Some Aspects of Tree-Indexed Processes

By

Marcus Sheehan

Thesis

Submitted to the University of Warwick

For the degree of

Doctor of Philosophy

Department of Statistics

May 2006

THE UNIVERSITY OF
WARWICK

PAGE NUMBERING AS ORIGINAL

Contents

0.1 Acknowledgements

I would like to thank my supervisors Jon Warren and Saul Jacka for their constant support and guidance, particularly Saul for agreeing to be my co-supervisor. I would also like to acknowledge the financial support provided by the EPSRC. I would like to thank my parents and grandparents for their encouragement throughout my PhD work and the Department of Statistics at Warwick for enabling me to make the best of this opportunity. Finally, I would like to thank Carol Wright for her support, encouragement and kind-heartedness.

0.2 Declaration

I declare that this thesis is based on my own research in accordance with the rules and regulations of the University of Warwick. The work is original except where indicated by specific references. This thesis has not been submitted for examination at any other university.

Abstract

Chapters 2 and 3 of this thesis are based on a paper in preparation, "Partial observations of a tree-indexed process". We begin, in Chapter 1, with an introduction to the theory of branching random walks. We then consider stochastic process indexed by Galton-Watson trees and highlight the connection between "speeds" for the corresponding branching random walks and large deviation theory. We relate all of this to a class of martingales obtained via a change of measure corresponding to changing the distribution of the tree-indexed random variables along a randomly chosen "line of descent" through the tree. We show that the issue of uniform integrability for these martingales - a much studied subject in its own right - boils down essentially to large deviation calculations.

In Chapter 2 we introduce "recovery problems". We observe some information about a tree-indexed collection of random variables and ask if it is possible to "recover" the original random variables in some suitable sense. We motivate this with a simple example on the integers and then describe the analogous problem on the binary tree, using the theory developed in Chapter 1 to derive conditions on the underlying probability parameters under which recovery is possible.

In Chapter 3 we turn our attention to "Recursive Distributional Equations" (or RDEs). Motivated by the work of Chapter 2 we investigate a recursion for random variables on the binary tree and then study the corresponding RDE in its own right by introducing the idea of tree-indexed solutions. We are able to give a fairly complete analysis of the RDE in the case where recovery of the tree-indexed random variables from Chapter 2 is possible but find that the non-recovery case is less tractable. Here we provide partial

results and make conjectures based on what we know to be true.

In Chapter 4 we further develop the theory of RDEs by introducing the notion of "endogeny". This relates to whether (tree-indexed) solutions to RDEs can be written as functions of the original data alone or whether there is some additional randomness coming from the system. We conclude the chapter with a particular example, the so-called "noisy veto voter model", and obtain conditions for endogeny in this setting by extending some recent work in this area.

Chapter 5 is based on a paper in preparation, "A recursive distributional equation on $[0, 1]$". The RDE in question is obtained from the noisy veto voter RDE of Chapter 4 via various transformations and conditioning. We make a thorough study of this new RDE by identifying all invariant distributions, the corresponding "basins of attraction" and addressing the issue of endogeny for the associated tree-indexed problem.

At the end of each chapter we discuss briefly possible extensions to the work and make clear any unresolved issues or simplifying assumptions that might be relaxed in future research.

Chapter 1

Large deviations and martingales

In this chapter we introduce the idea of a branching process, focusing particularly on multi-type Galton-Watson branching processes and the underlying tree structure. We discuss the Binary tree with "Bernoulli noise" as much of our work will focus on this specific example. We then study the boundary of the binary tree via large deviation calculations and relate this to martingales coming from the associated branching random walk.

1.1 Galton-Watson branching processes and trees

Let N be a non-negative integer-valued random variable. Set $Z_0 = 1, Z_1 = N$ and, inductively,

$$Z_{n+1} = \sum_{i=1}^{Z_n} N_i^{n+1},$$

where the N_i^n are independent copies of N. Then we say that Z_n is the number of individuals in generation n of a Galton-Watson branching process. The population starts with one individual and then each individual independently produces a random number of offspring according to N. The collection of all individuals forms the vertices of the associated Galton-Watson tree, with edges connecting parents to their offspring. It will be convenient, especially in the later context of the binary tree, to elucidate a system of labelling the vertices of such a tree. First consider a Galton-Watson tree with deterministic branching factor N, which we can write as

$$\Gamma = \bigcup_{n=0}^{\infty} \{0, 1, ..., N-1\}^n,$$

where $\emptyset \in \{0, ..., N-1\}^0$ denotes the "root". In this notation, the labels $0, 1, ..., N-1$ represent the offspring of an individual. The collection of vertices in level n of the tree is then denoted G_n and may be represented in terms of "addresses" as points in $\{0, 1, ..., N-1\}^n$. We define Γ_n to be

$$\Gamma_n = \bigcup_{k \leq n} \{0, 1, ..., N-1\}^k,$$

the set of vertices belonging to levels up to and including n. We write $|u|$ for the depth of a vertex u, so that $u \in G_n$ if and only if $|u| = n$. We denote the boundary (or limit set) of Γ by $\partial\Gamma = \{0, 1, ..., N-1\}^{\mathbb{Z}^+}$. Corresponding to each vertex $u \in \Gamma$ is a sequence of length $|u|$ with entries in $\{0, 1, ..., N-1\}$, which is its "address" in level $|u|$ of the tree. For example, the third daughter of the second daughter of the root (ordered from left to right say) is denoted (12). Writing $u = (u_1, u_2, ..., u_n)$ for a vertex in level n, we denote the concatenated sequence $(u_1, u_2, ..., u_n, j)$, where $j \in \{0, 1, ..., N-1\}$, corresponding to the possible daughters of u, by uj. A vertex, together with its daughter vertices, is called a "family". We say that a vertex v is an ancestor of a vertex u, written $v \leq u$, if there is a sequence of vertices $(w_1 = v, ..., w_n = u)$ with w_{i+1} a daughter of w_i for $i = 1, ..., n-1$. For every point $t \in \partial\Gamma$, there is an infinite sequence of vertices $(w_1, w_2, ..., w_n, ...)$, with $w_n \in G_n$ and w_{n+1} a daughter of w_n for every n, starting at the root and "ending" at t, listing all the ancestors of t. We call this an "infinite line of descent".

Note that in some work, such sequences are referred to as "spines" or "rays". A line of descent from an arbitrary vertex u is just a sequence of vertices $(w_1 = u, w_2, ..., w_n, ...)$, with w_{n+1} a daughter of w_n for every n and a line of descent between two vertices is defined in the obvious way. In the more general setting where N is a random variable, we embed the resulting Galton-Watson tree in an N^*-ary tree, where $N^* = \sup_\omega N(\omega)$ and adopt the same notation as for the deterministic tree (though obviously there may be some "ghost" vertices that don't actually exist). Note that if N has positive probability of being infinite then we embed in an infinite tree with infinite branching factor. This approach will be used in Chapter 5.

We now introduce the notion of "vertex type". It is assumed throughout that we are working on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\xi = (\xi_u; u \in \Gamma)$ be a stochastic process indexed by the vertices of Γ. We will assume that, under $\mathbb{P}$, the ξ_u are independent, identically distributed real random variables. The process ξ is an example of a "multi-type Galton-Watson branching process", in which a "mark" or "noise" is associated with each vertex. In the context of the associated random walk, one can also think of the marks as being "displacements" from parent vertices. This fairly general setup is standard in the sense that much work has been done in this setting. We will be particularly interested in the deterministic binary tree T, that is, the case where $N \equiv 2$. For us, the associated marks will be independent, identically distributed Bern(p) random variables on the two point set $\{-1, 1\}$. Key to this is the assumption of independence between the random variables associated with the vertices. A generalisation of this structure will be discussed in Chapter 2.

1.2 Large deviations and rate functions for random walks

In this section we consider large deviations for random walks via rate functions. This is an important concept that will play a major role in our later work on tree-indexed random walks.

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables defined on some common probability space and write μ for their common mean and σ^2 for their common variance (both of which are assumed finite). Let $S_n = X_1 + \ldots + X_n$ be the associated partial sum. By interpreting the X_i as displacements, we can think of S_n as being the position at time n of a random walk with independent, identically distributed increments distributed as the X_i. The Central Limit Theorem asserts that

$$\frac{S_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to a standard Normal random variable. This "quantifies" the probability of S_n differing from its mean μn by an amount of order $\sqrt{n}$. Large deviation theory deals with events where S_n differs from its mean μn by an amount of order n, that is to say, events outside the scope of the Central Limit Theorem. [24]. Of particular interest is the event

$$S_n \geq an, \qquad a > 0,$$

whose probability tends to zero as $n \to \infty$, and the "rate" at which this convergence occurs. Under certain conditions (see Cramer's Theorem below), the decay is exponential in n in that, for $a > 0$,

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(S_n \geq an)}{n} < 0,$$

where the quantity on the left is called the rate function. Cramer's Theorem makes this more precise.

Theorem 1. *(Cramer) Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with a well-defined moment generating function, that is, satisfying*

$$\varphi(t) = \mathbb{E}[e^{tX_i}] < \infty$$

for every $t \in \mathbb{R}$. Then

$$I(a) = \begin{cases} -\lim_{n\to\infty} \frac{\log \mathbb{P}(S_n \geq an)}{n}, & a \geq \mu \\ -\lim_{n\to\infty} \frac{\log \mathbb{P}(S_n \leq an)}{n}, & a \leq \mu \end{cases}$$

exists and satisfies

$$I(z) = \sup_{t\in\mathbb{R}}[zt - \log\varphi(t)].$$

The theorem says that the rate function I is the Legendre transform of $\log\varphi(t)$. Informally, this formula says that the value of $I(z)$ at a point z is given by the greatest possible difference between the curve $\log\varphi(t)$ and the line zt. Geometrically speaking, this is the point at which the gradient of $\log\varphi(t)$ is the same as that of zt, namely z. To determine this distance we can compute the corresponding equation of the tangent to $\log\varphi(t)$ and then the point at which this intersects the y-axis will give $-I(z)$.

Notice that, by the weak law of large numbers, given $\epsilon > 0$,

$$\mathbb{P}(|\frac{S_n}{n} - \mu| \leq \epsilon) \to 1$$

as $n \to \infty$. For $a < \mu$, we can choose ϵ to be sufficiently small that

$$\frac{S_n}{n} > \mu - \epsilon > a,$$

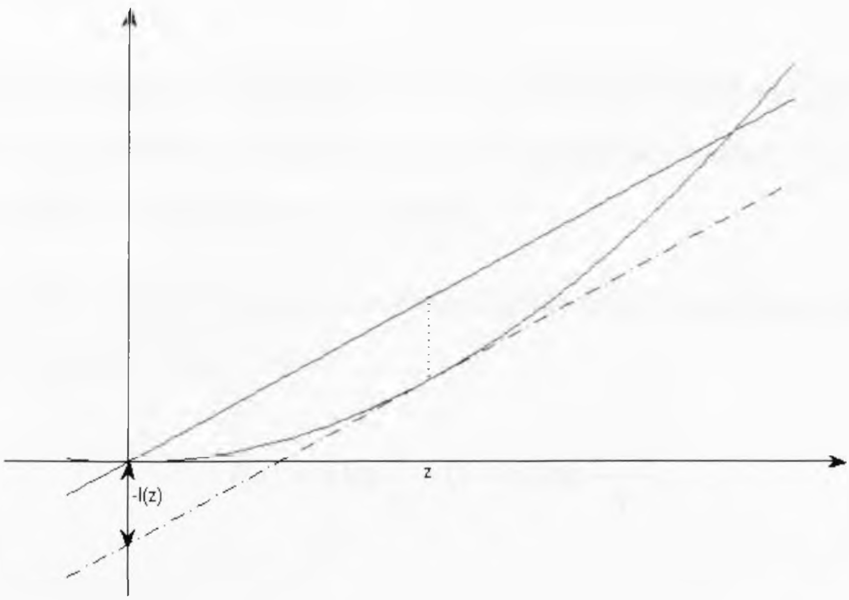

Figure 1.1: relationship between $\log \varphi$ and I

and hence

$$\mathbb{P}(\frac{S_n}{n} > a) \to 1,$$

as $n \to \infty$, so that, for $a < \mu$,

$$\frac{\log \mathbb{P}(S_n > na)}{n} \to 0$$

as $n \to \infty$. Similarly, for $a \geq \mu$, we have

$$\frac{\log \mathbb{P}(S_n \leq na)}{n} \to 0$$

as $n \to \infty$, hence our definition of I. Some key features of the rate function are that it is continuously differentiable and strictly convex at all point where it is finite with the properties that $I(a) \geq 0$ (with equality at the mean μ) and, furthermore, $I''(\mu) = 1/\sigma^2$. Indeed these two properties relate to the Strong Law of Large Numbers and Central Limit

Theorem respectively [24].

We will now compute the rate function for the random walk with independent Bern(p) increments as it will feature in later analysis. We present two methods: one via binomial coefficients and the other via Cramer's theorem.

Lemma 1. *The rate function I for the standard random walk S_n with independent Bern(p) increments is given, on* $[0, 1]$, *by*

$$I(a) = a \log \frac{a}{p} + (1-a) \log \frac{1-a}{q},$$

where $q = 1 - p$.

Proof. (via binomial coefficients) Throughout this proof we use (for convenience) the somewhat sloppy notation $na, n(1-a)$ when we really mean $\lfloor na \rfloor + 1$, $\lfloor n(1-a) \rfloor + 1$ respectively, where $\lfloor x \rfloor$ denotes the integer part of x. Now, we have

$$\mathbb{P}(S_n > na) = \sum_{k \geq na} \tbinom{n}{k} p^k q^{n-k}. \tag{1.1}$$

For $a > p$, we can bound (1.1) by noting that the greatest contribution to the summation comes from the first term:

$$\tbinom{n}{na} p^{na} q^{n(1-a)} \leq \mathbb{P}(S_n > na) \leq n \tbinom{n}{na} p^{na} q^{n(1-a)}.$$

Hence

$$\lim_{n\to\infty} \frac{\log \binom{n}{na} p^{na} q^{n(1-a)}}{n} \leq \lim_{n\to\infty} \frac{\log \mathbb{P}(S_n > na)}{n} \leq \lim_{n\to\infty} \left(\frac{\log n}{n} + \frac{\log \binom{n}{na} p^{na} q^{n(1-a)}}{n}\right),$$

and now, since $\frac{\log n}{n} \to 0$ as $n \to \infty$, it follows that

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(S_n > na)}{n} = \lim_{n\to\infty} \frac{\log \binom{n}{na} p^{na} q^{n(1-a)}}{n}.$$

We now claim that, for $a > 0$,

$$\frac{1}{n} \log \frac{n!}{(an)![(1-a)n]!} \to -a \log a - (1-a)\log(1-a)$$

as $n \to \infty$. By Stirling's asymptotic formula

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+1/2},$$

we have

$$\log n! \sim -n + (n + 1/2)\log n,$$

$$\log(an)! \sim -an + (an + 1/2)\log(1-a)n$$

and

$$\log[(1-a)n]! \sim -(1-a)n + [(1-a)n + 1/2]\log(1-a)n.$$

Hence

$$\log \frac{n!}{(an)![(1-a)n]!} \sim -n + an + (1-a)n + (n+1/2)\log n - (an + 1/2)\log n$$

$$-[(1-a)n + 1/2)\log n - an \log a - (1-a)n\log(1-a)$$

$$= O(\log n) - n[a\log a + (1-a)\log(1-a)].$$

The claim now follows by dividing by n and taking limits.

Applying this to our limit,we have then

$$\lim_{n\to\infty}\frac{\log\binom{n}{na}p^{na}q^{n(1-a)}}{n} = a\log p + (1-a)\log q - a\log a - (1-a)\log(1-a),$$

which tidies to

$$a\log\frac{p}{a} + (1-a)\log\frac{q}{1-a}.$$

The result now follows by interchanging the numerator and denominator of the logarithmic terms so as to take account of the minus sign in front of this quantity. It is easily seen that we obtain the same formula for $a \leq p$. □

Proof. (via Cramer's Theorem) Writing φ for the moment generating function of the Bern(p) random variables, we have

$$\log\varphi(t) = \log(pe^t + q)$$

so that the value of t for which $\log\varphi(t)$ has gradient z satisfies

$$\frac{pe^t}{pe^t + q} = z.$$

This gives

$$t = \log\frac{qz}{p(1-z)}$$

with corresponding value of $\log\varphi(t)$ given by

$$\log\frac{q}{1-z}$$

so that the tangent $T(t)$ to $\log \varphi(t)$ with gradient z has equation

$$T(t) - \log \frac{q}{1-z} = z(t - \log \frac{qz}{p(1-z)}).$$

Setting $t = 0$ we obtain

$$T(0) = \log \frac{q}{1-z} - z \log \frac{qz}{p(1-z)}$$

and now we want $-T(0)$ which simplifies to

$$z \log \frac{z}{p} + (1-z) \log \frac{1-z}{q},$$

precisely the formula we obtained using the binomial coefficients. □

1.3 Probability on ∂T, large deviations and cloud speed

We now return to the subject of the binary tree T and, in particular, its boundary. In the following section we will review the "normal" behaviour of the types along lines of descent and then investigate "abnormal" behaviour along such lines. By "normal" we mean that the proportions of marks equal to one obey the Strong Law of Large Numbers and by "abnormal" we mean that the proportions deviate from this. It should be clear, even at this stage, that the theory of large deviations discussed in the previous section will be relevant.

We begin with some notation. Let $\Omega = \{-1, 1\}^T$ so that the binary tree with Bernoulli noise can be thought of as a point in Ω. Notice that, in this context, the law L_ξ of $\xi = (\xi_u; u \in T)$ is the Bern(p) product measure on Ω. For $\xi \in \Omega$ and $t \in \partial T$, let $p_i(\xi, t)$

denote the limiting proportion (when it exists) of type $i \in \{-1, 1\}$ vertices observed along the line of descent corresponding to t in ξ. That is

$$p_i(\xi, t) = \lim_{n \to \infty} \frac{1}{n} \sum_{t \leq u} \mathbf{1}(\xi_u = i)$$

when this limit exists. The following two lemmas give us an idea of what to expect in terms of the limiting proportions of ones along "typical" lines of descent. By *Leb* on ∂T (or "uniform measure") we mean the Bern(1/2) product measure on $\{0, 1\}^{\mathbb{Z}^+}$. We can choose a point in ∂T according to Lebesgue measure via the following construction. We start at the root and with equal probability branch off to the left or right. We then branch off with equal probability to the left or right from our new position and so on. This continues indefinitely. By the time we "reach" the boundary ∂T, we are equally likely to be at any point.

Lemma 2. *Let* $P = \{(\omega, t) \in \overline{\Omega} \times \partial T : p_1(\xi(\omega), t) = p\}$. *Then* $\mathbb{P}(\omega : (\omega, t) \in P) = 1$ *for all* $t \in \partial T$.

Proof. This follows from the Strong Law of Large Numbers. For a fixed t, the law under $\mathbb{P}$ of the marked line of descent corresponding to t is the Bern(p) product measure on $\{-1, 1\}^{\mathbb{Z}^+}$. □

Lemma 3. *Let* $Q = \{\omega : Leb(t \in \partial T : p_1(\xi(\omega), t) = p) = 1\}$. *Then* $\mathbb{P}(Q) = 1$.

Proof. We consider the product measure $L_\xi \times Leb$ on the product space $\Omega \times \partial T$. Let P be as in the previous lemma. Then

$$(L_\xi \times Leb)(P) = \int\int_{\Omega \times \partial T} \mathbf{1}_P(\xi, t) d(L_\xi \times Leb)(\xi, t)$$

$$= \int_{\partial T} (\int_{\Omega} \mathbf{1}_P(\xi, t) dL_\xi(\xi)) dLeb(t)$$

(by Fubini's theorem)

$$= \int_{\partial T} L_\xi((\xi, t) \in P)dLeb(t) = 1$$

by the previous lemma. But, again by Fubini's theorem, we also have

$$\int\int_{\Omega\times\partial T} \mathbf{1}_P(\xi, t)d(L_\xi \times Leb)(\xi, t) = \int_\Omega(\int_{\partial T} \mathbf{1}_P(\xi, t)dLeb(t))dL_\xi(\xi)$$

$$= \int_\Omega Leb((\xi, t) \in P)dL_\xi(\xi),$$

and it now follows that $Leb(t : (\xi, t) \in P) = 1$ almost surely. □

The above lemmas give us an idea of what to expect at the boundary of the tree: the limiting proportion of type one vertices along a typical line of descent ought to be p, though this will not be the case for all lines of descent. With this in mind it is natural to wonder by how much the limiting proportions can deviate from p. Indeed which limiting proportions are possible and which not for a given value of p? In the following subsections we will encounter various inter-related mathematical frameworks within which we can ask such questions. We start with the notion of "cloud speed".

1.3.1 Cloud speed

In this section we state the classical theorem of Hammersley, Kingman and Biggins. This serves as preparation for the work of Chapter 2. We are guided in part by [42] and [16]. The setting is a general Galton-Watson tree as discussed at the beginning of the chapter. Recall that $(\xi_u; u \in \Gamma)$ is a collection of independent, identically distributed random variables associated with the vertices of a Galton-Watson tree Γ. Define the "tree-indexed random walk" $(S(\xi, u); u \in \Gamma)$ by

$$S(\xi, u) = \sum_{v \leq u} \xi_v.$$

Then the classical definition of cloud speed, as stated in [42] for example, is

$$s_{cloud} = \limsup_{n\to\infty} \max_{|u|=n} \frac{S(\xi,u)}{n}.$$

This is one of various "speeds" that can be associated with a tree-indexed process (see [42] for example) and it turns out that cloud speed (and indeed the other speeds) are almost surely constant. The following theorem makes this precise.

Theorem 2. *(Hammersley, Kingman, Biggins) Let* Γ *be a Galton-Watson tree with mean* $m > 1$. *Suppose that the vertices of* Γ *are labelled by independent, identically distributed random variables* ξ_u *satisfying*

1. ξ_u *is not almost surely constant,*

2. $\mathbb{E}[e^{\lambda\xi_u}] < \infty$ *for all* $\lambda \geq 0$.

Then on the event that Γ *survives, we have, almost surely,*

$$s_{cloud} = \sup\{s : I(s) \leq \log m\},$$

where I *is the rate function for the random walk with independent increments distributed as the* ξ_u.

It is worth pointing out at this stage a very quick and informal proof of one half of the theorem theorem as it shows clearly the connection between cloud speed and the rate function. Let M_n denote $\max_{|u|=n} S(\xi,u)$. Then, for a constant c,

$$\mathbb{P}(M_n > cn) = \mathbb{E}[\mathbf{1}_{(M_n>cn)}] \leq \mathbb{E}[\sum_{|u|=n} \mathbf{1}_{(S(\xi,u)>cn)}]$$

$$= m^n \mathbb{P}(S(\xi,u) > cn) \approx e^{-n(I(c)-\log m)}$$

for large n, where I is the rate function for independent increments distributed as the ξ_u. Now, by the Borell-Cantelli lemma, we have that

$$\mathbb{P}(M_n > nc \; infinitely \; many \; n) = 0$$

provided

$$\sum_n e^{-n(I(c)-\log m)} < \infty,$$

that is, provided

$$I(c) - \log m > 0.$$

In this case, it follows that

$$\limsup_n \frac{M_n}{n} \leq c$$

so that

$$s_{cloud} \leq \sup\{a : I(a) \leq \log m\}.$$

We will see later how to obtain the other half of the proof via martingale calculations.

By interpreting ξ_u as the displacement of a vertex u from its parent vertex, the tree-indexed random walk is a branching random walk and the theorem above says that, under the assumptions on the ξ_u, the position of the "rightmost" vertex moves linearly in time, at a rate given by the cloud speed. In Chapter 2 we will be interested in limiting proportions of types along lines of descent and for this reason we make the following definitions, based on the idea of cloud speed and in the context of the binary tree with Bernoulli noise. Let $\xi \in \Omega$. Define random variables $\overline{c}_{\pm}(\xi)$, the "upper cloud speed" of ξ corresponding to counts of ones or minus ones by, respectively,

$$\overline{c}_{+}(\xi) = \limsup_{n\to\infty} \max_{|u|=n} \frac{S^{+}(\xi,u)}{n}, \quad \overline{c}_{-}(\xi) = \limsup_{n\to\infty} \max_{|u|=n} \frac{S^{-}(\xi,u)}{n}$$

and $\underline{c}_{\pm}(\xi)$, the "lower cloud speed" of ξ corresponding to counts of ones or minus ones by, respectively,

$$\underline{c}_{+}(\xi) = \liminf_{n\to\infty} \min_{|u|=n} \frac{S^{+}(\xi, u)}{n}, \qquad \underline{c}_{-}(\xi) = \liminf_{n\to\infty} \min_{|u|=n} \frac{S^{-}(\xi, u)}{n},$$

where, in each case,

$$S^{+}(\xi, u) = \sum_{v\leq u} \mathbf{1}_{(\xi_v=1)}(\xi), \qquad S^{-}(\xi, u) = \sum_{v\leq u} \mathbf{1}_{(\xi_v=-1)}(\xi).$$

Notice that by the elementary properties of limsup and liminf we have $\overline{c}_{+}(\xi) + \underline{c}_{-}(\xi) = 1$ for all $\xi \in \Omega$. To obtain $\overline{c}_{+}$ we can now apply the theorem of Hammersley, Kingman and Biggins using the formula obtained earlier for the rate function I. Recall that I is given on $[0,1]$ by $I(a) = a\log\frac{a}{p} + (1-a)\log\frac{1-a}{q}$. Notice that in the case $p \geq 1/2, I(1) = -\log p \leq \log 2$ so that $\overline{c}_{+} = 1$ almost surely. By interchanging p and q in the formula for I we obtain the rate function $\tilde{I}$ corresponding to a random walk with independent Bern(q) increments. Now, $\overline{c}_{-}$ is just $\overline{c}_{+}$ with p and q interchanged so that ones and minus ones are swopped. Hence we can use the theorem to obtain $\overline{c}_{-}$ simply by interchanging p and q and it now follows that, in the case $p < 1/2$, we have $\underline{c}_{+} = 0$ almost surely since $\underline{c}_{+} = 1 - \overline{c}_{-}$. Informally, $\overline{c}_{\pm}$ corresponds to the greatest possible limiting proportions (along lines of descent) of type ± 1 vertices while $\underline{c}_{\pm}$ correspond to the least possible limiting proportions of type ± 1 vertices. In Chapter 2 we will see that determining conditions under which $\overline{c}_{\pm} < 1/2$ plays an important role.

1.4 Martingales: change of measure and spines

In this subsection we introduce a class of martingales, obtained as a change of measure, which turn out to be closely related to the cloud speed calculations from the previous section. We are working throughout on the binary tree with Bernoulli noise. Recall that $\mathbb{P}$ is the probability measure on $\overline{\Omega}$ under which the ξ_u are independent Bern(p) random variables on $\{-1,1\}$. Let $\mathbb{Q}$ denote the probability measure on $(\overline{\Omega},\mathcal{F})$ under which the ξ_u are independent, identically distributed Bern(p) random variables, except a long a random line of descent, chosen uniformly on the boundary ∂T (see our earlier discussion of Lebesgue measure on the boundary), along which the ξ_u are independent, identically distributed Bern(θ) random variables, independent of the ξ_u not on this line, where $\theta \in (0,1)$ and $\theta \neq p$. Let $\mathbb{P}_n, \mathbb{Q}_n$ denote, respectively, the restrictions of $\mathbb{P}, \mathbb{Q}$ to $\mathcal{F}_n$, the σ-algebra induced on $\overline{\Omega}$ by restricting to level n of the tree (see later for more details). The corresponding change of measure,

$$\Lambda_n^\theta(\xi) := \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$$

is, since it is a Radon-Nikodym derivative, a martingale with respect to $\mathcal{F}_n$. We will, nevertheless, verify this for our own purposes once we have obtained an appropriate representation. Strictly speaking, $\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$ is a function on $\overline{\Omega}$ and so we should write $\Lambda_n^\theta(\xi(\omega))$. We will say more about this later. This type of martingale has been studied extensively in the literature. See [37], [32], [36] for example. Indeed we will see later that Λ_n^θ is a special case of the martingales studied by Biggins in [5] and Lyons in [37]. Before moving on, we mention briefly the connection with the "spine approach", as seen in [19], [20], [21], [22] for example.

The way we defined $\mathbb{Q}$ corresponds to what is sometimes termed "size-biasing", in which one in some sense "engineers" the dynamics along a distinguished line of descent. One can view $\mathbb{Q}_n$ as the projection onto $\overline{\Omega}$ of a measure $\widetilde{\mathbb{Q}_n}$ on $\overline{\Omega} \times \partial T$. Indeed the probability measure $\mathbb{Q}$ can be obtained via the following construction. We begin with the root $\emptyset$, which gives birth to two daughter vertices (0,1). With equal probability, one of the daughter vertices is chosen. The chosen vertex is distributed as a Bern(1/2) random variable while the other vertex is distributed as a Bern(p) random variable, both on $\{-1, 1\}$. The vertex that was not chosen gives birth to two daughter vertices which themselves give birth to two vertices and so on. This continues indefinitely, with each vertex being distributed as a Bern(p) random variable on $\{-1, 1\}$. The vertex that was chosen also gives birth to two vertices and, as before, with equal probability one of these is chosen. The one that isn't chosen is distributed as a Bern(p) random variable on $\{-1, 1\}$ and continues giving birth to vertices with the Bern(p) distribution as before. The chosen vertex is distributed as a Bern(1/2) random variable and continues giving birth to two vertices, one of which is chosen with equal probability at each stage to have the Bern(1/2) distribution and so on. This also continues indefinitely. The measure $\widetilde{\mathbb{Q}}_n$ from above is the joint distribution of the random marked tree together with the distinguished line of descent. In this type of construction the chosen line of descent (which is uniformly chosen on the boundary ∂T since it is uniformly chosen on each level of the tree) is sometimes referred to as a "spine" (see [19], [36], [32]) for example. In [19], for example, this term is particularly apt as one can think of the chosen line of descent as being the "backbone" of the process from which all vertices are born. This type of process is also analagous to what Lyons, Pemantle and Peres term "size-biased trees" in [36], in which the authors obtain the martingale of the Kesten-Stigum theorem (i.e. the number of vertices in level n of the tree normed by the n th power of the mean number) via a change of measure in the same sort of way described above.

Proposition 1.

$$\Lambda_n^\theta(\xi) = \frac{1-\theta}{q} \sum_{|u|=n} \left(\frac{1-\theta}{2q}\right)^n \left(\frac{q\theta}{p(1-\theta)}\right)^{S^+(\xi,u)}.$$

Proof. For convenience, we work throughout on the canonical probability space. That is, we assume $\overline{\Omega} = \{-1, 1\}^T$ so that the ξ_u are "coordinate" maps with $\xi_u(\omega) = \omega_u$ giving the type of the vertex u, that $\mathbb{P}$ is the Bern(p) product measure on $\{-1, 1\}^T$, and that $\mathcal{F}$ is the σ-algebra generated by the ξ_u, i.e. $\mathcal{F} = \sigma(\xi_u; u \in T)$, which is, of course, the Borel σ-algebra on $\overline{\Omega}$ since it is generated by the coordinate maps. Now, to begin the proof, notice that the space $(\{-1, 1\}^T, \sigma(\xi_u; |u| \leq n))$, together with $\mathbb{P}_n$ or $\mathbb{Q}_n$, is discrete. We can therefore calculate Λ_n^θ by computing the probabilities of "fundamental" events and taking their quotient. In this case, the fundamental events in question are collections of marked trees with specified vertex types down to level n. Let

$$A = \{\xi : \xi_u = t_u; |u| \leq n\},$$

where the $t_u \in \{-1, 1\}$ are fixed. Define

$$N_i^n(\xi) = \sharp(u : \xi_u = i; |u| \leq n).$$

For a vertex $u \in G_n$, define

$$N_i^n(\xi, u) = \sharp(v : \xi_v = i, v \nleq u).$$

Then we have

$$\mathbb{P}_n(A) = p^{N_1^n(\xi)} q^{N_{-1}^n(\xi)}$$

and, since

$$\mathbb{Q}_n(A) = \frac{1}{2^n} \sum_{u \in G_n} \mathbb{Q}_n(A|u),$$

where $\mathbb{Q}_n(A|u)$ denotes the probability of A given that the distinguished line of descent (i.e. the line of descent along which the types are Bern($\frac{1}{2}$)) goes through u, we have

$$\mathbb{Q}_n(A) = \frac{1}{2^n} \sum_{|u|=n} p^{N_1^n(\xi,u)} q^{N_{-1}^n(\xi,u)} \theta^{S^+(\xi,u)} (1-\theta)^{S^-(\xi,u)},$$

where $S^-(\xi,u) = \sum_{v \le u} 1_{(\xi_v=-1)}(\xi)$. Dividing these probabilities gives

$$\Lambda_n^\theta = \frac{\mathbb{Q}_n(A)}{\mathbb{P}_n(A)} = \frac{1}{2^n} \sum_{|u|=n} (\frac{\theta}{p})^{S^+(\xi,u)} (\frac{1-\theta}{q})^{S^-(\xi,u)}$$

$$= \frac{1}{2^n} \sum_{|u|=n} (\frac{\theta}{p})^{S^+(\xi,u)} (\frac{1-\theta}{q})^{n+1-S^+(\xi,u)}$$

since $S^+(\xi,u) + S^-(\xi,u) = n+1$. Simplifying, we obtain

$$\Lambda_n^\theta = \frac{1}{2^n} \sum_{|u|=n} (\frac{q}{p}\frac{\theta}{1-\theta})^{S^+(\xi,u)} (\frac{1-\theta}{q})^n \frac{1-\theta}{q}$$

$$= \frac{1-\theta}{q} \sum_{|u|=n} (\frac{1-\theta}{2q})^n (\frac{q\theta}{p(1-\theta)})^{S^+(\xi,u)}.$$

□

Notice that although it might appear that the formula we have obtained applies only to the canonical space described at the beginning of the proof, this result does, in fact, hold for any abstract space. To see this, consider the following mapping from an arbitrary abstract space to the canonical space:

$$(\overline{\Omega}, \mathcal{F}, \mathbb{P}, \mathbb{Q}) \to^{\xi} (\Omega = \{-1,1\}^T, E = \sigma(\xi_u; u \in T), L_\xi, \nu),$$

where L_ξ is the law of ξ under $\mathbb{P}$ and ν is the law of ξ under $\mathbb{Q}$. Notice that the σ-algebra on the canonical space induces a σ-algebra on $\overline{\Omega}$, namely $\{\xi^{-1}(A) : A \in E)\}$ so that the corresponding restricted σ-algebra is given by $\mathcal{F}_n = \{\xi^{-1}(A) : A \in E_n\}$, where

$E_n = \sigma(\xi_u; |u| \leq n)$.

Claim 1. *Let $\mathbb{P}_n, \mathbb{Q}_n$ be the restrictions of $\mathbb{P}, \mathbb{Q}$ respectively to $\mathcal{F}_n$ as above. Then*

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} = \Lambda_n^\theta(\xi(\omega)).$$

Proof. The above is a function on $\overline{\Omega}$ and indeed it is $\mathcal{F}_n$-measurable since $\Lambda_n^\theta(\xi)$ is E_n measurable and ξ is $\mathcal{F}$-measurable. Now take $K \in E_n$ with $\xi^{-1}(K) = A \in \mathcal{F}_n$. Then

$$\mathbb{Q}(A) = \mathbb{Q}(\xi^{-1}(K)) = \nu(K)$$

$$= \int_K \Lambda_n^\theta dL_\xi$$

(by our result in the canonical case)

$$= \int_A \Lambda_n^\theta d\mathbb{P}.$$

□

Henceforth we will write Λ_n^θ for the Radon-Nikodym derivative and work with it in this form. That is, we will regard it as a function of ξ.

