Durham E-Theses

Limiting behaviour of random spatial graphs and asymptotically homogeneous RWRE

Andrew R. Wade

How to cite:

Andrew R. Wade (2005) Limiting behaviour of random spatial graphs and asymptotically homogeneous RWRE. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a https://etheses.durham.ac.uk/id/eprint/2776/ is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
https://etheses.durham.ac.uk

Limiting behaviour of random spatial graphs and asymptotically homogeneous RWRE

Andrew R. Wade

The copyright of this thesis rests with the author or the university to which it was submitted. No quotation from it, or information derived from it may be published without the prior written consent of the author or university, and any information derived from it should be acknowledged.

A Thesis presented for the degree of
Doctor of Philosophy

Statistics and Probability Group
Department of Mathematical Sciences
University of Durham
England

September 2005

0 5 MAY 2006

Dedicated to

My parents

and

Mike Bryant

Limiting behaviour of random spatial graphs and asymptotically homogeneous RWRE

Andrew R. Wade

Submitted for the degree of Doctor of Philosophy
September 2005

Abstract

We consider several random spatial graphs of the nearest-neighbour type, including the k-nearest neighbours graph, the on-line nearest-neighbour graph, and the minimal directed spanning tree. We study the large sample asymptotic behaviour of the total length of these graphs, with power-weighted edges. We give laws of large numbers and weak convergence results. We evaluate limiting constants explicitly.

In Bhatt and Roy's minimal directed spanning tree (MDST) construction on random points in $(0,1)^2$, each point is joined to its nearest neighbour in the south-westerly direction. We show that the limiting total length (with power-weighted egdes) of the edges joined to the origin converges in distribution to a Dickman-type random variable. We also study the length of the longest edge in the MDST.

For the total weight of the MDST, we give a weak convergence result. The limiting distribution is given a normal component plus a contribution due to boundary effects, which can be characterized by a fixed point equation. There is a phase transition in the limit law as the weight exponent increases.

In the second part of this thesis, we give criteria for ergodicity, transience and null recurrence for the random walk in random environment (RWRE) on $\mathbf{Z}^+ = \{0, 1, 2, \ldots\}$, with reflection at the origin, where the random environment is subject to a vanishing perturbation from the so-called Sinai's regime. Our results complement existing criteria for random walks in random environments and for Markov chains with asymptotically zero drift, and are significantly different to these previously studied cases. Our method is based on a martingale technique — the method of Lyapunov functions.

Declaration

The work in this thesis is based on research carried out in the Department of Mathematical Sciences, at the University of Durham, England. No part of this thesis has been submitted elsewhere for any other degree or qualification. It is all my own work unless referenced to the contrary in the text.

Parts of Chapters 2 to 6 are adapted from joint work with Mathew D. Penrose [108–110]. Parts of Chapter 8 are adapted from joint work with Mikhail Menshikov [99].

Copyright © 2005 Andrew Robert Wade.
The copyright of this thesis rests with the author. No quotation from it should be published in any format, including electronic and the Internet, without the author's prior written consent. All information derived from this thesis should be acknowledged appropriately.

Acknowledgements

I would like to sincerely thank both my supervisors, Mathew Penrose and Mikhail Menshikov, for all their help, support, ideas and encouragement. For Mathew's crucial help early on I am especially grateful. Thanks also to everybody in the Probability and Statistics group at Durham for a stimulating working environment, in particular to Iain MacPhee, Pauline Coolen-Schrijner, Ostap Hryniv, Lisa Müller, and, from outside the department, Serguei Popov, for useful conversations about my work.

Thanks to everybody – staff, students and secretaries – in the Maths department for making it a fun and friendly place to study. Thanks to all the postgraduate students for insightful discussions on a variety of themes. Thanks to Jonathan and Majid for restraint whilst in the office. Particular thanks are due to James Carlisle, José Sanchez-Loureda, Mike Watson, Owen Madden, Rich Whisker and Mark Short, for many japes.

Thanks to M. Imran for preparing this LaTeX thesis template. Thanks to Bob C for inspiration. Thanks to Sharry and the friday football gang. Thanks to Angelique for being a tolerant housemate, and to the Alder folk of Howlands Farm, 2002/3 harvest.

I gratefully acknowledge the support of the EPSRC, who funded my PhD.

Above all, I would like to thank my parents for making this possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Random spatial graphs (randomly distributed points in space, connected by edges according to some deterministic rule) are of considerable interest in applied probability, with applications to network modelling and statistical procedures. In particular, those graphs in which edges join *nearby* points, and which are therefore locally determined in some sense, have been the subject of study for some time. Examples include the Euclidean minimal spanning tree, the nearest-neighbour graph and its variants, and the geometric graph.

A related class of probabilistic objects is that of Euclidean combinatorial optimization problems, in which, typically, one considers a minimal-length combinatorial structure over random Euclidean points. Famous examples include the travelling salesman problem, minimal spanning tree, and minimal matching problem. The travelling salesman problem can be stated as follows: a salesman has to visit a given set of locations by making a journey in which he visits each location exactly once and returns to his starting point having visited each location. What is the shortest such journey?

The probability theory of such graphs is now well-developed, while several major open problems remain. In this thesis, we consider several examples of 'nearest-neighbour type' graphs, on random point sets in d-dimensional Euclidean space, $\mathbf{R}^d$. Our main focus will be obtaining limit theorems for the *total length* (with power-weighted edges) of these graphs, as the number of points becomes large. Our main results are of two types: laws of large numbers (LLNs) and convergence in distribution results.

In this chapter we discuss some of the history and motivation behind the study of such random spatial graphs, and describe known results in several important cases. For a review of some essential graph theoretic and probabilistic background, see Appendix A.

1.1 Random spatial graphs

In this section we give some examples of random spatial graphs. By $\mathbf{N}$ we denote the set of natural numbers $\{1, 2, 3, \ldots\}$.

For $d \in \mathbf{N}$, let $V \subset \mathbf{R}^d$ be a finite point set in Euclidean space. Let $\|\cdot\|$ denote some norm on $\mathbf{R}^d$; subsequently we will take the Euclidean norm, that is, for $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbf{R}^d$

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{d} |x_i|^2 \right)^{1/2}. \tag{1.1}$$

Given a graph $G = (V, E)$, let $w : E \to [0, \infty)$ be a *weight function* on edges of the graph $G = (V, E)$ (with undirected or directed edges). For example, for $V \subset \mathbf{R}^d$, with $\{\mathbf{x}, \mathbf{y}\} \in E$ one may take $w(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ for the norm $\|\cdot\|$. A case of particular interest is when the weight function on edges is given by power-weighted Euclidean distance, that is, we take $\|\cdot\|$ to be the Euclidean norm on $\mathbf{R}^d$ (given by (1.1)), and, for $\alpha \geq 0$, consider the weight function $w = w_\alpha$ given by

$$w_\alpha(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|^\alpha. \tag{1.2}$$

If $G = (V, E)$ is a graph, denote the total weight of the graph under weight function w by

$$w(G) := \sum_{e \in E} w(e).$$

We now list some examples of random spatial graphs. Some of these emerge from the study of combinatorial optimization problems; see [136] and [142] for relevant monographs. Note that all these graphs can be defined for general vertex sets V; we focus here on $V \subset \mathbf{R}^d$.

1.1.1 Probabilistic setting

Our spatial graphs will be random in that they will be defined on *random* point sets. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) random vectors on $\mathbf{R}^d$ with common density function f. Then, for $n \in \mathbf{N}$, we set

$$\mathcal{X}_n := \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}. \tag{1.3}$$

That is, $\mathcal{X}_n$ is a point process consisting of n i.i.d. random vectors on $\mathbf{R}^d$.

A particular choice for $\mathcal{X}_n$ that we will sometimes use is to take the density f to be the indicator of the unit d-cube, that is $f(\mathbf{x}) = 1$ for $\mathbf{x} \in (0, 1)^d$ and $f(\mathbf{x}) = 0$ otherwise.

In this case, each $\mathbf{X}_i$, $i \in \mathbf{N}$, is *uniformly* distributed on $(0,1)^d$, and we write $\mathbf{U}_i = \mathbf{X}_i$ for all i. In this case, we denote

$$\mathcal{U}_n := \{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n\}, \tag{1.4}$$

and call $\mathcal{U}_n$ the *binomial point process* consisting of n independent uniform random vectors in $(0,1)^d$.

For a bounded $A \subset \mathbf{R}^d$, and $\mathcal{X} \subset \mathbf{R}^d$ a locally finite point set, let $N(\mathcal{X}; A)$ denote the number of points of $\mathcal{X}$ that lie in A ($\mathcal{X}$ being locally finite means that this number is finite for bounded sets A), i.e.

$$N(\mathcal{X}; A) := \operatorname{card}(\mathcal{X} \cap A), \tag{1.5}$$

where card denotes cardinality. Then one sees the origin of the terminology 'binomial' for $\mathcal{U}_n$: if $A \subseteq (0,1)^d$, then the number of points of $\mathcal{U}_n$ that fall in A is binomially distributed $\operatorname{Bin}(n, |A|)$, where $|A|$ is the d-dimensional volume of A.

We also consider $\mathcal{P}_n$, the homogeneous Poisson process of intensity n on $(0,1)^d$. For the general definition and theory of Poisson processes, see Kingman's book [84]. One may characterise $\mathcal{P}_n$ as follows – $\mathcal{P}_n$ is a random countable subset of $(0,1)^d$ such that

(i) for any disjoint (Borel) subsets $A_1, \ldots, A_k$ of $(0,1)^d$, $N(\mathcal{P}_n; A_1), \ldots, N(\mathcal{P}_n; A_k)$ are mutually independent, and

(ii) for each i, $N(\mathcal{P}_n; A_i)$ has the Poisson distribution $\operatorname{Po}(n|A_i|)$.

The following simple relationship between $\mathcal{P}_n$ and $\mathcal{U}_n$ is very useful. This is the fact that the conditional distribution of $\mathcal{P}_n$ given $N(\mathcal{P}_n; (0,1)^d) = m$ is the same as the distribution of $\mathcal{U}_m$. Or, from the other direction, if $N(n)$ is a Poisson random variable with mean n, then $\mathcal{U}_{N(n)} \stackrel{\mathcal{D}}{=} \mathcal{P}_n$. Thus $\mathcal{P}_n$ has the same distribution as a binomial point process consisting of $N(n) \sim \operatorname{Po}(n)$ independent uniform random points. See, for example, p. 21 of [84].

We now give some important examples of spatial graphs.

1.1.2 The travelling salesman problem (TSP)

This famous problem in combinatorial optimization can be posed as follows: what is the length of the shortest closed path spanning V, such that each point of V is visited exactly once? Formally, a *closed tour* (or Hamiltonian cycle) is a closed path traversing each vertex in V exactly once. Let $\mathcal{T}^{d,w}(V)$ be the weight of the shortest closed tour on $V \subset \mathbf{R}^d$, under weight function w. Thus

$$\mathcal{T}^{d,w}(V) := \min_{T=(V,E_T)} \sum_{e \in E_T} w(e), \tag{1.6}$$

where the minimum is over all closed tours $T = (V, E_T)$.

1.1.3 The minimal matching (MM)

Suppose n is an even integer, and $V = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \subset \mathbf{R}^d$. Let $\mathcal{A}^{d,w}(V)$ denote the weight of the minimal matching on V, under weight function w. Let S_n denote the set of all permutations on the integers $1, 2, \ldots, n$. Then,

$$\mathcal{A}^{d,w}(V) := \min_{\sigma \in S_n} \sum_{i=1}^{n/2} w(\mathbf{x}_{\sigma(2i-1)}, \mathbf{x}_{\sigma(2i)}).$$

That is, the minimal matching pairs up points of V so that the total weight of the edges between paired points is minimal. If V consists of an *odd* number of elements, then the usual convention is that the minimal matching on V is given by the minimum-weight minimal matching on each of the n subsets of V consisting of $n-1$ vertices.

1.1.4 The minimal spanning tree (MST)

Given a finite vertex set V, we say the graph $T = (V_T, E_T)$ is a *spanning tree* on V if T is a tree, $V_T = V$ and T spans V (i.e. every member of V lies in an edge of E_T). Then a *minimal spanning tree* (MST) on V is a spanning tree on V with minimal total weight, under weight function w. Let $\mathcal{M}^{d,w}(V)$ denote the weight of a MST on $V \subset \mathbf{R}^d$ under weight function w. Thus, if, $T = (V, E_T)$ is a MST on V, then

$$\mathcal{M}^{d,w}(V) := \sum_{e \in E_T} w(e) \leq \sum_{e' \in E_{T'}} w(e') \tag{1.7}$$

for all spanning trees $T' = (V, E_{T'})$ on V. The MST is unique if the weights of each edge in the complete graph on V are distinct (see Proposition 1.1.1 below). Figure 1.1 below shows a realization of the MST on 50 simulated uniform random points in $(0,1)^2$.

The MST has applications in computer science, biology, physics and biochemistry (see e.g. [46,47]), as well as in statistics; see, for instance, [92] and the references therein. For statistical applications to multivariate non-parametric tests, see in particular [55,56,123]. For an early probabilistic paper on the MST, see [60].

The minimal spanning tree, as the travelling salesman problem and the minimal matching, is defined as a solution to a global optimization problem. However, the MST on finite vertex set V can be constructed in an essentially local manner, via the following 'greedy algorithm' due to Kruskal [88]. For practical implementation of MST algorithms, see [141].

(1) Let $k = 1$, $E_0 = \emptyset$ and $E'_0 = E$, the edge set of the complete graph $G = (V, E)$ on V.

(2) Choose $e_k \in E'_{k-1}$ such that $w(e_k) = \min\{w(e) : e \in E'_{k-1}\}$. Let $E_k = E_{k-1} \cup \{e_k\}$. Take

$$E'_k = \left\{e \in E'_{k-1} \setminus \{e_k\} : E_k \cup \{e\} \text{ has no cycles}\right\}.$$

(3) If $E'_k = \emptyset$ then set $E_T = E_k$ and stop. Else update $k \mapsto k + 1$ and return to (2).

The algorithm chooses edges (from the complete graph on V) one at a time from the shortest available provided no cycles are formed. When the algorithm is done, the graph $T = (V, E_T)$ is produced. Clearly T will be a spanning tree.

We assume that all the edges in the complete graph on V have distinct weights under w (as will occur with probability 1 in the random setting with weight function w_α, $\alpha > 0$, as given by (1.2)). Without this assumption, Kruskal's algorithm still produces an MST, but this MST may no longer be unique. The following result is due to Kruskal [88].

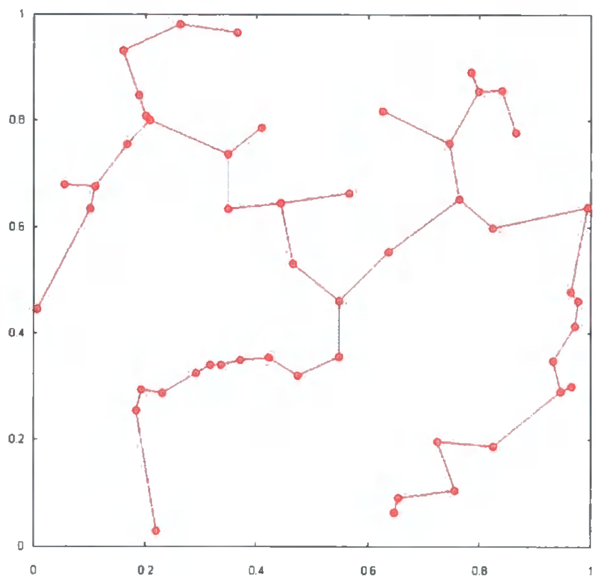
Proposition 1.1.1 *Suppose that the complete graph $G = (V, E)$ on V is such that the edges in E all have distinct weights under w. Then $T = (V, E_T)$ produced by the greedy algorithm above is the unique MST on V.*

Proof. Suppose $T' = (V, E_{T'})$ is a MST on V which has as many edges in common with $T = (V, E_T)$ as possible. Suppose that $T' \neq T$. Let $e = \{u_1, u_2\}$ denote the first edge of T which is not an edge of T'. Then T' contains a unique path from u_1 to u_2. This path contains at least one edge, $f = \{v_1, v_2\}$, say, that is not in T since T has no cycles. When e was selected via the algorithm as an edge of T, f must also have been a candidate. Hence $w(e) \leq w(f)$, by construction. But then $T^* = (V, E_{T^*})$, where $E_{T^*} := (E_{T'} \cup \{e\}) \setminus \{f\}$, is a spanning tree of V, and $w(T^*) = w(T') - w(f) + w(e) \leq w(T')$ so T^* must also be an MST. This tree T^* has more edges in common with T than T' does, contradicting the choice of T'. Hence $T = T'$ and T is an MST. □

1.1.5 The nearest-neighbour graph (NNG)

In the *nearest-neighbour (directed) graph* on vertex set V, each point v of V is connected, by a directed edge (v, u), to a nearest-neighbour $u \in V \setminus \{v\}$ of v; that is $w(v, u) \leq w(v, u')$ for all $u' \in V \setminus \{v\}$ (with u chosen arbitrarily to break any ties). In the *nearest-neighbour (undirected) graph* on V, two vertices u and v are joined by an edge $\{u, v\}$ if u is a nearest-neighbour of v and/or v is a nearest-neighbour of u.

