MARTIN WILLIAM BAXTER
PEMBROKE COLLEGE

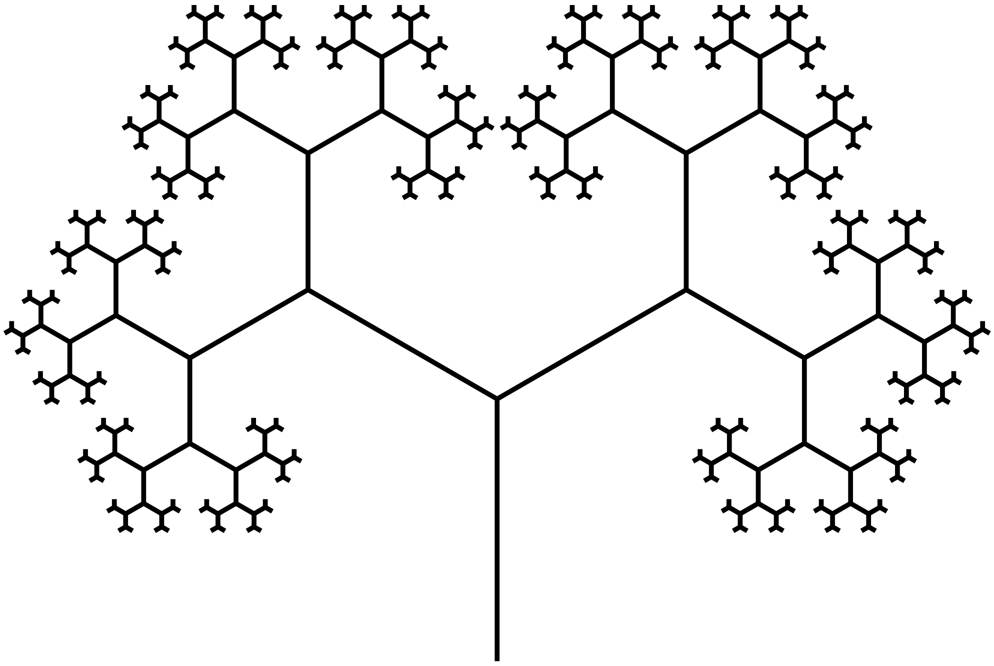
DISCOUNTED FUNCTIONALS OF MARKOV PROCESSES

A THESIS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE

DISCOUNTED FUNCTIONALS
OF MARKOV PROCESSES

MARTIN BAXTER M.A. was born in Edinburgh in 1968 and was educated there at Daniel Stewart's and Melville College and at Pembroke College Cambridge, where he has been elected to a Research Fellowship.



STATISTICAL LABORATORY

University of Cambridge

First Published 1993

Printed in the Statistical Laboratory, Cambridge
Set in Adobe Palatino by TEX

The illustration above is a possible visualization of the Ray-Knight compactification of a Markov chain on the Binary Tree from Chapter Five

Contents

Detailed Contents

Preface

This dissertation is submitted towards the degree of Doctor of Philosophy at the University of Cambridge. My work on it has been supervised by Prof David Williams, sometime of Clare College and now at the University of Bath, and by Dr James Norris of Churchill College.

My course of study has been supported by the United Kingdom Science and Engineering Research Council, and I am grateful to Pembroke College Cambridge and the Statistical Laboratory Cambridge for travel grants given during it.

I owe David Williams in particular a very large debt for his great enthusiasm for the subject, generosity of spirit, and continual and effective support. I must thank also James Norris, Chris Rogers, and Martin Barlow who have shown me many times in the past three years how probability should be done, as well as Ray Lickorish who taught me how to analyse mathematically. Less formal, but no less significant, were the innumerable tea-room round tables with a host of gifted research students and staff, especially Ben, Phil, Dave and Roland.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where explicitly approved by the Board of Graduate Studies and as stated in the Introduction.

Chapter Zero

Introduction

A tale told by an idiot,
Full of sound and fury,
Signifying nothing.

***Shakespeare**, Macbeth*

0.1 Introduction

Probability theory, from its very beginnings, has carried a strong whiff of determinism. Even Bernoulli [8] in 1713 thought that just as backward peoples still gambled on eclipses that science could now predict, some day gambling on coins and dice would seem equally primitive once mechanics was perfected (quoted in Gigerenzer et al. [20]). Nowadays randomness is accepted in its own right as a natural phenomenon, arising in a wide class of cases from the emergent uncertainty of large-scale deterministic systems to the quantum uncertainty of the very small. However, the feeling that randomness is untidy and that (almost) certain results are better ones lingers on. Indeed, sometimes, the theory of random events resembles Wittgenstein's philosophical ladder — we climb up it only to throw it away, to be left with a better view of the deterministic world.

One of the earliest, and still popular, methods of removing uncertainty from some random quantities was to take an average of them, in the hope that the stochastic fluctuations would cancel out. The simplest average is of course the sample mean, which weights each of the samples equally. We shall call averages of this type Cesàro averages. These averages are the subject of the Laws of Large Numbers, the Central Limit Theorem, and Large-Deviation results which give us certainty (almost sure limits), distributional information (asymptotic normality), and extreme value likelihoods. The importance of these truths in Statistics and elsewhere can be measured by the amount of effort expended on obtaining, generalizing and developing these results.

If the samples all have the same status, then the Cesàro average is the natural one, but if the symmetry between them is broken, then it is less so. We might think of our samples as being a time-series of values from some stochastic process. In, say, an economic context a value now is worth more than the same (nominal) value at a later time. Similarly in control theory, current events are privileged over discounted future events. Mathematically, we would take an average where the weights are not equal but depend on the time position of the sample. For exponential or geometric discounting it is natural to call such a sum the Abel average.

In this work, we find analogues of existing theorems for both Abel and more general averages, as well as completely new results about averages. Although much of the work has the advantageous property that it holds true for a general average and process whenever proved true for the Cesàro average, other parts are derived from different considerations using a fresh approach. In the latter case it has proved of considerable technical ease to work in the context of finite state Markov chains. It is clear what more general corresponding results would be, but we shall not attempt to prove them here.

In the last two Chapters, the focus shifts from time discounted integrals of the local time to more basic questions concerning the existence and continuity of the local times themselves for some particular Markov processes. Although the general theory has been much studied and understood, we will find straightforward methods where possible, as well as working within the existing framework. Chapter Five is concerned with the construction and analysis of Markov chains on the infinite binary tree, with reflection off the end of each infinite branch. We discover exact conditions for the regularity of boundary points, as well as for the existence of a jointly continuous version of the local time. This allows the construction of a Markov process on the Cantor set, which is quite different from the recently much-studied Brownian motion on a fractal type of process.

Finally, the ultimate Chapter creates a continuous space analogue of the tree process. In two-dimensions, a typical diffusion, such as Brownian motion in the plane, never revisits a point, so it does not have a local time. In Chapter Six we shall construct the local times of some particular two-dimensional diffusions on a special one-dimensional subspace, and show that they are jointly continuous in both time and space under some exact conditions. Conditions for recurrence of the process are also derived.

0.2 Outline

Our basic object of study will be the average at discount rate λ of the occupation measure of a stochastic process X, A_λ, where

$$A_\lambda := \lambda \int_0^\infty m(\lambda t)\delta_{X(t)}\, dt.$$

Here A_λ is an element of the space of probability measures on the state space of X, and m, the discount function, is a density on $\mathbb{R}^+$. The Cesàro case is when m is the indicator function of $[0, 1]$, and the Abel case is when $m(t)$ is e^{-t}. For X a diffusion, we generally abuse notation and write A_λ for $A_\lambda(\mathbb{R}^+)$, but for X a finite state Markov chain, we continue to treat A_λ as a measure (a point on a simplex).

Our principal questions concern the search for expressions for the following:

- Exact distribution of A_λ
- Symmetry characterizations of the distribution of A_λ
- Asymptotics of the distribution of A_λ (for fixed λ)
- PDE's governing the density of A_λ
- Strong Law and Central Limit Theorem for A_λ (as $\lambda \downarrow 0$)
- Large Deviation Principle for A_λ (as $\lambda \downarrow 0$)
- Finer density expansions for A_λ (as $\lambda \downarrow 0$)

Where we seek answers for various classes of discount (Cesàro, Abel, general) and for various processes (Brownian motion, Ornstein-Uhlenbeck, Markov chain, general).

Some of these questions have been studied before, especially in the Cesàro cases. For instance, we know that the Cesàro average of a Brownian motion has the arc-sine distribution (Lévy), and that of an Ornstein-Uhlenbeck process has the uniform distribution. Both these distributions are members of the beta family of distributions, and we shall see further such examples in the Abel cases (if only because of the paucity of analytic distributions on $[0, 1]$ for exact solutions to belong to).

The symmetric characterizations (for the Abel case only) not only provide speedy verification of the above results, but also give the distribution asymptotics through some careful analysis and use of Tauberian theorems. The Strong Law and the Central Limit Theorem follow immediately from existing theorems and other results derived here.

Of the great amount of recent work in the theory of large-deviations ever since Donsker and Varadhan [16], many notable papers, such as de Acosta [1] have concentrated purely on the discrete-time Cesàro average. Others, including Nummelin [29] and Jain [23], have looked at functions of the form $t^{-1} \int_0^t f(X_s)\, ds$, which can be made to fit our framework by taking X to be the Markovian space-time process associated with the original process. However, this approach leads to awkwardness in the conditions to be satisfied and the expressions derived, and does not lead to the exact forms which we will derive ourselves (albeit in a much more restricted case). Nevertheless, as remarked above, if the large-deviation property holds for the Cesàro case then we can extend it using our theorems to the other cases, so to our extend our own results to, say, infinite state Markov chains, the paper of Chen and Lu [14] and others will surely be useful.

0.3 Summary of Results

Let us give a flavour of the results which will be obtained.

Theorem (Large-Deviations). *Suppose that X is an irreducible Markov chain on a finite state-space S, with Q-matrix Q, and C_t is the Cesàro average defined as*

$$C_t := \frac{1}{t} \int_0^t \delta_{X_s}\, ds,$$

taking values in $M_1(S)$, the space of probability measures on S. Then the large-deviation property holds for C_t with rate function I defined on $M_1(S)$:

$$\limsup_{t\to\infty} t^{-1} \log \mathbb{P}(C_t \in F) \leqslant -\inf_{x\in F} I(x),$$
$$\liminf_{t\to\infty} t^{-1} \log \mathbb{P}(C_t \in G) \geqslant -\inf_{x\in G} I(x),$$

for F and G respectively closed and open subsets of $M_1(S)$. The function I is given by the expression

$$I(x) = \sup_{h\in(\mathbb{R}^+)^S} -\sum_{i,j\in S} x_i \frac{q_{ij} h_j}{h_i}.$$

Further, let m be any density on $[0,\infty)$ and A_λ be defined as the average

$$A_\lambda := \lambda \int_0^\infty m(\lambda t)\delta_{X_t}\,dt,$$

also taking values in $M_1(S)$. Then the large-deviation property holds for A_λ as λ goes to 0 with rate function K given by

$$K(x) = \sup_{v\in\mathbb{R}^S}\left(\langle v,x\rangle - \int_0^\infty \sup_{y\in M_1(S)}\big(\langle m(t)v,y\rangle - I(y)\big)\,dt\right).$$

Or more naturally,

$$K^*(v) = \int_0^\infty I^*(m(t)v)\,dt,$$

where $*$ denotes convex conjugation (Legendre transformation).

Theorem (Central Limit). *Under the same conditions, if m is of bounded variation and π is the invariant distribution of X, then*

$$\frac{A_\lambda - \pi}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0, H_K^{-1}(\pi)), \qquad \text{as } \lambda \downarrow 0,$$

where H_K is the Hessian of K on $M_1(S)$.

Theorem (Density asymptotics). *Under the same conditions, and if f_i^λ is the density of A_λ on $M_1(S)$, starting X in state i, then*

$$f_i^\lambda(x) \overset{\mathrm{w}}{\sim} z_i(x)e^{-K(x)/\lambda}(2\pi\lambda)^{-(n-1)/2}\big(\det H_K(x)\big)^{\frac{1}{2}}, \qquad \text{as } \lambda \downarrow 0,$$

where $z(x)$ is the (unique) positive eigenvector of $Q + \mathrm{diag}(m_0\nabla K(x))$, and where '$\overset{\mathrm{w}}{\sim}$' means that the ratio of the two sides tends weakly to 1 (in the sense of Corollary 4.17).

Theorem (Distribution asymptotics). *Let X be a stochastic process on $\mathbb{R}$ and let F_λ be the distribution of the Abel average A_λ*

$$A_\lambda := \lambda \int_0^\infty e^{-\lambda t} I_{\mathbb{R}^+}(X_t)\,dt.$$

(1) If X is Brownian motion, then F_λ is independent of λ and

$$F_\lambda(x) \sim \frac{2\sqrt{x}}{\pi\sqrt{\log(1/x)}} \qquad \text{as } x \downarrow 0.$$

(2) If X is the recurrent Ornstein-Uhlenbeck process generated by the SDE $dX = dB - \frac{1}{2}\gamma X\, dt$, and $X_0 = 0$, then

$$F_\lambda(x) \approx x^{\frac{1}{2}(1+\gamma/\lambda)}, \qquad \text{as } x \downarrow 0,$$

where $f(x) \approx g(x)$ means that f/g tends to a positive limit as x goes to 0.

(3) If X is a symmetrisable Markov chain then, replacing $I_{\mathbb{R}^+}$ by $I_{\{i\}}$ in the definition of A_λ, we have that

$$F_\lambda(x) \approx x^{\gamma/\lambda}, \qquad \text{as } x \downarrow 0,$$

where γ is the minimal positive eigenvalue of the submatrix of $(-Q)$ formed by deleting the ith row and ith column.

There are nonetheless many questions unanswered. It is clear that there should be analogues of the finite state Markov chain results not only for countable state processes but also for continuous space ones. In the economics context especially one would expect the discount shape itself to be random and possibly not independent of the process itself.

The two local time theorems about the discrete and continuous space cases respectively can be summarised as follows. Fuller details of their construction and behaviour can be found in Chapters Five and Six.

Theorem (Binary Tree process). *There is a Markov chain X on the binary tree which makes up-jumps from level n to level $n-1$ at rate μ_n, and left and right down-jumps from level n to level $n+1$ at rate $\frac{1}{2}\lambda_n$, and which reflects off the boundary on explosion. Then*

(1) X is positive recurrent $\iff \sum_n \pi_n < \infty$, and then

(2) X reaches the boundary in finite time $\iff \sum_n \frac{1}{\lambda_n \pi_n} < \infty$, and then

(3) any (and hence each) boundary point is regular $\iff \sum_n b_n < \infty$, and then

(4) there exists a jointly-continuous local time on $F \iff \sum_n \sqrt{\frac{1}{n} c_n} < \infty$, and then

(5) X has visited all the states of F by a finite time,

where $\pi_n := (\lambda_0 \ldots \lambda_{n-1})/(\mu_1 \ldots \mu_n)$, $b_n := \frac{2^n}{\lambda_n \pi_n}$, and $c_n := \sum_{r=n}^{\infty} b_r$.

Theorem (Continuous space process). *There is a process Z in the upper half-plane, whose vertical component Y is a Brownian motion reflecting off 0, and whose horizontal component X is governed by the stochastic differential equation $dX = \sigma(Y)\,dW$, where W is a Brownian motion independent of Y, and σ is a positive function which is locally square-integrable on $[0, \infty)$. Let $\mathbb{R}_0$ be the x-axis, the boundary of the domain of Z. If σ is slowly varying enough at 0 and infinity (see Corollary 6.9 and Theorem 6.11) then*

(1) any (and hence each) point of $\mathbb{R}_0$ is regular $\iff \displaystyle\int_0^1 \frac{dt}{t\sigma(t)} < \infty$*, and then*

(2) Z has a jointly-continuous local time on $\mathbb{R}_0$ $\iff \displaystyle\int_0^{1/2} \frac{\varphi(z)\,dz}{z\sqrt{\log(1/z)}} < \infty$*, and then*

(3) Z will visit all the points of any compact subset of $\mathbb{R}_0$ by a finite time $\iff \displaystyle\int_1^{\infty} \frac{dt}{t\sigma(t)} = \infty$,

where $\varphi^2(y) := \int_0^y \big(t\sigma(t)\big)^{-1}\,dt$.

We shall also discover how conditions (3) and (4) of the Binary Tree theorem can be seen as merely discrete versions of conditions (1) and (2) of the Continuous Space theorem.

Provenance and prior Publication

Chapters One and Two are joint work with David Williams, and articles based upon them have been published in the *Mathematical Proceedings of the Cambridge Philosophical Society* (Baxter and Williams [3] and Baxter and Williams [4]). Chapter One is mostly due to Williams, with the exceptions that part of the proof of Theorem 1.2, and the first half of the proof of Theorem 1.1 are due to Baxter. In Chapter Two, Proposition 2.4 is due to Williams, Lemma 2.2 is due to Baxter, and Theorems 2.1, 2.3, 2.5, and 2.6 are joint work.

An earlier version of Chapter Three has been published as Baxter [5].

Chapter Five has appeared in *Séminaire de Probabilitiés XXVI* (Baxter [6]).

The first three Sections of Chapter Six are being published in the Mathematical and Physical Sciences *Proceedings of the Royal Society* (Baxter [7]).

Chapter One
Symmetry Characterizations

To probe a hole we first use a straight stick to see how far it takes us. To probe the visible world we use the assumption that things are simple until they prove to be otherwise.

***E. H. Gombrich**, Art and Illusion*

1.1 Introduction and Summary

We begin by studying just one distribution and investigating ways of characterizing it using symmetry properties. The distribution, which began our initial study, is the exponential (or Abel) discount of the time spent by a Brownian motion in the half-line, or in other words a discounted version of the arc-sine law. Theorem 1.1 gives two symmetry properties which characterize this distribution, and Theorem 1.2 derives from them some asymptotic information about it. Section 1.5 begins a generalization process, which continues in Section 2.4 of Chapter Two, by deriving similar symmetry characterizations for symmetric diffusions, which involve the law of the diffusion's excursions from 0. In Sections 1.6 and 1.7, numerical analysis is used in an attempt, so far abortive, to identify the distribution, although values of its density at various points are calculated.

Let H be the function on $\mathbb{R}$ defined by

$$H(x) := \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geqslant 0. \end{cases}$$

Let $B = \{B_t : t \geqslant 0\}$ be Brownian motion on $\mathbb{R}$ starting at 0.

The layman's intuition (and ours) is that as $t \to \infty$, it should be the case that

$$H(B_t) \to \tfrac{1}{2} \text{ in some average sense.}$$

(For confirmation of the layman's intuition, see Bingham and Rogers [11].) Brownian motion has the scaling property that for $c \neq 0$, $\{c^{-1}B(c^2 t) : t \geqslant 0\}$ is also a Brownian motion starting at 0. It therefore follows that the distribution of the Cesàro average

$$C_t := t^{-1} \int_0^t H(B_u)du \quad (t > 0)$$

is independent of t. Lévy's arc-sine law states that for $0 \leqslant x \leqslant 1$,

$$\mathbb{P}(C_t \leqslant x) = \frac{2}{\pi} \arcsin(\sqrt{x}).$$

Abel averaging is generally 'stronger' than Cesàro averaging, (though not for bounded functions such as H). We would say that $H(B_t) \to \frac{1}{2}$ in Abel's sense if $A_\lambda \to \frac{1}{2}$ as $\lambda \downarrow 0$, where

$$A_\lambda := \int_0^\infty \lambda e^{-\lambda u} H(B_u)du \quad (\lambda > 0).$$

However, Brownian scaling shows that the distribution function F of A_λ is independent of λ.

The problem of calculating

$$F(x) := \mathbb{P}(A \leqslant x) \quad (0 \leqslant x \leqslant 1),$$

where

$$A := \int_0^\infty e^{-u} H(B_u) du, \tag{1.1}$$

proves to have several very strange aspects. It is important to realize that A is *not* the value C_ζ of C at an exponentially distributed time independent of B. The problem of calculating the distribution of A is a Feynman-Kač problem (see Section 1.3) about the space-time process $\{(t, B_t)\}$, and is in principle 'one dimension up' from the arc-sine law.

Since A is a random variable with values in [0,1], its distribution is determined by its moments

$$\mu_n := \mathbb{E}(A^n) = \int_0^1 x^n dF(x) \quad (n = 0, 1, 2, \ldots).$$

Clearly, the μ_n can be evaluated by calculating some complicated integrals involving the Brownian transition density function. In Section 1.2, we explain a 'double recursion' based on two 'symmetry' properties which allows us to find the first few μ_n easily.

The values of the first few moments of A are as follows:

$$\mu_0 = 1, \quad \mu_1 = \frac{1}{2}, \quad \mu_2 = \frac{1}{2\sqrt{2}}, \quad \mu_3 = \frac{3-\sqrt{2}}{4\sqrt{2}},$$
$$\mu_4 = \frac{3\sqrt{3}-\sqrt{2}-\sqrt{6}}{4\sqrt{2}},$$
$$\mu_5 = \frac{15\sqrt{3}-\sqrt{2}-5\sqrt{6}-10}{8\sqrt{2}},$$
$$\mu_6 = \frac{12\sqrt{2}-30\sqrt{3}-30\sqrt{5}+10\sqrt{6}-3\sqrt{10}+45\sqrt{15}-15\sqrt{30}}{16\sqrt{3}}.$$

It may well be the case therefore that there is no closed-form expression for F. The list of values of μ_n for $n \leqslant 6$ helps hint at the fact that even for moderate n (such as $n = 25$), rounding errors (even in 'double precision' programs) can cause serious problems in calculating μ_n numerically. See Section 1.6.

In case you are interested in trying to guess the probability density function f_A of A (and for an important reason mentioned at the end of this Section), we give in Table 1.1 values which we believe to be correct rounded to the accuracy shown. See Section 1.7. Values of the arc-sine density f_C are given for comparison.

Table 1.1: some values of f_A and f_C:

x	1/32	1/16	1/8	1/4	1/2
$f_A(x)$	1.44124	1.18554	0.999632	0.87253948	0.814977
$f_C(x)$	1.82944	1.31500	0.962479	0.73510519	0.636620

The 'number-theoretic' exact expressions for the μ_n give no idea of the asymptotic behaviour of the (μ_n) sequence, and hence do not help to determine the behaviour of $F(x)$ as $x \downarrow 0$. Nor are the values in Table 1.1 of much help.

Of course, one can use the arc-sine law as a comparison to obtain estimates on F such as

$$\limsup_{x\downarrow 0} \frac{F(x)}{\sqrt{x}} \leqslant \frac{2\sqrt{e}}{\pi}, \tag{1.2}$$

$$\liminf_{x\downarrow 0} \frac{F(x)\sqrt{\log 1/x}}{\sqrt{x}} \geqslant \frac{2}{\pi}. \tag{1.3}$$

Proof of (1.2) For every t,

$$A \geq e^{-t}\int_0^t H(B_s)ds = te^{-t}C_t,$$

so that

$$\mathbb{P}(A \leq x) \leq \mathbb{P}(C_t \leq xe^t/t) = (2/\pi)\arcsin\sqrt{xe^t/t}.$$

Since the minimum value of e^t/t is e, result (1.2) follows. □

Proof of (1.3) For every t, we have

$$A \leqslant \int_0^t H(B_u)du + e^{-t},$$

whence

$$\mathbb{P}(A \leqslant x) \geqslant \mathbb{P}[C_t \leqslant t^{-1}(x - e^{-t})] = \frac{2}{\pi} \arcsin \sqrt{(x - e^{-t})/t}\,.$$

Choose t to maximize $(x - e^{-t})/t$, so that

$$(1 + t)e^{-t} = x, \quad (x - e^{-t})/t = e^{-t} = x/(1 + t).$$

As $x \downarrow 0$, $t \sim \log 1/x$, and result (1.3) follows. □

The argument we have used to obtain (1.3) seems rather crude. However, it yields the best possible result, and although we cannot see but quite why this is so, we will meet this situation again for a class of distributions in Section 4.2.

Theorem 1.1 *The distribution function F of A satisfies the asymptotic relation:*

$$F(x) \sim \frac{2\sqrt{x}}{\pi\sqrt{\log 1/x}} \quad \textit{as } x \downarrow 0. \tag{1.4}$$

Equivalently (because of Tauberian theorems and (1.6a) — see later),

$$\mu_n \sim (\pi n \log n)^{-\frac{1}{2}} \quad (n \to \infty). \tag{1.5}$$

(As usual, the $\sim$ symbol signifies that the ratio of the two sides tends to 1.)

Proof of Theorem 1.1 In Section 1.4 □

Theorem 1.2 *The distribution of A is characterized by the following two 'symmetry' properties:*

$$A \textit{ and } 1 - A \textit{ have the same distribution,} \tag{1.6a}$$

$$\mathbb{E}h(-\alpha A) = -e^{-\alpha}\mathbb{E}h(\alpha A) \quad (\alpha \in \mathbb{R}), \tag{1.6b}$$

where, for $z \in \mathbb{R}$,

$$h(z) := \sum_{n=1}^{\infty} \frac{z^n\sqrt{n}}{n!} = \frac{1}{\sqrt{\pi}} \int_0^1 \frac{ze^{tz}\,dt}{\sqrt{\log 1/t}}\,. \tag{1.7}$$

Proof of Theorem 1.2 See Section 1.2. □

Hardy [21] showed that

$$h(z) \sim z^{\frac{1}{2}} e^z \qquad (z \to +\infty), \tag{1.8}$$

$$-h(-z) \sim (\pi \log z)^{-\frac{1}{2}} \qquad (z \to +\infty). \tag{1.9}$$

(Of course, because of (1.6a), we nowadays see both (1.8) and (1.9) as consequences of the 'Abelian' half of Karamata's 'Tauberian' Theorem — see Theorem 1.7.1' of Bingham et al. [10].)

Theorem 1.1 is proved in Section 1.4 by considering (in the light of Hardy's results) the behaviour of (1.6b) as $\alpha \to \infty$. One has to be rather careful in the use of uniform-integrability properties and the like to prove that

$$\mathbb{E}e^{-\alpha A} \sim (\pi \alpha \log \alpha)^{-\frac{1}{2}} \qquad (\alpha \to +\infty), \tag{1.10}$$

which, by Karamata's Tauberian Theorem proper, implies the desired result (1.4).

As was remarked before the statement of Theorem 1.1, it would be satisfying to understand in *probabilistic* terms why that theorem is true. In a sense here, we have too many trees and not enough wood. By the time we reach Chapter Four, the view may be clearer.

Table 1.1 helps emphasize that the terms in the asymptotic-series expansion for F near 0 (of which Theorem 1.1 provides the first) decrease only very slowly in magnitude.

1.2 Proof of Theorem 1.2

There is a very quick method for calculating the moments μ_n $(n \geqslant 0)$.

We allow our Brownian motion B to start at an arbitrary point x of $\mathbb{R}$, and write $\mathbb{P}^x$ and $\mathbb{E}^x$ for the corresponding probability and expectation. Thus $\mathbb{P} = \mathbb{P}^0$, $\mathbb{E} = \mathbb{E}^0$. For $\alpha \in \mathbb{R}$, define

$$\Phi(\alpha, x) := \mathbb{E}^x e^{\alpha A}. \tag{1.11}$$

We exploit an obvious symmetry

$$\Phi(\alpha, x) = \mathbb{E}^x e^{\alpha A} = \mathbb{E}^{(-x)} e^{\alpha(1-A)} = e^{\alpha}\Phi(-\alpha, -x). \tag{1.12}$$

Next, we note that if

$$T_0 := \inf\{t : B_t = 0\},$$

then, for $x < 0$, and $n \in \mathbb{Z}^+$, the strong Markov property shows that

$$\begin{aligned} \mathbb{E}^x(A^n) &= \mathbb{E}^x \left(\left\{ e^{-T_0} \int_0^\infty e^{-u} H\left(B_{T_0+u}\right) du \right\}^n \right) \\ &= \mathbb{E}^x \left(e^{-nT_0}\right) \mu_n = \mu_n e^{x\sqrt{2n}} \,. \end{aligned} \tag{1.13}$$

From (1.11) and (1.13), we obtain

$$\Phi(\alpha, x) = \sum_{n=0}^{\infty} \frac{\alpha^n \mu_n e^{x\sqrt{2n}}}{n!} \quad (x < 0). \tag{1.14}$$

All we need to know in order to calculate the moments μ_n $(n \geqslant 0)$ is that

for fixed α, $x \mapsto \Phi(\alpha, x)$ is continuous at $x = 0$, (1.15a)

for fixed α, $x \mapsto \Phi(\alpha, x)$ is differentiable at $x = 0$. (1.15b)

For proof, see Section 1.3.

On combining (1.12), (1.14) and (1.15), we derive

$$\sum_{n=0}^{\infty} \frac{\alpha^n \mu_n}{n!} = e^{\alpha} \sum_{n=0}^{\infty} \frac{(-\alpha)^n \mu_n}{n!}\,, \tag{1.16a}$$

$$\sum_{n=1}^{\infty} \frac{\alpha^n \mu_n \sqrt{2n}}{n!} = -e^{\alpha} \sum_{n=1}^{\infty} \frac{(-\alpha)^n \mu_n \sqrt{2n}}{n!}\,. \tag{1.16b}$$

Of course, (1.16a) just repeats the fact that A and $1 - A$ have the same distribution under $\mathbb{P}$. On comparing coefficients of α^n in equations (1.16), we obtain

$$\mu_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \mu_k, \tag{1.17a}$$

$$\mu_n = (\sqrt{2n})^{-1} \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \mu_k \sqrt{2k}\,, \tag{1.17b}$$

equation (1.17a) recursively giving μ_n for odd n, and (1.17b) giving μ_n for even n — with no contradictions in the other cases. (Of course, $\mu_0 = 1$.)

Deriving (1.6b) is now a very easy exercise, and is left to the demanding reader. □

1.3 The Feynman-Kač approach (modified)

The fact that $x \mapsto \Phi(\alpha, x)$ is differentiable at $x = 0$ was crucial in Section 1.2, and one of the advantages of the Feynman-Kač approach is that it allows us to prove this differentiability property. (Warning: for $\alpha \neq 0$, the map $x \mapsto \Phi(\alpha, x)$ is *not* twice differentiable at $x = 0$.)

Define

$$A_t := \int_t^\infty e^{-u} H(B_u) du = e^{-t} \int_0^\infty e^{-u} H(B_{u+t}) du$$

and note that

$$\begin{aligned} e^{\alpha A} - 1 = e^{\alpha A_0} - e^{\alpha A_\infty} &= - \int_0^\infty \frac{d}{dt} (e^{\alpha A_t}) dt \\ &= \alpha \int_0^\infty e^{-t} H(B_t) e^{\alpha A_t} dt. \end{aligned}$$

Now,

$$\begin{aligned} \mathbb{E}^x (e^{\alpha A_t} | B_t) &= \mathbb{E}^x \left(\exp \left\{ \alpha e^{-t} \int_0^\infty e^{-u} H(B_{u+t}) dt \right\} \middle| B_t \right) \\ &= \Phi(\alpha e^{-t}, B_t). \end{aligned}$$

Hence

$$\Phi(\alpha, x) := \mathbb{E}^x (e^{\alpha A}) = 1 + \alpha \int_0^\infty e^{-t} \mathbb{E}^x \{ H(B_t) \Phi(\alpha e^{-t}, B_t) \} dt,$$

so that we have the following integral equation for Φ:

$$\Phi(\alpha, x) = 1 + \alpha \int_{t=0}^\infty \int_{y=-\infty}^\infty e^{-t} H(y) \Phi(\alpha e^{-t}, y) p(t, x, y) dy dt, \tag{1.18}$$

where p denotes the Brownian transition density function. Thus

$$\Phi(\alpha, x) = 1 + \alpha \int_0^\infty \int_0^\infty e^{-t} \frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}} \Phi(\alpha e^{-t}, y) dy dt. \tag{1.19}$$

(The multiplicative property of the exponential function has allowed the collapsing of a dimension.) At least formally, we have

$$\frac{\partial \Phi}{\partial x} (\alpha, x) = \alpha \int_0^\infty \int_0^\infty e^{-t} \left\{ \frac{\partial}{\partial x} p(t, x, y) \right\} \Phi(\alpha e^{-t}, y) dt dy. \tag{1.20}$$

If, for $c \in \mathbb{R}$, T_c (or $T(c)$) denotes the hitting time $\inf\{t : B_t = c\}$, then

$$\mathbb{P}^0 (T_{y-x} \in dt) = \text{sgn}\,(y - x) \frac{\partial}{\partial x} p(t, x, y) dt.$$

Hence, the not-yet-proved formula (1.20) may be written:

$$\frac{\partial \Phi}{\partial x}(\alpha, x) = S(\alpha, x) := \alpha \mathbb{E}^0 \int_0^\infty \operatorname{sgn}(y-x) e^{-T(y-x)} \Phi(\alpha e^{-T(y-x)}, y) dy. \tag{1.21}$$

However the expression $S(\alpha, x)$ is very well behaved. Firstly, $\Phi(\alpha e^{-T(c)}, y)$ is bounded by $e^{|\alpha|}$; secondly,

$$\mathbb{E}^0 e^{-T(c)} = \exp(-|c|\sqrt{2});$$

and thirdly, each of the functions $\operatorname{sgn}(y-x)$, $T(y-x)$ is (almost surely) continuous in x at a *fixed* x different from y. The Dominated-Convergence Theorem shows that $S(\alpha, x)$ is therefore continuous in x. Moreover, we may now easily justify an application of Fubini's Theorem to show that, for $a < b$,

$$\int_a^b S(\alpha, x) dx = \Phi(\alpha, b) - \Phi(\alpha, a).$$

Since

$$\left(\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial t}\right) p = 0 \quad \text{and} \quad \left(\frac{\partial}{\partial t} + \alpha \frac{\partial}{\partial \alpha}\right)\left[\alpha e^{-t} \Phi(\alpha e^{-t}, y)\right] = 0,$$

it follows formally on applying $\frac{1}{2}\partial^2/\partial x^2$ to (1.18) and then performing an integration by parts that

$$\frac{1}{2}\frac{\partial^2 \Phi}{\partial x^2} + \alpha H(x) \Phi = \alpha \frac{\partial \Phi}{\partial \alpha} \quad (x \neq 0). \tag{1.22}$$

Of course, deriving the modified Feynman-Kač differential equation (1.22) adds nothing because we have already obtained the separation-of-variables solution at (1.14) and (1.12).

An outline of the classical Feynman-Kač method may be found in Williams [39].

1.4 Proof of Theorem 1.1

We begin by outlining the strategy.

We know from (1.9) that

$$-(\pi \log \alpha)^{\frac{1}{2}}\, h(-\alpha A) \to 1 \qquad\qquad (\alpha \to +\infty). \tag{1.23}$$

We shall prove below that

$$\textit{the family } \{(\pi \log \alpha)^{\frac{1}{2}}|h(-\alpha A)| : \alpha \geqslant 1\} \textit{ is uniformly integrable.} \tag{1.24}$$

(Williams [40] contains a reminder about uniform integrability.) It therefore follows that

$$-\mathbb{E}h(-\alpha A) \sim (\pi \log \alpha)^{-\frac{1}{2}} \qquad\qquad (\alpha \to +\infty). \tag{1.25}$$

(Indeed, given (1.23), results (1.24) and (1.25) are equivalent.) Moreover, property (1.6a) shows that

$$\begin{aligned} \mathbb{E}h(\alpha A) &= \mathbb{E}h(\alpha(1-A)) \\ &= \alpha^{\frac{1}{2}} e^{\alpha} \mathbb{E}[R(\alpha, 1-A)(1-A)^{\frac{1}{2}} e^{-\alpha A}] \end{aligned} \tag{1.26}$$

where

$$R(\alpha, 1-A) := \frac{h(\alpha(1-A))}{\{\alpha(1-A)\}^{\frac{1}{2}} e^{\alpha(1-A)}} \ . \tag{1.27}$$

Now, from (1.9),

$$R(\alpha, 1-A) \to 1 \quad \text{as } \alpha \to +\infty, \tag{1.28}$$

and results (1.26) and (1.28) make it plausible (we shall see shortly that it is true) that

$$\mathbb{E}h(\alpha A) \sim \alpha^{\frac{1}{2}} e^{\alpha} \mathbb{E}(e^{-\alpha A}) \qquad\qquad (\alpha \to +\infty). \tag{1.29}$$

From (1.25), (1.29) and (1.6b), we deduce (1.10) which implies Theorem 1.1.

Proof of (1.24) Set

$$g(z) := -h(-z) = \int_0^1 \frac{z e^{-zt}\, dt}{\sqrt{\pi \log 1/t}} \quad (z \in \mathbb{R}^+),$$

and for $\alpha \geqslant 1$, define

$$X_\alpha := -(\pi \log \alpha)^{\frac{1}{2}} h(-\alpha A) = (\pi \log \alpha)^{\frac{1}{2}} g(\alpha A).$$

We must show that the family $(X_\alpha : \alpha \geqslant 1)$ is uniformly integrable.