Lemma 4. *Λ_n^θ is a martingale with respect to the filtration $\mathcal{F}_n$.*

Proof. Notice firstly that Λ_n^θ is adapted to $\mathcal{F}_n$ (i.e. it is $\mathcal{F}_n$-measurable) and that it has finite (unit in fact) expectation. Now, to prove the martingale property, we have

$$\mathbb{E}[\Lambda_{n+1}^\theta|\mathcal{F}_n] = \mathbb{E}[\frac{1-\theta}{q}\sum_{|u|=n}(\frac{1-\theta}{2q})^{n+1}(\frac{q\theta}{p(1-\theta)})^{S^+(\xi,u)+1_{(\xi_{u_{n+1}}=1)}}|\mathcal{F}_n],$$

where $\xi_{u_{n+1}}$ is the type of a vertex in level $n+1$ of the tree. Hence

$$\mathbb{E}[\Lambda^\theta_{n+1}|\mathcal{F}_n] = \frac{1-\theta}{q}(\frac{1-\theta}{2q})^{n+1}\mathbb{E}[\sum_{|u|=n}(\frac{q\theta}{p(1-\theta)})^{S^+(\xi,u)}(\frac{q\theta}{p(1-\theta)})^{\mathbf{1}(\xi_{u_{n+1}}=1)}|\mathcal{F}_n]$$

(taking out the constant term)

$$= \frac{1-\theta}{q}(\frac{1-\theta}{2q})^{n+1}\sum_{|u|=n}(\frac{q\theta}{p(1-\theta)})^{S^+(\xi,u)}\mathbb{E}[(\frac{q\theta}{p(1-\theta)})^{\mathbf{1}(\xi_{u_{n+1}}=1)}|\mathcal{F}_n]$$

(since $\sum_{|u|=n}(\frac{q\theta}{p(1-\theta)})^{S^+(\xi,u)}$ is $\mathcal{F}_n$-measurable)

$$= \frac{1-\theta}{q}(\frac{1-\theta}{2q})^{n+1}\sum_{|u|=n}(\frac{q\theta}{p(1-\theta)})^{S^+(\xi,u)}\mathbb{E}[(\frac{q\theta}{p(1-\theta)})^{\mathbf{1}(\xi_{u_{n+1}}=1)}]$$

(since $\frac{q\theta}{p(1-\theta)})^{\mathbf{1}(\xi_{u_{n+1}}=1)}$ is independent of $\mathcal{F}_n$)

$$= \frac{1-\theta}{q}(\frac{1-\theta}{2q})^{n+1}\sum_{|u|=n}(\frac{q\theta}{p(1-\theta)})^{S^+(\xi,u)} \times 2(p\frac{q\theta}{p(1-\theta)}+q)$$

$$= \frac{1-\theta}{q}\sum_{|u|=n}(\frac{1-\theta}{2q})^n(\frac{q\theta}{p(1-\theta)})^{S^+(\xi,u)} = \Lambda^\theta_n.$$

□

1.5 Uniform integrability

A very natural question to ask at this stage is under what circumstances is the probability measure $\mathbb{Q}$ absolutely continuous with respect to $\mathbb{P}$ or indeed under what conditions are they equivalent? For absolute continuity we require that the martingale Λ^θ_n converges in L^1 while for equivalence we require the further property that the limit is almost surely positive. The L^1 convergence of martingales of this form is a classical subject, much studied by authors such as Biggins and Lyons. We will see in the next section that Λ_n^θ is, in fact, a martingale of the form covered by Biggins' Theorem [5]. For the moment we will advance our own methods.

The main idea of the proof of this section relies on the decomposition of Λ_n^θ as a sum of independent copies. This technique is by no means new and has been used, for example, by Neveu in [40]. We write Λ_n^θ in terms of

$$S(\xi, u) = \sum_{v \leq u} \xi_v$$

since, for the purposes of later work, it will be convenient to have it in this form. Noting that, for $u \in G_n$,

$$S^+(\xi, u) + S^-(\xi, u) = n + 1, \quad S^+(\xi, u) - S^-(\xi, u) = S(\xi, u),$$

so that

$$S^+(\xi, u) = \frac{n + 1 + S(\xi, u)}{2},$$

we have then

$$\Lambda_n^\theta = \sqrt{\frac{\theta(1-\theta)}{pq}} \sum_{|u|=n} (\frac{\theta(1-\theta)}{4pq})^{n/2} (\frac{q\theta}{p(1-\theta)})^{S(\xi,u)/2}.$$

1.5.1 Decomposition of Λ_n^θ

In this section we write Λ_n^θ as a sum of independent copies associated with the immediate daughter vertices of the root. This type of decomposition has been used by authors such as Neveu in [40]. Let $\Lambda_n^{\theta'}$ denote the probability ratio $\frac{d\mathbb{Q}}{d\mathbb{P}}$ restricted to the σ-algebra generated by the vertices to level n of the sub-tree rooted at 0 and $\Lambda_n^{\theta''}$ be defined similarly but rooted at 1. Let

$$T_L = \{u \in T : 0 \leq u\}, \quad T_R = \{u \in T : 1 \leq u\}.$$

We have then

$$\Lambda_n^\theta = \sqrt{\frac{\theta(1-\theta)}{pq}} \sum_{|u|=n} (\frac{\theta(1-\theta)}{4pq})^{n/2} (\frac{q\theta}{p(1-\theta)})^{S(\xi,u)/2},$$

$$\Lambda_n^{\theta'} = \sqrt{\frac{\theta(1-\theta)}{pq}} \sum_{|u|=n+1, u \in T_L} (\frac{\theta(1-\theta)}{4pq})^{n/2} (\frac{q\theta}{p(1-\theta)})^{S(\xi,u)/2-\xi_\emptyset/2},$$

$$\Lambda_n^{\theta''} = \sqrt{\frac{\theta(1-\theta)}{pq}} \sum_{|u|=n+1, u \in T_R} (\frac{\theta(1-\theta)}{4pq})^{n/2} (\frac{q\theta}{p(1-\theta)})^{S(\xi,u)/2-\xi_\emptyset/2}.$$

It is clear that $\Lambda_n^\theta, \Lambda_n^{\theta'}, \Lambda_n^{\theta''}$ all have the same distribution. Furthermore, it is now easy to see that we have the following decomposition:

$$\Lambda_n^\theta = \frac{\Lambda_0^\theta}{2} (\Lambda_{n-1}^{\theta'} + \Lambda_{n-1}^{\theta''}),$$

where

$$\Lambda_0^\theta = \sqrt{\frac{\theta(1-\theta)}{pq}} (\frac{q\theta}{p(1-\theta)})^{\xi_\emptyset/2}$$

is the value of the probability ratio at the root. Writing $\Lambda_\infty^\theta, \Lambda_\infty^{\theta'}, \Lambda_\infty^{\theta''}$ for the almost sure limits of $\Lambda_n^\theta, \Lambda_n^{\theta'}, \Lambda_n^{\theta''}$ (which exist by the Martingale Convergence Theorem), we also have

$$\Lambda_\infty^\theta = \Lambda_0^\theta (\frac{1}{2} \Lambda_\infty^{\theta'} + \frac{1}{2} \Lambda_\infty^{\theta''}) \quad a.s.,$$

with $\Lambda_\infty^\theta, \Lambda_\infty^{\theta'}, \Lambda_\infty^{\theta''}$ having the same distribution. We will see later that we can use the idea of one of the proofs in [40] to deduce when Λ_∞^θ is trivial (i.e. almost surely zero). An immediate consequence of the above decomposition is the following.

Proposition 2. *Either* $\Lambda_\infty^\theta = 0$ *almost surely or* $\Lambda_\infty^\theta > 0$ *almost surely.*

Proof. By the decomposition we have that $\Lambda_\infty^\theta = 0$ if and only if $\Lambda_\infty^{\theta'} = \Lambda_\infty^{\theta''} = 0$, forcing $\mathbb{P}(\Lambda_\infty^\theta = 0) \in \{0, 1\}$ since $\Lambda_\infty^{\theta'}, \Lambda_\infty^{\theta''}$ have the same distribution and are independent. □

We remark that one can use the same type of argument to deduce the analogous result for the Kesten-Stigum martingale. See [36] for example.

Proposition 3. Λ_n^θ *is uniformly integrable for* $\theta \in (0, 1)$ *satisfying*

$$I(\theta) < \log 2,$$

where, as usual, I is the rate function for the random walk with independent Bern(p) increments.

By considering the graph of the rate function I we observe the following. For $p < 1/2$, the proposition says that Λ_n^θ is uniformly integrable for $\theta \in [0, c^*)$, where c^* is the "upper cloud speed" (corresponding to ones), i.e. the almost sure value of $\overline{c}_+(\xi)$. For $p > 1/2$, the proposition says that Λ_n^θ is uniformly integrable for $\theta \in (c_*, 1]$, where c_* is the "lower cloud speed" (corresponding to ones), i.e. the almost sure value of $\underline{c}_+(\xi)$. On the other hand when $p = 1/2$, $I(0) = I(1) = \log 2$ so that the proposition says that Λ_n^θ is uniformly integrable for all $\theta \in (0, 1)$.

Lemma 5. *Let* $x, y \geq 0$*. Then* $(x + y)^r \leq x^r + x^{r-1}y + xy^{r-1} + y^r$ *for* $1 \leq r \leq 2$.

Proof. For $x, y \geq 0$ and $p \in [0, 1]$ we have the well known inequality

$$(x + y)^p \leq x^p + y^p.$$

The result follows by multiplying through by $(x + y)$ and writing $r = p + 1$. □

Proof. (of Proposition 3) We prove that for the stated values of θ, Λ_n^θ is bounded in L^r for some $r > 1$. Recall that

$$\Lambda_n^\theta = \frac{\Lambda_0^\theta}{2}(\Lambda_{n-1}^{\theta'} + \Lambda_{n-1}^{\theta''}),$$

where

$$\Lambda_0^\theta = \sqrt{\frac{\theta(1-\theta)}{pq}}(\frac{q\theta}{p(1-\theta)})^{\xi_\emptyset/2}$$

is the value of the probability ratio at the root. By mutual independence, we have then

$$\mathbb{E}[(\Lambda_n^\theta)^r] = (\frac{p^{1-r}\theta^r + q^{1-r}(1-\theta)^r}{2^r})(\mathbb{E}[(\Lambda_{n-1}^{\theta'} + \Lambda_{n-1}^{\theta''})^r]).$$

Now, by the previous claim,

$$\mathbb{E}[(\Lambda_n^\theta)^r] \leq K_r^\theta\big(\mathbb{E}[(\Lambda_{n-1}^{\theta'})^r] + \mathbb{E}[(\Lambda_{n-1}^{\theta'})^{r-1}]\mathbb{E}[\Lambda_{n-1}^{\theta''}] + \mathbb{E}[(\Lambda_{n-1}^{\theta''})^{r-1}]\mathbb{E}[\Lambda_{n-1}^{\theta'}] + \mathbb{E}[(\Lambda_n^{\theta''})^r]\big),$$

where

$$K_r^\theta = (2p)^{1-r}\theta^r + (2q)^{1-r}(1-\theta)^r.$$

We now appeal to dominance of L^p norms: for $1 \leq r \leq 2$, $||\cdot||_{L^1} \leq ||\cdot||_{L^{\frac{1}{r-1}}}$ and hence $\mathbb{E}[(\Lambda_n^\theta)^{r-1}] \leq \mathbb{E}[\Lambda_n^\theta]^{r-1} = 1$ so that

$$\mathbb{E}[(\Lambda_n^\theta)^r] \leq K_r^\theta[\mathbb{E}[(\Lambda_{n-1}^\theta)^r] + 1].$$

We thus have a recurrence relation in terms of $\mathbb{E}[(\Lambda_n^\theta)^r]$. The recurrence $a_{n+1} = K(a_n + 1)$, for K positive has solution $a_n = K^n(a_0 + 1) + K^{n-1} + ... + K$ and hence the solution tends to a finite limit provided $K < 1$. It follows that Λ_n^θ will be bounded in L^r if there exists an $r \in (1, 2)$ such that $K_r^\theta < 1$. Simplifying the expression for K_r^θ and setting it less than 1 gives

$$p^{1-r}\theta^r + q^{1-r}(1-\theta)^r - 2^{r-1} < 0.$$

Let

$$f(r) = p^{1-r}\theta^r + q^{1-r}(1-\theta)^r - 2^{r-1}.$$

Then $f(1) = 0$ and

$$f'(1) = \theta \log \frac{\theta}{p} + (1-\theta)\log \frac{1-\theta}{q} - \log 2 < 0$$

precisely for those θ satisfying

$$I(\theta) - \log 2 < 0.$$

By the continuity of f there exists an $r > 1$ with $f(r) < 0$. This completes the proof. □

We now verify that the limit Λ_∞^θ is almost surely positive.

Lemma 6. *Let*

$$\Lambda_\infty^\theta = \lim_{n\to\infty} \Lambda_n^\theta.$$

Then $\Lambda_\infty^\theta > 0$ almost surely whenever Λ_n^θ is uniformly integrable.

Proof. By uniform integrability, $\mathbb{E}[\Lambda_\infty^\theta] = 1$ and hence Λ_∞^θ is not almost surely zero. Since $\mathbb{P}(\Lambda_\infty^\theta = 0) \in \{0, 1\}$ by Proposition 2, we must have $\mathbb{P}(\Lambda_\infty = 0) = 0$. □

Since Λ_∞^θ is almost surely positive for θ such that Λ_n^θ is uniformly integrable, it now follows that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent for such values of θ. The obvious question now is what happens for the other values of $\theta \in (0,1)$? The following proposition addresses this issue. The idea of the proof is based on a proof in [40].

Proposition 4. *Let $\theta \in (0,1)$ be such that*

$$I(\theta) \geq \log 2.$$

Then Λ_n^θ converges, almost surely, to zero.

Proof. Notice first that for any $a > 0$, $b \geq 0$ and $r \in [0, 1]$, we have

$$\frac{1}{1-r} a[a^{r-1} - (a+b)^{r-1}] \leq \max(1, b)$$

and that

$$\lim_{r \uparrow 1} \frac{1}{1-r} a[a^{r-1} - (a+b)^{r-1}] = a \log \frac{a+b}{a},$$

so that, by dominated convergence,

$$\lim_{r \uparrow 1} \frac{1}{1-r} [\mathbb{E}[(\Lambda_\infty^{\theta'})^r] - \mathbb{E}[\Lambda_\infty^{\theta'}(\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''})^{r-1}]] = \mathbb{E}\left[\Lambda_\infty^{\theta'} \log \frac{\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''}}{\Lambda_\infty^{\theta'}}\right].$$

Now,

$$\mathbb{E}[(\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''})^r] = \mathbb{E}[(\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''})(\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''})^{r-1}]$$

$$= \mathbb{E}[\Lambda_\infty^{\theta'}(\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''})^{r-1}] + \mathbb{E}[\Lambda_\infty^{\theta''}(\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''})^{r-1}]$$

and hence, by symmetry,

$$\mathbb{E}[\Lambda_\infty^{\theta'}(\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''})^{r-1}] = \frac{1}{2}\mathbb{E}[(\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''})^r].$$

But by the decomposition of Λ_n^θ, we have

$$\mathbb{E}[(\Lambda'_\infty + \Lambda''_\infty)^r] = \frac{2^r \mathbb{E}[(\Lambda_\infty^\theta)^r]}{\mathbb{E}[(\Lambda_0^\theta)^r]}$$

so that

$$\mathbb{E}[\Lambda_\infty^{\theta'} \log \frac{\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''}}{\Lambda_\infty^{\theta'}}] = \lim_{r \uparrow 1} \frac{1}{1-r} \mathbb{E}[(\Lambda_\infty^\theta)^r](1 - \frac{2^{r-1}}{\mathbb{E}[(\Lambda_0^\theta)^r]}),$$

where

$$\mathbb{E}[(\Lambda_0^\theta)^r] = p^{1-r}\theta^r + q^{1-r}(1-\theta)^r.$$

Dealing with the term in brackets, in the limit we obtain

$$\lim_{r\uparrow 1} \frac{p(\frac{\theta}{p})^r + q(\frac{1-\theta}{q})^r - 2^{r-1}}{(1-r)[p(\frac{\theta}{p})^r + q(\frac{1-\theta}{q})^r]}$$

$$= \lim_{r\uparrow 1} \frac{p(\frac{\theta}{p})^r \log\frac{\theta}{p} + q(\frac{1-\theta}{q})^r \log\frac{1-\theta}{q} - 2^{r-1}\log 2}{-[p(\frac{\theta}{p})^r + q(\frac{1-\theta}{q})^r]}$$

$$= -\theta\log\frac{\theta}{p} - (1-\theta)\log\frac{1-\theta}{q} + \log 2. \tag{1.2}$$

That is,

$$\mathbb{E}[\Lambda_\infty^{\theta'} \log\frac{\Lambda_\infty^{\theta'} + \Lambda_\infty^{\theta''}}{\Lambda_\infty^{\theta'}}] = [-\theta\log\frac{\theta}{p} - (1-\theta)\log\frac{1-\theta}{q} + \log 2]\mathbb{E}[\Lambda_\infty^{\theta}]$$

$$= (\log 2 - I(\theta))\mathbb{E}[\Lambda_\infty^{\theta}].$$

Now, the expectations on the left and right hand side are both non-negative and so it follows that we must have $\mathbb{E}[\Lambda_\infty^\theta] = 0$ for values of θ such that $I(\theta) \geq \log 2$. This in turn forces $\Lambda_\infty^\theta = 0$ almost surely for such θ since Λ_n^θ is itself non-negative. □

The previous two results provide us with an interpretation of the upper cloud speed. We can think of the cloud speed as being (for given values of p and q) the largest possible value of θ such that the martingale Λ_n^θ tends to a trivial (i.e. almost surely zero) limit. The cloud speed is thus the "transition point" at which the limit of the corresponding martingale goes from being trivial to uniformly integrable. We are now also in a position to prove the second half of the theorem of Hammersley, Kingman and Biggins. Recall that, by the theorem of Hammersley, Kingman and Biggins, the cloud speed is given almost surely by

$$s^* = \sup\{a : I(a) \leq \log m\},$$

with I being the rate function for the associated random walk and m the mean of the offspring distribution (see an earlier section for full details). Recall also that we pointed out an easy method (involving the Borell-Cantelli lemma) for proving that the cloud speed is bounded above by s^*. We now claim that the uniform integrability arguments of this section enable us to conclude that the cloud speed is also bounded below by s^*, thus giving the desired equality. This ties in with our observation that the cloud speed is the value of θ corresponding to the "transition" of the Λ_n^θ martingales from being trivial to non-trivial. We write out the details below.

Under $\mathbb{Q}$, there exists (almost surely) a line of descent with limiting proportion θ of type one vertices (recall the construction of $\mathbb{Q}$). When $\mathbb{P}$ and $\mathbb{Q}$ are equivalent (i.e. when Λ_n^θ is uniformly integrable), it follows that the same is true under $\mathbb{P}$. Hence the cloud speed must be bounded below by θ whenever the martingale Λ_n^θ is uniformly integrable. The question now is how large can we make θ? The answer comes from an earlier calculation in which we remarked that the issue of whether or not Λ_n^θ is trivial depended on the sign of $\log 2 - I(\theta)$. For θ such that Λ_n^θ is uniformly integrable, we must have $\log 2 - I(\theta) > 0$, that is, $I(\theta) < \log 2$. So we can take θ to be as large as

$$\sup\{\theta : I(\theta) \leq \log 2\},$$

thus establishing that

$$s_{cloud} \geq \sup\{\theta : I(\theta) \leq \log 2\},$$

which generalises in a straightforward way to an arbitrary Galton-Watson tree with mean m.

1.6 Martingales again: Biggins's theorem

In this section we show that our martingales Λ_n^θ are of the form studied by Biggins in [5] and later Lyons in [37]. We elucidate the connection, state Biggin's Theorem relating to the convergence of such martingales and then re-derive our earlier results by making explicit calculations. The setup is taken directly from [37].

Recall the Galton-Watson tree Γ with corresponding branching factor $\mathbb{E}[N]$ and the associated collection of tree-indexed independent, identically distributed random variables $(\xi_u; u \in \Gamma)$. Let q be the extinction probability for the underlying tree Γ. For $\alpha \in \mathbb{R}$, define

$$\langle \alpha, N \rangle = \sum_{i=1}^{N} e^{-\alpha \xi_i}$$

and

$$m(\alpha) = \mathbb{E}[\langle \alpha, N \rangle],$$

where it is assumed that $m(0) > 1$ so that $q < 1$. Then if $m(\alpha) < \infty$ for some α, the sequence

$$W_n(\alpha) = \frac{\sum_{|u|=n} e^{-\alpha S(\xi,u)}}{m(\alpha)^n} \tag{1.3}$$

is a martingale and, being positive, has an almost surely finite limit by the Martingale Convergence Theorem, which we denote $W(\alpha)$. Writing

$$m'(\alpha) = \mathbb{E}[\sum_{i=1}^{N} \xi_i e^{-\alpha \xi_i}]$$

when this exists, we can now state Biggins' Theorem [5], which determines when $W(\alpha)$ is non-trivial.

Theorem 3. *(Biggins 1977) Suppose that $\alpha \in \mathbb{R}$ is such that $m(\alpha) < \infty$ and $m'(\alpha)$ exists and is finite. Then the following are equivalent:*

1. $\mathbb{P}(W(\alpha) = 0) < 1$;

2. $W_n(\alpha)$ *is uniformly integrable;*

3. $\mathbb{E}[\langle \alpha, N \rangle \log^+ \langle \alpha, N \rangle] < \infty$ *and* $\frac{\alpha m'(\alpha)}{m(\alpha)} < \log m(\alpha)$,

where $\log^+ x = \max(0, \log x)$.

This theorem may be regarded, in some sense, as an extension of the Kesten-Stigum Theorem, in which conditions for the convergence of the martingale obtained as the number of vertices in level n of the tree, normed by the n th power of the mean, are given. Notice that the condition "$m(\alpha) < \infty$" (necessary so that the corresponding sequence $W_n(\alpha)$ is a martingale) is essentially a condition on the moment generating function of the ξ_u. This should be viewed as the same type of condition governing the existence of the rate function I for the associated random walk (see an earlier section) and hence as a link between the two pieces of work. It is self-evident that Λ_n^θ is a martingale of the above form. The following result makes this apparent.

Lemma 7. *Our martingale Λ_n^θ is a special case of the Biggins martingales in that*

$$W_n(\alpha) = \mathbb{E}[W_0]\Lambda_n^\theta,$$

with

$$\alpha = \frac{1}{2} \log \frac{p(1-\theta)}{q\theta}.$$

Proof. Notice that since W_n is expressed in terms of $S(\xi. u)$, it is most convenient to work with the expression for Λ_n^θ in terms of $S(\xi, u)$. Recall that

$$\Lambda_n^\theta = \sqrt{\frac{\theta(1-\theta)}{pq}} \sum_{|u|=n} \left(\frac{\theta(1-\theta)}{4pq}\right)^{n/2} \left(\frac{q\theta}{p(1-\theta)}\right)^{S(\xi,u)/2}.$$

Comparing the exponent term in the martingales $W_n(\alpha)$ with the corresponding term in Λ_n^θ, we see immediately that to have equality we must have

$$e^{-\alpha} = \sqrt{\frac{q\theta}{p(1-\theta)}},$$

which gives the value of α stated. To see that this works, notice that

$$m\left(\frac{1}{2}\log\frac{p(1-\theta)}{q\theta}\right) = 2\left(p\sqrt{\frac{q\theta}{p(1-\theta)}} + q\sqrt{\frac{p(1-\theta)}{q\theta}}\right) = \sqrt{\frac{4pq}{\theta(1-\theta)}}$$

so that

$$\begin{aligned} W_n &= \frac{\sum_{|u|=n}\left(\sqrt{\frac{q\theta}{p(1-\theta)}}\right)^{S(\xi,u)}}{\left(\sqrt{\frac{4pq}{\theta(1-\theta)}}\right)^n} \\ &= \sum_{|u|=n}\left(\frac{\theta(1-\theta)}{4pq}\right)^{n/2}\left(\frac{q\theta}{p(1-\theta)}\right)^{S(\xi,u)/2} \\ &= \mathbb{E}[W_0]\Lambda_n^\theta, \end{aligned}$$

since

$$\mathbb{E}[W_0] = p\sqrt{\frac{q\theta}{p(1-\theta)}} + q\sqrt{\frac{p(1-\theta)}{q\theta}} = \sqrt{\frac{pq}{\theta(1-\theta)}}.$$

□

Notice that the "($\mathbb{E}[W_0]$)" term is a normalising constant. This shouldn't be too surprising when we remember where Λ_n^θ came from: as a Radon-Nikodym derivative (or change of measure), we must have $\mathbb{E}[\Lambda_n^\theta] = 1$ for every n. By the martingale property, this will be the case provided $\mathbb{E}[\Lambda_0^\theta] = 1$, which is why Λ_n^θ is equal to a normalised version of W_n.

Having established the connection between the martingales Λ_n^θ and $W_n(\alpha)$, we can use Biggins' theorem to verify our earlier result, in which we determined conditions under which Λ_n^θ is uniformly integrable.

Setting

$$\alpha = \frac{1}{2}\log\frac{p(1-\theta)}{q\theta},$$

we have

$$m(\alpha) = 2(pe^{-\alpha} + qe^{\alpha}) = 2(p\sqrt{\frac{q\theta}{p(1-\theta)}} + q\sqrt{\frac{p(1-\theta)}{q\theta}})$$
$$= 2\sqrt{\frac{pq}{\theta(1-\theta)}}$$

and that

$$m'(\alpha) = 2(-pe^{-\alpha} + qe^{\alpha}) = 2(1-2\theta)\sqrt{\frac{pq}{\theta(1-\theta)}}$$

so that Biggins' condition for uniform integrability "$\frac{\alpha m'(\alpha)}{m(\alpha)} < \log m(\alpha)$" becomes

$$\frac{1-2\theta}{2}\log\frac{p(1-\theta)}{q\theta} < \frac{1}{2}\log\frac{4pq}{\theta(1-\theta)}$$

which is satisfied if and only if

$$\frac{1}{2}\log\frac{p}{\theta} + \frac{1}{2}\log\frac{1-\theta}{q} + \theta\log\frac{\theta}{p} - \theta\log\frac{1-\theta}{q} - \frac{1}{2}(\log 2 + \log\frac{p}{\theta}) - \frac{1}{2}(\log 2 + \log\frac{q}{1-\theta}) < 0.$$

Simplifying, this inequality becomes

$$\theta\log\frac{\theta}{p} + (1-\theta)\log\frac{1-\theta}{q} - \log 2 < 0,$$

or the familiar

$$I(\theta) - \log 2 < 0$$

condition. Notice that since Biggins' theorem gives us equivalent conditions, we can deduce that for values of θ not satisfying the above inequality, Λ_n^θ is almost surely zero (negation of the first equivalent condition of the theorem), which tallies with our earlier result. It is worth noting that the calculations above and indeed some of the earlier calculations are greatly simplified by a prudent choice of θ. For $\theta = 1/2$, we have, for example, that $m'(\alpha) = 0$. In Chapter 2 we will introduce "recovery problems" and discuss a fundamental reason for studying the martingale $\Lambda_n^{1/2}$.

1.7 Asymptotic behaviour of $S^+(\xi, u)$

The theorem on cloud speed tells us about the asymptotic behaviour of $S^+(\xi, u)$. The purpose of this section is to draw attention to the fact that we can also make use of the convergence of the Λ_n^θ martingales. In a so-called "critical case" in Chapter 2 this approach will give us more information than can be deduced directly from the cloud speed arguments.

Being a positive martingale, we know that Λ_n^θ converges almost surely to a finite limit (Martingale Convergence Theorem). Writing Λ_n^θ as an exponential

$$\Lambda_n^\theta = \frac{1-\theta}{2q} \sum_{|u|=n} \exp\{S^+(\xi, u) \log \frac{q\theta}{p(1-\theta)} - n \log \frac{2q}{1-\theta}\},$$

we now note that we must have $\Lambda_n^\theta < M$ for every n, where M is an almost surely finite random variable. Since Λ_n^θ is a sum of positive terms, it follows that

$$\exp\{S^+(\xi,u)\log\frac{q\theta}{p(1-\theta)} - n\log\frac{2q}{1-\theta}\} < M < \infty \qquad a.s.$$

for every $u \in G_n$ and for all n, and indeed the same must be true for the exponent:

$$S^+(\xi,u)\log\frac{q\theta}{p(1-\theta)} - n\log\frac{2q}{1-\theta} < M < \infty \qquad a.s.$$

for every $u \in G_n$ and for all n. In particular, this statement also holds if we take the maximum over vertices in level n:

$$\max_{|u|=n}\{S^+(\xi,u)\log\frac{q\theta}{p(1-\theta)} - n\log\frac{2q}{1-\theta}\} < \infty \qquad a.s. \tag{1.4}$$

For $\theta > p$, $\log\frac{q\theta}{p(1-\theta)} > 0$, and so we may divide through by it:

$$\max_{|u|=n}\{S^+(\xi,u) - c_p^\theta n\} < \infty \qquad a.s.,$$

where

$$c_p^\theta := \frac{\log\frac{2q}{1-\theta}}{\log\frac{q\theta}{p(1-\theta)}}, \qquad \theta \neq p.$$

In the case $\theta < p$ there are two possibilities. When $\max(0, p-q) < \theta < p$, we have

$$\log\frac{q\theta}{p(1-\theta)} < 0, \qquad \log\frac{2q}{1-\theta} > 0,$$

while when $0 < \theta \leq \max(0, p-q)$, we have

$$\log\frac{q\theta}{p(1-\theta)} < 0, \qquad \log\frac{2q}{1-\theta} \leq 0.$$

Considering each case separately, it is easily seen that, for $\theta < p$,

$$\max_{|u|=n}\{c_p^\theta n - S^+(\xi, u)\} < \infty \qquad a.s.$$

In Chapter 2 we will be interested in when $c_p^\theta \in [0, 1/2]$ or $[1/2, 1]$ and will use statements of the above form to make deductions about the asymptotic behaviour of $S^+(\xi, u)$. As was the case earlier, the particular choice $\theta = 1/2$ will prove crucial. The following propositions give a geometric interpretation of c_p^θ and illustrate, once again, the strong connection between the martingales Λ_n^θ and cloud speed.

Proposition 5. *Let $T_\theta(x)$ be the tangent to the rate function I at θ, where it is assumed that $\theta > p$. Then c_p^θ satisfies*

$$T_\theta(c_p^\theta) = \log 2,$$

that is, c_p^θ is the x-coordinate of the point at which the tangent to the rate function at θ intersects the line $y = \log 2$.

Proof. We have

$$\frac{T_\theta(x) - I(\theta)}{x - \theta} = I'(\theta)$$

and hence $T_\theta(x) = \log 2$ if and only if x satisfies

$$I(\theta) + I'(\theta)(x - \theta) = \log 2$$

which re-arranges to give

$$x = \frac{\log 2 - I(\theta)}{I'(\theta)} + \theta$$

$$= \frac{\log 2 - \log \frac{1-\theta}{q}}{\log \frac{q\theta}{p(1-\theta)}}$$

$$= \frac{\log \frac{2q}{1-\theta}}{\log \frac{q\theta}{p(1-\theta)}} = c_p^\theta.$$

□

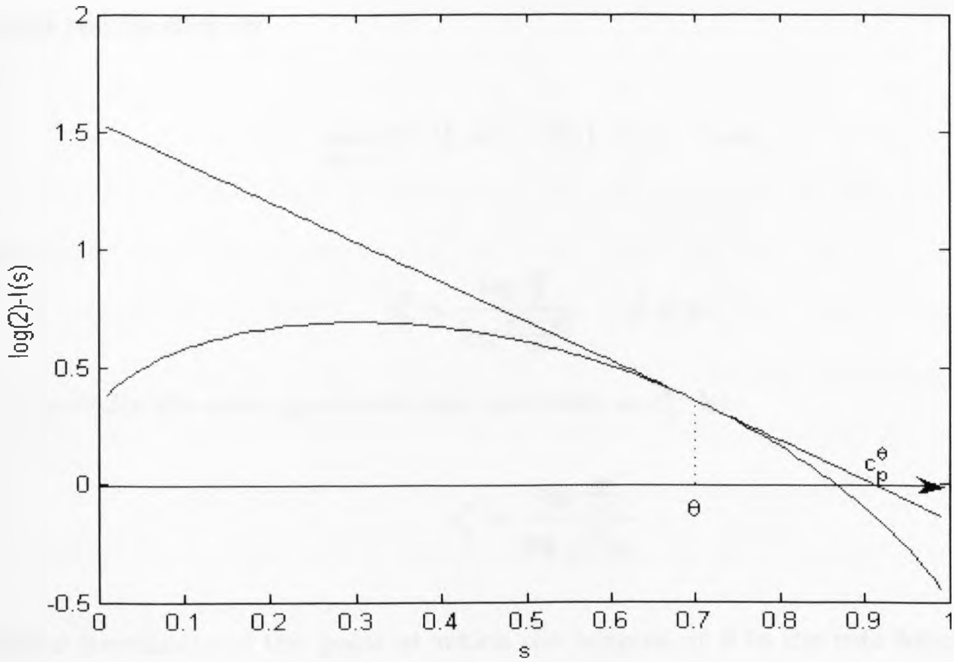

Figure 1.2: graphical interpretation of c_p^θ

Lemma 8. *The cloud speed* $\overline{c}_+(\xi)$ *is given, almost surely, by*

$$c^* = \inf_{\theta > p} c_p^\theta.$$

Proof. We have

$$\inf_{\theta > p} c_p^\theta = \sup\{\theta : \theta \leq c_p^\theta\}$$

$$= \sup\{\theta : \theta \leq \frac{\log 2 + \log q - \log(1-\theta)}{\log q + \log \theta - \log p - \log(1-\theta)}\}$$

$$= \sup\{\theta : \theta \log \frac{\theta}{p} + (1-\theta) \log \frac{1-\theta}{q} \leq \log 2\}$$

$$= \sup\{\theta : I(\theta) \le \log 2\},$$

which is (almost surely) the cloud speed. □

If we were to express Λ_n^θ in terms of $S^-(\xi, u)$ rather than $S^+(\xi, u)$ then we would end up with results such as

$$\max_{|u|=n}\{S^-(\xi, u) - d_p^\theta n\} < \infty \qquad a.s.,$$

where

$$d_p^\theta = \frac{\log \frac{2p}{\theta}}{\log \frac{p(1-\theta)}{q\theta}}, \qquad \theta \neq p$$

has essentially the same geometric interpretation as c_p^θ: let

$$e_p^\theta = \frac{\log \frac{2p}{1-\theta}}{\log \frac{p\theta}{q(1-\theta)}}$$

be the x-coordinate of the point at which the tangent at θ to the rate function $\widetilde{I}$ (corresponding to a random walk with independent Bern(q) increments - i.e. $\widetilde{I}$ is the same as I but with p and q interchanged) intersects the line $y = \log 2$. Then $d_p^\theta = e_p^{1-\theta}$. Indeed the analogous interpretations are true if we obtain the corresponding quantities by expressing Λ_n^θ in terms of $S(\xi, u)$.

A discussion of large deviation calculations would not be complete without mention of Hausdorff dimension. The result we state is perhaps the most refined of all, extending Hawkes' theorem [33], and has strong links with both cloud speed and our martingales.

1.8 Hausdorff dimension

We discuss Hausdorff dimension with great brevity as it doesn't really play a part in the later theory and any role it does play is, in essence, through our earlier calculations. A thorough discussion and consideration of the topic can be found in works such as [33] and [14].