Figure 1.1 shows a realization of the nearest-neighbour (directed) graph on 50 simulated uniform random points in $(0, 1)^2$. Nearest-neighbour graphs and related graphs are the main subject of this thesis. They will be introduced in detail in Chapter 2.



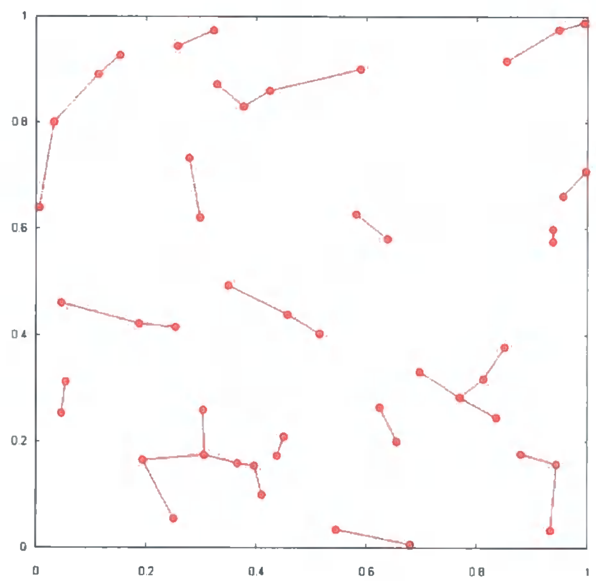


Figure 1.1: Realizations of the MST (left) and NNG (right), each on 50 simulated uniform random points in the unit square.

1.1.6 The geometric graph

In the specific sense employed in [104], a *geometric graph* is an undirected graph on vertex set $V \subset \mathbf{R}^d$ with an edge between any two vertices $\mathbf{u}, \mathbf{v} \in V$ if $||\mathbf{u} - \mathbf{v}|| \leq r$, for a fixed $r > 0$, and $|| \cdot ||$ some norm on $\mathbf{R}^d$.

The book [104] presents many results on random geometric graphs, as well as history and motivation.

1.1.7 Motivation and a brief history

Motivation for studying spatial graphs has originated in many areas. The most obvious motivation is the modelling of real-world spatial networks, such as communications networks, social networks, and, more recently, the internet.

An alternative set of random network models are the 'classical' random graphs of Erdös and Rényi, which have no spatial structure. In a typical Erdös-Rényi type scheme, a random graph on vertex set V is constructed by, independently for each pair of vertices, tossing a coin to determine whether an edge is included. Thus there is no spatial structure to the graph – an edge between two vertices is equally likely. Graphs with spatial content are often more desirable as models for real-world networks. For a thorough account of the theory of Erdös-Rényi random graphs, see [26]. See also [58].

An early paper on *infinite* random geometric graphs (in the plane) is [59]. Such models are now part of the modern theory of *continuum percolation*; see [97] for a survey.

The probabilistic limit theory of random spatial graphs began with the famous paper of Beardwood, Halton and Hammersley [18]. They proved a law of large numbers result for the TSP.

Theorem 1.1.1 *[18] Suppose $d \in \{2,3,\ldots\}$. Let $\mathcal{X}_n$ denote the point process consisting of n independent random points on $(0,1)^d$, with common density f supported by $(0,1)^d$ (a special case of (1.3)). Then, with weight function $w = w_1$ as defined at (1.2) (i.e., simple Euclidean length) and $\mathcal{T}^{d,w}$ the TSP length functional as defined at (1.6), as $n \to \infty$*

$$n^{(1-d)/d}\mathcal{T}^{d,w_1}(\mathcal{X}_n) \stackrel{\text{a.s.}}{\longrightarrow} C(d) \int_{(0,1)^d} f(\mathbf{x})^{(d-1)/d} \mathrm{d}\mathbf{x}, \tag{1.8}$$

where $C(d)$ is a positive constant that depends only on d.

Building on the ideas of [18], an extensive theory for proving results along the lines of (1.8) for a large number of problems in Euclidean optimization has been developed. This theory often makes use of a form of *subadditivity*. The general theory and several applications are presented in the monographs [136] and [142]. In particular, results of the form (1.8) are shown to also hold for the MST and minimal matching, and other problems in combinatorial optimization. The literature in this area is extensive; see [136, 142] for surveys, and also [3, 134, 135].

Convergence in distribution results are generally harder to obtain than laws of large numbers like (1.8). Avram and Bertsimas [11] give central limit theorems and rates of convergence for the NNG (amongst other graphs) using a dependency graph technique of Baldi and Rinott [14]. Here, the crucial idea is that the structure of the graph is essentially locally determined in some sense, and so, with high probability, the dependency structure of the graph has a finite range.

At first, the MST and TSP appear more complicated. Alexander [4] proved a central limit theorem (CLT) for the total length of the MST on Poisson points in $d = 2$, by a continuum percolation approach suggested by Ramey [119]. Kesten and Lee gave the following more general CLT for $d \geq 2$ and power-weighted edges. For $\sigma^2 > 0$ let $\mathcal{N}(0, \sigma^2)$ denote the normal distribution with mean 0 and variance σ^2.

Theorem 1.1.2 *[82] Suppose $d \in \{2, 3, \ldots\}$, and the weight function $w = w_\alpha$, $\alpha > 0$, as given by (1.2). Then, with $\mathcal{M}^{d,w}$ as defined by (1.7), as $n \to \infty$,*

$$n^{(2\alpha-d)/(2d)} \left(\mathcal{M}^{d,w_\alpha}(\mathcal{U}_n) - E[\mathcal{M}^{d,w_\alpha}(\mathcal{U}_n)] \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{\alpha,d}), \tag{1.9}$$

for some $\sigma^2_{\alpha,d} > 0$.

The CLT for the TSP is, in general, still open.

Kesten and Lee's approach is based on a martingale difference method, and a notion of *stabilization* – on the addition of a new point, only the 'nearby' configuration of the MST is affected. This fact has its roots in Kruskal's algorithm mentioned above. The methodology of stabilization was subsequently used by Lee [92, 93] to give an analogous result to (1.9) for the number of vertices with fixed degree in the MST.

Several papers by Penrose and Yukich [105, 106, 111–114] extended the stabilization technique to give general results in geometric probability. These include laws of large numbers [112] and central limit theorems [111, 113], with applications to a wide class of stabilizing functionals, including those concerned with the MST, nearest-neighbour graph, percolation, and Boolean models. More recent results deal with convergence of random *measures* in geometrical probability based on stabilization techniques, see for example [17, 107, 114].

In recent years, interest in random graph models has been considerable with regard to the modelling of real-world networks, and in particular the internet. The search for realistic models of the world wide web often involves replication of certain empirically observed characteristics, such as the so-called 'small worlds' phenomenon (see Watts [139]) and observed degree distributions, often involving some form of power law. Many of these models have a spatial component, and many models that hope to describe network evolution have an *on-line* structure; that is, vertices are added one at a time. A simple model along these lines is the *on-line nearest-neighbour graph*, which we study in this thesis (see Chapter 2).

For extensive surveys on network modelling (including some not completely rigorous results) see [103] and [44]. For recent rigorous mathematical results on world wide web graphs, see for example [19,27–29,36]. Motivation also comes from other communications networks (e.g., telecommunications) and drainage networks (see [122] for a hydrological overview). We describe this in more detail in Chapter 2, in connection with the so-called *minimal directed spanning tree*.

1.2 Random walk in random environment

Given an infinite sequence $\omega = (p_0, p_1, p_2, \ldots)$ such that, for some (small) $\delta > 0$, $\delta \leq p_i \leq 1 - \delta$ for all $i \in \{0, 1, 2, \ldots\}$, we consider $(\eta_t(\omega); t \in \mathbf{Z}^+)$ the nearest-neighbour random walk on $\mathbf{Z}^+ := \{0, 1, 2, \ldots\}$ defined as follows. Set $\eta_0(\omega) = a$ for some $a \in \mathbf{Z}^+$, and for $n = 1, 2, \ldots$,

$$
\begin{aligned}
P[\eta_{t+1}(\omega) = n - 1 | \eta_t(\omega) = n] &= p_n, \\
P[\eta_{t+1}(\omega) = n + 1 | \eta_t(\omega) = n] &= 1 - p_n =: q_n,
\end{aligned}
$$

and $P[\eta_{t+1}(\omega) = 0 | \eta_t(\omega) = 0] = p_0$, $P[\eta_{t+1}(\omega) = 1 | \eta_t(\omega) = 0] = 1 - p_0 =: q_0$. The given form for the reflection at the origin ensures that the Markov chain is *aperiodic*, which eases some technical complications.

We call the sequence of jump probabilities ω our *environment*. As an example, the case $p_i = 1/2$ for all i gives the symmetric simple random walk on $\mathbf{Z}^+$.

Of interest here is the case in which the sequence ω itself is random – in this case $\eta_t(\omega)$ is a *random walk in random environment* (or RWRE for short). Suppose the random environment ω is specified by random variables $(p_0, p_1, p_2, \ldots)$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Write $\mathbb{E}$ for expectation under $\mathbb{P}$.

The RWRE was first studied by Kozlov [87] and Solomon [133], in the case where the p_i, $i \geq 0$ form an i.i.d. sequence; in this case the random environment is *homogeneous*. (In fact, Kozlov and Solomon considered the RWRE on the whole of $\mathbf{Z}$ rather than $\mathbf{Z}^+$). Subsequently, the RWRE has been extensively studied; see for example [121] or [143] for surveys. The higher dimensional RWRE has also received much interest; it is not so well understood as the one-dimensional case. See [143].

Here we will be primarily concerned with the behaviour of $\eta_t(\omega)$ in terms of *recurrence* and *transience*. Recall that (given ω) $\eta_t(\omega)$ is recurrent if it eventually returns to 0 with probability one. Otherwise, it is transient. Let τ_0 denote the time of the first return of $\eta_t(\omega)$ to 0, i.e.

$$\tau_0 := \inf\{t \geq 1 : \eta_t(\omega) = 0\}.$$

Then, given ω, recurrence corresponds to $P[\tau_0 < \infty] = 1$, while transience corresponds to $P[\tau_0 < \infty] < 1$. It follows that, given ω, if $\eta_t(\omega)$ is transient, then $E[\tau_0] = \infty$ and $\eta_t(\omega) \xrightarrow{\text{a.s.}} \infty$.

Also recall that, given ω, if $\eta_t(\omega)$ is recurrent, then it is *null-recurrent* if $E[\tau_0] = \infty$ and *positive recurrent* (or *ergodic* if also irreducible and aperiodic) if $E[\tau_0] < \infty$. Equivalently, $\eta_t(\omega)$ is positive recurrent if and only if there exists a (unique, non-zero) stationary distribution $\pi_j, j \in \mathbf{Z}^+$. See any standard text on Markov chains, e.g. [52].

In the case of the i.i.d. random environment on $\mathbf{Z}^+$, Solomon [133] showed (essentially) that the transience/recurrence properties of $\eta_t(\omega)$ depend on $\zeta_1 = \mathbb{E}[\log(p_1/(1-p_1))]$. If $\zeta_1 < 0$, $\zeta_1 = 0$, $\zeta_1 > 0$ $\eta_t(\omega)$ is respectively transient, null-recurrent, ergodic for $\mathbb{P}$-almost every ω.

The case $\zeta_1 = 0$ is often known as *Sinai's regime.* Sinai [132] showed that, under this condition, roughly speaking, in the RWRE on $\mathbf{Z}$, $\eta_t(\omega)$ is of the order of $(\log t)^2$. Analogous results for the RWRE on $\mathbf{Z}^+$ were given by Golosov (see [63–65]).

Here we are primarily concerned with recurrence/transience criteria for the RWRE on $\mathbf{Z}^+$ in which the environment is a perturbation of Sinai's regime. We study this in Chapter 8.

1.3 Thesis outline

In this thesis we present limit theorems for several nearest-neighbour type graphs in $\mathbf{R}^d$. Most of our results are concerned with large-sample asymptotics for the total weight of the graph (with power-weighted edges) or of certain special subgraphs. These results include laws of large numbers, analogous to Theorem 1.1.1, and central limit theorems, analogous to Theorem 1.1.2. In several cases, our methods enable us to evaluate limiting constants (the analogues of $C(d)$ in (1.8) and $\sigma^2_{\alpha,d}$ in (1.9)) explicitly. This work is presented in Chapters 2–7.

We also present results on the limiting behaviour of the random walk in an asymptotically homogeneous random environment on $\{0, 1, 2, \ldots\}$; we give a complete classification in terms of transience, null-recurrence and ergodicity. This is presented in Chapter 8.

The particular graph that receives most attention is the *minimal directed spanning tree* (MDST), first studied by Bhatt and Roy [21], as a potential model for telecommunications or drainage networks. We present some new results, and potential avenues for further investigation. The MDST is described fully in Chapter 2; the version introduced by Bhatt and Roy places a (directed) edge from $\mathbf{x} \in (0,1)^2$ to $\mathbf{y} \in (0,1)^2$ if $\mathbf{x} \neq \mathbf{y}$ and both components of $\mathbf{x} - \mathbf{y}$ are nonnegative. The MDST bears similarities to the standard minimal spanning tree and the nearest neighbour graph, and some analogous results can be obtained, as well as some rather more specific results. Interesting boundary effects in the MDST can lead to strikingly different behaviour from other graphs.

The overview of the remainder of this thesis is as follows. In Chapter 2 we introduce the nearest-neighbour type graphs that we will consider for the remainder of the thesis, as well as some background, motivation and further references. We also present our laws of large numbers (along the lines of Theorem 1.1.1). Some of the material in this chapter is adapted from joint work with Mathew D. Penrose [109, 110].

In Chapter 3 we present some general results in geometric probability concerned with stabilizing functionals, which we then use to prove the laws of large numbers given in Chapter 2. We also give general central limit theorems, which we will use later (in Chapter 6). Some of the material in this chapter is adapted from joint work with Mathew D. Penrose [109].

Chapter 4 is concerned with the MDST on $\mathcal{U}_n$ and $\mathcal{P}_n$ in $d = 2$, in particular, the total weight of the 'rooted' edges, and the length of the longest edge. It turns out that the limit theory for these quantities can be described in terms of so-called Dickman distributions (which are discussed in Appendix C along with the Poisson-Dirichlet distribution from which they emerge; many of the results given there are well known). Chapter 4 concludes by proving weak convergence results for the 'rooted' edges and longest edge of the MDST on $\mathcal{U}_n$ and $\mathcal{P}_n$ in $d = 2$, stated in terms of Dickman-type distributions. Some of the material in this chapter is adapted from joint work with Mathew D. Penrose [108].

In Chapter 5 we consider *one dimensional* nearest-neighbour type graphs, on uniform random points in the unit interval $[0, 1]$. These graphs are of interest in their own right, as they admit more detailed analysis than their higher dimensional counterparts, but also will prove to be essential to the analysis of the boundary effects in the MDST on the unit square $[0, 1]^2$, which we undertake in Chapter 6. In Chapter 5, amongst our results are weak convergence results for the total weight of the graphs considered. The limits arising in these results are not necessarily normal; some are given in terms of solutions to distributional fixed-point equations. Thus we begin Chapter 5 with a discussion of the 'contraction method' for proving such 'divide-and-conquer' convergence results, as well as a discussion of the theory of Dirichlet spacings, which is an underlying theme in these one-dimensional nearest-neighbour type problems. Some of the material in this chapter is adapted from joint work with Mathew D. Penrose [109, 110].

Chapter 6 deals with convergence in distribution of the total weight of the random MDST in $[0,1]^2$. There are essentially two competing contributions to the limiting total weight – a normal component arising from those points away from the boundary, and a non-normal boundary effect. To analyse the boundary effects, we use the results of Chapter 5; the normal component is handled by the stabilization methodology of Chapter 3. The final limit theorem (Theorem 6.1.1) demonstrates a phase transition at a particular choice of weight exponent (that is, $\alpha = 1$ in w_α given by (1.2)). For $0 < \alpha < 1$, the normal contribution dominates, and we have a central limit theorem, while for $\alpha > 1$, the boundary effects dominate and we have a non-normal limit. When $\alpha = 1$, both effects contribute to the limit law. Some of the material in this chapter is adapted from joint work with Mathew D. Penrose [109].

In Chapter 7 we present some auxiliary results, some conclusions, and some possible directions for further investigation with respect to the material in Chapters 2 to 6.

Chapter 8 is concerned with a rather different topic - that of the complete classification of the one dimensional random walk in random environment (RWRE) in a perturbation of the so-called Sinai's regime. Some of the material in this chapter is adapted from joint work with M.V. Menshikov [99].

The Appendix contains complementary material of an auxiliary or technical nature, including essential technical background (Appendix A), as well as proofs that would otherwise interrupt the flow of the text (Appendix B). Some of this is adapted from joint work with Mathew D. Penrose [108, 109].