The function $g : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, and, from (1.9),

$$g(z) \sim (\pi \log z)^{-\frac{1}{2}} \quad (z \to \infty).$$

Hence, for some constants K_0 and K_1,

$$0 \leqslant g(z) \leqslant K_0 \quad (\forall z \geqslant 0), \quad 0 \leqslant (\pi \log z)^{\frac{1}{2}} g(z) \leqslant K_1 \quad (\forall z \geqslant 1).$$

Let $\alpha \geqslant 1$. If $A > \alpha^{-3/4}$ (so that $\alpha A \geqslant 1$), then

$$0 \leqslant X_\alpha \leqslant K_1 \left\{ \frac{\pi \log \alpha}{\pi \log \alpha A} \right\}^{\frac{1}{2}} \leqslant 2K_1,$$

since $\log(\alpha A) \geqslant (\log \alpha)/4$. To put this another way, we have

$$\{X_\alpha > 2K_1\} \subseteq \{A \leqslant \alpha^{-3/4}\}.$$

Thus

$$\mathbb{E}(X_\alpha^2;\, X_\alpha > 2K_1) \leqslant \mathbb{E}(X_\alpha^2;\, A \leqslant \alpha^{-3/4}) \leqslant (\pi \log \alpha) K_0^2 \mathbb{P}(A \leqslant \alpha^{-3/4}),$$

and it now follows from (1.2) that $\mathbb{E}(X_\alpha^2;\, X_\alpha > 2K_1) \to 0$ as $\alpha \to \infty$. Thus

$$K_2 := \sup_{\alpha \geqslant 1} \mathbb{E}(X_\alpha^2;\, X_\alpha > 2K_1) < \infty,$$

and $\sup_\alpha \mathbb{E}(X_\alpha^2) \leqslant 4K_1^2 + K_2 < \infty$. We have shown that the family $(X_\alpha : \alpha \geqslant 1)$ is even bounded in $\mathcal{L}^2$, whence it is certainly uniformly integrable. $\square$

Proof of (1.29) For $\alpha > 0$, define

$$\Phi(-\alpha) := \mathbb{E}(e^{-\alpha A}).$$

Then (1.29) is equivalent to the statement that

$$\mathbb{E}V(\alpha) \to 1 \quad \text{as } \alpha \to \infty,$$

where

$$V(\alpha) := R(\alpha, 1-A)(1-A)^{\frac{1}{2}}\Phi(-\alpha)^{-1}e^{-\alpha A}.$$

Recall that

$$R(\alpha, 1-A) = h(Z_\alpha)\left\{Z_\alpha^{\frac{1}{2}}e^{Z_\alpha}\right\}^{-1} \quad \text{where } Z_\alpha := \alpha(1-A).$$

Since $\log 1/(1-t) \geqslant t$ for $0 < t < 1$, we have, for $z > 0$,

$$h(z) = \int_0^1 \frac{ze^{(1-t)z}dt}{\sqrt{\pi \log 1/(1-t)}} \leqslant ze^z \int_0^\infty \frac{e^{-tz}dt}{\sqrt{\pi t}},$$

so that

$$h(z) \leqslant z^{\frac{1}{2}}e^z \quad (z > 0), \tag{1.30}$$

and

$$\mathbb{E}V(\alpha) \leqslant 1 \quad (\alpha > 0). \tag{1.31}$$

Next, note that for $0 < \eta < \delta < 1$ (and $\alpha > 0$),

$$\mathbb{E}\left(e^{-\alpha A};\, A \leqslant \eta\right) \geqslant e^{\alpha(\delta-\eta)}\mathbb{P}(A \leqslant \eta)\mathbb{E}\left(e^{-\alpha A};\, A > \delta\right),$$

whence it is obvious that

$$\Phi(-\alpha)^{-1}\mathbb{E}\left(e^{-\alpha A};\, A \leqslant \delta\right) \to 1 \quad (\alpha \to \infty). \tag{1.32}$$

For the moment fix δ with $0 < \delta < 1$. Let $\varepsilon > 0$ be given. By (1.8), we can choose $\alpha_0(\delta)$ such that whenever $\alpha \geqslant \alpha_0(\delta)$ and $A \leqslant \delta$, we have $R(\alpha, 1-A) \geqslant 1-\varepsilon$. Then, for $\alpha \geqslant \alpha_0(\delta)$, we have

$$\begin{aligned}\mathbb{E}V(\alpha) &\geqslant \mathbb{E}(V(\alpha);\, A \leqslant \delta) \\ &\geqslant (1-\varepsilon)(1-\delta)^{\frac{1}{2}}\Phi(-\alpha)^{-1}\mathbb{E}(e^{-\alpha A};\, A \leqslant \delta),\end{aligned}$$

so that, by (1.32),

$$\liminf \mathbb{E}V(\alpha) \geqslant (1-\varepsilon)(1-\delta)^{\frac{1}{2}}. \tag{1.33}$$

Since (1.33) is true for all ε and δ in (0,1) and since also (1.31) is true, the result follows. $\square$

1.5 A generalization

The 'mysterious' symmetry property (1.6b) finds its natural setting in excursion theory (see part 8 of chapter 6 of Rogers and Williams [33] for this theory), and may therefore be studied in very general contexts. Here (for those who know the jargon) is a first generalization of our earlier result to certain 1-dimensional diffusions.

Suppose for example that X is a diffusion process on $\mathbb{R}$ satisfying the stochastic differential equation

$$dX = \sigma(X)dB + b(X)dt, \quad X_0 = 0, \tag{1.34}$$

where σ and b are nice functions on $\mathbb{R}$, σ being positive and even, and b being odd. For $\lambda > 0$, define

$$A_\lambda := \int_0^\infty \lambda e^{-\lambda t} H(X_t)dt. \tag{1.35}$$

Then the distribution of A_λ is completely characterized by the following two symmetry properties:

$$A_\lambda \text{ and } 1 - A_\lambda \text{ have the same distribution,} \tag{1.36a}$$

$$\mathbb{E}h_\lambda(-\alpha A_\lambda) = -e^{-\alpha}\mathbb{E}h_\lambda(\alpha A_\lambda) \quad (\alpha \in \mathbb{R}), \tag{1.36b}$$

where

$$h_\lambda(z) = \int_0^1 z e^{zt} \nu[\lambda^{-1} \log 1/t, \infty]dt, \tag{1.37}$$

where ν is (some normalization of) the Lévy measure on $(0, \infty]$ which describes the durations of excursions from 0. If p is the transition density function of X, then

$$\int_{(0,\infty]} (1 - e^{-\theta t})\nu(dt) = \text{const.} \left\{ \int_0^\infty e^{-\theta t} p_t(0,0)dt \right\}^{-1}. \tag{1.38}$$

Excursion theory therefore reiterates the fundamental problem.

Result (1.36b) may be proved by the methods of this Chapter. Section 2.4 of Chapter Two will both derive (1.36b) from a more fundamental result and relate that proof to the methods used here by means of Lévy's formula.

1.6 Numerical analysis: finding the moments

The recursion scheme (in which $\mu_0 = 1$ and)

$$2^n \mu_n = \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} 2^k \mu_k, \qquad (n \text{ odd}), \tag{1.39a}$$

$$2^n \mu_n \sqrt{2n} = \sum_{k=0}^{n-1} (-1)^{k+1} \binom{n}{k} 2^k \mu_k \sqrt{2k} \qquad (n \text{ even}), \tag{1.39b}$$

is equivalent to, but somewhat more stable than, that provided by equations (1.17).

Proof of (1.39a) Equation (1.39a) just asserts that

$$\mathbb{E}[(2A-1)^n] = 0 \quad (n \text{ odd}). \qquad \square$$

Proof of (1.39b) The following Lemma with $a_n = \mu_n \sqrt{2n}$ shows that the equations at (1.39b) are linearly dependent on the even-n equations at (1.17b). (Section 1.5 makes it clear that the Lemma is relevant in other cases too.) $\square$

Lemma 1.3 *For a sequence $a = (a_n : n = 1, 2, \ldots)$ of real numbers, define*

$$E_r(a) := \sum_{k=1}^{2r} (-1)^k \binom{2r}{k} a_k + a_{2r} \quad (r \geqslant 1),$$

$$F_n(a) := \sum_{k=1}^{2n} (-1)^k \binom{2n}{k} 2^k a_k \quad (n \geqslant 1).$$

Then there exist unique constants $(c_{n,r} : n \geqslant r \geqslant 1)$ such that

$$F_n(a) = \sum_{r=1}^{n} c_{n,r} E_r(a) \quad (n \geqslant 1). \tag{1.40}$$

Proof of Lemma 1.3 Standard linear-dependence considerations show that it is enough to prove that (1.40) holds whenever a has the form

$$a_k = (\tfrac{1}{2} - \tfrac{1}{2}t)^k,$$

in which case (with obvious abuse of notation)

$$F_n(t) = t^{2n} - 1, \quad E_r(t) = 2^{-2r} \left\{ (1+t)^{2r} + (1-t)^{2r} \right\} - 1,$$

and linear independence of the polynomials (in t^2) $E_1, E_2, \ldots, E_n$ shows that, for some unique constants $c_{n,r}(n \geqslant r \geqslant 1)$ and $c_{n,0}(n \geqslant 1)$, we have

$$F_n(t) = \sum_{r=1}^{n} c_{n,r} E_r(t) + c_{n,0}.$$

However, when $t = 1$, $F_n(t) = 0$ and $E_r(t) = 0$ (for all $n \geqslant 1$ and $r \geqslant 1$). Hence $c_{n,0} = 0$ for every n. □

We had conjectured that the Lemma is true with

$$c_{n,r} = \binom{2n}{2r} 2^{2r-1} b_{n-r}, \tag{1.41}$$

where $(b_n : n = 0, 1, 2, \ldots)$ is the sequence of integers specified by

$$\sum_{k=0}^{n} \binom{2n+1}{2k} 2^{2n-2k} b_k = 1 \quad (n \geqslant 0). \tag{1.42}$$

The first proof of the Lemma was then obtained, and in this strong form, by Kevin Buzzard of Trinity College, Cambridge. He used a clever induction argument.

We then found the 'abstract' proof given above, and it reveals the whole structure. For if we define integers $(b_n : n = 0, 1, 2, \ldots)$ via

$$\sum_{n=0}^{\infty} b_n \frac{\theta^{2n}}{(2n)!} = \operatorname{sech}\theta, \tag{1.43}$$

and define $c_{n,r}$ as at (1.41), then the Lemma follows on comparing coefficients of θ^{2n} in the identity

$$2(\cosh\theta t - \cosh\theta) = \{\cosh(1+t)\theta + \cosh(1-t)\theta - \cosh 2\theta - 1\}\operatorname{sech}\theta.$$

It is now merely an exercise in manipulation to show that (1.39b) amounts to (1.6b) with 2α replacing α.

In order to calculate μ_n up to $n = 512$ (using (1.39)) and to invert the moments (as described in the next section) to obtain Table 1.1, we had to use *Mathematica* working initially to 1400-digit precision! The numbers in the 2nd and 3rd columns of Table 1.2 are rounded to the last occurring digit.

Table 1.2: some moments of A:

n	μ_n	$(n\log n)^{\frac{1}{2}}\mu_n - \pi^{-\frac{1}{2}}$
4	0.2355459516514725	−0.0095212550729781
8	0.1518057773661884	0.0549757768411142
16	0.0966634970812128	0.0796315382569828
32	0.0615894446082279	0.0844135009074081
64	0.0395051386283965	0.0803237892749407
128	0.0255645569083701	0.0729066024601239
256	0.0166932887256634	0.0647654709911452
512	0.0109919876599126	0.0570300854833243

1.7 Numerical analysis: moment-inversion formula

If X is a random variable with values in [0,1] and with nice probability density function f on [0,1], then, for $0 < x < 1$,

$$f(n,x) := n\binom{n}{[nx]}\left(D^{n-[nx]}m\right)_{[nx]} \to f(x) \quad (n \to \infty). \tag{1.44}$$

In (1.44),

$$m_k := \mathbb{E}(X^k) \quad (k = 0,1,2,\ldots),$$

and for a sequence $a = (a_k : k = 0,1,2,\ldots)$, we define the sequence Da via

$$(Da)_k := a_k - a_{k+1} \quad (k = 0,1,2,\ldots).$$

(See Exercise E18.5 in Williams [40], and extend it along the lines of Exercise E7.1 there.)

Let X have the Beta distribution with density

$$f(x) = \frac{\Gamma(\beta+\gamma)}{\Gamma(\beta)\Gamma(\gamma)}\, x^{\beta-1}(1-x)^{\gamma-1} \quad (0 \leqslant x \leqslant 1),$$

where $\beta > 0, \gamma > 0$. Then for $0 \leqslant i \leqslant n$,

$$n\binom{n}{i}(D^{n-i}m)_i = n\,\frac{\Gamma(n+1)}{\Gamma(\beta+\gamma+n)}\,\frac{\Gamma(\beta+i)}{\Gamma(i+1)}\,\frac{\Gamma(n-i+\gamma)}{\Gamma(n-i+1)}\,\frac{\Gamma(\beta+\gamma)}{\Gamma(\beta)\Gamma(\gamma)}\,,$$

and, since (see §4.42 of Titchmarsh [35]), for $a > 0$,

$$\Gamma(r+a)/\Gamma(r) = r^a(1 + \mathrm{O}(r^{-1})) \qquad (r \to \infty),$$

we have

$$f(n,x) = f(x) + \mathrm{O}(n^{-1}), \tag{1.45}$$

the error term being uniform over x in $(\delta, 1-\delta)$ for any δ in $(0, \frac{1}{2})$. It seems reasonable to suppose that (1.45) will hold when m is replaced by μ and f by the unknown density f_A of A. Indeed, there is fairly strong numerical evidence of an asymptotic expansion (for each x) of the form

$$f_A(n,x) \sim f_A(x) + \sum_k a_k(x) n^{-k},$$

and this idea was used in making the estimates at Table 1.1.

Chapter Two

Discounted Functionals

*In completing one discovery we never fail to get
an imperfect knowledge of others of which we
could have no idea before, so that we cannot solve
one doubt without creating several new ones.*

Joseph Priestley, *Experiments and Observations
on Different Kinds of Air*

2.1 A large-deviations problems

As in Chapter One, we are mainly concerned with exponentially discounted additive functionals. We find that the large-deviation behaviour of the average depends on the precise average used. We derive, in certain cases, a link (but not equality) between the Cesàro average and Abel average limits, and would expect that other averages would produce other limiting behaviours. We focus still on the exponentially discounted (Abel average) case, both because of its tractability and because of its frequent appearance in decision/control problems and models of financial markets. We do give in Section 2.4 the promised 'excursion' treatment of symmetry characterizations of the type intimated in Chapter One; and this new treatment is simpler, more illuminating and more general. First, however, we focus attention on a different kind of asymptotic behaviour from that studied in Chapter One, and on differential equations for exact results.

Let X be an irreducible continuous-parameter Markov chain on the finite set $S = \{1, 2, \ldots, n\}$, with Q-matrix Q. For $\lambda > 0$, define the random probability measure A_λ on S by setting

$$A_\lambda(i) := A_\lambda(\{i\}) := \int_0^\infty \lambda e^{-\lambda t} I_i(X_t)dt, \tag{2.1}$$

where $I_i(j) := \delta_{ij}$ $(i, j \in S)$. By the ergodic theorem, we have almost surely (a.s.)

$$A_\lambda(i) \to \pi_i \qquad \text{as } \lambda \downarrow 0, \tag{2.2}$$

where π is the invariant probability distribution for X. One is therefore obliged to ask: what is the associated large-deviation theory? One expects that, irrespective of where X starts, we have (as $\lambda \downarrow 0$) some kind of asymptotic formula:

$$\mathbb{P}(A_\lambda \in dx) \approx \exp\{-\lambda^{-1}K(x)\}\text{meas}(dx) \tag{2.3}$$

for some rate-function K on the set

$$M := \left\{(x_i)_{i\in S} : x_i \geqslant 0, \quad \sum x_j = 1\right\} \subseteq \mathbb{R}^n$$

of probability measures on S, where meas denotes the normalized Euclidean measure of total mass 1 on the $(n-1)$-simplex M. Indeed, in this case, one expects (2.3) to hold in the ultra-precise sense that if f_A^λ denotes the density of the distribution of A_λ relative to the measure meas, then

$$\lim_{\lambda\downarrow 0} \lambda \log f_A^\lambda(x) = -K(x), \qquad x \in M. \tag{2.4}$$

The standard precise form of the large-deviation principle is given in Theorem 2.1.

A heuristic appeal to 'the Laplace-Varadhan principle' suggests that for any function v on S,

$$\begin{aligned} \mathbb{E} \exp\left\{\int_0^\infty e^{-\lambda t} v(X_t)dt\right\} &= \mathbb{E}\exp\left\{\lambda^{-1}\sum v_i A_\lambda(i)\right\} \\ &\approx \int \exp\left\{\lambda^{-1}\Big(\sum v_i x_i - K(x)\Big)\right\} \text{meas}(dx). \end{aligned}$$

Again, one expects this to translate into precise form:

$$\lim_{\lambda\downarrow 0} \lambda \log \mathbb{E}\exp\left\{\int_0^\infty e^{-\lambda t} v(X_t)dt\right\} = \eta(v), \tag{2.5}$$

where

$$\eta(v) = \sup_{x\in M}\left\{\sum_i v_i x_i - K(x)\right\}. \tag{2.6}$$

The usual pattern of things in large-deviation theory would then give

$$K(x) = \sup_{v\in\mathbb{R}^n}\left\{\sum_i v_i x_i - \eta(v)\right\}. \tag{2.7}$$

Let us recall briefly the classical results on undiscounted occupation times in this chain setting. These form a minor part of the Donsker-Varadhan theory. (For marvellous treatments of the full theory, see Varadhan [36] and Deuschel and Stroock [15].) Let

$$C_t(i) := t^{-1}\int_0^t I_i(X_s)ds, \qquad t > 0. \tag{2.8}$$

The Feynman-Kač formula allows one to prove that

$$\lim_{t\uparrow\infty} t^{-1}\log \mathbb{E}\,\exp\left\{\int_0^t v(X_s)ds\right\} = \delta(v), \tag{2.9}$$

where

$$\delta(v) := \sup\{\mathbb{R}z : z \in \text{spect}(Q+V)\}, \tag{2.10}$$

where V denotes the diagonal matrix $\text{diag}(v_i)$ and $\text{spect}(\cdot)$ denotes spectrum (here, the set of eigenvalues). The Perron-Frobenius theorem shows that $\delta(v) \in \text{spect}(Q+V)$. Deeper is the fact, at 4.2.16 of Deuschel and Stroock [15], that (in the large-deviation sense of limit)

$$\lim_{t\uparrow\infty} t^{-1} \log f_C^t(x) = -I(x), \tag{2.11}$$

where f_C^t is the density of C_t relative to meas, and, from 4.2.17 of Deuschel and Stroock [15],

$$I(x) = \sup_v \left\{ \sum v_i x_i - \delta(v) \right\}. \tag{2.12}$$

Moreover, $I(x)$ has the alternative expression (Section 13 of Varadhan [36])

$$I(x) = \sup\Big\{ -\sum_i x_i (Qh)_i / h_i : h_i > 0, \forall i \Big\}. \tag{2.13}$$

In the case when X is *symmetrizable* in the usual sense that

$$\pi_i q_{ij} = \pi_j q_{ji}, \qquad \forall (i,j), \tag{2.14}$$

the optimal h for the right-hand side of (2.13) is given (modulo positive scalar multiples) by

$$h_j = (x_j/\pi_j)^{\frac{1}{2}}. \tag{2.15}$$

One can produce a heuristic argument which suggests that η is the convex function

$$\eta(v) = \int_0^1 \alpha^{-1} \delta(\alpha v) d\alpha, \tag{2.16}$$

but proving this in general is going to be one of the highpoints of Chapter Four and involves some technical detail. We now take a natural first step by proving the following theorem. Further development will be made in Sections 3.2 and 4.3.

Theorem 2.1 *Let X be symmetrizable. Then, irrespective of where X starts, the following results are true.*

(i) *Equation* (2.5) *holds with η as at* (2.16).

(ii) *Define K to be the convex conjugate of η as at (2.7). Then the large-deviation principle holds for A_λ with rate function K: in other words,* *for any closed subset F of M,*

$$\limsup_{\lambda\downarrow 0} \lambda \log \mathbb{P}(A_\lambda \in F) \leqslant -\inf\{K(x) : x \in F\},$$

for any open subset G of M,

$$\liminf_{\lambda\downarrow 0} \lambda \log \mathbb{P}(A_\lambda \in G) \geqslant -\inf\{K(x) : x \in G\}.$$

Proof of Theorem 2.1 In Section 2.5 □

Following Theorem 1.1, we can also derive a lemma about the behaviour of the marginal densities at 0, whose proof will enable us to prove the exact result of Theorem 2.3.

Lemma 2.2 *Let X be symmetrizable. Let $F_j^{(i)}$ be the marginal law of $A_\lambda(i)$ starting X at j, and let $Q^{[i]}$ be the substochastic matrix formed by setting the i^{th} row and i^{th} column of Q to 0. Then*

$$\begin{aligned} F_j^{(i)}(x) &= k\big(a_j + o(1)\big)x^{\gamma/\lambda} \qquad (j \neq i), \\ F_i^{(i)}(x) &= k\big(a_i + o(1)\big)x^{(\gamma/\lambda)+1} \qquad \text{where } o(1) \to 0 \text{ as } x \downarrow 0. \end{aligned} \tag{2.17}$$

Here k is a positive constant, and γ and the vector $(a_j)_{j\neq i}$ are respectively the minimal positive eigenvalue and non-negative eigenvector of $-Q^{[i]}$, and a_i is $\sum_{j\neq i} q_{ij}a_j/(\lambda+\gamma)$.

Proof of Lemma 2.2 In Section 2.6 □

2.2 Symmetry characterizations for exact results

The 'symmetry' referred to in the title of this subsection has nothing to do with symmetrizability of the chain X; rather, it has to do with 'symmetry in relation to change in starting position'.

We do not now assume that X is symmetrizable.

There is no natural coordinate system for our set M of probability measures on our state-space $S = \{1, 2, \ldots, n\}$. However each of the measures

$$dx_1 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_n$$

on the co-ordinate space $M_i := \{(x_j)_{j\neq i} : x_j \geqslant 0, \ \sum x_j \leqslant 1\}$ induce the same measure on M. For technical ease we must extend f^λ to a $\mathbb{R}^n$ neighbourhood of M, but the operators we shall use will be invariant to the extension chosen. Similarly on an edge of M, such as $M \cap \{x_i = 0\}$, all the measures

$$dx_1 \dots dx_{i-1}\, dx_{i+1} \dots dx_{j-1}\, dx_{j+1} \dots dx_n$$

on $M_j \cap \{x_i = 0\}$ induce the same measure (for each j), which we call $dx^{(i)}$.

It is possible to calculate the joint moments of $A_\lambda(1), \dots, A_\lambda(n-1)$ in a similar way to that used in Section 2.6. However, even in special cases, it is rather messy to use moments to verify the form of the joint density. The following theorem allows direct verification of the answer in certain circumstances.

Theorem 2.3 *The family $(f_1^\lambda, \dots, f_n^\lambda)$ is characterized by (the fact that each f_i^λ is non-negative and integrates to 1, and) the equations:*

$$\mathcal{L} f^\lambda = -\lambda^{-1} Q f^\lambda, \tag{2.18}$$

where $\mathcal{L}$ is the matrix differential operator

$$\mathcal{L} := \operatorname{diag}\Big(\sum_{j\neq i} (\partial_j - \partial_i) x_j\Big)_{i\in S}$$

using ∂_i to denote $\partial/\partial x_i$, with the boundary conditions that for each i:

$$\lim_{x\downarrow 0} \int_{M\cap\{y_i = x\}} f_i^\lambda(y)\, dy^{(i)} = 0. \tag{2.19}$$

Proof of Theorem 2.3 In Section 2.7 □

Note that Theorem 2.3 implies that

$$-(n-1)\lambda - \mathcal{L}(\lambda \log f_i^\lambda) + (n-1)(\lambda \log f_i^\lambda) = (Qf^\lambda)_i / f_i^\lambda. \tag{2.20}$$

These results make it plausible that if A_λ satisfies a large-deviations principle with rate function K, then

$$(\tilde{\mathcal{L}} - \partial_i) K = -g_i, \tag{2.21}$$

where $\tilde{\mathcal{L}} := \sum x_j \partial_j = \mathcal{L}_i + \partial_i - (n-1)$ and g is the plausible limit

$$g_i(x) = -\lim_{\lambda \downarrow 0} (Qf^\lambda)_i(x)/f_i^\lambda(x). \tag{2.22}$$

Note that it follows from (2.21) that

$$\sum_{k \leqslant n} x_k g_k(x) = 0. \tag{2.23}$$

Proposition 2.4 (Duality Principle) *Assume that A_λ satisfies the large-deviation principle associated with some rate function K and that η and K are related by the pair of Legendre-Fenschel transforms:*

$$\eta(v) = \sup_{y \in M} \left\{ \sum_i v_i y_i - K(y) \right\}, \tag{2.24}$$

$$K(x) = \sup_{w \in \mathbb{R}^n} \left\{ \sum_i w_i x_i - \eta(w) \right\}. \tag{2.25}$$

Assume further that (2.21) *and* (2.22) *hold.*

Let $x \in M$, and let $v := g.(x) \in \mathbb{R}^n$. Then the supremum in (2.24) *is attained when $y = x$, and the supremum at* (2.25) *is achieved when $w = v$. We therefore have, for $x \in M$,*

$$K(x) = -\eta(g.(x)), \tag{2.26}$$

$$(\text{grad}\, \eta)(g.(x)) = x. \tag{2.27}$$

Proof of Proposition 2.4 Extending (2.25) to the convex function $\tilde{K}(x)$ over x in $\mathbb{R}^n$, we can solve (2.24) by Lagrangian methods, using the fact that (2.21) becomes

$$\nabla \tilde{K} = (\tilde{\mathcal{L}} \tilde{K})\mathbf{1} + g, \qquad \text{where} \quad \tilde{\mathcal{L}} = \sum_{j=1}^n x_j \partial_j. \tag{2.28}$$

Then for the appropriate value of the Lagrange multiplier, $y = x$ is a local optimum (and hence a global optimum, by the convexity of $\tilde{K}$). Then (2.23) shows that the value of the supremum is $-K(x)$, confirming (2.26). The remainder of the result is self-evident.

Remarks (i) The function K on M attains its minimum value 0 at π. Thus, from (2.21), $g(\pi) = 0$, and, from (2.27), $(\text{grad}\,\eta)(0) = \pi$.

(ii) If we believe (2.16), then

$$(\text{grad}\,\eta)(v) = \int_0^1 (\text{grad}\,\delta)(\alpha v)d\alpha, \tag{2.29}$$

and so grad η inherits from grad δ the property that it takes its values in M.

(iii) Many of the most interesting questions remain unresolved.

2.3 *An example*

Suppose that π is a probability measure on S with $\pi_i > 0 \quad (\forall i)$, and that

$$q_{ij} = \pi_j - \delta_{ij} \qquad \text{for all } i, j.$$

Then X is symmetrizable with invariant measure π. From (2.15), we find that

$$I(x) = 1 - \left(\sum \sqrt{x_i \pi_i}\right)^2; \tag{2.30}$$

and from (2.10) we find that $\delta(v)$ is the unique root δ in $(\sup(v_i - 1), \infty)$ of

$$\sum \frac{\pi_k}{\delta + 1 - v_k} = 1. \tag{2.31}$$

Defining $\tilde{\eta}$, by

$$\tilde{\eta}(v) = \delta(v) - \sum \pi_k \log[\delta(v) + 1 - v_k], \tag{2.32}$$

we can differentiate and use (2.31) to see that

$$\frac{\partial}{\partial \alpha}\tilde{\eta}(\alpha v) = \frac{\delta(\alpha v)}{\alpha}. \tag{2.33}$$

Theorem 2.1 now tells us that $\eta = \tilde{\eta}$, and that the rate function K is obtained from η as at (2.7). The supremum is achieved (differentiate) at

$$v_k = g_k(x) = 1 - \frac{\pi_k}{x_k}, \tag{2.34}$$

with value

$$K(x) = \sum \pi_i \log(\pi_i / x_i), \tag{2.35}$$

the relative entropy 'the wrong way round'.

In fact, it is possible to calculate $f_A^\lambda(x)$ explicitly for this example: *if X starts at i, and we write $f_i^\lambda(x)$ for $f_A^\lambda(x)$ for this case, then*

$$f_i^\lambda(x) = \frac{x_i \pi_i^{-1} \Gamma(\lambda^{-1})}{\Gamma(n)} \prod_{j \in S} \frac{x_j^{(\pi_j/\lambda)-1}}{\Gamma(\pi_j/\lambda)} ; \tag{2.36}$$

and (of course) this result implies that (2.4) holds with K as at (2.35). The presence of $\Gamma(n)$ is due to the use of the normalized measure meas.

We can prove (2.36) by checking that the the (f_i^λ) satisfy (2.18) and (2.19). We can reduce (2.18) by changing co-ordinates from M to $(\mathbb{R}^+)^n$ (where (2.36) is also defined) to

$$(1+\partial_i) f_i^\lambda - \Big(\sum_j \partial_j x_j\Big) f_i^\lambda = \lambda^{-1} (Q f^\lambda)_i \qquad (1 \leqslant i \leqslant n). \tag{2.37}$$

2.4 Subordinators

Theorem 2.5 should be seen both as preparation for Theorem 2.6, and as an analogue of Theorem 2.3 for a special kind of finite-state semi-Markov process.

Theorem 2.5 *Let $\xi_1, \xi_2, \ldots$ be independent identically distributed (IID) random variables with values in $\{1, \ldots, m\}$ so that there are some $(p_k)_{k=1}^m$ with p_k in $(0,1)$, $\sum_k p_k = 1$, and $\mathbb{P}(\xi_i = k) = p_k$. For each k in $\{1, \ldots, m\}$, let $W_1^k, W_2^k, \ldots$ be IID random variables in $(0, \infty]$. (Note that the value $+\infty$ is allowed.) Suppose that all the variables introduced above are independent. For $i \in \mathbb{N} = \{1, 2, 3, \ldots\}$ and $n \in \mathbb{Z}^+ := \{0\} \cup \mathbb{N}$, define*

$$W_i := \sum_{k=1}^m I_k(\xi_i) W_i^k, \qquad S_n := \sum_{j=1}^n W_j,$$

so that $S_0 := 0$. For $\lambda > 0$, define

$$A_\lambda(k) := \sum_{n=1}^\infty I_k(\xi_n) \big(e^{-\lambda S_{n-1}} - e^{-\lambda S_n}\big).$$

Let

$$\varphi_k(\lambda) := p_k \mathbb{E}\big(1 - e^{-\lambda W_1^k}\big) \qquad (\lambda > 0), \tag{2.38}$$

$$h_\lambda^k(z) := \sum_{n=0}^\infty \frac{\varphi_k(n\lambda) z^n}{n!} \qquad (\lambda > 0,\, z \in \mathbb{R}). \tag{2.39}$$

For a m-vector v, the distribution of $v^\top A_\lambda$ is characterized by the symmetry property:

$$\sum_{k=1}^{m} e^{\alpha v_k}\, \mathbb{E}h_\lambda^k\big(\alpha(v^\top A_\lambda - v_k)\big) = 0 \qquad (\alpha \in \mathbb{R}). \tag{2.40}$$

Proof of Theorem 2.5 In Section 2.8 □

The idea is that we have a process $X = \{X_t : 0 \leqslant t < \infty\}$ such that

$$X = k \text{ on } (S_n, S_{n+1}) \text{ if } \xi_{n+1} = k.$$

Then

$$A_\lambda(k) = \int_0^\infty \lambda e^{-\lambda t} I_k(X_t)dt. \tag{2.41}$$

We now proceed to the continuous-parameter version of Theorem 2.5, which generalizes Section 1.5 of Chapter One.

Recall that a *subordinator* ρ is a right-continuous process $\{\rho_t : t \geqslant 0\}$ with non-decreasing paths and with stationary independent increments and such that $\rho_0 = 0$. If ρ is a subordinator, then for some constant $c \geqslant 0$ and some (Lévy) measure μ on $(0, \infty]$, we have for $\gamma > 0$,

$$\mathbb{E}\exp\{-\gamma\rho(t)\} = \exp\{-t\varphi(\gamma)\},$$

where

$$\varphi(\gamma) = c\gamma + \int_{(0,\infty]} (1 - e^{-\gamma\ell})\mu(d\ell).$$

Then

$$\rho(t) = ct + \sum_{s\leqslant t} (\Delta\rho)(s), \quad (\Delta\rho)(s) := \rho(s) - \rho(s-),$$

and the number of jumps of size greater than ℓ made by ρ during time-interval $[0, t]$ is Poisson with parameter $t\mu[\ell, \infty]$. We allow ρ to make infinite jumps; and of course $\rho = \infty$ from the time of the first infinite jump on. We call our subordinator ρ *driftless* if $c = 0$.

Theorem 2.6 *Suppose that $(\rho_k)_{k=1}^m$ are independent driftless subordinators with*

$$\mathbb{E}\exp\big\{-\gamma\rho_k(t)\big\} = \exp\big\{-t\varphi_k(\gamma)\big\}, \tag{2.42}$$

$$\varphi_k(\gamma) = \int_{(0,\infty]} \big(1 - e^{-\gamma\ell}\big)\mu_k(d\ell). \tag{2.43}$$

Define

$$\rho(t) := \sum_{k=1}^{m} \rho_k(t),$$
$$A_\lambda(k) := \sum_t \exp\{-\lambda\rho(t-)\}\Big[1 - \exp\{-\lambda(\triangle\rho_k)(t)\}\Big].$$

Then, for a m-vector v, the distribution of $v^\top A_\lambda$ is characterized by (2.40)*, where the functions h_λ^k are obtained from the φ_k of* (2.42) *via the analogue of formula* (2.39)*, whereupon*

$$h_\lambda^k(z) = \int_{t=0}^{1} ze^{zt}\mu_k\big[\lambda^{-1}\log(1/t), \infty\big]\,dt. \tag{2.44}$$

Proof of Theorem 2.6 In Section 2.8 □

The obvious application of Theorem 2.6 is to the case of Section 1.5 in which we have a nice diffusion process X on $\mathbb{R}$ with local time L_t at 0 and in which we set

$$\tau_t := \inf\{u : L_u > t\},$$
$$\rho_1(t) := \text{meas}\{s \leqslant \tau_t : X_s > 0\},$$
$$\rho_2(t) := \text{meas}\{s \leqslant \tau_t : X_s < 0\}.$$

Then it is well known that ρ_1 and ρ_2 are independent subordinators. We now have

$$A_\lambda(1) = \int_0^\infty \lambda e^{-\lambda t} I_{(0,\infty)}(X_t)dt.$$

Equation (2.40) now becomes

$$\mathbb{E}h_\lambda^2(\alpha A_\lambda(1)) = -e^{\alpha}\mathbb{E}h_\lambda^1\big(-\alpha(1 - A_\lambda(1))\big), \tag{2.45}$$

which together with (2.44) takes us back to the results of Section 1.5.

It is interesting to compare the method of this paper with that of Chapter One in the light of *Lévy's formula:*

$$\varphi_1(\gamma) = \text{const.}\frac{\partial}{\partial x}\mathbb{E}^x(e^{-\gamma T_0})\big|_{x=0+}, \tag{2.46}$$

where $T_0 := \inf\{t : X_t = 0\}$ and $\mathbb{E}^x$ denotes the law of X starting from x. This formula confirms the equivalence of the equation derived from (1.15b) and the equation (2.45). For discussion of Lévy's formula, see Section 6.2 of Itô and McKean [22].