The following result is a consequence of a theorem of Lalley and Selke in [33]. In their work, the setting is far more general than ours in that the types associated with the vertices of the underlying Galton-Watson tree are not assumed independent and identically distributed. This is one of the reasons for our not giving a more detailed account: a complete description of the problem would, at the very least, require a digression into rate functions for Markov chains, a subject which is remote from our concerns. For this reason we simply state the result of the theorem as it relates to our setup and then indicate how it is consistent with our earlier calculations.

Let $\xi = (\xi_1, \xi_2, ...), \xi' = (\xi'_1, \xi'_2, ...) \in \{-1, 1\}^{\mathbb{Z}^+}$. Then we define a metric d by $d(\xi, \xi') = 2^{-n(\xi,\xi')}$, where $n(\xi, \xi')$ is the smallest positive integer such that $\xi_n \neq \xi'_n$. Let μ_θ be the Bern(θ) product measure on $\{-1, 1\}^{\mathbb{Z}^+}$. We say that a sequence $\omega = (\omega_1, \omega_2, ...) \in \{-1, 1\}^{\mathbb{Z}^+}$ is μ_θ-generic if every finite sequence (or "word") $x = (x_1, x_2, ..., x_n)$, with entries in $\{-1, 1\}$, occurs with limiting relative frequency $\mu_\theta(\{\omega : \omega_1 = x_1, ..., \omega_n = x_n\})$. The following theorem tells us about the Hausdorff dimension of the subset of the boundary of T corresponding to μ_θ generic sequences.

Theorem 4. *(consequence of Lalley's and Selke's theorem, [33]) Let $\Sigma_{\mu_\theta}(\xi)$ be a random subset of ∂T corresponding to μ_θ generic lines of descent. Then if $I(\theta) > \log 2$, $\Sigma_{\mu_\theta}(\xi)$ is almost surely empty while if $I(\theta) \leq \log 2$, then, almost surely, in the metric d, the*

Hausdorff dimension of $\Sigma_{\mu_\alpha}(\xi)$ *is given by*

$$\dim_H(\Sigma_{\mu_\theta}(\xi)) = 1 - \frac{I(\theta)}{\log 2}.$$

See the earlier sketch of $\log(2) - I$ for an idea of how $\dim_H$ varies with θ. Now, let $\Sigma_\theta(\xi)$ be a random subset of ∂T corresponding to marked lines of descent with limiting proportion θ of type one marks. Then $\Sigma_{\mu_\theta}(\xi) \subseteq \Sigma_\theta(\xi)$ and hence the Hausdorff dimension stated for Σ_{μ_θ} is a lower bound for that of $\Sigma_\theta(\xi)$. Heuristically, we expect that there is in fact equality since the set of sequences having the correct proportion of ones but which are not generic is very "small" compared to the set of generic sequences. The appearance of the rate function I in the calculations illustrates the connection between the various ideas discussed in this chapter.

We conclude this section and chapter with a very brief discussion of $\{-1,1\}^T$ as a metric space. The metric is a sequence metric analogous to the metric d introduced in relation to Hausdorff dimension, except that we work in terms of the levels of the tree. Define a metric D on $\{-1,1\}^T$ via

$$D(\xi,\xi') = 2^{-n(\xi,\xi')},$$

where $n(\xi,\xi')$ is the first level of the tree (relative to the root) in which ξ and ξ' differ (that is, $n(\xi,\xi') = k$ if and only if there exists a vertex $u \in G_k$ with $\xi_u \neq \xi'_u$). In this setting the open ϵ-ball centered at ξ consists of those marked trees which agree with ξ on the first $\lfloor \frac{-\log \epsilon}{\log 2} \rfloor$ levels of the tree, where we have assumed that $0 < \epsilon < 1$. The metric space $(\{-1,1\}^T, D)$ is complete and separable. Furthermore the metric D induces a topology on $\{-1,1\}^T$, which is in fact the product topology.

Chapter 2

Recovery problems

In example 4.7 of [43], Tsirelson and Vershik describe a recursive process on a binary tree and an associated continuous product of probability spaces. Consideration of the random variable associated with the root leads to the following multi-type branching process. In this process, each individual has either one or two offspring and we denote the possible types by S and C. A type C individual has, with equal probability, either two type C offspring or two type S offspring. A type S individual has either a type S and a type C offspring, with probability $1/9$, or just one type S offspring, with probability $8/9$. Starting with an individual of unknown type, a random tree is constructed from the branching process in the usual way. Suppose that we observe the tree but not the types. Can we deduce the type of the individual at root?

We do get some information about the types from the shape of the tree: each time we observe only one offspring we know that the parent has type S. Indeed each time we are able to deduce the types of two offspring we are able to deduce the type of the parent. In this way we are able to construct some of the types associated with the tree, which may or may not include the type at root. If we regard the types as possible "explanations" for the observed tree, an obvious question to ask is whether there is more than one consistent explanation. It may be that there are several explanations consistent with the local rules but that all but one can be ruled out on probabilistic grounds on the infinite tree. Going back to the example in [43], being able to recover the the types from the shape of the tree is equivalent to the existence of a cyclic vector for the Von Neumann algebra generated by the projectors of the continuous product.

The recovery problem we study in this chapter is somewhat simpler than the one just described. It was devised as a model that displayed the same sort of phenomena but is easier to study. We build up to the model and motivate its analysis by first describing an analogous problem on the integers.

2.1 A reconstruction problem on the integers

We present here a very simple example of the type of problem to be studied later. The setup is as follows. Consider an integer-indexed sequence $X = (X_u; u \in \mathbb{Z})$, of independent, identically distributed Bern(p) random variables taking values in the two-point set $\{-1, 1\}$. Define another integer-indexed sequence $Y = (Y_u; u \in \mathbb{Z})$ of random variables by setting $Y_u = X_u X_{u+1}$ for every $u \in \mathbb{Z}$. Suppose now that we "observe" the Y_u. That is, suppose we are given the sequence Y. What information can we deduce about X from Y? Is it possible to "recover" the X_u from the Y_u? We will see below that it is indeed possible to recover the X_u (i.e. determine explicit values), provided $p \neq 1/2$.

Proposition 6. *The values of Y determine uniquely the values of X almost surely, provided $p \neq 1/2$.*

Proof. To see this, suppose we observe an arbitrary Y_k. We "recover" the X_u as follows. Let $X_k = x = \pm 1$. The choice is arbitrary (see later). Now, observe Y_{k+1}. If $Y_{k+1} = 1$, then set $X_{k+1} = x$. If $Y_{k+1} = -1$, then set $X_{k+1} = -x$. Now observe Y_{k+2} and do the same: if $Y_{k+2} = 1$ then set $X_{k+2} = X_{k+1}$ but if $Y_{k+2} = -1$, then set $X_{k+2} = -X_{k+1}$. One can continue in this way, extending to the indices before k in precisely the same fashion. That is, inductively, we set $X_{k\pm n} = X_{k\pm n\mp 1}$ if $Y_{k\pm n} = 1$ and $X_{k\pm n} = -X_{k\pm n\mp 1}$ if $Y_{k\pm n} = -1$, for any $n \in \mathbb{Z}$. We have thus defined a sequence $(X_u; u \in \mathbb{Z})$ with the property that $X_u X_{u+1} = Y_u$ for every $u \in \mathbb{Z}$, that is we have obtained a possible reconstruction for the X_u consistent with the observed Y_u. Of course we made an arbitrary choice $X_k = x$ at the beginning. The sequence resulting from the initial choice $X_k = -x$ also satisfies $X_u X_{u+1} = Y_u$ for every $u \in \mathbb{Z}$ and is therefore also a possible "reconstruction" for X consistent with the observed Y.

Having found two possible reconstructions for X consistent with Y we would like to rule out one of the possibilities. Thankfully, the strong law of large numbers comes to the rescue.

It is a special case of the Strong Law of Large Numbers that the limiting proportion u such that $X_u = 1$ is almost surely equal to p. Hence we are able to "distinguish" between the two possible reconstructions for X. Denote the two possible reconstructions by X, X'. Then, by virtue of the reconstruction, we have $X = -X'$. Now, almost surely, either X or X' (but not both!) has limiting proportion p of ones (and therefore proportion $1 - p$ of minus ones) by the above claim which means that the other has, almost surely, proportion $1 - p$ of ones since the minus sign interchanges the plus and minus ones. We are therefore able, almost surely, to distinguish between the two possible reconstructions on the basis of only one being consistent with the strong law of large numbers. That is, unless $p = 1/2$, in which case both possible reconstructions are consistent. □

We now present an analogous formulation for the above problem on the binary tree and make use of the theory from Chapter 1 to obtain a solution.

2.2 A recovery problem on the binary tree T

As with the problem on the integers, let $\xi = (\xi_u; u \in T)$ be a T-indexed collection of independent, identically distributed Bern(p) random variables taking values in $\{-1, 1\}$. Define random variables $(\eta_u; u \in T)$ by

$$\eta_u = \xi_u \xi_{u0} \xi_{u1}, \quad u \in T.$$

For $u \in T$ we think of the nodes $u, u0, u1$ as comprising a "family". In this setting we say that a node u has type ξ_u and we think of η_u as being the corresponding family type. As before, we are interested in knowing whether we can in some sense"recover" or "reconstruct" the vertex types from the family types. It is immediately clear that the present situation is far more complex than before: one need only interchange minus and plus ones along an infinite marked line of descent from the root and the resulting tree-indexed sequence will have the same family types as the original on account of the fact that any family that intersects this line of descent does so in precisely two places and therefore the effect of interchanging the signs cancels. There are uncountably many marked infinite lines of descent corresponding to uncountably many points in the boundary of T and hence one could potentially obtain uncountably many different "explanations" consistent with observed family types. The results from Chapter 1 are, nevertheless, encouraging and motivate further investigation.

2.2.1 The recovery case $pq \le 1/16$

Recall the work on cloud speed and the rate function I for the random walk with Bern(p) increments. We have the following simple result.

Lemma 9. *Let $pq < 1/16$ with $p < 1/2$. Then $\overline{c}_+(\xi) < 1/2$ almost surely. Let $pq < 1/16$ with $p > 1/2$. Then $\overline{c}_-(\xi) < 1/2$ almost surely.*

Proof. We use our earlier characterisation of $\overline{c}_+$ in terms of the supremum of a rate function. Recall that for $p > 1/2$, we have $\overline{c}_+ = 1$ almost surely. It is therefore necessary to impose the constraint $p < 1/2$. Under this constraint, since $I(a)$ is increasing for $a \ge p$, we have, by the theorem of Hammersley, Kingman and Biggins, $\overline{c}_+ < 1/2$ if and only if $I(1/2) > \log(2)$. Using our formula for $I(a)$, we require then

$$\frac{1}{2}\log\frac{1}{2p} + \frac{1}{2}\log\frac{1}{2q} > \log 2$$

which simplifies to requiring

$$\log\frac{1}{4pq} > \log 4.$$

After exponentiating, the condition becomes $pq < 1/16$. The proof of the second part of the lemma now follows immediately by interchanging p and q. □

Informally, all possible limiting proportions of type one (or minus one) vertices are less than $1/2$. If would follow then, that after "flipping" in the way described above (along an infinite marked line of descent in order to obtain a different tree-indexed sequence with the same family types as the original), the corresponding proportion of ones (or minus ones) would be greater than $1/2$ and, therefore, implausible. The same type of result is also motivated by the Hausdorff dimension calculations. Recall $\Sigma_\theta(\xi)$, the subset of the boundary corresponding to lines of descent with limiting proportion θ of type one vertices and the corresponding formula Hausdorff dimension given in terms of the rate function I from Chapter 1.

Claim 2. *Let $pq < 1/16$. Then the possible values α of (limiting) proportions (i.e. the α for which $\Sigma_\alpha(\xi)$ is non-empty) lie on the same side of 1/2.*

Proof. This is the same calculation as for the cloud speed in the previous lemma. □

It seems then that we are able to argue that recovery is possible for $pq < 1/16$ (though we will need to make this precise). It turns out that $pq = 1/16$ is the "critical case" and for this we will need to use arguments involving the asymptotic behaviour of $S^+(\xi, u)$, discussed briefly in Chapter 1.

Proposition 7. *Let $pq \leq 1/16$ with $p \neq 0, 1$. When $p < 1/2$ we have*

$$\max_{|u|=n}\{S^+(\xi, u) - \frac{n}{2}\} \to -\infty \qquad a.s.$$

whereas when $p > 1/2$ we have

$$\max_{|u|=n}\{\frac{n}{2} - S^+(\xi, u)\} \to -\infty \qquad a.s.$$

Proof. This is a consequence of the argument advanced in the section on the asymptotic behaviour of $S^+(\xi, u)$ in Chapter 1. Setting $\theta = 1/2$ in the Λ_n^θ martingales gives

$$\Lambda_n^{1/2} \equiv \Lambda_n = \frac{1}{2q}\sum_{|u|=n}\frac{1}{(4q)^n}(\frac{q}{p})^{S^+(\xi,u)} = \frac{1}{2q}\exp\{\log\frac{q}{p}S^+(\xi, u) - n\log 4q\}$$

and

$$c_p^{1/2} \equiv c_p = \frac{\log 4q}{\log\frac{q}{p}}.$$

Suppose that $pq < 16$ with $p < 1/2$ (we will deal with the "critical" case later). Then it is easily seen that $c_p \in (0, 1/2)$. It follows that $n/2 > c_p n$ and hence we have

$$S^+(\xi, u) - \frac{n}{2} = S^+(\xi, u) - c_p n - (\frac{1}{2} - c_p)n,$$

where $1/2 - c_p > 0$. Now, since

$$\max_{|u|=n}\{S^+(\xi, u) - c_p n\} < \infty \quad a.s.$$

it follows that

$$\max_{|u|=n}\{S^+(\xi, u) - \frac{n}{2}\} \to -\infty \quad a.s.$$

since $(1/2 - c_p)n \to \infty$ as $n \to \infty$. Similarly, when $pq < 1/16$ with $p > 1/2$, we have $c_p \in (1/2, 1)$. We also have

$$\frac{n}{2} - S^+(\xi, u) = c_p n - S^+(\xi, u) - (c_p - \frac{1}{2})n,$$

where $c_p - 1/2 > 0$. Now,

$$\max_{|u|=n}\{c_p n - S^+(\xi, u)\} < \infty \quad a.s.$$

so that

$$\max_{|u|=n}\{\frac{n}{2} - S^+(\xi, u)\} \to -\infty \quad a.s.$$

since $(c_p - 1/2)n \to \infty$ as $n \to \infty$. The case $pq = 1/16$ is more delicate, requiring us to use the fact that Λ_n converges to zero almost surely. By Proposition [?] from Chapter 1, Λ_n converges to zero almost surely precisely when

$$I(1/2) \geq \log 2,$$

that is, when $pq \le 1/16$ so that, in particular, this is the case when $pq = 1/16$. We deduce that, for any $u \in G_n$,

$$\exp\{S^+(\xi, u)\log\frac{q}{p} - n\log 4q\} \to 0 \quad a.s.$$

so that, for any $u \in G_n$,

$$S^+(\xi, u)\log\frac{q}{p} - n\log 4q \to -\infty \quad a.s.$$

and hence

$$\max_{|u|=n}\{S^+(\xi, u)\log\frac{q}{p} - n\log 4q\} \to -\infty \quad a.s.$$

The result now follows by dividing through by $\log\frac{q}{p}$ and treating the cases where this is positive and negative separately, as before. $\square$

This suggests the following.

Lemma 10. *Let*

$$U = \{\omega : \max_{|u|=n}\{S^+(\xi(\omega), u) - \frac{n}{2}\} \to -\infty\}.$$

Let $\eta : \overline{\Omega} \to \{-1, 1\}^T$ *be the family type mapping. That is,* $\eta(\omega)(\eta_u(\omega); u \in T)$, *where* $\eta_u(\omega) = \xi_u(\omega)\xi_{u0}(\omega)\xi_{u1}(\omega)$. *Then the restriction of* η *to* U *is injective when* $pq \le 1/16$ *with* $p < 1/2$.

Proof. We begin by convincing ourselves of the truth of the following: $\xi \in U$ implies

$$S^+(\xi, u(n,t)) - \frac{n}{2} \to -\infty$$

for all $t \in \partial T$, where $u(n,t)$ is a vertex u in level n (i.e. $|u| = n$), on the line of descent "ending" at t. This follows from that fact that

$$S^+(\xi, u(n,t)) - \frac{n}{2} \leq \max_{|u|=n}\{S^+(\xi, u(n,t)) - \frac{n}{2}\}$$

for every n and for all $t \in \partial T$.

Now, let $\xi, \xi' \in U$. We argue that $\xi \neq \xi' \Rightarrow \rho(\xi) \neq \rho(\xi')$. Suppose that ξ and ξ' differ on a vertex $u^* \in T$ (that is, $\xi_{u^*} \neq \xi'_{u^*}$). Suppose further that $\eta(\xi) = \eta(\xi')$. Then it must be the case that $\xi_v = -\xi'_v$ for all $w \leq v$ (recall that $w \leq v$ means that v is a descendant of w), where $m = |w| \leq |u^*|$. This is due to the fact that having changed a vertex type there are two ways in which the family types may be maintained. The first is by changing parent and daughter vertices, that is, changing along a line of descent. But one can also change sister vertex type or types at some level on or above $|u^*|$. In this way we must have $\xi_v = -\xi'_v$ for $w \leq v$, with w the first time (relative to u^*) that we change a sister rather than parent vertex. Denote this distinguished line of descent, starting at w, by $\hat{t}$. Define, for $|w| \leq |u|$,

$$S^+(\xi, w, u) = \sum_{w \leq v \leq u} 1_{(\xi_v = 1)}(\xi)$$

and

$$S^-(\xi, w, u) = \sum_{w \leq v \leq u} 1_{(\xi_v = -1)}(\xi).$$

Then we have $S^+(\xi, w, u) \leq S^+(\xi, u)$ and hence by our opening remark,

$$S^+(\xi, w, u(n, \hat{t})) - \frac{n}{2} \rightarrow -\infty.$$

We now observe that

$$S^+(\xi', w, u(n, \hat{t})) = S^-(\xi, w, u(n, \hat{t}))$$

and use the fact that, for $u \in G_n$,

$$S^+(\xi, w, u) + S^-(\xi, w, u) = n + 1 - m$$

to see that

$$n + 1 - m - S^+(\xi', w, u(n, \hat{t})) - \frac{n}{2} \to -\infty$$

which is equivalent to having

$$S^+(\xi', w, u(n, \hat{t})) - \frac{n}{2} \to \infty,$$

that is,

$$S^+(\xi, u(n, t)) - \frac{n}{2} \to \infty,$$

contradicting the fact that $\xi' \in U$. □

Lemma 11. *Let*

$$V = \{\omega : \max_{|u|=n}\{\frac{n}{2} - S^+(\xi, u)\} \to -\infty\}.$$

Then the restriction of η *to* V *is injective when* $pq \leq 1/16$ *with* $p > 1/2$.

Proof. The argument is analogous to the previous one, though here we use the fact that $S^+(\xi, w) < \infty$, that is, only finitely many ones occur before w. □

We have found sets of probability one on which the family type map is injective. More precisely, when $pq \leq 1/16$, either U or V has probability one and η is injective on the set in question. Notice that this is an improvement on the results that would be obtained via cloud speed (or indeed Hausdorff dimension) as it takes care of the case $pq = 1/16$.

The advantage of the result obtained via the martingale theory over the cloud speed calculation is one of finer detail. In the case $pq = 1/16$, the cloud speed is almost surely 1/2. That is,

$$\max_{|u|=n} S(\xi, u) = \frac{n}{2} + o(n).$$

Our result suggests that the coefficient of the second order term is negative. See [8] for details of the second order term in the context of branching Brownian motion.

We now have the ingredients required to prove the following result which makes precise the idea that the family and individual vertex types carry the same amount of "information" when $pq \leq 1/16$. We will formulate this in terms of the σ-algebras generated by the individual and family types. For this reason we need to clarify precisely what we mean by the "σ-algebras generated". Let

$$\xi = (\xi_u; u \in T), \qquad \eta = (\eta_u; u \in T).$$

Then $\xi, \eta : \overline{\Omega} \to \{-1,1\}^T$. The σ-algebras generated by ξ, η will depend on the σ-algebras with which we equip the target space $\{-1,1\}^T$. For the purposes of this work (and throughout) we will assume that $\{-1,1\}^T$ is equipped with the Borel σ-algebra $\mathcal{B}$ so that

$$\sigma(\xi) = \{\xi^{-1}A : A \in \mathcal{B}\}$$

and similarly for $\sigma(\eta)$. Notice that it makes sense to talk about the Borel σ-algebra on $\{-1,1\}^T$ because it has a metric structure, discussed at the end of Chapter 1.

Before stating and proving the result we have in mind we recall a corollary of Kuratowski's Theorem concerning the image of Borel sets under injective mappings. Recall the metric structure of $\{-1,1\}^T$ via the metric D from Chapter 1.

Theorem 5. *(Kurtawoski's Theorem, [31]) Let X_1, X_2 be complete separable metric spaces and $E_1 \subseteq X_1, E_2 \subseteq X_2$ subsets with E_1 Borel. If $E_2 = \phi(E_1)$, where ϕ is a measurable injective map from X_1 into X_2, then E_2 is Borel.*

Theorem 6. *Suppose that $pq \leq 1/16$, where $q = 1 - p$. Then the σ-algebras generated by the individual vertex and family types are the same up to null sets. That is, their completions are equal.*

Proof. We will deal with the case $p < 1/2$ and leave the reader to make the analogous argument for $p > 1/2$ (which is the same as that given with the substitution of V for U). Given $A \in \sigma(\xi_u; u \in T)$, we exhibit an $A' \in \sigma(\eta_u; u \in T)$ such that $\mathbb{P}(A \triangle A') = 0$. It is sufficient to do this for the canonical space $\overline{\Omega} = \{-1, 1\}^T$ which we equip with the Borel σ-algebra (the justification of this result in the general setting is discussed after the proof). For $A \in \sigma(\xi_u; u \in T)$, let

$$A' = \{\xi : \eta(\xi) \in \eta(U \cap A)\}.$$

Then $A' = \eta^{-1}B$, where $B = \eta(U \cap A)$. Now, A is Borel since the coordinate maps ξ_u generate the Borel σ-algebra and hence $\eta(U \cap A)$ is Borel by Kuratoski's Theorem since η is injective on U. It follows that $A' \in \sigma(\eta) = \sigma(\eta_u; u \in T)$. We now claim that $U \cap A = U \cap A'$. To see this, suppose that $\xi \in U \cap A$. We have then $\eta(\xi) \in \eta(U \cap A)$ so that $\xi \in A'$ and hence $\xi \in U \cap A$. On the other hand, suppose that $\xi \in U \cap A'$. Then $\eta(\xi) \in \eta(U \cap A)$ and, by the injectivity of η on U, this implies that $\xi \in U \cap A$. Hence $A \triangle A' \subseteq U^c$ so that $\mathbb{P}(A \triangle A') = 0$. □

The reason for working on the canonical space in the above proof is that it enabled us to apply Kurtawoski's Theorem. We now show that the conclusion of the theorem holds for any arbitrary probability space. This type of technical detail is similar to the issue discussed in Chapter 1 in relation to defining the change of measure Λ_n^θ on non-canonical spaces.

Our abstract probability space is $(\overline{\Omega}, \mathcal{F}, \overline{\mathbb{P}})$, where we have written the bar over the probability to avoid confusion with the canonical probability measure. Let $\overline{\xi}, \overline{\eta} : (\overline{\Omega}, \mathcal{F}) \to (\{-1,1\}^T, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra, denote the non-canonical individual and family types so that $\overline{\eta} = \eta \circ \overline{\xi}$. Our aim is to extend the result of Theorem 6: given an $\overline{A} \in \sigma(\overline{\xi})$, we seek an $\overline{A}' \in \sigma(\overline{\eta})$ such that $\overline{\mathbb{P}}(\overline{A} \Delta \overline{A}') = 0$. Let $\overline{A} = \overline{\xi}^{-1} A$, where $A \in \mathcal{B}$ is arbitrary. Define $\overline{A}' = \overline{\eta}^{-1} B$, where, as in the proof of the theorem in the canonical case, $B = \eta(U \cap A)$. Then

$$\overline{\mathbb{P}}(\overline{A} \setminus \overline{A}') = \overline{\mathbb{P}}(\overline{\xi}^{-1} A \setminus \overline{\eta}^{-1} B)$$

$$= \overline{\mathbb{P}}(\overline{\xi}^{-1} A \setminus (\overline{\xi}^{-1} \circ \eta^{-1}) B)$$

$$= \overline{\mathbb{P}}(\omega : \overline{\xi}(\omega) \in A \setminus \eta^{-1} B) = \mathbb{P}(A \setminus \eta^{-1} B)$$

$$= \mathbb{P}(A \setminus A') = 0,$$

by the result in the canonical case, where $\mathbb{P}$ is the canonical probability measure or, equivalently, the law of $\overline{\xi}$ under $\overline{\mathbb{P}}$. It is easily verified by making essentially the same calculation that $\overline{\mathbb{P}}(\overline{A}' \setminus \overline{A}) = 0$ and the result now follows.

Intuitively, the "$pq \leq 1/16$" condition is telling us that "recovery" is possible provided p and q are sufficiently different. $pq \leq 1/16$ (with $p \neq 0, 1$) corresponds to

$$p \in (0, \frac{1}{2} - \frac{\sqrt{3}}{4}) \cup (\frac{1}{2} + \frac{\sqrt{3}}{4}, 1),$$

that is, p is either very large (so that q is very small) or very small (so that q is very large).

Notice also that the above recovery result concerning the equivalence of the σ-algebras generated by the individual and family types isn't a million miles away from the recovery result we obtained on the integers. If we write S for event that the limiting proportion of ones in X converges to p then, as we have seen, S has probability one by the strong law of large numbers. Furthermore, the family type map is injective on S when $p \neq 1/2$ since, of the two possible reconstructions, one will have (almost surely) limiting proportion p of ones while the other has (almost surely) limiting proportion $1-p$. We can therefore conclude, by running the above argument, that the completions of the σ-algebras generated by X and the Y are the same, provided $p \neq 1/2$.

2.2.2 Non-recovery: the case $pq > 1/16$

We saw in the last section that, when $pq \leq 1/16$, there exist sets of probability one on which the family type map is injective. We then showed that this implies, via the equivalence of the (completions of) σ-algebras generated by the individual vertex and family types, that the "information" carried by the vertex and family types is, in some sense, the same. From the section on cloud speed (or Hausdorff dimension) we know that, when $pq > 1/16$, there exist almost surely marked trees having lines of descent with proportion $1/2$ of type one vertices. One ought then to be able to swap the signs of the vertices along this line in order to obtain another marked tree with the same family types as the original, making recovery seemingly impossible in this case. From the section on cloud speed, we can deduce that when $pq > 1/16$, $c_+(\xi) \geq 1/2$ almost surely but then contradictions of the type we derived in the case $pq < 1/16$ will not be possible. The rest of this section will use martingale methods following on from the previous chapter to develop results that can be used collectively to prove the following theorem.

Theorem 7. *Let $pq > 1/16$. Then we have strict inclusion between the σ-algebras gen-*

erated by the individual vertex and family types:

$$\overline{\sigma(\xi_u; u \in T)} \supsetneq \overline{\sigma(\eta_u; u \in T)}.$$

Recall the probability measure $\mathbb{Q}$ (with $\theta = 1/2$) from chapter one as the probability under which the ξ_u are independent, identically distributed Bern(p) random variables, except along a random line of descent, uniformly chosen on the boundary of the tree, along which the ξ_u are independent $Bern(1/2)$ random variables, independent of the random variables not on this distinguished line of descent. Our strategy will involve a "coupling" argument. That is, we will construct marked trees ξ, ξ' having the same law under $\mathbb{Q}$ and with the same family types.

Claim 3. *Let $t \in \partial T$ be chosen uniformly. Define a marked tree ξ by giving ξ_u the Bern(1/2) distribution for $u \leq t$ and the Bern(p) distribution for $u \not\leq t$. Define another marked tree ξ' by $\xi'_u = -\xi_u$ for $u \leq t$ and $\xi'_u = \xi_u$ otherwise. Then ξ, ξ' have the same distribution under $\mathbb{Q}$ and $\rho(\xi) = \rho(\xi')$.*

This is straightforward to see and is made explicit only because it plays a role in the following proposition.

Proposition 8.

$$\overline{\sigma(\xi_u; u \in T)}^{\mathbb{Q}} \supsetneq \overline{\sigma(\eta_u; u \in T)}^{\mathbb{Q}}.$$

Proof. Let $\xi_\emptyset$ denote the type at root. The event $A = \{\omega : \xi_\emptyset(\omega) = 1\}$ is contained in $\overline{\sigma(\xi_u; u \in T)}^{\mathbb{Q}}$. We will show that it is not contained in the other σ-algebra. Suppose, for a contradiction, that $A \in \overline{\sigma(\eta_u; u \in T)}^{\mathbb{Q}}$. Then there exists an $H \in \mathcal{B}$ such that $\eta^{-1}(H) \triangle A$ is $\mathbb{Q}$-null. Notice also that $\eta^{-1}(H)$ has a disjoint decomposition: $\eta^{-1}(H) = C \uplus D$, with $C = \{\omega : \xi_\emptyset(\omega) = 1, \eta(\omega) \in H\}$ and $D = \{\omega : \xi_\emptyset(\omega) = -1, \eta(\omega) \in H\}$, so that $D \subseteq A \cup N$ for some $\mathbb{Q}$-null set N. Now, since $D \cap A = \emptyset$, we have $D \subseteq N$ so that $\mathbb{Q}(D) = 0$.

On the other hand, we observe that $\mathbb{Q}(C \uplus D) = \mathbb{Q}(A) = 1/2$ and now we must have $\mathbb{Q}(C) = \mathbb{Q}(D)$ by the previous claim ($\xi(\omega) \in C$ if and only if $\xi'(\omega) \in D$ and therefore $\mathbb{Q}(\{\omega : \xi(\omega) \in C\}) = \mathbb{Q}(\{\omega : \xi'(\omega) \in D\}) = \mathbb{Q}(\{\omega : \xi(\omega) \in D\}))$ so that $\mathbb{Q}(D) = 1/4$, a contradiction. □

Notice that the existence of a line of descent along which the proportion of ones is 1/2 was vital: key to the above proof was the fact that ξ, ξ' have the same distribution under $\mathbb{Q}$. Had we chosen any other value of θ in our martingales Λ_n^θ then this would not have been the case. We see then that the choice $\theta = 1/2$ is crucial for the non-recovery part of the argument and we will see that for this value the two ends of the argument "meet". That is, we can prove that when we're not in the recovery regime then we are in the non-recovery regime and vice-versa. It was clear from the previous section (via the problems with determining when Λ_∞^θ is trivial) that this is not the case for other values of θ.

Proof. (Theorem 7) The result follows provided $\mathbb{P}$ and $\mathbb{Q}$ are equivalent when $pq > 1/16$. This is a corollary of Proposition 3 from Chapter 1: setting $\theta = 1/2$ the condition for uniform integrability of Λ_n is $I(1/2) < \log 2$, i.e. $pq > 1/16$. That is, when $pq > 1/16$, Λ_n converges in L^1 to a finite limit which is almost surely positive. □

2.3 Further work and unresolved problems

In this section we discuss briefly some of the ways in which one might extend the results of the previous sections. For example, one might consider having more types, working on a d-ary tree rather than the binary tree, defining the family type in a different way and so on. We introduce and discuss some of these possibilities, giving, in some cases, motivational results. We also mention briefly a more "classical" type of reconstruction problem on trees.

More types

Suppose that the ξ_u may take one of K types, where $K = 1, 2, ...$. Our immediate question is which (if any) of the results of the previous sections translate directly to this more general setting. We will not explore the type of family type mapping that might be appropriate in this context but illustrate instead the sort of arguments - extending the two type case from Chapter 2 - that might be exploited. We suspect that much of the work involving martingales obtained via a change of measure should work along the same lines as before.

Let $\mathbb{P}^K$ be the probability measure under which the ξ_u are independent, identically distributed Bern$(p_1, ..., p_K)$ random variables and $\mathbb{Q}^K$ the probability measure under which the ξ_u are independent, identically distributed Bern$(p_1, ..., p_K)$ random variables, except along a random line of descent, chosen uniformly on the boundary of the tree, along which the ξ_u are independent, identically distributed Bern$(1/K, ..., 1/K)$ random variables, independent of the ξ_u not on this line of descent. Writing

$$S^{t_i}(\xi, u) = \sum_{v \preceq u} \mathbf{1}_{(\xi_v = t_i)}(\xi)$$

for the number of type t_i vertices ($i = 1, ..., K$) observed along the line of descent to the vertex u, it is easily seen that, with obvious notation, (see derivation of the corresponding expression for two types in an earlier section for more details)

$$\Lambda_n^K = \frac{d\mathbb{Q}_n^K}{d\mathbb{P}_n^K} = \frac{1}{2^n} \sum_{|u|=n} (\frac{1}{Kp_1})^{S^{t_1}(\xi,u)} ... (\frac{1}{Kp_K})^{S^{t_K}(\xi,u)}.$$

We could now try to write this in a form consistent with Biggins' Theorem and then deduce results about uniform integrability and so on. It is perhaps more convenient, however, to appeal to the decomposition argument used in the previous section. The decomposition of the change of measure is a property of the underlying branching structure (i.e. the binary tree) and so the same decomposition holds here:

$$\Lambda_n^K = \frac{\Lambda_0^K}{2}(\Lambda_{n-1}^{K'} + \Lambda_{n-1}^{K''}),$$

where $\Lambda_{n-1}^{K'}, \Lambda_{n-1}^{K''}$ are (as before - see previous section for precise formulation) the corresponding changes of measure rooted at the vertices $0, 1$ respectively. Arguing along the same lines as before, we can now find a condition on $p_1, ..., p_K$ such that Λ_n^K is bounded in L^r for some $r > 1$. We write the argument out in brief, omitting some of the calculations as the analogous argument can be seen in the previous section.

We have

$$\mathbb{E}[(\Lambda_0^K)^r] = p_1(\frac{1}{Kp_1})^r + ... + p_K(\frac{1}{Kp_K})^r$$

so that

$$\mathbb{E}[(\Lambda_n^K)^r] \leq \frac{p_1(\frac{1}{Kp_1})^r + ... + p_K(\frac{1}{Kp_K})^r}{2^{r-1}}(\mathbb{E}[(\Lambda_{n-1}^K)^r] + 1).$$

Writing

$$\Theta_r = \frac{p_1(\frac{1}{Kp_1})^r + ... + p_K(\frac{1}{Kp_K})^r}{2^{r-1}},$$

we seek an $r \in (1, 2)$ such that $\Theta_r < 1$. Re-arranging, this inequality becomes

$$g(r) := \sum_{k=1}^{K}(\frac{1}{K})^r p_k^{1-r} - 2^{r-1} < 0.$$

Now,

$$g(1) = \sum_{k=1}^{K}\frac{1}{K} = 1$$

and

$$g'(1) = -\frac{1}{K}\sum_{k=1}^{K} \log Kp_k - \log 2 = -\log 2K(p_1...p_K)^{1/K}$$

so that $g'(1) < 0$ provided

$$p_1...p_K > (\frac{1}{2K})^K.$$

Notice that when $K = 2$ this gives the familiar condition "$pq > 1/16$".

It is not clear that this martingale would be useful (in the way Λ_n was with two types) in determining whether or not recovery is possible. Re-examining proposition 9 from the previous section, we suspect that under conditions on the family type mapping, we ought to be able to obtain strict inclusion between the completions (with respect to $\mathbb{Q}^K$) of the σ-algebras generated by the family and individual types which would then extend to the completions under $\mathbb{P}^K$ via the uniform integrability argument. Although the details are sketchy in this general setting, we do have, however, a concrete example with three types, where the family type is given by the product of the individual types as with our earlier work. Here, we argue that it is not possible to recover the individual vertex types from the family types for certain values of the parameters but the recovery part of the argument is more difficult and we obtain only a partial result.

