0 and stays there for an exponentially distributed time T_0 with expectation p_0, i.e. $X(t) = 0$ for all $t \in [0, T_0)$. Then we flip a coin and move to $Y_1 = \pm 1$ with equal probability, and wait again for

an exponential time T_1, which is independent of T_0 and has expectation p_{Y_1}. In other words $X(t) = Y_1$ for $t \in [T_0, T_0 + T_1)$.

Suppose now we have just jumped to a level n and it was our kth jump. Then we stay there for an (independent) exponentially distributed time T_k with expectation p_n, before making the next jump of height ± 1, chosen independently and with equal probability. This means, heuristically, that we spend a long time when we jump to levels near 0 and only a short time at levels away from 0.

Because the symmetric random walk returns to 0 infinitely often and the times spent there are an independent, identically distributed sequence of times with positive expectation, the total time spent in 0 is infinite (by the strong law of large numbers) and hence we have defined the process on the whole time axis. The process is indeed a local martingale, to see this formally let

$$S_k = T_0 + \cdots + T_{k-1}$$

be the time of the kth jump. Then $\{S_k\}$ is a sequence of stopping times increasing to infinity. Moreover if $Y_1, Y_2, \ldots$ is the i.i.d. sequence of jump heights, then

$$X(t \wedge S_k) = \sum_{i=1}^{k} Y_i \mathbf{1}_{\{S_i \le t\}} .$$

For $s < t$ and $F \in \mathcal{F}(s)$ (the natural filtration), the event $F \cap \{s < S_i \le t\}$ is independent of Y_i. Thus

$$\int_F \Big\{ X(t \wedge S_k) - X(s \wedge S_k) \Big\} \, d\mathbb{P} = \sum_{i=1}^{k} \mathbb{E}\Big[Y_i \mathbf{1}_{\{s < S_i \le t\} \cap F} \Big] = 0 \,,$$

hence $\mathbb{E}[X(t \wedge S_k) \,|\mathcal{F}(s)] = X(s \wedge S_k)$, proving the martingale property of X^{S_k}. We give a heuristic argument why $\{X(t)\}$ is not a martingale: Let $s < t$ and suppose we are given the information that $X(s)$ has a very large value. If the process were a martingale, then this large value would be the expected value of $X(t)$. But it is not, because we know that the process spends only a small amount of time at the high values and, in fact, most of the time is spent near zero, so that the expected value for $X(t)$ given the unusual information about $X(s)$ is below the value of $X(s)$. A rigorous argument (using Markov chain theory) can be found in (4.2.6) in v. Weizsäcker/Winkler.

The examples of local martingales might not convince you that this class contains natural examples which fail to be martingales, but the next theorem shows a remarkable advantage of working with local martingales: If we look at *continuous* local martingales, we get uniform integrability for free. For example, it is worth considering Brownian motion as a local martingale and using a reducing sequence, such that the stopped Brownian motions are uniformly integrable martingales.

Theorem A.4. *Suppose $\{X(t) : t \ge 0\}$ is a continuous local martingale. Then the sequence*

$$T_n = \inf\{t \ge 0 : |X(t)| > n\}$$

is always reducing. In particular, we can find a sequence, which reduces $\{X(t)\}$ to a bounded (or uniformly integrable) martingale.

Proof. Suppose $0 < s < t$. If $\{S_n\}$ is a reducing sequence, then we apply the optional stopping theorem to $\{X^{S_n}(t)\}$ at times $s \wedge T_m$ and $t \wedge T_m$ and obtain, using the first remark of this section,

$$\mathbb{E}[X(t \wedge T_m \wedge S_n) | \mathcal{F}(s \wedge T_m \wedge S_n)] = X(s \wedge T_m \wedge S_n).$$

Multiplying by $\mathbf{1}_{\{T_m>0,S_n>0\}} \in \mathcal{F}(s \wedge T_m \wedge S_n)$ we obtain

$$\mathbb{E}[X(t \wedge T_m \wedge S_n)\mathbf{1}_{\{S_n>0,T_m>0\}}|\mathcal{F}(s \wedge T_m \wedge S_n)] = X(s \wedge T_m \wedge S_n)\mathbf{1}_{\{S_n>0,T_m>0\}}.$$