2.5 Proof of Theorem 2.1

As X is symmetrizable, there is an invariant distribution π, such that Q is a symmetric operator in the Hilbert space defined by the inner product:

$$\langle u, v \rangle := \sum_{i=1}^{n} \pi_i u_i v_i.$$

We use **1** to denote the constant vector which has norm 1. Given a vector v, we set $B_\lambda := \int_0^\infty e^{-\lambda t} v(X_t)\, dt$, and $\varphi_i(\alpha, \lambda) := \mathbb{E}_i(e^{\alpha B_\lambda})$. If we set $B_\lambda(t) := \int_0^t e^{-\lambda s} v(X_s)\, ds$, and we define M to be a martingale, by

$$M_t := \mathbb{E}(e^{\alpha B_\lambda} \mid \mathcal{F}_t), \tag{2.47}$$

then by the strong Markov property

$$M_t = e^{\alpha B_\lambda(t)} \mathbb{E}_{X_t}\left(e^{\alpha e^{-\lambda t} B_\lambda}\right) = e^{\alpha B_\lambda(t)} \varphi_{X_t}(\alpha e^{-\lambda t}, \lambda). \tag{2.48}$$

Applying Itô's formula to (2.48) gives

$$dM_t = e^{\alpha B_\lambda(t)} \left\{ \alpha e^{-\lambda t} v(X_t) \varphi_{X_t}(\alpha e^{-\lambda t}, \lambda) - \alpha \lambda e^{-\lambda t} \frac{\partial \varphi}{\partial \alpha} + Q\varphi \right\} dt + d(\text{mart}), \tag{2.49}$$

whence we deduce that

$$\alpha \lambda \frac{\partial \varphi}{\partial \alpha} = (Q + \alpha V)\varphi, \tag{2.50}$$

where V is diag(v). To find the asymptotics of φ, as λ tends to 0, we discount φ, by defining (with δ as in (2.10))

$$\psi_i(\alpha, \lambda) := \varphi_i(\alpha, \lambda) \exp\left(-\frac{1}{\lambda} \int_0^\alpha \frac{\delta(\beta v)}{\beta}\, d\beta \right), \tag{2.51}$$

and (2.50) becomes:

$$\alpha \lambda \frac{\partial \psi}{\partial \alpha} = -R(\alpha)\psi, \tag{2.52}$$

where $R(\alpha) := \delta(\alpha v) I - (Q + \alpha V)$. The Perron-Frobenius Theorem, of which there is a good treatment in Seneta [34], tells us that $R(\alpha)$ has a zero eigenvalue, and that all other eigenvalues have positive real part. As $R(\alpha)$ is symmetric with respect to the inner product, it is diagonalizable with all eigenvalues but the zero eigenvalue real and positive. Theorem II.1.10 of Kato [24] allows us a smooth (in fact holomorphic in a complex neighbourhood of $\mathbb{R}$) choice of both eigenvalues and orthonormal eigenvectors, which we denote $(\mu_i(\alpha))_{i=1}^n$ and $(y_i(\alpha))_{i=1}^n$ respectively. As $R(\alpha)$ is positive semidefinite, equation (2.52) suggests that as λ gets small, ψ will track y_n (the eigenvector corresponding to 0, with $y_n(0) = \mathbf{1}$). Certainly (2.52) bounds $\|\psi\|$ by 1 because $\psi(0,\lambda) = \mathbf{1}$ and

$$\langle \psi, \psi \rangle' = -\frac{2}{\alpha\lambda} \langle R(\alpha)\psi, \psi \rangle \leqslant 0. \tag{2.53}$$

We can write ψ as $\sum_i \xi_i(\alpha,\lambda) y_i(\alpha)$, knowing that the (ξ_i) are bounded by 1 and the $(y_i'(\alpha))$ are bounded by some K for all α in $[0,1]$. Equation (2.52) becomes:

$$\frac{\partial \xi_i}{\partial \alpha} + \frac{\mu_i(\alpha)\xi_i(\alpha,\lambda)}{\lambda\alpha} = \langle \psi, y_i' \rangle \leqslant K. \tag{2.54}$$

Now for any i not equal to n, we deduce that

$$\frac{\partial}{\partial \alpha}(\xi_i f_i) \leqslant K f_i, \qquad \text{where} \quad f_i(\alpha) := \exp\Big(-\frac{1}{\lambda}\int_\alpha^1 \frac{\mu_i(\beta)}{\beta}\, d\beta\Big). \tag{2.55}$$

Hence

$$\xi_i(\alpha,\lambda) \leqslant K \int_0^\alpha \exp\Big(-\frac{1}{\lambda}\int_\beta^\alpha \frac{\mu_i(\gamma)}{\gamma}\, d\gamma\Big)\, d\beta \leqslant \frac{K}{l}\lambda, \tag{2.56}$$

where l is a lower positive bound for the $\mu_i(\alpha)/\alpha$ for all α in $(0,1]$. The same is true of the $-\xi_i$, so (2.56) actually provides a bound on the modulus of ξ_i. Considering (2.54) with i equal to n shows us that

$$\Big| \frac{\partial \xi_n}{\partial \alpha} \Big| = \Big| \sum_{i \neq n} \xi_i(\alpha,\lambda)\langle y_i, y_n' \rangle \Big| \leqslant K'\lambda, \qquad \text{for some } K',$$

$$\text{and hence} \qquad |\xi_n(\alpha,\lambda) - \xi_n(0,\lambda)| \leqslant K'\alpha\lambda. \tag{2.57}$$

We can now deduce that

$$\lim_{\lambda \to 0} \psi_i(\alpha,\lambda) = y_n(\alpha)(i) > 0. \tag{2.58}$$

The positivity of $y_n(\alpha)(i)$ follows from the Perron-Frobenius Theorem. Finally we can say that

$$\lim_{\lambda \to 0} \lambda \log \mathbb{E}_i \exp\Big(\int_0^\infty e^{-\lambda t} v(X_t)\, dt\Big) = \lim_{\lambda \to 0} \lambda \log \varphi_i(1,\lambda)$$

$$= \lim_{\lambda \to 0} \lambda \Big(\log \psi_i(1, \lambda) + \frac{1}{\lambda} \int_0^1 \frac{\delta(\alpha v)}{\alpha} \, d\alpha \Big)$$

$$= \int_0^1 \frac{\delta(\alpha v)}{\alpha} \, d\alpha = \eta(v). \tag{2.59}$$

Theorem II.2 of Ellis [18], with $Y_\lambda(i) := \int_0^\infty e^{-\lambda t} I_i(X_t) \, dt$ and $c := \eta$, is exactly the second part of the theorem. □

2.6 *Proof of Lemma 2.2*

Conditioning on the first jump allows us to decompose $A_\lambda(i)$ as

$$(A_\lambda(i)|X_0 = i) = 1 - e^{-\lambda \mathcal{E}(q_i)} \big(1 - (\tilde{A}_\lambda(i)|X_0 = k)\big) \quad \text{w.p. } \frac{q_{ik}}{q_i}, \ (k \neq i) \tag{2.60a}$$

$$\text{and} \quad (A_\lambda(i)|X_0 = j) = e^{-\lambda \mathcal{E}(q_j)} (\tilde{A}_\lambda(i)|X_0 = k) \quad \text{w.p. } \frac{q_{jk}}{q_j} \quad (k \neq j \neq i), \tag{2.60b}$$

where $\mathcal{E}(q)$ denotes an exponential random variable with parameter q, and $\tilde{A}_\lambda(i)$ is distributed as $A_\lambda(i)$ and is independent of the exponential variables. ('w.p.' means *with probability*.) In terms of the distributions, equations (2.60) assert that

$$F_i^{(i)}(x) = \frac{1}{\lambda}(1-x)^{q_i/\lambda} \int_0^x \sum_{k \neq i} q_{ik} F_k^{(i)}(y)(1-y)^{-(1+q_i/\lambda)} \, dy, \tag{2.61a}$$

$$\text{and} \quad F_j^{(i)}(x) = x^{q_j/\lambda} \Big(1 + \frac{1}{\lambda} \int_x^1 \sum_{k \neq j} q_{jk} F_k^{(i)}(y) y^{-(1+q_j/\lambda)} \, dy \Big), \quad (i \neq j). \tag{2.61b}$$

Thus the $(F_j^{(i)})$ are smooth on $(0,1)$, and

$$\lambda(x-1) \frac{\partial F_i^{(i)}}{\partial x} = -(QF^{(i)})_i, \tag{2.62a}$$

$$\text{and} \qquad \lambda x \frac{\partial F_j^{(i)}}{\partial x} = -(QF^{(i)})_j, \qquad (i \neq j). \tag{2.62b}$$

Changing the variable in (2.62) to

$$g(t) := F^{(i)}(e^{-\lambda t}),$$

$$\text{gives} \qquad \frac{\partial g}{\partial t} = \big(Q^{[i]} + R - (e^{\lambda t} - 1)^{-1} S \big) g, \tag{2.63}$$

$$\text{where} \qquad R_{jk} := \delta_{ik}(1 - \delta_{ij}) q_{ji}, \quad \text{and} \quad S_{jk} := \delta_{ij} q_{ik}.$$

As X is symmetrizable, we can split $Q^{[i]}$ up into irreducible blocks $I_1, \ldots, I_r$ forming a disjoint partition of $\{1, \ldots, n\} \backslash \{i\}$. By the Perron-Frobenius theorem, each block I_s has an eigenvector $a^{(s)}$ which is strictly positive on I_s and zero elsewhere, with corresponding eigenvector $-\gamma_s < 0$. We assume that $\gamma_1 = \gamma$ is minimal amongst the (γ_s). As $Q^{[i]}$ is self-adjoint, it has n orthogonal eigenvectors, $n-1$ of which have i^{th}-coordinate 0 and are in common with $Q^{[i]} + R$, which also has **1** as a zero eigenvector. The orthogonality also tells us that each $a^{(s)}$ is the unique non-negative eigenvector in I_s. The Levinson theorem, very clearly presented in Eastham [17], says that, as $Q^{[i]} + R$ is diagonalizable, a basic solution g to (2.63) will satisfy

$$e^{\theta t} g(t) \to b \qquad \text{as } t \to \infty, \tag{2.64}$$

where b is an eigenvector of $Q^{[i]} + R$ and $-\theta$ is the corresponding eigenvalue. As g is non-negative and goes to 0, the dominant term must involve one of the $(a^{(s)}, \gamma_s)$. Let X^* be the process X conditioned never to jump out of $I_1 \cup \{i\}$, which will have its discounted occupation time $A^*_\lambda(i) \geqslant A_\lambda(i)$ in the obvious stochastic-domination sense. Thus by the above (as there is only one block and it has $(a^{(1)}, \gamma_1)$ as an eigenpair)

$$F_j^{(i)}(x) \geqslant F_j^{(i)*}(x) \sim k a_j^{(1)} x^{\gamma_1/\lambda} \qquad \text{as } x \downarrow 0 \quad (j \in I_1).$$

We see that the dominant term is that of I_1, so proving the first line of (2.17). The equation at (2.61a) gives the second. (Note that even if γ is common to some of the I_s then there will be some a in the eigenspace for which (2.17) holds.) $\square$

Remark We can use (2.62) again to get the stronger analogue of (2.17) for the marginal densities $(f_j^{(i)})$:

$$\begin{aligned} f_j^{(i)}(x) &= k(\gamma/\lambda)\big(a_j + o(1)\big) x^{\gamma/\lambda - 1} \qquad (j \neq i), \\ f_i^{(i)}(x) &= k(\gamma/\lambda + 1)\big(a_i + o(1)\big) x^{\gamma/\lambda} \qquad \text{where } o(1) \to 0 \text{ as } x \downarrow 0. \end{aligned} \tag{2.65}$$

Thus, for the symmetrizable case, we know the rate of convergence in (2.19).

2.7 Proof of Theorem 2.3

We posit densities $(f_i^\lambda(y))_{i=1}^n$ for A_λ under the $\mathbb{P}_i$ on the $(n-1)$-simplex $M := \{(y_1,\dots,y_n) : y_i \geqslant 0, \quad \sum_i y_i = 1\}$, and dy is the measure induced by any of the co-ordinate measures. Writing (2.50) at $\alpha = \lambda$ in terms of the densities gives:

$$\lambda \int_M \sum_{j=1}^n v_j y_j f_i^\lambda(y) e^{v^\top y}\,dy = \int_M \sum_{j=1}^n q_{ij} f_j^\lambda(y) e^{v^\top y}\,dy + \lambda \int_M v_i f_i^\lambda(y) e^{v^\top y}\,dy. \tag{2.66}$$

Fixing i, and $v_n = 0$, and setting $g_j(x) := (x_j - \delta_{ij}) f_i^\lambda(x)$, an application of Stokes' Theorem gives us

$$-\sum_{j=1}^n \int_{M\cap\{y_j=0\}} g_j(y) e^{v^\top y}\,dy^{(j)} - \sum_{j=1}^n \int_M \frac{\partial g_j}{\partial x_j} e^{v^\top y}\,dy = \frac{1}{\lambda}\int_M (Qf^\lambda)_i(y) e^{v^\top y}\,dy. \tag{2.67}$$

Now using (2.62) we can deduce from

$$\int_0^x \frac{1}{x_j - \delta_{ij}} \int_{M\cap\{y_j=x_j\}} g_j(y)\,dy^{(j)}\,dx_j = F_i^{(j)}(x), \tag{2.68}$$

$$\text{that} \qquad \int_{M\cap\{y_j=x\}} g_j(y)\,dy^{(j)} = -\frac{1}{\lambda}(QF^{(j)})_i(x) \to 0 \qquad \text{as } x \downarrow 0. \tag{2.69}$$

Hence, as $e^{v^\top y}$ is bounded and g_j has constant sign, the boundary terms of equation (2.67) vanish, to show that

$$-\sum_{j=1}^n \int_M \frac{\partial g_j}{\partial x_j} e^{v^\top y}\,dy = \frac{1}{\lambda}\int_M (Qf^\lambda)_i(y) e^{v^\top y}\,dy. \tag{2.70}$$

By the uniqueness of Laplace transforms,

$$-\sum_{j=1}^n \frac{\partial}{\partial x_j}\big((x_j - \delta_{ij}) f_i^\lambda(x)\big) = \frac{1}{\lambda}(Qf^\lambda)_i(x), \tag{2.71}$$

which is merely another way of writing (2.18). Conversely if f^λ satisfies (2.18) and (2.19), then (2.66) must hold, so (2.50) holds with $\varphi_i := \int_M e^{\alpha v^\top y/\lambda} f_i^\lambda(y)\,dy$. We can use (2.53) to show that the solution to (2.50) is unique, and so f_i^λ is the law of A_λ given $X_0 = i$. □

2.8 Proofs of Theorem 2.5 and Theorem 2.6

Note firstly that $h_\lambda^k(z)$ as defined by (2.38) and (2.39) satisfies

$$h_\lambda^k(z) = p_k e^z \mathbb{E}\Big(1 - \exp\big\{-z(1-e^{-\lambda W_1^k})\big\}\Big). \tag{2.72}$$

Then (2.40) can be seen simply to express A_λ conditional on ξ_1, as in

$$A_\lambda = \big(1 - e^{-\lambda W_1^k}\big)e_k + e^{-\lambda W_1^k}\tilde{A}_\lambda, \quad \text{w.p. } p_k, \tag{2.73}$$

where $\tilde{A}_\lambda$ also has the A_λ distribution and is independent of the W_1's, and e_k is the unit positive vector in the k^{th} direction. Rewriting (2.40) in terms of the moments (μ_n) of $v^\top A_\lambda$, we derive that

$$\mu_n = -\Big(\sum_{k=1}^m \varphi_k(n\lambda)\Big)^{-1} \sum_{r=0}^{n-1}\binom{n}{r}\mu_r \sum_{s=0}^{n-r}\binom{n-r}{s}(-1)^s \sum_{k=1}^m v_k^{n-r}\varphi_k\big((r+s)\lambda\big). \tag{2.74}$$

The symmetry property therefore gives us the moments, and as is well known (for a bounded distribution) sufficient unto the law are the moments thereof. □

Now we turn our attention towards Theorem 2.6. For motivation and further explanation of excursion theory, see Part VI.8 of Rogers and Williams [33]. Let ϵ be positive. We aim to approximate A_λ by a Theorem 2.5-type scenario by discarding all jumps of the ρ-process of size less than ϵ. Formally we write

$$\begin{aligned}
\rho_k^\epsilon(t) &:= \sum_{u\leqslant t} \triangle\rho_k(u) I(\triangle\rho_k(u) \geqslant \epsilon), \\
\rho^\epsilon(t) &:= \sum_{k=1}^m \rho_k^\epsilon(t), \\
\text{and} \quad A_\lambda^\epsilon(k) &:= \sum_t e^{-\lambda\rho^\epsilon(t-)}\Big(1 - e^{-\lambda\triangle\rho_k^\epsilon(t)}\Big).
\end{aligned}$$

In the language of Theorem 2.5, $p_k = \mu_k[\epsilon,\infty]/\mu[\epsilon,\infty]$ where $\mu = \sum_k \mu_k$, and

$$\varphi_k^\epsilon(\lambda) = p_k \int_{[\epsilon,\infty]} (1 - e^{-\lambda x})\,\mu_k(dx)/\mu_k[\epsilon,\infty], \tag{2.75}$$

$$\text{and} \quad \sum_{k=1}^m e^{\alpha v_k}\mathbb{E}h_\lambda^{\epsilon,k}\big(\alpha(v^\top A_\lambda^\epsilon - v_k)\big) = 0. \tag{2.76}$$

We can renormalise the φ^ϵ so that $\varphi_k^\epsilon(\lambda) = \int_{[\epsilon,\infty]}(1-e^{-\lambda x})\,\mu_k(dx)$, and then

$$|\varphi_k^\epsilon(\lambda) - \varphi_k(\lambda)| = \int_{(0,\epsilon)} (1-e^{-\lambda x})\,\mu_k(dx), \quad \text{and} \tag{2.77}$$

$$|h_\lambda^{\epsilon,k}(z) - h_\lambda^k(z)| \leqslant \int_{(0,\epsilon)} \left(e^z - e^{ze^{-\lambda x}}\right)\mu_k(dx) \leqslant ze^z \int_{(0,\epsilon)} (1-e^{-\lambda x})\,\mu_k(dx). \tag{2.78}$$

Thus $h_\lambda^{\epsilon,k} \to h_\lambda^k$ uniformly on bounded intervals. Writing

$$A_\lambda^\epsilon(k) = \sum_{t:\triangle\rho_k(t)>0} I(\triangle\rho(t) \geqslant \epsilon)\left(e^{-\lambda\rho^\epsilon(t-)} - e^{-\lambda\rho^\epsilon(t)}\right),$$

$$\text{and} \quad 1 - A_\lambda^\epsilon(k) = \sum_{t:\,\substack{\triangle\rho(t)>0 \\ \triangle\rho_k(t)=0}} I(\triangle\rho(t) \geqslant \epsilon)\left(e^{-\lambda\rho^\epsilon(t-)} - e^{-\lambda\rho^\epsilon(t)}\right),$$

and applying Fatou's lemma to each line gives us that $A_\lambda^\epsilon(k) \to A_\lambda(k)$ as ϵ goes to 0 (almost surely). Combining this, the control on h^ϵ and the Bounded Convergence Theorem, we can take the limit of (2.76) to obtain (2.40). The remarks around (2.74) still apply to show that A_λ is determined by the symmetry equation. Finally using equations (2.39) and (2.40) and carefully integrating by parts, we see that

$$h_\lambda^k(z) = \int_{(0,\infty]} (e^z - e^{ze^{-\lambda s}})\,\mu_k(ds) = \int_0^\infty z\lambda e^{-\lambda s} e^{ze^{-\lambda s}} \mu_k[s,\infty]\,ds, \tag{2.79}$$

whence we can deduce (2.44). $\square$

Chapter Three

Asymmetric Markov chains

You can only find truth with logic if you
have already found truth without it.

***G. K. Chesterton**, The Man who was Orthodox*

3.1 Introduction and abstract

In Chapter One we began a study of Abel averages, $A_\lambda(i) := \lambda \int_0^\infty e^{-\lambda t} I_i(X_t)\, dt$, as opposed to the oft-studied Cesàro averages $C_t(i) := t^{-1} \int_0^t I_i(X_s)\, ds$. In Chapter Two, we studied the large-deviation behaviour of these averages. In the case where X is an irreducible Markov chain on a finite state-space $S = \{1, \ldots, n\}$, we observed that

$$\begin{aligned} &C_t \to \pi \qquad \text{as } t \to \infty, \\ \text{and} \qquad &A_\lambda \to \pi \qquad \text{as } \lambda \downarrow 0, \end{aligned}$$

where π is the invariant distribution of X. We noted that

$$\lim_{t\to\infty} t^{-1} \log \mathbb{E} \exp\Big\{ \int_0^t v(X_s)\, ds \Big\} = \delta(v), \tag{3.1}$$

where v is an n-vector, $\delta(v) := \sup\{\mathrm{Re}(z) : z \in \mathrm{spect}(Q+V)\}$, and where spect$(\cdot)$ denotes spectrum (here the set of eigenvalues), Q is the Q-matrix of X, and V denotes the diagonal matrix diag(v_i). It is also true that the large-deviation property holds for C_t with rate function I defined on $M := \{(x_i)_{i\in S} : x_i \geqslant 0,\ \sum_j x_j = 1\}$. That is that

$$\begin{aligned} &\limsup_{t\to\infty} t^{-1} \log \mathbb{P}(C_t \in F) \leqslant -\inf_{x\in F} I(x), \\ \text{and} \qquad &\liminf_{t\to\infty} t^{-1} \log \mathbb{P}(C_t \in G) \geqslant -\inf_{x\in G} I(x), \end{aligned} \tag{3.2}$$

for F and G respectively closed and open subsets of M. These two limits, δ and I, are related by convex conjugation (Legendre transform), in that

$$\begin{aligned} &I(x) = \delta^*(x) := \sup_{v\in\mathbb{R}^n} \Big\{ \sum_i v_i x_i - \delta(v) \Big\}, \\ \text{and} \qquad &\delta(v) = I^*(v) := \sup_{x\in M} \Big\{ \sum_i v_i x_i - I(x) \Big\}. \end{aligned} \tag{3.3}$$

In Chapter Two we were able to derive similar results for the Abel average when X was symmetrisable. We can now extend this result both to the general X, and also to the family of discounted averages, G^γ, parameterised by $\gamma \geqslant 0$, where

$$\begin{aligned} &G_t^\gamma(i) := t^{-1}(1+\gamma) \int_0^t (1 - s/t)^\gamma\, I_i(X_s)\, ds, \\ \text{and} \qquad &G_t^\gamma \to \pi \qquad \text{as } t \to \infty. \end{aligned} \tag{3.4}$$

3.2 The results

We first extend Theorem 2.1 of Chapter Two to all (including asymmetric) finite chains

Theorem 3.1 *Let X be an honest irreducible finite-state Markov chain, with Q-matrix Q. Then, irrespective of where X starts*

$$\lim_{\lambda \downarrow 0} \lambda \log \mathbb{E} \exp\Big\{ \int_0^\infty e^{-\lambda t} v(X_t)\, dt \Big\} = \eta(v) := \int_0^1 \alpha^{-1} \delta(\alpha v)\, d\alpha, \tag{3.5}$$

and the large-deviation property holds for A^λ with rate $K := \eta^*$, and $\eta = K^*$.

Proof of Theorem 3.1 Our programme is essentially to follow the proof in Chapter Two, but is complicated by the asymmetry of Q breaking the usual link between the spectrum of Q and its action as a bilinear form $\langle \cdot, Q \cdot \rangle$, where $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{R}^n$.

As in Chapter Two we define

$$B_\lambda := \int_0^\infty e^{-\lambda t} v(X_t)\, dt,$$

$$\text{and} \qquad \varphi_i(\alpha, \lambda) := \mathbb{E}_i(e^{\alpha B_\lambda}),$$

and deduce from applying Itô's formula to the martingale $M_t := \mathbb{E}(\exp(\alpha B_\lambda) | \mathcal{F}_t)$ that

$$\alpha \lambda \frac{\partial \varphi}{\partial \alpha} = (Q + \alpha V) \varphi. \tag{3.6}$$

We again discount φ by defining ψ by

$$\psi_i(\alpha, \lambda) := \varphi_i(\alpha, \lambda) \exp\Big\{ -\frac{1}{\lambda} \int_0^\alpha \beta^{-1} \delta(\beta v)\, d\beta \Big\},$$

which satisfies the differential equation

$$\alpha \lambda \frac{\partial \psi}{\partial \alpha} = -R(\alpha) \psi, \tag{3.7}$$

where $R(\alpha) := \delta(\alpha v) I - (Q + \alpha V)$ has a spectrum with non-negative real part. Our solution to this equation follows from facts deriving from the Perron-Frobenius Theorem about non-negative or essentially non-negative (non-negative off the diagonal) matrices. These may be found in I.7 of Kato [24] or Seneta [34]. The points we shall use most are as follows. A non-negative matrix has a maximal modulus eigenvalue which is real and positive, and whose corresponding eigenvector is non-negative. If the matrix is irreducible (in the stochastic sense) the eigenvalue is simple and the eigenvector is both positive and the only non-negative eigenvector.

In the symmetric case we could decompose R into acting on orthogonal eigenspaces with all but the zero eigenvalue positive. In the general asymmetric case, we still have all but one eigenvalue positive, but the matrix need not be diagonalisable and the eigenspaces are not all mutually orthogonal. Nevertheless, we can produce a partial orthogonal decomposition by adapting theorem I.7.13 of Kato [24] to give the following.

Theorem 3.2 (Adapted from Kato) *Let T be an essentially non-negative and irreducible matrix. Let δ be its principal eigenvalue. Then there exists a real diagonal matrix F with positive elements such that $S := F^{-1}TF - \delta I$ has a simple eigenvalue zero, with an orthogonal eigenprojection P, and that, for some positive σ*

$$\langle Sx, x\rangle \leqslant -\sigma\|(I-P)x\|^2. \tag{3.8}$$

Proof of Theorem 3.2 Following Kato exactly at this stage, we may assume that T is non-negative, and we know that there exist strictly positive eigenvectors y and y^* of T and $T^\top$ respectively, both with the maximal eigenvalue δ. We set F to be the diagonal matrix diag$(\sqrt{y_i/y_i^*})$, z to be the vector $(z_i) = (\sqrt{y_i y_i^*})$, and B the non-negative matrix $F^{-1}TF$. Then $Bz = B^\top z = \delta z$, and so z is the strictly positive eigenvector of the symmetric non-negative matrix $B^\top B$ with associated eigenvalue δ^2 which is maximal. Kato's theorem I.6.49 tells us that P, the eigenprojection for B associated with z, is self-adjoint and so orthogonal. As $I = P \oplus (I-P)$ is orthogonal we can decompose $\mathbb{R}^n$ as $V_+ \oplus V_-$, the orthogonal sum of two B and $B^\top$ invariant subspaces. Let B_- be the restriction of B to V_-, and then $(B^\top B)_- = (B_-)^\top(B_-)$ and the second greatest eigenvalue of $B^\top B$ (which is real and positive) is the greatest eigenvalue of $(B^\top B)_-$ and is strictly less than δ^2. We shall call this eigenvalue δ_-^2. We have $\|(B^\top B)_-\| = \delta_-^2$, so $\|B_-\| = \delta_-$, and

$$\begin{aligned} &\langle Bx_-, x_-\rangle \leqslant \delta_-\|x_-\|^2, \qquad \text{where } x_- = (I-P)x,\\ \text{and}\qquad &\langle Bx_+, x_+\rangle = \delta\|x_+\|^2, \qquad \text{where } x_+ = Px. \end{aligned}$$

Now $S = F^{-1}TF - \delta I = B - \delta I$, so we have

$$\langle Sx, x\rangle = \langle Sx_+, x_+\rangle + \langle Sx_-, x_-\rangle \leqslant -(\delta - \delta_-)\|x_-\|^2,$$

and the result is proved with σ set equal to $\delta - \delta_-$, which is positive. □

We now apply Theorem 3.2 with $T = T(\alpha) = Q + \alpha V$ and δ equal to $\delta(\alpha v)$. So, for each α in $[0, 1]$, there exist $F(\alpha)$, $P(\alpha)$, $S(\alpha)$, and $\sigma(\alpha)$ as in the theorem. We choose $y(0) = \mathbf{1}$ and $y^*(0) = \pi$, the invariant distribution of Q. By Kato [4] II.1–4, the functions $\delta(\alpha v)$, $y(\alpha)$, $y^*(\alpha)$ and $P(\alpha)$ are (or can be chosen) smooth in α, and $\delta_-(\alpha)$ is at least continuous. We deduce that $F(\alpha)$ is smooth, and $\sigma(\alpha)$ is continuous, so is bounded away from 0 by some positive σ_0 for all α in $[0, 1]$.

Changing bases appropriately we set

$$\chi(\alpha, \lambda) := F^{-1}(\alpha)\psi(\alpha, \lambda),$$

so that $\|\chi(0, \lambda)\| = 1$, and transform (3.7) into

$$\alpha\lambda\frac{\partial\chi}{\partial\alpha} = S(\alpha)\chi(\alpha, \lambda) + \alpha\lambda G(\alpha)\chi(\alpha, \lambda), \tag{3.9}$$

where $G(\alpha) := -F'(\alpha)F^{-1}(\alpha)$ is diagonal, and because $S = -F^{-1}RF$. We can bracket the line above with χ to find that

$$\alpha\lambda\partial_\alpha\langle\chi, \chi\rangle = 2\langle\chi, S\chi\rangle + 2\alpha\lambda\langle\chi, G\chi\rangle,$$

and so

$$\partial_\alpha\langle\chi, \chi\rangle \leqslant 2K_G\langle\chi, \chi\rangle,$$

where $K_G := \sup_\alpha \|G(\alpha)\|$. We deduce that

$$\partial_\alpha(\exp(-2K_G\alpha)\langle\chi, \chi\rangle) \leqslant 0,$$

and so

$$\langle\chi, \chi\rangle(\alpha, \lambda) \leqslant \exp(2K_G\alpha) \leqslant K_\chi \tag{3.10}$$

for all α in $[0, 1]$ and all positive λ. Writing

$$\chi = \chi_+ + \chi_-, \qquad (\chi_-(0, \lambda) = 0),$$

$$\text{then} \qquad \alpha\lambda\chi_-' = S\chi_- + \alpha\lambda\big((I - P)G - P'\big)\chi.$$

We can get a bound on χ_- using a similar method to that used for the universal bound on χ, to see that

$$\partial_\alpha \langle \chi_-, \chi_- \rangle \leqslant -\frac{2\sigma_0}{\alpha\lambda} \langle \chi_-, \chi_- \rangle + K_1,$$

where $K_1 = (K_G + K_P)K_\chi$ and $K_P = \sup_\alpha \|P'(\alpha)\|$, and hence

$$\partial_\alpha \left(\langle \chi_-, \chi_- \rangle \alpha^{2\sigma_0/\lambda} \right) \leqslant K_1 \alpha^{2\sigma_0/\lambda}.$$

Whence we find by integrating that

$$\langle \chi_-, \chi_- \rangle(\alpha, \lambda) \leqslant \lambda \frac{K_1 \alpha}{2\sigma_0 + \lambda} \leqslant K_2 \lambda, \qquad \text{for } \lambda > 0,\, \alpha \in [0, 1]. \tag{3.11}$$

Finally we consider χ_+, the part of χ annhilated by S, for which

$$\partial_\alpha \chi_+ = (P' + PG)\chi_+ + (P' + PG)\chi_-,$$

and writing $\chi_+(\alpha, \lambda)$ as $\xi(\alpha, \lambda)z(\alpha)$, so that $\xi(0, \lambda) = 1$, we see that

$$\xi' z = -\xi P F^{-1}(Fz)' + (P' + PG)\chi_-.$$

Now if we replace y by $y(\alpha)\exp(-\int_0^\alpha \langle y', y^* \rangle \langle y, y^* \rangle^{-1}(\beta)\, d\beta)$ to ensure that $\langle y', y^* \rangle = 0$, and similarly for y^*, we then have $\|z(\alpha)\| = 1$ and $PF^{-1}(Fz)' = 0$. This simplifies the previous differential equation to

$$\begin{aligned} \xi' z &= (P' + PG)\chi_-, \\ \text{and so} \qquad |\xi'(\alpha, \lambda)| &\leqslant (K_P + K_G)\sqrt{K_2 \lambda}. \end{aligned} \tag{3.12}$$

Thus $\lim_{\lambda \downarrow 0} \chi(\alpha, \lambda) = z(\alpha)$, and $\lim_{\lambda \downarrow 0} \psi(\alpha, \lambda) = F(\alpha)z(\alpha) = y(\alpha)$ which is strictly positive, so the first part of the result holds.

Note. In the symmetric case, $y^*(\alpha) = \operatorname{diag}(\pi)y(\alpha)$ for all α, and $z(\alpha) = \operatorname{diag}(\sqrt{\pi})y(\alpha)$. Thus the $\|z\| = 1$ normalisation condition is just the previous $\|y\|_\pi = 1$ normalisation, where $\langle u, v \rangle_\pi = \sum_i \pi_i u_i v_i$.

For the large-deviation result, as in Chapter Two, we use theorem II.2 of Ellis [18] with $Y_\lambda(i) := \int_0^\infty e^{-\lambda t} I_i(X_t)\, dt$, and $c := \eta$. The fact that $\eta = K^*$ comes from the duality of convex conjugation coupled with η inheriting the convexity of δ. $\square$

In fact, we now have enough to consider another average, if for no other reason than to emphasise the point that each average we can think of has different large-deviation behaviour. We look at the "γ-discounted" average ($\gamma \geqslant 0$) defined by (3.4)

$$G_t^\gamma(i) := t^{-1}(1+\gamma)\int_0^t (1-s/t)^\gamma\, I_i(X_s)\,ds.$$

Then we can replicate our previous work with the following.

Theorem 3.3 *Let X be an honest irreducible finite-state Markov chain, with Q-matrix Q. Then, irrespective of where X starts,*

$$\lim_{t\to 0} t^{-1}\log \mathbb{E}\exp\Big\{(1+\gamma)\int_0^t (1-s/t)^\gamma\, v(X_s)\,ds\Big\} = \delta_\gamma(v) := \int_0^1 \delta((1+\gamma)\alpha^\gamma v)\,d\alpha, \tag{3.13}$$

and G_t^γ has the large deviation property with rate $I_\gamma := \delta_\gamma^*$, and $\delta_\gamma = I_\gamma^*$.

Proof of Theorem 3.3 If $\gamma = 0$, G_t^0 is just the Cesàro average C_t for which we already have the result, so we now assume that $\gamma > 0$. We define

$$\varphi_i(\alpha, t) := \mathbb{E}_i \exp\Big\{\alpha(1+\gamma)\int_0^t (1-s/t)^\gamma\, v(X_s)\,ds\Big\},$$

and define M_u to be the martingale

$$\begin{aligned} M_u :&= \mathbb{E}\Big(\exp\Big\{\alpha(1+\gamma)\int_0^t (1-s/t)^\gamma\, v(X_s)\,ds\Big\}\,\Big|\,\mathcal{F}_u\Big) \\ &= \exp\Big\{\alpha(1+\gamma)\int_0^u (1-s/t)^\gamma v(X_s)\,ds\Big\}\,\varphi_{X_u}\big(\alpha(1-u/t)^\gamma, t-u\big). \end{aligned}$$

Applying Itô's formula to the second line above, we deduce that

$$\frac{\alpha\gamma}{t}\frac{\partial\varphi}{\partial\alpha} + \frac{\partial\varphi}{\partial t} = (Q + \alpha(1+\gamma)V)\varphi. \tag{3.14}$$

We now define the discounted function ψ as

$$\psi_i(\alpha,\lambda) := \varphi_i(\alpha, \alpha^{1/\gamma}/\lambda)\exp\Big\{-\lambda^{-1}\int_0^{\alpha^{1/\gamma}} \delta((1+\gamma)\beta^\gamma v)\,d\beta\Big\},$$

which satisfies the differential equation

$$\lambda\gamma\alpha^{1-1/\gamma}\frac{\partial\psi}{\partial\alpha} = -R(\alpha(1+\gamma))\psi. \tag{3.15}$$

This is solved in exactly the same way as we solved (3.7) in Theorem 3.1, with the only exception being on the bound of $\langle \chi_-, \chi_- \rangle$ which is derived from the differential inequality

$$\partial_\alpha \langle \chi_-, \chi_- \rangle \leqslant -\frac{2\sigma_0 \alpha^{1/\gamma}}{\lambda \gamma \alpha} \langle \chi_-, \chi_- \rangle + K_1.$$

From which we deduce that

$$\partial_\alpha \Big(\langle \chi_-, \chi_- \rangle \exp(2\sigma_0 \alpha^{1/\gamma}/\lambda) \Big) \leqslant K_1 \exp(2\sigma_0 \alpha^{1/\gamma}/\lambda),$$

and hence by integrating that

$$\|\chi_-\|^2(\alpha, \lambda) \leqslant K_1 \int_0^\alpha \exp\big(-2\sigma_0(\alpha^{1/\gamma} - \beta^{1/\gamma})/\lambda\big)\, d\beta.$$

The right-hand side of this expression tends to 0 uniformly on α in $[0, 1]$ as $\lambda \downarrow 0$. The rest goes through as before to show that $\lim_{\lambda \downarrow 0} \psi_i(\alpha, \lambda) = y(\alpha)$, and hence that

$$\lim_{t \to \infty} t^{-1} \log \varphi_i(1, t) = \int_0^1 \delta((1+\gamma)\alpha^\gamma v)\, d\alpha. \qquad \square$$

Chapter Four

General discounts

All philosophers who find
Some favourite system to their mind,
In every point to make it fit
Will force all nature to submit.