An example with three types

Plus one and minus one are the square roots of unity and so it is perhaps natural to consider the cube roots of unity which we label, for convenience, as $t_1 = 1, t_2 = e^{2\pi i/3}, t_3 = e^{-2\pi i/3}$. As before, we define the family type to be the product of the types of the family members and we write η for the family type map. Recalling the notation we adopted above, we have then

$$\Lambda_n^3 = \frac{d\mathbb{Q}^3}{d\mathbb{P}^3} = \frac{1}{2^n}\sum_{|u|=n}(\frac{1}{3p})^{S^{t_1}(\xi,u)}(\frac{1}{3q})^{S^{t_2}(\xi,u)}(\frac{1}{3r})^{S^{t_3}(\xi,u)},$$

where we have written $p = p_1, q = p_2, r = p_3$ for the probabilities of types 1, 2, 3 respectively. It follows that Λ_n^3 is uniformly integrable provided $pqr > (1/6)^3$ and, furthermore, since $\lim_{n\to\infty}\Lambda_n^3$ is almost surely non-zero, we have that $\mathbb{Q}^3$ and $\mathbb{P}^3$ are equivalent when $pqr > (1/6)^3$. We now modify the argument from Proposition 7 to show that this gives us the non-recovery part of the argument.

Proposition 9. *Let $pqr > (1/6)^3$. Then*

$$\overline{\sigma(\xi_u; u\in T)}^{\mathbb{Q}^3} \supsetneq \overline{\sigma(\eta_u; u\in T)}^{\mathbb{Q}^3}.$$

Proof. We begin with a "coupling" in the spirit of the two type case. Let $t \in \partial T$ be chosen uniformly. Define a marked tree $\xi = (\xi_u; u \in T)$ by giving ξ_u the Bern(1/3) distribution for $u \leq t$ and the Bern(p) distribution for $u \not\leq t$. Define another marked tree ξ' by setting, for vertices $u \leq t$ such that $|u| = n$,

$$\xi'_u = \begin{cases} t_2\xi_u, & n = 0, 2, 4, ... \\ t_3\xi_u, & n = 1, 3, 5, ... \end{cases}$$

and $\xi'_u = \xi_u$ otherwise. Then it is easily seen that ξ, ξ' have the same distribution under $\mathbb{Q}^3$ and that $\eta(\xi) = \eta(\xi')$.

Now, let $A = \{\xi : \xi_\emptyset = 1\}$. Then $A \in \overline{\sigma(\xi_u; u\in T)}^{\mathbb{Q}^3}$. We will show that it is not contained in the other σ-algebra. Suppose that $A \in \overline{\sigma(\eta_u; u\in T)}^{\mathbb{Q}^3}$. We have $\sigma(\eta_u; u \in T) = \eta^{-1}(\mathcal{B})$ and hence there exists a set $H \in \mathcal{B}$ such that $\eta^{-1}(H)\triangle A$ is

$\mathbb{Q}^3$-null. Notice also that $\eta^{-1}(H)$ has a disjoint decomposition, $\eta^{-1}(H) = D_1 \uplus D_2 \uplus D_3$, with $D_i = \{\xi : \xi_\emptyset = t_i, \eta(\xi) \in H\}$, so that $D_2 \uplus D_3 \subseteq A \cup N$ for some $\mathbb{Q}^3$-null set N. Now, since $D_i \cap A = \emptyset$ for $i = 2, 3$, we have $D_2 \uplus D_3 \subseteq N$ so that $\mathbb{Q}(D_i) = 0$ for $i = 2, 3$. Recalling our marked trees ξ, ξ', observe that

$$\xi \in D_1 \Rightarrow \xi' \in D_2, \quad \xi \in D_2 \Rightarrow \xi' \in D_3, \quad \xi \in D_3 \Rightarrow \xi' \in D_1,$$

from which we deduce that

$$\mathbb{Q}^3(D_1) = \mathbb{Q}^3(D_2) = \mathbb{Q}^3(D_3).$$

Since $\mathbb{Q}^3(D_1 \uplus D_2 \uplus D_3) = \mathbb{Q}^3(A) = 1/3$, it now follows that $\mathbb{Q}^3(D_i) = 1/9$ for $i = 1, 2, 3$, contradicting the fact that D_2 and D_3 are $\mathbb{Q}^3$-null. □

Combining this result with the equivalence of $\mathbb{Q}^3$ and $\mathbb{P}^3$ we now have that

$$\overline{\sigma(\xi_u; u \in T)}^{\mathbb{P}^3} \supsetneq \overline{\sigma(\eta_u; u \in T)}^{\mathbb{P}^3}$$

when $pqr > (1/6)^3$.

The recovery part of the argument

By noting that

$$\sum_{i=1}^{3} S^{t_i}(\xi, u) = n + 1$$

we can write Λ_n^3 in three possible ways: in terms of S^{t_1} and S^{t_2}, in terms of S^{t_2} and S^{t_3} or in terms of S^{t_1} and S^{t_3}. For brevity we will express Λ_n^3 in terms of S^{t_1} and S^{t_2} and then point out how the argument would work if we had expressed it in terms of the other variables. We have then

$$\Lambda_n^3 = \frac{1}{3r}\sum_{|u|=n} \exp\{\log\frac{r}{p}S^{t_1}(\xi,u) + \log\frac{r}{q}S^{t_2}(\xi,u) - n\log 6r\}.$$

The idea is now the same as with the two type argument: we wish to obtain a

$$\text{“}\max_{|u|=n}\{S^{t_1}(\xi,u) + S^{t_2}(\xi,u) - \frac{n}{2}\} \to -\infty \qquad a.s.\text{”}$$

type result. Before we get to grips with how to obtain this, we will first verify that this is the right type of result.

Lemma 12. *The family type map η is injective on*

$$S = \{\xi : \max_{|u|=n}\{S^{t_1}(\xi,u) + S^{t_2}(\xi,u) - \frac{n}{2} \to -\infty\}\}.$$

Proof. We will only sketch the argument as it is analogous to the two type case. Let $\xi, \xi' \in S$. Then, as with the two type case, if ξ, ξ' differ on a vertex u, then it must be the case that $\xi'_v = t_j\xi_v$ for all $u \leq v$, where t_j corresponds to changing daughter vertices so as to maintain the family types amongst the families descended from u. Along this distinguished line of descent $t*$, we have

$$S^{t_1}(\xi, u(t*)) + S^{t_2}(\xi, u(t*) - \frac{n}{2} \to -\infty$$

Now, a type can be changed to either of the other two and hence we have

$$S^{t_1}(\xi', u) \leq S^{t_2}(\xi, u) + S^{t_3}(\xi, u),$$

$$S^{t_2}(\xi', u) \leq S^{t_1}(\xi, u) + S^{t_3}(\xi, u),$$

$$S^{t_3}(\xi', u) \leq S^{t_1}(\xi, u) + S^{t_2}(\xi, u).$$

Hence

$$S^{t_3}(\xi', u(t*)) - \frac{n}{2} \to -\infty$$

or

$$n + 1 - S^{t_1}(\xi', u(t*)) - S^{t_2}(\xi', u(t*)) - \frac{n}{2} \to -\infty$$

which holds if and only if

$$S^{t_1}(\xi', u(t*)) + S^{t_2}(\xi', u(t*)) - \frac{n}{2} \to \infty,$$

contradicting the fact that $\xi' \in S$. □

We now need to argue that S has probability 1 under certain conditions on p, q, r. As with the two type case, we have

$$\max_{|u|=n}\{\log \frac{r}{p} S^{t_1}(\xi, u) + \log \frac{r}{q} S^{t_2}(\xi, u) - n \log 6r\} < \infty \quad a.s.,$$

else Λ_n^3 would fail to converge and now it is just a case of conducting careful analysis. By taking $r > p$ we can divide through by $\log \frac{r}{p}$ and by taking $p \geq q$ we can make the coefficient of $S^{t_2}(\xi, u)$ greater than or equal to 1. The result now follows provided we can control the term

$$\frac{\log 6r}{\log \frac{r}{p}}$$

to be less than or equal to 1/2. This corresponds to having $pr \leq 1/36$. That is, S has probability 1 provided $pr \leq 1/36$ with $r > p \geq q$. Of course when $r < p$, we would be dividing through by a negative quantity and so we would then have to work with the set

$$S' = \{\xi : \max_{|u|=n}\{\frac{n}{2} - S^{t_1}(\xi, u) - S^{t_2}(\xi, u) \to -\infty\}\}.$$

It is readily seen that η is also injective on S' and that S' has probability 1 provided $pq \leq 1/36$ with $r < p$ and $p \geq q$. Recall now that we could have written Λ_n^3 in three different ways, depending on which of the $S^{t_j}(\xi, u)$ we use in the representation. Furthermore, in our argument above we divided through by $\log \frac{r}{p}$ but could equally have divided through by $\log \frac{r}{q}$ and then argued analogously. The analysis holds equally for each such representation and for division (provided it's permitted) by either of the coefficients of the S^{t_j}. It follows that there exist sets of probability 1 on which η is injective when

$$pr \leq 1/36, p \geq q, r \neq p, \quad qr \leq 1/36, q \geq p, r \neq q,$$

or when

$$pq \leq 1/36, q \geq r, p \neq q, \quad pr \leq 1/36, r \geq q, p \neq r,$$

or when

$$pq \leq 1/36, p \geq r, q \neq p, \quad qr \leq 1/36, r \geq p, q \neq r.$$

We can simplify these conditions to give

$$pq \leq 1/36, p \geq r; q \geq r, p \neq q,$$

$$pr \leq 1/36, p \geq q; r \geq q, p \neq r,$$

$$rq \leq 1/36, q \geq p; r \geq p, q \neq r.$$

This isn't quite the recovery result we hoped for, namely "$pqr \leq (1/6)^3$". As with many arguments, it is often relatively straightforward to prove that something is not true by finding a counterexample. In contrast, to prove that something is true requires a proof that works in generality. Here, we were able to construct marked trees having the same distribution under $\mathbb{Q}^3$ and with the same family types. This enabled us to complete the non-recovery part of the argument. Going back to the two type case, recall that if we change the type of a vertex then to maintain the corresponding family type we must also change one (but not both) of the types of the other two vertices comprising the family. This is not the case with the three type case. In some cases we can change both of the other vertex types and still maintain the family type. For example, consider

$$\xi_u = e^{2\pi i/3}, \quad \xi_{u0} = 1, \quad \xi_{u1} = e^{-2\pi i/3}$$

so that the corresponding family type is $\eta_u = 1$. Now suppose we change ξ_u to $\xi'_u = 1$. Then there are three ways in which we can maintain the family type:

$$\xi'_{u0} = \xi'_{u1} = 1; \quad \xi'_{u0} = e^{2\pi i/3}, \xi'_{u1} = e^{-2\pi i/3}; \quad \xi'_{u0} = e^{-2\pi i/3}, \xi'_{u1} = e^{2\pi i/3}.$$

In the first, we have changed ξ_{u1}, in the second ξ_{u0}, and in the third both ξ_{u0} and ξ_{u1}. This means that if two marked trees have the same family types then they can differ in a more complicated way than was the case with two types. We suspect that we ought in fact to have equality between the completions of the σ-algebras generated by the individual and family types when $pqr \leq (1/6)^3$ but there is more work to be done here.

Non-binary branching structure: working on the d-ary tree

If we assume that the types are independent, identically distributed Bern(p) random variables and that the family types are defined in the usual way (i.e. as the product of the individual types) then we saw what happened in the case $d = 1$ with the reconstruction problem on the integers. Reconstruction turned out to be possible for all but one value of the parameter p. We then saw what happened on the binary tree: reconstruction is possible provided p is either very large or very small. In fact it turns out that this problem isn't very interesting for $d > 2$ for the only thing that differs is the mean branching factor which plays only a minor role in the theory. Looking at the cloud speed calculation, for example, the only modification necessary is to write

$$\text{“}\sup\{a : I(a) \leq \log d\}\text{”}$$

rather than "... $\leq \log 2$". Indeed it is easily seen that reconstruction is possible provided $pq \leq 1/4d^2$.

Conditional independence of types

We now relax the assumption that the $(\xi_u; u \in T)$ be independent, imposing instead conditional independence. We assume still that the types are either -1 or 1 and that we are working on the binary tree. Suppose that the root type $\xi_\emptyset$ is chosen according to an arbitrary distribution π. Conditionally on $\xi_\emptyset = x \in \{-1, 1\}$, pick (ξ_0, ξ_1) according to a distribution $P_x(.,.)$, depending on x. To construct the next level, pick offspring(ξ_{00}, ξ_{01}) and (ξ_{10}, ξ_{11}) independently and according to P_{ξ_0} and P_{ξ_1} respectively. We continue to generate the entire marked tree $(\xi_u; u \in T)$ in this fashion. The family type remains the same as before, namely the product of the individual types. This type of multi-type branching process has been considered in more generality by Biggins and Kyprianou in [7]. There, a random tree is embedded in one in which every individual has a countably infinite number of offspring, with the type of an individual taken from a very general set, conditional on the parent type.

We begin by establishing the martingales in this setting, analogous to the ones considered in Chapter 1. That is, we determine conditions under which the sum

$$M_n = \sum_{|u|=n} a(\xi_u)\alpha^{S_n(\xi u)}\beta^n$$

is a martingale.

Lemma 13. *Let $\alpha, \beta \in \mathbb{R}$ and $a : \{-1, 1\} \to \mathbb{R}$ be a function. Then the sum*

$$M_n = \sum_{|u|=n} a(\xi_u)\alpha^{S_n(\xi,u)}\beta^n$$

is a martingale with respect to the σ-algebra generated by the types to level n of the tree provided that the eigenvalue problem

$$P_\alpha \underline{a} = \frac{1}{2\beta}\underline{a}$$

is satisfied, where $\underline{a} = (a(-1), a(+1))$ and

$$P_\alpha = \begin{pmatrix} P^\alpha(-1,-1) & P^\alpha(-1,+1) \\ P^\alpha(+1,-1) & P^\alpha(+1,+1) \end{pmatrix}.$$

Here, $P^\alpha(x, x') = p(x, x')\alpha^{x'}$, where $p(x, x')$ is the probability that $\xi_0 = x'$ given $\xi_\emptyset = x$ (which we assume is also the probability that $\xi_1 = x'$ given $\xi_\emptyset = x$ - i.e. left/right symmetry).

Proof. It is enough to show that $\mathbb{E}[M_1|\mathcal{F}_0] = M_0$. We have

$$M_0 = a(\xi_\emptyset)\alpha^{\xi_\emptyset}$$

and

$$M_1 = \{a(\xi_0)\alpha^{\xi_0} + a(\xi_1)\alpha^{\xi_1}\}\alpha^{\xi_\emptyset}\beta$$

so that

$$\mathbb{E}[M_1|\mathcal{F}_0] = \begin{cases} \alpha\beta[2a(1)P^\alpha(+1,+1) + 2a(-1)P^\alpha(+1,-1)], & \xi_\emptyset = 1 \\ \alpha^{-1}\beta[2a(+1)P^\alpha(-1,+1) + 2a(-1)P^\alpha(-1,-1)], & \xi_\emptyset = -1 \end{cases}$$

We now define $\underline{a}$ and P_α as in the statement of the lemma and equate $\mathbb{E}[M_1|\mathcal{F}_0]$ to M_0. □

In what follows we show that this martingale may be obtained as a change of measure between the "standard" offspring distribution described above and one under which a uniformly chosen line of descent has a modified distribution. Let $\mathbb{P}$ be the law of the standard offspring distribution. Let $\widetilde{\mathbb{P}}$ be the probability measure under which the nodes are distributed according to the standard offspring distribution except for those along a distinguished spine and their immediate siblings. Along this spine, given the vertex u_n in generation $n \geq 0$, the law of its offspring distribution with respect to the standard law has Radon-Nikodym derivative $M_1(\alpha, u_n)$, where $M_1(\alpha, u_n)$ is the version of $M_1(\alpha)$ obtained by treating u_n as the root. Furthermore, the vertex u_{n+1} is chosen from the offspring of u_n with probability $p_\xi(u_{n+1})$, where

$$p_\xi(u_{n+1}) = \frac{a(\xi_{u_{n+1}})\alpha^{\xi_{u_{n+1}}}}{a(\xi_{u_n0})\alpha^{\xi_{u_n0}} + a(\xi_{u_n1})\alpha^{\xi_{u_n1}}}.$$

We have

$$M_1(\alpha, u_n) = c_{u_n}\beta\alpha^{\xi_{u_n}}\{a(\xi_{u_n0})\alpha^{u_n0} + a(\xi_{u_n1})\alpha^{u_n1}\},$$

and hence if we denote by $P_x(y, z)$ the standard offspring distribution, that is the probability of a parent of type x having offspring of types y, z and by $Q_x(y, z)$ the corresponding "modified" distribution along the spine, then we have

$$\frac{Q_x(y, z)}{P_x(y, z)} = f(x, y, z),$$

where $f(u_n, u_n0, u_n1) = M_1(\alpha, u_n)$. Another way of looking at the construction is as follows. $\widetilde{\mathbb{P}}$ is the projection of $\widetilde{\mathbb{P}}^*$ onto the space of marked trees, where $\widetilde{\mathbb{P}}^*$ is the law of the random marked tree together with a distinguished spine. Having chosen the type at root according to some arbitrary distribution π, choose its offspring according to the law Q, where Q is as above. Now choose the first vertex (after the root) to be on our distinguished spine according to p_ξ and assign the types of its offspring according to Q while assigning the types of the offspring of the other vertex (not on the spine) according to P. Continuing in this way we obtain a marked tree, some types distributed according to P and others to Q. In doing so we also "generate" a distinguished spine.

Lemma 14. *Let* $\mu_n = \frac{d\widetilde{\mathbb{P}}_n}{d\mathbb{P}_n}$ *be the Radon-Nikodym derivative of the restrictions of the measures to level* n *of the tree. Then* $\mu_n = M_n(\alpha)$.

Proof. We have

$$\mathbb{P}_n(\xi) = \pi(\xi_\emptyset) \sum_{|v|<n} P_{\xi_v}(\xi_{v0}, \xi_{v1}).$$

Let $\widetilde{\mathbb{P}_n}^*$ be the probability measure on $\overline{\Omega} \times \partial T$ corresponding to the way in which we generated a marked tree together with a distinguished spine (described above). Then

$$\widetilde{\mathbb{P}_n}^*(\xi, u) = \pi(\xi_\emptyset) \prod_{v\leq u, |v|<n} Q_{\xi_v}(\xi_{v0}, \xi_{v1}) \prod_{v\nleq u, |v|<n} P_{\xi_v}(\xi_{v0}, \xi_{v1}) \prod_{v\leq u, |v|>0} p_\xi(v).$$

Writing Q in terms of M_1 and tidying up a little gives

$$\widetilde{\mathbb{P}_n}^*(\xi, u) = \pi(\xi_\emptyset)\beta^n \prod_{|v|<n} P_{\xi_v}(\xi_{v0}, \xi_{v1}) \prod_{v\leq u, |v|<n} c_v \alpha^{\xi_v} \prod_{v\leq u, |v|>0} a(\xi_v)\alpha^{\xi_v}$$

$$= \pi(\xi_\emptyset)\beta^n \prod_{|v|<n} P_{\xi_v}(\xi_{v0}, \xi_{v1}) \prod_{v\leq u, |v|<n} \frac{1}{a(\xi_v)} \prod_{v\leq u, |v|>0} a(\xi_v)\alpha^{\xi_v}$$

$$= \pi(\xi_\emptyset)\beta^n \prod_{|v|<n} P_{\xi_v}(\xi_{v0}, \xi_{v1}) \frac{a(\xi_u)}{a(\xi_\emptyset)} \alpha^{S_n(u)-\xi_\emptyset}$$

$$= \pi(\xi_\emptyset)\beta^n \prod_{|v|<n} P_{\xi_v}(\xi_{v0}, \xi_{v1}) a(\xi_u) \alpha^{S_n(u)} \frac{1}{\alpha^{\xi_\emptyset} a(\xi_\emptyset)}$$

$$= \pi(\xi_\emptyset)\beta^n \prod_{|v|<n} P_{\xi_v}(\xi_{v0}, \xi_{v1}) a(\xi_u) \alpha^{S_n(u)} c_\emptyset.$$

Summing over nodes in level n we obtain

$$\widetilde{\mathbb{P}_n}(\xi) = \sum_{|u|=n} \prod_{|v|<n} P_{\xi_v}(\xi_{v0}, \xi_{v1}) c_\emptyset a(\xi_u) \alpha^{S_n(u)} \beta^n,$$

and now when we divide this by $\mathbb{P}_n(\xi)$ we obtain

$$\mu_n = \sum_{|u|=n} c_\emptyset a(\xi_u) \alpha^{S_n(u)} \beta^n = M_n(\alpha),$$

as claimed. □

In terms of the issue of recovering types, we would like the process along the spine to be "invariant under flipping". More precisely, we would like the distribution of types along the spine to be the same when we interchange the types (recall how the argument worked for two types). One way this could be achieved is if the process along the spine were a Markov chain with symmetric transition matrix. The question of interest is then whether there exists a value of α for which this is the case. One approach to the non-recovery part of the argument would be to relate the martingales M_n to those studied by Kyprianou and Sani in [32]. The authors establish a matrix $M(\theta)$ which is analagous to our P_α but with $\alpha = \exp(-\theta)$. There they deal with a multi-type branching random walk having p types whereas we are interested in only two. We specialise their notation accordingly to illustrate the connection between their work and ours. Following their setup, let $e(\alpha)$ be the maximum eigenvalue of P_α (which exists by the Perron-Frobenius theorem provided P_α is positive regular) and $v(\alpha) = (v_{-1}(\alpha), v_{+1}(\alpha))$ be the corresponding right normalised eigenvector. Define, for $i \in \{-1, 1\}$,

$$W_i^n(\alpha) = \sum_{|u|=n} \frac{v_{\xi_u}(\alpha)\alpha^{S_n(\xi,u)}}{v_i(\alpha)e(\alpha)^n}.$$

This is a martingale with resect to the σ-algebra generated by the types to level n of the tree and we see immediately its connection to M_n: our constant β is $1/e(\alpha)$ and the function a is the appropriate entry of the right eigenvector corresponding to $e(\alpha)$. The authors go on to show that these martingales can in fact be obtained as a change of measure and give a description analogous to ours. Among the results of the paper is a criterion for uniform integrability which could potentially be used (in the sort of way we have illustrated with two and three types) to prove that under certain circumstances recovery of the individual types from the family types is not possible. As with some of our other work, this is an open problem.

We conclude this section with a brief description of a different type of reconstruction problem on trees that has been widely studied.

Classical reconstruction on trees

We use [39] and [26] as our main references for this. The problem take place on a deterministic d-ary tree (Galton-Watson tree where the number of offspring is always d). The type at root is chosen from some finite set according to an arbitrary probability distribution. With the value at root fixed, each vertex iteratively chooses its type from the one of its parent by an application of a Markov transition rule, with all such applications being independent. The basic question of interest is then typically whether the boundary of the tree contains any information about the type at root. The formulation is abstract but essentially boils down to whether the value at root is "encoded" into the boundary in a significant way. See [39] for full details. It turns out that in such problems it is the eigenvalues of the transition matrix for the associated Markov chain that are crucial for deciding whether or not the problem is "solvable". This type of problem is more akin to the conditional independence example just given, though the type of recovery studied is mathematically different. It might be interesting to investigate further the possible links between the two types of problem.

Chapter 3

Recursive Distributional Equations

In this chapter we introduce the idea of a "recursive distributional equation" (RDE). We begin with some background and then study a particular example coming from the work of the previous chapters. This leads us to the study of a particular RDE. We obtain a complete understanding of this equation in the case $pq \leq 1/16$ (where p, q are the probability parameters from Chapter 1) and obtain partial results for the more difficult case $pq > 1/16$, drawing particular attention to the special case $p = 1/2$.

The initial setup and discussion of tree-indexed solutions will set the scene for later chapters. We are guided throughout by the work of Aldous and Bandyopadhyay in [1].

3.1 Introduction: the basic idea

In a variety of problems, particularly in applied probability, distributional equations of the form

$$X = g(\xi_i, X_i; i = 1, ..., \mathbf{N}) \tag{3.1}$$

are often of interest. Here, the ξ_i and g are known, $\mathbf{N}$ may be random or deterministic and the X_i are independent copies of a random variable with some unknown distribution. In some work the ξ_i are referred to as "innovations" or "noise". The primary aim is to determine whether there exists, in the context of (3.1), a distribution μ such that if the X_i are independent with distribution μ, then X also has distribution μ. Writing $\mathcal{P}$ for the set of probability measures on a space A, suppose we are given a joint distribution for a family of random variables $(\xi_i; i = 1, ..., \mathbf{N})$ and an A-valued function g with appropriate domain (we will make this precise in the next section). Then we can define a mapping $T : \mathcal{P} \to \mathcal{P}$ by setting $T(\mu)$ to be the distribution of $g(\xi_i, X_i; i = 1, ..., \mathbf{N})$, where the X_i are independent with distribution μ and independent of the ξ_i and $\mathbf{N}$. We can now ask about the existence (and uniqueness) of fixed points, that is, distributions μ for which $T(\mu) = \mu$. We can also iterate: starting with independent copies $X_i^{(0)}(\mu)$ of a random variable with some arbitrary distribution μ, we apply the distributional equation (3.1) to obtain mutually independent random variables $X_i^{(1)}(\mu)$ with distribution $T(\mu)$. Applying the equation again we obtain mutually independent random variables $X_i^{(2)}(\mu)$ with distribution $T(T(\mu))$ and so on. Inductively we write $T^n(\mu)$ for the distribution of the mutually independent random variables $X_i^{(n)}(\mu)$, obtained by applying the distributional equation n times. By iterating in this way we think of (3.1) as being a recursive distributional equation (RDE) of the form

$$X_j^{(n+1)} = g(\xi_i, X_i^{(n)}; i = 1, ..., \mathbf{N}) \tag{3.2}$$

so that invariant distributions for the RDE correspond to fixed points of the induced map T or, equivalently, solutions to the distributional equation (3.1). This can be viewed in terms of an underlying "tree structure". We will say more about this later.

In this context we can also talk about the basin of attraction of a fixed point μ: for which distributions ν do we have $T^n(\nu) \to \mu$ as $n \to \infty$, in some suitable sense? Questions such as these have been studied fairly widely by authors such as Aldous [1], Biggins [5] and Liu [38]. This type of work is also related to the study of "random hierarchical systems", as studied by authors such as [27]. A comprehensive survey of what is known about RDEs can be found in [1], with particular attention being drawn to equations in which g is essentially a maximum function. A subclass of RDEs which is well understood is the linear type. We will define precisely what we mean by this and present a summary of the relevant theory in a later section.

3.2 The precise setup

We now specify a precise setup for the study of RDEs. This is taken from Aldous and Bandhapadhyay in [1]. In this chapter we will be working on the binary tree and so this level of detail is unnecessary but will be useful preparation for later chapters. Let $(A, \mathcal{A})$ be a measurable space, let $\mathcal{P}(A)$ be the set of probability measures on $(A, \mathcal{A})$ and let $(B, \mathcal{B})$ be another measurable space. Define

$$B^* = B \times \bigcup_{0 \le n < \infty} A^n,$$

where the union is disjoint and A^n denotes product space with the usual interpretation that A^0 is a singleton (point). Let $g : B^* \to A$ be a measurable map and let ν be a probability measure defined on $B \times (\mathbb{Z}_+ \cup \{+\infty\})$. We assume throughout (for simplicity) that the random variable $\mathbf{N}$ is finite. Define a map $T : \mathcal{P}(A) \to \mathcal{P}(A)$ by defining $T(\mu)$ to be the distribution of $g(\xi, X_i; 1 \le i \le \mathbf{N})$, where

1. the $(X_i; i \ge 1)$ are independent with common distribution μ;

2. $(\xi, \mathbf{N})$ has distribution ν;

3. the random variables in 1 and 2 are independent.

Notice that this is consistent with (3.1) by writing $\xi = (\xi_i)_{1 \leq i < \infty}$ and thinking of (ξ_i) as a single random element. The setup given in [1] does allow for the possibility that $\mathbf{N}$ may be infinite. A complication may arise here in that g may not be well-defined on the whole of $B \times A^{\infty}$. See [1] for full details.

3.2.1 Recursive tree process

Suppose that we associate the random A-valued variables X_i in (3.1) with the vertices of a Galton-Watson tree Γ having offspring distribution $\mathbf{N}$. In the distributional equation

$$X = g(\xi_i, X_i; i = 1, ..., \mathbf{N}),$$

we can think of X as being the random variable associated with a parent vertex, determined by the independent X_i associated with its offspring and some random noise ξ associated with the parent. This extends in an obvious way to earlier generations of the tree via iterating (3.2) and we call the resulting structure a recursive tree process. This leads to the idea of a "tree-indexed" solution to the RDE (3.2). We say that $(X_u; u \in \Gamma)$ is a tree-indexed solution to the RDE (3.2) if

1. for every n, the random variables $(X_u; |u| = n)$ are independent and identically distributed, having as distribution a fixed point of the induced map T;

2. for every $u \in \Gamma$, $X_u = g(\xi_u, X_{ui}; i = 1, ..., N_u)$, where (ξ_u, N_u) has the distribution ν from the previous section, independently as u varies;

3. for every n, the random variables $(X_u; |u| = n)$ and $(\xi_u, N_u; |u| \leq n-1)$ are independent.

Notice that we can also talk about tree-indexed solutions down to level n of the tree. Let $\Gamma_n = \{u \in \Gamma : |u| \leq n\}$. Then we say that $(X_u; u \in \Gamma_n)$ is a tree-indexed solution to level n if

1. for $1 \leq m \leq n$, the random variables $(X_u; |u| = m)$ are independent and identically distributed, having as distribution a fixed point of the induced map T;

2. for every $u \in \Gamma_n, X_u = g(\xi_u, X_{ui}; i = 1, ..., N_u)$, where (ξ_u, N_u) has the distribution ν from the previous section, independently as u varies;

3. for every $1 \leq m \leq n$, the random variables $(X_u; |u| = m)$ and $(\xi_u, N_u; |u| \leq m - 1)$ are independent.

Tree-indexed solutions will not be investigated in this chapter but will play an important role in Chapters 4 and 5. Nevertheless, the natural tree structure is particularly convenient as an indexing set for iterating and will prove to be a useful tool in some of the proofs of this chapter.

3.2.2 Solutions, invariant measures and basins of attraction

By invariant measure for (or solution to) the RDE (3.2) we mean a solution to the corresponding distributional equation (3.1) or a fixed point of the induced map T. We now make precise the idea of a basin of attraction of a solution to the RDE (3.2). For a solution μ to the RDE (3.2), we say that a distribution ν is in the basin of attraction of μ if $T^n(\nu)$ converges weakly to μ (strictly speaking we mean in the weak* sense), by which we mean that

$$\int f dT^n(\nu) \to \int f d\mu$$

for every bounded continuous function on A (which we assume is endowed with the structure of a metric space). This is, of course, equivalent to saying that the iterates $X_i^{(n)}(\nu)$

converge in distribution to a random variable with distribution μ.

3.3 Some examples of RDEs

3.3.1 Stable laws

A well known problem is that of finding independent, identically distributed random variables X_1, X_2 on $\mathbb{R}$ such that the random variable

$$X = \frac{1}{\sqrt{2}}(X_1 + X_2)$$

has the same distribution as X_1, X_2 or belongs to the same class of distribution. The non-trivial solutions in this case are centred normal distributions. The trivial solution is just the unit mass at zero. We could of course regard this as an RDE. We have, by iteration,

$$X_i^{(n)} = \frac{X_1^{(0)} + \dots + X_{2^n}^{(0)}}{2^{\frac{n}{2}}}$$

so that if the $X_i^{(0)}$ have finite mean m and finite positive variance σ^2, then $X_i^{(n)}$ is asymptotically Normal with mean $\sqrt{n}m$ and variance σ^2. It follows that any distribution with mean 0 and variance σ^2 is in the basin of attraction of the normal solution with mean 0 and variance σ^2.

3.3.2 Population size of a Galton-Watson process

Let Z be the total population of a Galton-Watson branching process with offspring distribution N. Then it is easily seen that Z satisfies the RDE

$$Z = 1 + \sum_{i=1}^{N} Z_i. \tag{3.3}$$

Writing H for the probability generating function of N, the existence of an invariant distribution for this RDE is equivalent to finding a generating function G such that

$$G(s) = sH(G(s)), \quad s \in [-1, 1].$$

In simple cases this provides a practical means of obtaining the solution (indeed we will make use of this type of argument in Chapter 5). For example, if N takes value 2 with probability p and value 0 with probability $q = 1 - p$, then G satisfies a quadratic:

$$psG(s)^2 - G(s) + q = 0,$$

from which we are able (with a little analysis) to obtain G. When $\mathbb{E}[N] \leq 1$ and $\mathbb{P}(N = 1) < 1$, it can be shown that the total population size of a Galton-Watson branching process with offspring distribution N is the unique solution to the RDE (3.3). See [1]

3.3.3 Smoothing transformations

Recall that the martingale Λ_n from the previous chapters satisfies the equation

$$\Lambda_n = \frac{\Lambda_0}{2}(\Lambda'_{n-1} + \Lambda''_{n-1}),$$

where Λ'_n, Λ''_n are independent copies of Λ_n and

$$\Lambda_0 = \frac{1}{\sqrt{4pq}}(\frac{q}{p})^{\xi_\emptyset/2}.$$

Thus one solution to the RDE

$$X = \frac{1}{\sqrt{16pq}}(\frac{q}{p})^{\xi/2}(X_1 + X_2), \tag{3.4}$$

where ξ takes value 1 with probability p and value -1 with probability $q = 1 - p$, is the law of the martingale limit Λ_∞. (3.4) is an example of a linear RDE. Linear RDEs are of the form

$$X = \sum_i \xi_i X_i. \tag{3.5}$$

The induced map T corresponding to the process of obtaining a random variable X via (3.5) is sometimes referred to as a "smoothing transformation" since X is a weighted average of the X_i. This type of RDE has been studied extensively in [1], [5], [38] and there are fairly comprehensive results concerning the existence and uniqueness of invariant distributions and basins of attraction for these distributions. We will study the RDE (3.4) in more detail and summarise some of the important aspects of this type of RDE in a later section on linear RDEs.

3.4 Tree-indexed recursions

Recall from Chapter 2 the independent Bern(p) individual types $(\xi_u; u \in T)$ on $\{-1, 1\}^T$ and the corresponding family types $(\eta_u; u \in T)$, where, for $u \in T$,

$$\eta_u = \xi_u \xi_{u0} \xi_{u1}.$$

In this chapter we will be interested in the conditional probability

$$p_\emptyset := \mathbb{P}(\xi_\emptyset = 1 | \eta_u; u \in T).$$

Of course when $pq \leq 1/16$, the (completions of the) σ-algebras generated by the random variables $(\xi_u; u \in T)$ and $(\eta_u; u \in T)$ are the same and hence $p_\emptyset$ will be, almost surely,

$\mathbf{1}(\xi_\emptyset = 1)$.