Chapter 2

Nearest-neighbour type graphs and laws of large numbers

2.1 Introduction

Graphs constructed on random point sets consisting of independent random points in the unit d-cube ($d \in \mathbf{N}$), formed by joining nearby points according to some deterministic rule, have recently received considerable interest. Such graphs include the geometric graph, the minimal-length spanning tree, and the nearest neighbour graph and its relatives. Many aspects of the large-sample asymptotic theory for such graphs, which are locally determined in a certain sense, are by now quite well understood. See for example [104, 111, 113, 136, 142].

In this chapter we introduce the nearest-neighbour type graphs in which we are interested and present laws of large numbers (in the L^p sense) for the total length (with power-weighted edges) of the k-nearest neighbours (directed) graph, the j-th nearest neighbour (directed) graph, the minimal directed spanning tree, and the on-line nearest-neighbour graph in $\mathbf{R}^d$, $d \in \mathbf{N}$. We give the limiting constants, in the case of uniform random points in $(0,1)^d$, explicitly. We prove our laws of large numbers in Chapter 3, after introducing some general methodology.

In Chapter 5 we deal specifically with the case $d = 1$. In the one-dimensional case, we can often obtain more detailed results, including weak convergence results. The one-dimensional cases are of interest in their own right, in relation to certain fragmentation or interval splitting problems, as we shall see in Chapter 5.

Given a locally finite point set $\mathcal{X} \subset \mathbf{R}^d$, $d \in \mathbf{N}$, and a positive integer k, the k-nearest neighbours (undirected) graph on $\mathcal{X}$, denoted k-NNG($\mathcal{X}$), is the graph with vertex set $\mathcal{X}$ obtained by including $\{x, y\}$ as an edge whenever $y \in \mathcal{X}$ is one of the k nearest neighbours of $x \in \mathcal{X}$, and/or x is one of the k nearest neighbours of y. The k-nearest neighbours (directed) graph on $\mathcal{X}$, denoted k-NNG$'(\mathcal{X})$, is the graph with vertex set $\mathcal{X}$ in which each point is connected (by a directed edge) to each of its k nearest neighbours.

Nearest-neighbour graphs and nearest-neighbour distances in $\mathbf{R}^d$, $d \geq 1$, are of interest in several areas of applied science, including the social sciences, geography and ecology, where proximity data are often important. In the analysis of multivariate data, in particular by non-parametric statistics, nearest-neighbour graphs and near-neighbour distances have found many applications, including goodness of fit tests, classification, regression, noise estimation, density estimation, dimension identification, and the two-sample and multi-sample problems; see for example [22, 31, 48, 56, 68, 69] and references therein.

We also consider the *on-line nearest-neighbour graph* (or ONG for short). This can be described as follows. The ONG is constructed on n points arriving sequentially in $\mathbf{R}^d$ by connecting each point to its nearest neighbour amongst the preceding points in the sequence.

The ONG was apparently introduced in [19] as a simple growth model of the world wide web graph (for $d = 2$). When $d = 1$, the ONG is related to certain fragmentation processes, which are of separate interest in relation to, for example, molecular fragmentation (see e.g. [20], and references therein). The ONG in $d = 1$ is dealt with in more detail in Chapter 5. A central limit theorem for the ONG is given in [106].

In this chapter (Theorem 2.3.1) we give new LLNs for the random ONG in $(0, 1)^d$, $d \in \mathbf{N}$. Later on, in Chapter 5, we also give some more detailed properties of the random ONG when $d = 1$, and identify the limiting distribution of the centred total length of the graph. This distribution is described in terms of a distributional fixed-point equation reminiscent of those encountered in, for example, the analysis of stochastic 'divide-and-conquer' or recursive algorithms. Such fixed-point distributional equalities, and the recursive algorithms from which they arise, have received considerable attention recently; see, for example, [2, 102, 124, 125].

The *minimal directed spanning tree* (or MDST for short) was introduced by Bhatt and Roy in [21]. In its structure, the MDST on n random points in the unit square resembles both the standard minimal spanning tree and the nearest neighbour graph for point sets in the plane, with the extra twist that all edges must be oriented in a south-westerly direction, so that there exists a unique directed path from each vertex to the root placed at the origin. This feature gives rise to significant boundary effects and hence to asymptotic properties which are qualitatively different from those for many of the previously considered graphs.

In this chapter (Theorem 2.4.1) we give laws of large numbers for the total length of the random MDST in $(0,1)^d$, $d \in \mathbf{N}$. Other results on the MDST are presented in subsequent chapters. We also give some preliminary results on the construction of the MDST in Section 2.4.1 of this chapter.

In Chapter 4 we identify the limiting distributions (for large n) for the total length of rooted edges, and also for the maximal length of all edges in the tree. These limit distributions have been seen previously in analysis of the Poisson-Dirichlet distribution and elsewhere; they are expressed in terms of Dickman's function, and their properties are discussed in some detail in Chapter 4. In Chapter 6 we give results for the *total length* of the MDST in the unit square.

We formally define the graphs of interest and present our laws of large numbers in the following sections. Let $\mathcal{X}$ be a finite point set in $\mathbf{R}^d$, and let $\|\cdot\|$ be the Euclidean norm. Write card$(\mathcal{X})$ for the cardinality (number of elements) of $\mathcal{X}$. Let $\mathbf{0}$ denote the origin of $\mathbf{R}^d$. For $d \in \mathbf{N}$, let

$$v_d := \pi^{d/2} \left[\Gamma\left(1 + (d/2)\right)\right]^{-1}, \tag{2.1}$$

the volume of the unit d-ball (see, e.g., [73], equation (6.50)).

Define w to be a weight function on Euclidean edges, assigning weight $w(\mathbf{x}, \mathbf{y})$ to the edge between $\mathbf{x} \in \mathbf{R}^d$ and $\mathbf{y} \in \mathbf{R}^d$, such that $w : \mathbf{R}^d \times \mathbf{R}^d \to [0, \infty)$. We will take the weight function to be power-weighted Euclidean distance, as given by (1.2), for some $\alpha \geq 0$.

We will take the point set $\mathcal{X}$ to be *random*, in particular we take $\mathcal{X} = \mathcal{X}_n$, where $\mathcal{X}_n = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$, for $\mathbf{X}_1, \ldots, \mathbf{X}_n$ independent random vectors in $\mathbf{R}^d$ with common density function f. For some of our results, we assume one of the following conditions on f - either

(C1) f is supported by a convex polyhedron in $\mathbf{R}^d$ and is bounded away from 0 and infinity on its support; or

(C2) For $0 < \alpha < d$, $\int_{\mathbf{R}^d} f(\mathbf{x})^{(d-\alpha)/d} d\mathbf{x} < \infty$, and $\int_{\mathbf{R}^d} |\mathbf{x}|^r f(\mathbf{x}) d\mathbf{x} < \infty$ for some $r > d/(d-\alpha)$.

In many cases, we take $f(\mathbf{x}) = 1$ for $\mathbf{x} \in (0,1)^d$ and $f(\mathbf{x}) = 0$ otherwise, in which case we denote $\mathcal{X}_n = \mathcal{U}_n$, the binomial point process consisting of n independent uniform random vectors on $(0,1)^d$.

Note that, with probability one, $\mathcal{X}_n$ has unique inter-point distances so that all the nearest-neighbour type graphs on $\mathcal{X}_n$ that we consider are almost surely unique.

2.2 k-nearest neighbours, j-th nearest-neighbour graphs

Let $j \in \mathbf{N}$. A point $\mathbf{x} \in \mathcal{X}$ has a j-th *nearest neighbour* $\mathbf{y} \in \mathcal{X} \setminus \{\mathbf{x}\}$ if $\text{card}(\{\mathbf{z} : \mathbf{z} \in \mathcal{X} \setminus \{\mathbf{x}\}, \|\mathbf{z} - \mathbf{x}\| < \|\mathbf{y} - \mathbf{x}\|\}) = j - 1$.

Let $j \in \mathbf{N}$. In the j-th nearest neighbour (directed) graph on $\mathcal{X}$, denoted j-th NNG$'(\mathcal{X})$, each point of $\mathcal{X}$ is joined by a directed edge to its j-th nearest neighbour only.

Let $k \in \mathbf{N}$. In the k-nearest neighbours (directed) graph on $\mathcal{X}$, denoted k-NNG$'(\mathcal{X})$, each point of $\mathcal{X}$ is joined by a directed edge to its first k nearest neighbours in $\mathcal{X}$ (i.e. each of its j-th nearest neighbours for $j = 1, 2, \ldots, k$). Clearly the 1-th NNG$'$ and 1-NNG$'$ coincide, and in this case we have the standard nearest neighbour (directed) graph.

Figure 2.1 shows realizations of the j-th NNG$'$ (with $j = 3$) and k-NNG$'$ (with $k = 5$) each on 50 simulated uniform random points in $(0,1)^2$.

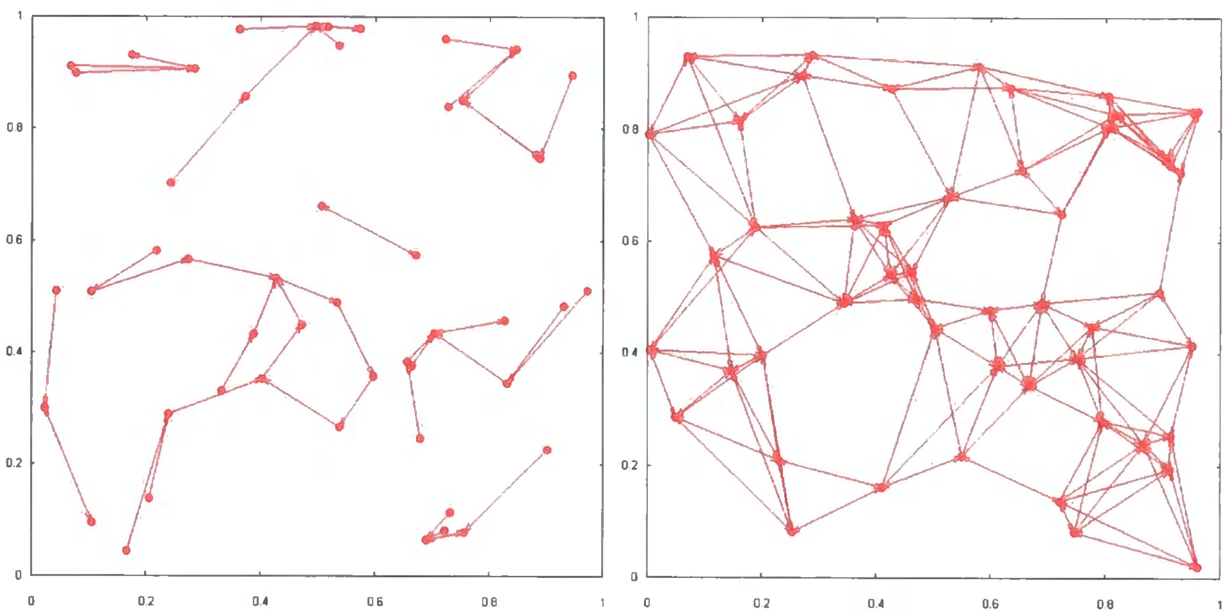


Figure 2.1: Realizations of the 3-rd NNG$'$ (left) and 5-NNG$'$ (right), each on 50 simulated uniform random points in the unit square.

We also consider the k-nearest neighbours (undirected) graph on $\mathcal{X}$, denoted k-NNG$(\mathcal{X})$, in which an undirected edge is placed between points $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ if $\mathbf{x}$ is one of the k nearest neighbours of $\mathbf{y}$, and/or $\mathbf{y}$ is one of the k nearest neighbours of $\mathbf{x}$.

The total length of the random k-nearest neighbours (directed) graph satisfies a central limit theorem; see [22] for $k = 1$, and [11, 17, 111, 114] for general k. Laws of large numbers for the total length of the random k-nearest neighbours (directed) graph are given by McGivney [96] and Yukich (Theorem 8.3 of [142]), and with more general results by Penrose and Yukich [113], but the limiting constants are not given. Here we evaluate these constants explicitly. Partial and related results also appear in [11, 49, 101, 115].

Let $\mathcal{N}_j^{d,\alpha}(\mathcal{X}_n)$, $\mathcal{N}_{\leq k}^{d,\alpha}(\mathcal{X}_n)$ denote respectively the total weight of the j-th nearest neighbour (directed) graph, k-nearest neighbours (directed) graph on $\mathcal{X}_n \subset \mathbf{R}^d$, for $d \in \mathbf{N}$, under weight function w_α, for $\alpha \geq 0$. Note that

$$\mathcal{N}_{\leq k}^{d,\alpha}(\mathcal{X}_n) = \sum_{j=1}^{k} \mathcal{N}_j^{d,\alpha}(\mathcal{X}_n). \tag{2.2}$$

Theorems 2.2.1 and 2.3.1 below feature the constant $C(d, \alpha, k)$, defined for $d \in \mathbf{N}$, $\alpha \geq 0$ and $k \in \mathbf{N}$ by

$$C(d, \alpha, k) := v_d^{-\alpha/d} \frac{d}{d+\alpha} \frac{\Gamma(k + 1 + (\alpha/d))}{\Gamma(k)}. \tag{2.3}$$

In Table 2.1 below we present some values for $C(d, \alpha, k)$ with $\alpha = 1$ and $d \in \mathbf{N}$ for some small values of k. Also, $C(1, \alpha, 1) = 2^{-\alpha}\Gamma(1 + \alpha)$. Proposition 2.2.1 below gives some

$C(d,1,k)$	$d=1$	$d=2$	$d=3$	$d=4$
$k=1$	0.5	0.5	0.553960	0.608140
$k=2$	1.5	1.25	1.292574	1.368315
$k=3$	3	2.1875	2.154290	2.223512
$k=4$	5	3.28125	3.111752	3.149976
$k=5$	7.5	4.511719	4.149003	4.134343

Table 2.1: Some values of $C(d, 1, k)$ for small d, k, given to six decimal places.

asymptotic formulae for $C(d, \alpha, k)$.

Proposition 2.2.1 *Let $k \in \mathbf{N}$, $\alpha \geq 0$ and $d \in \mathbf{N}$. We have*

$$\lim_{d\to\infty} d^{-\alpha/2} C(d,\alpha,k) = \frac{k}{(2e\pi)^{\alpha/2}}; \tag{2.4}$$

$$\lim_{k\to\infty} k^{-1-(\alpha/d)} C(d,\alpha,k) = \frac{d}{d+\alpha} \frac{\Gamma(1+(d/2))^{\alpha/d}}{\pi^{\alpha/2}}; \tag{2.5}$$

$$\begin{aligned} \lim_{d\to\infty}\lim_{k\to\infty} d^{-\alpha/2} k^{-1-(\alpha/d)} C(d,\alpha,k) &= \lim_{k\to\infty}\lim_{d\to\infty} d^{-\alpha/2} k^{-1} C(d,\alpha,k) \\ &= (2e\pi)^{-\alpha/2}. \end{aligned} \tag{2.6}$$

Proof. First we prove (2.4), and so consider the case where $k \in \mathbf{N}$ and $\alpha \geq 0$ are fixed and $d \to \infty$. Recall that Stirling's formula gives, for $x \in \mathbf{R}$, as $x \to \infty$

$$\Gamma(1+x) \sim (2\pi)^{1/2} x^{x+(1/2)} e^{-x}, \tag{2.7}$$

where $\sim$ means that the ratio of the two sides tends to 1 in the limit. From (2.7) we have, as $d \to \infty$

$$\Gamma(1+(d/2))^{(\alpha/d)} \sim (2e)^{-\alpha/2} d^{\alpha/2}.$$

Since $\Gamma(x)$ is continuous for $x > 0$, we have

$$\lim_{d\to\infty} \frac{\Gamma(k+1+(\alpha/d))}{\Gamma(k)} = \frac{\Gamma(k+1)}{\Gamma(k)} = k,$$

so from (2.3) we obtain

$$C(d,\alpha,k) \sim (2e\pi)^{-\alpha/2} d^{\alpha/2} k,$$

as $d \to \infty$, which gives (2.4).

Next we prove (2.5). Suppose that $d \in \mathbf{N}$ and $\alpha \geq 0$ are fixed and let $k \to \infty$. Then, by (2.7),

$$\frac{\Gamma(k+1+(\alpha/d))}{\Gamma(k)} \sim \frac{(k+(\alpha/d))^{k+(\alpha/d)+(1/2)}}{(k-1)^{k-(1/2)}} e^{-1-(\alpha/d)}.$$

Then we have

$$(k+(\alpha/d))^{k+(\alpha/d)+(1/2)} \sim k^{k+(1/2)+(\alpha/d)} e^{\alpha/d},$$

and

$$(k-1)^{k-1/2} \sim k^{k-1/2} \cdot ((k-1)/k)^{k-1/2} \sim k^{k-1/2} e^{-1}.$$

Thus we obtain

$$C(d,\alpha,k) \sim k^{1+(\alpha/d)} \Gamma(1+(d/2))^{\alpha/d} \pi^{-\alpha/2} \frac{d}{d+\alpha},$$

as $k \to \infty$, which gives (2.5).