Jonathan Swift

4.1 Introduction

In this Chapter we will push forward the three major strands of inquiry that we have developed in the proceeding Chapters. The generality that we will obtain should be sufficient to satisfy the demands of a reader just searching for a convenient theorem, though reference will be made to earlier sections when a proof is essentially the same as one before.

In the next Section, we investigate symmetry characterizations and distribution asymptotics for the Abel discounted occupation times of Ornstein-Uhlenbeck processes. This is in the spirit of Chapter One, which studied the Brownian motion case (itself a driftless Ornstein-Uhlenbeck process), and also of Section 2.4 of Chapter Two, which looked at subordinators. The two important results from these Chapters are recapped to aid the casual reader and to reintroduce the concepts that we shall use.

For general results about the large-deviations behaviour of discounted occupation times of processes, in Section 4.3 we will build on Section 2.1 of Chapter Two and on Chapter Three. Previously we knew that the large-deviation property held for the Abel discounted average of a general finite-state Markov chain, and we will fully extend this to completely general discounts of chains and partially extend further to a wide class of Markov processes. We discover that although the large-deviation rate function of the discounted average can be written in terms of that for the Cesàro, the rate is often different for a different discount. We also derive results about the smoothness and finiteness of the rate function which are used in the next Section to prove a central limit theorem.

Finally we shall go beyond the limited approximation precision of the large-deviation property and give an asymptotic expansion of the density of the distribution itself, following from the density studies of Section 2.2. This again is now performed for general discounted averages of finite-state Markov chains. We also notice a pattern in the differential equations we worked with, and hypothesize about their full solutions and other generalizations.

4.2 Symmetry Characterizations

The initial results of this section have previously appeared in Chapters One and Two, but are reproduced here to give the beginnings of the story. They also provide us with our only exact handle on the behaviour of the discounts for diffusions and related processes.

We start by recalling that a *subordinator* (an increasing Lévy process) ρ is a right-continuous process $\{\rho_t : t \geqslant 0\}$ with non-decreasing paths and with stationary independent increments and such that $\rho_0 = 0$. If ρ is a subordinator, then for some constant $c \geqslant 0$ and some (Lévy) measure μ on $(0, \infty]$, we have for $\gamma > 0$,

$$\mathbb{E}\exp\{-\gamma\rho(t)\} = \exp\{-t\varphi(\gamma)\},$$

where
$$\varphi(\gamma) = c\gamma + \int_{(0,\infty]}(1 - e^{-\gamma\ell})\mu(d\ell).$$

Then
$$\rho(t) = ct + \sum_{s\leqslant t}(\Delta\rho)(s), \quad (\Delta\rho)(s) := \rho(s) - \rho(s-),$$

and the number of jumps of size greater than ℓ made by ρ during time-interval $[0, t]$ is Poisson with parameter $t\mu[\ell, \infty]$. We allow ρ to make infinite jumps; and of course $\rho = \infty$ from the time of the first infinite jump on. We call our subordinator ρ *driftless* if $c = 0$.

We suppose that $(\rho_k)_{k=1}^m$ are independent driftless subordinators, and let $\rho(t) := \sum_{k=1}^m \rho_k(t)$. The idea is that we have a process $X = \{X_t : 0 < t < \infty\}$ such that

$$X = k \ \text{ on } \ \big[\rho(t-), \rho(t)\big) \qquad \text{and} \qquad \triangle\rho(t) = \triangle\rho_k(t), \quad \text{if } \triangle\rho_k(t) > 0.$$

Then the discounted occupation measure A_λ is given by

$$A_\lambda(k) := \int_0^\infty \lambda e^{-\lambda t} I_k(X_t)\, dt = \sum_t \exp\big(-\lambda\rho(t-)\big)\Big(1 - \exp\big(-\lambda\triangle\rho_k(t)\big)\Big).$$

Theorem 4.1 *Suppose that (ρ_k), ρ and A_λ are as above, and that each ρ_k can be represented as*

$$\mathbb{E}\exp\big\{-\gamma\rho_k(t)\big\} = \exp\big\{-t\varphi_k(\gamma)\big\}, \tag{4.1}$$

$$\varphi_k(\gamma) = \int_{(0,\infty]}\big(1 - e^{-\gamma\ell}\big)\mu_k(d\ell), \tag{4.2}$$

$$\textit{and define}\qquad h_\lambda^k(z) := \sum_{n=0}^{\infty} \frac{\varphi_k(n\lambda) z^n}{n!}, \tag{4.3}$$

$$\textit{whereupon}\qquad h_\lambda^k(z) = \int_{t=0}^{1} z e^{zt} \mu_k\big[\lambda^{-1}\log(1/t), \infty\big]\, dt. \tag{4.4}$$

Then, for a m-vector v, the distribution of $v^\top A_\lambda$ is characterized by the following identity

$$\sum_{k=1}^{m} \exp(\alpha v_k)\, \mathbb{E}\, h_\lambda^k\big(\alpha(v^\top A_\lambda - v_k)\big) = 0. \tag{4.5}$$

Proof of Theorem 4.1 Theorem 2.6 of Chapter Two. □

As in Chapter Two, we can apply Theorem 4.1 to a well-behaved diffusion process X on $\mathbb{R}$ with local time L_t at 0. We define two subordinators ρ_1 and ρ_2 as

$$\tau_t := \inf\{u : L_u > t\}, \quad \rho_1(t) := \mathrm{Leb}\{s \leqslant \tau_t : X_s > 0\}, \quad \rho_2(t) := \mathrm{Leb}\{s \leqslant \tau_t : X_s < 0\},$$

where Leb is the Lebesgue measure on $\mathbb{R}$. It is known that ρ_1 and ρ_2 are independent, so the discounted occupation of the positive half-line,

$$A_\lambda(1) = \int_0^\infty \lambda e^{-\lambda t} I_{(0,\infty)}(X_t)\, dt,$$

can be characterized by the equation

$$\mathbb{E}h_\lambda^2(\alpha A_\lambda(1)) = -e^\alpha \mathbb{E}h_\lambda^1\big(-\alpha(1 - A_\lambda(1))\big). \tag{4.6}$$

The functions h_λ^1 and h_λ^2 are defined via (4.4), where μ_1 and μ_2 are the Lévy measures on $(0,\infty]$ describing the length of excursions from 0 into $\mathbb{R}^+$ and $\mathbb{R}^-$ respectively.

In the particular case of Brownian motion, we have that h_λ^1 and h_λ^2 have the common value

$$h_\lambda(z) = \sqrt{\lambda} \int_0^1 \frac{z e^{zt}}{\sqrt{\log 1/t}}\, dt,$$

so that (4.6) becomes Theorem 1.2 of Chapter One. Finally we can use this symmetry relation to derive asymptotic results about the distribution of A_λ (which is independent of λ).

Theorem 4.2 *The distribution function F of A satisfies the asymptotic relation*

$$F(x) \sim \frac{2\sqrt{x}}{\pi\sqrt{(\log 1/x)}} \qquad \textit{as } x \downarrow 0. \tag{4.7}$$

And the moments (μ_n) *of* A *satisfy*

$$\mu_n \sim (\pi n \log n)^{-\frac{1}{2}} \qquad \text{as } n \to \infty. \tag{4.8}$$

(Where the symbol $\sim$ indicates that the ratio of the sides converges to 1.)

Proof of Theorem 4.2 Theorem 1.1 of Chapter One. $\square$

We can also perform these calculations for processes which can be written as deterministic time/scale-changes of Brownian motion, such as the Brownian bridge and Ornstein-Uhlenbeck processes. Let us recall that the Brownian bridge X_t can be defined either as the unique solution to the stochastic differential equation (SDE)

$$dX_t = dB_t - \frac{X_t}{1-t}\,dt, \quad X_0 = 0, \tag{4.9}$$

or explicitly as a process with the same distribution as

$$X_t = (1-t)B\big(t/(1-t)\big), \quad t \in [0,1], \tag{4.10}$$

where B is a Brownian motion. The Ornstein-Uhlenbeck (OU) process, Y_t, starting at 0, is governed by the SDE

$$dY_t = dB_t - \tfrac{1}{2}\gamma Y_t\,dt, \tag{4.11}$$

but can also be expressed as

$$Y_t = e^{-\gamma t/2} B\big((e^{\gamma t} - 1)/\gamma\big), \tag{4.12}$$

where B is again a Brownian motion. Note further that the Cesàro average of the Brownian bridge can be written as

$$\int_0^1 I_{(0,\infty)}(X_t)\,dt = \int_0^\infty \frac{1}{(1+t)^2} I_{(0,\infty)}(B_t)\,dt,$$

and the Abel averages of the OU process can be written as

$$\int_0^\infty \lambda e^{-\lambda t} I_{(0,\infty)}(Y_t)\,dt \overset{\mathcal{D}}{=} \begin{cases} \displaystyle\int_0^\infty \frac{\lambda/\gamma}{(1+t)^{1+\lambda/\gamma}} I_{(0,\infty)}(B_t)\,dt & \text{if } \gamma > 0, \\ \displaystyle\int_0^1 \frac{\lambda}{|\gamma|}(1-t)^{\lambda/|\gamma|-1} I_{(0,\infty)}(B_t)\,dt & \text{if } \gamma < 0. \end{cases}$$

We observe both that the Abel average of a positive recurrent OU process at discount rate $\lambda = \gamma$ has the same distribution as the Cesàro average of the Brownian bridge, and that the Abel average of a transient OU process at discount rate $\lambda = |\gamma|$ has the same distribution as the Cesàro average of Brownian motion.

Theorem 4.3 *(1) Let F_β be the distribution of the average*

$$A_\beta := \int_0^\infty \frac{1}{\beta(1+t)^{1+1/\beta}} I_{(0,\infty)}(B_t)\,dt, \quad (\beta > 0) \tag{4.13}$$

then $F_\beta(x) \sim c_\beta x^{\frac{1}{2}(1+\beta)}$ as $x \downarrow 0$, for some positive constant c_β.

(2) Let $F_{-\beta}$ be the distribution of the average

$$A_{-\beta} := \int_0^1 \beta^{-1}(1-t)^{(1/\beta)-1} I_{(0,\infty)}(B_t)\,dt, \quad (\beta > 0) \tag{4.14}$$

then $F_{-\beta}(x) \sim 2\sqrt{\beta x}/\pi$ as $x \downarrow 0$.

Note: The family of distributions (F_β) is actually continuous in β at 0, with F_0 being the distribution of the exponential discount of Brownian motion itself.

Proof of Theorem 4.3 Following the approach of Chapter One, we can bound A_β above and below by expressions involving the Cesàro average of Brownian motion itself,

$$C_t := \frac{1}{t}\int_0^t I_{(0,\infty)}(B_s)\,ds,$$

which we know to have the arc-sine distribution $F_{-1}(x) = \mathbb{P}(C_t \leqslant x) = \frac{2}{\pi}\sin^{-1}\sqrt{x}$. These bounds are summarized in the following lemma:

Lemma 4.4 *With F_β defined as above*

$$\begin{array}{rcccccc}
\frac{2}{\pi}(1+1/\beta)^{-\frac{1}{2}(1+\beta)} & \leqslant & \varliminf\limits_{x\downarrow 0} \frac{F_\beta(x)}{x^{\frac{1}{2}(1+\beta)}} & \leqslant & \varlimsup\limits_{x\downarrow 0} \frac{F_\beta(x)}{\sqrt{x}} & \leqslant & \frac{2}{\pi}(1+\beta)^{\frac{1}{2}(1+1/\beta)}, \\
\frac{2\sqrt{\beta}}{\pi} & \leqslant & \varliminf\limits_{x\downarrow 0} \frac{F_{-\beta}(x)}{\sqrt{x}} & \leqslant & \varlimsup\limits_{x\downarrow 0} \frac{F_{-\beta}(x)}{\sqrt{x}} & \leqslant & \frac{2}{\pi}(1-\beta)^{-\frac{1}{2}(1/\beta-1)}, \\
\frac{2}{\pi}(1-1/\beta)^{-\frac{1}{2}(\beta-1)} & \leqslant & \varliminf\limits_{x\downarrow 0} \frac{F_{-\beta}(x)}{\sqrt{x}} & \leqslant & \varlimsup\limits_{x\downarrow 0} \frac{F_{-\beta}(x)}{\sqrt{x}} & \leqslant & \frac{2\sqrt{\beta}}{\pi},
\end{array}$$

where the first line holds for all positive β, the second for all β in $(0,1)$, and the third for all β in $(1,\infty)$. Here $\varliminf$ and $\varlimsup$ are used to denote lim inf *and* lim sup *respectively.*

Proof of Lemma 4.4 Fix β positive. For every t,

$$A_\beta \geqslant \frac{1}{\beta(1+t)^{1+1/\beta}} \int_0^t I_{(0,\infty)}(B_s)\,ds = \frac{tC_t}{\beta(1+t)^{1+1/\beta}},$$

so $F_\beta(x) = \mathbb{P}(A_\beta \leqslant x) \leqslant \mathbb{P}(C_t \leqslant \beta x(1+t)^{1+1/\beta}/t) = \frac{2}{\pi}\sin^{-1}\sqrt{\beta x(1+t)^{1+1/\beta}/t}$. The minimum value of $\beta(1+t)^{1+1/\beta}/t$ is achieved at $t = \beta$, and the upper bound for F_β follows.

For every t,

$$A_\beta \leqslant \beta^{-1} \int_0^t I_{(0,\infty)}(B_s)\,ds + \frac{1}{(1+t)^{1/\beta}} = \beta^{-1} t C_t + \frac{1}{(1+t)^{1/\beta}},$$

so $F_\beta(x) = \mathbb{P}(A_\beta \leqslant x) \geqslant \mathbb{P}(C_t \leqslant \beta(x - (1+t)^{-1/\beta})/t)$. The maximum value of $\beta(x-(1+t)^{-1/\beta})/t$ is asymptotically the same as $(x/(1+1/\beta))^{1+\beta}$, whence the lower bound for F_β. The other two pairs of bounds are similar. $\square$

The corresponding excursion length measures for the diffusions are

$$\mu[t,\infty] = t^{-\frac{1}{2}} \qquad \text{Brownian motion,}$$
$$\mu_\gamma[t,\infty] = \left(\frac{e^{\gamma t}-1}{\gamma}\right)^{-\frac{1}{2}} \qquad \text{OU process.}$$

Hence

$$h_\beta(z) = \int_0^1 z e^{zt} \mu_{(\gamma=\beta\lambda)}[\lambda^{-1}\log(1/t),\infty]\,dt = \begin{cases} \displaystyle\int_0^1 \frac{z e^{zt}\,dt}{\sqrt{t^{-\beta}-1}} & \text{if } \beta > 0, \\ \displaystyle\int_0^1 \frac{z e^{zt}\,dt}{\sqrt{1-t^{-\beta}}} & \text{if } \beta < 0. \end{cases} \tag{4.15}$$

We can use this to verify that the Cesàro average of a bridge has the uniform distribution merely by checking that the α^{n+1} components of $\mathbb{E}h_1(\alpha X)$ and $-e^\alpha \mathbb{E}h_1(-\alpha X)$ agree for each $n \geqslant 0$, where X is uniform. That is, confirming that

$$\int_0^1 \int_0^1 \frac{x^{n+1}t^n}{\sqrt{t^{-1}-1}}\,dx\,dt = \int_0^1 \int_0^1 \frac{x(1-xt)^n}{\sqrt{t^{-1}-1}}\,dx\,dt.$$

Generally, we can investigate the behaviour of F_β near 0, but we need to discover the asymptotics of the functions h_β, which is done in the subsequent lemma:

Lemma 4.5 *With h_β as defined above*

$$h_\beta(z) \sim \sqrt{\frac{\pi z}{|\beta|}} e^z \qquad \text{as } z \to +\infty,$$
$$-h_\beta(-z) \sim \begin{cases} \Gamma(1+\beta/2) z^{-\beta/2} & \text{if } \beta > 0 \\ 1 & \text{if } \beta < 0 \end{cases} \qquad \text{as } z \to +\infty.$$

Proof of Lemma 4.5 Firstly

$$h_\beta(z) = \int_0^1 \frac{z e^{zt}\, dt}{\sqrt{|t^{-\beta} - 1|}} = \sqrt{z} e^z \int_0^z \frac{e^{-u}\, du}{\sqrt{z|(1-u/z)^{-\beta} - 1|}}.$$

Now $z|(1-u/z)^{-\beta} - 1| \geqslant (1 \wedge |\beta|)u$, and the left-hand side of this inequality tends to $|\beta|u$ as z goes to infinity. Hence by the Dominated Convergence theorem, we have the first line of the result.

Secondly, if β is positive, then

$$-h_\beta(-z) = z^{-\beta/2} \int_0^{\sqrt{z}} \frac{u^{\beta/2} e^{-u}\, du}{\sqrt{1-(u/z)^\beta}} + \int_{1/\sqrt{z}}^1 \frac{z e^{-zt}}{\sqrt{t^{-\beta} - 1}}.$$

The latter of the right-hand side terms is bounded by $z e^{-\sqrt{z}} \int_0^1 (t^{-\beta} - 1)^{-\frac{1}{2}}\, dt$, which is of smaller order than $z^{-\beta/2}$ as z goes to infinity. The integrand of the former term is dominated by $u^{\beta/2} e^{-u} (1 - z^{-\beta/2})^{-\frac{1}{2}}$ and tends to $u^{\beta/2} e^{-u}$ as z goes to infinity.

If β is negative, then we can similarly decompose $-h_\beta(-z)$ as

$$-h_\beta(-z) = \int_0^{\sqrt{z}} \frac{e^{-u}\, du}{\sqrt{1-(u/z)^{|\beta|}}} + \int_{1/\sqrt{z}}^1 \frac{z e^{-zt}}{\sqrt{1-t^{|\beta|}}}\, dt,$$

whence the result. □

We see from (4.6) that $e^{-\alpha}\mathbb{E}h_\beta(\alpha A_\beta) = -\mathbb{E}h_\beta(-\alpha A_\beta)$. We can prove in exactly the same way as in Chapter One that

$$\mathbb{E}h_\beta(\alpha A_\beta) \sim \sqrt{\pi\alpha/|\beta|}\, e^\alpha \mathbb{E}(e^{-\alpha A_\beta}), \quad \text{as } \alpha \to \infty. \tag{4.16}$$

If β is negative then $-h_\beta(-\alpha A_\beta)$ tends to 1 as α goes to infinity, and as h_β is bounded on $\mathbb{R}^-$, we can deduce that $-\mathbb{E}h_\beta(-\alpha A_\beta) \to 1$. Hence

$$\begin{aligned} \mathbb{E}(e^{-\alpha A_\beta}) &\sim \sqrt{\frac{|\beta|}{\pi\alpha}} \quad \text{as } \alpha \to \infty, \\ \text{and} \quad F_\beta(x) &\sim \frac{2}{\pi}\sqrt{|\beta| x} \quad \text{as } x \downarrow 0. \end{aligned} \tag{4.17}$$

Here we are using a Tauberian theorem adapted from (§8, 5.3) of Widder [38], which says that for a non-negative random variable X with distribution function F_X,

$$\mathbb{E}e^{-\alpha X} \sim \frac{k}{\alpha^\gamma} \quad \text{as } \alpha \to \infty \quad \Rightarrow \quad F_X(x) \sim \frac{kx^\gamma}{\Gamma(\gamma+1)} \quad \text{as } x \downarrow 0, \tag{4.18}$$

for positive constants γ and k.

For β positive, life is harder. Set I to be the set of 'good' values of β. That is $I := \{\beta > 0 : F_\beta(x) \sim c_\beta x^{\frac{1}{2}(1+\beta)}\}$.

Lemma 4.6 *Fix β positive. If $F_\beta(x) \leqslant kx^\gamma$, for some $\gamma > \beta/2$, and some $k > 0$, then β is in I.*

Proof of Lemma 4.6 For any positive ϵ, choose K large enough so that

$$|\Gamma(1+\beta/2) + z^{\beta/2} h_\beta(-z)| < \epsilon, \quad (z \geqslant K).$$

Then

$$\begin{aligned}
-\mathbb{E}h_\beta(-\alpha A_\beta) &= \mathbb{E}(-h_\beta(-\alpha A_\beta); A_\beta < K/\alpha) + \mathbb{E}(-h_\beta(-\alpha A_\beta); A_\beta \geqslant K/\alpha) \\
&=: E_1(\alpha) + E_2(\alpha).
\end{aligned}$$

Now $0 \leqslant E_1(\alpha) \leqslant (\sup -h_\beta(-x))F_\beta(K/\alpha) \leqslant c\alpha^{-\gamma}$. Also

$$\begin{aligned}
\mathbb{E}(A_\beta^{-\beta/2}) = \int_0^1 x^{-\beta/2}\, dF_\beta(x) &= \left[x^{-\beta/2}F_\beta(x)\right]_0^1 + \tfrac{1}{2}\beta \int_0^1 x^{-(1+\beta/2)} F_\beta(x)\, dx \\
&\leqslant (1-0) + \tfrac{1}{2}k\beta \int_0^1 x^{(\gamma-\beta/2)-1}\, dx < \infty.
\end{aligned}$$

So hence $E_2(\alpha)$ looks like $\alpha^{-\beta/2}(\Gamma(1+\beta/2) \pm \epsilon)\mathbb{E}(A_\beta^{-\beta/2})$, as α gets large, and thus E_2 dominates E_1. So we have that

$$\mathbb{E} - h_\beta(-\alpha A_\beta) \sim \Gamma(1+\beta/2)\mathbb{E}(A_\beta^{-\beta/2})\alpha^{-\beta/2}. \tag{4.19}$$

And thus

$$\begin{aligned}
\mathbb{E}(e^{-\alpha A_\beta}) &\sim \sqrt{\frac{\beta}{\pi}}\Gamma(1+\beta/2)\mathbb{E}(A_\beta^{-\beta/2})\alpha^{-\frac{1}{2}(1+\beta)} \quad \text{as } \alpha \to \infty, \\
\text{and so by (4.18)} \quad F_\beta(x) &\sim \frac{2}{\pi}\sqrt{\beta}\Gamma(1+\beta/2)\mathbb{E}(A_\beta^{-\beta/2})x^{\frac{1}{2}(1+\beta)} \quad \text{as } x \downarrow 0,
\end{aligned} \tag{4.20}$$

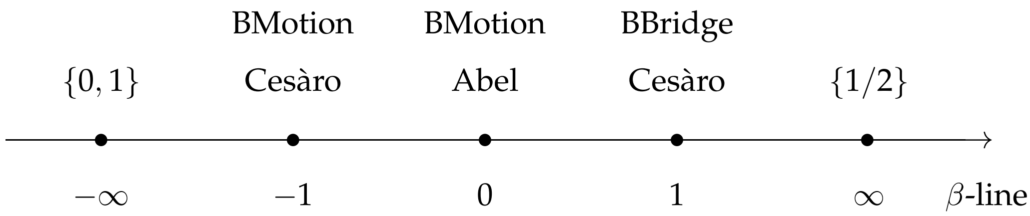
that is, that β is in I. $\square$

We can now complete the proof of Theorem 4.3. From Lemma 4.4, we have that F_β is dominated by $x^{\frac{1}{2}}$, so Lemma 4.6 tells us that the interval $(0,1)$ is contained in I. Now for $\delta < \beta$,

$$\frac{1}{\beta(1+t)^{1+1/\beta}} \geqslant \frac{\delta}{\beta}\frac{1}{\delta(1+t)^{1+1/\delta}}, \quad \text{for all } t.$$

So $A_\beta \geqslant (\delta/\beta)A_\delta$, and hence $F_\beta(x) \leqslant F_\delta(x\beta/\delta)$. Now if δ is already in I and β is in the open interval $(\delta, 1+\delta)$, then F_β is dominated by $F_\delta \approx x^{\frac{1}{2}(1+\delta)}$, and so β is in I by Lemma 4.6. Thus the theorem is proved. $\square$

As we have seen before, A_0 is the distribution of the Abel average of Brownian motion, A_{-1} is the distribution of the Cesàro average of Brownian motion (the arc-sine distribution), and A_1 is the distribution of the Cesàro average of the Brownian bridge (the uniform distribution). For an Ornstein-Uhlenbeck process with parameter γ as in (4.11), its Abel average at rate λ has the distribution of A_β where $\beta = \gamma/\lambda$. As λ goes to 0, or equivalently as $|\gamma|$ goes to infinity, A_β tends either to the symmetric distribution on the end-points $\{0,1\}$ or to the point $\frac{1}{2}$, according to whether the process is transient or recurrent. We can summarize this information with the aid of the following diagram of the 'β-spectrum'.



We note that amongst symmetric distributions on $[0,1]$, $A_{-\infty}$ has the greatest variance, and A_∞ has the least. We shall see evidence shortly that the variance is decreasing in β. By equating coefficients of α in $\mathbb{E}h_\beta(\alpha A_\beta) = -e^\alpha \mathbb{E}h_\beta(-\alpha A_\beta)$, we see that

$$\int_0^1 f_\beta(x) \int_0^1 \frac{x(xt)^n}{\sqrt{|t^{-\beta}-1|}}\,dx\,dt = \int_0^1 f_\beta(x) \int_0^1 \frac{x(1-xt)^n}{\sqrt{|t^{-\beta}-1|}}\,dx\,dt, \tag{4.21}$$

where f_β is the density of F_β. Now we can perform the t-integration as follows

$$\int_0^1 \frac{t^{r-1}\,dt}{\sqrt{|t^{-\beta}-1|}} = \begin{cases} \dfrac{1}{\beta}\displaystyle\int_0^1 u^{r/\beta-\frac{1}{2}}(1-u)^{-\frac{1}{2}}\,du = \dfrac{\Gamma(\frac{r}{\beta}+\frac{1}{2})\Gamma(\frac{1}{2})}{r\Gamma(\frac{r}{\beta})} & \text{if } \beta > 0, \\ \dfrac{1}{|\beta|}\displaystyle\int_0^1 u^{r/|\beta|-1}(1-u)^{-\frac{1}{2}}\,du = \dfrac{\Gamma(\frac{r}{|\beta|}+1)\Gamma(\frac{1}{2})}{r\Gamma(\frac{r}{|\beta|}+\frac{1}{2})} & \text{if } \beta < 0. \end{cases} \tag{4.22}$$

Hence we can deduce a generalization of (1.17) of Chapter One.

Proposition 4.7 *Fix β real. Let μ_n be the nth moment of A_β. Then*

$$\mu_n = \sum_{k=0}^{n}\binom{n}{k}(-1)^k\mu_k \quad \textit{and} \quad M_n\mu_n = \sum_{k=1}^{n}\binom{n}{k}(-1)^{k+1}M_k\mu_k,$$

$$\textit{where} \quad M_n := \begin{cases} \frac{\Gamma(\frac{n}{\beta}+\frac{1}{2})}{\Gamma(\frac{n}{\beta})} & \textit{if } \beta > 0, \\ \frac{\Gamma(\frac{n}{|\beta|}+1)}{\Gamma(\frac{n}{|\beta|}+\frac{1}{2})} & \textit{if } \beta < 0. \end{cases}$$

Proof of Proposition 4.7 Follows immediately from the symmetry of A_β and from (4.21) and (4.22). □

This allows us to calculate the variance explicitly as

$$\mathrm{Var}(A_{-\beta}) = \frac{1}{4}\left(\frac{\Gamma(\frac{1}{\beta})\Gamma(\frac{2}{\beta}+\frac{1}{2})}{\Gamma(\frac{1}{\beta}+\frac{1}{2})\Gamma(\frac{2}{\beta})} - 1\right),$$

$$\mathrm{Var}(A_{\beta}) = \frac{1}{4}\left(\frac{2\Gamma(\frac{1}{\beta}+\frac{1}{2})\Gamma(\frac{2}{\beta})}{\Gamma(\frac{1}{\beta})\Gamma(\frac{2}{\beta}+\frac{1}{2})} - 1\right).$$

Were A_β to be a member of the beta family of distributions, with densities

$$f_{\beta_1,\beta_2}(x) = \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}x^{\beta_1-1}(1-x)^{\beta_2-1},$$

then the symmetry of A_β together with the asymptotics of Theorem 4.3 would give A_β the Beta$(\frac{1}{2}(1+\beta), \frac{1}{2}(1+\beta))$ distribution for positive β and the Beta$(\frac{1}{2}, \frac{1}{2})$ distribution (arc-sine) for negative β. Let us let B_β be this hypothesised distribution, that is it has the symmetric Beta-distribution with parameter $\frac{1}{2}$ for negative β and parameter $\frac{1}{2}(1+\beta)$ for positive β.

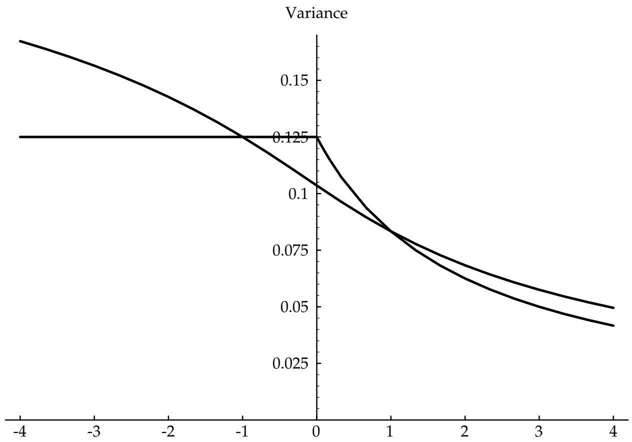


Figure 4.1 Graphs of $\mathrm{Var}(A_\beta)$ *(smooth) and* $\mathrm{Var}(B_\beta)$ *(crooked) against* β

Making a graph of the variances numerically (Figure 4.1, drawn by *Mathematica*) we can believe that although $\mathrm{Var}(A_\beta)$ is indeed decreasing in β, it only intersects with the hypothesized beta distribution variance at the points $\beta = \pm 1$, the two points which are already known, and at no others. Identification of the other distributions involved remains an open question.

4.3 Large Deviations

We now work with general discount shapes and positive recurrent processes. Let X be a stochastic process with state-space E and invariant distribution π. In the Cesàro case we would expect some sort of ergodic theorem such as

$$C_t(F) := \frac{1}{t}\int_0^t I_F(X_s)\,ds \to \pi(F) \qquad \text{as } t \to \infty, \tag{4.23}$$

for all measurable subsets F of E. Then C_t takes values in $M_1(E)$, the space of probability measures on E. We might also have a large-deviation result, which we can think of for the moment as the slogan

$$\text{“}\mathbb{P}(C_t \in H) \approx \exp\bigl(-t \inf_{\nu \in H} I(\nu)\bigr) \quad \text{as } t \to \infty\text{''} \qquad H \subset M_1(E),$$

for some rate function I, with $I(\pi) = 0$. The space $M_1(E)$ and the continuous bounded functions on E, $C_b(E)$ are in duality via the bracket, $\langle v, \nu\rangle = \int_E v(x)\nu(dx)$. A related slogan is that of the Laplace transform

$$\text{“}\mathbb{E}\exp\bigl(\langle v, C_t\rangle t\bigr) \approx \exp\bigl(t\delta(v)\bigr) \quad \text{as } t \to \infty\text{''} \qquad v \in C_b(E),$$

where δ and I are related by Legendre transformation (convex conjugation), in that

$$I(\nu) = \delta^*(\nu) := \sup_{v \in C_b(E)} \big(\langle v, \nu \rangle - \delta(v) \big), \tag{4.24}$$

$$\delta(v) = I^*(v) := \sup_{\nu \in M_1(E)} \big(\langle v, \nu \rangle - I(\nu) \big), \tag{4.25}$$

Our program will be to study, for a discount density m, the average

$$A_\lambda(F) := \int_0^\infty \lambda m(\lambda t) I_F(X_t)\, dt. \tag{4.26}$$

We will show that $A_\lambda \to \pi$ as λ goes to 0, and that the large-deviation principle holds with rate K whose Legendre transform η is given by the equation

$$\eta(v) = \int_0^\infty \delta\big(m(t)v\big)\, dt. \tag{4.27}$$

This is actually the same as the η-equation at (2.16) in Chapter Two (with discount $m_t = e^{-t}$) and at Theorem 3.3 (with discount $m_t = (1+\gamma)(1-t)^\gamma$), but (4.27) is a more natural formulation.

Standard set-up. Let X be an ergodic Feller-Dynkin Markov process on a locally compact Polish space E, with generator L. We define the Cesàro average C_t and the general average A_λ by (4.23) and (4.26) respectively. Then C_t and A_λ converge to π, the invariant distribution of X, with respect to the weak topology on $M_1(E)$, that is, in the sense of (4.23). A sufficient condition for the former limit is that, as in 8.11.2 of Bingham et al. [10], π is a limiting distribution of the transition semigroup (P_t). The latter limit follows from the former by a similar L^1-continuity argument to that which will be used in the proof of Proposition 4.8. Deuschel and Stroock [15] show that under an assumption of uniform ergodicity the large-deviation property holds for C_t with rate function I defined on $M_1(E)$. That is that

$$\begin{aligned} &\limsup_{t\to\infty} t^{-1} \log \mathbb{P}(C_t \in F) \leqslant - \inf_{\nu \in F} I(\nu), \\ \text{and} \qquad &\liminf_{t\to\infty} t^{-1} \log \mathbb{P}(C_t \in G) \geqslant - \inf_{\nu \in G} I(\nu), \end{aligned} \tag{4.28}$$

for F and G respectively closed and open subsets of $M_1(E)$. We learn from 4.2.17 of Deuschel and Stroock [15] that

$$\lim_{t\to\infty} t^{-1} \log \mathbb{E} \exp \int_0^t v(X_s)\, ds = \delta(v), \tag{4.29}$$

where δ and I are convex functions satisfying (4.24) and (4.25). Further there are, by 4.2.27 and 4.2.38 of Deuschel and Stroock [15], explicit expressions for δ and I as

$$\delta(v) = \lim_{t\to\infty} t^{-1} \log \|P_t^v\|_{\mathrm{op}}, \quad \text{where } P_t^v f(x) := \mathbb{E}_x\big(\exp(\textstyle\int_0^t v(X_s)\,ds) f(X_t)\big), \tag{4.30}$$

$$I(\nu) = \sup\left\{ -\int_E \frac{Lf(x)}{f(x)}\,\nu(dx) : f \geqslant 1,\ f \in \mathrm{Dom}(L) \right\}. \tag{4.31}$$

As in Chapter Two we shall be particularly interested in the case where X is a Markov chain on a finite state-space S with Q-matrix Q. Then $\delta(v) = \sup\{\mathrm{Re}(z) : z \in \mathrm{spect}(Q+V)\}$, where V denotes the diagonal matrix $\mathrm{diag}(v)$ and $\mathrm{spect}(\cdot)$ denotes spectrum (here the set of eigenvalues). This expression for δ also holds in the general Markov process setting, if the generator L is π-symmetric.

We begin by proving a result whose first part is similar to one remarked by Kifer [27] in the context of the large-deviations of the averages of dynamical systems, but it is the second part which will be more useful in our further work. In earlier Chapters we derived a differential equation by the self-similarity of discount shapes such as e^{-t}, but it is enough to study the shifts of the discount along the time-axis, which provides a useful one-dimensional parameterisation.

Proposition 4.8 *Suppose that X is an FD Markov process on a space E, with generator L, and m is any density on $[0,\infty)$, and for x in E and v in $C_b(E)$ the limit $\delta(v) = \lim_\lambda \lambda \log \mathbb{E}_x \exp \int_0^{1/\lambda} v(X_s)\,ds$ exists uniformly in x on E. If we define φ by*

$$\varphi(x,t,\lambda,v) = \mathbb{E}_x \exp \int_0^\infty \theta_t m(\lambda s) v(X_s)\,ds, \tag{4.32}$$

where θ_t is the shift operator $\theta_t f(s) = f(t+s)$, then

$$\lim_{\lambda\downarrow 0} \lambda \log \varphi(x,0,\lambda,v) = \eta(v) := \int_0^\infty \delta(m_t v)\,dt. \tag{4.33}$$

Further, $\varphi(\cdot,t,\lambda,v)$ is in the domain of L and $\varphi(x,\cdot,\lambda,v)$ is differentiable and

$$-\frac{\partial\varphi}{\partial t} = \lambda^{-1}(L + m_t V)\varphi. \tag{4.34}$$

Proof of Proposition 4.8 We prove the first part using continuity arguments. If we define

$$H(\lambda,x,\alpha) := \lambda \log \mathbb{E}_x \exp \int_0^{1/\lambda} \alpha v(X_s)\,ds, \tag{4.35}$$

then $\sup_x |H(\lambda, x, \alpha) - \delta(\alpha v)|$ goes to 0 as λ does. We start by proving the general average limit for m of the form

$$m = \sum_{i=1}^{n} c_i I(a_i, b_i),$$

where $\{(a_i, b_i)\}$ are disjoint intervals in $\mathbb{R}^+$ and $c_i > 0$. Set

$$Y_i^\lambda := \exp\left(\int_{a_i/\lambda}^{b_i/\lambda} c_i v(X_s)\, ds - \lambda^{-1}(b_i - a_i)\delta(c_i v)\right), \tag{4.36}$$

and define $y_i^\lambda(x) := \mathbb{E}_x Y_i^\lambda$. Then for λ sufficiently small $|\lambda \log y_i^\lambda(x)| < \epsilon$ uniformly in x. Thus

$$\begin{aligned}\frac{\mathbb{E}_x \exp \int_0^\infty m(\lambda t) v(X_t)\, dt}{\exp \lambda^{-1} \int_0^\infty \delta(m_t v)\, dt} &= \mathbb{E}_x(Y_1^\lambda \dots Y_n^\lambda) = \mathbb{E}_x(Y_1^\lambda \dots Y_{n-1}^\lambda y_n^\lambda(X_{a_n/\lambda})) \\ &\leqslant e^{\epsilon/\lambda}\mathbb{E}_x(Y_1^\lambda \dots Y_{n-1}^\lambda) \leqslant \dots \leqslant e^{n\epsilon/\lambda},\end{aligned}$$

and so we have the right upper bound. Similarly we have the lower bound.