Theorem 8. *Let $pq > 1/16$. Define*

$$p_u = \mathbb{P}(\xi_u = 1 | \eta_v; v \in T_u),$$

where T_u is the sub-tree rooted at u. Let

$$s_u = \frac{p_u}{1 - p_u}$$

be the associated "odds ratio", which is well defined since $p_u \in (0, 1)$ almost surely. Then the s_u satisfy the recursion

$$s_u = \frac{p}{q}\left(\frac{1 + s_{u0}s_{u1}}{s_{u0} + s_{u1}}\right)^{\eta_u}.$$

Proof. We consider

$$\mathbb{P}(\xi_\emptyset = 1 | \eta_\emptyset = 1, \eta_u = a_u; 1 \leq |u| \leq n),$$

where $a_u \in \{-1, 1\}$. That is, we begin by considering the probability of the root having type one, conditional on the family types to level n of the tree and on the family type at root being 1 (later we will do the same calculation conditional on the family type at root being -1). For ease of notation, we will write this probability as N/D, where

$$N = \mathbb{P}(\xi_\emptyset = 1, \eta_\emptyset = 1, \eta_u = a_u; 1 \leq |u| \leq n)$$

(the numerator) and

$$D = \mathbb{P}(\eta_\emptyset = 1, \eta_u = a_u; 1 \leq |u| \leq n)$$

(the denominator). We have $\xi_\emptyset = \eta_\emptyset = 1 \Rightarrow \xi_0 = \xi_1 = 1$ or $\xi_0 = \xi_1 = -1$ so that

$$N = \mathbb{P}\big((\xi_\emptyset = 1) \cap (\xi_0 = 1) \cap (\xi_1 = 1) \cap \bigcap_{1 \le u \le n} (\eta_u = a_u)\big)$$

$$+\mathbb{P}\big((\xi_\emptyset = 1) \cap (\xi_0 = -1) \cap (\xi_1 = -1) \cap \bigcap_{1 \le u \le n} (\eta_u = a_u)\big).$$

We now re-write this in order to emphasise independence between various terms. Recall the notation

$$T_L = \{u \in T : u \preceq 0\}, \qquad T_R = \{u \in T : u \preceq 1\}$$

for vertices to the left and right of the root. We will specialise this notation for our purposes by writing

$$T_L^n = \{u \in T : u \preceq 0, 1 \le |u| \le n\}, \qquad T_R^n = \{u \in T : u \preceq 1, 1 \le |u| \le n\}.$$

Then

$$N = \mathbb{P}\big((\xi_\emptyset = 1) \cap [(\xi_0 = 1) \cap \bigcap_{u \in T_L^n} (\eta_u = a_u)] \cap [(\xi_1 = 1) \cap \bigcap_{u \in T_R^n} (\eta_u = a_u)]\big)$$

$$+\mathbb{P}\big((\xi_\emptyset = 1) \cap [(\xi_0 = -1) \cap \bigcap_{u \in T_L^n} (\eta_u = a_u)] \cap [(\xi_1 = -1) \cap \bigcap_{u \in T_R^n} (\eta_u = a_u)]\big).$$

By independence, we have then that

$$\frac{N}{p} = \mathbb{P}\big((\xi_0 = 1) \cap \bigcap_{u \in T_L^n} (\eta_u = a_u)\big)\mathbb{P}\big((\xi_1 = 1) \cap \bigcap_{u \in T_R^n} (\eta_u = a_u)\big)$$

$$+\mathbb{P}\big((\xi_0 = -1) \cap \bigcap_{u \in T_L^n} (\eta_u = a_u)\big)\mathbb{P}\big((\xi_1 = -1) \cap \bigcap_{u \in T_R^n} (\eta_u = a_u)\big).$$

Now, the denominator D is given by $D = N + N'$, where

$$N' = \mathbb{P}\big((\xi_0 = -1) \cap (\eta_\emptyset = 1) \cap \bigcap_{1 \le u \le n} (\eta_u = a_u)\big).$$

By an argument analogous to that given above, we have

$$\frac{N'}{q} = \mathbb{P}\big((\xi_0 = -1) \cap \bigcap_{u \in T_L^n} (\eta_u = a_u)\big)\mathbb{P}\big((\xi_1 = 1) \cap \bigcap_{u \in T_R^n} (\eta_u = a_u)\big)$$

$$+\mathbb{P}\big((\xi_0 = 1) \cap \bigcap_{u \in T_L^n} (\eta_u = a_u)\big)\mathbb{P}\big((\xi_1 = -1) \cap \bigcap_{u \in T_R^n} (\eta_u = a_u)\big).$$

We now divide both numerator N and denominator D of our conditional probability by $\mathbb{P}(\eta_u = a_u; u \in T_L^n)\mathbb{P}(\eta_u = a_u; u \in T_R^n)$ to obtain the following expressions respectively:

$$\mathbb{P}(\xi_0 = 1 | \eta_u = a_u; u \in T_L^n)\mathbb{P}(\xi_1 = 1 | \eta_u = a_u; u \in T_R^n) \tag{3.6}$$

$$+ \quad \mathbb{P}(\xi_0 = -1 | \eta_u = a_u; u \in T_L^n)\mathbb{P}(\xi_1 = -1 | \eta_u = a_u; u \in T_R^n),$$

and

$$(3.6) \quad + \quad \mathbb{P}(\xi_0 = 1 | \eta_u = a_u; u \in T_L^n)\mathbb{P}(\xi_1 = -1 | \eta_u = a_u; u \in T_R^n)$$

$$+ \quad \mathbb{P}(\xi_0 = -1 | \eta_u = a_u; u \in T_L^n)\mathbb{P}(\xi_1 = 1 | \eta_u = a_u; u \in T_R^n).$$

Now, for $u \in T, |u| \le n$, let

$$p_u^n = \mathbb{P}(\xi_u = 1 | \eta_v; |v| \le n).$$

We have then

$$\frac{N}{\mathbb{P}(\eta_u = a_u; u \in T_L^n)\mathbb{P}(\eta_u = a_u; u \in T_R^n)} = p(p_0^n p_1^n + (1 - p_0^n)(1 - p_1^n))$$

and

$$\frac{D}{\mathbb{P}(\eta_u = a_u; uT_L^n)\mathbb{P}(\eta_u = a_u; u \in T_R^n)} = p(p_0^n p_1^n + (1-p_0^n)(1-p_1^n)) + q(p_1^n(1-p_0^n) + p_0^n(1-p_1^n))$$

for those points ω such that

$$\eta_\emptyset(\omega) = 1, \eta_u(\omega) = a_u, \quad 1 \leq |u| \leq n.$$

Since the a_u are arbitrary these hold as equalities between random variables on the event $\{\eta_\emptyset = 1\}$. Hence on the event $\{\eta_\emptyset = 1\}$,

$$(3.7) \qquad p_\emptyset^n = \frac{N}{D} = \frac{p(p_0^n p_1^n + (1-p_0^n)(1-p_1^n))}{p(p_0^n p_1^n + (1-p_0^n)(1-p_1^n)) + q(p_1^n(1-p_0^n) + p_0^n(1-p_1^n))}.$$

By mimicking the arguments above we find that, on the event $\{\eta_\emptyset = -1\}$,

$$(3.8) \qquad p_\emptyset^n = \frac{p(p_1^n(1-p_0^n) + p_0^n(1-p_1^n))}{p(p_1^n(1-p_0^n) + p_0^n(1-p_1^n)) + q(p_0^n p_1^n + (1-p_0^n)(1-p_1^n))}.$$

Now, the sequence $(p_\emptyset^n)_n$ is a uniformly integrable martingale with respect to the σ-algebra generated by the family types to level n of the tree and hence we have, almost surely,

$$p_\emptyset^n \to p_\emptyset, \qquad n \to \infty,$$

justifying the above expressions in the limit as $n \to \infty$. That is,

$$(3.9) \qquad p_\emptyset = P_1 \mathbf{1}(\eta_\emptyset = 1) + P_2 \mathbf{1}(\eta_\emptyset = -1) \quad a.s.,$$

where

$$P_1 = \frac{p(p_0p_1 + (1-p_0)(1-p_1))}{p(p_0p_1 + (1-p_0)(1-p_1)) + q(p_1(1-p_0) + p_0(1-p_1))} \tag{3.10}$$

and

$$P_2 = \frac{p(p_1(1-p_0) + p_0(1-p_1))}{p(p_1(1-p_0) + p_0(1-p_1)) + q(p_0p_1 + (1-p_0)(1-p_1))}. \tag{3.11}$$

Now, we want to write $p_\emptyset$ in terms of

$$s_\emptyset = \frac{p_\emptyset}{1-p_\emptyset}$$

and so we need to make sure that $p_\emptyset \in (0,1)$ almost surely, else $s_\emptyset$ will not be well defined. Let

$$\alpha = \mathbb{P}(p_\emptyset = 0), \qquad \beta = \mathbb{P}(p_\emptyset = 1).$$

By considering (3.10) and (3.9) we see that, on the event $\{\eta_\emptyset = 1\}$, $p_\emptyset = 0$ if and only if $p_0 = 0, p_1 = 1$ or $p_0 = 1, p_1 = 0$. Since $p_\emptyset, p_0, p_1$ all have the same distribution and p_0, p_1 are independent, this gives

$$\mathbb{P}(p_\emptyset = 1 | \eta_\emptyset = 1) = 2\alpha\beta.$$

Similarly, by considering (3.11) and (3.9), we see that, on the event $\{\eta_\emptyset = -1\}$, $p_\emptyset = 0$ if and only if $p_0 = p_1 = 0$ or $p_0 = p_1 = 1$. Hence

$$\mathbb{P}(p_\emptyset = 0 | \eta_\emptyset = -1) = \alpha^2 + \beta^2.$$

Multiplying these conditional probabilities by the probabilities of the events $\{\eta_\emptyset = 1\}$ and $\{\eta_\emptyset = -1\}$ and summing gives us $\mathbb{P}(p_\emptyset = 0)$:

$$\alpha = (p^3 + 3pq^2)2\alpha\beta + (q^3 + 3p^2q)(\alpha^2 + \beta^2). \tag{3.12}$$

It is easily seen that, on the event $\{\eta_\emptyset = 1\}$, $p_\emptyset = 1$ if and only if $p_0 = p_1 = 0$ or $p_0 = p_1 = 1$ and that, on the event $\{\eta_\emptyset = -1\}$, $p_\emptyset = 1$ if and only if $p_0 = 0, p_1 = 1$ or $p_0 = 1, p_1 = 0$. Arguing along the same lines as before, this gives

$$\beta = (p^3 + 3pq^2)(\alpha^2 + \beta^2) + (q^3 + 3p^2q)2\alpha\beta. \tag{3.13}$$

Adding (3.12) and (3.13) gives

$$\alpha + \beta = (\alpha + \beta)^2(p + q)^3 = (\alpha + \beta)^2$$

so that $\alpha + \beta = 0$ or 1. If $\alpha + \beta = 0$, it follows that $\alpha = \beta = 0$. If $\alpha + \beta = 1$, $p_\emptyset$ is concentrated on $\{0, 1\}$. In this case, since $p_\emptyset$ has mean p, it follows that $p_\emptyset = \mathbf{1}(E)$ almost surely, where $\mathbb{P}(E) = p$. Hence $\mathbf{1}(\xi_\emptyset = 1)$ is measurable with respect to the completion of $\sigma(\eta_u; u \in T)$. From the work of Chapter 2 we know that when $pq \leq 1/16$, this is indeed the case but that when $pq > 1/16$, this fails to be true. Hence we must have $p_\emptyset \in (0, 1)$ when $pq > 1/16$. Writing $p_\emptyset$ in terms of $s_\emptyset$ we have then

$$s_\emptyset = \frac{p}{q}\left(\frac{1 + s_0 s_1}{s_0 + s_1}\mathbf{1}_{(\eta_\emptyset = 1)} + \frac{s_0 + s_1}{1 + s_0 s_1}\mathbf{1}_{(\eta_\emptyset = -1)}\right),$$

or, equivalently,

$$s_\emptyset = \frac{p}{q}\left(\frac{1 + s_0 s_1}{s_0 + s_1}\right)^{\eta_\emptyset}.$$

Notice that there is nothing special about the root in these arguments. The same applies to any arbitrary vertex $u \in T$:

$$s_u = \frac{p}{q}(\frac{1 + s_{u0}s_{u1}}{s_{u0} + s_{u1}})^{\eta_u}. \tag{3.14}$$

□

Remark 1. *The expression for s_u can be extended to the case $pq \leq 1/16$ by interpreting it suitably at infinity.*

The problem with equation (3.14) is the following. We would like to regard the equation as an RDE for the unknown distribution but the fact that η_u depends on s_{u0} and s_{u1} makes life difficult. We would like, therefore, to re-write this equation in terms of independent random variables. This will be our next task.

Theorem 9. *Define tree-indexed random variables $(t_u; u \in T)$ by $t_u = s_u^{-\xi_u}$. Then the t_u satisfy the recursion*

$$t_u = (\frac{q}{p})^{\xi_u} \frac{t_{u0} + t_{u1}}{1 + t_{u0}t_{u1}}, \tag{3.15}$$

where the random variables on the right hand side are independent.

Proof. This is simply a matter of making the appropriate substitutions in the eight possible cases for the values of $(\xi_u, \xi_{u0}, \xi_{u1})$:

$$(1,1,1), (1,1,-1), (1,-1,-1), (-1,-1,-1), (-1,-1,1), (-1,1,1), (1,-1,1), (-1,1,-1).$$

We make the calculation in two of the cases to illustrate the idea. For $(1,1,1)$, we have $\eta_u = 1$ so that

$$\frac{1}{t_u} = \frac{p}{q}\frac{1 + \frac{1}{t_{u0}}\frac{1}{t_{u1}}}{\frac{1}{t_{u0}} + \frac{1}{t_{u1}}}$$

which gives

$$t_u = \frac{q}{p}\frac{t_{u0} + t_{u1}}{1 + t_{u0}t_{u1}}.$$

For $(-1,-1,-1)$ we have $\eta_u = -1$ so that

$$t_u = \frac{p}{q}\left(\frac{1 + t_{u0}t_{u1}}{t_{u0} + t_{u1}}\right)^{-1}$$

$$= \frac{p}{q}\frac{t_{u0} + t_{u1}}{1 + t_{u0}t_{u1}}.$$

Indeed it is easily verified that we end up with a multiplicative factor of $\frac{q}{p}$ whenever $\eta_u = 1$ and a factor of $\frac{p}{q}$ whenever $\eta_u = -1$. □

At this stage it is natural to make a summary - based on the results of Chapter 2 - of the results we know to be true for the RDE corresponding to (3.15). That is, for the RDE on $[0, \infty)$ given by

$$X = \left(\frac{q}{p}\right)^{\xi}\frac{X_1 + X_2}{1 + X_1X_2}, \tag{3.16}$$

where ξ takes value 1 with probability p and value -1 with probability $q = 1 - p$.

Theorem 10. *Let δ_x denote the point mass at x given by*

$$\delta_x(A) = \begin{cases} 1 & x \in A, \\ 0 & \textit{otherwise}. \end{cases}$$

The degenerate solution δ_0 is a solution to the RDE (3.16) for all values of p. In the case $pq > 1/16$, there exists a non-trivial solution whose distribution is that of $t_\emptyset$. Let ν be the probability measure $p\delta_{\frac{q}{p}} + q\delta_{\frac{p}{q}}$. Then in the case $pq \leq 1/16$, ν is in the basin of attraction of the degenerate solution while in the case $pq > 1/16$, ν is in the basin of attraction of the non-trivial solution.

Proof. We know that the distribution of $t_\emptyset$ is a solution to the RDE (3.16). In the case $pq \leq 1/16$ we have $p_\emptyset = \mathbf{1}(\xi_\emptyset = 1)$ almost surely. Hence $s_\emptyset = \frac{p_\emptyset}{1-p_\emptyset}$ takes value 0 with probability q and value ∞ with probability p and therefore $t_\emptyset = 0$ almost surely. On the other hand we know from Chapter 2 that when $pq > 1/16$, $p_\emptyset$ has a non-trivial distribution. It follows that the same is true for $s_\emptyset$ and therefore $t_\emptyset$.

Let $t_u^n = (s_u^n)^{-\xi_u}$. Then, since $p_\emptyset^n \to p_\emptyset$ almost surely, we have $t_\emptyset^n \to t_\emptyset$ almost surely, where the distribution of $t_\emptyset^n$ is just $T^n(\nu)$ for a suitable choice of ν. Since $t_\emptyset^n$ corresponds to knowing the family types down to level n of the tree, we may obtain it by inserting conditional probabilities in level n of the tree and then iterating to obtain $t_\emptyset$. More precisely, in level n of the tree we insert at each vertex the conditional probability of a type one given just the family type at that vertex. To make life easier, we can start one level below so that there are no family types on which to condition. Let w be a vertex in this level. Then the probability that $\xi_w = 1$ without conditioning on any family types is $\mathbb{P}(\xi_w = 1) = p$. The associated odds ratio is then $s_w = \frac{p}{q}$ so that $t_w = \frac{p}{q}$ if $\xi_w = -1$ and $t_w = \frac{q}{p}$ if $t_w = 1$. Hence the distribution of t_w is $p\delta_{\frac{q}{p}} + q\delta_{\frac{p}{q}}$. Now, since $t_\emptyset^n \to t_\emptyset$ almost surely, we have $t_\emptyset^n \to t_\emptyset$ in distribution. It follows then that ν is in the basin of attraction of the solution corresponding to the law of $t_\emptyset$, which is trivial/non-trivial depending on whether $pq \leq 1/16$ or $pq > 1/16$. □

3.5 A study of the special case $p = 1/2$

The case $p = 1/2$, in which the "$(\frac{q}{p})^{\xi_u}$" term disappears, is relatively straightforward and we are able to give a complete classification of the behaviour of the RDE (3.16). In this case the RDE reads

$$X = \frac{X_1 + X_2}{1 + X_1 X_2}. \tag{3.17}$$

The following theorem collects together all of the results we will obtain for the RDE.

Theorem 11. *The RDE (3.17) has only two solutions, one being $\delta_{\{0\}}$, the unit mass at 0, and the other being $\delta_{\{1\}}$, the unit mass at 1. Of the two it is $\delta_{\{1\}}$ which is attracting, with basin of attraction all distributions on $[0, \infty)$ except $\delta_{\{0\}}$.*

To prove the theorem we need several intermediate results.

Lemma 15. *Let $c > 1$ be a constant. Let X be a random variable whose law is a solution of the RDE (3.17). Then $\mathbb{P}(X \in [1/c, c]) \in \{0, 1\}$.*

Proof. Suppose that we start in some arbitrary level of the tree with independent copies of some random variable Y. Recall the notation $Y_i^{(n)}(\nu)$ for a random variable with distribution $T^n(\nu)$, obtained by applying the RDE n times to obtain the corresponding random variables n levels above where we started. Throughout this proof we will use the abbreviated notation Y_n, Y_n' to denote independent random variables with distribution $T^n(\nu)$, where we have suppressed the dependence on the arbitrary distribution ν, so that, in particular, $Y_0 = Y$. Consider now the level sets of the function

$$\frac{x + y}{1 + xy}$$

or, equivalently, the graph

$$\frac{x + y}{1 + xy} = c,$$

whose equation is

$$y = \frac{c - x}{1 - cx}.$$

Now,

$$\frac{x+y}{1+xy} > c$$

provided

$$y > \frac{c-x}{1-cx}$$

for $x < 1/c$ and

$$y < \frac{c-x}{1-cx}$$

for $x > 1/c$. Formally, the regions satisfying these inequalities are given by

$$A = \{(x,y) \in [0,\infty)^2 : x < 1/c, y > \frac{c-x}{1-cx}\}, \quad B = \{(x,y) \in [0,\infty)^2 : x > 1/c, y < \frac{c-x}{1-cx}\}.$$

Let C and D be the regions

$$C = \{(x,y) \in [0,\infty)^2 : x < 1/c, y > c\}, \quad D = \{(x,y) \in [0,\infty)^2 : x > c, y < 1/c\}.$$

Then we have

$$A \cup B \subseteq C \cup D$$

and hence

$$\mathbb{P}(Y_{n+1} > c) = \mathbb{P}((Y_n, Y_n') \in A \cup B) \leq \mathbb{P}((Y_n, Y_n') \in C \cup D) = 2\mathbb{P}(Y_n > c)\mathbb{P}(Y_n' < 1/c).$$

We now obtain a similar estimate for $\mathbb{P}(Y_{n+1} < 1/c)$. Consider the graph

$$\frac{x+y}{1+xy} = \frac{1}{c},$$

whose equation is

$$y = \frac{\frac{1}{c} - x}{1 - \frac{x}{c}}.$$

In this case,

$$\frac{x+y}{1+xy} < \frac{1}{c}$$

provided

$$y < \frac{\frac{1}{c} - x}{1 + \frac{x}{c}}$$

for $x < c$ and

$$y > \frac{\frac{1}{c} - x}{1 - \frac{x}{c}}$$

for $x > c$. Now,

$$\{(x,y) \in [0,\infty)^2 : x < c, y < \frac{\frac{1}{c} - x}{1 + \frac{x}{c}}\} \subseteq \{(x,y) \in [0,\infty)^2 : x, y < 1/c\}$$

and

$$\{(x,y) \in [0,\infty)^2 : x > c, y > \frac{\frac{1}{c} - x}{1 + \frac{x}{c}}\} \subseteq \{(x,y) \in [0,\infty)^2 : x, y > c\}.$$

Hence

$$\mathbb{P}(Y_{n+1} < 1/c) \leq \mathbb{P}(Y_n > c)^2 + \mathbb{P}(Y_n' < 1/c)^2.$$

Since Y_n, Y_n' have the same distribution, we can write the above as a system of inequalities for

$$a_n = \mathbb{P}(Y_n > c), \qquad b_n = \mathbb{P}(Y_n < 1/c).$$

We have

$$a_{n+1} \leq 2a_n b_n, \qquad b_{n+1} \leq a_n^2 + b_n^2,$$

so that, by adding the sequences,

$$a_{n+1} + b_{n+1} \leq (a_n + b_n)^2.$$

Hence $a_n + b_n \to 0$ provided $a_0 + b_0 < 1$, that is, provided $\mathbb{P}(Y > c) + \mathbb{P}(Y < 1/c) < 1$. Under these circumstances, it follows that

$$\mathbb{P}(Y_n > c) \to 0, \quad \mathbb{P}(Y_n < 1/c) \to 0.$$

Hence if the distribution of Y is a solution to the RDE and puts mass on $[1/c, c]$ it must in fact be concentrated on $[1/c, c]$. The result now follows. □

Proof. (of first part of Theorem 11) Let X be a random variable whose distribution is a solution to the RDE (3.17). Define a map $\varphi : (1, \infty) \to \mathbb{R}$ by

$$\varphi(c) = \mathbb{P}(X \in [1/c, c]).$$

Then, by the previous lemma, φ takes only values 0 and 1 and, furthermore, since

$$1 < c \leq d \Rightarrow [1/c, c] \subseteq [1/d, d],$$

we have

$$1 < c \leq d \Rightarrow \varphi(c) \leq \varphi(d),$$

so that φ is (weakly) increasing. It follows then that φ is either identically zero, identically one, or else it jumps from zero to one at some value $c = c_0 > 1$. Now, if $\varphi \equiv 0$ then the distribution of X is concentrated on $[0, 1/c) \cup (1/c, \infty)$, which we can shrink to the point $\{0\}$ by making c arbitrarily large. We conclude that, in this case, the distribution of X is $\delta_{\{0\}}$. If, on the other hand, $\varphi \equiv 1$ then we conclude that the distribution of X is $\delta_{\{1\}}$ since we can shrink the interval $[1/c, c]$ to the point $\{1\}$ by taking c to be arbitrarily close to 1. The final possibility is that φ jumps from 0 to 1 at $c = c_0$. It follows that the distribution of X must be $a\delta_{\{c\}} + b\delta_{\{1/c\}}$, for non-negative constants a, b such that $a + b = 1$. We need to check that this isn't a solution to the RDE.

Applying the RDE (3.17) to the random variable with distribution $a\delta_{\{c_0\}} + b\delta_{\{1/c_0\}}$, we obtain a random variable with distribution

$$(a^2 + b^2)\delta_{\{2c_0/(1+c_0^2)\}} + 2ab\delta_{\{c_0/2+1/2c_0\}}.$$

For the distribution of X to be a solution to the RDE we would have to have

$$\frac{2c_0}{1 + c_0^2} = \begin{cases} c_0 \\ \frac{1}{c_0} \end{cases} \tag{3.18}$$

both of which force $c_0 = 1$ so that X has distribution $\delta_{\{1\}}$ as before. This completes the proof. □

Proof. (second part of Theorem 11) We assume that $\mu \neq \delta_{\{1\}}$ since this case is trivial. Under the assumption of the theorem we also have that $\mu \neq \delta_{\{0\}}$ and so we may deduce that there exists a constant $c_0 > 1$ such that μ puts mass on the interval $[1/c_0, c_0]$. Now, by the previous lemma, as we iterate μ, the iterates $T^n(\mu)$ put more and more mass on this interval and hence $T(\mu)$ will place even more mass on $[1/c_0, c_0]$. It follows then that there exists a $c_1 < c_0$ such that $T(\mu)$ puts mass on $[1/c_1, c_1]$ (notice that we cannot have $c_1 = c_0$ because this would mean that μ is concentrated on $c_0, 1/c_0$ which, as we saw, forces $c_0 = 1$). Applying the same argument, we deduce that there exists a $c_2 < c_1$ such that $T^2(\mu)$ puts mass on $[1/c_2, c_2]$ and so on. In this way we obtain a decreasing sequence of real numbers $(c_n)_{n\geq 0}$, bounded below by one, and with the property that $T^n(\mu)$ puts mass on the interval $[1/c_n, c_n]$ for every n. Now, since c_n is decreasing and bounded below, it follows that the sequence has a limit, c. Suppose that $c > 1$. It would follow then that, in the limit as $n \to \infty$, $T^n(\mu)$ is concentrated on $1/c$ and c. We saw earlier that this is not a solution to the RDE (3.17) when $c \neq 1$. We must therefore have $c = 1$ so that, in the limit as $n \to \infty$, $T^n(\mu)$ puts all of its mass on the singleton $\{1\}$. □

Remark 2. *Notice that Theorem (11) generalises a result concerning the independence of individual and family types in the case* $p = 1/2$. *We know (since* $pq = 1/4 > 1/16$*) that* $p_\emptyset$ *is not almost surely zero. It follows then from Theorem (11) that, in the case* $p = 1/2$, $t_u = 1$ *almost surely so that* $s_\emptyset = 1$ *almost surely and therefore* $p_\emptyset = 1/2$ *almost surely. Thus knowing the family types tell us nothing about the type at root in the special case* $p = 1/2$. *This is a generalisation of conditioning on only one family type. We have*

$$\mathbb{P}(\xi_\emptyset = 1|\eta_\emptyset) = \begin{cases} \frac{p^3+pq^2}{p^3+3pq^2}, & \eta_\emptyset = 1 \\ \frac{2p^2q}{3p^q+q^3}, & \eta_\emptyset = -1 \end{cases} \tag{3.19}$$

so that, when $p = 1/2$,

$$\mathbb{P}(\xi_\emptyset = 1|\eta_\emptyset) = 1/2 \quad a.s.$$

3.6 Some analysis of the RDE (3.16): the general case $p \neq 1/2$

Recall the RDE (3.16):

$$X = (\frac{q}{p})^\xi \frac{X_1 + X_2}{1 + X_1 X_2},$$

where ξ takes value 1 with probability p and value -1 with probability $q = 1 - p$. We consider throughout solutions which are finite.

Lemma 16. *Let X be a non-negative random variable whose distribution is a non-trivial solution to the RDE (3.16). Write $\phi(x) = \mathbb{P}(X > x) + \mathbb{P}(X < 1/x)$ for the sum of the tails of the distribution of X and let $K = \max(\frac{q}{p}, \frac{p}{q}) > 1$. Then, for some $\alpha, \beta > 0$, ϕ is bounded by an exponential function:*

$$\phi(x) \leq e^{-\alpha x^{\beta}}$$

for all $x > K$.

Proof. We extend the idea of the proof of Lemma 15. Let Y, Y_n, Y_n' be defined as there but in relation to the RDE (3.17). Suppose further that Y has the same distribution as X so that the same is true for Y_n, Y_n' for every n. The idea is to iterate copies of the solution in order to understand its behaviour. Consider the equation

$$K(\frac{x+y}{1+xy}) = c,$$

where $c > K$, or, equivalently,

$$y = \frac{\frac{c}{K} - x}{1 - \frac{cx}{K}}.$$

By the same argument as that of Lemma 15 but with c replaced by $\frac{c}{K}$,

$$\mathbb{P}(Y_{n+1} > c) \leq 2\mathbb{P}(Y_n > c/K)\mathbb{P}(Y_n' < K/c)$$

and

$$\mathbb{P}(Y_{n+1} < 1/c) \leq \mathbb{P}(Y_n > cK)^2 + \mathbb{P}(Y_n' < 1/cK)^2.$$

Now, for simplicity, assume that $q > p$ (we will consider the case $p > q$ in a moment). Let K_n denote the random variable that takes value K with probability p and value $1/K$ with probability q, independently of Y_n, and of each other for different n. We think of K_n

as being the noise associated with Y_n. Then we have

$$\mathbb{P}(Y_{n+1} > c | K_n = K)\mathbb{P}(K_n = K) \leq 2p\mathbb{P}(Y_n > c/K)\mathbb{P}(Y_n' < K/c),$$

$$\mathbb{P}(Y_{n+1} > c | K_n = 1/K)\mathbb{P}(K_n = 1/K) \leq 2q\mathbb{P}(Y_n > cK)\mathbb{P}(Y_n' < 1/cK),$$

and

$$\mathbb{P}(Y_{n+1} < 1/c | K_n = K)\mathbb{P}(K_n = K) \leq p\mathbb{P}(Y_n > cK)^2 + p\mathbb{P}(Y_n' < 1/cK)^2,$$

$$\mathbb{P}(Y_{n+1} < 1/c | K_n = 1/K)\mathbb{P}(K_n = 1/K) \leq q\mathbb{P}(Y_n > c/K)^2 + q\mathbb{P}(Y_n' < K/c)^2.$$

Adding these probabilities gives

$$\mathbb{P}(Y_{n+1} > c) + \mathbb{P}(Y_{n+1} < 1/c) \leq 2p\mathbb{P}(Y_n > c/K)\mathbb{P}(Y_n < K/c) + 2q\mathbb{P}(Y_n > cK)\mathbb{P}(Y_n < 1/cK)$$

$$+p\mathbb{P}(Y_n > cK)^2 + p\mathbb{P}(Y_n < 1/cK)^2$$

$$+q\mathbb{P}(Y_n > c/K)^2 + q\mathbb{P}(Y_n < K/c)^2.$$

Remembering that Y_n has the same distribution as X, we can write

$$\phi(c) \leq 2p\mathbb{P}(Y_n > c/K)\mathbb{P}(Y_n < K/c) + 2q\mathbb{P}(Y_n > cK)\mathbb{P}(Y_n < 1/cK)$$

$$+p\mathbb{P}(Y_n > cK)^2 + p\mathbb{P}(Y_n < 1/cK)^2$$

$$+q\mathbb{P}(Y_n > c/K)^2 + q\mathbb{P}(Y_n < K/c)^2.$$

Writing everything in terms of ϕ we have

$$\phi(c) \leq p\phi(cK)^2 + 2(q - p)\mathbb{P}(Y_n > cK)\mathbb{P}(Y_n < 1/cK)$$

$$+q\phi(c/K)^2 - 2(q-p)\mathbb{P}(Y_n > c/K)\mathbb{P}(Y_n < K/c).$$

It follows now, since $q > p$, $c, K > 1$ and $c/K > 1$, that

$$\phi(c) \leq p\phi(cK)^2 + q\phi(c/K)^2. \tag{3.20}$$

It is easily seen that if $p > q$ then the same is true but with p and q interchanged:

$$\phi(c) \leq q\phi(cK)^2 + p\phi(c/K)^2. \tag{3.21}$$

Now, let $f_m = \phi(K^m)$ so that $(f_m)_{m\geq 1}$ is a decreasing sequence. Then, using (3.20),

$$\phi(K^m) \leq p\phi(K^{m+1})^2 + q\phi(K^{m-1})^2 = pf_{m+1}^2 + qf_{m-1}^2$$

$$\leq pf_{m-1}^2 + qf_{m-1}^2 = f_{m-1}^2.$$

The same is clearly true if we use (3.21). Inductively,

$$f_m \leq f_1^{2^m},$$

but it may be the case that $f_1 = 1$ in which case this bound is useless. We have already seen that the distribution of X isn't concentrated on the two points $\{0\}$ and $\{\infty\}$ (we have assumed in particular that it is not $\delta_{\{0\}}$) and hence there exists an $L \geq 1$ such that $\phi(K^L) < 1$. We can write f_m in terms of f_L in this case:

$$f_m \leq f_L^{2^{m-L}},$$

where now $f_L < 1$. We write $x = K^m$ to obtain

$$\phi(x) \leq f_L^{2^{\log(x)/\log(K)-L}} = f_L^{x^{\log(2)/\log(K)}\times 2^{-L}} = e^{-\alpha x^\beta},$$

which holds for all $m > 1$, that is, for $\log x/\log K > 1$ or $x > K$, where $\alpha, \beta > 0$ are constants. □

Lemma 17. *Any solution to the RDE (3.16) has finite mean.*

Proof. We make use of the previous lemma. Let X be a random variable whose distribution is a solution to the RDE (3.16). Then

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > \lambda)d\lambda = \int_0^K \mathbb{P}(X > \lambda)d\lambda + \int_K^\infty \mathbb{P}(X > \lambda)d\lambda$$

$$\leq \int_0^K d\lambda + \int_K^\infty e^{-\alpha\lambda^\beta}d\lambda$$

(by the previous lemma)

$$= K + \int_K^\infty e^{-\alpha\lambda^\beta}d\lambda < \infty,$$

since $\alpha, \beta > 0$. □

In the next section we will combine this with a result about linear RDEs to deduce an important fact about the RDE (3.16) in the case $pq \leq 1/16$.

3.6.1 Linear approach to the non-linear RDE (3.16): basin of attraction of $\delta_{\{0\}}$

One obvious approach to tackling our more complicated non-linear RDE is to examine the linear part of the RDE in the numerator. More precisely, we claim the following.

Lemma 18. *Consider the linear RDE*

$$Y = (\frac{q}{p})^{\xi}(Y_1 + Y_2), \tag{3.22}$$

where ξ (as usual) takes value 1 with probability p and value -1 with probability $q = 1-p$. Let $Y_i^{(n)}(\nu)$ denote the iterates of a distribution ν under (3.22) and $X_i^{(n)}(\nu)$ the corresponding iterates under (3.16). Suppose that $X_i^{(0)} = Y_i^{(0)}$. Then $X_i^{(n)}(\nu) \leq Y_i^{(n)}(\nu)$ for all $n \geq 1$.

Proof. Define maps f and g by

$$f(x,y) = \frac{x+y}{1+xy}, \qquad g(x,y) = x+y.$$

Then $f(x,y) \leq g(x,y)$ for all $x, y \geq 0$. Let f^n denote the n-fold composition of f with itself and similarly for g^n and set $f^0 = g^0 = c$, for c some non-negative constant (where to form the composition we consider the mappings $f^*, g^* : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f^*(x,y) = (f(x,y), f(x,y)), g^*(x,y) = (g(x,y), g(x,y))$ and then set $f^n(x,y)$ to be the x-coordinate of $(f^*)^n(x,y)$ and similarly for g^n). We have $f^1 \leq g^1$. Suppose that this inequality holds for $n = N \geq 1$. That is, suppose $f^N(x,y) \leq g^N(x,y)$ for all $x, y \geq 0$. Then, for any $x, y \geq 0$,

$$f^{N+1}(x,y) = f(f^N(x,y)) \leq g(f^N(x,y)) \leq g(g^N(x,y)) = g^{N+1}(x,y).$$

Hence $f^n(x,y) \leq g^n(x,y)$ for all $x, y \geq 0$ and for all $n \geq 1$. The result now follows because for any point at which we evaluate $X_i^{(n)}(\nu), Y_i^{(n)}(\nu)$, we can apply the above with $f^0 = X_i^{(0)}(\nu), g^0 = Y_i^{(0)}(\nu)$, both evaluated at the point in question. □

Theorem 12. *Let $pq \leq 1/16$. Then the RDE (3.16) has the trivial distribution $\delta_{\{0\}}$ as unique solution. Furthermore, this solution is attracting, with basin of attraction contain-*

ing all distributions of finite mean.