Finally, (2.6) follows from (2.4) and (2.5). □

Our first main result is Theorem 2.2.1 below, in which we give laws of large numbers for $\mathcal{N}_j^{d,\alpha}(\mathcal{X}_n)$ and $\mathcal{N}_{\leq k}^{d,\alpha}(\mathcal{X}_n)$, complete with explicit expressions for the limiting constants in the case where $\mathcal{X}_n = \mathcal{U}_n$. We prove Theorem 2.2.1 in Section 3.3.1.

Let supp(f) denote the support of f. Recall the conditions (C1) and (C2) given just before the start of this section. Under (C1), supp(f) is a convex polyhedron; under (C2), supp(f) is $\mathbf{R}^d$.

Theorem 2.2.1 *Let $d \in \mathbf{N}$. Suppose the weight function is w_α given by (1.2). The following results hold, with $p = 2$, for $\alpha \geq 0$ if the density function f satisfies condition (C1), and, with $p = 1$, for $0 \leq \alpha < d$ if f satisfies condition (C2).*

(a) For j-th NNG′ on $\mathbf{R}^d$ with weight function w_α, we have, as $n \to \infty$,

$$n^{(\alpha-d)/d}\mathcal{N}_j^{d,\alpha}(\mathcal{X}_n) \xrightarrow{L^p} v_d^{-\alpha/d}\frac{\Gamma(j+(\alpha/d))}{\Gamma(j)}\int_{\mathrm{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d}d\mathbf{x}. \tag{2.8}$$

In particular,

$$n^{(\alpha-d)/d}\mathcal{N}_j^{d,\alpha}(\mathcal{U}_n) \xrightarrow{L^p} v_d^{-\alpha/d}\frac{\Gamma(j+(\alpha/d))}{\Gamma(j)}.$$

(b) For k-NNG′ on $\mathbf{R}^d$ with weight function w_α, we have, as $n \to \infty$,

$$n^{(\alpha-d)/d}\mathcal{N}_{\leq k}^{d,\alpha}(\mathcal{X}_n) \xrightarrow{L^p} C(d,\alpha,k)\int_{\mathrm{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d}d\mathbf{x}. \tag{2.9}$$

In particular,

$$n^{(\alpha-d)/d}\mathcal{N}_{\leq k}^{d,\alpha}(\mathcal{U}_n) \xrightarrow{L^p} C(d,\alpha,k). \tag{2.10}$$

Remarks. (a) If we use a different norm on $\mathbf{R}^d$ than the Euclidean, Theorem 2.2.1 remains valid with v_d appropriately redefined to be the volume of the unit d-ball in the chosen norm.

(b) Laws of large numbers for the k-NNG′ total length functional can be found in [96]. Theorem 8.3 of [142] gives a LLN (with complete convergence) for $\mathcal{N}_{\leq k}^{d,1}(\mathcal{X}_n)$. Our Theorem 2.2.1, without the explicit constants, follows from Theorem 2.4 of [113]. In none of these are the limiting constants evaluated explicitly. Avram and Bertsimas (Theorem 7 of [11]) attribute a result on the limiting expectation (and hence the constant in the law of large numbers) for the j-th NNG′$(\mathcal{U}_n)$ in $d = 2$, with $\alpha = 1$, to Miles [101] (see also page 101 of [142]). From (2.8) we have

$$n^{(\alpha-d)/d}\mathcal{N}_j^{d,\alpha}(\mathcal{U}_n) \xrightarrow{L^p} v_d^{-\alpha/d}\frac{\Gamma(j+(\alpha/d))}{\Gamma(j)}. \tag{2.11}$$

The constant in [11] is given as

$$\frac{1}{2}\pi^{-1/2}\sum_{i=1}^{j}\frac{\Gamma(i-(1/2))}{\Gamma(i)},$$

which simplifies to $\pi^{-1/2}\Gamma(j+(1/2))/\Gamma(j)$, and so agrees with the $\alpha = 1$, $d = 2$ case of (2.11).

(c) Related results are the asymptotic expressions for expectations of j-th nearest neighbour distances in finite point sets given in [115] and [49]. The results in [115] are consistent with the $\alpha = 1$ case of our (2.10). The result in [49] includes general α and certain non-uniform densities, although their conditions on f are more restrictive than our (C1); the result is consistent with (2.9). Also, [49] gives (equation (6.4)) a weak LLN (with convergence in probability) for the empirical mean k-nearest neighbour *distance.* With Theorem 2.4 of [113], the results in [49] yield LLNs for the total weight of the j-th NNG′ and k-NNG′ only when $d-1<\alpha<d$ (due to the rates of convergence given in [49]). Our methods yield LLNs for any $\alpha > 0$ under (C1), and also encompass a wider class of density functions f. It may be possible to obtain rates of convergence in (2.8) and (2.9) by adapting the methods of [49].

(d) Finally, we note that when we take $\alpha = 0$, the functional $\mathcal{N}_j^{d,0}(\mathcal{X}_n)$ simply counts the number of points in $\mathcal{X}_n$ with j-th nearest neighbours, and so $\mathcal{N}_j^{d,0}(\mathcal{X}_n) = n$ a.s. for $j \leq n-1$, and $\mathcal{N}_j^{d,0}(\mathcal{X}_n) = 0$ a.s. for $j \geq n$. Similarly, $\mathcal{N}_{\leq k}^{d,0}(\mathcal{X}_n)$ counts the total number of j-th nearest neighbours for $j = 1, 2, \ldots, k$ for all points of $\mathcal{X}_n$, and so $\mathcal{N}_j^{d,0}(\mathcal{X}_n) = kn$ a.s. provided $k \leq n-1$, and $\mathcal{N}_j^{d,0}(\mathcal{X}_n) = n(n-1)$ a.s. provided $k \geq n$. These observations are consistent with the $\alpha = 0$ cases of (2.8) and (2.9).

From our results on nearest neighbours (directed) graphs, it is possible to obtain results for nearest neighbours (undirected) graphs, in which if x is a nearest neighbour of y and vice versa, then the edge between x and y is only counted once. As an example, we give the following result for the total weight of the standard nearest neighbour (undirected) graph in $\mathbf{R}^d$.

Let $\mathcal{Z}^{d,\alpha}(\mathcal{X}_n)$ denote the total weight, with weight function w_α $\alpha > 0$, of the nearest neighbour (undirected) graph on $\mathcal{X}_n \subset \mathbf{R}^d$, $d \in \mathbf{N}$. Recall the definition of v_d, the volume of the unit d-ball, from (2.1). For $d \in \mathbf{N}$, let ω_d be the volume of the union of two unit d-balls with centres unit distance apart.

Theorem 2.2.2 *Let $d \in \mathbf{N}$. Suppose the weight function is w_α as given by (1.2). Suppose $\alpha \geq 0$ and that the density function f satisfies condition (C1). Then, as $n \to \infty$,*

$$n^{(\alpha-d)/d} \mathcal{Z}^{d,\alpha}(\mathcal{X}_n) \xrightarrow{L^2} \Gamma(1+(\alpha/d)) \left(v_d^{-\alpha/d} - \frac{1}{2} v_d \omega_d^{-1-(\alpha/d)} \right) \int_{\mathrm{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d} \mathrm{d}\mathbf{x}. \quad (2.12)$$

In particular, when $d = 2$ we have, for $\alpha > 0$

$$n^{(\alpha-2)/2} \mathcal{Z}^{2,\alpha}(\mathcal{U}_n) \xrightarrow{L^2} \Gamma(1+(\alpha/2)) \left(\pi^{-\alpha/d} - \frac{\pi}{2} \left(\frac{6}{8\pi + 3\sqrt{3}} \right)^{1+(\alpha/2)} \right), \quad (2.13)$$

and when $d = 2$, $\alpha = 1$, we get

$$n^{-1/2} \mathcal{Z}^{2,1}(\mathcal{U}_n) \xrightarrow{L^2} \frac{1}{2} - \frac{1}{4} \left(\frac{6\pi}{8\pi + 3\sqrt{3}} \right)^{3/2} \approx 0.377508. \quad (2.14)$$

Finally, when $d = 1$, $\alpha = 1$, we have $\mathcal{Z}^{1,1}(\mathcal{U}_n) \xrightarrow{L^2} 7/18$ as $n \to \infty$.

We give the proof of Theorem 2.2.2 following the proof of Theorem 2.2.1 in Section 3.3.1.

Remark. It should be possible to obtain more general results for the undirected graphs k-NNG$(\mathcal{X}_n)$ $(k = 2, 3, \ldots)$ using our methods, and by modifying the methods of Henze [68] for the fraction of points that are the l-th nearest neighbour of their own k-th nearest neighbour.

2.3 The on-line nearest-neighbour graph

We now consider the *on-line nearest-neighbour graph* (or ONG for short). Let d be a positive integer. Suppose $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are points in $(0,1)^d$, arriving sequentially; the ONG on vertex set $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ is formed by connecting each point $\mathbf{X}_i$, $i = 2, 3, \ldots, n$ to its nearest neighbour amongst the preceding points in the sequence (i.e. $\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}$), using the lexicographic ordering on $\mathbf{R}^d$ to break any ties. The resulting graph is a tree, which we call the ONG on $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$.

For our results on the ONG, we restrict our analysis to the case in which the points $\mathbf{X}_1 = \mathbf{U}_1, \mathbf{X}_2 = \mathbf{U}_2, \ldots$ are uniformly distributed on $(0,1)^d$. One could consider more general distributions, as in the previous section.

The ONG is of interest as a natural growth model for random spatial graphs; in particular it has been used (with $d = 2$) in the context of the world wide web graph (see [19]). In [106], stabilization techniques were used to prove that the total length (suitably scaled) of the ONG on uniform random points in $(0,1)^d$ for $d > 4$ converges in distribution to a normal random variable. It is suspected that a central limit theorem also holds for $d = 2, 3, 4$. On the other hand, when $d = 1$, the limit is not normal, as demonstrated by Theorem 5.2.2 (ii) in Chapter 5.

Figure 2.2 shows a realization of the ONG on 50 simulated uniform random points in the unit interval. Figure 2.3 below shows realizations of the planar and three-dimensional ONG, each on 50 simulated random points.

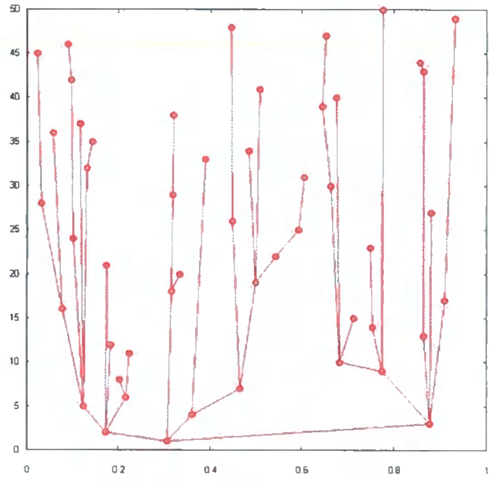


Figure 2.2: Realization of the ONG on 50 simulated random points in the unit interval. The vertical axis gives the order in which the points arrive, and their position is given by the horizontal axis.

In order to obtain our law of large numbers (Theorem 2.3.1 below), we modify the setup of the ONG slightly. Let $\mathcal{U}_n$ be a *marked* random finite point process in $\mathbf{R}^d$, consisting of n independent uniform random vectors in $(0,1)^d$, where each point $\mathbf{U}_i$ of $\mathcal{U}_n$ carries a random mark $T(\mathbf{U}_i)$ which is uniformly distributed on $[0,1]$, independent of the other marks and of the point process $\mathcal{U}_n$. The points are listed in increasing order of mark, i.e. the marks represent time of arrival. With this ordering, we connect each point of $\mathcal{U}_n$ to the nearest point that precedes it in the ordering, if such a point exists, to obtain a graph that we call the ONG on the marked point set $\mathcal{U}_n$. This definition extends to infinite but locally finite point sets.

Clearly the ONG on the marked point process $\mathcal{U}_n$ has the same distribution as the ONG (with the first definition) on a sequence $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n$ of independent uniform

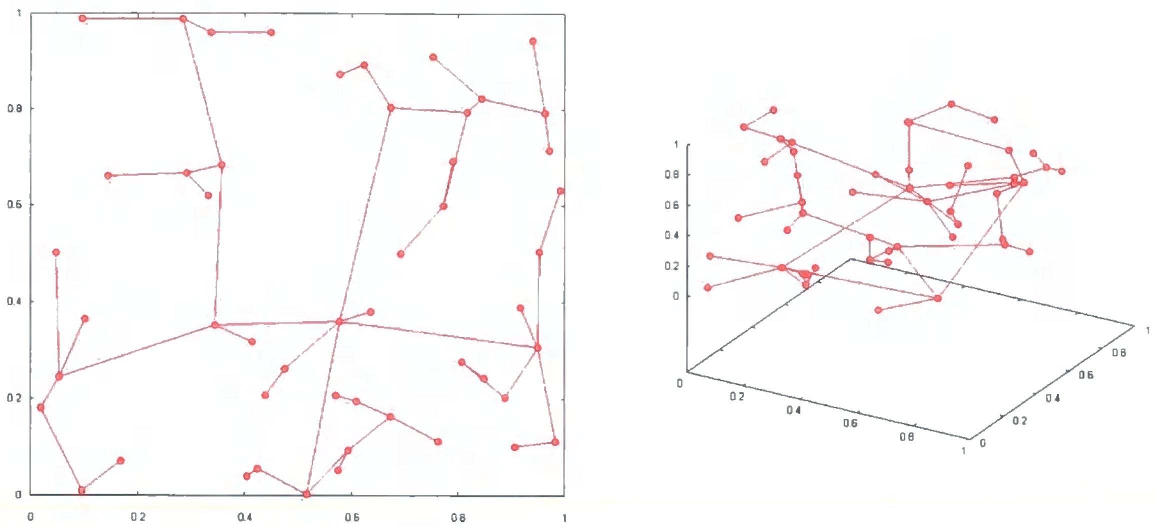


Figure 2.3: Realizations of the ONG on 50 simulated uniform random points in the unit square (left) and the unit cube (right).

points on $(0,1)^d$.

For $d \in \mathbf{N}$ and $\alpha \geq 0$, let $\mathcal{O}^{d,\alpha}(\mathcal{U}_n)$ denote the total weight, with weight function w_α, of ONG$(\mathcal{U}_n)$. Our results for the ONG in general dimensions are as follows, and constitute a law of large numbers for $\alpha < d$, a distributional convergence result for $\alpha > d$, and asymptotic behaviour of the mean for $\alpha = d$.

Theorem 2.3.1 *Suppose $d \in \mathbf{N}$, and the weight function is w_α as given by (1.2).*

(i) Suppose $0 \leq \alpha < d$. For the ONG, *as $n \to \infty$, with $C(d, \alpha, k)$ as given by (2.3), we have*

$$n^{(\alpha-d)/d}\mathcal{O}^{d,\alpha}(\mathcal{U}_n) \xrightarrow{L^1} \frac{\alpha}{d-\alpha} C(d,\alpha,1) \tag{2.15}$$

(ii) Suppose $\alpha > d$. Then, as $n \to \infty$,

$$\mathcal{O}^{d,\alpha}(\mathcal{U}_n) \longrightarrow W(d,\alpha), \tag{2.16}$$

where the convergence is in L^p, $(p \in \mathbf{N})$, and almost sure, and $W(d,\alpha)$ is a nondegenerate, nonnegative random variable with $E[(W(d,\alpha))^k] < \infty$ for $k = 1, 2, 3, \ldots$.

(iii) Suppose $\alpha = d$. Then, as $n \to \infty$,

$$E[\mathcal{O}^{d,d}(\mathcal{U}_n)] = v_d^{-1} \log n + O(1). \tag{2.17}$$

In particular (2.17) implies that $E[\mathcal{O}^{1,1}(\mathcal{U}_n)] = (1/2)\log(n) + O(1)$, a result given more precisely in the convergence of expectations version of Theorem 5.2.2 (ii) in Chapter 5. We prove Theorem 2.3.1 (i) in Section 3.3.2 and Theorem 2.3.1 (ii) and (iii) in Section 3.3.3.

2.4 The minimal directed spanning tree (MDST)

In Bhatt and Roy's MDST construction [21], each point $\mathbf{x}$ of a finite (random) subset $\mathcal{S}$ of $(0,1)^d$ (in [21], $d = 2$) is connected by a directed edge to the nearest $\mathbf{y} \in \mathcal{S} \cup \mathbf{0}$ such that $\mathbf{y} \neq \mathbf{x}$ and $\mathbf{y} \preccurlyeq^* \mathbf{x}$, where $\mathbf{y} \preccurlyeq^* \mathbf{x}$ means that each component of $\mathbf{x} - \mathbf{y}$ is nonnegative ($\preccurlyeq^*$ is the "coordinate-wise" partial order on $\mathbf{R}^d$). Of interest is the behaviour of the length of the graph, or of various parts of the graph.

The original motivation for studying the MDST comes from communications and drainage networks (see the discussion at the end of this section, and also [21, 108, 122]). The constraint on the direction of the edges can lead to significant boundary effects due to the possibility of long edges occurring near the lower and left boundaries of the unit square (see Chapter 6). Another difference between the MDST and the standard minimal spanning tree and nearest-neighbour graph for point sets in the plane is the fact that there is no uniform upper bound on vertex degrees in the MDST.