Let us now define $I_\lambda(m) := \int_0^\infty m(\lambda t) v(X_t)\, dt$, $L_\lambda(m) := \lambda \log \mathbb{E} \exp I_\lambda(m)$, and $J(m) := \int_0^\infty \delta(m_t v)\, dt$. Then $\lambda |I_\lambda(m_1) - I_\lambda(m_2)|$ is bounded by $\|v\|_\infty \|m_1 - m_2\|_1$ uniformly in ω, and hence $|L_\lambda(m_1) - L_\lambda(m_2)|$ has the same bound. Because $|\delta(v_1) - \delta(v_2)| \leqslant |v_1 - v_2|$, the same bound also dominates $|J(m_1) - J(m_2)|$. This L^1-continuity and (careful) application of monotone class theorems let us generalize firstly to all bounded m of compact support and then to all m in $L^1(\mathbb{R}^+)$.

For the differential equation, we need only apply the Feynman-Kač formula to the space-time process $Y_t := (X_t, \tau_t)$, where $\tau_t = \tau_0 + t$, which has generator $L + \partial_t$. Taking $\lambda = 1$ for simplicity, for v in $C_b(E)$, we define v_Y on $C_b(E \times \mathbb{R}^+)$ by $v_Y(x, t) := m_t v(x)$, and

$$A_t := \int_0^t v_Y(Y_s)\, ds = \int_0^t \theta_{\tau_0} m(s) v(X_s)\, ds. \tag{4.37}$$

Without loss of generality we can assume that v is non-negative, because if (4.33) and (4.34) hold for some v, then they hold for all vectors of the form $v + \alpha\mathbf{1}$, where $\mathbf{1}$ is the constant vector $(1, 1, \dots, 1)$. This shifting identity follows from the fact that $\delta(v + \alpha\mathbf{1}) = \delta(v) + \alpha$, a property which η inherits. Then the semigroup P^v defined by

$$P_t^v f(y) = \mathbb{E}_{Y_0 = y}\big(e^{A_t} f(Y_t)\big)$$

has generator $L^v := L + \partial_t + m_t V$, as seen in, for example, III.39 of Williams [39]. Then if we set $\varphi((x,t)) := \varphi(x,t,1,v)$, which is continuous in t, we have that

$$P_t^v \varphi(Y_0) = \mathbb{E}_{Y_0}\left(e^{A_t}\varphi(Y_t)\right) = \mathbb{E}_{Y_0}\left(\mathbb{E}(\exp A_\infty | \mathcal{F}_t)\right) = \varphi(Y_0). \tag{4.38}$$

Thus $t^{-1}(P_t^v - I)\varphi = 0$, implying that φ is in the domain of L^v and is annihilated by it. The equation $L^v\varphi = 0$ is exactly (4.34). □

We note that in the case of X a standard Brownian motion and the exponential discount $m_t = e^{-t}$ and $v(x) = I(x > 0)$, then (4.34) is equation (1.22) of Chapter One.

Corollary 4.9 *Suppose that X is an irreducible Markov chain on a finite state-space S, with Q-matrix Q, and m is any density on $[0,\infty)$, and A_λ is defined by (4.26), then the large-deviation property analogue of (4.28) holds for A_λ with rate function K,*

$$\limsup_{\lambda\to 0} \lambda \log \mathbb{P}(A_\lambda \in F) \leqslant -\inf_{\nu\in F} K(\nu), \quad \textit{and} \quad \liminf_{\lambda\to 0} \lambda \log \mathbb{P}(A_\lambda \in G) \geqslant -\inf_{\nu\in G} K(\nu), \tag{4.39}$$

for F and G respectively closed and open subsets of M. The rate function K relates to the η of (4.33) through the following equations:

$$K(x) = \sup_{v\in\mathbb{R}^S} \langle v, x\rangle - \eta(v), \tag{4.40}$$

$$\eta(v) = \sup_{x\in M} \langle v, x\rangle - K(x), \tag{4.41}$$

where $M := M_1(S) = \{(x_i)_{i=1}^n : \sum_i x_i = 1,\ x \geqslant 0\}$.

Proof of Corollary 4.9 We are in the context of Proposition 4.8 because X will satisfy condition (Ũ) of 4.2.7 of Deuschel and Stroock [15], which is sufficient for the limit δ to exist as required by the theorem. The large-deviation property and (4.40) come from theorem II.2 of Ellis [18]. In his language, t is our v, Y_n is our A_λ, $c_n(\cdot)$ is our $\lambda \log \varphi(x, 0, \lambda, \cdot)$, and $c(\cdot)$ is our $\eta(\cdot)$. As η is defined and differentiable on the whole of $\mathbb{R}^S$, it meets Ellis' 'steep' hypothesis. From (4.33), η inherits the (strict) convexity and differentiability of δ, which gives (4.41). □

We complete this section with a pair of results about the large-deviation rate function K. The former of these is in the spirit of Proposition 2.4 of Chapter Two and identifies the points where the various suprema in Legendre transforms (4.40) and (4.41) are achieved. This leads to central limit results and the major result of the next Section.

Proposition 4.10 *Under the conditions of Corollary 4.9, K is finite, twice differentiable and strictly convex on* $\mathrm{Int}(M)$*, and the supremum of (4.40) is attained uniquely (up to multiples of* $\mathbf{1}$*) at $v = \nabla K(x)$, and the supremum of (4.41) is attained uniquely at $x = \nabla\eta(v)$.*

Proof of Proposition 4.10 It is immediate from its definition that $\delta(v)/\|v\|_\infty \to 1$ as $\|v\|_\infty \to \infty$ with $v \geqslant 0$. But as also $|\delta(v)| \leqslant \|v\|_\infty$, the Dominated Convergence theorem gives us that $\eta(v)/\|v\|_\infty$ goes to 1 as well. Take $x \in \mathrm{Int}(M)$ and suppose there exists a sequence of vectors (v_n) such that

$$\langle v_n, x\rangle - \eta(v_n) \to \infty.$$

Without loss of generality we can replace (v_n) by $(v_n - (\min_i v_n(i))\mathbf{1})$, because $\eta(v + \alpha\mathbf{1}) = \eta(v) + \alpha$, and thus assume that the (v_n) are positive, with at least one zero co-ordinate. The sequence must still get infinitely large, but

$$\langle v_n, x\rangle - \eta(v_n) \leqslant \|v\|_\infty \Big((1 - \min_i x_i) - \eta(v_n)/\|v_n\|_\infty\Big), \tag{4.42}$$

which is large and negative for large n, contradicting our supposition of $K(x) = \infty$.

As remarked in Corollary 4.9, η inherits the smoothness and the strict convexity on $\mathbf{1}^\perp$ of δ. Its continuity means that the supremum must be attained at some finite point $\hat{v}(x)$, and the convexity gives the uniqueness. The differentiability shows that the maximizing $\hat{v}$ will be the solution of $\nabla\eta(v) = x$. We can expand $\hat{v}$ around an x as $\hat{v}(x+\epsilon) = \hat{v}(x) + H_\eta^{-1}\epsilon$ to see that

$$K(x+\epsilon) = K(x) + \langle\hat{v}(x), \epsilon\rangle + \tfrac{1}{2}\epsilon^\top H_\eta^{-1}\epsilon + o(\epsilon^2). \tag{4.43}$$

Thus K is twice differentiable, $\nabla K(x) - \hat{v}(x)$ is a multiple of $\mathbf{1}$, and K is locally (and hence globally) strictly convex. (Technical note: we are regarding H_η, the Hessian of η, as an automorphism of $\mathbf{1}^\perp$.) By the above $x = \nabla\eta(v)$ is a solution of (4.41), and the strict convexity of K shows it is unique. □

In the simple example studied in Section 2.3 of Chapter Two, the rate function was calculated exactly as $K(x) = \sum \pi_i \log(\pi_i/x_i)$, which is infinite on the boundary of M, whilst the Cesàro rate function I is finite everywhere. Note that we can see that I is finite in the general Corollary 4.9 situation by considering equation (4.31). We shall have further remarks about this example in the next section, but for the moment we derive a necessary and sufficient condition for K to be everywhere finite or infinite.

Proposition 4.11 *Under the conditions of Corollary 4.9, the rate function K is either everywhere finite or everywhere infinite on the boundary of M according as to whether the support of the discount function m is of finite or infinite (Lebesgue) length.*

Proof of Proposition 4.11 Firstly let us define V_+ to be the space of elements of $(\mathbb{R}^+)^n$ which have at least one zero component. We note that $\lambda V_+ = V_+$ for any positive λ, which is a feature we shall use later. For x in Int(M), we take v_x to be the unique choice in V_+ of the $v = \nabla K(x)$ in Proposition 4.10. In fact the pair $(\nabla K, \nabla \eta)$ represents a homeomorphism between Int(M) and V_+. Then $x = \nabla \eta(v_x)$, so by taking the gradient of (4.33), we can write x as

$$x = \int_0^\infty m_t \nabla \delta(m_t v_x)\, dt.$$

Then because $\langle v, \nabla\delta(v)\rangle = \delta(v) + I(\nabla\delta(v))$ for all v in $\mathbb{R}^n$,

$$\langle v_x, x\rangle = \int_0^\infty \Big(\delta(m_t v_x) + I\big(\nabla\delta(m_t v_x)\big)\Big)\, dt = \eta(v_x) + \int_0^\infty I\big(\nabla\delta(m_t v_x)\big)\, dt.$$

As v_x is the optimal v in (4.40), we can express $K(x)$ as

$$K(x) = \int_0^\infty I\big(\nabla\delta(m_t v_x)\big)\, dt. \tag{4.44}$$

Thus, for an upper bound,

$$K(x) \leqslant \sup_{y\in M} I(y) \int_0^\infty I_{\{m_t>0\}}\, dt = \sup_{y\in M} I(y)\ \mathrm{Leb\,supp}(m),$$

and so K is bounded on all of M if the support of m, supp(m), is compact. The rate function I is only 0 in M at π, and $\nabla\delta$ only takes the value π in V_+ at 0. Thus from (4.44) we have the lower bound

$$K(x) \geqslant \mathrm{Leb}\big\{t : m_t \geqslant \|v_x\|_\infty^{-1}\big\}\ \inf\big\{I\big(\nabla\delta(v)\big) : v \in V_+,\ \|v\|_\infty \geqslant 1\big\} > 0.$$

Now as x tends towards ∂M, the boundary of M, the vector v_x tends to infinity in V_+. So if m has unbounded support, then $K(x)$ tends to infinity as x tends to ∂M. The intuition, of course, is that X can with positive probability avoid hitting a certain state for all times in a finite length set but not for all times in an infinite length set. □

4.4 More exact results for Markov chains

Our aim is to obtain a sharper version of (4.33) for finite Markov chains, and then to derive more terms of the asymptotic expansion of the density of A_λ.

The initial case studied in Chapter Two was of a symmetrizable (reversible) Markov chain and a smooth discount density m. It turns out that m need only be of bounded variation (see below), but for technical ease we shall give the proof first in the case where m is also absolutely continuous.

More interestingly, the symmetrizability is seen now to have only been needed to make one of the eigenvalues of Q real and its corresponding eigenvector orthogonal to the others. This in fact happens automatically because every (non-diagonal) element of Q is non-negative (we say that Q is *essentially non-negative*). The following theorem collects all the facts about non-negative matrices that we will need.

Theorem 4.12 *Let R be an essentially non-negative $n \times n$ matrix. Let δ be its principal eigenvalue (the one with greatest real part). Then δ is itself real, and its corresponding eigenvector is non-negative and no other is positive. If, in addition, R is irreducible (in the stochastic sense), then δ is simple, its eigenvector is strictly positive and no other is non-negative, and there exists a real diagonal matrix F with positive elements such that $S := \delta I - F^{-1}RF$ has a simple eigenvalue zero, with an orthogonal eigenprojection P, and that, for some positive σ*

$$\langle Sx, x\rangle \geqslant \sigma \|(I-P)x\|^2, \qquad \textit{for all } x \in \mathbb{R}^n. \tag{4.45}$$

(Where $\langle\ ,\ \rangle$ and $\|\ \|$ are the standard inner product and its norm on $\mathbb{R}^n$, and an orthogonal projection P satisfies $P^\top = P^2 = P$.)

Proof of Theorem 4.12 For the first parts see the Perron-Frobenius theorem in, for example, theorem 1.5 of Seneta [34] or theorems I.7.5 and I.7.10 of Kato [24]. For the

existence of F see Theorem 3.2, which itself is adapted from theorem I.7.13 of Kato [24].
□

We recall that the *variation* of a measurable function $x : [0,\infty) \to \mathbb{R}$ on $[a,b]$ is defined as

$$V_x(a,b) := \sup \sum_{i=1}^{n} |x(t_i) - x(t_{i-1})|, \tag{4.46}$$

where the supremum is taken over all partitions: $a = t_0 < t_1 < \ldots < t_n = b$ of $[a,b]$. We say that x is of finite variation (FV) if $V_x(0,t)$ is finite for all t, and that x is of bounded variation (BV) if $V_x(0,\infty) := \lim_{t\to\infty} V_x(0,t)$ is finite. An absolutely continuous BV function is the partial integral of a function in $L^1(0,\infty)$. (My thanks to James Norris for correcting a previous mis-statement here.)

We can now begin by strengthening (4.33):

Theorem 4.13 *Let X be an honest irreducible Markov chain on a finite set S, with Q-matrix Q. Let m be a non-negative absolutely continuous density on $[0,\infty)$ of bounded variation. Then, if X starts in state i and v is in $\mathbb{R}^S$,*

$$\mathbb{E}_i \exp \int_0^\infty m(\lambda t) v(X_t)\, dt = e^{\eta(v)/\lambda}\big(w_i(m_0 v) + o(1)\big), \tag{4.47}$$

where $\eta(v)$ is as in (4.33) and $w(v)$ is the positive eigenvector of $Q + V$ and $o(1)$ tends to 0 locally uniformly in v as λ goes to 0.

Proof of Theorem 4.13

The chain has an invariant distribution π, but we do not need to assume that Q is π-symmetric. We will aim to get a uniform bound for all v in some compact subset V_K of $\mathbb{R}^S$, and for a fixed m such that $V_m(0,\infty) \leqslant K_V$. Since Proposition 4.8 gives us the asymptotic exponential size of φ, it is sensible to discount it by the same, by defining

$$\psi_i(t,\lambda,v) := \exp\left(-\lambda^{-1}\int_0^\infty \delta(\theta_t m_s v)\, ds\right)\varphi(i,t,\lambda,v). \tag{4.48}$$

Then ψ satisfies the vector differential equation transformed from (4.34)

$$\partial_t \psi = \lambda^{-1} R(m_t v)\psi, \qquad \psi(\infty,\lambda) = \mathbf{1}, \tag{4.49}$$

where $R(v)$ is $\delta(v)I - (Q+V)$ which has a simple eigenvalue at 0, and all its other eigenvalues have positive real part. From Theorem 4.12, there exists a real diagonal matrix $F(v)$ with positive elements, such that $S(v) := F^{-1}(v)R(v)F(v)$ has an orthogonal eigenprojection $P(v)$ onto the space spanned by the strictly positive eigenvector $y(v)$ corresponding to the eigenvalue zero. Further there exists a positive $\sigma(v)$ such that (4.45) holds, that is

$$\langle S(v)x, x\rangle \geqslant \sigma(v)\|(I - P(v))x\|^2 \qquad \text{for all } x, \tag{4.50}$$

where $\langle\ ,\ \rangle$ and $\|\ \|$ are the standard inner product and its norm on $\mathbb{R}^S$. Kato [24], or otherwise, tells us that R, S, F, P, y and σ are smooth in v with bounded derivatives on V_K. Let $\sigma_0 := \inf_{v\in V_K} \sigma(v)$, which is positive. We now fix v, although our bounds will still be uniform, and write R_α for $R(\alpha v)$, and so on. We can choose the normalisation of F uniquely such that $F_0 = \operatorname{diag}(\pi_i^{-1/2})$ and $P_\alpha F_\alpha^{-1} F'_\alpha y_\alpha = 0$, and by choosing $\|y(v)\| = 1$ we ensure that $P_\alpha y'_\alpha = 0$ and $y_0 = \sqrt{\pi}$.

As in Chapter Three, we change bases appropriately by defining

$$\chi(t,\lambda) := F^{-1}_{m(t)}\psi(t,\lambda). \tag{4.51}$$

The differential equation (4.49) now becomes

$$\partial_t\chi = \lambda^{-1}S_{m(t)}\chi + J_{m(t)}\chi m', \qquad \chi(\infty,\lambda) = \sqrt{\pi}, \tag{4.52}$$

where $J_\alpha := -F_\alpha^{-1}F'_\alpha$. Then by taking the inner product of (4.52) with χ we can produce a differential inequality in the norm of χ,

$$\tfrac{1}{2}\partial_t\|\chi\|_t^2 \geqslant 0 - K_1\|\chi\|_t^2|m'_t|, \qquad \text{so that} \qquad \|\chi\|_t \leqslant \exp\left(K_1\int_t^\infty |m'_s|\,ds\right), \tag{4.53}$$

where $K_1 := \sup\|J_\alpha\|$, the supremum taken over the range $\alpha \in [0, K_V]$ and $v \in V_K$. Whence we deduce that χ is uniformly bounded in t and λ by $K_\chi := \exp(K_1K_V)$. Now we split χ up according to the decomposition $I = P_\alpha \oplus (I - P_\alpha)$, and define $\chi_-(t) := (I - P_{m(t)})\chi_t$. We differentiate χ_-, using (4.52), to get

$$\partial_t\chi_- = \lambda^{-1}S_m\chi_- + (I - P_m)J_m\chi m' - P'_m\chi m', \qquad \chi_-(\infty,\lambda) = 0. \tag{4.54}$$

Taking the inner product of this with χ_- itself, we derive the inequality

$$\tfrac{1}{2}\partial_t\|\chi_-\|_t^2 \geqslant \lambda^{-1}\sigma_0\|\chi_-\|_t^2 - K_\chi(K_1+K_2)\|\chi_-\|_t|m_t'|, \tag{4.55}$$

where $K_2 := \sup\|P_\alpha'\|$, with the supremum taken over the same range as K_1. Which we can integrate to get the upper bound

$$\|\chi_-\|_t \leqslant (K_1+K_2)K_\chi\int_t^\infty e^{-\sigma_0(s-t)/\lambda}|m_s'|\,ds. \tag{4.56}$$

And so we see that $\chi_-(t,\lambda)$ tends to 0 as λ tends to 0 for all finite t, though note that the convergence is not necessarily uniform in t. Finally we consider the component of χ in the y_m direction, $\xi(t,\lambda) := \langle\chi(t,\lambda), y_{m(t)}\rangle$, which is governed by the differential equation obtained from (4.52)

$$\partial_t\xi_t = \langle\chi_-, Jy+y'\rangle m_t', \qquad \xi(\infty,\lambda)=1, \tag{4.57}$$

where we used the fact that $PJy = Py' = 0$. Then

$$|\xi(t,\lambda)-1| \leqslant (K_1+K_2)\int_t^\infty \|\chi_-(s,\lambda)\|\,|m_s'|\,ds, \tag{4.58}$$

which, by the Dominated Convergence theorem, tends to 0 uniformly in t as λ goes to 0. So

$$\varphi(i,0,\lambda,v) = \exp(\eta(v)/\lambda)\big((Fy)_i(m_0v)+o(1)\big), \tag{4.59}$$

and $F(v)y(v) = w(v)$, where $w(v)$ is the positive eigenvector of $Q+V$, with the normalisation that $w(0) = \mathbf{1}$ and $\partial_\alpha w(\alpha v)$ is orthogonal to the positive eigenvector of $Q^\top + \alpha V$. In the case where Q is π-symmetrizable ($\pi_i q_{ij} = \pi_j q_{ji}$), then the normalisation condition becomes $\|w\|_\pi = 1$, where $\|v\|_\pi^2 = \sum \pi_i v_i^2$. $\square$

The next theorem removes the restriction that m need be continuous, but takes us into the technicalities of FV functions. The casual reader can pass this by without disadvantage.

An FV function can be written as the difference of two increasing functions, that is

$$x_t = x_0 + x_t^+ - x_t^-,$$

where $\quad x_t^+ := \tfrac{1}{2}(x_t + V_x(0,t) - x_0) \quad$ and $\quad x_t^- := \tfrac{1}{2}(x_0 + V_x(0,t) - x_t)$.

And so x has only countably many discontinuities (though they may even be dense), and thus can be taken to be an R-function (right-continuous with left limits). We shall take all our functions to be R-functions. We adapt the calculus from the left-continuous integrands of V.18 of Rogers and Williams [33] (changing some signs) to give the formulae

(Decomposition) $$x = x_0 + x^c + x^a$$

(Integration by parts) $$d(xy) = x\,dy + y\,dx - \triangle x \triangle y$$

(Itô's formula) $$d(f(x)) = f'(x)\,dx^c + \triangle(f(x))$$

where x and y are FV and f is C^1, $\triangle x_t$ is $x_t - x_{t-}$, and x^c and x^a denote the continuous and purely discontinuous parts of x respectively. There is an expression for x^a as $\sum_{0<s\leqslant t} \triangle x_s$. As x^+ and x^- are increasing they induce positive σ-finite Lebesgue-Stieltjes measures on $(0,\infty)$, via $x^+(a,b] = x^+_b - x^+_a$. So we can associate x with the (signed) measure of their difference. We write $dx_t = dx^+_t - dx^-_t$. We will also use the notation $|dx_t|$ for $dx^+_t + dx^-_t = dV_x$. The differential expressions above are symbolic, being merely shorthand for integral expressions.

We will also use an FV exponential result in that if x is BV and

$$\text{if} \quad dx_t \geqslant -x_t|dy_t|, \quad \text{then} \quad x_t \leqslant x_\infty \prod_{s>t}(1+|\triangle y_s|) \exp V_{y^c}(t,\infty). \tag{4.60}$$

Another useful result follows from integration by parts, in that

$$\tfrac{1}{2} d\|x_t\|^2 = \langle x_t, dx_t\rangle - \tfrac{1}{2}\|\triangle x_t\|^2. \tag{4.61}$$

Theorem 4.14 *Theorem 4.13 remains true if m is a discontinuous non-negative density of bounded variation.*

Proof of Theorem 4.14 We follow the proof of Theorem 4.13 exactly down to (4.51), except that we take R, F S and P to be functions of t rather than v. We write $G := F^{-1}$, $w := Fy$ and $w^* := Gy$, and take the normalisation that $\|y\| = 1$ ($\iff$ $\langle w, w^*\rangle = 1$) and $\langle w, dw^*\rangle = 0$ ($\iff$ $\langle dw, w^*_-\rangle = 0$). Note that all these functions are BV. Then (4.52) becomes

$$d\chi = \lambda^{-1} S \chi\,dt + dG\,F\chi. \tag{4.62}$$

$$\text{So} \quad \tfrac{1}{2}d\|\chi_t\|^2 \geqslant \langle dG\,F\chi, \chi\rangle_t - \tfrac{1}{2}\|\triangle GF\chi\|_t^2 \geqslant K_1\|\chi_t\|^2\|dG\|, \tag{4.63}$$

by (4.61) and (4.60), for some constant K_1. Hence $\|\chi_t\|$ is uniformly bounded in t by some constant K_χ. Now using (4.62) and (4.61) we again work with the components of χ orthogonal to y,

$$\begin{aligned}\tfrac{1}{2}d\|\chi_-\|_t^2 &\geqslant \frac{\sigma_0}{\lambda}\|\chi_-\|_t^2\,dt + \langle dG\,F\chi - dP\,\chi + \triangle P\triangle\chi, \chi_-\rangle_t - \tfrac{1}{2}\|\triangle(P\chi)\|_t^2 \\ &\geqslant \frac{\sigma_0}{\lambda}\|\chi_-\|_t^2\,dt - K_2(\|dG_t\| + \|dP_t\|),\end{aligned}$$

for some constant K_2. So a result of the same form as (4.56) holds. Finally we find that

$$d\xi = \xi\langle w, dw^*\rangle + \langle F\chi_-, dw^*\rangle,$$

and we have a bound similar to that of (4.58), because $\langle w, dw^*\rangle = 0$. Explicitly, w_t is the positive eigenvector of $Q + m_tV$ with the normalisation that $w_\infty = \mathbf{1}$ and dw_t is orthogonal to the positive eigenvector of $Q^\top + m_{t-}V$. □

We can calculate an exact expression for w.

Lemma 4.15 *Let $y(v)$ be the positive eigenvector of $Q + V$ of constant norm, with $y(0) = \mathbf{1}$. If $w_t^0 := y(m_tv)$ and w_t^* is the positive eigenvector of $Q^\top + m_tV$ satisfying $\langle w_t^0, w_t^*\rangle = 1$, then*

$$w_t = w_t^0\prod_{s>t}(1 + \langle\triangle w_s^0, w_{s-}^*\rangle)\exp\int_t^\infty \langle dw_s^{0,c}, w_s^*\rangle.$$

Further $w_t = w_t(v, m)$ is continuous in v.

Proof of Lemma 4.15 If we set $w_t = r_tw_t^0$, then

$$dw_t = dr_t^c\,w_t^0 + r_t\,dw_t^0 + \triangle r_tw_{t-}^0 \quad \text{so} \quad \langle dw_t, w_{t-}^*\rangle = dr_t + r_t\langle dw_t^0, w_{t-}^*\rangle.$$

An application of (4.60) gives the expression for w. Elementary perturbation results, in for example Kato [7], tell us that y is smooth in v and so

$$dw_t^0(v) = \langle v, \nabla\rangle y(m_tv)\,dm_t^c + \triangle(w_t^0(v)).$$

Thus the difference $dw_t^0(v) - dw_t^0(u)$ can be written as

$$\Big(\langle v-u, \nabla\rangle y(m_tv) + \langle u, \nabla\rangle(y(m_tv) - y(m_tu))\Big)\,dm_t^c + \triangle(y(m_tv) - y(m_tu)).$$

$$\text{Hence} \qquad |dw_t^0(v) - dw_t^0(u)| \leqslant K_1\|v-u\|\,|dm_t^c| + K_2\|v-u\|\,|\triangle m_t|,$$

where $K_1 := \sup_{K_V} \|\nabla y(v)\| + V_K \sup_{K_V} \|v\| \, \|\nabla y(v) - \nabla y(u)\| / \|v - u\|$ and some constant K_2. Hence w_t is (Lipschitz) continuous in v. $\square$

Theorem 4.16 *Let X be an honest irreducible Markov chain on a finite set S, with Q-matrix Q. Let m be a non-negative density on $[0, \infty)$ of bounded variation. Then where f_i^λ is the density of A_λ on M under the law starting X at i, the (f_i^λ) can be written as*

$$f_i^\lambda(x) = e^{-K(x)/\lambda} (2\pi\lambda)^{-(n-1)/2} \big(\det H_K(x)\big)^{1/2} z_i(x) r_i^\lambda(x), \tag{4.64}$$

where K is as defined by (4.40), H_K denotes its Hessian taken with respect to M, $z(x)$ is the positive eigenvector of $Q + \mathrm{diag}(m_0 \nabla K(x))$, and the residue term r^λ goes to 1 as λ goes to 0, in the sense that

$$\limsup_{\lambda \downarrow 0} \int_F r_i^\lambda(x + \sqrt{\lambda} y)\, dy \leqslant |F|, \quad \textit{and} \quad \liminf_{\lambda \downarrow 0} \int_G r_i^\lambda(x + \sqrt{\lambda} y)\, dy \geqslant |G|, \tag{4.65}$$

for all x in $\mathrm{Int}(M)$, and for F and G respectively closed and open bounded subsets of $\mathbf{1}^\perp$.

Notes: (1) We take the Hessian regarding K as a function on an open subset of $\mathbb{R}^{n-1}$, that is $K(x_1, \ldots, x_{n-1}, 1 - \sum_i^{n-1} x_i)$. See the example at the end of this Chapter.

(2) Unfortunately we would really like to prove the result that

$$\mathbb{P}_i(A_\lambda \in H) \sim \int_H e^{-K(x)/\lambda} (2\pi\lambda)^{-(n-1)/2} \big(\det H_K(x)\big)^{1/2} z_i(x)\, dx, \tag{4.66}$$

for suitable H, as λ goes to 0. This could be proved if the integrand in our control of r^λ was $r_i^\lambda(x + \lambda y)$ rather than $r_i^\lambda(x + \sqrt{\lambda} y)$.

Proof of Theorem 4.16 Theorem 4.13 can be taken as saying that for v in $\mathbb{R}^S$,

$$f_i^{v,\lambda}(x) := \exp\Big(\lambda^{-1}\big(\langle v, x\rangle - \eta(v)\big)\Big) f_i^\lambda(x) / w_i(v) \tag{4.67}$$

is (asymptotically) a density on M, where $w_i(v)$ is the $w_0(v, m)(i)$ of Lemma 4.15, and our $z_i(x)$ will be $w_i(\nabla K(x))$. If A_λ^v under $\mathbb{P}_i$ has the law $f_i^{v,\lambda}$, then we can derive a central limit result by considering $Z_\lambda^v := (A_\lambda^v - \nabla\eta(v))/\sqrt{\lambda}$. We see that for u in $\mathbb{R}^S$,

$$\begin{aligned} \mathbb{E}_i\big(\exp\langle u, Z_\lambda^v\rangle\big) = \exp\Big(&\lambda^{-1}\big(\eta(v + \sqrt{\lambda} u) - \eta(v) - \langle \sqrt{\lambda} u, \nabla\eta(v)\rangle\,\big)\Big) \\ &\ldots \Big(w_i(v + \sqrt{\lambda} u) + o(1)\Big) / w_i(v), \end{aligned} \tag{4.68}$$

using the local uniformity in v of the convergence of $o(1)$. As η inherits the smoothness of δ, we can expand it about v as

$$\eta(v+\sqrt{\lambda}u)=\eta(v)+\langle\sqrt{\lambda}u,\nabla\eta(v)\rangle+\tfrac{1}{2}\lambda u^\top H_\eta(v)u+o(\lambda), \tag{4.69}$$

and hence deduce that

$$\lim_{\lambda\downarrow 0}\mathbb{E}_i\big(\exp\langle u,Z_\lambda^v\rangle\big)=\exp\big(\tfrac{1}{2}u^\top H_\eta(v)u\big). \tag{4.70}$$

In other words

$$Z_\lambda^v\xrightarrow{\mathcal{D}}N(0,H_\eta(v)).$$

We can think of $f_i^{v,\lambda}$ as the distribution of A_λ conditioned in some way to converge to $\nabla\eta(v)$, but we do not make this formal. Proposition 4.10 provides the interpretation of $\nabla\eta(v)$ as the maximizing x in the Legendre transform.

Recall that a sequence of laws (ν_n) on a Polish space E converges to a law ν with respect to the weak topology on $M_1(E)$ if $\langle v,\nu_n\rangle\to\langle v,\nu\rangle$ for all v in $C_b(E)$. Billingsley [9], 2.1, shows that this is equivalent to each of the following

$$\begin{aligned} &\limsup_{n\to\infty}\nu_n(F)\leqslant\nu(F) && F\text{ closed in }E,\\ &\liminf_{n\to\infty}\nu_n(G)\geqslant\nu(G) && G\text{ open in }E,\\ \text{and}\quad &\lim_{n\to\infty}\nu_n(H)=\nu(H) && H\text{ in }E\text{ with }\nu(\partial H)=0.\end{aligned}$$

Setting $x=\nabla\eta(v)$, we recall from Proposition 4.10 that $\nabla K(x)$ is v, up to a multiple of $\mathbf{1}$. The asymptotics of the density of Z_λ^v are given by

$$f_i^{Z,\lambda}(y):=\lambda^{(n-1)/2}f_i^{v,\lambda}(x+\sqrt{\lambda}y)\sim(2\pi)^{-(n-1)/2}\det H_K(x)^{\frac{1}{2}}e^{-\frac{1}{2}y^\top H_K(x)y}\,r_i^\lambda(x+\sqrt{\lambda}y),$$

because

$$\begin{aligned}&\langle v,x+\sqrt{\lambda}y\rangle-\eta(v)-K(x+\sqrt{\lambda}y)\\ &\quad=\langle v,x\rangle-\eta(v)-K(x)+\sqrt{\lambda}\langle v-\nabla K(x),y\rangle-\tfrac{1}{2}\lambda y^\top H_K(x)y+o(\lambda)\\ &\quad=\lambda(-\tfrac{1}{2}y^\top H_K(x)y+o(1)).\end{aligned}$$

The normal distribution $N(0,H_\eta(v))$ itself has density

$$f(y)=(2\pi)^{-(n-1)/2}\det H_K(x)^{\frac{1}{2}}e^{-\frac{1}{2}y^\top H_K(x)y},$$

as Proposition 4.10 tells us that $H_K = H_\eta^{-1}$ on M. By Lemma I.45.1 of Williams [39], if H is bounded and $|\partial H| = 0$ then

$$\int I_H(y)(f(y))^{-1} f_i^{Z,\lambda}(y)\,dy \to |H|, \qquad \text{or} \qquad \int_H r_i^\lambda(x+\sqrt{\lambda}y)\,dy \to |H|. \tag{4.71}$$

Hence by the equivalence of the above expressions for weak convergence, the result is proved. □

The following Corollary is intended in the way of a remark, and was the original statement of Theorem 4.16, but is now seen to be weaker, although perhaps a more natural formulation.

Corollary 4.17 *Under the conditions of Theorem 4.16,*

$$\limsup_{\lambda\downarrow 0} \int_F r_i^\lambda(x)\,dx \leqslant |F| \qquad \textit{and} \qquad \liminf_{\lambda\downarrow 0} \int_G r_i^\lambda(x)\,dx \geqslant |G|, \tag{4.72}$$

for F closed in $\mathrm{Int}(M)$ *and G open in* $\mathrm{Int}(M)$*. In other words, $r_i^\lambda(x)\,dx$ converges weakly to dx on* $\mathrm{Int}(M)$*.*

Proof of Corollary 4.17 Take G open in $\mathrm{Int}(M)$, δ small and positive with $G_\delta := \{y \in G : B(y,\delta) \subseteq G\}$, and B a ball around 0, then by Fatou's lemma and Fubini's theorem

$$\begin{aligned} |G_\delta|\,|B| &= \int_{G_\delta} \left(\liminf_{\lambda\downarrow 0} \int_B r_i^\lambda(x+\sqrt{\lambda}y)\,dy \right) dx \\ &\leqslant \liminf_{\lambda\downarrow 0} \int_{G_\delta} \int_B r_i^\lambda(x+\sqrt{\lambda}y)\,dy\,dx \leqslant \left(\liminf_{\lambda\downarrow 0} \int_G r_i^\lambda(x)\,dx \right) |B|. \end{aligned}$$

Letting δ tend to 0, we have one of our bounds. For some F closed in $\mathrm{Int}(M)$, we need $\int_B r_i^\lambda(x+\sqrt{\lambda}y)\,dy$ to be uniformly bounded on F and for λ near 0. It is, and the bound is

$$\sup_{x\in F} (2\pi)^{(n-1)/2} \det H_K(x)^{-\frac{1}{2}} \sup_{y\in B} e^{\frac{1}{2} y^\top H_K(x) y} < \infty.$$

Working with this F and with $F^\delta := \{y \in M : d(y,F) \leqslant \delta\}$, we can show in a similar way that

$$\limsup_{\lambda\downarrow 0} \int_F r_i^\lambda(x)\,dx \leqslant |F^\delta|,$$

and hence we are home. □

Some remarks.

(a) Were the r_i^λ to be equicontinuous (or some such condition) we would have that $r_i^\lambda(x) \to 1$ for all x and hence that $f_i^\lambda(x)/f_j^\lambda(x) \to z_i(x)/z_j(x)$ and $\lim -(Qf^\lambda)_i/f_i^\lambda$ differs from $m_0\nabla K(x)$ only by a multiple of $\mathbf{1}$, as in Section 2.2 of Chapter Two, where the choice of $\nabla K(x)$ in $\ker(\delta)$ was called $g(x)$.