Proof. We begin with the case $pq < 1/16$. Define a mapping U by setting $U(\nu)$ to be the distribution of

$$(\frac{q}{p})^{\xi}(Y_1 + Y_2),$$

where Y_1, Y_2 are independent with distribution ν, independent of ξ. Then we claim that $U^n(\nu)$ converges weakly to $\delta_{\{0\}}$ provided $pq < 1/16$ and ν has finite mean.

Consider the recursive tree process associated with (3.22), that is,

$$Y_u = (\frac{q}{p})^{\xi_u}(Y_{u0} + Y_{u1}), \qquad u \in T.$$

We now consider $\mathbb{E}[Y_u^r]$ and iterate. We have

$$\mathbb{E}[Y_u^r] = [p(\frac{q}{p})^r + q(\frac{p}{q})^r]\mathbb{E}[(Y_{u0} + Y_{u1})^r] \leq 2[p(\frac{q}{p})^r + q(\frac{p}{q})^r]\mathbb{E}[Y_{u0}^r]$$

for $r \in [0, 1]$. Let

$$T(r) = 2[p(\frac{q}{p})^r + q(\frac{p}{q})^r].$$

Then

$$T(1/2) = 4\sqrt{pq} < 1$$

if $pq < 1/16$. The significance of the value of $1/2$ is that it is the unique point in $[0, 1]$ which minimises T:

$$T'(r) = 2[p(\frac{q}{p})^r \log\frac{q}{p} + q(\frac{p}{q})^r \log\frac{p}{q}] = 0$$

if and only if $r = 1/2$ so that

$$\inf_{r\in[0,1]} T(r) = T(1/2) = 4\sqrt{pq}.$$

Now, by iterating $\mathbb{E}[Y_u^{1/2}]$, we conclude that, when $pq < 1/16$, Y_u converges to zero in $L^{1/2}$, provided our initial input random variables are in $L^{1/2}$. That is, $U^n(\nu)$ tends to a point mass at zero provided $\nu \in L^{1/2}$. Since $L^1 \subseteq L^{1/2}$, it now follows that $U^n(\nu)$ converges weakly to $\delta_{\{0\}}$.

The critical case $pq = 1/16$, (in which $T(1/2) = 1$ and $\mathbb{E}[Y_u^{\frac{1}{2}}] \leq \mathbb{E}[Y_{u0}^{\frac{1}{2}}]$) is far more delicate and follows from Theorem 1.5 of [38]. We will say more about this when we give a precise description of the basin of attraction of $\delta_{\{0\}}$ under linear RDEs. Combining the cases $pq < 1/16$ and $pq = 1/16$, by the bounding argument (Lemma 18), it now follows that, when $pq \leq 1/16$, the basin of attraction of the trivial solution $\delta_{\{0\}}$ under the RDE (3.16) contains all those distributions with finite mean (finite 1/2 mean in fact). It also follows that any non-trivial solution to the RDE (3.16) must have infinite mean, else the linear bound established converges in distribution to the trivial solution, forcing the same to be true for the RDE (3.16). But we have already proved that any solution to the RDE (3.16) must have finite mean (Lemma 17). It follows that, in the case $pq \leq 1/16$, $\delta_{\{0\}}$ is the only solution to the RDE (3.16). □

3.7 More about linear RDEs

Recall from the introduction that by "linear" we mean RDEs of the form

$$X = \sum_{i=1}^{N} \xi_i X_i.$$

For the theory we present, the ξ_i need not be independent. We state below a theorem which brings together the main results concerning the existence and uniqueness of solutions to linear RDEs. We assume that $N < \infty$ is non-random for the sake of simplicity. See, for example, [1] and [38] for a more general treatment.

Theorem 13. *(abridged from [1], [38]) Suppose that $\xi_i > 0$ for $1 \le i \le N$ and that $\sum_{i=1}^{N} \xi_i$ has finite k th moment for some $k > 1$. Let*

$$t(x) = \mathbb{E}[\sum_{i=1}^{N} \xi_i^x].$$

Suppose that $t(0) > 1$ and that $\inf_{x\in[0,1]} t(x) \le 1$. Then the RDE

$$X = \sum_{i=1}^{N} \xi_i X_i$$

has an invariant distribution X with $\mathbb{P}(X = 0) < 1$. If $t(1) = 1$ and $t'(1) \le 0$ then the same conclusion holds and this solution is unique up to a multiplicative constant in its argument. If, in addition, $t'(1) < 0$, then $\mathbb{E}[X] < \infty$.

As a simple illustrative example, this theorem can be used to deduce facts about the linear RDE (3.4) for Λ_n. We sketch the details below. Recall that

$$\Lambda_n = \frac{\Lambda_0}{2}(\Lambda'_{n-1} + \Lambda''_{n-1}),$$

where

$$\frac{\Lambda_0}{2} = \frac{1}{\sqrt{16pq}}(\frac{q}{p})^{\xi_\emptyset/2}.$$

In our case,

$$\begin{aligned} t(x) &= 2(p(\frac{1}{\sqrt{16pq}}\sqrt{\frac{q}{p}})^x + q(\frac{1}{\sqrt{16pq}}\sqrt{\frac{p}{q}})^x) \\ &= 2(p(\frac{1}{4p})^x + q(\frac{1}{4q})^x). \end{aligned}$$

We have then $t(1) = 1$ and

$$t'(1) = \frac{1}{2}\log\frac{1}{16pq}$$

so that $t'(1) < 0$ provided $pq > 1/16$. We conclude from the theorem above that the

RDE (3.4) has a non-trivial solution when $pq \geq 1/16$ (when $pq = 1/16$ we have $t(1) = 1$ and $t'(1) = 0$), which has finite mean when $pq > 1/16$. This agrees with the uniform integrability result of Chapter 2: we know that when $pq > 1/16$, Λ_n is uniformly integrable and therefore Λ_∞ has unit expectation. The above calculation tells us that when $pq > 1/16$ there are only two solutions to the RDE - the trivial solution and the non-trivial one covered by the theorem; the non-trivial distribution must correspond to the distribution of Λ_∞ and so it must have finite mean when $pq > 1/16$. Notice that this type of calculation is very similar to the rth mean calculations carried out both in this chapter and in Chapter 1. It should come as no surprise then that the function t is also related to the rate function I from Chapter 1, illustrating the connection between the ideas.

Theorem 14. *(Liu, [38]) Suppose that $\sum_{i=1}^N \xi_i$ has finite k th moment for some $k > 1$ and that $t(\alpha) = 1$ and $t'(\alpha) \leq 0$ for some $\alpha \in (0,1)$. For a distribution ν on $[0,\infty)$, if*

$$(3.23) \qquad \begin{cases} \lim_{x\to\infty} x^\alpha \nu(x,\infty) = 0, & t'(\alpha) < 0 \\ \lim_{x\to\infty} \frac{x^\alpha \nu(x,\infty)}{\log x} = 0, & t'(\alpha) = 0 \end{cases}$$

then $T^n(\nu)$ converges to $\delta_{\{0\}}$. Furthermore, if the above limits are infinite or take a non-zero finite value then ν is not in the basin of attraction of $\delta_{\{0\}}$.

It follows almost immediately from the theorem that distributions of finite mean are in the basin of attraction of $\delta_{\{0\}}$, for if a distribution ν has mean $m < \infty$, then

$$\nu(x,\infty) \leq \frac{m}{x}, \qquad x > 0,$$

by Markov's inequality. Hence

$$x^\alpha \nu(x,\infty) \leq m x^{\alpha-1}, \quad \frac{x^\alpha \nu(x,\infty)}{\log x} \leq \frac{m x^{\alpha-1}}{\log x},$$

both of which tend to 0 as $x \to \infty$ since $\alpha < 1$. The following lemma shows how Theorem 14 can be applied to obtain a full description of the basin of attraction of $\delta_{\{0\}}$ under the linear RDE. For the linear RDE (3.22), the map t is given by

$$t(x) = 2(p(\frac{q}{p})^x + q(\frac{p}{q})^x). \tag{3.24}$$

Notice that this is the same as the map T introduced in the proof of Theorem 12.

Lemma 19. *Let t be as given in (3.24). Then for $pq \leq 1/16$, there exists an $\alpha \in (0,1)$ such that $t(\alpha) = 1$ and $t'(\alpha) \leq 0$.*

Proof. We know from the proof of Theorem (12) that $t'(x) = 0$ if and only if $x = 1/2$ and that $t(1/2) = 4\sqrt{pq}$. Hence when $pq = 1/16$, $t(1/2) = 1$ and $t'(1/2) = 0$. Now, when $pq < 1/16$, the minimum of t moves downwards and hence we have $t(1/2) < 1, t(0) = 2$, so that there is an $\alpha \in (0, 1/2)$ with $t(\alpha) = 1$. Recall that

$$t'(x) = 2(p(\frac{q}{p})^x - q(\frac{p}{q})^x) \log \frac{q}{p},$$

which is easily seen to be less than 0 for $x \in (0, 1/2)$. □

Hence we are able to apply Theorem (14) to the linear RDE (3.22) to obtain a complete description of the basin of attraction of the trivial solution $\delta_{\{0\}}$. This applies equally to the non-linear RDE (3.16): by the bounding arguments, the basin of attraction of the trivial solution under the RDE (3.16) also contains of all of the distributions in the corresponding basin of attraction under the linear RDE (3.22). As a final remark, notice that the linear RDE (3.22) may be obtained from the RDE for Λ_n^θ by setting $\theta = q$. It is therefore possible to study the RDE (3.22) via the RDE for Λ_n^θ, whose behaviour we understand to some extent from Chapter 2. This further illustrates the connection between linear RDEs and martingales alluded to earlier.

3.8 A stable class for the RDE (3.16): the case $pq > 1/16$

We understand the RDE (3.16) fairly well in the case $pq \leq 1/16$. In the case $pq > 1/16$ we know that there exists a non-trivial solution but know very little about it thus far. In what follows we identify a class of random variables that is stable under the distributional equation corresponding to the RDE (3.16).

To begin with we need to spell out precisely what we mean by "stable" in this context. Let C be a class of random variables and $f(x_1, ..., x_n)$ a real-valued function of n real variables. Then we say that the class C is stable under the distributional equation

$$X = f(X_1, ..., X_n)$$

if, for independent, identically distributed random variables $X_1, ..., X_n$, each with law $\mu \in C$, the law of the random variable

$$X = f(X_1, ..., X_n)$$

also belongs to C.

We return briefly to the simple case $p = 1/2$ for motivation. In this case, the RDE reads

$$X = \frac{X_1 + X_2}{1 + X_1 X_2}. \tag{3.25}$$

Suppose that X_1, X_2 are concentrated on two points $x, 1/x$, where $x > 0$. Suppose further that $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 1$. Then X_1, X_2 have distribution

$$\frac{\delta_x}{x+1} + \frac{x\delta_{\frac{1}{x}}}{x+1}.$$

The random variable X, given by (3.25), can take three possible values:

$$X_1 = X_2 = x \Rightarrow X = \frac{2x}{1+x^2},$$

$$X_1 = x, X_2 = \frac{1}{x} \ (or\ X_1 = \frac{1}{x}, X_2 = x) \Rightarrow X = \frac{x + \frac{1}{x}}{2} = \frac{1+x^2}{2x},$$

$$X_1 = X_2 = \frac{1}{x} \Rightarrow X = \frac{\frac{2}{x}}{1+\frac{1}{x^2}} = \frac{2x}{1+x^2}.$$

Notice that the first and third of these possibilities are the same and that the second is their reciprocal. The corresponding distribution for X is therefore

$$\frac{2x}{1+x^2}\delta_{\frac{x^2+1}{(x+1)^2}} + \frac{1+x^2}{2x}\delta_{\frac{2x}{(1+x)^2}}$$

and it is easily seen that $\mathbb{E}[X] = 1$. We conclude, therefore, that, in the case $p = 1/2$, the class of distributions with unit expectation that are concentrated on reciprocal non-negative finite points is stable under our RDE.

In the general setting of the RDE (3.16), if the law of X_1, X_2 is concentrated on reciprocal points $x, \frac{1}{x}$ for some $x > 0$, then the law of X turns out to be concentrated on the four points

$$\frac{q}{p}\frac{2x}{1+x^2}, \quad \frac{p}{q}\frac{2x}{1+x^2}, \quad \frac{q}{p}\frac{1+x^2}{2x}, \quad \frac{p}{q}\frac{1+x^2}{2x}.$$

Notice that although the corresponding distribution isn't concentrated on two reciprocal points (as in the case $p = 1/2$), the first possible value is the reciprocal of the fourth and the same is true for the second and third values. This suggests that a "continuous" analogue of such distributions may be stable in the general setting. A continuous analogue can be formulated in terms of Radon-Nikodym derivatives as follows. For a discrete random variable X concentrated on points $x, \frac{1}{x}$, $(x > 0)$, and having unit expectation, we have

$$\frac{\mathbb{P}(X = x)}{\mathbb{P}(X = \frac{1}{x})} = \frac{1}{x}$$

and hence a continuous analogue is a random variable X on $(0, \infty)$, whose law μ_X satisfies

$$\frac{d\mu_X}{d\mu_{\frac{1}{X}}}(s) = \frac{1}{s}.$$

Definition 1. *We write $\mathcal{G}$ for the class of such random variables. That is,*

$$\mathcal{G} = \{X \in \mathcal{MF}(0, \infty) : \frac{d\mu_X}{d\mu_{\frac{1}{X}}}(s) = \frac{1}{s}\},$$

where $\mathcal{MF}(0, \infty)$ random variables taking values in $(0, \infty)$.

Remark 3. *Notice that*

$$\int t d\mu_X = \int d\mu_{1/X} = 1.$$

Hence random variables in $\mathcal{G}$ have unit expectation which is consistent with the discrete case.

For later work it will be useful to characterise random variables belonging to $\mathcal{G}$ in terms of expectation. We assume that $(0, \infty)$ is equipped with the Borel σ-algebra. Let f be a bounded measurable function on $(0, \infty)$ and $X \in \mathcal{G}$. Then we have

$$\int f(t) d\mu_X = \int f(t)/t d\mu_{1/X} = \int t f(1/t) d\mu_X$$

or, equivalently, $\mathbb{E}[f(X)] = \mathbb{E}[Xf(1/X)]$, with the expectations being meaningful by virtue of f being bounded. Observe also that if it is the case that

$$\int f(t)d\mu_X = \int tf(1/t)d\mu_X$$

for every bounded measurable function f, then $X \in \mathcal{G}$, since the class of bounded measurable functions contains the indicators of the Borel sets on $(0, \infty)$. Hence $X \in \mathcal{G}$ if and only if $\mathbb{E}[f(X)] = \mathbb{E}[Xf(1/X)]$ for every bounded measurable function f.

Theorem 15. *Random variables in $\mathcal{G}$ are stable under the RDE (3.16). More precisely, given independent $X_1, X_2 \in \mathcal{G}$, the random variable X given by (3.16) belongs to $\mathcal{G}$.*

Proof. We start by proving that if $X, Y \in \mathcal{G}$ are independent then $XY \in \mathcal{G}$. Our first observation is that if X, Y are independent then $XYf(1/XY)$ and $f(XY)$ are integrable for bounded measurable f. Hence by the expectation property established for random variables belonging to $\mathcal{G}$,

$$\begin{aligned}\mathbb{E}[XYf(1/XY)] &= \mathbb{E}\mathbb{E}[X\frac{Y}{X}f(X/Y)|Y] = \mathbb{E}[Yf(X/Y)] \\ &= \mathbb{E}\mathbb{E}[Y\frac{1}{Y}f(XY)|X] = \mathbb{E}[f(XY)].\end{aligned}$$

We now prove that for X, Y as before,

$$\frac{X+Y}{1+XY} \in \mathcal{G}.$$

Clearly $f(\frac{X+Y}{1+XY})$ is integrable and it follows that $\frac{X+Y}{1+XY}f(\frac{1+XY}{X+Y})$ is integrable because

$$\frac{X(\omega)+Y(\omega)}{1+X(\omega)Y(\omega)} \leq X(\omega)+Y(\omega)$$

for every ω and $X+Y$ is integrable. We have then

$$\mathbb{E}[\frac{X+Y}{1+XY}f(\frac{1+XY}{X+Y})] = \mathbb{E}[\frac{X}{1+XY}f(\frac{1+XY}{X+Y})] + \mathbb{E}[\frac{Y}{1+XY}f(\frac{1+XY}{X+Y})]$$

$$= \mathbb{E}\mathbb{E}[X\frac{1/X}{1+Y/X}f(\frac{1+Y/X}{1/X+Y})|Y] + \mathbb{E}\mathbb{E}[Y\frac{1/Y}{1+X/Y}f(\frac{1+X/Y}{X+1/Y})|X]$$

$$= \mathbb{E}[\frac{X}{X+Y}f(\frac{X+Y}{1+XY})] + \mathbb{E}[\frac{Y}{X+Y}f(\frac{X+Y}{1+XY})]$$

$$= \mathbb{E}[f(\frac{X+Y}{1+XY})].$$

The result now follows since $(\frac{q}{p})^{\xi} \in \mathcal{G}$. □

3.9 The density approach

Before presenting some conjectural results concerning the RDE (3.16) we look briefly at the properties that the pdf of a solution to the RDE (3.16) would have to have. It is stressed that we have not and will not prove that a solution has a density relative to Lebesgue measure. What follows are some results listing the properties of the density of a solution under the assumption that the solution admits a density. Of course we have already seen that when $pq \leq 1/16$ the only solution is $\delta_{\{0\}}$ and so this approach is tailored to the case $pq > 1/16$.

Lemma 20. *Let X be a continuous random variable on $[0, \infty)$ with density f_X. Let Y be a discrete random variable taking positive values y_1 and y_2 with probabilities α, β respectively. Then the random variable XY has density g given by*

$$g(u) = \frac{\alpha}{y_1} f_X(\frac{u}{y_1}) + \frac{\beta}{y_2} f_X(\frac{u}{y_2}).$$

Proof. We have

$$\mathbb{P}(XY \leq x|Y = y_1)\mathbb{P}(Y = y_1) = \alpha\mathbb{P}(y_1 X \leq x) = \alpha\mathbb{P}(X \leq \frac{x}{y_1}) = \alpha \int_0^{\frac{x}{y_1}} f_X(t)dt$$

and similarly

$$\mathbb{P}(XY \leq x|Y = y_2)\mathbb{P}(Y = y_2) = \beta\mathbb{P}(y_2 X \leq x) = \beta\mathbb{P}(X \leq \frac{x}{y_2}) = \beta \int_0^{\frac{x}{y_2}} f_X(t)dt.$$

Adding gives

$$\mathbb{P}(XY \leq x) = \alpha \int_0^{\frac{x}{y_1}} f_X(t)dt + \beta \int_0^{\frac{x}{y_2}} f_X(t)dt.$$

Substituting $u = y_1 t$ in the first integral and $u = y_2 t$ in the second gives

$$\mathbb{P}(XY \leq x) = \int_0^x \{\frac{\alpha}{y_1} f_X(\frac{u}{y_1}) + \frac{\beta}{y_2} f_X(\frac{u}{y_2})\}du.$$

□

Proposition 10. *Suppose that a solution to the RDE (3.16) has density g relative to Lebesgue measure. Then g satisfies the integral equation*

$$g(x) = \frac{p}{q}\int_0^\infty g(y)g(\frac{qy - px}{pxy - q})|\frac{1-y^2}{(\frac{pxy}{q} - 1)^2}|dy + \frac{q}{p}\int_0^\infty g(y)g(\frac{py - qx}{qxy - p})|\frac{1-y^2}{(\frac{qxy}{p} - 1)^2}|dy.$$

Proof. Let X, Y be independent random variables on $[0, \infty)$ having density g. Let $f_{X,Y}$ be the joint density of X, Y. Let $V = X, W = \frac{X+Y}{1+XY}$ so that if lower case letters represent values taken by the corresponding upper case random variables, then $x = v$ and $y = \frac{v-w}{vw-1}$. Define

$$J = \begin{pmatrix} \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \\ \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w} \end{pmatrix}.$$

Then, by the change of variables formula, we have

$$f_{V,W}(v,w) = f_{X,Y}(v, \frac{v-w}{vw-1})|detJ|.$$

For us,

$$J = \begin{pmatrix} 1 & \frac{w^2-1}{(vw-1)^2} \\ 0 & \frac{1-v^2}{(vw-1)^2} \end{pmatrix}$$

and hence

$$f_W(w) = \int_0^\infty f_{X,Y}(v, \frac{v-w}{vw-1})|\frac{1-v^2}{(vw-1)^2}|dv.$$

Now X, Y are assumed to be independent with common density g and hence this becomes

$$f_W(w) = \int_0^\infty g(v)g(\frac{v-w}{vw-1})|\frac{1-v^2}{(vw-1)^2}|dv.$$

The result now follows by applying the previous lemma with $y_1 = \frac{q}{p}, \alpha = p, y_2 = \frac{p}{q}, \beta = q$. $\square$

Notice that we have already remarked that the density g would also have to satisfy the functional equation

$$g(x) = \frac{g(\frac{1}{x})}{x^3}. \tag{3.26}$$

It turns out that there is a neat way of obtaining densities satisfying this functional relationship. Let $g_1(t)$ be a positive function defined on $[0,1]$. Define

$$g_2(t) = \frac{g_1(1/t)}{t^3}$$

on $[1,\infty)$. Then we claim that the density g given by

$$g(t) = \begin{cases} g_1(t), & t \in [0,1]; \\ g_2(t), & t \in [1,\infty) \end{cases}$$

satisfies the functional relationship (3.26), where, if necessary, we scale $g_1(t)$ so that

$$\int_0^\infty g(t)dt = 1.$$

As an example, if we take $g_1(t) = \lambda > 0$, then $g_2(t) = g_1(1/t)/t^3 = \lambda/t^3$, so that, in order to be a density, we must have

$$\lambda(\int_0^1 dt + \int_1^\infty \frac{1}{t^3}dt) = 1,$$

that is, we must have $\lambda = 2/3$. Now, the density g defined by

$$g(t) = \begin{cases} \frac{2}{3}, & t \in [0,1]; \\ \frac{2}{3t^3}, & t \in [1,\infty) \end{cases}$$

satisfies (3.26). Notice that it is clear that if $g(t)$ is a density then so too is $\frac{g(\frac{1}{t})}{t^3}$ since

$$\int_0^\infty \frac{g(\frac{1}{t})}{t^3}dt = \int_0^\infty ug(u)du = 1,$$

since we have already seen that stable random variables have unit expectation.

3.10 Summary of results and conjectures

We conclude this chapter by summarising what we know to be true for the RDE (3.16) and then making some conjectures.

3.10.1 The case $pq \leq 1/16$

We have proven that the trivial solution $\delta_{\{0\}}$ is the unique solution to the RDE (3.16) in this case and we have established that its basin of attraction contains all finite mean distributions. We then extended this via the theory given in [38].

3.10.2 The case $pq > 1/16$

The trivial distribution $\delta_{\{0\}}$ is still a solution in this case but we know that $t_\emptyset$ is also a solution and that it is non-trivial when $pq > 1/16$. In particular we have seen that in the special case $p = 1/2$, there are precisely two solutions, the other being $\delta_{\{1\}}$. We conjecture that in the general case $pq > 1/16$ there are always two solutions with the non-trivial one being the distribution of $t_\emptyset$. We have also identified distributions in the basin of attraction of the non-trivial solution (Theorem 12) and conjecture that it is in fact far larger, encompassing essentially all distributions on $[0, \infty)$, in keeping with the fact that this is so in the case $p = 1/2$.

Chapter 4

Endogeny

4.1 Introduction

In this chapter we introduce the important concept of endogeny for a tree-indexed solution to an RDE. We then study a particular example (the "noisy veto voter model") by applying a recent result of Warren in [44] concerning necessary and sufficient conditions for endogeny in the context of the binary tree and extend this result to arbitrary branching factor. We conclude the chapter by transforming the noisy veto voter RDE into an RDE which is of interest in its own right. This new RDE forms the basis of the work of Chapter 5.

4.2 The notion of endogeny

Recall the notion of a recursive distributional equation (RDE) from Chapter 3. There, we considered RDEs of the form

$$X = g(\xi_i, X_i; i = 1, ..., \mathbf{N}), \tag{4.1}$$

where **N** may be random or infinite. Recall also the associated recursive tree process (or RTP) in which we think of X as being a value associated with a vertex in a Galton-Watson tree, determined by the values X_i associated with its daughter vertices and some noise ξ associated with the parent. Since **N** may be infinite it is convenient to embed the random tree with offspring distribution **N** inside a tree with infinite branching factor. Let Γ_∞ denote the infinite tree with infinite branching factor. Then each vertex $u \in \Gamma_\infty$ gives rise to an infinite number of daughter vertices, the first N_u of which are "alive", the remaining being "dead", where N_u is an independent copy of **N**. By considering the collection of live vertices we obtain an embedding of the random tree in the infinite tree. Recall from Chapter 3 that $Y = (Y_u; u \in \Gamma_\infty)$ is a "tree-indexed" solution to the RDE (4.1) if

1. for every n, the random variables $(Y_u; |u| = n)$ are independent and identically distributed, having as distribution a fixed point of the induced map T;

2. for every $u \in \Gamma_\infty, Y_u = g(\xi_u, Y_{ui}; i = 1, ..., N_u)$, where (ξ_u, N_u) has the distribution ν (see precise setup of RDEs in Chapter 3), independently as u varies;

3. for every n, the random variables $(Y_u; |u| = n)$ and $(\xi_u, N_u; |u| \leq n - 1)$ are independent.

Notice that these conditions determine the joint law of Y. This means that a tree-indexed solution is also stationary in the strong sense, that is, a tree-indexed solution is translation invariant with respect to the root (if we consider the collection $Y^v = (Y_u; u \in (\Gamma_\infty)_v)$, where $(\Gamma_\infty)_v$ is the sub-tree rooted at v, then Y^v has the same distribution as Y for any $v \in \Gamma_\infty$). In this setting it is natural to wonder whether the "solution at root" $Y_\emptyset$ depends only on the initial data. This leads to the idea of "endogeny", a concept introduced in [1] that will play an important role in this chapter and the next.

Definition 2. *We say that a tree-indexed solution Y to the RDE (4.1) is endogenous if $Y_\emptyset$ is measurable with respect to $\sigma(\xi_u, N_u; u \in \Gamma_\infty)$. Writing Γ for the random tree* *with offspring distribution* $\boldsymbol{N}$*, this is equivalent to* $Y_\emptyset$ *being measurable with respect to* $\sigma(\xi_u, N_u; u \in \Gamma)$*. See [1].*

Informally speaking, endogeny means that there is no additional randomness in the system, "located at the boundary". In this sense our concerns are somewhat different from before. In Chapter 3 we were interested in identifying solutions to RDEs and making statements about uniqueness and basins of attraction and so on. In this chapter and the next we will be interested in whether the corresponding tree-indexed solutions are endogenous. For the example we study in this chapter the existence of a solution is rather trivial and it is the issue of endogeny that is of interest. We may at times use equivalent definitions of endogeny according to context. The following lemma justifies this.

Lemma 21. *Let* $Y = (Y_u; u \in \Gamma_\infty)$ *be a tree-indexed solution to the RDE (4.1). Then the following are equivalent:*

1. $Y_\emptyset$ *is measurable with respect to* $\sigma(\xi_u, N_u; u \in \Gamma_\infty)$*;*
2. Y_u *is measurable with respect to* $\sigma(\xi_v, N_v; v \in \Gamma_\infty)$ *for any* $u \in \Gamma_\infty$*;*
3. Y_u *is measurable with respect to* $\sigma(\xi_v, N_v; v \in (\Gamma_\infty)_u)$ *for any* $u \in \Gamma_\infty$*.*

Proof. Clearly both (2) and (3) imply (1). We deduce that (1) implies (3) from the fact that Y is strongly stationary (i.e. we think of u as being the root). Finally, it follows that (3) implies (2) from the fact that $\sigma(\xi_v, N_v; v \in (\Gamma_\infty)_u) \subseteq \sigma(\xi_v, N_v; v \in \Gamma_\infty)$ for any $u \in \Gamma_\infty$. □

Remark 4. *Notice that if a tree-indexed solution is endogenous then the property "*$(Y_u; |u| = n)$ *is independent of* $(\xi_u, N_u; |u| \leq n-1)$ *for every* n*" is automatic: for every* $u \in \Gamma_\infty$*,* Y_u *is measurable with respect to* $\sigma(\xi_v, N_v; v \in (\Gamma_\infty)_u)$ *and hence is independent of the* ξ_u *and* N_u *in levels* $n-1$ *and above.*

4.3 A simple example

We discuss briefly an example taken from [1]. With notation as in Chapter 3, suppose we are working on $A = \{0, 1\}$. Define $T : \mathcal{P}(A) \to \mathcal{P}(A)$ by

$$T(\mu) = \text{Bern}(1/2) \tag{4.2}$$

for all μ. Thus Bern($\frac{1}{2}$) is the unique fixed point of T. Now suppose that $N = 2$ so that we are working on the binary tree and that the ξ_u have Bern($\frac{1}{2}$) distribution. Set

$$g(a, x_1, x_2) = a.$$

This induces the map T given in (4.2). In the associated tree-indexed solution, each X_u has the Bern($\frac{1}{2}$) distribution and it is easily seen that $Y_\emptyset = \xi_\emptyset$ so that endogeny holds.

Consider now the von Neumann random bit extractor. This is a function $g^* : \{0, 1\}^\infty \to \{0, 1\}$ which, when applied to a Bern(p) sequence ($0 < p < 1$), produces a Bern($\frac{1}{2}$) sequence. Set

$$g(a, x_1, x_2, \ldots) = \begin{cases} a, & x_1 = x_2 = \ldots; \\ g^*(x_1, x_2, \ldots) & otherwise. \end{cases}$$

Take $N = \infty$ so that we are working on the infinite tree and let the ξ_u have the Bern($\frac{1}{2}$) distribution. Then the induced T is given by (4.2). For the associated tree-indexed solution in which the X_u have the Bern($\frac{1}{2}$) distribution, the ξ_u are never used and so $Y_\emptyset$ is in fact independent of $\sigma(\xi_u, N_u; u \in \Gamma)$, causing endogeny to fail. This shows that, in general, we cannot tell whether or not endogeny holds by looking at the induced map T alone, even when the solution is unique.

4.4 Noisy veto voter model

Consider the RDE on $\{-1, 1\}$ given by

$$X = \xi(X_1 \wedge X_2), \tag{4.3}$$

where ξ takes value 1 with probability p and value -1 with probability $q = 1 - p$. We wish to investigate endogeny for a tree-indexed solution.

The $(Y_u; u \in T)$ must satisfy the recursion

$$Y_u = \xi_u(Y_{u0} \wedge Y_{u1}), \qquad u \in T, \tag{4.4}$$

where ξ_u is an independent copy of ξ.

We will show that there exists an invariant measure for the RDE (4.3) and investigate a criterion for endogeny established by Warren in [44]. Finally, by conditioning in the right way, we will re-write the RDE as one in which the noise is incorporated into the underlying branching structure. The study of this RDE will be the subject of Chapter 5.

4.4.1 Existence of an invariant measure for (4.3)

Let μ be a probability measure on $\{-1, 1\}$ so that $\mu(1)$ denotes the probability of 1. Consider (4.3). There are two ways in which we can have $Y = 1$: if $Y_1 \wedge Y_2 = 1$ then we need $\xi = 1$, whereas if $Y_1 \wedge Y_2 = -1$, we require $\xi = -1$. Hence in order for μ to be invariant, we need

$$\mu(1) = p\mu(1)^2 + q(1 - \mu(1)^2) = (p - q)\mu(1)^2 + q$$

so that $\mu(1)$ satisfies the quadratic

$$(p-q)\mu(1)^2 - \mu(1) + q = 0. \tag{4.5}$$

Lemma 22. *There exists a unique invariant measure on $\{-1,1\}$ for the RDE (4.3).*

Proof. We have seen that the probability measure μ is invariant for (4.3) provided it satisfies equation (4.5). Let

$$F(x) = (p-q)x^2 - x + q.$$

Then $F(0) = q > 0$ and $F(1) = p - 1 = -q < 0$ so that F has at least one root in $(0,1)$. We claim that this is unique. We have

$$F(1/2) = \frac{q-p}{4} - \frac{1}{2} + q = \frac{p - q + 4q - 2(p+q)}{4} = \frac{q-p}{4}.$$

Hence $F(1/2) < 0$ if $p > 1/2$, $F(1/2) = 0$ if $p = 1/2$ and $F(1/2) > 0$ if $p < 1/2$. It now follows from this, and the fact that F is continuous on $[0,1]$ with $F(0) = q, F(1) = -q$, that if $p < 1/2$ then F has a root in $(1/2,1)$ and cannot have one in $[0,1/2]$; that if $p = 1/2$ then $1/2$ is the unique root (so that $\mu(1) = 1/2$ is the invariant measure); and that if $p > 1/2$ then F has a root in $[0,1/2)$ and cannot have one in $[1/2,1]$. □

4.5 A criterion for endogeny

In this section we state a generalisation of a result of Warren in [44], giving necessary and sufficient conditions for endogeny of a tree-indexed solution to an RDE. Consider the recursion

$$X_u = \phi(X_{u0}, X_{u1}, \dots, X_{u(d-1)}, \xi_u), \qquad u \in \Gamma_d,$$

where the X_u take values in a finite space S, the "noise" terms ξ_u take values in a space E, Γ_d is the deterministic d-ary tree and where ϕ is symmetric in its first $d-1$ arguments. We suppose that the ξ_u are independent with common law ν and that there exists a measure μ which is invariant for the above recursion (i.e. μ is a solution of the associated RDE). Let $u_0 = \emptyset, u_1, u_2, ...$ be an infinite line of descent. For $n \leq 0$, define $X_n = X_{u_{-n}}$. Then, under the invariant measure μ, the law of the sequence $(X_n; n \leq 0)$, which, by the symmetry of ϕ does not depend on the choice of sequence of vertices chosen, is that of a stationary Markov chain. Let P^2 be the transition matrix of a Markov chain on S^2, given by

$$P^2((x_1, x_1'), A \times A') = \int_S \int_E \mathbf{1}(\phi(x_1, x_2, ..., x_d, z) \in A, \phi(x_1', x_2, ..., x_d, z) \in A') d\nu(z) d\mu(x_2) ... d\mu(x$$

Let P^- be the restriction of P^2 to non-diagonal terms and ρ the largest eigenvalue of the matrix corresponding to P^-. Write $\mathcal{H}_0$ for all L^2 random variables measurable with respect to $X_\emptyset$ and $\mathcal{K}$ for the L^2 random variables measurable with respect to $(\xi_u; u \in \Gamma_d)$.

Theorem 16. *(Generalisation of Theorem 1, [44]) The tree-indexed solution to the RDE associated with*

$$\xi_u = \phi(\xi_{u0}, \xi_{u1}, ..., \xi_{u(d-1)}, \epsilon_u),$$

corresponding to the invariant measure μ, is endogenous if and only if $d\rho \leq 1$. In the critical case $d\rho = 1$ endogeny holds provided P^- is irreducible and $\mathcal{H}_0 \cap \mathcal{K}^\perp = \{0\}$.