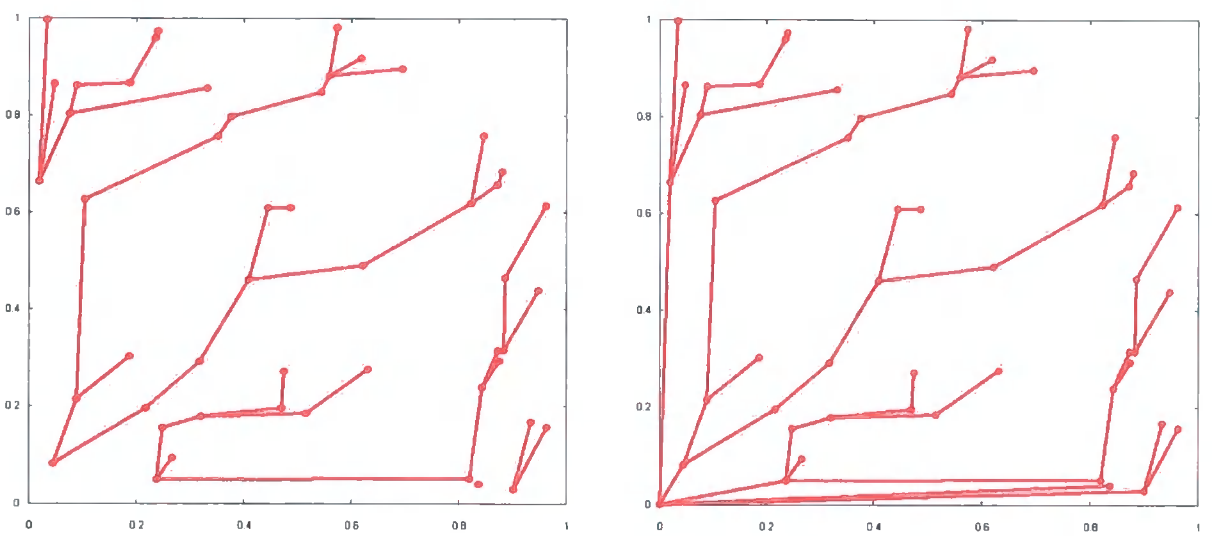


Figure 2.4: Realizations of the MDSF (left) and MDST on 50 simulated uniform random points in the unit square, under the partial ordering $\preccurlyeq^*$.

We consider a more general definition of the MDST than that used by Bhatt and Roy.

In particular, we consider a general class of partial orderings, including $\preccurlyeq^*$, and take d to be any nonnegative integer. We consider properties of the MDST on random points in $(0,1)^d$ as the number of points becomes large. We also consider the *minimal directed spanning forest* (MDSF), which is the MDST with the edges incident to $\mathbf{0}$ removed; this model seems more appropriate for general partial orders. In [21], Bhatt and Roy mention that the total length of the graph is an object of considerable interest, although they restrict their analysis to the length of the edges joined to the origin (see also Chapter 4). A first order result for the total length of the MDST or MDSF is a law of large numbers; we present this in Theorem 2.4.1 for a family of MDSFs indexed by partial orderings on $\mathbf{R}^d$, which include $\preccurlyeq^*$ as a special case.

Second order results, i.e., weak convergence results for the distribution of the total length, suitably centred and scaled, in the case where $d = 2$ and the partial order is $\preccurlyeq^*$, are presented in Chapter 6.

In Chapter 4, we give results dealing with the weight of the edges joined to the origin (with $d = 2$ under $\preccurlyeq^*$), including weak convergence results, in which the limiting distributions are given in terms of some generalized Dickman distributions. Subsequently, it has been shown [12] that this two dimensional case is rather special – in higher dimensions the corresponding limits are normally distributed. Chapter 4 also deals with the maximum edge length of the MDST, and the maximum edge length of those edges incident to the origin (which was also dealt with in [21]).

The MDST is defined formally in the next section. Motivation comes from the modelling of communications or drainage networks. The communications model considered in [21] goes as follows. Consider a network of radio masts, each of which can receive signals only from masts to the south-west. Suppose a source transmitter is positioned at the origin of the plane, and a network of masts is positioned in the first quadrant. Then the graph of the transmission network can be viewed as a directed spanning tree. For convenience, the direction of the edges is taken to be from receiver to transmitter, so that all the directed paths eventually meet at the origin. We restrict the model to a single link into each receiver, which we may justify by asserting that once the first connection has been established, further links may be ignored for many purposes. Various characteristics of the resulting graph are then of interest.

The same graph may be considered as a model for drainage networks, following the spirit of Rodriguez-Iturbe and Rinaldo in [122]; again, see [21]. The idea is that water is allowed to run off an inclined bounded field, forming several drainage channels. These channels eventually merge so that all the water flows out of the field at the lowest point on the boundary. Given any particular landscape geometry, this situation is fairly unpleasant to model directly, so we study a model that maintains the essential features of the above system while being much simpler to handle.

In one model proposed in [122], given a fixed number n of points which the stream network (graph) must contain as nodes, the optimal configuration is achieved by minimizing the quantity $\sum_i Q_i^{1/2} L_i$ where L_i is the length and Q_i the discharge of stream (edge) i. If we assume that Q_i is fixed for all i (and so the flows are non-additive), and flow is constrained to be in a south-westerly direction, the optimum configuration on a set of points is given by the construction we consider here. Another viewpoint is to consider the catchment of the network, which will depend on the total length of the channels.

Understanding these networks for large systems may be difficult: by investigating the behaviour of the MDST on random points we hope to shed light on their 'typical' behaviour.

2.4.1 Construction and basic results

Suppose V is a finite nonempty set endowed with a partial ordering $\preccurlyeq$, that is a binary relation $\preccurlyeq$ between elements of V such that (see e.g. [86])

(i) $\preccurlyeq$ is reflexive, i.e. $u \preccurlyeq u$ for all $u \in V$;

(ii) $\preccurlyeq$ is transitive, i.e. if $u \preccurlyeq v$ and $v \preccurlyeq w$ then $u \preccurlyeq w$ for all $u, v, w \in V$;

(iii) If, for $u, v \in V$, $u \preccurlyeq v$ and $v \preccurlyeq u$ then $u = v$.

The partial ordering induces a directed graph $G = (V, E)$ on V, with vertex set V and edge set E consisting of all elements (directed edges) $(v, u) \in E, u \neq v$ that are ordered pairs of elements of V such that $u \preccurlyeq v$.

Definition 2.4.1 *A* minimal element, *or* sink, *is a vertex $v_0 \in V$ for which there exists no $v \in V \setminus \{v_0\}$ such that $v \preccurlyeq v_0$. Let V_0 denote the set of all sinks of V; observe that V_0 cannot be empty.*

Definition 2.4.2 *A* directed spanning subgraph *(DSS) of G is a subgraph $H = (V, E_H)$ of (V, E), such that, for each vertex $v \in V \setminus V_0$, there exists at least one directed path in H that starts at v and ends at some sink $u \in V_0$.*

Definition 2.4.3 *A* directed spanning forest *(DSF) T on V is a DSS on V such that, for each vertex $v \in V \setminus V_0$, there exists a unique directed path in T that starts at v and ends at some sink $u \in V_0$. In the case where V_0 consists of a single sink, we refer to any DSF on V as a* directed spanning tree *(DST) on V.*

We can show easily that the DST is in fact a tree and the DSF is in fact a forest, and that the DST and DSF have other graph-theoretic properties. We do this by proving the following equivalence result.

Proposition 2.4.1 *The following properties of a DSS $H = (V, E_H)$ of $G = (V, E)$ are equivalent:*

(i) H is a DSF (or DST), i.e., for each vertex $v \in V \setminus V_0$ there exists a unique sequence of directed edges in E_H of the form $(v, v_k), (v_k, v_{k-1}), \ldots, (v_1, v_0)$ where $k \in \{0, 1, \ldots\}$, and some $v_0 \in V_0$.

(ii) For each vertex $v \in V \setminus V_0$ there exists exactly one edge of the form $(v, u) \in E_H$ (where $u \preccurlyeq v$).

(iii) Ignoring the orientation of the edges, H is a forest whose components are in one-to-one correspondence with the elements of V_0 (in particular, if V_0 contains a single sink, H is a tree).

In the case where V has a single sink, the following is also equivalent to any one of the above statements; otherwise it is implied by any one of the above:

(iv) There are no cycles (disregarding the directedness of the edges) in H.

Proof. Consider (i). By uniqueness, we must have that any two vertices v_1, v_2 that are joined to the same sink v_0 must lie in the same connected component of H, rooted at v_0. Hence, ignoring orientation, condition (i) implies (iii).

Condition (ii) can be restated as there exists no branching point in H, i.e., there do not exist distinct vertices $u, u', v \in V$ such that (v, u) and (v, u') are both edges of H. Since, by definition of a DSS, H contains at least one directed path from v to a sink, we have that (ii) is equivalent to (i) – this path is unique.

It also clear that (iii) implies (ii). Also, it is easy to see that (ii) implies (iv), and in the case of a single sink, (iv) implies (iii), for example. □

Recall that a weight function on the edges of a (directed) graph (V, E) is a function $w : E \to [0, \infty)$.

Definition 2.4.4 *Suppose that V is a partially ordered finite set and that the induced graph $G = (V, E)$ carries a weight function w. A* minimal directed spanning forest *(MDSF) on V (or, equivalently, on G), is a directed spanning forest T on V with edge set $E_T \subseteq E$ such that*

$$w(T) := \sum_{e \in E_T} w(e) = \min \left\{ \sum_{e \in E_{T'}} w(e) : T' = (V, E_{T'}) \text{ a DSF on } V \right\}. \quad (2.18)$$

If V has a single sink, then any minimal directed spanning forest on V is called a minimal directed spanning tree *(MDST) on V.*

Thus, a MDSF on V is defined as a solution to a global optimization problem. However, the following simple result shows that, when all edge weights are distinct, a MDSF can be constructed simply in a 'local' manner, reminiscent of Kruskal's greedy algorithm [88] for finding the minimal spanning tree in an undirected graph.

Definition 2.4.5 *We say that $v \in V$ has a* directed nearest neighbour *$u \in V \setminus \{v\}$ if $u \preccurlyeq v$ and $w(v, u) \leq w(v, u')$ for all $u' \in V \setminus \{v\}$ such that $u' \preccurlyeq v$.*

Note that in the random setting (see below), the directed nearest neighbour of a point is almost surely unique, and so the MDSF (or MDST) is almost surely uniquely defined.

Proposition 2.4.2 *Suppose that V is a partially ordered finite set with set of sinks denoted V_0 and that the induced graph $G = (V, E)$ carries a weight function w. For each $v \in V \setminus V_0$, let n_v denote a directed nearest neighbour of v (chosen arbitrarily if v has more than one directed nearest neighbour). Let $M = (V, E_M)$ be the DSS of V obtained by taking*

$$E_M := \{(v, n_v) : v \in V \setminus V_0\}.$$

Then M is a MDSF on V.

Proof. Let $T = (V, E_T)$ be an arbitrary DSF on V. Then, for every $v \in V \setminus V_0$, there exists a unique element of V, denoted u_v such that $(v, u_v) \in E_T$ (uniqueness follows from Proposition 2.4.1). Necessarily $u_v \preccurlyeq v$, and by definition of directed nearest neighbours we have

$$w(M) = \sum_{v \in V \setminus V_0} w(v, n_v) \leq \sum_{v \in V \setminus V_0} w(v, u_v) = w(T)$$

for every DSF T. Thus, M is a MDSF on V. □

From the properties described above, it is clear how to construct the MDSF (or MDST) on a given a finite point set V: Join each point $v \in V \setminus V_0$ to a directed nearest neighbour n_v by the edge (v, n_v). Given a MDSF on V, it is therefore easy to construct a MDSF on $V' := V \cup \{v'\}$ for some $v' \notin V$, via an 'add and delete' procedure. Let $M = (V, E_M)$ be a MDSF on V (for example, as obtained by the procedure in Proposition 2.4.2). Then the following add/delete procedure produces a MDSF $M' = (V', E_{M'})$ on V': we add the edge from v' to a directed nearest neighbour (if one exists), and replace previous edges from points that now have v' as a directed nearest neighbour in V' with the edge from those points to v'. More precisely, let $n_{v'}$ denote a directed nearest neighbour of v' in V, if one exists. If such an $n_{v'}$ exists, let $e' := (v', n_{v'})$. Denote by W the (possibly empty) set of vertices $v \in V$ such that a directed nearest neighbour of v in V' is v'. For $v \in W$, let u_v denote the directed nearest neighbour of v in V (that is, *not* including v'). Then

$$E_{M'} = \left(E_M \cup \{e'\} \cup \bigcup_{v \in W} \{(v, v')\} \right) \setminus \left(\bigcup_{v \in W} \{(v, u_v)\} \right). \tag{2.19}$$

Then (by Proposition 2.4.2) M' is a MDSF on V'.

While the statements above apply to any partially ordered set V with weights defined for all induced edges, we will be subsequently concerned only with the case where V is a randomly generated subset of $\mathbf{R}^d$, $d \in \mathbf{N}$, and where the partial ordering and weight function are as follows.

The partial order

For what follows, we consider a general type of partial ordering of $\mathbf{R}^2$, denoted $\overset{\theta,\phi}{\preccurlyeq}$, specified by the angles, $\theta \in [0, 2\pi)$ and $\phi \in (0, \pi]$. For $\mathbf{x} \in \mathbf{R}^2$, let $C_{\theta,\phi}(\mathbf{x})$ be the closed cone with vertex $\mathbf{x}$ and boundaries given by the rays from $\mathbf{x}$ at angles θ and $\theta + \phi$, measuring anticlockwise from the upwards vertical. The partial order is such that, for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{R}^2$,

$$\mathbf{x}_1 \overset{\theta,\phi}{\preccurlyeq} \mathbf{x}_2 \text{ iff } \mathbf{x}_1 \in C_{\theta,\phi}(\mathbf{x}_2).$$

We shall use $\preccurlyeq^*$ as shorthand for the special case $\overset{\pi/2,\pi/2}{\preccurlyeq}$, which is of particular interest, as in [21]. This is the coordinate-wise partial order on $\mathbf{R}^2$, for which $u \preccurlyeq^* v$ for $u = (u_1, u_2), v = (v_1, v_2)$ if and only if $u_1 \leq v_1$ and $u_2 \leq v_2$. Many of our results are specific to $\preccurlyeq^*$, as in Chapters 4 and 6.

For general $d \in \mathbf{N}$ we consider two special partial orders on $\mathbf{R}^d$. The first we again denote $\preccurlyeq^*$, for which we have $\mathbf{y} \preccurlyeq^* \mathbf{x}$ for $\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$ if and only if each component of $\mathbf{x} - \mathbf{y}$ is nonnegative – this is the coordinate-wise partial order, coinciding with the above definition of $\preccurlyeq^*$ in $d = 2$.

The second we denote by $\preccurlyeq_*$, such that $\mathbf{y} \preccurlyeq_* \mathbf{x}$ for $\mathbf{x}, \mathbf{y} \in \mathbf{R}^d$ if and only if the *first* component of $\mathbf{x} - \mathbf{y}$ is nonnegative. We refer to this as the single-coordinate partial order. In $d = 2$, this corresponds to $\overset{0,\pi}{\preccurlyeq}$. For aesthetic reasons, however, we will take $\preccurlyeq_*$ to be instead the (for our purposes equivalent) partial order $\overset{\pi/2,\pi}{\preccurlyeq}$ in $d = 2$, so that all edges run north-south in $(0,1)^2$. Note that when $d \geq 2$ the binary relation $\preccurlyeq_*$ is not, strictly speaking, a partial order on $\mathbf{R}^d$ ($\mathbf{z} \preccurlyeq_* \mathbf{x}$ and $\mathbf{x} \preccurlyeq_* \mathbf{z}$ does not imply that $\mathbf{x} = \mathbf{z}$). However, in the random setting we consider, $\preccurlyeq_*$ *will* be a true partial order on our point sets, with probability one.

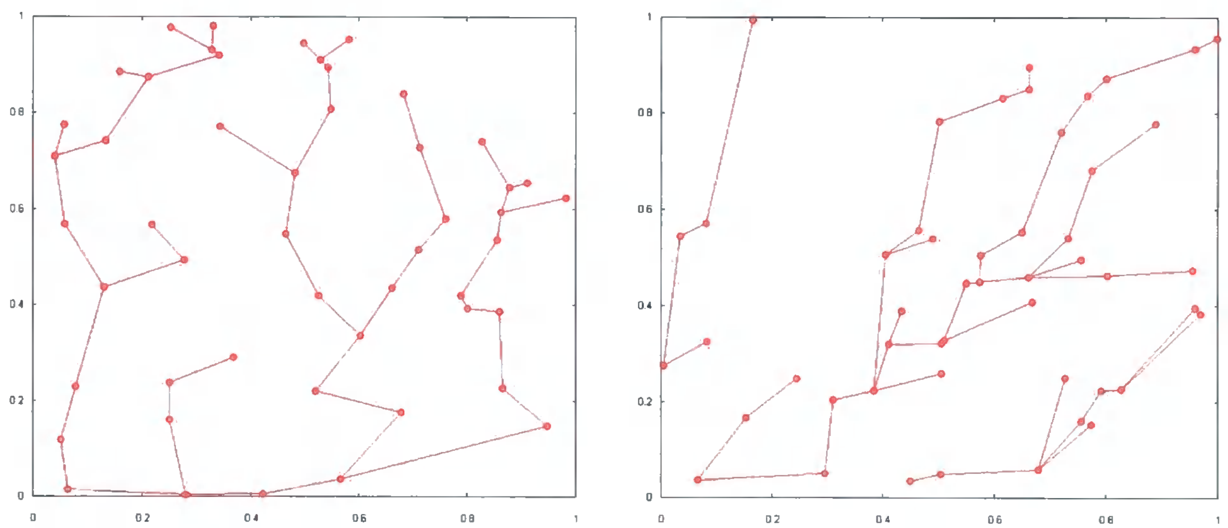


Figure 2.5: Realizations of the MDST under $\preccurlyeq_*$ (left) and MDSF under $\preccurlyeq^*$ on 50 simulated uniform random points in $(0,1)^2$.