(b) Note that the proof of Theorem 4.16 gives us a central limit theorem for A_λ as

$$Z_\lambda := (A_\lambda - \pi)/\sqrt{\lambda} \xrightarrow{\mathcal{D}} N(0, H_\eta(0)). \tag{4.73}$$

Taking a Taylor expansion of δ about 0 and integrating we discover that $H_\eta(0) = \sigma^2 H_\delta(0)$, where $\sigma = \|m\|_{L^2}$ which is finite because m is in both L^1 and L^∞.

Example. (This case was first studied in Section 2.3) Suppose we have a Markov chain which is symmetric and space-homogeneous, with Q-matrix $q_{ij} := \pi_j - \delta_{ij}$, where π is a distribution on a finite set S. The Cesàro large-deviation rate function is $I(x) = 1 - (\sum \sqrt{\pi_i x_i})^2$, and the exponentially discounted large-deviation rate is $K(x) = \sum \pi_i \log(\pi_i/x_i)$. We found then that $\delta(v)$ is the unique root δ in $(\max_i(v_i - 1), \infty)$ of

$$\sum_{i\in S} \frac{\pi_i}{\delta + 1 - v_i} = 1,$$

and that η is given by $\eta(v) = \delta(v) - \sum_i \pi_i \log(\delta(v) + 1 - v_i)$. We find now that

$$\begin{aligned}
\nabla_i I(x) &= 1 - \sqrt{\frac{\pi_i}{x_i}}\Big(\sum_{j\in S}\sqrt{\pi_j x_j}\Big),\\
\nabla_i \delta(v) &= \frac{\pi_i}{(\delta(v)+1-v_i)^2}\Big/\sum_{j\in S}\frac{\pi_j}{(\delta(v)+1-v_j)^2},\\
\nabla_i K(x) &= 1 - \frac{\pi_i}{x_i},\\
\text{and}\quad \nabla_i \eta(v) &= \frac{\pi_i}{\delta(v)+1-v_i}.
\end{aligned}$$

Here we chose ∇I and ∇K to be in the kernel of δ. The distribution of A_λ can be calculated explicitly to be a multidimensional β-distribution with density

$$f_i^\lambda(x) = \frac{x_i}{\pi_i}\Gamma(\lambda^{-1})\prod_{j\in S}\frac{x_j^{(\pi_j/\lambda)-1}}{\Gamma(\pi_j/\lambda)}.$$

Note that the Hessian of K on M is not the same as that derived from the extension of K to $\mathbb{R}^S$, but by using any of the following co-ordinate schemes:

$$
\begin{aligned}
&K_i : \mathbb{R}^{S\setminus\{i\}} \longrightarrow \mathbb{R} \qquad \text{for each } i \in S\\
\text{where}\qquad &(x_j)_{j\neq i} \longmapsto K(x_1,\dots,x_{i-1},1-\sum_{j\neq i} x_j, x_{i+1},\dots,x_n)\\
\text{or}\qquad &K_0 : \mathbb{R}^n \longrightarrow \mathbb{R}\\
\text{where}\qquad &x \longmapsto K(x+(1-\mathbf{1}^\top x)\mathbf{1}/n)+\tfrac{1}{2}(\mathbf{1}^\top x)^2.
\end{aligned}
$$

What is happening here is that our choice of basis for evaluating the Hessian corresponds to our choice of basis for integrating which was made back at the start of Section 2.2 of Chapter Two. The K_0 representation projects onto M and adds a strictly convex term which is perpendicular to M. This representation is more natural, though cumbersome to calculate with, and can be shown equivalent to any of the others by verifying that the change of basis matrix has determinant one. Thus the Hessian (in the K_n realisation) and its determinant are given by

$$
H_{K_n}(x)_{ij} = \frac{\pi}{x_i^2}\delta_{ij} + \frac{\pi_n}{x_n^2} \qquad \text{and} \qquad \det(H_K(x)) = \Big(\prod_{i\in S}\frac{\pi_i}{x_i^2}\Big)\sum_{j\in S}\frac{x_j^2}{\pi_j}.
$$

The normalisation of the eigenvector $z_i(x)$ is that $\|z\|_\pi = 1$, so it is given by

$$
z_i(x) = \frac{x_i}{\pi_i}\Big(\sum_{j\in S}\frac{x_j^2}{\pi_j}\Big)^{-1/2}.
$$

(The corresponding vector for the Cesàro case is $\sqrt{x_i/\pi_i}(\sum_j \sqrt{\pi_j x_j})$.) We can now calculate the residual functions using Stirling's formula

$$
\Gamma(x) = \sqrt{2\pi}x^{x-\frac{1}{2}}e^{-x}(1+O(x^{-1})),
$$

where $|O(x^{-1})| \leqslant K/x$ as $x\to\infty$ for some constant K. It is thus discovered that the residual functions $r_i^\lambda(x)$ can be calculated and are found to be independent of both i and x, and are of size $1+O(\lambda)$.

Hypothesis 4.18 We recall from Theorem 2.3 that in the set-up of Theorem 4.16 with the exponential discount ($m_t = e^{-t}$), the density f^λ satisfies the vector differential equation

$$
\mathcal{L}f^\lambda = -\lambda^{-1}Qf^\lambda, \tag{4.74}
$$

where $\mathcal{L}$ is the matrix differential operator $\mathcal{L} = \operatorname{diag}(\sum_{j\neq i}(\partial_j - \partial_i)x_j)_{i\in S}$. Here we have changed the domain of f^λ from a subset of $\mathbb{R}^{n-1}$ equivalent to M, to a neighbourhood of M in $\mathbb{R}^n$ by extension. The operator $\mathcal{L}$ is invariant to the extension chosen. If we discount f^λ by the known large-deviation rate function K, that is by defining g^λ by

$$f^\lambda(x) = e^{-K(x)/\lambda}(2\pi\lambda)^{-(n-1)/2} g^\lambda(x)$$

then

$$\mathcal{L}g^\lambda = \lambda^{-1} R\big(\nabla K(x)\big) g^\lambda.$$

This compares with equation (4.49) which said that

$$\partial_t \psi = \lambda^{-1} R(m_t v)\psi,$$

where ψ is the discount of φ as defined by (4.48). The matrix $R(v)$ has a simple eigenvalue 0 and all other eigenvalues have positive real part. We saw that ψ tended to a multiple of the 0-eigenvector of $R(m_t v)$ as λ went to 0, and also that g^λ tended (in some sense) to $z(x)$, which was the 0-eigenvector of $R(\nabla K(x))$. We can formulate an analogue of (4.74) for the general discount case as follows.

Let us write $A_{\lambda,t}$ for $\lambda M_t^{-1}\int_0^\infty \theta_t m(\lambda s)\delta_{X_s}\,ds$, where $M_t := \int_0^\infty \theta_t m(s)\,ds$, and $f_i^{\lambda,t}$ for the density of $A_{\lambda,t}$ if X starts in state i. Then $A_{\lambda,t}$ will satisfy the large-deviation property with rate function K_t, where

$$K_t = \eta_t^* \quad \text{where} \quad \eta_t(v) := \int_0^\infty \delta(M_t^{-1}\theta_t m(s)v)\,ds,$$

and we write $f^{\lambda,t}$ as

$$f^{\lambda,t}(x) = e^{-K_t(x)/\lambda}(2\pi\lambda)^{-(n-1)/2} g^{\lambda,t}(x).$$

Then

$$M_t^{-1} m_t \mathcal{L} g^{\lambda,t} + \partial_t g^{\lambda,t} = \lambda^{-1} R\big(M_t^{-1} m_t \nabla K_t(x)\big) g^{\lambda,t}.$$

Again Theorem 4.16 tells us that $g^{\lambda,t}$ tends (in some sense) to the 0-eigenvector, z, of the matrix R. We hypothesize that the convergence is in fact pointwise.

Chapter Five

Processes on the Binary Tree

If you wish to advance into the infinite,
explore the finite in all directions.

Goethe, *Maximen und Reflexionen*

5.1 Introduction and Summary

We consider a Markov chain on the nodes of the binary tree, I:

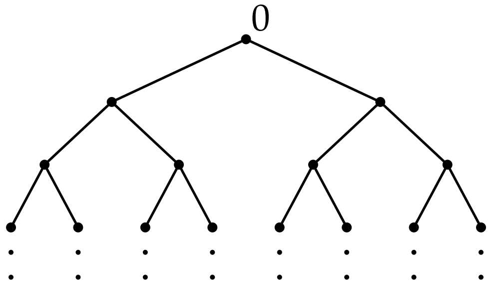


Figure 5.1 Graph of the Binary Tree

By choice of jump rates and the relative up-down weightings, we can ensure that the chain is reversible, positive recurrent, and able to hit infinity and return in finite time. Our basic structural assumption, on which we shall lean heavily, is of lateral symmetry — that is that jump rates depend on the state only through its level and that the process is equally likely to go left as right on any down jump.

As we shall see, Rogers and Williams [32] allows us the existence of the chain with reflection at its boundary. The Ray-Knight compactification can be thought of as:

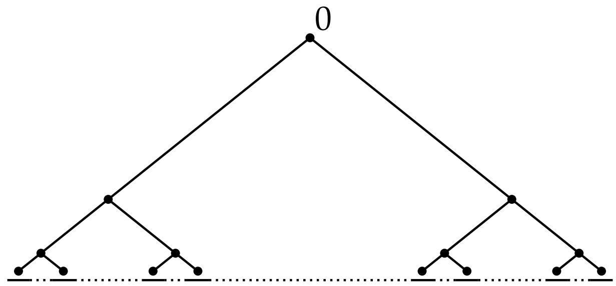


Figure 5.2 Ray-Knight Compactification

Let I be the points of the nodes of the graph, C be the Cantor set of limit points of I, and F be $I \cup C$, the Ray-Knight compactification of I.

Given the F-valued process X, the projection process $n(X_t)$, the level of X_t, is also a Markov chain (by the symmetry) and is a birth-death process on the non-negative integers with the one-point compactification at infinity. This level process reflects from infinity, as the projection of the time-truncation of X is the same process as the time-truncation of the reflecting birth-death process.

The chain can be fully quantified by the jump rates from level n.

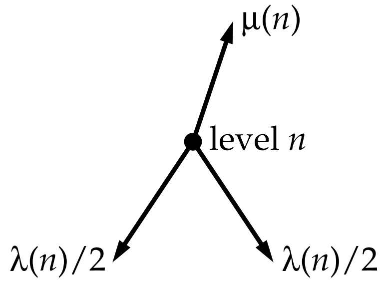


Figure 5.3 Graph process

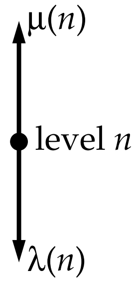


Figure 5.4 Projected BD process

The up-jump rate is μ_n and the left and right down-jump rates are both $\frac{1}{2}\lambda_n$, and we shall define q_n to be $\lambda_n + \mu_n$. We put $\pi_n := (\lambda_0 \ldots \lambda_{n-1})/(\mu_1 \ldots \mu_n)\pi_0$, the invariant measure for the BD process, choosing π_0 to make π a distribution if it is a finite measure.

Let us give a formal statement of the construction of the chain on the tree:

Theorem 5.1 *There exists a symmetrizable transition matrix $P(t)$ and a Feller resolvent R_λ on F, and a strong Markov F-valued honest R-process X with law $P(t)$ such that X makes jumps from points of I as given by Figure 5.3, and reflects off the boundary C. Further the $P(t)$ semigroup is Feller-Dynkin (FD), and for all the boundary points ξ, $\lambda r_\lambda(\xi, I) = 1$.*

Proof of Theorem 5.1 The proof is given in Section 5.3.

NOTATION: We shall denote states of the graph process on I by i, j and so on, and states of the level process by n. For example the invariant measure of the graph process, ν, is given by $\nu_i = 2^{-n(i)}\pi_{n(i)}$, where $n(i)$ is the level of state i, and π is the invariant measure of the birth-death level process as given above.

We can think in terms of the level process as being the time-change of a Brownian motion reflecting at each end of a bounded interval, $[0, a]$. This is made rigorous later (Section 5.4). The set of times at which this Brownian process is at 0 is a random Cantor set, which in particular is uncountable and perfect. Our first question is inspired by noticing that when the graph process first hits the boundary, it must almost immediately return uncountably often to the boundary. But there are uncountably many other boundary points nearby. Does it return to the same point of the boundary ?

That is, is the boundary point regular ? One can think of a bolt of lightning which bounces back off the ground and back down again repeatedly. From all the possible bits of ground, how can it ever find again the exact spot where it first hit ?

Assuming (wisely) that we can arrange it so that all boundary points are regular (and thus have individual local times), our second question is: can we find a jointly-continuous version of the local time on F ?

The precise answers to all these questions are contained in the following theorem:

Theorem 5.2 *Let X be the graph process as constructed above. Then*

(1) X is positive recurrent $\iff \sum_n \pi_n < \infty$, and then

(2) X reaches the boundary in finite time $\iff \sum_n \frac{1}{\lambda_n \pi_n} < \infty$, and then

(3) any (and hence each) boundary point is regular $\iff \sum_n b_n < \infty$, and then

(4) there exists a jointly-continuous local time on F $\iff \sum_n \sqrt{\frac{1}{n} c_n} < \infty$, and then

(5) X has visited all the states of F by a finite time,

where $b_n := \frac{2^n}{\lambda_n \pi_n}$, and $c_n := \sum_{r=n}^{\infty} b_r$.

Corollary 5.3 *If (1)–(3) hold then*

$$\sum_n n^{1+\epsilon}\, b_n < \infty \quad \Rightarrow \quad \text{there exists a jointly-continuous local time on } F,$$

$$\text{and} \quad \sum_n n b_n = \infty \quad \Rightarrow \quad \text{there does not.}$$

We shall see in the next and final Chapter how the conditions (3) and (4) can be seen as translations of results for continuous space Lévy processes.

We now have enough to construct a boundary (Cantor set) valued process, which we examine in Section 5.5.

5.2 Proof of Theorem 5.2

This section contains the mathematics of the proof, the next contains the arithmetic.

Proof of Parts (1) and (2) These parts of the theorem are basic Markov chain theory. See, for example, 4-3 of Wolff [7]. □

Proof of Part (3) Rogers and Williams [32] have given us a standard honest π-symmetric transition matrix function $P(t) = (p_{ij}(t))_{i,j\in I}$, such that $\mathbb{P}_i(X_t = j) = p_{ij}(t)$. We define its Laplace transform $R(\lambda)$ and a ν-normalised symmetric resolvent u_λ by

$$\begin{aligned} r_{ij}(\lambda) &:= \int_0^\infty e^{-\lambda t} p_{ij}(t)\,dt, \\ \text{and} \quad u_\lambda(i,j) = u_\lambda(j,i) &:= r_{ij}(\lambda)/\nu_j \end{aligned} \tag{5.1}$$

respectively. It is known that a boundary point ξ is regular if and only if (for any and hence all λ) u_λ has a continuous extension to (ξ,ξ) and $u_\lambda(\xi,\xi) < \infty$. By the definition of the Ray-Knight compactification, in for example III.57 of Williams [39], we see that u_λ has a finite continuous extension to $F \times I$, and hence by symmetry to $F \times F \backslash \{(\xi,\xi) : \xi \in F\backslash I\}$.

Let us put a partial order on F by saying $x < y$ for x, y in F, if $n(x) < n(y)$ and x is one of the points between y and the root of the tree 0. We say that x is *above* y, and that y is *below* x. For any pair x and y, we let $x \wedge y$ be the $<$-greatest point which is above both of them. Pick an i above ξ and let $I_i := \{j \in I : i \not< j\}$ be the set of all points not below i. For any k below i, and for any j in I_i, all paths from j to k must pass through i, so that the strong Markov property implies that

$$u_\lambda(j,k) = \mathbb{E}_j\left(e^{-\lambda H(i)}\right) u_\lambda(i,k),$$

where $H(i)$ is the time to first hit state i. Letting k tend to $\xi \in C$, we find that

$$u_\lambda(j,\xi) = \mathbb{E}_j\left(e^{-\lambda H(i)}\right) u_\lambda(i,\xi).$$

Then

$$\begin{aligned} \sum_{j\in I_i} \nu_j u_\lambda(j,\xi) &= \sum_{j\in I_i} \nu_j \mathbb{E}_j\left(e^{-\lambda H(i)}\right) u_\lambda(i,\xi), \\ \text{and so} \qquad u_\lambda(i,\xi) &= \frac{\lambda \sum_{j\in I_i} \nu_j u_\lambda(j,\xi)}{\lambda \sum_{j\in I_i} \nu_j \mathbb{E}_j\left(e^{-\lambda H(i)}\right)}. \end{aligned} \tag{5.2}$$

As i goes to ξ, $I_i \uparrow I$, and the numerator in (5.2) tends upwards to $\lambda r_\lambda(\xi, I)$ which is 1 by Theorem 5.1. Therefore

$$u_\lambda(\xi,\xi) = \frac{1}{\lambda \sum_I \nu_j \mathbb{E}_j \left(e^{-\lambda H(\xi)}\right)}, \tag{5.3}$$

and $u_\lambda(\xi,\xi) < \infty$ if and only if $H(\xi) < \infty$ (a.s.).

We let the "up-jump time", V_n, be a random variable distributed as the time to hit level $(n-1)$ starting at level n, and let the "left-down-jump time", T_n, be a random variable distributed as the time to hit the point below and to the left of a start point on level n. Then we can control the means and variances of these in the following theorem.

Theorem 5.4 *If (1) and (2) hold then*

$$\mathbb{E}V_n = \frac{\pi[n]}{\lambda_{n-1}\pi_{n-1}}, \qquad \mathrm{Var}(V_n) = \frac{1}{\lambda_{n-1}\pi_{n-1}} \sum_{r=n}^{\infty} \left(\frac{\pi[r]^2}{\lambda_{r-1}\pi_{r-1}} + \frac{\pi[r+1]^2}{\lambda_r \pi_r} \right),$$

$$\mathbb{E}T_n = \frac{2^{n+1} - \pi[n+1]}{\lambda_n \pi_n}, \quad \mathrm{Var}(T_n) \leqslant K \frac{2^n}{\lambda_n \pi_n} \sum_{r=0}^{n} \frac{2^r}{\lambda_r \pi_r}, \quad \textit{for some } K,$$

where $\pi[r] := \pi(\{r, r+1, \ldots\})$.

We defer this proof till Section 5.3, but the method of calculation in each case is just to find the minimal non-negative solution to a system of equations induced by conditioning on the first jump. We find that means are enough for upper bounds and sufficiency, but we need control away from 0, that is variance information, for lower bounds and necessity.

Proof of Sufficiency of (3) If $\sum_n b_n < \infty$ (where $b_n = 2^n/\lambda_n \pi_n$) then Theorem 5.4 shows that $\mathbb{E}_0(H(\xi)) = \sum_n \mathbb{E}T_n < \infty$ and so ξ is a regular boundary point. □

Proof of Necessity of (3) Conversely, if $\sum_n b_n = \infty$ we use the following lemma:

Lemma 5.5 (Lower-Bound Lemma) *If* $X : \Omega \to [0, \infty]$ *is a random variable such that* $\mathbb{E}(X^2) \leqslant K\mathbb{E}(X) < \infty$ *for some* K*, then*

$$\mathbb{E}\left(1 - e^{-X}\right) \geqslant \alpha \mathbb{E}(X) \tag{5.4}$$

$$\textit{where} \quad \alpha = \alpha(K) = \frac{1 - e^{-4K}}{8K}.$$

Proof of Lemma 5.5 The proof is given in Section 5.3, and uses concavity coupled with the variance control. □

By Theorem 5.4 (noticing that $(\mathbb{E}T_n)^2$ is of no greater order than $\mathrm{Var}(T_n)$) we have that

$$\mathbb{E}(T_n^2) \leqslant K d_n \mathbb{E}(T_n),$$

(for some new K) where $d_n = \sum_{r=0}^n b_r$. And so the Lower Bound Lemma 5.5 tells us that

$$\begin{aligned}\mathbb{E}\left(1 - e^{-T_n}\right) &\geqslant \frac{1 - e^{-4Kd_n}}{8Kd_n}\mathbb{E}(T_n) \\ &\geqslant \alpha \frac{b_n}{d_n} \quad \text{for some } \alpha > 0,\end{aligned}$$

as $d_n \to \infty$ as $n \to \infty$. Kronecker's Lemma tells us that $\sum_n \frac{b_n}{d_n} = \infty$, and we deduce that $\sum_n \mathbb{E}(1 - e^{-T_n}) = \infty$, and hence $\mathbb{E}_0(e^{-H(\xi)}) = \prod_n \mathbb{E}(e^{-T_n}) = 0$, giving us the necessity of condition (3). Here we have used, and will use again, the useful analysis lemma that for a sequence (x_n) in $(0, 1)$, $\sum_n (1 - x_n)$ is finite if and only if $\prod_n x_n$ is positive. □

Proof of Part (4) Given (1)–(3), we can assume an individual local time $L(t, x)$ for each point x of F. For (4) we use an excellent paper of Marcus and Rosen [28], which uses an Isomorphism theorem of Dynkin between the Markov chain on the graph, and zero-mean Gaussian process on the graph with covariance equal to the 1-potential density $u_1(\cdot, \cdot)$. Their theorems 2 and 8.1 together state that

Theorem 5.6 (Marcus and Rosen) *Let X be a strongly symmetric standard Markov Process with continuous 1-potential density u_1. Let $L = \{L(t, x) : t \in \mathbb{R}^+,\ x \in F\}$ be the joint local time of X, then L is continuous a.s. if and only if there exists a probability measure m on F such that*

$$\sup_{x \in F} \int_0^\delta \left[\log \frac{1}{m(B_d(x, r))}\right]^{1/2} dr \longrightarrow 0 \qquad \text{as } \delta \to 0, \tag{5.5}$$

where $B_d(x, r)$ is the radius-r closed ball centered on x under the metric d, where

$$d(x, y) = \left[u_1(x, x) + u_1(y, y) - 2u_1(x, y)\right]^{1/2}. \tag{5.6}$$

X is *strongly symmetric* if u_1 exists as a symmetric π-density for the Laplace transform of $P(t)$, which here is true. Our first step is to show (in Section 5.3)

Lemma 5.7 *Given (1)–(3), for x, y in F, with $n(x) < n(y)$, then*

$$\alpha \mathbb{E}_{x \wedge y}(H(y)) \leqslant d^2(x,y) \leqslant A\mathbb{E}_{x \wedge y}(H(y)),$$

for some universal constants α and A.

Proof of Lemma 5.7 Given in Section 5.3. □

As already hinted, the upper bound follows from simple inequalities and knowledge of the means, whilst the lower is derived from variance control and the more subtle analysis of the Lower Bound Lemma 5.5. The lemma tells us that d is "equivalent" to the sequence (c_n), and we essentially translate condition (5.5) of Theorem 5.6 into a statement about (c_n), preserving both necessity and sufficiency.

Proof of Sufficiency of (4) We elect m to be $\frac{1}{2}p + \frac{1}{2}c$, where p is the probability on I giving mass $2^{-(2n+1)}$ to each point on level n, and c is the Cantor distribution (Hausdorff measure) on C. Our condition holds, that is

$$\sum_n \sqrt{\frac{1}{n} c_n} < \infty \qquad \text{where} \quad c_n := \sum_{r=n}^{\infty} b_r, \quad \text{and} \quad b_n := \frac{2^n}{\lambda_n \pi_n}.$$

Putting $x = \xi \in C$, we can deduce from Theorem 5.4 and Lemma 5.7 that (for new α and A)

$$\alpha c_{n(y \wedge \xi)} \leqslant d^2(\xi, y) \leqslant A c_{n(y \wedge \xi)}. \tag{5.7}$$

Thus for r chosen to lie in $\sqrt{Ac_n} \leqslant r < \sqrt{Ac_{n-1}}$, then $I_n := \{y \geqslant \xi_n\} \subset B_d(\xi, r)$, where ξ_n is the point on level n above ξ. So

$$m(B_d(\xi, r)) \geqslant m(I_n) \geqslant \tfrac{1}{2} c(I_n) = 2^{-(n+1)},$$

and

$$\begin{aligned}
\int_0^{\sqrt{Ac_n}} \left[\log \frac{1}{m(B_d(\xi,r))}\right]^{1/2} dr &\leqslant \sum_{r=n}^{\infty} \sqrt{A \log 2}\left(\sqrt{c_r} - \sqrt{c_{r+1}}\right)\sqrt{r+2} \\
&\leqslant K\left(\sqrt{nc_n} + \sum_{r=n}^{\infty}\left(\sqrt{r+1} - \sqrt{r}\right)\sqrt{c_r}\right) \\
&\leqslant K\left(\sqrt{nc_n} + \sum_{r=n}^{\infty}\sqrt{\frac{1}{r} c_r}\right).
\end{aligned}$$

By the monotonicity of (c_n), $\sqrt{nc_n} < 2\sum_{r=\lceil n/2\rceil}^{\infty}\sqrt{\frac{1}{r}c_r}$ which goes to 0 as n goes to infinity, so that the whole right-hand side goes to 0 as we wish.

We also have to get a similar result when $x = i \in I$. Let N be $n(i)$, the level of i, and let i_n be the point on level n above i, for $n \leqslant N$. As before, for r such that $\sqrt{Ac_n} \leqslant r < \sqrt{Ac_{n-1}}$, then $\{y \geqslant i_n\} \subset B_d(i,r)$. In addition for $0 < r < \sqrt{Ac_N}$, then $\{i\} \subset B_d(i,r)$, so $m(B_d(i,r)) \geqslant m\{i\} = 2^{-(2N+1)}$. And so

$$\int_0^{\sqrt{Ac_M}}\left[\log\frac{1}{m(B_d(i,r))}\right]^{1/2} dr \leqslant K'\begin{cases}\sqrt{Mc_M}+\sqrt{Nc_N}+\sum_{r=M}^{N-1}\sqrt{\frac{1}{r}c_r} & M<N,\\ \sqrt{Nc_M} & M \geqslant N,\end{cases}$$

which goes to 0 uniformly in N as $M \to \infty$. Thus the sufficiency is proved. □

Proof of Necessity of (4) For this we use the lower bound for d. If m is any probability measure on F, set ξ_0 to be 0, and recursively define ξ_{n+1} to be the point immediately below ξ_n which has no more m-mass in its subtree than the other point immediately below ξ_n, so that $m\{y \geqslant \xi_n\} \leqslant 2^{-n}$. We set $\xi := \lim_n \xi_n$ be a boundary point. If r is such that $\sqrt{\alpha c_n} \leqslant r < \sqrt{\alpha c_{n-1}}$, then $B_d(\xi,r) \subset \{y \geqslant \xi_n\}$, so

$$\begin{aligned}\int_0^{\sqrt{\alpha c_N}}\left[\log\frac{1}{m(B_d(\xi,r))}\right]^{1/2} dr &\geqslant \sum_{n=N}^{\infty}\sqrt{\alpha\log 2}\left(\sqrt{c_n}-\sqrt{c_{n+1}}\right)\sqrt{n+1}\\ &\geqslant k\sum_{n=N+1}^{\infty}\sqrt{\frac{1}{n}c_n}, \quad \text{for some } k>0.\end{aligned}$$

□

Proof of Part (5) Fix $\xi \in C$ and let $A_n := \{\zeta \in C : \zeta > \xi_n\}$, where ξ_n is the point on level n above ξ. Let $\tau_\xi(t) := \inf\{s : L^X(s,\xi) > t\}$, and set

$$p(n,t) := \mathbb{P}_\xi(L^X(\tau_\xi(t),\cdot) > 0 \text{ on } A_n).$$

The function p is monotone in each co-ordinate, and as L^X is jointly-continuous $\lim_{n\to\infty} p(n,t) = 1$, $(t > 0)$. Thus p is positive for some (and hence all) n, and the strong Markov property gives us that

$$1 - p(0,Nt) \leqslant (1-p(0,t))^N,$$

whence we deduce that $\lim_{t\to\infty} p(0,t) = 1$. If we now set $C_t := \{\zeta \in C : L^X(t,\zeta) > 0\}$, which is open as L^X is continuous, we have proved that $C_t \uparrow C$ as $t \to \infty$, and by the compactness of C, we deduce that $C_T = C$ for some finite T. □

5.3 Various Proofs

Proof of Theorem 5.1 The time-truncation arguments of Rogers and Williams [32] let us take the limit of finite chains on the tree which reflect at level n to give us, their theorem 9.13, a symmetrizable transition matrix $P(t)$ and Feller resolvent R_λ on I. Section III.81 of Williams [39] extends these functions to the Ray-Knight compactification F, and constructs the process X. By symmetry, either all or none of the boundary points are 'relevant' points in the sense of that section, that is that a point ξ is relevant if $\lambda r_\lambda(\xi, I) = 1$. Irrelevant points cannot be visited by X, but X does visit the boundary on explosion, so implying that all boundary points, ξ, are relevant, in other words that $\lambda r_\lambda(\xi, I) = 1$.

Finally, to show that $P(t)$ is FD, we have only to work to prove that

$$P_t f(x) \to f(x) \qquad \text{as } t \downarrow 0,$$

for any f in $C(F)$, the space of continuous functions on F, and for any x in F. This is easy for x in I, so let us assume that $x = \xi$ is in C. Given f in $C(F)$ and any positive ϵ, then there exists a level n such that if we let ξ_n be the point on level n above ξ and let I_n be the set of points below or equal to ξ_n, then $|f(\xi) - f(x)| < \epsilon$ for all x in I_n. Then as, starting at ξ, we cannot leave I_n without passing through ξ_n,

$$P_t(\xi, I_n) \geqslant e^{-q_n t}. \quad \text{And hence} \quad |P_t f(\xi) - f(\xi)| < \epsilon + 2\|f\|_\infty \big(1 - e^{-q_n t}\big),$$

which goes to ϵ as t goes to 0. As ϵ is arbitrary, we can deduce that $P(t)$ is FD. □

Proof of Corollary 5.3 Firstly, if $\sum_n n^{(1+\epsilon)} b_n < \infty$, then by Hölder's inequality

$$\sum_n \sqrt{\frac{1}{n} c_n} \leqslant \left(\sum_n \frac{1}{n^{1+\epsilon}} \right)^{1/2} \left(\sum_n n^\epsilon c_n \right)^{1/2} < \infty,$$

$$\text{because} \qquad \sum_n n^\epsilon c_n = \sum_n b_n \left(\sum_{r=1}^n r^\epsilon \right) \leqslant \sum_n n^{(1+\epsilon)} b_n < \infty.$$

Secondly, if $\sum_n n b_n = \infty$, then as

$$\sum_{n=1}^N n b_n = \sum_{n=1}^N \sum_{r=n}^N b_r \leqslant \sum_{n=1}^N c_n,$$

we see that $\sum_n c_n = \infty$. We consider the sequence (d_n), defined by $d_n := \sqrt{c_n/n}$, and look at the set

$$A := \{n : d_n \geqslant 1/n\} = \{n : c_n \geqslant d_n\}.$$

If A is finite, then (d_n) is eventually more than (c_n) so its sum diverges. If A is infinite, there exists an increasing sequence (n_i) in A, so that by the monotonicity of (d_n)

$$\sum_n d_n \geqslant \sum_i \frac{n_i - n_{i-1}}{n_i} = \sum_i \left(1 - \frac{n_{i-1}}{n_i}\right) = \infty, \qquad \text{as} \quad \prod_i \frac{n_{i-1}}{n_i} = 0. \qquad \square$$

Proof of Theorem 5.4 Let k_n be set to $\mathbb{E}(V_n)$, the expected time to jump-up one level from n. We can expand V_n conditionally on the first jump as

$$V_n = \mathcal{E}(q_n) + \begin{cases} 0 & \text{with prob. } \mu_n/q_n \\ V_{n+1} + \tilde{V}_n & \text{with prob. } \lambda_n/q_n, \end{cases} \tag{5.8}$$

where $\tilde{V}_n$ has the same distribution as V_n, $\mathcal{E}(\alpha)$ is exponentially distributed with rate α, and all variables on the right-hand side are independent. We know the (k_n) are the minimal non-negative solutions to

$$\begin{aligned} k_n &= \frac{1}{q_n} + \frac{\lambda_n}{q_n}(k_n + k_{n+1}), \qquad (n \geqslant 1), \\ \text{or} \qquad (\lambda_{n-1}\pi_{n-1}k_n) &= \pi_n + (\lambda_n \pi_n k_{n+1}) \end{aligned}$$

(using $\mu_n \pi_n = \lambda_{n-1}\pi_{n-1}$). This has the required solution

$$k_n = \frac{\pi[n]}{\lambda_{n-1}\pi_{n-1}}.$$

Similarly the variance sequence $(\mathrm{Var}(V_n))$ will satisfy

$$\begin{aligned} \mathrm{Var}(V_n) &= \frac{1}{q_n^2} + \frac{\lambda_n}{q_n}\big[\mathrm{Var}(V_{n+1}) + \mathrm{Var}(V_n)\big] + \frac{\lambda_n \mu_n}{q_n^2}\big[\mathbb{E}(V_{n+1} + V_n)\big]^2, \\ \text{or} \qquad \big[\lambda_{n-1}\pi_{n-1}\,\mathrm{Var}(V_n)\big] &= \big[\lambda_n \pi_n\,\mathrm{Var}(V_{n+1})\big] + \frac{\pi[n]^2}{\lambda_{n-1}\pi_{n-1}} + \frac{\pi[n+1]^2}{\lambda_n \pi_n}, \end{aligned}$$

which has the desired solution.

Now we let h_n be equal to $\mathbb{E}(T_n)$, the expected time to down-jump one level from n to a particular point. We can decompose T_n as

$$T_0 = \mathcal{E}(\lambda_0) + \begin{cases} 0 & \text{with prob. } \frac{1}{2} \\ V_1 + \tilde{T}_0 & \text{with prob. } \frac{1}{2}, \end{cases}$$

$$T_n = \mathcal{E}(q_n) + \begin{cases} 0 & \text{with prob. } \tfrac{1}{2}\lambda_n/q_n \\ V_{n+1} + \tilde{T}_n & \text{with prob. } \tfrac{1}{2}\lambda_n/q_n \qquad (n \geqslant 1) \\ T_{n-1} + \tilde{T}_n & \text{with prob. } \mu_n/q_n, \end{cases} \tag{5.9}$$

where $\tilde{T}_n$ has the same distribution as T_n. So (h_n) is the minimal solution to

$$\begin{aligned} h_0 &= \frac{1}{\lambda_0} + \tfrac{1}{2}h_0 + \tfrac{1}{2}\frac{\pi[1]}{\lambda_0\pi_0} \\ h_n &= \frac{1}{q_n} + \frac{\mu_n + \frac{1}{2}\lambda_n}{q_n}h_n + \frac{\mu_n}{q_n}h_{n-1} + \frac{\frac{1}{2}\lambda_n}{q_n}\frac{\pi[n+1]}{\lambda_n\pi_n} \qquad (n \geqslant 1), \end{aligned}$$

or

$$\begin{aligned} \lambda_0\pi_0 h_0 &= 2 - \pi[1] \\ (\lambda_n\pi_n h_n) &= 2(\lambda_{n-1}\pi_{n-1}h_{n-1}) + \pi_n + \pi[n] \qquad (n \geqslant 1). \end{aligned}$$

We can use induction to show that

$$(\lambda_n\pi_n h_n) = 2^{n+1} - \pi[n+1], \qquad (n \geqslant 0)$$

whence the result.