As $n \to \infty$, $\mathcal{F}(s \wedge T_m \wedge S_n) \uparrow \mathcal{F}(s \wedge T_m)$ and

$$X(r \wedge T_m \wedge S_n)\mathbf{1}_{\{S_n>0,T_m>0\}} \to X(r \wedge T_m)\mathbf{1}_{\{T_m>0\}},$$

for all $r > 0$. Because the sequence is dominated by m we find for the conditional expectations, using Dominated Convergence,

$$\mathbb{E}\Big[X(t \wedge T_m)\mathbf{1}_{\{T_m>0\}}\Big|\mathcal{F}(s \wedge T_m)\Big] = X(s \wedge T_m)\mathbf{1}_{\{T_m>0\}} \text{ almost surely,}$$

which proves the statement. □

Remark: The proof also works for every sequence S_n of stopping times, which increases to infinity, but is smaller than T_n.

One more advantage of local martingales in comparison to martingales is that they can be defined easily on random time intervals $[0, \tau)$. Whereas the concept of martingale on $[0, \tau)$ is meaningless, because for large t the random variable $X(t)$ is not defined on the whole space Ω, the following definition of a local martingale on $[0, \tau)$ is very natural:

Definition A.5 (local martingale)**.** *Suppose τ is a random time. The process $\{X(t) : t \in [0, \tau)\}$ is a* local martingale *if there exists a sequence $T_n \uparrow \tau$ of stopping times such that $\{X^{T_n}(t) : t \geq 0\}$ is a martingale.*

Let us now explore the relationship between local martingales and martingales. We show that we can change the time of a local martingale to obtain a martingale. Recall our example, which is constructed by taking a martingale (symmetric random walk) and distorting the time scale, so that the martingale property is violated, but the local martingale property still holds.

Theorem A.6. *Suppose $\{X(t) : t \in [0, \tau)\}$ is a continuous local martingale on an arbitrary, possibly random, time interval. Then there exists a time change $\gamma : [0, \infty) \to [0, \tau)$ such that each $\gamma(t)$ is a stopping time and the process $\{X(\gamma(t)) : t \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}(\gamma(t))\}$.*

Proof. If T_n are the reducing stopping times as in Theorem A.4, then define $\gamma : [0, \infty) \to [0, \tau)$ by

$$\gamma(t) = \begin{cases} t - (k-1) & \text{if } T_{k-1} + (k-1) \leq t \leq T_k + (k-1) \\ T_k & \text{if } T_k + (k-1) \leq t \leq T_k + k. \end{cases}$$

Now optional stopping comes into play. Let $n = [t] + 1$. As $\gamma(t) \leq T_n \wedge t$ we have $X(\gamma(t)) = \mathbb{E}[X(T_n \wedge n)|\mathcal{F}(\gamma(t))]$ and hence

$$\mathbb{E}[X(\gamma(t))|\mathcal{F}(\gamma(s))] = \mathbb{E}\Big[X(T_n \wedge n)|\mathcal{F}(\gamma(s))\Big] = X(\gamma(s)) \text{ almost surely,}$$

as desired. □

One hopes that integrability conditions ensure that local martingales are martingales. The following is a positive result in this direction.

Theorem A.7. *Suppose $\{X(t) : t \geq 0\}$ is a local martingale and, for every $t > 0$,*

$$\mathbb{E}\Big[\sup_{0 \leq s \leq t} |X(s)|\Big] < \infty\,,$$

then $\{X(t) : t \geq 0\}$ is a martingale.

Proof. Clearly, $\mathbb{E}|X(t)| < \infty$. Now, if $\{T_n\}$ is a reducing sequence,

$$\mathbb{E}\Big[X^{T_n}(t)\,\Big|\mathcal{F}(s \wedge T_n)\Big] = X^{T_n}(s) \text{ almost surely.}$$

By our assumption we can let $n \to \infty$ and use Dominated Convergence, observing that our condition makes sure that $X^{T_n}(t)$ is dominated by an integrable function. The limiting equation is the martingale property of $\{X(t)\}$. □

From this we easily deduce the following important corollary.

Corollary A.8. *A bounded local martingale is a martingale.*