Note that in $d = 1$, $\preccurlyeq^*$ and $\preccurlyeq_*$ both equate to the total order relation $\leq$. Where required, the symbol $\preccurlyeq$ will denote an arbitrary partial order on $\mathbf{R}^d$.

Remarks. Outside the given ranges of θ and ϕ, the relation $\overset{\theta,\phi}{\preccurlyeq}$ is no longer a partial order, and we do not consider this case in this thesis. If we allow $\pi < \phi \leq 2\pi$ we obtain a directional relation between points in the plane, but this will not be a partial order, since we do not have that if $u \preccurlyeq v$ and $v \preccurlyeq w$ then $u \preccurlyeq w$ for all $u, v, w \in V$. This follows from the fact that the statement

$$C_{\theta,\phi}(\mathbf{x}_1) \subseteq C_{\theta,\phi}(\mathbf{x}_2) \ \forall \mathbf{x}_1 \in C_{\theta,\phi}(\mathbf{x}_2)$$

holds if and only if $0 < \phi \leq \pi$. We do not consider $\pi < \phi \leq 2\pi$ in detail. However, one could define a graph, by analogy with the MDSF, such that each point $v \in V \setminus V_0$ is joined to a 'directed nearest neighbour' with the relation $\overset{\theta,\phi}{\preccurlyeq}$ replacing the partial order. In fact, our laws of large numbers (Theorem 2.4.1) will still hold in this case. However, the graph in general will not be a forest (it may contain cycles, for instance) and the graph theoretic results given above will not hold. One particular case of interest is the case $\phi = 2\pi$, which leads in fact to the standard (directed) nearest neighbour graph (as described in Section 2.2); in this case, Theorem 2.4.1 coincides with the $d = 2$ case of Theorem 2.2.1. We do not permit here the case $\phi = 0$, in which case we almost surely have a disconnected point set.

2.4.2 The random MDST/F in $\mathbf{R}^d$

The weight function is given by power-weighted Euclidean distance, as defined at (1.2). From now on, we shall assume that $V \subset \mathbf{R}^d$ is given by $V = \mathcal{S}$, or sometimes $V = \mathcal{S}^0$, where $\mathcal{S}$ is generated in a *random* manner and $\mathcal{S}^0 := \mathcal{S} \cup \{\mathbf{0}\}$. The random point set $\mathcal{S}$ will usually be either the set of points given by a homogeneous Poisson point process $\mathcal{P}_n$ of intensity n on the unit cube $(0,1)^d$, or the point process $\mathcal{X}_n$ consisting of n independent random vectors in $\mathbf{R}^d$ with common density function f. In the case where f is the indicator function of $(0,1)^d$, we write $\mathcal{U}_n$ for $\mathcal{X}_n$, now the binomial point process consisting of n independent uniformly distributed random vectors on $(0,1)^d$.

We sometimes consider the MDST on $\mathcal{S}^0$ rather than $\mathcal{S}$ when the partial order is $\preccurlyeq^*$ and $\mathcal{S} = \mathcal{U}_n$ or $\mathcal{S} = \mathcal{P}_n$ (following Bhatt and Roy in the $d = 2$ case).

It follows from Proposition 2.4.2 and the discussion thereafter (see (2.19)) that, under partial order $\preccurlyeq^*$, if $V \subset (0,1)^d$ with set of minimal elements V_0, and $M = (V, E_M)$ is a MDSF on V, we have that $M' = (V \cup \{\mathbf{0}\}, E_{M'})$ is a MDSF (in fact, a MDST) on $V \cup \{\mathbf{0}\}$ if

$$E_{M'} = E_M \cup \bigcup_{v \in V_0} \{(v, \mathbf{0})\}.$$

Note that in this random setting, with probability one each point of $\mathcal{S}$ has a unique directed nearest neighbour, so that V has a unique MDSF, which does not depend on the choice of α.

Denote by $\mathcal{L}^{d,\alpha}(V)$ the total weight of all the edges in the MDSF on $V \subset \mathbf{R}^d$, under weight function w_α, $\alpha > 0$ as given by (1.2). Also (when V is random) let $\tilde{\mathcal{L}}^{d,\alpha}(V) := \mathcal{L}^{d,\alpha}(V) - E[\mathcal{L}^{d,\alpha}(V)]$, the centred total weight.

Our first result presents laws of large numbers for the total edge weight in $d = 2$ for the general partial order $\overset{\theta,\phi}{\preccurlyeq}$ and general $0 < \alpha < 2$. Recall that $\mathcal{X}_n$ is the point process of n independent random points in $\mathbf{R}^d$ with common density f, and that when f is the indicator function of $(0,1)^d$ we write $\mathcal{U}_n = \mathcal{X}_n$ for the binomial point process.

Theorem 2.4.1 *Suppose $d = 2$, and the weight function is w_α as given by (1.2). Suppose $0 < \alpha < 2$, and that the density function f satisfies condition (C1). Under the general partial order $\overset{\theta,\phi}{\preccurlyeq}$, with $0 \leq \theta < 2\pi$ and $0 < \phi \leq \pi$, it is the case that, as $n \to \infty$,*

$$n^{(\alpha/2)-1}\mathcal{L}^{2,\alpha}(\mathcal{X}_n) \xrightarrow{L^1} (2/\phi)^{\alpha/2}\Gamma(1+\alpha/2)\int_{\mathrm{supp}(f)} f(\mathbf{x})^{(2-\alpha)/2}d\mathbf{x}. \tag{2.20}$$

In particular, as $n \to \infty$

$$n^{(\alpha/2)-1}\mathcal{L}^{2,\alpha}(\mathcal{U}_n) \xrightarrow{L^1} (2/\phi)^{\alpha/2}\Gamma(1+\alpha/2), \tag{2.21}$$

and when the partial order is $\preccurlyeq^*$, (2.21) remains true with the addition of the origin, i.e. with $\mathcal{U}_n$ replaced by $\mathcal{U}_n^0$.

Remark. In the special case $\alpha = 1$, the limit in (2.21) is $\sqrt{\pi/(2\phi)}$. This limit is 1 when $\phi = \pi/2$. Also, for $\phi = 2\pi$ we have the standard nearest neighbour (directed) graph (that is, every point is joined to its nearest neighbour by a directed edge), and this limit is then $1/2$. This result (for $\alpha = 1, \phi = 2\pi$) is stated without proof (and attributed to Miles [101]) in [11]. See Theorem 2.2.1 and Remark (b) thereafter.

The next result presents laws of large numbers for the total edge weight in general $d \in \mathbf{N}$ for the special partial orders $\preccurlyeq^*$ and $\preccurlyeq_*$, with $0 < \alpha < d$.

Theorem 2.4.2 *Suppose $d \in \mathbf{N}$, and the weight function is w_α as given by (1.2). Suppose also that $0 < \alpha < d$, and the density function f satisfies condition (C1). Under the partial order $\preccurlyeq^*$ we have that, as $n \to \infty$,*

$$n^{(\alpha/d)-1}\mathcal{L}^{d,\alpha}(\mathcal{X}_n) \xrightarrow{L^1} 2^\alpha\Gamma(1+(\alpha/d))v_d^{-\alpha/d}\int_{\mathrm{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d}d\mathbf{x}. \tag{2.22}$$

In particular, as $n \to \infty$,

$$n^{(\alpha/d)-1}\mathcal{L}^{d,\alpha}(\mathcal{U}_n) \xrightarrow{L^1} 2^\alpha\Gamma(1+(\alpha/d))v_d^{-\alpha/d}, \tag{2.23}$$

and in addition (2.23) remains true with the addition of the origin, i.e. with $\mathcal{U}_n$ replaced by $\mathcal{U}_n^0$. Also, when the partial order is $\preccurlyeq_*$, we have that, as $n \to \infty$,

$$n^{(\alpha/d)-1}\mathcal{L}^{d,\alpha}(\mathcal{X}_n) \xrightarrow{L^1} 2^{\alpha/d}\Gamma(1+(\alpha/d))v_d^{-\alpha/d}\int_{\mathrm{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d}d\mathbf{x}. \tag{2.24}$$

In particular, as $n \to \infty$,

$$n^{(\alpha/d)-1}\mathcal{L}^{d,\alpha}(\mathcal{U}_n) \xrightarrow{L^1} 2^{\alpha/d}\Gamma(1+(\alpha/d))v_d^{-\alpha/d}. \tag{2.25}$$

Remark. In the uniform case, when $d = 2$, $\alpha = 1$ the limits on the right hand side of (2.23) and (2.25) are 1 and $1/\sqrt{2}$ respectively (agreeing with the relevant cases of Theorem 2.4.1). When $d = 1$, the limits on the right hand side of (2.23) and (2.25) both reduce to $\Gamma(1+\alpha)$ (the partial orders both reduce to the total order $\leq$ here).

2.4.3 Record values and rooted vertices in the MDST

Here we consider the minimal elements, under $\preccurlyeq^*$, of random points in $\mathbf{R}^d$. There is a natural link with the theory of record values when $d = 2$, we discuss this here. Note that the minimal elements in the point set $\mathcal{X} \subset (0,1)^d$ are exactly those vertices that are connected to the origin in the MDST on $\mathcal{X} \cup \{\mathbf{0}\}$. This fact will help to explain the appearance of the so-called Dickman distribution in Chapter 4.

For $d \in \mathbf{N}$, let $M_d^*(\mathcal{X})$ denote the number of minimal elements under the ordering $\preccurlyeq^*$ of a point set $\mathcal{X} \subset \mathbf{R}^d$. When $d = 2$, there is a well-known connection (see, for example, [21]) between $M_2^*(\mathcal{X})$ and the *record values* in a sample of size n. Given a sequence of real numbers $(x_1, x_2, \ldots, x_n)$, we say that x_i is a *lower record value* if $x_i \leq x_j$ for $j = 1, \ldots, i-1$ (thus x_1 is always a lower record). Note that often the notion of records is defined with a *strict* inequality. For continuous distributions, as we will consider, the two definitions are equivalent, almost surely.

Let $N_R((x_1, x_2, \ldots, x_n))$ denote the number of lower records in $(x_1, \ldots, x_n) \in \mathbf{R}^n$. The following well-known result gives the connection between M_2^* and N_R. The result is essentially the same as Lemma 2.1 in [21].

Lemma 2.4.1 *Let $\mathcal{Y}_n$ be a finite point set in $\mathbf{R}^2$ with* $\mathrm{card}(\mathcal{Y}_n) = n$*, consisting of points (x_i, y_i), $i = 1, \ldots, n$. Let $((x_1, y_1),\ldots,(x_n, y_n))$ be an arrangement of the points of $\mathcal{Y}_n$, such that $y_1 \leq y_2 \leq \cdots \leq y_n$. Then*

$$M_2^*(\mathcal{Y}_n) = N_R((x_1, x_2, \ldots, x_n)).$$

Proof. To see the relationship between record values and $M_2^*(\cdot)$, consider the second dimension of $\mathbf{R}^2$ to be time. Then, for $\mathbf{x} = (x, t_x)$, $\mathbf{y} = (y, t_y)$ points of $\mathcal{Y}_n$, $\mathbf{x} \preccurlyeq^* \mathbf{y}$ is equivalent to $x \leq y$ and $t_x \leq t_y$. Arrange the points of $\mathcal{Y}_n$ in order of nondecreasing y-coordinate ('time') as $((x_1, t_1), (x_2, t_2), \ldots, (x_n, t_n))$. Then, for x_i, $i \in \mathbf{N}$ a record value of $(x_1, x_2, \ldots, x_n)$, we have $x_i \leq x_j$ for all $1 \leq j \leq i$. That is, $(x_i, t_i) \preccurlyeq^* (x_j, t_j)$ for all j, i.e. (x_i, t_i) is a minimal element of $\mathcal{Y}_n$. Thus $M_2^*(\mathcal{Y}_n)$ is exactly the number of lower records in $(x_1, x_2, \ldots, x_n)$. □

When the components of $(X_1, \ldots, X_n)$ are i.i.d., many properties of the number of records are known, and hence (when $\mathcal{X}_n$ is the point process consisting of n i.i.d. random vectors in $\mathbf{R}^2$) the corresponding properties of $M_d^*(\mathcal{X}_n)$ follow immediately. For example, the following result of [120] (see also Theorems 1.1 and 2.1 of [21]):

Lemma 2.4.2 *Let $(U_1, U_2, \ldots, U_n)$ be a sequence of independent uniform random variables on $(0, 1)$. Let $\mathcal{U}_n$ be a binomial point process on $(0, 1)^2$. Then the following results hold, with $Y_n := N_R((U_1, \ldots, U_n))$ or $Y_n := M_2^*(\mathcal{U}_n)$:*

(i) As $n \to \infty$,

$$(\log n)^{-1} Y_n \xrightarrow{\text{a.s.}} 1.$$

(ii) As $n \to \infty$,

$$(\log n)^{-1/2}(Y_n - \log n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

(iii)

$$\limsup_{n\to\infty} \left[\frac{Y_n - \log n}{(2 \log n)^{1/2} (\log \log \log n)^{1/2}} \right] = 1, \text{ a.s.}$$

(iv)

$$\liminf_{n\to\infty} \left[\frac{Y_n - \log n}{(2 \log n)^{1/2} (\log \log \log n)^{1/2}} \right] = -1, \text{ a.s.}$$

Note that Lemma 2.4.2 holds for general i.i.d. random variables with continuous distribution. For example, that this is true in the N_R case can be seen by the following argument. Let $X_1, \ldots, X_n$ be independent random vectors in $\mathbf{R}$ with common distribution F. Suppose that the inverse $F^{-1}(x)$ of F is well-defined for all $x \in (0, 1)$. Let $U_1, \ldots, U_n$ be i.i.d. uniform on $(0, 1)$. Then $(F^{-1}(U_1), \ldots, F^{-1}(U_n))$ has the same distribution as $(X_1, \ldots, X_n)$ (and any permutation thereof). Also, by the properties of F^{-1}, $N_R((F^{-1}(U_1), \ldots, F^{-1}(U_n)))$ is the same as $N_R((U_1, \ldots, U_n))$. Thus $N_R((X_1, \ldots, X_n))$ has the same distribution as $N_R((U_1, \ldots, U_n))$.

See also [45], p. 45 and p. 102, for proofs of parts (i) and (ii) of Lemma 2.4.2. For further results on records, see, for example [30]. We will revisit record values in Chapter 4, where *sums* of records will be related to some of our Dickman-type distributions.

For this section, we need results about $E[M_d^*(\mathcal{U}_n)]$ for general $d \geq 2$. It is worth noting that $M_d^*(\mathcal{X}_n)$ and related statistics are of independent interest (see e.g. [15]) as a so-called distribution-free quantity – the distribution of $M_d^*(\mathcal{X}_n)$ does not depend on the underlying common (continuous) distribution of the points of $\mathcal{X}_n$, provided that the underlying d-dimensional density f is a pure product, i.e. the coordinates of each point are independent.

From the records argument given above, we can immediately get the following for $d = 2$. For $k \in \mathbf{N}$, and for $a < b$ and $c < d$ let $\mathcal{U}_{k,(a,b]\times(c,d]}$ denote the point process consisting of k independent random vectors uniformly distributed on the rectangle $(a, b] \times (c, d]$. Then

$$E[M_2^*(\mathcal{U}_{k,(a,b]\times(c,d]})] = E[M_2^*(\mathcal{U}_k)] = 1 + (1/2) + \cdots + (1/k) \leq 1 + \log k. \tag{2.26}$$

The first equality in (2.26) comes from some obvious scaling which shows that the distribution of $M_2^*(\mathcal{U}_{k,(a,b]\times(c,d]})$ does not depend on a, b, c, d. For the second equality in (2.26), we note that in an i.i.d. sample from a uniform distribution, the kth observation will be a (lower) record with probability $1/k$; see Theorem 2.3 of [15] or the proof of Theorem 1.1(a) of [21].

The next result deals with $E[M_d^*(\mathcal{U}_n)]$ for general d. The result was first obtained in [15] (Lemma 2.4.3 below is the case $r = 1$ of (3.39) in [15]); an asymptotic expression of the from (2.27) was also given in [89], see the discussion and references in [13]. For more sophisticated asymptotic results on $E[M_d^*(\mathcal{U}_n)]$, see [13]. The asymptotic result in (2.27) is also true (but not very informative) when $d = 1$, for in that case $M_1^*(\mathcal{U}_n) = 1$ almost surely.