Finally $\mathrm{Var}(T_n)$ will be the minimal non-negative solution to the following equations:

$$\begin{aligned} \mathrm{Var}(T_0) &= \frac{1}{\lambda_0^2} + \tfrac{1}{2}\big[\mathrm{Var}(V_1) + \mathrm{Var}(T_0)\big] + \tfrac{1}{4}\big[\mathbb{E}(V_1 + T_0)\big]^2 \\ \mathrm{Var}(T_n) &= \frac{1}{q_n^2} + \frac{\frac{1}{2}\lambda_n}{q_n}\left[\mathrm{Var}(T_n) + \mathrm{Var}(V_{n+1})\right] + \frac{\mu_n}{q_n}\left[\mathrm{Var}(T_n) + \mathrm{Var}(T_{n-1})\right] \\ &\quad + \frac{\frac{1}{2}\lambda_n(\mu_n + \frac{1}{2}\lambda_n)}{q_n^2}\left[\mathbb{E}(T_n + V_{n+1})\right]^2 + \frac{\lambda_n\mu_n}{q_n^2}\left[\mathbb{E}(T_n + T_{n-1})\right]^2 \\ &\quad - \frac{\lambda_n\mu_n}{q_n^2}\mathbb{E}(T_n + V_{n+1})\mathbb{E}(T_n + T_{n-1}), \qquad (n \geqslant 1), \end{aligned}$$

which, on setting u_n to equal $2^{-n}\lambda_n\pi_n\,\mathrm{Var}(T_n)$, can be rearranged to give

$$\begin{aligned} u_0 &= \frac{(1+\pi_0)^2}{\lambda_0\pi_0} + 2\sum_{r=0}^{\infty}\frac{\pi[r+1]^2}{\lambda_r\pi_r} < \infty \\ u_n &= u_{n-1} + 2^{-n}\sum_{r=n}^{\infty}(1 + I_{(r>n)})\frac{\pi[r+1]^2}{\lambda_r\pi_r} + 4\left(\frac{2^n - \pi[n+1]}{\lambda_n\pi_n} + \frac{2^{n-1} - \pi[n]}{\lambda_{n-1}\pi_{n-1}}\right) \\ &\quad + \frac{2^{1-n}\pi_n}{q_n} + \frac{\mu_n}{q_n}\frac{\pi[n+1]^2}{\lambda_n\pi_n} + \frac{\lambda_n}{q_n}\frac{\pi[n]^2}{\lambda_{n-1}\pi_{n-1}} + \frac{2\pi[n]\pi[n+1]}{\lambda_n\pi_n + \lambda_{n-1}\pi_{n-1}} \quad (n \geqslant 1). \end{aligned}$$

Hence

$$u_n \leqslant u_{n-1} + 2^{1-n}\left(\sum_{r=n}^{\infty}\frac{\pi[r+1]^2}{\lambda_r\pi_r}\right) + 4\left(\frac{2^n}{\lambda_n\pi_n} + \frac{2^{n-1}}{\lambda_{n-1}\pi_{n-1}}\right)$$
$$+ \frac{2^{1-n}\pi_n}{q_n} + \left(\frac{\pi[n+1]^2}{\lambda_n\pi_n}\right) + 3\left(\frac{\pi[n]^2}{\lambda_{n-1}\pi_{n-1}}\right)$$

So, remembering that (1) and (2) hold

$$u_n \leqslant A + B\sum_{r=0}^{n}\frac{2^r}{\lambda_r\pi_r} \leqslant K\sum_{r=0}^{n}\frac{2^r}{\lambda_r\pi_r} \qquad (n \geqslant 0),$$

for some constants A, B, and K. This delivers the required result. □

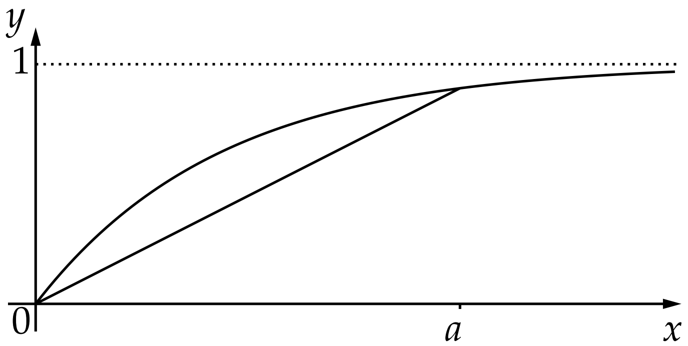


Figure 5.5 The concave function $y = 1 - e^{-x}$ and a subchord

Proof of Lower Bound Lemma 5.5 The function $f(x) = 1 - e^{-x}$ is concave on the bounded interval $[0, a]$ (Fig. 5), so

$$f(x) \geqslant x\left(\frac{1-e^{-a}}{a}\right), \qquad \text{for } x \in [0, a].$$

Now

$$\mathbb{E}\big(f(X)\big) \geqslant \mathbb{E}\big(f(X); X \leqslant a\big) \geqslant \left(\frac{1-e^{-a}}{a}\right)\mathbb{E}(X; x \leqslant a), \tag{5.10}$$

and by Hölder's inequality

$$\mathbb{E}(X; X > a) = \|X\, I_{(X>a)}\|_1 \leqslant \|X\|_2 \mathbb{P}(X > a)^{1/2} \leqslant \big(K\mathbb{E}(X)\mathbb{P}(X > a)\big)^{1/2}.$$

Further $a\mathbb{P}(X > a) \leqslant \mathbb{E}(X)$, so we deduce that

$$\mathbb{E}(X; X > a) \leqslant (K/a)^{1/2}\mathbb{E}(X).$$

Choosing a to be $4K$, then $\mathbb{E}(X; X \leqslant a) \geqslant \frac{1}{2}\mathbb{E}(X)$ and (5.10) becomes

$$\mathbb{E}\big(f(X)\big) \geqslant \left(\frac{1-e^{-4K}}{8K}\right)\mathbb{E}(X). \qquad \square$$

Proof of Lemma 5.7 We can write $d^2(x,y)$ as

$$d^2(x,y) = u_1(x,x) + u_1(y,y) - \mathbb{E}_x(e^{-H(y)})u_1(y,y) - \mathbb{E}_y(e^{-H(x)})u_1(x,x).$$

For the upper bound, we use the fact that $(1-e^{-x}) \leqslant x$ to show that

$$d^2(x,y) \leqslant \left[\sup_{z\in F} u_1(z,z)\right]\left(\mathbb{E}_x\big(H(y)\big) + \mathbb{E}_y\big(H(x)\big)\right).$$

We can split the expected hitting times into four summands, two being sums of (V_n)'s from x and y to $x \wedge y$, and two of (T_n)'s from $x \wedge y$ to x and y. Theorem 5.4 tells us that the largest will be the down time from $x \wedge y$ to y, so

$$d^2(x,y) \leqslant 4\left[\sup_{z\in F} u_1(z,z)\right]\mathbb{E}_{x\wedge y}\big(H(y)\big).$$

For the upper bound we throw away some terms to reveal that

$$\begin{aligned} d^2(x,y) &\geqslant \left[\inf_{z\in F} u_1(z,z)\right]\left(\mathbb{E}_x\big(1-e^{-H(y)}\big)\right) \\ &\geqslant \left[\inf_{z\in F} u_1(z,z)\right]\left(\mathbb{E}_{x\wedge y}\big(1-e^{-H(y)}\big)\right). \end{aligned}$$

The function $u_1(z,z)$ is a continuous positive function on the compact space F, so the sup and the inf are finite and positive. Remembering that $\sum \frac{2^n}{\lambda_n \pi_n}$ is finite, Theorem 5.4 and the Lower Bound Lemma 5.5 together give us that

$$\mathbb{E}_{x\wedge y}\big(1-e^{-H(y)}\big) \geqslant \alpha\mathbb{E}_{x\wedge y}\big(H(y)\big),$$

for some positive α. $\square$

5.4 Time Substitution

We aim to tie the Markov chain on the nodes, F, of the graph together with a Brownian diffusion on the graph, G, comprising of F and the edges. We can construct the diffusion by building up excursions from a point on level n as follows.

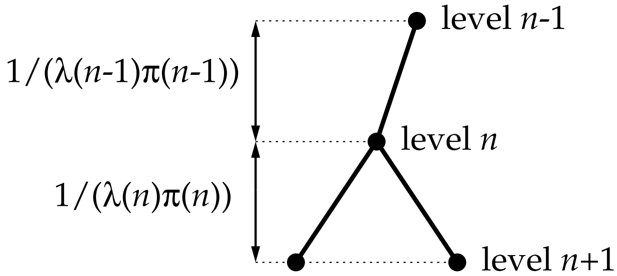


Figure 5.6 Construction of diffusion

(We are only going to define the height and current edge of the process, the horizontal position being thus determined.) Given a reflecting Brownian motion, we take its excursions from 0 in order and make them excursions from our start point. The edge that each excursion follows is randomly selected according to the law assigning probability $\frac{1}{2}$ to going up, and probability $\frac{1}{4}$ to each of the down edges. Run this process until it hits levels $(n-1)$ or $(n+1)$, then repeat starting from the new node. We identify level n with the height $x_n := \sum_{r \geqslant n} \frac{1}{\lambda_r \pi_r}$ (Fig. 5.6).

The height process then becomes a reflecting Brownian motion on the finite interval $[0, \sum_n \frac{1}{\lambda_n \pi_n}]$. We notate the G-valued process as $(\tilde{X}_t)$, and the height process as $(\tilde{Y}_t)$. Then Trotter's Theorem allows us a jointly-continuous local time $\tilde{L}^Y$ for the height process. For a good treatment of local times see V.3 of Blumenthal and Getoor [12]. We can then time change $\tilde{Y}$ via

$$A_t := \sum_n \pi_n \tilde{L}^Y(t, x_n)$$
$$\tau_t := \inf\{s \geqslant 0 : A_s > t\}.$$

We note that A is continuous and (weakly) increasing; τ is right-continuous and strictly increasing; $A(\tau_t) = t$; and $\tau(A_t) \geqslant t$ with equality if and only if t is a point of right increase of A. We time change the diffusion by setting Y_t to be $\tilde{Y}(\tau_t)$, which by III.37 of Williams [39] is a strong Markov process on the support of A ($\{0\} \cup \{x_n : n \in \mathbb{N}\}$).

The local time of $\tilde{Y}$ at a level before it hits an adjacent level (the holding time of the Y-process) is exponentially distributed, by the strong Markov property, and with the right normalisation of $\tilde{L}^Y$, our choice of (x_n) has ensured that the jump rates of Y agree with those of the BD-chain. In fact they are the same process. We can then define the local time of Y, L^Y, on $\mathbb{R}^+ \times \mathbb{N}$ by

$$L^Y(t,n) := \frac{1}{\pi_n}\int_0^t I_n(Y_s)\,ds, \tag{5.11}$$

and notice that by change of variable

$$\tilde{L}^Y(t,x_n) = \frac{1}{\pi_n}\int_0^t I_{x_n}(\tilde{Y}_s)\,dA_s = \frac{1}{\pi_n}\int_{J_+\cap[0,t]} I_{x_n}(\tilde{Y}_s)\,dA_s = L^Y(A_t,n),$$

where J_+ is the set of the points of right-increase of A, which is all but countably many points of the set of points of increase of A.

We can also time change the G-diffusion by τ to produce $X_t := \tilde{X}(\tau_t)$, which is similarly a Markov chain on F with the same jump-rates as the process we studied in previous sections. Consideration of the time-truncation arguments of Rogers and Williams [32] should convince that the processes are the same. Given that conditions (1)–(4) of Theorem 5.2 hold, we can construct a jointly-continuous local time, L^X for X on F. It then follows that $\tilde{L}^X := \{L^X(A_t,x)\}_{x\in F}$ is a local time for $\tilde{X}$ at the points F in G. We can extend $\tilde{L}^X$ by interpolation on the edges to be continuous on G.

It is now possible to construct processes on the boundary, C, via time changes of X and $\tilde{X}$ induced by

$$A^\partial_t := \int_C L^X(t,\xi)\,c(d\xi) = L^Y(t,\infty),$$

$$\text{and} \qquad \tilde{A}^\partial_t := \int_C \tilde{L}^X(t,\xi)\,c(d\xi) = \tilde{L}^Y(t,0),$$

with τ^∂ and $\tilde{\tau}^\partial$ respectively the right-continuous inverses. This gives us the strong Markov R-processes $Z_t := X(\tau^\partial_t)$ and $\tilde{Z}_t := \tilde{X}(\tilde{\tau}^\partial_t)$. By the continuity of the local times,

$$A^\partial(A_t) = \lim_{n\to\infty} 2^{-n}\sum_{x\in\text{level } n} L^X(A_t,x) = \lim_{n\to\infty} 2^{-n}\sum_{x\in\text{level } n} \tilde{L}^X(t,x) = \tilde{A}^\partial_t.$$

For any (A, τ)-type pair, $\tau_t < s \iff t < A_s$, hence

$$\tilde{\tau}^{\partial}_t < s \iff t < \tilde{A}^{\partial}_s \iff t < A^{\partial}(A_s) \iff \tau^{\partial}_t < A_s \iff \tau(\tau^{\partial}_t) < s,$$

whence we can deduce that $\tilde{\tau}^{\partial}_t = \tau(\tau^{\partial}_t)$ and that $Z_t = \tilde{Z}_t$. The process Z also has a jointly-continuous local time L^Z, given by

$$L^Z(t, \xi) = L^X(\tau^{\partial}_t, \xi) = \tilde{L}^X(\tilde{\tau}^{\partial}_t, \xi).$$

In summary we can say that Figure 5.7 commutes. We have thus produced the same process by taking local time on the boundary of both the chain and the diffusion, which allows us to work with whichever is more appropriate for the current problem.

$$\begin{tikzcd}
\tilde{Y} & \tilde{X} \arrow[l, "n"'] \arrow[r, "\tilde{\tau}^{\partial}"] \arrow[d, "\tau"] & \tilde{Z} \arrow[d, equal] \\
Y \arrow[from=u, "\tau"] & X \arrow[l, "n"'] \arrow[r, "\tau^{\partial}"] & Z
\end{tikzcd}$$

Figure 5.7 A commutative diagram of processes

5.5 The Boundary Process

We now finally turn our attention to the boundary process Z. We know that the graph processes (both chain and diffusion) spend no intervals of time on the boundary, but rather the set of times at which they visit the boundary is a Cantor set obtained by removing the open excursion intervals from the time axis. The process will (almost surely) not be back in its original position at the right-hand endpoints of these intervals — even though it will return to its original position uncountably often almost immediately. As the boundary Cantor set is totally disconnected, we see that Z must be a very discontinuous process. In fact Z is discontinuous at a dense, though countable, set of times.

Lemma 5.8 *Z is FD.*

Proof of Lemma 5.8 By adapting the argument at the end of III.38 of Williams [39], we can show that Z is FD if $\mathbb{E}_{\xi}(1 - e^{-\tilde{H}(\eta)})$ goes to 0 as $\eta \to \xi$ in C, where $\tilde{H}(\eta) :=$

$\inf\{t \geqslant 0 : Z_t = \eta\}$. Now $\tilde{H}(\eta)$ is almost surely a point of right-increase of A^∂, so $\tilde{H}(\eta) = A^\partial(H(\eta))$. We can write $H(\eta)$ as

$$H(\eta) \stackrel{\mathcal{D}}{=} U_n := \sum_{r=n}^{\infty}(T_r + V_{r+1}), \qquad \text{where } n = n(\xi \wedge \eta),$$

$$\text{and} \qquad A^\partial(H(\eta)) \stackrel{\mathcal{D}}{=} A^*(U_n) := \sum_{r=n}^{\infty} A_r^*,$$

with A_r^* the local time on the boundary notched up while a version of the process did an up-down $T_r + V_{r+1}$. Then $U_n \downarrow 0$, $A^*(U_n) \downarrow 0$, and thus $(1 - e^{-A^*(U_n)}) \downarrow 0$ as $\eta \to \xi$, giving the result. □

By VI.28 of Rogers and Williams [33], there exists a *Lévy system* (N, H) for Z. In our case $H_t = t$, and N as usual is a *kernel*, that is a function

$$N : (C, \mathcal{B}(C)) \longrightarrow [0, \infty],$$

such that $N(\cdot, \Gamma)$ is $\mathcal{B}(C)$-measurable, for all Γ in $\mathcal{B}(C)$,

and $N(\xi, \cdot)$ is a σ-finite measure on $\mathcal{B}(C)$, for all ξ in C.

In addition $N(\xi, \{\xi\}) = 0$ for all ξ in C, and N has the Lévy property, in the sense that for any non-negative borel-measurable function f on $C \times C$ with $f(\xi, \xi) = 0$ for all ξ in C, then

$$M_t^f := \sum_{s \leqslant t} f(Z_{s-}, Z_s) - \int_{(0,t]} ds \int_C N(Z_{s-}, d\xi) f(Z_{s-}, \xi)$$

is a martingale, if the expectation of either term is finite. We can think of $N(\xi, d\eta)$ as the rate at which jumps from ξ to $d\eta$ of Z occur.

We can calculate this directly using excursion theory, and we will not need any more than is in Rogers [31]. By thinking of the diffusion height process, proposition 2 of Rogers [31] tells us that the rate of excursions from ξ in C to level n or above is

$$\frac{1}{x_n} = \frac{1}{\sum_{r=n}^{\infty} a_r}, \qquad \text{where} \quad a_r := \frac{1}{\lambda_r \pi_r}.$$

(The factor $\frac{1}{2}$ is lost because we have reflection at the boundary so all our excursions go up.) Therefore the rate of excursions from ξ which have their furthest extent on level n is simply the difference

$$\Delta_n := \frac{1}{x_n} - \frac{1}{x_{n-1}} = \begin{cases} \frac{a_{n-1}}{x_n x_{n-1}} & n \geqslant 1 \\ \frac{1}{x_0} & n = 0. \end{cases}$$

The chance that such an excursion ends up in $d\eta \subset C$ is then (by symmetry) exactly $2^n c(d\eta)$ if $\xi \wedge \eta$ is on or below level n, and 0 if not. We deduce that the rate of excursions from ξ to $d\eta$ is given by the following:

Lemma 5.9

$$N(\xi, d\eta) = c(d\eta)\left(\frac{1}{x_0} + \sum_{n=1}^{n(\xi\wedge\eta)} 2\frac{b_{n-1}}{x_n x_{n-1}}\right).$$

Proof of Lemma 5.9 For ξ in C, let ξ_n be the point on level n above ξ, and let A_n be the set $\{\zeta \in C : \zeta > \xi_n\}$ of points of C below ξ_n. For any subset B of $C \backslash A_n$, we notice that $N(\cdot, B)$ is constant on A_n by symmetry. We can set $f(x,y) := I_{A_n}(x) I_B(y)$, and $T := \inf\{t : Z_t \notin A_n\}$. Then $\mathbb{E}(M_T^f) = 0$ implies that $N(\xi, B) = \big(\mathbb{E}(T)\big)^{-1}\mathbb{P}(Z_T \in B)$. In other words, $N(\xi, B)$ is the rate of T-jumping multiplied by the chance that a jump goes into B. all of which is just the rate of jumping into B. So

$$N(\xi, d\eta) = \sum_{n=0}^{n(\xi\wedge\eta)} \Delta_n 2^n c(d\eta),$$

and the result is proved. □

Example In the geometric case, with

$$\lambda_n = \alpha^n, \qquad \mu_n = \alpha^n/\gamma, \qquad \text{and} \quad \pi_n = \Big(\frac{\alpha-\gamma}{\alpha}\Big)(\gamma/\alpha)^n,$$

$$\text{then} \quad \frac{1}{\lambda_n \pi_n} = \frac{\alpha\gamma^{-n}}{\alpha-\gamma}, \qquad x_n = \frac{\alpha\gamma^{-(n-1)}}{(\alpha-\gamma)(\gamma-1)}, \qquad b_n = \frac{\alpha(2/\gamma)^n}{\alpha-\gamma},$$

$$\text{and} \quad N(\xi, d\eta) = c(d\eta)\big(A(2\gamma)^n + B\big) \qquad \text{where } n = n(\xi\wedge\eta);\ A, B > 0.$$

The conditions of Theorem 5.2 translate as (1) $\alpha > \gamma$; (2) $\gamma > 1$; (3)&(4) $\gamma > 2$, and we assume that all these hold. Then H_n, the first time to leave A_n, will be exponentially distributed with rate

$$N(\xi, A_n^c) = \frac{A(\gamma^n - 1)}{2(\gamma-1)} + B(1 - 2^{-n}).$$

We can form an analogue of the walk dimension of a diffusion as

$$\lim_{n\to\infty} \frac{\log \mathbb{E}(H_n)}{-n} = \log\gamma.$$

This can be seen as a measure of the asymptotic neighbourhood escape rate of the process. As γ gets larger it takes longer to escape as the downward pressure inhibits larger excursions. The normal scaling logarithm in the denominator is missing as there is no obviously natural metric on C. □

There is (at least) one easy generalisation of the chain, keeping the same basic structure, by taking the M-ary tree with M down edges, all equally likely, from each node (Fig. 5.8). In the case of $M = 3$, the Ray-Knight compactification can be thought of as a tree-like graph (Fig. 5.9).

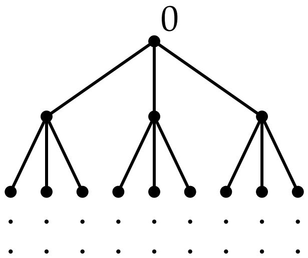


Figure 5.8 Ternary Tree

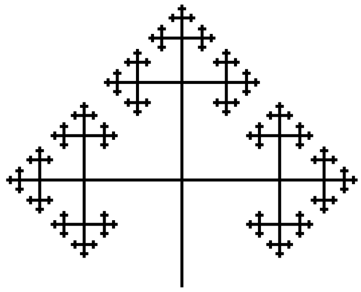


Figure 5.9 Ray-Knight compactification

Everything thus described still holds, with the alteration of the down-jump time line of Theorem 5.4 to

$$\mathbb{E}T_n = \frac{M^{n+1} - \pi[n+1]}{\lambda_n \pi_n}, \quad \mathrm{Var}(T_n) \leqslant K \frac{M^n}{\lambda_n \pi_n} \sum_{r=0}^{n} \frac{M^r}{\lambda_r \pi_r}, \quad \textit{for some } K.$$

And Theorem 5.2 holds with $b_n := M^n/\lambda_n \pi_n$. The Lévy kernel N is as stated above for these new values of (b_n), and the number 2 replaced by M.

Chapter Six

Two-dimensional Local Time

"How often have I said to you that when you
have eliminated the impossible, whatever remains,
however improbable, must be the truth."
— Sherlock Holmes

***Sir Arthur Conan Doyle**, The Sign of Four*

6.1 Introduction

For a general two-dimensional diffusion, it is of course impossible to find a local time at each point. However, we shall manage to construct a diffusion with a continuous local time on a special one-dimensional subspace. We are inspired by the process on the Cantor set in Chapter Five, as viewed as the boundary of the binary tree. In that case, the basic process was a Markov chain; and in our current continuous state-space case, the exact necessary and sufficient conditions are slightly more involved, although we shall develop easily verified sufficient conditions.

We shall work within the upper half-plane $S := \mathbb{R} \times \mathbb{R}^+$ in $\mathbb{R}^2$, and we shall focus our attention on the bottom edge $\mathbb{R}_0 := \mathbb{R} \times \{0\}$, and on the point $\mathbf{0} = (0,0)$ in particular. We define a two-dimensional diffusion $Z_t = (X_t, Y_t)$ by taking Y_t to be a reflecting Brownian motion in $\mathbb{R}^+$, and X_t to be governed by the SDE

$$dX_t = \sigma(Y_t)\,dB_t, \tag{6.1}$$

where B is a Brownian motion independent of Y, and σ is a non-negative function on $\mathbb{R}^+$. As X is a continuous local martingale, we can write it as the time change of a new Brownian motion $\tilde{B}$,

$$X_t = \tilde{B}\Big(\int_0^t \sigma^2(Y_s)\,ds\Big). \tag{6.2}$$

Now the time change $\int_0^t \sigma^2(Y_s)\,ds$ can be written in terms of the local time, $L^Y = L^Y(t,y)$ of Y at y by time t, as $\int_0^\infty \sigma^2(y) L^Y(t,y)\,dy$. Thus a necessary and sufficient condition for X to be well defined by (6.2) is simply that

$$\int_0^K \sigma^2(y)\,dy < \infty, \quad \text{for all finite } K. \tag{6.3}$$

The alert reader might point out that a time change of the process Z would result in its horizontal component being a standard Brownian motion independent of the vertical process Y which is now governed by the SDE

$$dY_t = \frac{dB_t}{\sigma(Y_t)}. \tag{6.4}$$

Despite the simplicity of this representation, our original formulation will prove more tractable, but either case contains the same moral. To make Z return to a point on $\mathbb{R}_0$, either X will have to be moving very quickly, or Y must move very slowly.

We can see a simulation of this process to get an idea of how it behaves. Figure 6.1 shows a particular sample path with $\sigma^2(y) = y^{-0.9}$, starting at $(0, 1)$, and for ease of inspection, we condition the Y co-ordinate to move deterministically towards 0.

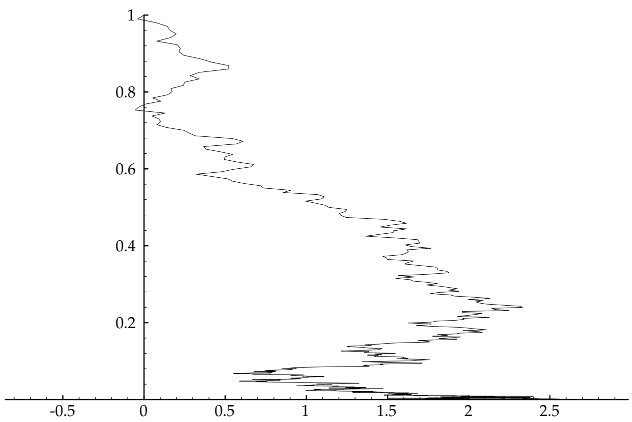


Figure 6.1 The Z process moves towards the x-axis

We can believe from this picture that Z could actually build up a local time around about the point $(2, 0)$ where it hits the x-axis. We can start to prove this rigorously in the next section.

6.2 Calculations

We shall denote the probability transition density of Z by $p_t^Z(w, z)$ for w and z in S, and as everything in sight is reversible, p_t^Z is symmetric with respect to the uniform invariant measure. Recall that a *Lévy process* is a R-process (right-continuous with left limits) with stationary independent increments. If X is a Lévy process started at 0, then

$$\mathbb{E}\exp(i\theta X_t) = \exp\bigl(-t\psi(\theta)\bigr)$$
$$\text{where} \quad \psi(\theta) = ic\theta + \tfrac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}} \left(1 - e^{i\theta y} + \frac{i\theta y}{1 + y^2}\right) \nu(dy),$$

and where $c \in \mathbb{R}$ is the drift, $\sigma \geqslant 0$ is the Brownian component, and ν is a measure on $\mathbb{R}$ such that $\int (1 \wedge y^2)\, \nu(dy) < \infty$ and $\nu(\{0\}) = 0$. Further, if X is purely discontinuous and symmetric ($p_t(0, G) = p_t(0, -G)$) then $c = \sigma = 0$ and ψ is real, so that ψ has Lévy-Khinchine representation

$$\psi(\theta) = \int_{\mathbb{R}} \big(1 - \cos(\theta y)\big)\, \nu(dy).$$

Pausing only to remember that a point x is *regular* if $\mathbb{P}_x(T_x = 0) = 1$, where $T_x := \inf\{t > 0 : X_t = x, \text{ or } X_{t-} = x\}$ is the first time to return to x, we can call on the following condition for regularity:

Theorem 6.1 (Bretagnolle and Kesten) *Let $\tilde{X}$ be a symmetric Lévy process with probability transition density $p_t(x, y) = p_t(0, |x - y|)$, λ-resolvent u_λ, given by $u_\lambda(x, y) := \int_0^\infty e^{-\lambda t} p_t(x, y)\, dt$, which is monotone in $|x - y|$, and cumulant $\chi = \chi(\theta)$ which is real, non-negative and even, such that*

$$\mathbb{E}_0 \exp(i\theta \tilde{X}_t) = \exp(-t\chi(\theta)). \tag{6.5}$$

Then setting $C := \{x : \mathbb{P}_0(T_x < \infty) > 0\}$, exactly one of the following holds:

Either (i) $C = \emptyset$, $u_\lambda(0, 0) = \infty$, $\int_0^\infty (\lambda + \chi(\theta))^{-1}\, d\theta = \infty$ $(\forall \lambda)$, and 0 is not regular,

or (ii) $C = \mathbb{R}$, $u_\lambda(0, 0) < \infty$, $\int_0^\infty (\lambda + \chi(\theta))^{-1}\, d\theta < \infty$, $u_\lambda(0, x) = k_\lambda \mathbb{E}_0(\exp(-\lambda T_x))$ for some finite k_λ $(\forall \lambda)$, and 0 is a regular point.

Proof of Theorem 6.1 From Bretagnolle [13] and Kesten [26]. Note that if 0 is regular, then $\mathbb{P}_0(T_0 < \infty) = 1$, so that $0 \in C$, and so (ii) must hold. We can also replace their condition ($k_\lambda < \infty$) by ($u_\lambda(0, 0) < \infty$), because of the monotonicity of u_λ. This is the result in the more familiar form: if u_λ is bounded, then it is continuous, so the point is regular. Of course, all the points are regular if and only if any one of them is. □

Before we can calculate the resolvent and apply Theorem 6.1, we need to calculate the probability transition density, and our first step is to prove a lemma about the maximum height of a scaled Brownian excursion.

Lemma 6.2 *Let $W = (W_u)_{u=0}^1$ be a scaled Brownian excursion of length 1, let $W^* = \sup_u W_u$, and let δ be positive, then*

$$\mathbb{E}\big((W^*)^\delta\big) < \infty. \tag{6.6}$$

Proof of Lemma 6.2 The distribution of W^* was found explicitly by Kennedy [25], but it is easier for us to use a remark of his, later proved and made intuitive by Vervaat [37], that we can write W^* as $W^* = V^* + (-V)^*$ where $V = (V_u)_{u=0}^1$ is a Brownian bridge with $V_0 = V_1 = 0$. By using the observation that for all positive x and y

$$(x+y)^\delta \leqslant \begin{cases} x^\delta + y^\delta & \text{if } \delta \leqslant 1, \\ 2^{\delta-1}(x^\delta + y^\delta) & \text{if } \delta > 1, \end{cases} \tag{6.7}$$

it is sufficient to show that $\mathbb{E}\big((V^*)^\delta\big)$ is finite. By change of variable, we can write V in terms of a Brownian motion B as

$$V\Big(\frac{1}{1+t}\Big) = \frac{B_t}{1+t} \qquad (t \geqslant 0),$$

and thus $\mathbb{P}(V^* \geqslant c) = \mathbb{P}(\exists t : B_t = c + ct) = \exp(-2c^2)$. So we can calculate $\mathbb{E}\big((V^*)^\delta\big)$ as $2^{-\delta/2}\Gamma(1+\delta/2)$. □

The height process Y has a local time $L^Y(t,y)$, and we set τ to be the right-continuous inverse to $L^Y(t,0)$,

$$\tau_t := \inf\{s : L^Y(s,0) > t\}. \tag{6.8}$$

Then a time-changed process $\tilde{X}_t$ defined by $\tilde{X}_t := X(\tau_t)$ is another Markov process on $\mathbb{R}$. We can calculate the transition density function p_t of $\tilde{X}$, which is also symmetric, as is its 1-resolvent u_1. There are analogues for Z of the following theorems about $\tilde{X}$ which directly prove that Z has a local time on $\mathbb{R}_0$, but we cannot deduce its continuity. Although our path is not the direct route to the local time, we will end up in a stronger position.

Theorem 6.3 *The process $\tilde{X}$ has symmetric transition density function p_t given by*

$$p_t(z,x) = \mathbb{E}\big((2\pi\Sigma_t)^{-\frac{1}{2}} \exp(-(x-z)^2/2\Sigma_t)\big), \tag{6.9}$$

$$\textit{where} \qquad \Sigma_t := \int_0^{\tau_t} \sigma^2(Y_s)\, ds.$$

In addition the functions $p_t(0,x)$ and $u_1(0,x)$ are continuous and monotone for x positive. Further, if $\sigma^2(y) = y^{-\alpha}$ for some α in $(0,1)$, then $u_1(0,0)$ is finite and $u_1(0,x)$ is thus continuous at $x=0$, so each point of $\mathbb{R}$ is regular for $\tilde{X}$.

Proof of Theorem 6.3 Without loss of generality, we can take $z = 0$. Let $\mathcal{G}$ be the filtration induced by $Y(\tau_t)$, that is, $\mathcal{G}_t := \sigma(Y_s : 0 \leqslant s \leqslant \tau_t)$, then Σ_t is $\mathcal{G}_t$-measurable, and $\tilde{X}_t$ conditional on $\mathcal{G}_t$ has a normal distribution with zero mean and variance Σ_t. Thus

$$\frac{1}{\epsilon}\mathbb{P}_0(\tilde{X}_t \in [x, x+\epsilon] \,|\, \mathcal{G}_t) = \frac{1}{\epsilon\sqrt{2\pi\Sigma_t}} \int_x^{x+\epsilon} e^{-y^2/2\Sigma_t}\, dy \,\uparrow\, \frac{e^{-x^2/2\Sigma_t}}{\sqrt{2\pi\Sigma_t}}$$

with monotone convergence on the right as ϵ tends downwards to 0 (for x non-negative). So by the Monotone Convergence theorem the probability density result follows.

As $ue^{-x^2u^2/2} \leqslant e^{-\frac{1}{2}}/x$ for all u, we note that the right-hand side limit above is continuous and monotone in positive x, and dominated by $(2\pi e)^{-\frac{1}{2}}/x$. Thus $p_t(0, x)$ is also continuous and monotone in positive x and dominated by k/x. Dominated Convergence now shows that $u_1(0, x)$ inherits these same properties.

Given that $\sigma^2(y) = y^{-\alpha}$, we need to calculate an upper bound for $(\Sigma_t)^{-\frac{1}{2}}$, so we first need a lower bound for Σ_t itself. We shall begin by remembering some scaling properties of Brownian motion. If Y is our reflecting Brownian motion in $\mathbb{R}^+$, then for any positive c the new process $\tilde{Y}$ given by

$$\tilde{Y}_t := c^{-1} Y_{c^2 t}$$

is another reflecting Brownian motion. The corresponding local time $L^{\tilde{Y}}$ and L^Y, and their inverses $\tilde{\tau}$ and τ are related by

$$L^{\tilde{Y}}(t, y) = c^{-1} L^Y(c^2 t, cy), \quad \text{and} \quad \tilde{\tau}_t = c^{-2}\tau_{ct}.$$

We can see that the corresponding Σ and $\tilde{\Sigma}$ satisfy

$$\tilde{\Sigma}_t = c^{\alpha-2}\Sigma_{ct}.$$

By choosing c to be t^{-1}, we see that

$$\Sigma_t \stackrel{\mathcal{D}}{=} t^{2-\alpha}\Sigma_1. \tag{6.10}$$

We shall now focus on obtaining bounds for Σ_1. Let T be the length of the longest excursion of Y in the interval $[0, \tau_1]$. Now we know (Rogers [31]) that

$$\mathbb{P}(T \leqslant s) = \mathbb{P}(\#\{\text{excursions of length} \geqslant s \text{ in } [0, \tau_1]\} = 0) = \exp(-a/\sqrt{s}),$$

for a constant a. This longest excursion will be distributed as $(\sqrt{T}W(u/T))_{u=0}^{T}$, where W is a scaled Brownian excursion of length 1 independent of T. Thus, because σ is decreasing,

$$\Sigma_1 \geqslant T^{1-\alpha/2}(W^*)^{-\alpha},$$

where W^* is the supremum of W. So, for β positive,

$$\mathbb{E}(\Sigma_1^{-\beta}) \leqslant \mathbb{E}(T^{-\beta+\alpha\beta/2})\mathbb{E}\big((W^*)^{\alpha\beta}\big) < \infty, \tag{6.11}$$

because $\mathbb{E}(T^{-\gamma}) = a^{-2\gamma}\Gamma(1+2\gamma)$ is finite, and all moments of W^* exist by Lemma 6.2. In particular, (6.10) and (6.11) give us the bound that $\mathbb{E}(\Sigma_t^{-\frac{1}{2}}) \leqslant kt^{-1+\alpha/2}$ for some constant k. So $p_t(0,0)$ is dominated by a multiple of $t^{-1+\alpha/2}$, which is integrable with respect to $e^{-t}\,dt$ over $(0,\infty)$. Thus $u_1(0,0)$ is finite, and regularity follows immediately from Theorem 6.1. □

6.3 Local Times

For the present, we take $\sigma^2(y) = y^{-\alpha}$ for some α in $(0,1)$. As in V.3.13 of Blumenthal and Getoor [12], we may now deduce the existence of a local time of $\tilde{X}$, $\tilde{L} = \{\tilde{L}(t,x) : t \in \mathbb{R}^+,\ x \in \mathbb{R}\}$. We shall, similarly to Chapter Five, use theorems in Marcus and Rosen [28] to ensure that $\tilde{L}$ is jointly continuous in time and space. Firstly we have to obtain some bounds on differences of u_1.