The theorem above is a simple technical generalisation of Warren's result. The additional criteria given for endogeny in the critical case are analogous to the conditions under which a Galton-Watson branching process with mean one becomes extinct. See [44] for further discussion and the technical reasons behind these conditions.

We now apply the theorem to the noisy veto voter model on the binary tree. In this setting, the map ϕ is given by

$$\phi(\xi_u, X_{u0}, X_{u1}) = \xi_u(X_{u0} \wedge X_{u1})$$

and the corresponding condition for endogeny is "$2p \leq 1$".

Lemma 23. *For the noisy veto voter RDE on the binary tree, P^- is given by*

$$P^- = \begin{pmatrix} \mu(1)p & \mu(1)q \\ \mu(1)q & \mu(1)p \end{pmatrix}.$$

Proof. With ϕ as given, consider the transition from $(-1, 1)$ to $(-1, 1)$. Under the equation

$$x = z(x_0 \wedge x_1), \tag{4.6}$$

the pair $(-1, x_1)$ maps to $z(-1 \wedge x_1) = -z$ and we therefore require $z = 1$. Under (4.6), and with $z = 1$, the pair $(1, x_1)$ maps to $1 \wedge x_1 = x_1$. Hence we need $x_1 = 1$. Combining, $(-1, 1)$ maps to $(-1, 1)$ provided $x_1 = z = 1$, whose probability is $\mu(1)p$. Consider now the transition from $(-1, 1)$ to $(1, -1)$. As before, $(-1, x_1)$ maps to $z(-1 \wedge x_1) = -z$ so that we require $z = -1$. Now, with $z = -1$, $(1, z_1)$ maps to $-(1 \wedge x_1) = -x_1$ and hence we also require $x_1 = 1$. Combining, we require $x_1 = 1$ and $z = -1$, whose probability is $\mu(1)q$. It is easily seen that the probability of moving from $(1, -1)$ to $(-1, 1)$ is the same as that of moving from $(-1, 1)$ to $(1, -1)$ and similarly that the probability of moving from $(1, -1)$ to $(1, -1)$ is the same as that of moving from $(-1, 1)$ to $(-1, 1)$. □

Theorem 17. *The solution to the RDE (4.3) is endogenous if and only if $p \geq 1/2$.*

Proof. Notice that

$$\begin{pmatrix} \mu(1)p & \mu(1)q \\ \mu(1)q & \mu(1)p \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \mu(1)p + \mu(1)q \\ \mu(1)p + \mu(1)q \end{pmatrix} = \begin{pmatrix} \mu(1) \\ \mu(1) \end{pmatrix}$$

so that $\binom{1}{1}$ is a positive eigenvector of P^- with corresponding eigenvalue $\mu(1)$. It follows by Perron-Frobenius theory that $\mu(1)$ is the maximal eigenvalue ρ. The criterion for endogeny from the theorem in the non-critical case is therefore $2\mu(1) < 1$ or $\mu(1) < 1/2$, which corresponds to the case $p > 1/2$ (see the earlier analysis). For the critical case $\mu(1) = p = 1/2$, observe firstly that P^- is clearly irreducible. For the second criterion, let $X \in \mathcal{H}_0 \cap \mathcal{K}^{\perp}$. Then $X = f(X_\emptyset)$ for some L^2 function f and $\mathbb{E}[XY] = 0$ for all $Y \in \mathcal{K}$. Taking $Y = 1 \in \mathcal{K}$, we deduce that $\mathbb{E}[X] = 0$. Since $X_\emptyset$ takes only values -1 and 1, we can write

$$X = a\mathbf{1}(X_\emptyset = 1) + b\mathbf{1}(X_\emptyset = -1)$$

for constants a and b. In the case $p = 1/2$ we must have that $a + b = 0$ (of course if either $a = 0$ or $b = 0$ then there is nothing to prove so we assume that $a \neq 0$) so that we can re-write X, without loss of generality, as

$$X = \mathbf{1}(X_\emptyset = 1) - \mathbf{1}(X_\emptyset = -1).$$

Now, let $Y = \mathbf{1}(\xi_\emptyset = 1) \in \mathcal{K}$. Then

$$\mathbb{E}[XY] = \mathbb{E}[\mathbf{1}(\xi_\emptyset = 1)(\mathbf{1}(X_\emptyset = 1) - \mathbf{1}(X_\emptyset = -1))]$$

$$= \mathbb{E}[2.\mathbf{1}(\xi_\emptyset = 1)(\mathbf{1}(X_0 = X_1 = 1) - 1)]$$

$$= 2((1/2)^3 - 1) = -1/4,$$

which contradicts the assumption. It follows that $\mathcal{H}_0 \cap \mathcal{K}^\perp = \{0\}$. □

4.5.1 An equivalent problem on $\{0, 1\}$

Consider the RDE on $\{0, 1\}$ given by

$$X = \begin{cases} X_1 X_2, & \xi = 1 \\ 1 - X_1 X_2 & \xi = 0 \end{cases} \tag{4.7}$$

where ξ takes value 1 with probability p and value 0 with probability $q = 1 - p$. It is easily seen that there exists an invariant measure μ for (4.7) and that it is precisely the invariant measure for (4.3) but with " $-$ 1" re-labelled as "0". We have derived a condition for endogeny for (4.3) (which therefore holds for (4.7) too) but would like to gain an understanding of what the solutions actually look like. We claim that studying the RDE (4.7) is the same as studying the RDE (4.3).

The endogenous case $p \geq 1/2$

Let

$$E_u^{n+1} = \mathbb{E}[X_u | \xi_{uv}; |v| \leq n]$$

and

$$E_u = \mathbb{E}[X_u | \xi_v; v \in T_u].$$

Then the sequence $(E_u^n)_n$ is a bounded martingale and $E_u^n \to E_u$ almost surely for any $u \in T$. In the endogenous case $p \geq 1/2$, X_u is measurable with respect to $\sigma(\xi_v; v \in T_u)$ and hence $E_u = X_u$. We have then a martingale sequence E_u^n which converges to the endogenous solution.

We can say a bit more about this martingale sequence. The recursion corresponding to (4.7) is

$$X_u = \begin{cases} X_{u0}X_{u1}, & \xi_u = 1 \\ 1 - X_{u0}X_{u1} & \xi_u = 0 \end{cases} \tag{4.8}$$

which we can re-write as

$$X_u = \xi_u X_{u0} X_{u1} + (1 - \xi_u)(1 - X_{u0}X_{u1}).$$

It is easily seen that the E_u^n satisfy essentially the same recursion:

$$E_u^n = \xi_u E_{u0}^{n-1} E_{u1}^{n-1} + (1 - \xi_u)(1 - E_{u0}^{n-1} E_{u1}^{n-1})$$

with $E_u^0 = \mu(1)$ for any $u \in T$. The boundary condition $E_u^0 = u(1)$ comes from the fact that if we consider the first n levels of the tree and then start iterating one level below the boundary then the conditional expectation is just the expectation of X_u since there is nothing on which to condition.

The non-endogenous case $p < 1/2$

Consider the recursion (4.8). Suppose now that we distinguish the vertices $u \in T$ such that $\xi_u = 0$ (think of them as being coloured red for example) and then remove the part of the tree descended from such vertices. In this way, each vertex now has either 0 or 2 offspring, the probability of the latter being p. The mean number of offspring is hence $2p$ so that in the non-endogenous case $p < 1/2$ the process dies out almost surely. We can therefore start at the boundary of the tree and iterate back up to the root. Notice that this applies equally to $p = 1/2$, the critical value for endogeny. We think of the collection of red marks as being a random tree embedded within the binary tree. We see that, conditional on $\xi_\emptyset = 0$, (4.8) becomes (since $\xi_u = 1$ for the remaining non-red vertices)

$$X_\emptyset = 1 - \prod_{\textit{red vertices } u} X_u,$$

or, for a general vertex u,

$$X_u = 1 - \prod_{i=1}^{D_u} X_{ui},$$

where D_u denotes an independent copy of the total number of red vertices. This can be interpreted in terms of yet another tree, this time with branching factor equal to the total number of red vertices in the original tree. The RDE corresponding to this recursion is of interest in its own right and will be studied in detail in Chapter 5.

We conclude this section by looking more closely at the total number of red vertices D. We obtain an RDE which its distribution satisfies and then derive a formula for its generating function. It turns out (Chapter 5) that analysis of the RDE

$$X = 1 - \prod_{i=1}^{D} X_i \tag{4.9}$$

relies heavily on properties of the generating function of D.

The random variable D corresponds to the number of red vertices in the above construction. Recall that a vertex u is marked red if $\xi_u = 0$. Thus we can think of D as being the total number of deaths in a Galton-Watson branching process in which each vertex has (independently) either 0 or 2 daughters, the probability of the former being q and the latter $p = 1 - q$. In Chapter 3 we saw an RDE for the total offspring distribution of a Galton-Watson process. In the course of the proof of the next theorem we write down a similar RDE for the total number of deaths.

Theorem 18. *Consider a Galton-Watson process in which each vertex has either 0 offspring (with probability q) or 2 offspring (with probability $p = 1 - q$). Then the total number of deaths D has generating function G given, when $p \leq 1/2$, by*

$$G(s) = \frac{1 - \sqrt{1 - 4pqs}}{2p}.$$

Proof. Suppose that we are in the more general setting of a Galton-Watson process with offspring distribution M. Then D satisfies

$$D = \sum_{i=1}^{M} D_i + \mathbf{1}(M = 0),$$

where the indicator is present so that, in the case where there are no births, the number of deaths is recorded as one, namely that of the root. As in the statement of the theorem, G is the generating function of D. Let H be the generating function of M. Then we have

$$\mathbb{E}[s^{D_1 + \ldots + D_M + \mathbf{1}(M=0)} | M = m] = \begin{cases} s, & m = 0 \\ G(s)^m, & m \geq 1 \end{cases}$$

so that

$$\mathbb{E}[\mathbb{E}[s^{D_1 + \ldots + D_M + \mathbf{1}(M=0)} | M = m]] = s\mathbb{P}(M = 0) + \sum_{m \geq 1} G(s)^m \mathbb{P}(M = m)$$

$$= s\mathbb{P}(M = 0) + H(G(s)) - \mathbb{P}(M = 0).$$

Hence

$$G(s) = H(G(s)) + \mathbb{P}(M = 0)(s - 1),$$

where for the example we are considering,

$$H(s) = q + ps^2.$$

We have then

$$pG(s)^2 - G(s) + sq = 0.$$

Solving, we find that for $p \leq 1/2$,

$$G(s) = \frac{1 - \sqrt{1 - 4pqs}}{2p}.$$

□

Remark 5. *Notice that when $p > 1/2$ there is positive probability that the process never becomes extinct and hence a total count is not possible. Notice also that $G(0) = 0$, i.e. the probability that there are no deaths is 0. When we analyse the RDE (4.9) in Chapter 5 this will be an integral assumption.*

Chapter 5

An RDE on the unit interval

In this chapter we study the RDE derived from the noisy veto voter model of Chapter 4. We give a completely general treatment by considering it as an RDE in its own right.

5.1 Introduction

Let N be a random variable, with probability generating function H, taking values in $\overline{\mathbb{Z}}_+ = \{0, 1, ...; \infty\}$. It is assumed throughout that $H(0) = 0$ (i.e. that N is almost surely positive). Write $\mathbb{D}$ for the set of distributions on $[0, 1]$. Recall the following RDE from the study of the noisy veto voter model in Chapter 4:

$$X = 1 - \prod_{i=1}^{N} X_i. \tag{5.1}$$

As we have remarked before in relation to the study of RDEs, a convenient generalisation is the so-called "tree-indexed" problem, in which we think of the X_i as being marks associated with a Galton-Watson branching process. Recall also from Chapter 4 the notion of endogeny in the context of tree-indexed solutions; a tree-indexed solution is said to be endogenous if it is measurable with respect to the underlying noise.

As in Chapter 4, it is convenient to work on an infinite tree with infinite branching factor and regard the random tree with offspring distribution N as being embedded within it. We now elucidate this further for our current purposes. An initial ancestor (in level zero), which we denote $\emptyset$, gives rise to a countably infinite number of daughter vertices (which form the members of the first generation), each of which gives rise to an infinite number of daughters (which form the members of the second generation), and so on. As usual we write $uj, j = 0, 1, 2, \ldots$, for the daughters of a vertex u. We write $\mathbf{T}$ for the collection of all relatives of the root (i.e. $\mathbf{T} = \bigcup_{n=0}^{\infty}(\mathbb{Z}_+)^n$) and think of it as being partitioned by depth, that is, as being composed of levels or generations, in the way described. Associated to each vertex $u \in \mathbf{T}$ is an independent copy N_u of N, telling us the (random) number of offspring produced by u. In this way we think of some vertices as being "alive" (relative to $\emptyset$) and others as being "dead": each vertex u has infinitely many daughters $u1, u2, \ldots \in \mathbf{T}$, with the vertices $u1, u2, \ldots, uN_u$ being alive and $\{uj : j > N_u\}$ dead. We can now write our RDE (5.1) as a recursion on the vertices of $\mathbf{T}$ and iterate:

$$X_u = 1 - \prod_{i=1}^{N_u} X_{ui}, \qquad u \in \mathbf{T}. \tag{5.2}$$

The advantage of the embedding now becomes clear: we can talk about the RDE at any vertex in the infinite tree (which is mathematically convenient) and yet, because the product only runs over the live vertices relative to u, the random tree with branching factor N is encoded into the RDE as "noise".

5.2 The discrete and conditional probability solutions

Lemma 24. *There exists a unique probability measure on $\{0, 1\}$ which is invariant for the RDE (5.1).*

Proof. Let X be a random variable whose distribution is invariant for (5.1). Let $\alpha = \mathbb{P}(X = 1)$. We have then $\mathbb{P}(X = 0) = 1 - \alpha$ and

$$\mathbb{P}(X_i = 1; i = 1, ..., N) = \sum_n \mathbb{P}(X_i = 1; i = 1, ..., n | N = n)\mathbb{P}(N = n)$$

$$= \sum_n \alpha^n \mathbb{P}(N = n) = \mathbb{E}[\alpha^N].$$

Now, $X = 0$ if and only if $X_i = 1$ for $i = 1, ..., N$ and $N \geq 1$. Hence a necessary and sufficient condition for invariance is

$$1 - \alpha = \mathbb{E}[\alpha^N] = H(\alpha).$$

Now, let

$$K(x) := H(x) + x - 1.$$

Since H is a generating function and $H(0) = 0$, we have $K(0) = -1 < 0$ and $K(1) = 1 > 0$ so that K is guaranteed to have a zero in $(0, 1)$. Notice that this is unique since the mapping $x \mapsto H(x) + x$ is strictly increasing. □

We can now deduce that there exists a tree-indexed solution on $\{0,1\}^{\mathbf{T}}$ to the RDE (5.1) by virtue of Lemma 6 of [1]. Alternatively this can be argued via the Kolmogorov Extension Theorem.

Theorem 19. *Let $S = (S_u; u \in \boldsymbol{T})$ be a tree-indexed solution on $\{0,1\}^{\boldsymbol{T}}$ to our RDE (5.1), which we will henceforth refer to as the "discrete solution" (i.e. the S_u have the invariant distribution on $\{0,1\}$). Let $C_u = \mathbb{P}(S_u = 1 | N_v; v \in \boldsymbol{T})$. Then $C = (C_u; u \in \boldsymbol{T})$ is also a tree-indexed solution to the RDE (5.1).*

Proof. To verify the relationship between the random variables, we have, writing $\underline{N} =$

$(N_u; u \in \mathbf{T})$ to ease the notation,

$$
\begin{aligned}
C_u = \mathbb{E}[\mathbf{1}(S_u = 1)|\underline{N}] &= \mathbb{E}[S_u|\underline{N}] \\
&= \mathbb{E}[1 - \prod_{i=1}^{N_u} S_{ui}|\underline{N}] \\
&= 1 - \mathbb{E}[\prod_{i=1}^{N_u} S_{ui}|\underline{N}] \\
&= 1 - \prod_{i=1}^{N_u} \mathbb{E}[S_{ui}|\underline{N}] \\
&= 1 - \prod_{i=1}^{N_u} C_{ui}
\end{aligned}
$$

by conditional independence. To verify stationarity, let

$$C_u^n = \mathbb{P}(C_u = 1|N_u; |u| \leq n).$$

Then the sequence $(C_u^n)_{n\geq 1}$ is a uniformly integrable martingale and so converges almost surely to a limit which must be C_u. Now, we can write C_u^n as

$$C_u^n = 1 - \prod_{i=1}^{N_u} C_{ui}^n = 1 - \prod_{i=1}^{N_u}(1 - \prod_{j=1}^{N_{ui}}(...(1 - (\mu^1)^{f(N_u)})...)) \to C_u \qquad a.s.,$$

where the exponent $f(N_u)$ is a function of N_u and $\mu^1 = \mathbb{E}[S_u]$. Now, $(C_u^n; u \in \mathbf{T})$ is stationary since each C_u^n is an explicit function of N_u, which is itself stationary. Since C_u is the (almost sure) limit of a sequence of stationary random variables, it follows that $C = (C_u; u \in \mathbf{T})$ is stationary. Notice that the conditional probabilities solution C is automatically endogenous since C_u is $\sigma(N_v; v \in \mathbf{T}_u)$-measurable for every $u \in \mathbf{T}$ and hence $(C_u; |u| = n)$ is independent of $(N_u; |u| \leq n - 1)$ by Lemma 21 of Chapter 4. The independence of the collection $(C_u; |u| = n)$ follows from the fact that the $(S_u; |u| = n)$ are independent. □

Remark 6. *Notice that if S is endogenous then $C = S$ almost surely so that if S and C do not coincide then S cannot be endogenous.*

5.3 The moment equation and uniqueness of solutions

Much of the work of this chapter relies heavily on the analysis of the following equation.

Theorem 20. *Any solution to the RDE (5.1) must have moments $(m_n)_{n \geq 0}$ satisfying the equation*

$$H(m_n) - (-1)^n m_n = \sum_{k=0}^{n-1} \binom{n}{k} (-1)^k m_k, \tag{5.3}$$

where $m_n^{1+1/n} \leq m_{n+1} \leq m_n$ and $m_0 = 1$.

Proof. Let X be a random variable whose distribution is invariant for the RDE (5.1) and write $m_k = \mathbb{E}[X^k]$. Applying the RDE (5.1) once to $(1 - X)^n$ we have

$$\mathbb{E}[(1 - X)^n] = \mathbb{E}[\prod_{i=1}^{N} X_i^n] = H(m_n).$$

On the other hand, by expanding $(1 - X)^n$ we obtain

$$\mathbb{E}[(1 - X)^n] = \mathbb{E}[\sum_{k=0}^{n} \binom{n}{k} (-1)^k X_i^k]$$

$$= \sum_{k=0}^{n} \binom{n}{k} (-1)^k m_k,$$

so that

$$H(m_n) = \sum_{k=0}^{n} \binom{n}{k}(-1)^k m_k.$$

The condition $m_{n+1} \leq m_n$ follows from the fact that the distribution is on $[0, 1]$. The other condition follows from the monotonicity of L^p norms: $(\mathbb{E}X^n)^{1/n} \leq (\mathbb{E}X^{n+1})^{1/(n+1)}$. □

5.3.1 Example: the binary tree

As an example, if the random variable N had generating function $H(x) = x^2$ (i.e. $N \equiv 2$), then the moment equation tells us that the first moment m_1 of an invariant distribution for the RDE (5.1) satisfies

$$m_1^2 + m_1 - 1 = 0$$

so that $m_1 = (\sqrt{5} - 1)/2$. For the second moment m_2 we have

$$m_2^2 - m_2 - (2 - \sqrt{5}) = 0$$

so that $m_2 = m_1$ or m_1^2 and so on. In fact the two possible moment sequences turn out to be $m_0 = 1, m_n = (\sqrt{5}-1)/2$ for $n \geq 1$ or $m_0 = 1, m_1 = (\sqrt{5}-1)/2, m_n = m_1^n$ for $n \geq 2$.

We now state the main result of this chapter.

Theorem 21. *Let $S = (S_u; u \in \boldsymbol{T})$ and $C = (C_u; u \in \boldsymbol{T})$ be, respectively, the discrete solution and corresponding conditional probability solution to the RDE (5.1). Let $\mu^1 = \mathbb{E}[S_u] = \mathbb{E}[C_u]$. Then*

1. *S is endogenous if and only if $H'(\mu^1) \leq 1$;*

2. *C is the unique endogenous solution;*

3. *The only invariant distributions for the RDE (5.1) are those of $S_\emptyset$ and $C_\emptyset$.*

The proof of the theorem relies on several lemmas. For part (1) we use a generalisation of a result of Warren [44] (given in Chapter 4) by first truncating N, and then take limits. For parts (2) and (3) we advance our own arguments by making use of the moment equation (5.3). The proof of parts (2) and (3) in the case $H'(\mu^1) \leq 1$ is straightforward and given below.

Proof. (of parts (2) and (3) of Theorem 21 in the case $H'(\mu^1) \leq 1$)

Let $(m_n)_{n\geq 0}$ be the moments of an invariant distribution. The moment equation (5.3) says that m_1 is a root of the equation

$$H(x) + x = 1.$$

Since H is increasing this has unique solution $x = \mu^1$ (and therefore any invariant distribution will have first moment μ^1). For $n = 2$ the moment equation says that m_2 is a root of

$$H(x) - x = 1 - 2m_1. \tag{5.4}$$

Let m_* denote the minimum of the function $H(x) - x$. Then $H'(m_*) = 1$ and, since $H'(\mu^1) \leq 1$, we have $\mu^1 \leq m_*$. Now, μ^1 is a root of (5.4) since the discrete solution on $\{0, 1\}$ satisfies the recursion and hence its moments must satisfy the moment equation. The other possible root m_2 is to the right of m_* and so greater than μ^1 and hence cannot be a moment. Hence we must have $m_1 = m_2 = \mu^1$. We conclude that, under the conditions of the theorem, any solution Z to the RDE (5.1) must satisfy $\mathbb{E}[Z] = \mathbb{E}[Z^2]$ or, equivalently, $\mathbb{E}[Z(1 - Z)] = 0$ so that $Z(1 - Z) = 0$ almost surely and thus Z is concentrated on $\{0, 1\}$. □

5.3.2 Bounded branching factor

Lemma 25. *Define* $N^n = min(n, N)$ *and denote its generating function by* H_n*. Then* N^n *is bounded and*

1. $N^n \to N$ *almost surely;*
2. $H_n(s) \geq H(s)$ *for all* $s \in [0, 1]$*;*
3. $H_n \to H$ *uniformly on* $[0, 1]$*;*
4. $H'_n \to H'$ *uniformly on compact subsets of* $[0, 1)$*.*

Proof. The first part of the lemma is easily seen by writing

$$N^n = N\mathbf{1}_{(N \leq n)} + n\mathbf{1}_{(N>n)}.$$

For the second part, note that $N^n \leq N$ and therefore $s^{N^n} \geq s^N$ for $s \in [0, 1]$. We have then

$$H_n(s) = \mathbb{E}s^{N^n} \geq \mathbb{E}s^N = H(s), \qquad s \in [0, 1].$$

For the third property, we have

$$\mathbb{E}[|s^{N^n} - s^N|] \leq \mathbb{E}[\mathbf{1}_{(N>n)}|s^n - s^N|]$$

$$\leq \mathbb{E}[\mathbf{1}_{(N>n)} \sup_n |s^n - s^N|]$$

$$\leq \mathbb{P}(N > n) \times const.,$$

where the constant does not depend on n. Hence

$$\mathbb{E}[|s^{N^n} - s^N|] \to 0$$

uniformly. For the uniform convergence of H'_n notice that we can differentiate a power series term-by-term to obtain another power series which converges inside the radius of convergence of the original. We have then

$$\mathbb{E}[|N^n s^{N^n-1} - N s^{N-1}|] \leq \mathbb{E}[\mathbf{1}_{(N>n)}|ns^{n-1} - Ns^{N-1}|].$$

Now note that for s in a compact subset of $[0,1)$ we have $s \leq 1-\epsilon$ for some $\epsilon > 0$. Hence

$$|ns^{n-1} - Ns^{N-1}| \leq n(1-\epsilon)^{n-1} \to 0$$

as $n \to \infty$, giving the desired uniform convergence. □

Lemma 26. *Assume that $H'(\mu^1) > 1$. Let $C^n_u = \mathbb{P}(S_u = 1|N^n_u; u \in \boldsymbol{T})$ denote the conditional probabilities solution for N^n. Let $\mu^k_n = \mathbb{E}[(C^n_u)^k]$ denote the corresponding k th moment and let $\mu^k = \mathbb{E}[C^k]$. Let $\mu^2_{n,m}$ denote the root of the (modified) equation for the second moment*

$$H_n(x) - x = 1 - \mu^1_m - \mu^1_n \tag{5.5}$$

to the left of the minimum of $H_n(x) - x$ (i.e. the lesser of the two possible roots). Then $\mu^k_n \to \mu^k$ for $k = 1, 2$ and $\mu^2_{n,m} \to \mu^2$.

Proof. For the case $k = 1$ consider the graphs of the functions $H_n(x) + x$ and $H(x) + x$. We have $H_n(x) \geq H(x)$ for all $x \geq 0$ and for all $n \geq 1$ so that μ^1_n is bounded above by μ^1 for every n. Furthermore, since H_n increases to H pointwise on $[0,1]$, it follows that the μ^1_n are increasing. The μ^1_n must therefore have a limit, which we will denote $\widehat{\mu}$. H is continuous and so $H(\mu^1_n) \to H(\widehat{\mu})$ and, furthermore, since $H_n \to H$ uniformly on compact subsets of $[0,1)$, we have $H_n(\mu_n^1) \to H(\widehat{\mu})$. Hence

$$1 = H_n(\mu_n^1) + \mu_n^1 \to H(\widehat{\mu}) + \widehat{\mu},$$

so that $\widehat{\mu}$ is a root of $H(x) + x = 1$. We must therefore have $\widehat{\mu} = \mu^1$ since this equation has only one root.

For the case $k = 2$ we consider the graphs of $H_n(x) - x$ and $H(x) - x$. Let μ_n^* be the minimum of $H_n(x) - x$ and μ^* the minimum of $H(x) - x$. We first show that $\mu_n^* \to \mu^*$, then that $\mu_n^2 \to \mu^2$ and finally that $\mu_{n,m}^2 \to \mu^2$ as $\min(n,m) \to \infty$.

Notice first that H_n' converges to H' uniformly on compact subsets of $[0,1)$ by Lemma 25, so the result follows if we can show that the sequence (μ_n^*) is bounded away from 1. But $\mu_n^1 \to \mu^1$ and $\mu_n^* < \mu_n^1$ so the conclusion follows.

To show that $\mu_n^2 \to \mu^2$ we argue that μ^2 is the only limit point of the sequence $(\mu_n^2)_{n \geq 1}$. Notice that, since $\mu_n^1 \to \mu^1$ and μ_n^2 satisfies

$$H_n(\mu_n^2) - \mu_n^2 = 1 - 2\mu_n^1,$$

the only possible limit points of the sequence $(\mu_n^2)_{n \geq 1}$ are μ^1 and μ^2. But $\mu_n^2 \leq \mu_n^* \to \mu^* < \mu^1$, so that in fact the only possible limit point is μ^2; the μ_n^2 are bounded and must therefore converge to μ^2.

We conclude the proof by showing that μ^2 is the only limit point of the sequence $(\mu_{n,m}^2)$. Notice that because we have assumed that $H'(\mu^1) > 1$, the equation $H(x) - x = 1 - 2\mu^1$ has two roots and, since μ_m^1, μ_n^1 both converge to μ^1 and H_n converges uniformly to H, it therefore follows that the equation $H_n(x)-x = 1-\mu_n^1-\mu_n^1$ also has two roots for sufficiently large m, n. Now, since $\mu_m^1, \mu_n^1 \to \mu^1$ as $\min(n,m) \to \infty$ and $\mu_{n,m}^2$ satisfies (5.5), the only possible limit points of the sequence $(\mu_{n,m}^2)_{m,n\geq 1}$ are μ^1 and μ^2. But $\mu^2 < \mu^* < \mu^1$ when $H'(\mu^1) > 1$ and $\mu_{n,m}^2 < \mu_n^*$ so that if $\mu_{n,m}^2$ converges, we have $\lim \mu_{n,m}^2 \leq \lim \mu_n^* = \mu^* < \mu^1$. Hence μ^2 is in fact the only limit point. □

Remark 7. *Notice that the method of the proof can be extended to prove that $\mu_n^k \to \mu^k$ for any k. For k odd, we consider the graphs of $H_n(x)+x, H(x)+x$ and for k even the graphs of $H_n(x)-x, H(x)-x$. For every $k \geq 2$ we can apply the argument involving limit points to deduce the desired convergence; for k odd this is particularly straightforward as there is only one limit point, corresponding to the equation $H(x)+x =$ constant having only one root.*

Proposition 11. *C_u^n converges to C_u in L^2 when $H'(\mu^1) > 1$.*

Proof. Let $m \geq n$. Define $E_{m,n} = \mathbb{E}[(C_u^m - C_u^n)^2]$. Expanding, we obtain

$$E_{m,n} = \mu_m^2 + \mu_n^2 - 2r_{m,n},$$

where $r_{m,n} = \mathbb{E}[C_u^m C_u^n]$. On the other hand, by applying the RDE (3.2) once, we obtain

$$E_{m,n} = \mathbb{E}[(\prod_{i=1}^{N_u^n} C_{ui}^n - \prod_{i=1}^{N_u^m} C_{ui}^m)^2]$$

$$= H_m(\mu_m^2) + H_n(\mu_n^2) - 2\mathbb{E}[\prod_{i=1}^{N_u^m} C_{ui}^m \prod_{i=1}^{N_u^n} C_{ui}^n].$$

Since $m \geq n$ we can bound $E_{m,n}$ above and below by truncating after m and n terms respectively:

$$H_m(\mu_m^2) + H_n(\mu_n^2) - 2H_m(r_{m,n}) \leq E_{m,n} \leq H_m(\mu_m^2) + H_n(\mu_n^2) - 2H_n(r_{m,n}).$$

Using the upper bound we have

$$2H_n(r_{m,n}) \leq H_m(\mu_m^2) + H_n(\mu_n^2) - E_{m,n} = H_m(\mu_m^2) + H_n(\mu_n^2) - \mu_m^2 - \mu_n^2 + 2r_{m,n}.$$

The moment equation (5.3) tells us that $H_m(\mu_m^2) - \mu_m^2 = 1 - 2\mu_m^1$ and that $H_n(\mu_n^2) - \mu_n^2 = 1 - 2\mu_n^1$. Hence

$$2H_n(r_{m,n}) \leq 1 - 2\mu_m^1 + \mu_m^2 + 1 - 2\mu_n^1 + \mu_n^2 - \mu_m^2 - \mu_n^2 + 2r_{m,n}$$

so that, on simplifying,

$$H_n(r_{m,n}) - r_{m,n} \leq 1 - \mu_m^1 - \mu_n^1.$$

Recall that, in the case $H'(\mu^1) > 1$, the equation $H_n(x) - x = 1 - \mu_m^1 - \mu_n^1$ has two roots, the lesser of which we denoted $\mu_{m,n}^2$. Let $\mu_{m,n}^1$ be the other (larger) root. Then $\mu_{n,m}^2 \leq r_{m,n} \leq \mu_{n,m}^1$ for all m, n and hence $\liminf_{m\to\infty} r_{m,n} \geq \mu^2$ since $\mu_{n,m}^2 \to \mu^2$ by the previous lemma.

On the other hand, Holder's inequality tells us that $r_{m,n} \leq \sqrt{\mu_m^2 \mu_n^2}$ and so it follows that $\limsup_{m\to\infty} r_{m,n} \leq \mu^2$ since $\mu_m^2, \mu_n^2 \to \mu^2$ by the previous lemma. Hence $r_{m,n} \to \mu^2$ as $n \to \infty$ and

$$E_{m,n} \to \lim_{m,n\to\infty} \mu_m^2 + \mu_n^2 - 2r_{m,n} = \mu^2 + \mu^2 - 2\mu^2 = 0,$$

showing that (C_u^n) is Cauchy in L^2. It now follows, by the completeness of L^2, that C_u^n converges. Since C_u^n is $\sigma(N)$-measurable, the limit L of the C_u^n must also be $\sigma(N)$-

measurable. To verify that this is the conditional probability solution, notice that

$$\mathbf{1}(E)C^n_\emptyset = (1 - \prod_{i=1}^{N^n_\emptyset} C^n_i)\mathbf{1}(E)$$

$$= (1 - \prod_{i=1}^{N_\emptyset} C^n_i)\mathbf{1}(E),$$

where $E = \{N_\emptyset \leq n\}$. As $n \to \infty$, the probability of E tends to 1; furthermore, since the C^n_i converge in L^2, they do so in probability. There exists, therefore, a subsequence which converges almost surely so that, in the limit,

$$L_\emptyset = 1 - \prod_{i=1}^{N_\emptyset} L_i \quad a.s.$$

Thus L is an endogenous solution to the RDE. It follows that L must be the conditional probability solution C. □

Theorem 22. *Consider the RDE*

$$X_u = 1 - \prod_{i=1}^{N^n_u} X_{ui}. \tag{5.6}$$

Then, by Lemma 24, there exists an invariant probability measure on $\{0, 1\}$ *for (5.6). Let* μ^1_n *denote the probability of a 1 under this invariant measure. Then the corresponding tree-indexed solution is endogenous if and only if* $H'_n(\mu^1_n) \leq 1$.

Proof. Recall Theorem 16 from Chapter 4 and the associated setup. Let $N^* = ess \sup N^n < \infty$ be a deterministic bound for N^n. We can then think of the random tree with branching factor N^n as being embedded in an N^*-ary tree. Each vertex has N^* daughter vertices, with a uniformly chosen random subset of size N^n being "alive" (the remaining being "dead"). In this context our RDE reads

$$X = 1 - \prod_{live\ u} X_u.$$

We now need to compute the transition probabilities required for Theorem 16. Consider first the transition from $(0,1)$ to $(1,0)$. The first coordinate automatically maps to 1 and the second maps to 0 provided all of the inputs not on the distinguished line of descent are equal to 1. The conditional probability of the vertex on the distinguished line of descent being alive is N^n/N^* since there are N^* vertices, of which N^n are alive. The probability of the remaining $N^n - 1$ vertices each taking value 1 is $(\mu_n^1)^{N^n-1}$ and so the probability of moving from $(0,1)$ to $(1,0)$, conditional on N^n, is just

$$\mathbf{1}(N^n \geq 1)\frac{(\mu_n^1)^{N^n-1}N^n}{N^*}.$$

Taking expectation, the required probability is

$$\mathbb{E}[\mathbf{1}(N^n \geq 1)\frac{(\mu_n^1)^{N^n-1}N^n}{N^*}] = \frac{\mathbb{E}[\mathbf{1}(N^n \geq 1)N^n(\mu_n^1)^{N^n-1}]}{N^*} = \frac{H_n'(\mu_n^1)}{N^*}.$$

The probability of moving from $(1,0)$ to $(0,1)$ is the same as that from $(0,1)$ to $(1,0)$ by symmetry. Hence P^- is given by

$$P^- = \begin{pmatrix} 0 & \frac{H_n'(\mu_n^1)}{N^*} \\ \frac{H_n'(\mu_n^1)}{N^*} & 0 \end{pmatrix},$$

whose largest eigenvalue ρ is $\frac{H_n'(\mu_n^1)}{N^*}$. By Theorem 16, the criterion for endogeny is $N^*\rho \leq 1$, i.e. $H_n'(\mu_n^1) \leq 1$, provided that, in the critical case $H_n'(\mu_n^1) = 1$, we verify the stated non-degeneracy conditions.