Lemma 2.4.3 *let $d \in \{2, 3, \ldots\}$. Then, for $n \in \mathbf{N}$, with $i_1 := n$,*

$$E[M_d^*(\mathcal{U}_n)] = \sum_{i_2=1}^{n} \sum_{i_3=1}^{i_2} \cdots \sum_{i_d=1}^{i_{d-1}} \frac{1}{i_2 i_3 \cdots i_d} = \frac{(\log n)^{d-1}}{(d-1)!} + O((\log n)^{d-2}), \tag{2.27}$$

as $n \to \infty$.

Proof. The idea is similar to that used in [15], but here we proceed by induction. Assume that (2.27) holds for $d = k$ for some $k \in \{2, 3, \ldots\}$. Let $(\mathbf{U}_1, \ldots, \mathbf{U}_n)$ be a list of the points of $\mathcal{U}_n$ in $\mathbf{R}^{k+1}$, in order of nondecreasing first co-ordinate. Write $\mathbf{U}_i = (U_1^i, U_2^i, \ldots, U_{k+1}^i)$, $i = 1, \ldots, n$. Then $U_1^1 \leq U_1^2 \leq \cdots \leq U_1^n$. We have that $\mathbf{U}_i \preccurlyeq^* \mathbf{U}_j$ in $(0,1)^{k+1}$ is then equivalent to $i \leq j$ and $(U_2^i, \ldots, U_n^i) \preccurlyeq^* (U_2^j, \ldots, U_n^j)$ in $(0,1)^k$. That is, $\mathbf{U}_j$ is a minimal element of $\mathcal{U}_n$ in $(0,1)^{k+1}$ if and only if $(U_2^j, \ldots, U_n^j)$ is a minimal element of $\{(U_2^i, \ldots, U_n^i),\ 1 \leq i \leq j\}$. Thus the probability that $\mathbf{U}_j$ is minimal in $\mathcal{U}_n$ is the same as the probability that a randomly chosen element of j i.i.d. uniform random vectors in $(0,1)^k$ is minimal. By assumption, that probability is

$$\frac{1}{j} \sum_{i_2=1}^{j} \sum_{i_3=1}^{i_2} \cdots \sum_{i_k=1}^{i_{k-1}} \frac{1}{i_2 i_3 \cdots i_k}.$$

Thus

$$\begin{aligned} E[M_{k+1}^*(\mathcal{U}_n)] &= \sum_{j=1}^{n} \frac{1}{j} \sum_{i_2=1}^{j} \sum_{i_3=1}^{i_2} \cdots \sum_{i_k=1}^{i_{k-1}} \frac{1}{i_2 i_3 \cdots i_k} \\ &= \sum_{i_2=1}^{n} \sum_{i_3=1}^{i_2} \cdots \sum_{i_{k+1}=1}^{i_k} \frac{1}{i_2 i_3 \cdots i_k i_{k+1}}, \end{aligned}$$

after a suitable relabelling. Hence, if (2.27) holds for $d = k$, it also holds for $d = k+1$. Further, (2.27) holds for $d = 2$ by a simple argument (see (2.26)). Hence the exact statement in (2.27) follows for $d \in \{2, 3, \ldots\}$. The asymptotic statement then follows easily (or can be seen to hold inductively by the above argument). □

Chapter 3

General results in geometric probability and proofs of laws of large numbers

Notions of *stabilizing* functionals of point sets have recently proved to be a useful basis for a general methodology for establishing limit theorems for functionals of random point sets in $\mathbf{R}^d$. In particular, Penrose and Yukich [111,113] provide general central limit theorems and laws of large numbers for stabilizing functionals. In fact we shall obtain our laws of large numbers presented in Chapter 2 (Theorems 2.2.1, 2.2.2, 2.3.1 (i), 2.4.1, and 2.4.2) by application of a result from [113]. Also, one might hope to apply the results in [111] to obtain the central limit theorem (see Chapter 6) for edges away from the boundary in the MDSF and MDST. However, the results of [111] are not directly applicable in this case; we need an extension of the general result in [111]. It is these general results that we describe in the present section.

For our general results, we use the following notation. Let $d \in \mathbf{N}$. For $\mathcal{X} \subset \mathbf{R}^d$, constant $a > 0$, and $\mathbf{y} \in \mathbf{R}^d$, let $\mathbf{y} + a\mathcal{X}$ denote the transformed set $\{\mathbf{y} + a\mathbf{x} : \mathbf{x} \in \mathcal{X}\}$. Let $\text{diam}(\mathcal{X}) := \sup\{\|\mathbf{x}_1 - \mathbf{x}_2\| : \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}\}$, and recall that $\text{card}(\mathcal{X})$ denotes the cardinality of $\mathcal{X}$. Recall that $\mathbf{0}$ denotes the origin of $\mathbf{R}^d$.

For $\mathbf{x} \in \mathbf{R}^d$ and $r > 0$, let $B(\mathbf{x}; r)$ denote the closed Euclidean ball with centre $\mathbf{x}$ and radius r, and let $Q(\mathbf{x}; r)$ denote the corresponding l_∞ ball, i.e., the d-cube $\mathbf{x} + [-r, r]^d$. For bounded measurable $R \subset \mathbf{R}^d$ let $|R|$ denote the Lebesgue measure of R, let ∂R denote the topological boundary of R and for $r > 0$, set $\partial_r R := \cup_{\mathbf{x} \in \partial R} Q(\mathbf{x}; r)$, the r-neighbourhood of the boundary of R.

3.1 A general law of large numbers

Let $\xi(\mathbf{x};\mathcal{X})$ be a measurable $\mathbf{R}_+$-valued function defined for all pairs $(\mathbf{x},\mathcal{X})$, where $\mathcal{X} \subset \mathbf{R}^d$ is finite and $\mathbf{x} \in \mathcal{X}$. Assume ξ is *translation invariant*, that is, for all $\mathbf{y} \in \mathbf{R}^d$, $\xi(\mathbf{y}+\mathbf{x};\mathbf{y}+\mathcal{X}) = \xi(\mathbf{x};\mathcal{X})$. When $\mathbf{x} \notin \mathcal{X}$, we abbreviate the notation $\xi(\mathbf{x};\mathcal{X}\cup\{\mathbf{x}\})$ to $\xi(\mathbf{x};\mathcal{X})$.

For our general law of large numbers, we use a notion of stabilization defined as follows. For any locally finite point set $\mathcal{X} \subset \mathbf{R}^d$ and any $\ell \in \mathbf{N}$ define

$$\xi^+(\mathcal{X};\ell) := \sup_{k\in\mathbf{N}} \left(\operatorname*{ess\,sup}_{\ell,k} \{\xi(\mathbf{0};(\mathcal{X}\cap B(\mathbf{0};\ell))\cup\mathcal{A}\} \right), \text{ and}$$

$$\xi^-(\mathcal{X};\ell) := \inf_{k\in\mathbf{N}} \left(\operatorname*{ess\,inf}_{\ell,k} \{\xi(\mathbf{0};(\mathcal{X}\cap B(\mathbf{0};\ell))\cup\mathcal{A}\} \right);$$

where $\operatorname{ess\,sup}_{\ell,k}$ is the essential supremum, with respect to Lebesgue measure on $\mathbf{R}^{dk}$, over sets $\mathcal{A} \subset \mathbf{R}^d \backslash B(\mathbf{0};\ell)$ of cardinality k. Define the *limit* of ξ on $\mathcal{X}$ by

$$\xi_\infty(\mathcal{X}) := \limsup_{k\to\infty} \xi^+(\mathcal{X};k).$$

We say the functional ξ *stabilizes* on $\mathcal{X}$ if

$$\lim_{k\to\infty} \xi^+(\mathcal{X};k) = \lim_{k\to\infty} \xi^-(\mathcal{X};k) = \xi_\infty(\mathcal{X}). \tag{3.1}$$

We say that ξ is *homogeneous of order* α if for all finite $\mathcal{X} \subset \mathbf{R}^d$ and all $r \in (0,\infty)$, $\xi(r\mathbf{x};r\mathcal{X}) = r^\alpha \xi(\mathbf{x};\mathcal{X})$.

For $\tau \in (0,\infty)$, let $\mathcal{H}_\tau$ be a homogeneous Poisson process of intensity τ on $\mathbf{R}^d$. The following general law of large numbers is due to Penrose and Yukich, and is obtained from Theorem 2.1 of [113] together with equation (2.9) of [113] (the homogeneous case). We shall use it to prove our Theorems 2.2.1, 2.2.2, 2.3.1 (i), 2.4.1, and 2.4.2.

Lemma 3.1.1 *[113] Suppose $q = 1$ or $q = 2$. Suppose ξ is almost surely stabilizing on $\mathcal{H}_1$, with limit $\xi_\infty(\mathcal{H}_1)$, and suppose that ξ is homogeneous of order α. Let f be a probability density function on $\mathbf{R}^d$, and let $\mathcal{X}_n$ be the point process consisting of n independent random d-vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ with common density f. If ξ satisfies the moments condition*

$$\sup_{n\in\mathbf{N}} E\left[\xi\left(n^{1/d}\mathbf{X}_1; n^{1/d}\mathcal{X}_n\right)^p\right] < \infty, \tag{3.2}$$

for some $p > q$, then as $n \to \infty$,

$$n^{-1}\sum_{\mathbf{x}\in\mathcal{X}_n} \xi(n^{1/d}\mathbf{x}; n^{1/d}\mathcal{X}_n) \xrightarrow{L^q} E[\xi_\infty(\mathcal{H}_1)] \int_{\operatorname{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d}\,d\mathbf{x},$$

and the limit is finite.

3.2 General central limit theorems

In the course of the proof of Theorem 6.1.1, we shall use a modified form of a general central limit theorem obtained for functionals of geometric graphs by Penrose and Yukich [111]. We recall the setup of [111]. As in Section 3.1, let $\xi(\mathbf{x}; \mathcal{X})$ be a translation invariant real-valued functional defined for finite $\mathcal{X} \subset \mathbf{R}^d$ and $\mathbf{x} \in \mathcal{X}$. Then ξ induces a translation invariant functional $H(\mathcal{X}; S)$ defined on all finite point sets $\mathcal{X} \subset \mathbf{R}^d$ and all Borel-measurable regions $S \subseteq \mathbf{R}^d$ by

$$H(\mathcal{X}; S) := \sum_{\mathbf{x} \in \mathcal{X} \cap S} \xi(\mathbf{x}; \mathcal{X}). \tag{3.3}$$

It is this 'restricted' functional that interests us here, while [111] is concerned rather with the global functional $H(\mathcal{X}; \mathbf{R}^d)$. In our particular application (the length of edges of the MDST on random points in a square), the global functional fails to satisfy the conditions of the central limit theorems in [111], owing to boundary effects. Here we generalize the result in [111] to the 'restricted' functional $H(\mathcal{X}; S)$. It is this generalized result that we can apply to the MDST, when we take S to be a region 'away from the boundary' of the square in which the random points are placed.

We use a notion of stabilization for H which is related to, but not equivalent to, the notion of stabilization of ξ used in Section 3.1. Loosely speaking, ξ is stabilizing if when a point inserted at the origin into a homogeneous Poisson process, only nearby Poisson points affect the inserted point; for H to be stabilizing we require also that the inserted point affects only nearby points.

For $B \subseteq \mathbf{R}^d$, let $\Delta(\mathcal{X}; B)$ denote the 'add one cost' of the functional H on the insertion of a point at the origin,

$$\Delta(\mathcal{X}; B) := H(\mathcal{X} \cup \{\mathbf{0}\}; B) - H(\mathcal{X}; B).$$

Let $\mathcal{P} := \mathcal{H}_1$ (a homogeneous Poisson point process of unit intensity on $\mathbf{R}^d$). Let $\mathcal{Q}_n := \mathcal{P} \cap R_n$ (the restriction of $\mathcal{P}$ to R_n). Adapting the ideas of [111], we make the following definitions.

Definition 3.2.1 *We say the functional H is* strongly stabilizing *if there exist almost surely finite random variables R (a* radius of stabilization*) and $\Delta(\infty)$ such that, with probability 1, for any $B \supseteq B(\mathbf{0}; R)$,*

$$\Delta(\mathcal{P} \cap B(\mathbf{0}; R) \cup \mathcal{A}; B) = \Delta(\infty), \ \forall \textit{ finite } \mathcal{A} \subset \mathbf{R}^d \setminus B(\mathbf{0}; R).$$

We say that the functional H is *polynomially bounded* if, for all $B \ni \mathbf{0}$, there exists a constant β such that for all finite sets $\mathcal{X} \subset \mathbf{R}^d$,

$$|H(\mathcal{X}; B)| \leq \beta\,(\mathrm{diam}(\mathcal{X}) + \mathrm{card}(\mathcal{X}))^{\beta}. \tag{3.4}$$

We say that H is *homogeneous of order* α if for all finite $\mathcal{X} \subset \mathbf{R}^d$ and Borel $B \subseteq \mathbf{R}^d$, and all $r \in \mathbf{R}$, $H(r\mathcal{X}; rB) = r^{\alpha} H(\mathcal{X}; B)$.

Let (R_n, S_n), for $n = 1, 2, \ldots$, be a sequence of ordered pairs of bounded Borel subsets of $\mathbf{R}^d$, such that $S_n \subseteq R_n$ for all n. Assume that for all $r > 0$, $n^{-1}|\partial_r R_n| \to 0$ and $n^{-1}|\partial_r S_n| \to 0$ (the *vanishing relative boundary condition*). Assume also that $|R_n| = n$ for all n, and $|S_n|/n \to 1$ as $n \to \infty$; that S_n tends to $\mathbf{R}^d$, in the sense that $\cup_{n\geq 1} \cap_{m \geq n} S_m = \mathbf{R}^d$; and that there exists a constant β such that $\mathrm{diam}(R_n) \leq \beta n^{\beta}$ for all n (the *polynomial boundedness* condition on $(R_n, S_n)_{n\geq 1}$). Subject to these conditions, the choice of $(R_n, S_n)_{n\geq 1}$ is arbitrary.

Let $\mathbf{U}_{1,n}, \mathbf{U}_{2,n}, \ldots$ be i.i.d. uniform random vectors on R_n. Let

$$\mathcal{U}_{m,n} = \{\mathbf{U}_{1,n}, \ldots, \mathbf{U}_{m,n}\}$$

(a binomial point process), and for Borel $A \subseteq \mathbf{R}^d$ with $0 < |A| < \infty$, let $\mathcal{U}_{m,A}$ be the binomial point process of m i.i.d. uniform random vectors on A.

Let $\mathcal{R}$ be the collection of all pairs (A, B) with $A, B \subset \mathbf{R}^d$ of the form $(A, B) = (\mathbf{x} + R_n, \mathbf{x} + S_n)$ with $\mathbf{x} \in \mathbf{R}^d$ and $n \in \mathbf{N}$. That is, $\mathcal{R}$ is the collection of all the (R_n, S_n) and their translates.

We say that the functional H satisfies the *uniform bounded moments condition* on $\mathcal{R}$ if

$$\sup_{(A,B)\in\mathcal{R}:\mathbf{0}\in A} \left(\sup_{|A|/2 \leq m \leq 3|A|/2} \{E[\Delta(\mathcal{U}_{m,A}; B)^4]\} \right) < \infty. \tag{3.5}$$

We now state the general results, which extend those of Penrose and Yukich (Theorem 2.1 and Corollary 2.1 in [111]).

Theorem 3.2.1 *Suppose that H is strongly stabilizing, is polynomially bounded (3.4), and satisfies the uniform bounded moments condition (3.5) on $\mathcal{R}$. Then there exist constants s^2, t^2, with $0 \leq t^2 \leq s^2$, such that as $n \to \infty$,*

(i) $n^{-1}\mathrm{Var}\,(H(\mathcal{Q}_n; S_n)) \to s^2$;

(ii) $n^{-1/2}\,(H(\mathcal{Q}_n; S_n) - E[H(\mathcal{Q}_n; S_n)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2)$;

(iii) $n^{-1}\mathrm{Var}\left(H\left(\mathcal{U}_{n,n}; S_n\right)\right) \to t^2$;

(iv) $n^{-1/2}\left(H\left(\mathcal{U}_{n,n}; S_n\right) - E\left[H\left(\mathcal{U}_{n,n}; S_n\right)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, t^2)$.

Also, s^2 and t^2 are independent of the choice of the (R_n, S_n). Further, if the distribution of $\Delta(\infty)$ is nondegenerate, then $s^2 \geq t^2 > 0$.

We present the proof of Theorem 3.2.1, which is similar to the proof of Theorem 2.1 in [111], in Appendix B.