Theorem 6.4 *Let $\sigma^2(y) = y^{-\alpha}$ for some α in $(0,1)$, and let d be a metric defined on $\mathbb{R}$ by*

$$d(x,y) := \big(u_1(x,x) + u_1(y,y) - 2u_1(x,y)\big)^{\frac{1}{2}}, \tag{6.12}$$

then, with $\beta := \alpha/(4-2\alpha)$, and for some positive constants c and k, then

$$c|x-y|^{\beta} \leqslant d(x,y) \leqslant k|x-y|^{\beta} \qquad (|x-y| \leqslant 1). \tag{6.13}$$

Proof of Theorem 6.4 As $d^2(x,y) = 2\big(u_1(0,0) - u_1(0,y-x)\big)$, we will calculate $u_1(0,0) - u_1(0,x)$ which, from (6.9) and (6.10) is given by

$$u_1(0,0) - u_1(0,x) = \mathbb{E}\int_0^{\infty} e^{-t}t^{-1+\alpha/2}(2\pi\Sigma_1)^{-\frac{1}{2}}\big(1 - \exp(-t^{\alpha-2}x^2/2\Sigma_1)\big)\,dt.$$

We can break up the integral into two parts over the pair of intervals $[0, x^{2/(2-\alpha)}]$ and $[x^{2/(2-\alpha)}, \infty)$. The former of these has an upper bound of

$$I_1 \leqslant \int_0^{x^{2/(2-\alpha)}} e^0 t^{-1+\alpha/2} \mathbb{E}(2\pi\Sigma_1)^{-\frac{1}{2}}\, dt = kx^{2\beta}.$$

Remembering that $\big(1 - \exp(-y)\big) \leqslant y$, we can bound the latter as

$$I_2 \leqslant \int_{x^{2/(2-\alpha)}}^{\infty} e^0 t^{-3+3\alpha/2} x^2 \mathbb{E}(8\pi\Sigma_1^3)^{-\frac{1}{2}}\, dt = kx^{2\beta},$$

using (6.11) to give the finiteness of $\mathbb{E}(\Sigma_1^{-3/2})$. Hence $d(x, y) \leqslant k|x - y|^{\beta}$, where $\beta = \alpha/(4 - 2\alpha)$. We can find a lower bound for d by merely considering the first integral (taking $x \leqslant 1$) as

$$I_1 \geqslant \int_0^{x^{2/(2-\alpha)}} e^{-1} t^{-1+\alpha/2} \mathbb{E}\big((2\pi\Sigma_1)^{-\frac{1}{2}}(1 - \exp(-1/2\Sigma_1))\big)\, dt \geqslant cx^{2\beta}. \qquad \square$$

We can now use theorems 2 and 8.4 of Marcus and Rosen [28], which say that

Theorem 6.5 (Marcus and Rosen) *If X is a symmetric Markov process on $\mathbb{R}$ with continuous 1-resolvent u_1 and local time $L = \{L(t, x) : t \in \mathbb{R}^+,\ x \in \mathbb{R}\}$, and if there exists a non-negative non-decreasing function $\hat{\sigma}$ such that*

$$d(x, y) \leqslant \hat{\sigma}(|x - y|), \qquad (x, y \in \mathbb{R})$$

$$\textit{and} \qquad \int_0^{\frac{1}{2}} \frac{\hat{\sigma}(z)\, dz}{z(\log(1/z))^{\frac{1}{2}}} < \infty,$$

then the local time L is continuous on $\mathbb{R}^+ \times \mathbb{R}$ almost surely.

We can see immediately that, with $\hat{\sigma}(z) = kz^{\beta}$, the local time $\tilde{L}$ of $\tilde{X}$ is continuous on $\mathbb{R}^+ \times \mathbb{R}$. Alternatively, we could apply the results of Barlow [2] to the Lévy process $\tilde{X}$. Taking his theorems B(a) and 1, we have that

Theorem 6.6 (Barlow) *For a general σ^2 satisfying (6.3), there exists a real function $\chi = \chi(\theta)$ such that*

$$\mathbb{E}\exp(i\theta\tilde{X}_t) = \exp(-t\chi(\theta)), \tag{6.14}$$

and then all points are regular for $\tilde{X}$ if and only if

$$\int_0^{\infty} \frac{d\theta}{1 + \chi(\theta)} < \infty, \tag{6.15}$$

and, if so, then there exists a jointly-continuous version of the local time of $\tilde{X}$ on $\mathbb{R}$ if and only if

$$\text{and} \quad \int_0^{\frac{1}{2}} \frac{\varphi(z)\,dz}{z\sqrt{\log(1/z)}} < \infty, \tag{6.16}$$

$$\text{where} \quad \varphi^2(y) := \int_0^\infty \frac{(1-\cos(\theta y))\,d\theta}{1+\chi(\theta)}. \tag{6.17}$$

Proof of Theorem 6.6 As $\tilde{X}$ is spatially-symmetric, χ is real. In this case the Lévy-Khinchine representation of χ is

$$\chi(\theta) = \int_0^\infty \big(1-\cos(\theta y)\big)\,\nu(dy), \tag{6.18}$$

where ν is a measure on $\mathbb{R}^+$ satisfying $\int(1 \wedge y^2)\,\nu(dy) < \infty$, $\nu(\{0\}) = 0$. The condition (6.15) is condition (0.4) of Barlow [2] and guarantees the existence of a local time at each point. In this case it is also sufficient to give that

$$\int_0^1 y\,\nu(dy) = \infty, \tag{6.19}$$

which is condition (0.5) of Barlow [2], and ensures that the local time is continuous at each fixed point. We can see that this holds, using the fact that $(1-\cos x) \leqslant x$, because

$$\begin{aligned}\chi(\theta) &= \int_0^1 \big(1-\cos(\theta y)\big)\,\nu(dy) + \int_1^\infty \big(1-\cos(\theta y)\big)\,\nu(dy)\\ &\leqslant \theta \int_0^1 y\,\nu(dy) + 2\nu[1,\infty].\end{aligned}$$

As the latter term is finite, if (6.19) did not hold then (6.15) would not either. The necessity of (6.15) comes from Theorem 6.1. Everything else will now follow from Barlow [2], because the function $\varphi(y)$ inherits the monotonicity of $u_1(0,y)$. □

Note that the function $\varphi(y)$ of Theorem 6.6 above is exactly the function $d(0,y)$ of Theorem 6.5. The process Σ is also a subordinator, so $\mathbb{E}\exp(-\theta\Sigma_t) = \exp(-t\psi(\theta))$ for a non-negative function ψ. Remembering that $\tilde{X}_t$ conditional on the height process has a normal distribution with zero mean and variance Σ_t, then

$$\mathbb{E}\exp(i\theta\tilde{X}_t) = \mathbb{E}\exp(-\tfrac{1}{2}\theta^2\Sigma_t) = \exp\big(-t\psi(\tfrac{1}{2}\theta^2)\big).$$

Thus we have an identity between χ and ψ, as $\chi(\theta) = \psi(\frac{1}{2}\theta^2)$. In the case where $\sigma^2(y) = y^{-\alpha}$, then $\psi(\theta) \approx \theta^{1/(2-\alpha)}$, and $\chi(\theta) \approx \theta^{1+2\beta}$ for large θ. The critical cases are going to be when the function σ^2 is slowly varying at 0. Recall that a function f is slowly varying if $f(\lambda x)/f(x)$ tends to 1 as x tends to a limit (usually either 0 or infinity), for each positive λ. We can re-express the necessary and sufficient conditions in this case.

Proposition 6.7 *If σ^2 satisfies (6.3), tends to infinity at 0, and is slowly varying at 0, then there exists a unique (up to asymptotic equivalence) function $l = l(\theta)$, slowly varying at infinity, such that*

$$l^2(\theta)\sigma^2\big(l(\theta)/\sqrt{\theta}\big) \to 1 \qquad \text{as } \theta \to \infty. \tag{6.20}$$

Then $\tilde{X}$ has a local time on $\mathbb{R}$ if and only if

$$\int_1^\infty \frac{l(\theta)\, d\theta}{\theta} < \infty, \tag{6.21}$$

and, if so, then the function $\varphi = \varphi(y)$ of (6.17) is asymptotically equal to

$$\varphi^2(y) \approx \int_{1/y^2}^\infty \frac{l(\theta)\, d\theta}{\theta}, \qquad \text{as } y \to 0, \tag{6.22}$$

where $f(x) \approx g(x)$ means that the lim sup *and the* lim inf *of f/g are finite and positive respectively.*

Proof of Proposition 6.7 From the definition of Σ_t at (6.9) and our expressions for excursions in the proof of Theorem 6.3, we can write a jump of Σ as

$$\triangle\Sigma_t = \triangle\tau_t \int_0^1 \sigma^2\big((\triangle\tau_t)^{\frac{1}{2}} W_s\big)\, ds, \tag{6.23}$$

where W is a scaled Brownian excursion of length 1. Now (1.5.6 of Bingham et al. [10]) for every positive δ (small) and K (large), there exists t_0 such that

$$\frac{\sigma^2(ty)}{\sigma^2(t)} \leqslant y^{-\delta}, \qquad t \leqslant t_0,\ y \in (0, K],$$

which is integrable with respect to dy on $(0, K]$, and the left-hand side goes to 1 as t goes to 0. So, by the Dominated Convergence theorem,

$$\int_0^1 \sigma^2(tW_s)\, ds = \int_0^{W^*} \sigma^2(ty)L(1, y)\, dy \sim \sigma^2(t) \qquad \text{as } t \downarrow 0.$$

Thus $\triangle\Sigma_t \sim (\triangle\tau_t)\sigma^2\big((\triangle\tau_t)^{\frac{1}{2}}\big)$ as $\triangle\tau_t$ gets small. So for large θ, writing μ for the Lévy measure of Σ,

$$\psi(\theta) = \int_0^\infty (1 - e^{-\theta s})\,\mu(ds) \sim \int_0^\infty \big(1 - e^{-\theta t\sigma^2(\sqrt{t})}\big)\frac{dt}{t^{3/2}},$$

because the Lévy measure for $(\triangle\tau_t)$ has density $t^{-3/2}\,dt$. The function $l = l(\theta)$ exists by 1.5.13 of Bingham et al. [10], and if t is set to be $l^2(\theta)/\theta$, then $\theta t\sigma^2(\sqrt{t})$ is asymptotically 1 as θ gets large. So now

$$\begin{aligned}\psi(\theta) &\approx \theta\int_0^{l^2(\theta)/\theta} \sigma^2(\sqrt{t})t^{-\frac{1}{2}}\,dt + \int_{l^2(\theta)/\theta}^\infty \frac{dt}{t^{3/2}}\\ &\approx l(\theta)\sqrt{\theta}\int_0^1 \sigma^2(ul(\theta)/\sqrt{\theta})\,du + \frac{\sqrt{\theta}}{l(\theta)}.\end{aligned}$$

As before $\int_0^1 \sigma^2(ul(\theta)/\sqrt{\theta})\,du$ varies asymptotically with $\sigma^2(l(\theta)/\sqrt{\theta})$ as θ grows large, so by (6.20), $\psi(\theta) \approx \sqrt{\theta}/l(\theta)$, and thus $\chi(\theta) = \psi(\frac{1}{2}\theta^2) \approx \theta/l(\theta^2)$. Condition (6.15) holds if and only if

$$\int_1^\infty \frac{d\theta}{\chi(\theta)} \approx \int_1^\infty \frac{l(\theta^2)\,d\theta}{\theta} = \tfrac{1}{2}\int_1^\infty \frac{l(\theta)\,d\theta}{\theta}$$

is finite. Now we can break up the expression (6.17) for φ into two integrals I_1 and I_2 as

$$\varphi^2(y) = \int_0^{1/y} \frac{\big(1-\cos(\theta y)\big)\,d\theta}{1+\chi(\theta)} + \int_{1/y}^\infty \frac{\big(1-\cos(\theta y)\big)\,d\theta}{1+\chi(\theta)}.$$

Then we can approximate the integrals asymptotically as

$$\begin{aligned}I_1(y) &\approx y^2\int_0^{1/y} \theta l(\theta^2)\,d\theta = \int_0^1 l(u/y^2)\,du \sim l(1/y^2),\\ I_2(y) &\approx \int_{1/y}^\infty \frac{l(\theta^2)\,d\theta}{\theta} = \int_1^\infty \frac{l(\theta/y^2)\,d\theta}{\theta}, \qquad \text{as } y \downarrow 0.\end{aligned}$$

So I_2/I_1 behaves like $\int_1^\infty l(\theta/y^2)/l(1/y^2)\theta^{-1}\,d\theta$ which goes to infinity as y goes to 0. Hence the result follows. □

Let us return for a moment to the binary tree of the discrete setting of Chapter Five. Recalling the model of it introduced at the start of Section 5.4, we have the height process a reflecting Brownian motion observed at the points $x_n := \sum_{r=n}^\infty a_r$, where $a_n := \frac{1}{\lambda_n \pi_n}$. We will embed the discrete process in the plane by making the horizontal distance between adjacent points on level n equal to 2^{-n}. This is the distance necessary to preserve the topology and to put all the points of $[0,1]$ in the boundary. We remember, from (6.2) and (6.4), that σ is merely the ratio of expected infinitessimal X-movements to expected infinitessimal Y-movements. Then, on level n as Y moves a_n, X moves 2^{-n}, so

$$\sigma(x_n) = b_n^{-1}, \qquad \text{where } b_n := \frac{2^n}{\lambda_n \pi_n}.$$

As this is a heuristic derivation, we will make the additional simplifying assumption that b_n is decreasing. Then $a_r \leqslant 2^{n-r} a_n$ $(r \geqslant n)$, so $x_n \leqslant 2a_n$, and thus $a_n \approx x_n$. Setting $t_n := x_n^2$ and $\theta_n := 2^{2n} \approx \big(t_n \sigma^2(\sqrt{t_n})\big)^{-1}$, then $l(\theta_n) = (t_n\theta_n)^{1/2} \approx b_n$. So, by approximating integrals with infinite sums,

$$\int_1^\infty \frac{l(\theta)\, d\theta}{\theta} \approx \sum_{n=0}^\infty \Big(\frac{\theta_{n+1}}{\theta_n} - 1\Big) l(\theta_n) \approx 3 \sum_{n=0}^\infty b_n,$$

$$\text{and} \quad \varphi^2(\theta_n^{-1/2}) = \int_{\theta_n}^\infty \frac{l(\theta)\, d\theta}{\theta} \approx \sum_{r=n}^\infty b_r.$$

$$\text{Thus} \quad \int_0^{1/2} \frac{\varphi(z)\, dz}{z\sqrt{\log(1/z)}} = \sqrt{\log 2} \int_1^\infty \frac{\varphi(2^{-t})\, dt}{\sqrt{t}} \approx \sum_{n=1}^\infty \sqrt{\tfrac{1}{n} \textstyle\sum_{r=n}^\infty b_r}.$$

We see now that the conditions (6.21) and (6.16) for existence and continuity of local times in the continuous case translate back into lines (3) and (4) of Theorem 5.2. This makes the promised connection between the summation conditions of the discrete case and our current application of the known conditions in the continuous case.

We can also perform calculations about the Lévy kernel and walk dimension as we did in Section 5.5, in that:

Proposition 6.8 *The process $\tilde{X}$ has a Lévy kernel N as defined in Section 5.5, such that:*

(i) For σ^2 slowly varying, as in Proposition 6.7,

$$N(0, dx) \approx \frac{dx}{x^2 l(x^{-2})},$$

and (ii) for $\sigma^2(y) = y^{-\alpha}$, for some α in $(0,1)$ and with $\beta = \alpha/2(2-\alpha)$,

$$N(0, dx) \approx \frac{dx}{x^{2+2\beta}}.$$

Then, setting $H_x := \inf\{t > 0 : |\tilde{X}_t| \geqslant x\}$*, the time to exit from the* x*-ball, then* $(\mathbb{E}H_x)^{-1} = \int_x^\infty N(0, dx)$*, and so*

$$\lim_{x \downarrow 0} \frac{\log \mathbb{E}H_x}{\log x} = \begin{cases} 1 & \textit{in case (i),} \\ 1 + 2\beta & \textit{in case (ii).} \end{cases}$$

Proof of Proposition 6.8 If μ is the Lévy measure of Σ, then

$$N(0, dx) = dx \int_0^\infty \frac{e^{-x^2/2s}\mu(ds)}{\sqrt{2\pi s}}.$$

In case (i), we replace s with $t\sigma^2(\sqrt{t})$ and replace $\mu(ds)$ with the Lévy measure for τ, $t^{-3/2}\,dt$. A change of variable to $u = \sqrt{t}/(xl(x^{-2}))$, splitting the resulting integral up into two parts over $(0, 1)$ and $(1, \infty)$, will then finish the calculation as in Proposition 6.7.

In case (ii), we need only note that $\triangle\Sigma \sim (\triangle\tau)^{1-\alpha/2}$ for small increments, so that $\mu(ds) \sim s^{-(3/2+\beta)}\,ds$. The rest is easy calculus. $\square$

Our final comment on the continuity of local times of $\tilde{X}$ before we move back to looking at the original Z process, is to find sufficient conditions purely in terms of σ itself:

Corollary 6.9 (Sufficient conditions) *If* σ^2 *satisfies (6.3), and is slowly varying tending to infinity at 0, and further satisfies*

$$\frac{\sigma\big(t\sigma(t)\big)}{\sigma(t)} \to 1 \qquad \textit{as } t \to 0, \tag{6.24}$$

then $\tilde{X}$ *has a local time on* $\mathbb{R}$ *if and only if*

$$\int_0^1 \frac{dt}{t\sigma(t)} < \infty, \tag{6.25}$$

and also the function φ *of (6.17) behaves asymptotically as*

$$\varphi^2(y) \approx \int_0^y \frac{dt}{t\sigma(t)}, \qquad \textit{as } y \to 0. \tag{6.26}$$

Proof of Corollary 6.9 Writing (6.24) as $\sigma(\sqrt{t}\sigma(\sqrt{t})) \sim \sigma(\sqrt{t})$ as t goes to 0, and setting $t = l^2(\theta)/\theta$, where l is as at (6.20), then $\theta \sim (t\sigma^2(\sqrt{t}))^{-1}$ and so

$$l(\theta) \sim \frac{1}{\sigma(\sqrt{t})} \sim \frac{1}{\sigma(1/\sqrt{\theta})} \qquad \text{as } \theta \to \infty.$$

Thus

$$\int_1^\infty \frac{l(\theta)\,d\theta}{\theta} \approx \int_1^\infty \frac{d\theta}{\theta\sigma(1/\sqrt{\theta})} = 2\int_0^1 \frac{dt}{t\sigma(t)},$$

and also

$$\varphi^2(y) \approx \int_{1/y^2}^\infty \frac{l(\theta)\,d\theta}{\theta} \sim 2\int_0^y \frac{dt}{t\sigma(t)}, \qquad \text{as } y \to 0.$$

Thus the result is proved. □

Example In the case where $\sigma^2(y) = \big(\log(1/y)\big)^{2(1+\delta)}$, then $l(\theta) \approx (\log\theta)^{-(1+\delta)}$, and (6.21) and (6.25) hold if and only if δ is greater than 0. Then the function $\varphi(y) \approx \big(\log(1/y)\big)^{-\delta/2}$, and (6.16) will hold if and only if δ is greater than 1.

Armed with a jointly-continuous local time for $\tilde{X}$, we can now proceed to find one for Z itself.

Theorem 6.10 *If $\tilde{L}$ exists and is continuous, then the process $Z_t = (X_t, Y_t)$ has a local time L^Z on $\mathbb{R}_0 := \mathbb{R}\times\{0\}$ given by*

$$L^Z(t,(x,0)) = \tilde{L}(L^Y(t,0),x), \tag{6.27}$$

and thus L^Z is jointly continuous on $\mathbb{R}^+ \times \mathbb{R}_0$.

Proof of Theorem 6.10 Define $L(t,x) := \tilde{L}(L^Y(t,0),x)$ which is a continuous function, increasing in t for each x and only increases when $Y_t = 0$ and $X_t = x$. Recall from Blumenthal and Getoor [12] that the local time of a process X at a point x is the unique (up to normalisation) continuous increasing function $l(t)$ such that $l(0) = 0$, l increases at t only if $X_t = x$, and is additively Markovian in the sense that

$$l(t+s,\omega) = l(t,\omega) + l(s,\theta_t\omega),$$

where θ_t is the shift operator defined by $X_u(\theta_t\omega) = X_{u+t}(\omega)$. In our case

$$\tilde{L}(t+s,x,\omega) = \tilde{L}(t,x,\omega) + \tilde{L}(s,x,\theta(\tau_t)\omega),$$

$$\text{and}\qquad L^Y(t+s,0,\omega) = L^Y(t,0,\omega) + L^Y(s,0,\theta_t\omega).$$

Thus, because in addition $\tilde{L}(s,x,\theta_u\omega)$ is constant for u in any interval of the form $[\tau_{t-},\tau_t]$, we can write $L(t+s,x,\omega)$ as $L(t,x,\omega) + L(s,x,\theta_t\omega)$. So L is indeed the local time of Z on $\mathbb{R}_0$. □

We complete this section with a collection of results about Z and L^Z which are now straightforward.

Theorem 6.11 *Let $L^Z = \{L^Z(t,(x,0)) : t \in \mathbb{R}^+,\ x \in \mathbb{R}\}$ be a continuous version of the local time of Z on $\mathbb{R}_0$. Then*

(a)
$$\int_{\mathbb{R}} L^Z(t,(x,0))\,dx = L^Y(t,0) \qquad (t \geqslant 0). \tag{6.28}$$

(b) (Occupation density formula) For g a bounded Borel measurable function,

$$\int_0^t g(X_s)\,L^Y(ds,0) = \int_{\mathbb{R}} g(x)L^Z(t,(x,0))\,dx \qquad (t \geqslant 0). \tag{6.29}$$

(c) If $\tilde{X}$ is recurrent then Z_t visits all the points in K by some finite time, for each compact K in $\mathbb{R}_0$.

Proof of Theorem 6.11

(a) This is immediate from (6.27) and the fact that $\int_{\mathbb{R}} \tilde{L}(t,x)\,dx = t$.

(b) For almost all times s, if $L^Y(ds,0) > 0$ then $\tau(L^Y(s,0)) = s$. Thus

$$\int_0^t g(X_s)L^Y(ds,0) = \int_0^t g\Big(X\big(\tau(L^Y(s,0))\big)\Big)L^Y(ds,0)$$

$$= \int_0^{L^Y(t,0)} g(\tilde{X}_u)\,du = \int_{\mathbb{R}} g(x)\tilde{L}(L^Y(t,0),x)\,dx = \int_{\mathbb{R}} g(x)L^Z(t,(x,0))\,dx.$$

(c) As the local time is continuous, compactness arguments will give this result immediately as long as $\tilde{X}$ is recurrent (see the next Section). □

6.4 Recurrence

We are working here with what is sometimes called 'point-recurrence' rather than 'interval-recurrence'. Point-recurrence means simply that the process will hit points ($\mathbb{P}_0(T_x < \infty) = 1$), whilst interval-recurrence merely means that all neighbourhoods of a particular point will eventually be hit.

The representations used in Barlow [2], can be equally effective in producing necessary and sufficient conditions for the process $\tilde{X}$ to be recurrent. In essence, whereas

σ had to get large at 0 quickly enough for a local time to exist and be continuous, σ must either stay bounded or not get large too quickly at infinity for the process to be recurrent. Intuitively, if σ is large at infinity, then when Y goes on a large excursion, X moves a very long way away and may never return. We can produce a composite theorem which is an analogue of Theorem 6.6, Proposition 6.7 and Corollary 6.9. In our case, point-recurrence implies regularity as the set C of Theorem 6.1 is equal to $\mathbb{R}$ if the process hits points.

Theorem 6.12 *For a general σ^2 satisfying (6.3), there exists a real function $\chi = \chi(\theta)$ such that*

$$\mathbb{E}\exp(i\theta\tilde{X}_t) = \exp(-t\chi(\theta)).$$

Then the process $\tilde{X}$ is (point-)recurrent if and only if both all points are regular for it, and

$$\int_0^1 \frac{d\theta}{\chi(\theta)} = \infty. \tag{6.30}$$

Further, if σ^2 is slowly varying at infinity, then there exists a unique (up to asymptotic equivalence) function $h = h(\theta)$, slowly varying at 0, such that

$$h^2(\theta)\sigma^2\big(h(\theta)/\sqrt{\theta}\big) \to 1 \qquad \text{as } \theta \to 0. \tag{6.31}$$

Then $\tilde{X}$ is recurrent if and only if

$$\int_0^1 \frac{h(\theta)\,d\theta}{\theta} = \infty. \tag{6.32}$$

Further, if σ also satisfies

$$\frac{\sigma\big(t\sigma(t)\big)}{\sigma(t)} \to 1, \qquad \text{as } t \to \infty, \tag{6.33}$$

then $\tilde{X}$ is recurrent if and only if

$$\int_1^\infty \frac{dt}{t\sigma(t)} = \infty. \tag{6.34}$$

Proof of Theorem 6.12 From Theorem 6.1, $\mathbb{E}_0(\exp(-\lambda T_x)$ is proportional to $u_\lambda(0,x)$ as x varies, so

$$\mathbb{E}_0\big(e^{-\lambda T(x)}\big) = \frac{u_\lambda(0,x)}{u_\lambda(0,0)},$$

and so $\tilde{X}$ is recurrent if and only if $u_\lambda(0,x) \sim u_\lambda(0,0)$ as λ goes to 0, for all x. Now, by (0.10) of Barlow [2], as $\tilde{X}$ is symmetric

$$u_\lambda(0,x) = \frac{1}{\pi}\int_0^\infty \frac{\cos(\theta x)\, d\theta}{\lambda + \chi(\theta)},$$

so as λ tends down to 0,

$$u_\lambda(0,0) - u_\lambda(0,x) = \frac{1}{\pi}\int_0^\infty \frac{\big(1-\cos(\theta x)\big)\, d\theta}{\lambda + \chi(\theta)} \uparrow \frac{1}{\pi}\int_0^\infty \frac{\big(1-\cos(\theta x)\big)\, d\theta}{\chi(\theta)}. \qquad (6.35)$$

We want to show that the right hand side of the above line is finite. Firstly we note that $1-\cos(x)$ is of size x^2 for small x. Precisely we have that $ax^2 \leqslant 1-\cos(x) \leqslant \frac{1}{2}x^2$ for all x in $[0,1]$, for some a. Thus, for $\theta \leqslant 1$, from (6.18)

$$\chi(\theta) \geqslant a\theta^2 \int_0^1 y^2\, \nu(dy) \geqslant a_1\theta^2,$$

for some constant a_1, and so

$$\frac{1-\cos(\theta x)}{\chi(\theta)} \leqslant \frac{x^2}{2\alpha_1} \quad \text{for } \theta \leqslant (1 \wedge x^{-1}).$$

Finally, as χ is increasing, we have for $\theta \geqslant 1$ that

$$\frac{1-\cos(\theta x)}{\chi(\theta)} \leqslant \frac{4}{a_2(1+\chi(\theta))},$$

where a_2 is the constant $(1 \wedge \chi(1))$. As $\tilde{X}$ has a local time, (6.15) holds, so we can deduce that the integral on the right-hand side of (6.35) is finite. Now $u_\lambda(0,0) \uparrow u_0(0,0) = \int_0^\infty \chi^{-1}(\theta)\, d\theta$, so if (6.30) holds, then this is infinite and the ratio of $u_\lambda(0,x)$ to $u_\lambda(0,0)$ indeed tends to 1. Conversely if (6.30) does not hold, then that and (6.15) together show that $u_0(0,0)$ is finite, and so the ratio tends to a limit strictly less than 1.

To achieve (6.32), we now aim to repeat our calculations of Proposition 6.7. Again, by 1.5.6 of Bingham et al. [10], for every δ (small), there exists T_0 (large) such that

$$\frac{\sigma^2(u)}{\sigma^2(v)} \leqslant (u/v)^\delta \vee (v/u)^\delta, \qquad u, v \geqslant T_0.$$

Then

$$\frac{\sigma^2(ty)}{\sigma^2(t)} I(ty \geqslant T_0) \leqslant (y^\delta \vee y^{-\delta}), \qquad t \geqslant T_0,$$

which is integrable with respect to dy on $(0, W^*]$, and the left-hand side goes to 1 as t goes to infinity. Also

$$\int_0^{T_0/t} \frac{\sigma^2(ty)}{\sigma^2(t)}\,dy = \frac{1}{t\sigma^2(t)} \int_0^{T_0} \sigma^2(u)\,du,$$

which goes to 0 as t goes to infinity. Hence we can deduce that

$$\int_0^1 \sigma^2(tW_s)\,ds = \int_0^{W^*} \sigma^2(ty)L(1,y)\,dy \sim \sigma^2(t) \qquad \text{as } t \to \infty,$$

and so, by (6.23), $\triangle\Sigma_t \sim (\triangle\tau_t)\sigma^2\big((\triangle\tau_t)^{\frac{1}{2}}\big)$ as $\triangle\tau_t$ gets large. We decompose the representation of ψ into three integrals as (for K large)

$$\psi(\theta) = \int_0^K (1 - e^{-\theta s})\,\mu(ds) + \int_K^{h^2(\theta)/\theta} (1 - e^{-\theta s})\,\mu(ds) + \int_{h^2(\theta)/\theta}^{\infty} (1 - e^{-\theta s})\,\mu(ds),$$

which we denote by I_1, I_2 and I_3. For small θ

$$I_1 \leqslant \theta \int_0^K s\,\mu(ds),$$

$$I_2 \approx \theta \int_K^{h^2(\theta)/\theta} t\sigma^2(\sqrt{t})\frac{dt}{t^{3/2}} = \sqrt{\theta}h(\theta) \int_{\sqrt{K\theta}/h(\theta)}^1 \sigma^2(uh(\theta)/\sqrt{\theta})\,du \approx \sqrt{\theta}/h(\theta),$$

$$I_3 \approx \int_{h^2(\theta)/\theta}^{\infty} \mu(ds) = \sqrt{\theta}/h(\theta).$$

So $\psi(\theta) \approx \sqrt{\theta}/h(\theta)$ for small θ, and so $\chi(\theta) \approx \theta/h(\theta^2)$. Thus (6.30) holds if and only if

$$\int_0^1 \frac{h(\theta^2)\,d\theta}{\theta} = \tfrac{1}{2} \int_0^1 \frac{h(\theta)\,d\theta}{\theta} = \infty.$$

Finally, if (6.33) holds, then $h(\theta) \sim \sigma^{-1}(1/\sqrt{\theta})$ for small θ and (6.34) follows as in Corollary 6.9. □

We note that condition (6.30) is exactly the necessary and sufficient condition of theorem 16.2 of Port and Stone [30] required for the process to be interval-recurrent. Thus, the first part of our theorem simply asserts that the process is point-recurrent if and only if it is both interval-recurrent and regular. Intuitively, backed up by (6.15) and (6.30), regularity is a local property depending on small jumps (σ near zero) and the behaviour of χ at infinity, independent of interval-recurrence which is a non-local property depending on long jumps (σ near infinity) and the behaviour of χ at zero.

References

Knowledge is of two kinds. We know
a subject ourselves, or we know where
we can find information upon it.

Dr Samuel Johnson

Atque inter silvas Academi quaerere verum

***Horace**, Epistles*

[1] A. DE ACOSTA. Large deviations for vector-valued functionals of a Markov chain: lower bounds. *Ann. Probab.* **16** (1988), 925–960.

[2] M. T. BARLOW. Necessary and sufficient conditions for the continuity of local time of Lévy processes. *Ann. Probab.* **16** (1988), 1389–1427.

[3] M. W. BAXTER and D. WILLIAMS. Symmetry characterizations of certain distributions, 1. *Math. Proc. Cambridge Philos. Soc.* **111** (1992), 387–397.

[4] M. W. BAXTER and D. WILLIAMS Symmetry characterizations of certain distributions, 2: Discounted additive functionals and large deviations. *Math. Proc. Cambridge Philos. Soc.* **112** (1992), 599–611.

[5] M. W. BAXTER. Symmetry characterizations of certain distributions, 3: Discounted additive functionals and large deviations for a general finite-state Markov chain. *Math. Proc. Cambridge Philos. Soc.* **113** (1993), 381–386.

[6] M. W. BAXTER. Markov processes on the boundary of the binary tree. In *Séminaire de Probabilités XXVI* (ed. J.Azéma, P.A.Meyer & M.Yor), Lecture Notes in Mathematics 1526 (Springer-Verlag, 1993), pp. 210–224.

[7] M. W. BAXTER. The local time of a two-dimensional diffusion. *Proc. Roy. Soc. Lond.* A (submitted).

[8] J. BERNOULLI. *Ars Conjectandi* (Basel, 1713).

[9] P. BILLINGSLEY. *Convergence of Probability Measures* (Wiley, 1968).

[10] N. H. BINGHAM, C. M. GOLDIE and J. L. TEUGELS. *Regular Variation*, Encyclopedia of Mathematics and its Applications, vol. 27 (Cambridge University Press, 1987).

[11] N. H. BINGHAM and L. C. G. ROGERS. Summability methods and almost sure convergence. In *Almost Everywhere Convergence II: Proc. Second Intnl. Conf., Evanston Ill.*, ed. A. Bellow and R. L. Jones (Academic Press 1991), pp. 69–83.

[12] R. M. BLUMENTHAL and R. K. GETOOR. *Markov Processes and Potential Theory* (Academic Press 1968).

[13] J. BRETAGNOLLE. Resultats de Kesten sur les processus à accroissements indépendants. In *Séminaire de Probabilités V* (ed. A.Dold & B.Eckmann), Lecture Notes in Mathematics 191 (Springer-Verlag, 1971), pp. 21-36.

[14] M. F. CHEN and Y. G. LU. Large deviations for Markov chains. *Acta Math. Sci.* **10** (1990), 217–228.

[15] J-D. DEUSCHEL and D. W. STROOCK. *Large deviations* (Academic Press, 1989).

[16] M. D. DONSKER and S. R. S. VARADHAN. Asymptotic evaluation of certain Markov process expectations for large time: I–IV. *Comm. Pure Appl. Math.* **28** (1975), 1–47 and 279–301; **29** (1976), 389–461; **36** (1983), 183–212.

[17] M. S. P. EASTHAM. *The asymptotic solution of linear differential systems* (Oxford University Press, 1989).

[18] R. S. ELLIS. Large deviations for a general class of random vectors. *Ann. Probab.* **12** (1984), 1–12.

[19] S. FANG. Grandes déviations pour le processus d'Ornstein Uhlenbeck. *C. R. Acad. Sci. Paris, Série I: Math.* **314** (1992), 291–294.

[20] G. GIGERENZER, Z. SWIJTINK, T. PORTER, L. DASTON, J. BEATTY, and L. KRÜGER. *The Empire of Chance* (Cambridge University Press, 1989).

[21] G. H. HARDY. On the zeroes of certain classes of integral Taylor series. Part II: On the integral function

$$\sum_{n=0}^{\infty} \frac{x^n}{(n+a)^s n!}$$

and other similar functions. *Proc. London Math. Soc.* (2), **2** (1905), 401–431.

[22] K. ITÔ and H. P. MCKEAN. *Diffusion Processes and their Sample Paths* (Springer-Verlag, 1965).

[23] N. C. JAIN. Large deviation lower bounds for additive functionals of Markov processes. *Ann. Probab.* **18** (1990), 1071–1098.

[24] T. KATO. *A Short Introduction to Perturbation Theory for Linear Operators* (Springer-Verlag, 1982).

[25] D. P. KENNEDY. The distribution of the maximum Brownian excursion. *J. Appl. Probab.* **13** (1976), 371–376.

[26] H. KESTEN. Hitting probabilities of single points for processes with stationary independent increments. *Mem. Amer. Math. Soc.* **109** (1969).

[27] Y. KIFER. Averaging in Dynamical Systems and Large Deviations. *Inventiones Math.* (preprint)

[28] M. B. MARCUS and J. ROSEN. Sample Path Properties of the Local Times of Strongly Symmetric Markov Processes via Gaussian Processes. *Ann. Probab.* **20** (1992), 1603–1684.

[29] E. NUMMELIN. Large deviations for functionals of stationary processes. *Probab. Th. Rel. Fields* **86** (1990), 387–401.

[30] S. C. PORT and C. J. STONE. Infinitely divisible processes and their potential theory. *Ann. Inst. Fourier* **21** (1971), 2:157–275 and 4:179-267.

[31] L. C. G. ROGERS. A Guided Tour through Excursions. *Bull. London Math. Soc.* **21** (1989), 305–341.

[32] L. C. G. ROGERS and D. WILLIAMS. Construction and Approximation of Transition Matrix Functions. *Analytic and Geometric Stochastics*, suppl. to *Adv. Appl. Probab.* **18** (1986), 133–160.

[33] L. C. G. ROGERS and D. WILLIAMS. *Diffusions, Markov Processes and Martingales, Vol.2: Itô calculus* (Wiley, 1987).

[34] E. SENETA. *Non-negative Matrices, an Introduction to Theory and Applications* (Allen & Unwin, 1973).

[35] E. C. TITCHMARSH. *The Theory of Functions* (2nd edition), (Oxford University Press, 1939).

[36] S. R. S. VARADHAN. *Large deviations and applications* (SIAM, 1984).

[37] W. VERVAAT. A relation between Brownian bridge and Brownian excursion. *Ann. Probab.* **7** (1979), 143–149.

[38] D. V. WIDDER. *An Introduction to Transform Theory* (Academic Press, 1971).

[39] D. WILLIAMS. *Diffusions, Markov Processes and Martingales, Vol.1: Foundations* (Wiley, 1979).

[40] D. WILLIAMS. *Probability with Martingales* (Cambridge University Press, 1991).

[41] R. W. WOLFF. *Stochastic Modelling and the Theory of Queues* (Prentice-Hall 1989).