It is easily seen that P^- is irreducible. For the other criterion, let $X \in \mathcal{H}_0 \cap \mathcal{K}^\perp$ so that $X = f(X_\emptyset)$ for some L^2 function f and $\mathbb{E}[XY] = 0$ for all $Y \in \mathcal{K}$. Taking $Y = 1$, we obtain $\mathbb{E}[X] = 0$. Writing X in the form

$$X = a\mathbf{1}(X_\emptyset = 1) + b\mathbf{1}(X_\emptyset = 0),$$

where a, b are constants, we obtain

$$X = a\mathbf{1}(X_\emptyset = 1) - \frac{a\mu_n^1}{1-\mu_n^1}\mathbf{1}(X_\emptyset = 0).$$

For convenience we will scale by taking $a = 1$ (we assume that $X \neq 0$):

$$X = \mathbf{1}(X_\emptyset = 1) - \frac{\mu_n^1}{1-\mu_n^1}\mathbf{1}(X_\emptyset = 0).$$

Now, take $Y = \mathbf{1}(N_\emptyset = 1) \in \mathcal{K}$. Then

$$\mathbb{E}[XY] = \mathbb{E}[\mathbf{1}(N_\emptyset = 1)\{\mathbf{1}(X_\emptyset = 1) - \frac{\mu_n^1}{1-\mu_n^1}\mathbf{1}(X_\emptyset = 0)\}]$$

$$= \mathbb{E}[\mathbf{1}(N_\emptyset = 1)\{\mathbf{1}(X_1 = 0) - \frac{\mu_1^n}{1-\mu_n^1}\mathbf{1}(X_1 = 1)\}]$$

$$= \mathbb{P}(N = 1)[1 - \mu_n^1 - \frac{(\mu_n^1)^2}{1-\mu_n^1}]$$

$$= \frac{1-2\mu_n^1}{1-\mu_n^1}\mathbb{P}(N = 1).$$

Now, by the strict convexity of H_n, we have

$$H_n(1/2) < \frac{H_n(0) + H_n(1)}{2} = 1/2$$

so that $H_n(1/2) + 1/2 < 1$ and therefore $\mu_n^1 > 1/2$. Hence $1 - 2\mu_n^1 < 0$ and $\mathbb{E}[XY] < 0$.

This contradicts the assumption that $X \in \mathcal{H}_0 \cap \mathcal{K}^{\perp}$. □

Proof. (of remainder of Theorem 21) We begin by proving part (1) of the theorem. We have already proved that if $H'(\mu^1) \leq 1$ then the distribution of S coincides with that of C. Since C is endogenous it follows that S is also endogenous. For the reverse implication, we prove that $H'(\mu^1) > 1$ implies S is not endogenous.

By Theorem 22 we know that the RDE (5.6) has two invariant distributions if and only if $H_n'(\mu_n^1) > 1$. Furthermore we know that C_u^n converges to C_u in L^2 when $H'(\mu^1) > 1$ and hence $\mu_n^2 \to \mu^2 \neq \mu^1$ so that S and C have different second moments. It now follows that S is not endogenous since if S does not have the same distribution as C it cannot be endogenous.

It remains to prove parts (2) and (3) of Theorem 21 in the case $H'(\mu^1) > 1$. We argue that any solution X that doesn't have the singular distribution on $\{0, 1\}$ must be the conditional probabilities solution. Our strategy is to show (via the moment equation (5.3)) that the moments of X are equal to those of C. We have already shown that the first moment of any stationary distribution for the RDE (5.1) is μ^1. For the second moment, we consider the moment equation (5.3) with $n = 2$. The equation $H(x) - x = 1 - 2m_1$ has two possible roots. Since S is not endogenous (by part (1) of the theorem), the second moment of C cannot be μ^1. Hence by assumption the second moment of X must be equal to the second moment of C. The third moment of X is a root of

$$H(x) + x = 1 - 3\mu^1 + 3m_2^C.$$

Again, this has at most one root and this has to be the third moment of C. Similarly, the fourth moment is a root of

$$H(x) - x = 1 - 4\mu^1 + 6m_2^C - 4m_3^C.$$

Notice that because $H'(\mu^1) > 1$ we must have μ^1 to the right of m_*, the minimum of the function $H(x) - x$. Since the equation $H(x) - x = 1 - 2\mu_1$ has roots μ^1 and the second moment of C, with $H'(\mu^1) > 1$, it follows that the second moment of C is to the left of m_*. The fourth moment must also therefore be to the left of m_* and for this reason we can discount the greater of the two possible roots. The other root (the lesser of the two) must be the fourth moment of C. We can now extend these arguments to cover all odd and even n and conclude that the moments of X in the case $H'(\mu^1) > 1$ are equal to those of C. Since $[0, 1]$ is bounded, this sequence of moments determines a unique distribution which is therefore that of C. [15] □

5.4 Basins of attraction

In the preceding section we showed that the endogenous solution corresponding to a bounded approximation of N converges in distribution to the endogenous solution corresponding to N. Now we consider the "basin of attraction" of the endogenous solution. That is, we ask for what initial distributions does the corresponding "solution at root" $X_\emptyset$ converge (in some sense) to the endogenous solution.

Definition 3. *Let ς be the law of the endogenous solution. Suppose that we insert independent, identically distributed random variables with law ν at level n of the tree and apply the RDE to obtain the corresponding solution $X_u^n(\nu)$ (with law $T^{n-|u|}(\nu)$) at vertex u. Then we define the basin of attraction B of the endogenous solution to be*

$$B = \{\nu \in \mathbb{D} : T^n(\nu) \to^{weak*} \varsigma\},$$

which is, of course, equivalent to the set of distributions ν for which $X_u^n(\nu)$ converges in law to the endogenous solution C_u.

5.4.1 The unstable case $H'(\mu^1) > 1$

Lemma 27. *Suppose that ν has mean μ^1 and that $H'(\mu^1) > 1$. Then*

1. *$\nu \in B$;*

2. *$X_u^n(\nu) \to^{L^2} C_u$, the endogenous solution, for any ν satisfying (1) other than the singular measure on $\{0,1\}$.*

Proof. Let $E_k = \mathbb{E}[X_u^n(\nu)^2]$, where $k = n - |u|$, and let $r_k = \mathbb{E}[C_u X_u^n(\nu)]$. Then

$$\mathbb{E}[(X_u^n(\nu) - C_u)^2] = E_k - 2r_k + \mu^2.$$

Now,

$$E_k = \mathbb{E}[(1 - 2\prod_{i=1}^{N_u} X_{ui}^n(\nu) + \prod_{i=1}^{N_u} X_{ui}^n(\nu)^2)]$$
$$= 1 - 2H(\mu^1) + H(E_{k-1}).$$

This is a recursion for E_k with at most two fixed points (recall that the equation $H(x)-x =$ *const.* has at most two roots). Recalling the moment equation (5.3), these are easily seen to be μ^1, μ^2, the first and second moments of the endogenous solution. We have assumed that ν is not the singular distribution and so its second moment (i.e. E_0) must be strictly less that μ^1. Now, under the assumption that ν is not the singular solution, μ^1, μ^2 lie either side of the minimum μ^* of $H(x)-x = 2H(\mu^1)-1$ and $H'(\mu^*) = 1$ so that $H'(\mu^2) < 1$. Hence μ^2 is stable and it now follows that E_k converges to μ^2.

The recursion for r_k is essentially the same as that for E_k:

$$\mu^2 - r_k = H(\mu^2) - H(r_{k-1}).$$

This has μ^2 as a fixed point and, since

$$r_0 = \mathbb{E}[C_u X_u(\nu)] \leq \sqrt{\mathbb{E}[C_u^2]\mathbb{E}[X_u(\nu)^2]} < \sqrt{\mu^1 \mu^1} = \mu^1,$$

we are in the same situation as with E_k. That is, we start to the left of μ^1 and, because $H'(\mu^2) < 1$, we conclude that μ^2 is stable (i.e. the other potential fixed point of the recursion for r_k isn't) and it now follows that r_k converges to μ^2 under the assumptions of the lemma. Hence

$$\mathbb{E}[(X_u^n(\nu) - C_u)^2] = E_k - 2r_k + \mu^2 \to 0.$$

□

Theorem 23. *Let $\delta(m)$ denote the singular distribution on $\{0,1\}$ with mean m. Then*

$$B = \{\nu \in \mathbb{D} : \int x d\nu(x) = \mu^1 \textit{ and } \nu \neq \delta(\mu^1)\}.$$

That is, B is precisely the set of distributions on $[0,1]$ with the correct mean (except the singular distribution with mean μ^1).

Proof. We have already shown that

$$\{\nu \in \mathbb{D} : \int x d\nu(x) = \mu^1 \textit{ and } \nu \neq \delta(\mu^1)\} \subseteq B.$$

By the definition of B, $\nu \in B$ if for any bounded continuous function f on $[0, 1]$,

$$\int f dT^n \nu \to \int f d\mu,$$

and now, since the identity is bounded on $[0, 1]$, we conclude that

$$\mathbb{E} X_u^n(\nu) \to \mathbb{E} C_u,$$

so that to $\nu \in B$ only if the mean of $T^n(\nu)$ converges to μ^1. From the moment equation (5.3), the mean of $X_u^n(\nu)$ is obtained by iterating

$$t \mapsto 1 - H(t)$$

n times, starting with the mean of ν. This mapping has unique fixed point μ^1 and, since $H'(\mu^1) > 1$, it is not attracting. It follows that the only way we can have convergence in mean is if we start with the correct mean, that is, if ν has mean μ^1. Hence

$$B \subseteq \{\nu \in \mathbb{D} : \int x d\nu(x) = \mu^1 \quad \textit{and } \nu \neq \delta(\mu^1)\}.$$

$\square$

5.4.2 The stable case $H'(\mu^1) \leq 1$

Proposition 12. *Let $B(\mu^1)$ be the basin of attraction of μ^1 under the iterative map for the first moment, $t \mapsto 1 - H(t)$. Then*

$$B = \{\nu \in \mathbb{D} : \int x d\nu(x) \in B(\mu^1)\}.$$

Consider once again $\mathbb{E}[(X_u^n(\nu) - C_u)^2]$. Let $m_k^\theta = \mathbb{E}X_u^n(\nu)^\theta$, where $k = n - |u|$. Then

$$m_k^2 = \mathbb{E}(1 - 2\prod_{i=1}^{N_u} X_{ui}^n(\nu) + \prod_{i=1}^{N_u} X_{ui}^n(\nu)^2)$$

$$= 1 - 2H(m_{k-1}^1) + H(m_{k-1}^2).$$

Recalling that $r_k = \mathbb{E}[C_u X_u^n(\nu)]$, we have

$$r_k = \mathbb{E}[(1 - \prod_{i=1}^{N_u} C_{ui})(1 - \prod_{i=1}^{N_u} X_{ui}^n(\nu))]$$

$$= \mathbb{E}[(1 - \prod_{i=1}^{N_u} C_{ui} - \prod_{i=1}^{N_u} X_{ui}^n(\nu) + \prod C_{ui} X_{ui}^n(\nu))]$$

$$= 1 - H(\mu^1) - H(m_{k-1}^1) + H(r_{k-1}).$$

We now turn our attention to analysing the dynamics of m_k^2 and r_k. We will concentrate on the equation for m_k^2 as the equation for r_k is essentially the same. By assumption, m_k^1 converges to μ^1 and so we may approximate m_k^1, for $k \geq k_\epsilon$ (say), by $\mu^1 \pm \epsilon$, for some small $\epsilon > 0$.

Lemma 28. *The dynamical system l_k defined by the recursion*

$$l_k = 1 - 2H(\mu^1 + \epsilon) + H(l_{k-1}), \qquad l_{k_\epsilon} = m_{k_\epsilon}^2,$$

is a lower bound for m_k^2 for all $k \geq k_\epsilon$, where k_ϵ is a positive integer such that

$$|m_k^1 - \mu^1| < \epsilon, \qquad k \geq k_\epsilon.$$

Proof. We have $l_{k_\epsilon} \leq m_{k_\epsilon}^2$. Suppose that the statement is true for some $K \geq k_\epsilon$, that is, suppose that $l_K \le m_K^2$. We have, since H is increasing, that

$$1 - 2H(\mu^1 + \epsilon) < 1 - 2H(m_K^1),$$

and that $H(l_K) \le H(m_K^2)$, so that $l_{K+1} \le m_{K+1}^2$. By induction, this holds for any $k \ge k_\epsilon$. □

Lemma 29. *Let f be a continuous, strictly increasing function on an interval $[a,b] \subseteq [0,1]$ with $f(a) > a, f(b) > b$ and such that there exists $y \in (a,b)$ with $f(y) < y$. Suppose that f has fixed points $p, q \in (a,b)$ with $p \neq q$. Let $g(\delta)$ be a positive, increasing continuous function in $\delta > 0$ with the property that $g(\delta)$ converges to zero as δ tends to zero. Then, for sufficiently small δ, the map $f - g$ has fixed points $p_\delta, q_\delta \in (a,b)$ and p_δ, q_δ converge to p, q respectively as δ tends to zero.*

Proof. We have then three distinguished points a, y, b such that

$$f(a) > a, f(y) < y, f(b) > b.$$

Since g is continuous, we can make $f(a) - g(\delta), f(y) - g(\delta), f(b) - g(\delta)$ arbitrarily close to $f(a), f(y), f(b)$ by choosing δ to be sufficiently small. Hence, for sufficiently small δ, we have

$$f(a) - g(\delta) > a, f(y) - g(\delta) < y, f(b) - g(\delta) > b,$$

and therefore the map $f - g(\delta)$ must still have two fixed points, one "close" to p, which we denote p_δ and the other "close" to q, which we denote q_δ. Now, it is easily seen that, as δ tends to zero, the sequence p_δ is increasing, bounded above by p so that it has a limit. This limit is a fixed point of f (since g converges to zero) and therefore must be p. Similarly, as δ tends to zero, the sequence q_δ is decreasing, bounded below by q and its limit is also a fixed point which has to be q. □

Lemma 30. *Let*

$$f_\epsilon(x) = 1 - 2H(\mu^1 + \epsilon) + H(x), \qquad x \in [0,1].$$

Then, for sufficiently small $\epsilon > 0$, f_ϵ *has two fixed points.*

Proof. By the strict convexity of H we have

$$H(\frac{x+y}{2}) < \frac{H(x) + H(y)}{2}. \qquad x, y \in [0,1]$$

so that

$$H(\frac{1}{2}) < \frac{H(0) + H(1)}{2} = \frac{1}{2}$$

and hence $H(1/2) + 1/2 < 1$. It now follows that $\mu^1 > 1/2$ so that $H(\mu^1) = 1 - \mu^1 < 1/2$. This means that $f_0(0) = 1 - 2H(\mu^1) > 0$ and $f_0(1) = 2(1 - H(\mu^1)) = 2\mu^1 > 1$. Hence f_0 must have two fixed points, one of which we know is μ^1. Since $H'(\mu^1) \leq 1$, this other point must be greater that μ^1. We can now apply the previous claim to the fixed points. In the critical case $H'(\mu^1) = 1$, the graph of f_0 "touches" the identity at μ^1. It follows, since f_ϵ is f_0 moved downwards by an amount depending on ϵ, that f_ϵ will intersect the identity in two places, one either side of μ^1. □

The above claim tells us that f_ϵ has a fixed point close to μ^1, which we will denote $\mu^1(\epsilon)$ and another one close to the other fixed point p of f_0, which we will denote $p(\epsilon)$. Of course in the critical case $H'(\mu^1) = 1$, both $\mu^1(\epsilon)$ and $p(\epsilon)$ are "close" to μ^1.

Lemma 31. l_k *converges to* $\mu^1(\epsilon)$.

Proof. We have $l_k = f_\epsilon^{k-k_\epsilon}(l_{k_\epsilon})$ and so we need only verify that l_{k_ϵ} is in the basin of attraction of $\mu^1(\epsilon)$ and that $\mu^1(\epsilon)$ is stable. We know that

$$f_\epsilon(\mu^1 + \epsilon) < \mu^1 + \epsilon$$

since $1-H(\mu^1+\epsilon) < 1-H(\mu^1) = \mu^1$ and so it must be the case that $\mu^1+\epsilon \in (\mu^1(\epsilon), p(\epsilon))$. It now follows that $l_{k_\epsilon} < p(\epsilon)$ since $l_{k_\epsilon} \leq m^1_{k_\epsilon} < \mu^1+\epsilon$. In the strictly stable case $H'(\mu^1) < 1$, the stability of $\mu^1(\epsilon)$ follows from the fact that $\mu^1(\epsilon)$ converges to μ^1 as ϵ tends to zero (by the previous lemma) and therefore $\mu^1(\epsilon)$ can be made arbitrarily close to μ^1 by choosing ϵ to be sufficiently small. This means that for sufficiently small ϵ, $H'(\mu^1(\epsilon)) < 1$ by the continuity of H'. In the critical case $H'(\mu^1) = 1$, we have $\mu^1(\epsilon) < \mu^1 < p(\epsilon)$, so that $H'(\mu^1(\epsilon)) < 1$. In either case it now follows that $f_\epsilon^{k-k_\epsilon}(l_{k_\epsilon})$ converges to $\mu^1(\epsilon)$. □

Proof. (of Proposition 13) The preceding lemmas tell us that

$$\liminf_{k\to\infty} m^2_k \geq \lim_{k\to\infty} l_k = \mu^1(\epsilon).$$

Letting ϵ tend to zero, we obtain

$$\liminf_{k\to\infty} m^2_k \geq \mu^1.$$

The fact that $m^2_k \leq m^1_k$ for every k gives us the corresponding inequality for the lim sup:

$$\limsup_{k\to\infty} m^2_k \leq \lim_{k\to\infty} m^1_k = \mu^1.$$

We conclude that m^2_k converges to μ^1.

Now,

$$\mathbb{E}[(X^n_u(\nu) - C_u)]^2 = m^2_k - 2r_k + \mu^2,$$

so that $\mathbb{E}[(X^n_u(\nu) - C_u)^2] \to 0$, remembering that in the stable case the singular solution and endogenous solution coincide (i.e. $\mu^1 = \mu^2$). We have now shown that

$$\{\nu \in \mathbb{D} : \int x d\nu(x) \in B(\mu^1)\} \subseteq B,$$

and the necessity for convergence in mean ensures that we have the reverse inclusion. This completes the proof. □

5.5 Outside the basin of attraction

In this section we examine what happens if we iterate distributions with mean outside the basin of attraction of the endogenous solution under the map for the mean $F : t \mapsto 1 - H(t)$.

Definition 4. *We say that a map f has an n-cycle p if $f^n(p) = p$, where f^n denotes the n-fold composition of f with itself.*

It is easily seen that F can have only one and two cycles. This is because the iterated map $F^2 : t \mapsto 1 - H(1 - H(t))$ is increasing in t and hence can have only one cycles; this corresponds to F having one or two cycles. Notice that the fixed points (or one cycles) of F^2 come in pairs: if p is a fixed point of F^2 then so too is $1 - H(p)$.

In what follows we distinguish two cases in terms of stability. In the stable case we show that the basin of attraction of the endogenous solution under T^2 is those distributions with the correct mean. In the unstable case we show that the basin of attraction is those distributions having mean in the basin of attraction of a fixed point of F^2. In both cases this is entirely analogous to the results we obtained earlier for the basin of attraction of the endogenous solution under T. Throughout what follows we will write

B^2 for the basin of attraction of the endogenous solution under T^2:

$$B^2 = \{\nu \in \mathbb{D} : T^{2n}(\nu) \to^{weak*} \varsigma\}.$$

We will also write μ_+^1 for a fixed point of F^2 and μ_-^1 for its "complimentary" fixed point (recall that they come in pairs):

$$\mu_+^1 = 1 - H(1 - H(\mu_+^1)), \quad \mu_-^1 = 1 - H(\mu_+^1).$$

5.5.1 The unstable case $H'(\mu_+^1)H'(\mu_-^1) > 1$

Lemma 32. *Suppose that $H'(\mu_+^1)H'(\mu_-^1) > 1$. Then*

$$B^2 = \{\nu \in \mathbb{D} : \int x d\nu(x) = \mu_+^1 \;\; \text{and} \;\; \nu \neq \delta(\mu_+^1)\}.$$

Proof. We have

$$m_{2k}^2 = 1 - 2H(1 - H(\mu_+^1)) + H(1 - 2H(\mu_+^1) + H(m_{2(k-1)}^2)).$$

The map M given by

$$t \mapsto 1 - 2H(1 - H(\mu_+^1)) + H(1 - 2H(\mu_+^1) + H(t))$$

has fixed point μ_+^1 and, since $H'(\mu_+^1)H'(\mu_-^1) > 1$, it must have another fixed point p since, by the strict convexity of H, if μ_+^1 were the only fixed point then M would be tangential to the identity at μ_+^1 (i.e. we would have $H'(\mu_+^1)H'(\mu_-^1) = 1$). Now, since ν is assumed not to be the singular distribution, its second moment m_0^2 must be strictly less than μ_+^1. Now, H is strictly increasing and so it follows that $H'(p) < 1$ and hence m_{2k}^2 converges to $p < \mu_+^1$ under the stated assumption. The recursion for r_k is the same as that for m_k^2 (see previous section) and so we can deduce that r_{2k} also converges to p. Hence

$$\mathbb{E}[(X_u^{2n}(\nu) - C_u)^2] = m_{2k}^2 - 2r_{2k} + p \to 0.$$

Recall from the proof of Theorem 23 that the necessity for convergence in mean gives the reverse inclusion. □

5.5.2 The stable case $H'(\mu_+^1)H'(\mu_-^1) \leq 1$

Proposition 13. *Suppose that m_{2k}^1 converges to μ_+^1. Then m_{2k}^2 also converges to μ_+^1.*

The recursion for m_{2k}^2 is

$$m_{2k}^2 = 1 - 2H(1 - H(m_{2(k-1)}^1)) + H(1 - 2H(m_{2(k-1)}^1) + H(m_{2(k-1)}^2)).$$

We follow the strategy we used earlier of establishing a lower bound and arguing that this lower bound converges to μ_+^1.

Lemma 33. *Define a sequence L_k by*

$$L_k = 1 - 2H(1 - H(\mu_+^1 - \epsilon)) + H(1 - 2H(\mu_+^1) + H(L_{k-1})), \qquad L_{k_\epsilon} = m_{2k_\epsilon}^2,$$

where k_ϵ is a positive integer such that

$$|m_{2k}^1 - \mu_+^1| < \epsilon, \qquad k \geq k_\epsilon.$$

Then L_k bounds m_{2k}^2 below for all $k \geq k_\epsilon$.

Proof. Define a map T_c by

$$T_c(x) = 1 - 2H(1-x) + H(c-2x).$$

Then we claim that T_c is increasing for $x \geq c-1$. This is easily seen by differentiating with respect to x and using the convexity of H. We now prove that m^2_{2k} is bounded below by l_k for all $k \geq k_\epsilon$, where l_k is given by

$$l_k = 1 - 2H(1 - H(\mu^1_+ - \epsilon)) + H(1 - 2H(\mu^1_+ - \epsilon) + H(l_{k-1})), \qquad l_{k_\epsilon} = m^2_{2k_\epsilon}.$$

Now, we have $m^2_{2k_\epsilon} \geq l_{k_\epsilon}$. Suppose that the statement is true for some arbitrary $K \geq k_\epsilon$, that is, $m^2_{2K} \geq l_K$. Since $H(m^2_{2K}) \geq H(l_K)$, we can use the result above for the map T_c with $x = H(m^1_{2K})$ and $c = 1 + H(m^2_{2K})$ (the condition "$x \geq c - 1$" is satisfied because $H(m^1_{2K}) \geq H(m^2_{2K})$), to conclude that $m^2_{2(K+1)} \geq l_{K+1}$. Hence $m^2_{2k} \geq l_k$ for any $k \geq k_\epsilon$ by induction. To complete the proof we now show that $l_k \geq L_k$ for every $k \geq k_\epsilon$. We have $l_{k_\epsilon} \geq L_{k_\epsilon}$. If $l_K \geq L_K$ for some arbitrary $K \geq k_\epsilon$, then it holds that $l_{K+1} \geq L_{K+1}$ since H is increasing. Again, by induction, the result holds for any $k \geq k_\epsilon$. □

Proof. (of Proposition 13) The argument is now essentially the same as it was for the un-iterated map in the stable case. Define a map $g_\epsilon : [0, \widehat{x}] \to \mathbb{R}$ by

$$g_\epsilon(x) = 1 - 2H(1 - H(\mu^1_+ - \epsilon)) + H(1 - 2H(\mu^1_+) + H(x)),$$

where $\widehat{x}$ is the unique point in $[0,1]$ satisfying the equation $H(x) = 2H(\mu^1_+)$. The reason for defining g_ϵ on this interval is to ensure that is well-defined. Now, we know that g_0 has fixed point μ^1_+ and, as before, we have $g_0(0) > 0$, $g_0(\widehat{x}) = 2\mu^1_+ > \widehat{x}$. It follows that g_0 has two fixed points, with the gradient of g_0 at these fixed points either greater than or less than 1 (as before if the gradient were 1 then g_0 would be tangential to the identity and there would be only one fixed point). If we have stability in that $H'(\mu_+^1)H'(\mu_-^1) < 1$, then for the other fixed point p we must have $g_0'(p) > 1$ so that $p > \mu^1$. Since g_ϵ is obtained from g_0 by moving down the graph of g_0 by a constant depending on ϵ, it now follows from Lemma 29 that g_ϵ will always have two fixed points (in the critical case g_0 touches the identity and so g_ϵ will intersect the identity either side of μ_+^1; we will denote these points $\mu_+^1(\epsilon)$ and $p(\epsilon)$ respectively). Again, by Lemma 29, it is easily seen that $\mu_+^1(\epsilon)$ converges to μ_+^1 as ϵ tends to zero.

Now, we have $L_k = g_\epsilon(L_{k-1}) = g_\epsilon^{k-k_\epsilon}(L_{k_\epsilon})$, with $L_{k_\epsilon} = m_{2k_\epsilon}^2 < \mu_+^1 + \epsilon$, where g_ϵ^n is the n-fold composition of g_ϵ with itself. Suppose now that μ_+^1 is strictly stable so that $g_0'(\mu_+^1) = H'(\mu_+^1)H'(1 - H(\mu_+^1)) < 1$ and hence, for sufficiently small ϵ (remember $\mu_+^1(\epsilon)$ converges to μ_+^1 and so we can make the two arbitrarily close by choosing ϵ to be sufficiently small), $g_\epsilon'(\mu^1(\epsilon)) = H'(\mu_+^1(\epsilon))H'(1 - H(\mu_+^1(\epsilon))) < 1$ by the continuity of H', so that $\mu_+^1(\epsilon)$ is stable for sufficiently small ϵ. This means that $\mu_+^1(\epsilon)$ has a (non-trivial) basin of attraction. Since $\mu_+^1(\epsilon)$ and $\mu_+^1 + \epsilon$ both converge to μ_+^1 as ϵ tends to 0, we can make them arbitrarily close by choosing ϵ to be sufficiently small. It follows then that $\mu_+^1 + \epsilon$ is in the basin of attraction of $\mu_+^1(\epsilon)$ for sufficiently small ϵ. Since $L_{k_\epsilon} < \mu_+^1 + \epsilon$, it also follows that L_{k_ϵ} is also in the basin of attraction of $\mu_+^1(\epsilon)$ for sufficiently small ϵ. We conclude that L_k converges to $\mu_+^1(\epsilon)$ in the strictly stable case. In the so-called "critical case" $H'(\mu_+^1)H'(\mu_-^1) = 1$, since g_0 is tangent to the identity at μ_+^1, it follows that $m_{2k}^1 \leq \mu_+^1$ for all k and hence $L_{k_\epsilon} \leq \mu_+^1 < p(\epsilon)$. By virtue of the fact that $\mu_+^1(\epsilon) < \mu_+^1$, we have $H'(\mu_+^1(\epsilon)) < 1$ and it now follows that L_k converges to $\mu_+^1(\epsilon)$. To finish off the proof, we establish the relevant inequality for the lim inf:

$$\liminf_{k\to\infty} m_{2k}^2 \geq \lim_{k\to\infty} L_k = \mu_+^1(\epsilon).$$

Letting ϵ tend to 0, we obtain

$$\liminf_{k\to\infty} m_{2k}^2 \geq \lim_{\epsilon\to 0} \mu_+^1(\epsilon) = \mu_+^1.$$

The corresponding inequality for the lim sup follows from the fact that $m_{2k}^2 \leq m_{2k}^1$ for every k. We have proved that if m_{2k}^1 converges to μ_+^1 then so too does m_{2k}^2. □

5.6 Further work

In this chapter we have made a thorough study of the RDE (5.1) under the assumption that the random variable N is almost surely positive. The main reason for making this assumption was that it made much of the analysis involved in the proofs more straightforward than would otherwise have been the case. It would be interesting, however, to work through the proofs without this assumption to see whether it really is necessary or merely convenient. In relation to the original noisy veto voter model from which this RDE was derived, it would be interesting to extend the problem from the binary tree to a deterministic tree with different branching factor or even to a random tree.

Bibliography

[1] Aldous D.J. and Bandyopadhyay A. (2005) "A survey of max-type recursive distributional equations", The Annals of Applied Probability, volume 15, number 2, 1047-1110.

[2] Athreya K.B. (1971) "A note on a functional equation arising in Galton-Watson branching processes", Journal of Applied Probability 8, 589-598.

[3] Athreya K.B. and Ney P.E. (1972) "Branching Processes", Springer, Berlin.

[4] Bandyopadhyay A. (2002) "Bivariate uniqueness in the logistic recursive distributional equation", Technical report 629, Department of Statistics, University of California, Berkeley.

[5] Biggins, J.D. (1977) "Martingale convergence in the branching random walk", Journal of Applied Probability, 14, 25-37.

[6] Biggins J.D. (1977) "Chernoff's theorem in the branching random walk", Journal of Applied Probability, 14, 630-636.

[7] Biggins J.D. and Kyprianou A.E. (2004) "Measure change in multi-type branching", Advances in Applied Probability, 36 (2), 544-581.

[8] Bramson, M. (1978) "Maximal displacement of branching brownian motion", Comm. Pure Appl. Math. 31, no. 5, 531-581.

[9] Bramson M., Ney P. and Tao J. (1992) "The population composition of a multitype branching random walk", The Annals of Applied Probability, volume 2, number 3, 575-596.

[10] Burd G.A., Waymire E.C., (2000) "Independent random cascades on Galton-Watson trees", Proceedings of the American Mathematical Society, Volume 128, Number 9, 2753-2761.

[11] Caliebe A. (2003) "Symmetric fixed points of a smoothing transformation", Advances in Applied Probability 35, 377-394.

[12] Collet P. and Koukiou F. (1992) "Large deviations for multiplicative chaos", Commun. Math. Phys. 147, 329-342.

[13] Durrett R. and Liggett T.M. (1983), "Fixed points of the smoothing transformation", Z. Wahrscheinlichkeitstheorie verw. Gebiete 64, 275-301.

[14] Evans S. N. (1992) "Polar and non-polar sets for a tree-indexed process", Annals of Probability, volume 20, issue 2, 579-590.

[15] Feller W. (1971) "Introduction to probability theory and its applications", volume 2, Wiley.

[16] Gantert, N. (2000) "The maximum of a branching random walk with semi-exponential increments", The Annals of Probability, volume 28, number 3, 1219-1229.

[17] Guivarc'h, Y. "Sur une extension de la notion de loi semi-stable" (1990), Annales de l'I.H.P., section B, tombe 26, 261-285.

[18] Hammersley J. M. (1974) "Postulates for subadditive processes", The Annals of Probability, volume 2, number 4, 652-680.

[19] Hardy R. and Harris S.C. (2004) "A new formulation of the spine approach to branching diffusions", Mathematics Preprint, University of Bath. 0404.

[20] Hardy R. and Harris S.C. (2004) "Spine proofs for L^p-convergence of branching-diffusion martingales, Mathematics Preprint, University of Bath. 0405.

[21] Hardy R. and Harris S.C. (2004) "A spine proof of a lower bound for a typed branching diffusion", Mathematics Preprint, University of Bath. 0408.

[22] Hardy R. and Harris S.C. (2004) "A spine proof of a large-deviations principle for branching brownian motion", S.P.A. (accepted for publication).

[23] Hilderbrand M. (2005) "On a question of Chung, Diaconis and Graham", Mathematics, abstract math.PR/0508427.

[24] den Hollander, F. (1991) "Large deviations", Mathematics subject classification 60-01, 60F10, 60K35.

[25] Hsing T., Husler J. and Reiss R.-D. (1996), "The extremes of a triangular array of normal random variables", The Annals of Applied Probability, volume 6, number 2, 671-686.

[26] Janson S. and Mossel E. (2004) "Robust reconsruction on trees is determined by the second eigenvalue", The Annals of Probability, volume 32, number 3B, 2630-2649.

[27] Jordan J. (2003) "Renormalisation of random hierarchical systems", Ph.D thesis, University of Oxford, UK.

[28] Kahane J.-P. and Peyiere J. "Sur certaines martingales de Benoit Mandelbrot" (1976), Advances in Mathematics 22, 131-145.

[29] Kesten H. and Stigum B.P. (1966) "A limit theorem for multidimensional Galton-Watson processes", Ann. Math. Statist. 37, 12111223.

[30] Kingman J.F.C. (1975), "The first birth problem for an age-dependent branching process", The Annals of Probability, volume 3, number 5, 790-801.

[31] Parthasarathy K.R. (1967) "Probability measures on metric spaces", New York Academic Press.

[32] Kyprianou A. E. and Sani R. (2001) "Martingale convergence and the functional equation in the multi-type branching random walk, Bernoulli 7(4), 593-604.

[33] Lalley S. P. and Selke T. (2000) "An extension of Hawkes's Theorem on the Hausdorff dimension of a Galton-Watson tree", Probab. Relat. Fields 116, 41-56.

[34] Le Gall, J.-F. (2006) "Random Real Trees", Annales de la facult des sciences de Toulouse Sr. 6, 15 (1), 35-62.

[35] Lyons R. and Pemantle R. (1992) "Random walk in a random environment and first-passage percolation on trees", Annals of Probability, 20, 125-136.

[36] Lyons R., Pemantle R. and Peres Y. (1995) "Conceptual proofs of $L \log L$ criteria for mean behaviour of braching processes", Annals of Applied Probability, 23, 1125-1138.

[37] Lyons, R. (1998) "A simple path to Biggins' martingale convergence for branching random walks", in Classical and Modern Branching Processes, IMA Vols. In Maths And Its Applications, 84, 217-222.

[38] Liu Q. (1998) "Fixed points of a genralised smoothing transformation and applications to the branching random walk", Advances in Applied Probability, 30, 85-112.

[39] Mossel E. (2001) "Reconstruction on trees: beating the second eigenvalue", The Annals of Applied Probability, 11 (1), 295-300.

[40] Neveu, J. (1987) "Multiplicative martingales for spatial branching processes", Seminar on Stochastic Processes, Progr. Probab. Statist. 15, 223-242.

[41] Pemantle R. (1994) "Sharpness of second moment criteria for branching and tree-indexed processes", Mathematics, abstract math.PR/0404090.

[42] Peres Y. (1997) "Probability on trees: an introductory climb", 193-280, Lecture Notes in Math, 1717, Springer, New York.

[43] Tsirelson B.S. and Vershik A.M. (1998) "Examples of nonlinear continuous tensor products of measure spaces and non-fock factorisations", Reviews in Mathematical Physics, volume 10, number 1, 81-145.

[44] Warren J. (2005) "Dynamics and endogeny for recursive processes on trees", Mathematics, abstract math.PR/0506038.

[45] Williams D. (2001) "Weighing the odds: a course in probability and statistics", Cambridge University Press.