Let R_0 be a fixed bounded Borel subset of $\mathbf{R}^d$ with $|R_0| = 1$ and $|\partial R_0| = 0$. Let $(S_{0,n}, n \geq 1)$ be a sequence of Borel sets with $S_{0,n} \subseteq R_0$ such that $|S_{0,n}| \to 1$ as $n \to \infty$ and for all $r > 0$ we have $|\partial_{n^{-1/d}r} S_{0,n}| \to 0$ as $n \to \infty$

Let $\mathcal{R}_0$ be the collection of all pairs of the form $(\mathbf{x} + n^{1/d}R_0, \mathbf{x} + n^{1/d}S_{0,n})$ with $n \geq 1$ and $\mathbf{x} \in \mathbf{R}^d$. Let $\mathcal{U}_n$ be the binomial point process of n i.i.d. uniform random vectors on R_0, and let $\mathcal{P}_n$ be a homogeneous Poisson point process of intensity n on R_0.

Corollary 3.2.1 *Suppose H is strongly stabilizing, satisfies the uniform bounded moments condition on $\mathcal{R}_0$, is polynomially bounded and is homogeneous of order α. Then with s^2, t^2 as in Theorem 3.2.1 we have that, as $n \to \infty$*

(i) $n^{(2\alpha/d)-1}\mathrm{Var}\left(H\left(\mathcal{P}_n; S_{0,n}\right)\right) \to s^2$;

(ii) $n^{(\alpha/d)-1/2}\left(H\left(\mathcal{P}_n; S_{0,n}\right) - E\left[H\left(\mathcal{P}_n; S_{0,n}\right)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, s^2\right)$;

(iii) $n^{(2\alpha/d)-1}\mathrm{Var}\left(H\left(\mathcal{U}_n; S_{0,n}\right)\right) \to t^2$;

(iv) $n^{(\alpha/d)-1/2}\left(H\left(\mathcal{U}_n; S_{0,n}\right) - E\left[H\left(\mathcal{U}_n; S_{0,n}\right)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, t^2\right)$.

Proof. The corollary follows from Theorem 3.2.1 by taking $R_n = n^{1/d}R_0$ and $S_n = n^{1/d}S_{0,n}$ (or suitable translates thereof), and scaling, since H is homogeneous of order α. □

3.3 Proofs of laws of large numbers

In this section we prove the laws of large numbers given in Theorems 2.2.1, 2.2.2, 2.3.1 (i), 2.4.1 and 2.4.2, by applying the general result presented in Section 3.1. Although not a LLN result, we also prove (in Section 3.3.3) Theorem 2.3.1 (ii) and (iii).

3.3.1 Proof of Theorems 2.2.1 and 2.2.2

We now derive our law of large numbers for the total length of the random k-NNG$'$ and j-th NNG$'$, and the nearest neighbour (undirected) graph.

We apply Lemma 3.1.1 to obtain a law of large numbers for $\mathcal{N}^{d,\alpha}_{\leq k}(\mathcal{X}_n)$, with certain conditions on f, α. This method enables us to evaluate the limit explicitly, unlike methods based on the subadditivity of the functional (see [142]).

For $j \in \mathbf{N}$, let $d_j(\mathbf{x}; \mathcal{X})$ be the (Euclidean) distance from $\mathbf{x} \in \mathcal{X}$ to its j-th nearest neighbour, if such a neighbour exists, or zero otherwise. For each j, d_j is translation invariant.

In the proofs that follow, we make use of the following integral. For a, b, c nonnegative real constants,

$$\int_0^\infty r^a e^{-cr^b} \mathrm{d}r = \frac{1}{b} c^{-(a+1)/b} \Gamma\left((a+1)/b\right). \tag{3.6}$$

Proof of Theorem 2.2.1. In applying Lemma 3.1.1 to the j-th NNG$'$ and k-NNG$'$ functionals, we take $\xi(\mathbf{x}; \mathcal{X}_n)$ to be $[d_j(\mathbf{x}; \mathcal{X}_n)]^\alpha$, where $\alpha \geq 0$. Note that ξ is translation invariant and homogeneous of order α. It was shown in [113] (see Theorem 2.4 in [113]) that the j-th NNG$'$ total weight functional ξ satisfies the conditions of Lemma 3.1.1 in the following two cases: (i) with $q = 2$, if the function f satisfies condition (C1), and $\alpha \geq 0$; and (ii) with $q = 1$, if the function f satisfies condition (C2), and $0 \leq \alpha < d$. (In fact, in [113] this is proved for the k-NNG$'$ functional $\sum_{j=1}^k [d_j(\mathbf{x}; \mathcal{X}_n)]^\alpha$, but this implies that the conditions also hold for the j-th NNG$'$ functional $[d_j(\mathbf{x}; \mathcal{X}_n)]^\alpha$).

First we take ξ to be the j-th NNG$'$ functional $\xi(\mathbf{x}; \mathcal{X}_n) = [d_j(\mathbf{x}; \mathcal{X}_n)]^\alpha$. Our functional ξ is stabilizing on $\mathcal{H}_1$, with limit $\xi_\infty(\mathcal{H}_1) = [d_j(\mathbf{0}; \mathcal{H}_1)]^\alpha$. Also, the moment condition (3.2) is satisfied for some $p > 1$ (if f satisfies condition (C2) and $\alpha < d$) or $p > 2$ (if f satisfies (C1)), and so we can apply Lemma 3.1.1 with $q = 1$ or $q = 2$ respectively. So by Lemma 3.1.1, using the fact that ξ is homogeneous of order α, we have that

$$\begin{aligned} n^{(\alpha/d)-1} \mathcal{N}^{d,\alpha}_j(\mathcal{X}_n) &= n^{-1} \sum_{\mathbf{x} \in \mathcal{X}_n} \xi(n^{1/d}\mathbf{x}; n^{1/d}\mathcal{X}_n) \\ &\xrightarrow{L^q} E[\xi_\infty(\mathcal{H}_1)] \int_{\mathrm{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d} \mathrm{d}\mathbf{x}. \end{aligned} \tag{3.7}$$

We now need to evaluate the expectation on the right hand side of (3.7). With v_d the volume of the unit d-ball as given by (2.1),

$$\begin{aligned} P\left[\xi_\infty(\mathcal{H}_1) > r\right] &= P\left[d_j(\mathbf{0};\mathcal{H}_1) > r^{1/\alpha}\right] = \sum_{i=0}^{j-1} P\left[\mathrm{card}(\{B(\mathbf{0};r^{1/\alpha}) \cap \mathcal{H}_1\}) = i\right] \\ &= \sum_{i=0}^{j-1} \exp\left(-v_d r^{d/\alpha}\right) \frac{(v_d r^{d/\alpha})^i}{i!}. \end{aligned}$$

So

$$E\left[\xi_\infty(\mathcal{H}_1)\right] = \int_0^\infty P\left[\xi_\infty(\mathcal{H}_1) > r\right] \mathrm{d}r = \int_0^\infty \sum_{i=0}^{j-1} \exp\left(-v_d r^{d/\alpha}\right) \frac{(v_d r^{d/\alpha})^i}{i!} \mathrm{d}r.$$

Then, interchanging the order of summation and integration, and using (3.6), we obtain

$$\begin{aligned} E\left[\xi_\infty(\mathcal{H}_1)\right] &= \sum_{i=0}^{j-1} v_d^i \frac{\alpha}{di!} v_d^{\left(-\frac{di+\alpha}{d}\right)} \Gamma\left(\frac{di+\alpha}{d}\right) \\ &= v_d^{-\alpha/d} \sum_{i=1}^{j} \frac{\alpha}{d} \frac{\Gamma\left(i-1+(\alpha/d)\right)}{(i-1)!} = v_d^{-\alpha/d} \frac{\Gamma\left(j+(\alpha/d)\right)}{\Gamma(j)}. \end{aligned} \tag{3.8}$$

Now from (3.8), (2.1) and (3.7) we obtain the j-th NNG′ result (2.11).

By (2.2), the k-NNG′ result (2.9) follows from (2.11) with

$$C(d,\alpha,k) = v_d^{-\alpha/d} \sum_{j=1}^{k} \frac{\Gamma\left(j+(\alpha/d)\right)}{\Gamma(j)} = v_d^{-\alpha/d} \frac{d}{d+\alpha} \frac{\Gamma(k+1+(\alpha/d))}{\Gamma(k)}. \quad \square$$

Proof of Theorem 2.2.2. Observe first that the nearest neighbour (directed) graph counts the weights of edges from points that are nearest neighbours of their own nearest neighbours twice, while the nearest neighbour (undirected) graph counts such weights only once. Hence the total weight of the undirected graph is given by the weight of the directed graph, minus half of the contribution to the directed graph from edges between points that are mutual nearest neighbours.

Let $q(\mathbf{x};\mathcal{X})$ be the functional that is the distance from $\mathbf{x} \in \mathcal{X}$ to its nearest neighbour in $\mathcal{X} \setminus \{\mathbf{x}\}$ if $\mathbf{x}$ is a nearest neighbour of its own nearest neighbour, and zero otherwise. Recall that $d_1(\mathbf{x};\mathcal{X})$ is the distance from $\mathbf{x} \in \mathcal{X}$ to its nearest neighbour. Then define

$$\xi'(\mathbf{x};\mathcal{X}) = [d_1(\mathbf{x};\mathcal{X})]^\alpha - \frac{1}{2}[q(\mathbf{x};\mathcal{X})]^\alpha.$$

Then $\sum_{\mathbf{x}\in\mathcal{X}} \xi'(\mathbf{x},\mathcal{X})$ is the total weight of the edges in the standard nearest neighbour (directed) graph on $\mathcal{X}$, minus half of the total weight from points that are nearest neighbours of their own nearest neighbours; this is then the total weight of the nearest neighbour (undirected) graph. Note that ξ' is translation invariant, and it is homogeneous of order α.

One can check that the functional ξ' is stabilizing on the Poisson process $\mathcal{H}_1$, using similar arguments to those for the j-th NNG$'$ and k-NNG$'$ functionals. Also (see [113]) if condition (C1) holds then ξ' satisfies the moments condition (3.2) for some $p > 2$, for all $\alpha > 0$.

Let $\mathbf{e}_1$ be a vector of unit length in $\mathbf{R}^d$. For $d \in \mathbf{N}$, let $\omega_d := |B(\mathbf{0};1) \cup B(\mathbf{e}_1;1)|$, the volume of the union of two unit d-balls with centres unit distance apart.

Now we apply Lemma 3.1.1 with $q = 2$. We have

$$\begin{aligned} n^{(\alpha/d)-1}\mathcal{Z}^{d,\alpha}(\mathcal{X}_n) &= n^{-1}\sum_{\mathbf{x}\in\mathcal{X}_n} \xi'(n^{1/d}\mathbf{x}; n^{1/d}\mathcal{X}_n) \\ &\xrightarrow{L^2} E[\xi'_\infty(\mathcal{H}_1)]\int_{\mathrm{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d}\mathrm{d}\mathbf{x}, \end{aligned} \tag{3.9}$$

where

$$E[\xi'_\infty(\mathcal{H}_1)] = E[(d_1(\mathbf{0};\mathcal{H}_1))^\alpha] - \frac{1}{2}E[(q(\mathbf{0};\mathcal{H}_1))^\alpha].$$

We need to evaluate $E[(q(\mathbf{0};\mathcal{H}_1))^\alpha]$. Let $\mathbf{X}$ denote the nearest point of $\mathcal{H}_1$ to $\mathbf{0}$. Then

$$\begin{aligned} P[q(\mathbf{0};\mathcal{H}_1) \in \mathrm{d}r] &= P[\{|\mathbf{X}| \in \mathrm{d}r\} \cap \{\mathcal{H}_1 \cap (B(\mathbf{0};r) \cup B(\mathbf{X};r)) = \{\mathbf{X}\}\}] \\ &= d v_d r^{d-1} e^{-v_d r^d} e^{-(\omega_d - v_d) r^d}\mathrm{d}r = d v_d r^{d-1} e^{-\omega_d r^d}\mathrm{d}r. \end{aligned}$$

So using (3.6) we obtain

$$E[(q(\mathbf{0};\mathcal{H}_1))^\alpha] = \int_0^\infty d v_d r^{d-1+\alpha} e^{-\omega_d r^d}\mathrm{d}r = v_d \omega_d^{-1-(\alpha/d)}\Gamma(1 + (\alpha/d)). \tag{3.10}$$

Then from (3.9) with (3.10) and the $j = 1$ case of (3.8) we obtain (2.12). By some calculus, we obtain $\omega_2 = (4\pi/3) + (\sqrt{3}/2)$ (the area of the union of two unit discs with centres unit distance apart in $\mathbf{R}^2$), which yields (2.13) and (2.14) from the $d = 2$ case of (2.12). Finally, we obtain the statement for $\mathcal{Z}^{1,1}(\mathcal{U}_n)$ from the $d = 1$ case of (2.12) with the fact that $\omega_1 = 3$. □

3.3.2 Proof of Theorem 2.3.1 (i)

We now derive our law of large numbers for the total weight of the random ONG($\mathcal{U}_n$), where $\mathcal{U}_n$ is the binomial point process of n independent uniformly distributed points $(\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n)$ on $(0,1)^d$, for $d = 1, 2, \ldots$, each point $\mathbf{U}_i$ bearing a mark $T(\mathbf{U}_i)$ distributed uniformly in the interval $[0,1]$, independently of the other marks and the distribution of the points.

We apply Lemma 3.1.1 to obtain a law of large numbers for $\mathcal{O}^{d,\alpha}(\mathcal{U}_n)$, $0 < \alpha < d$. Once again, the method enables us to evaluate the limit explicitly. In applying Lemma 3.1.1 to the ONG functional, we take $f(\mathbf{x})$ (the underlying probability density function in the lemma) to be 1 for $\mathbf{x} \in (0,1)^d$ and zero elsewhere. Define $D(\mathbf{x};\mathcal{X})$ to be the distance from point $\mathbf{x}$ with mark $T(\mathbf{x})$ to its nearest neighbour in $\mathcal{X}$ amongst those points $\mathbf{y} \in \mathcal{X}$ that have mark $T(\mathbf{y})$ such that $T(\mathbf{y}) < T(\mathbf{x})$, if such a neighbour exists, or zero otherwise. We take $\xi(\mathbf{x};\mathcal{X})$ to be $[D(\mathbf{x};\mathcal{X})]^\alpha$. Note that again ξ is translation invariant and homogeneous of order α.

Lemma 3.3.1 *The* ONG *functional ξ is stabilizing on $\mathcal{H}_1$.*

Proof. Although the notion of stabilization there is somewhat different, the lemma follows by the same argument as given at the start of the proof of Theorem 3.6 of [106]. □

Lemma 3.3.2 *Let $0 \leq \alpha < d$, and let $p > 1$ with $\alpha p < d$. Then the* ONG *functional ξ satisfies the moments condition (3.2).*

Proof. Let U_n denote the rank of the mark of $\mathbf{U}_1$ amongst the marks of all the points of $\mathcal{U}_n$, so that U_n is distributed uniformly over the integers $1, 2, \ldots, n$. Setting $S_{d,n} := (0, n^{1/d})^d$, we have, by conditioning on U_n,

$$\begin{aligned} E\left[\xi(n^{1/d}\mathbf{U}_1; n^{1/d}\mathcal{U}_n)^p\right] &= n^{-1}\sum_{i=1}^{n} E\left[d_1(n^{1/d}\mathbf{U}_1; n^{1/d}\mathcal{U}_i)^{p\alpha}\right] \\ &= n^{-1}\sum_{i=1}^{n}(n/i)^{p\alpha/d} E\left[d_1(i^{1/d}\mathbf{U}_1, i^{1/d}\mathcal{U}_i)^{p\alpha}\right]. \end{aligned} \quad (3.11)$$

The last expectation in (3.11) is bounded by a constant independent of i, by the argument for equation (6.4) of [106]. Hence the final expression in (3.11) is bounded by a constant times

$$n^{(p\alpha-d)/d}\sum_{i=1}^{n} i^{-p\alpha/d},$$

which is bounded by a constant. □

Proof of Theorem 2.3.1 (i). Let $d \in \mathbf{N}$. Set $f(\cdot)$ to be the indicator of the unit cube $(0,1)^d$. Take ξ to be the ONG functional $\xi(\mathbf{x};\mathcal{U}_n) = [D(\mathbf{x};\mathcal{U}_n)]^\alpha$. By Lemmas 3.3.1 and 3.3.2, with this choice of f, our functional ξ is homogeneous of order α, stabilizing on $\mathcal{H}_1$, with limit $\xi_\infty = [D(\mathbf{0}; \mathcal{H}_1)]^\alpha$, and satisfies the moment condition (3.2) for some $p > 1$, and so we can apply Lemma 3.1.1 with $q = 1$. So by Lemma 3.1.1, we have that

$$n^{(\alpha/d)-1}\mathcal{O}^{d,\alpha}(\mathcal{U}_n) = n^{-1}\sum_{\mathbf{x}\in\mathcal{U}_n}[D(n^{1/d}\mathbf{x}; n^{1/d}\mathcal{U}_n)]^\alpha \xrightarrow{L^1} E[\xi_\infty(\mathcal{H}_1)].$$

Now, for $0 < u < 1$, the points of $\mathcal{H}_1$ with lower mark than u form a homogeneous Poisson point process of intensity u, and hence by conditioning on the mark of the point at $\mathbf{0}$, we have

$$\begin{aligned} E[\xi_\infty(\mathcal{H}_1)] &= \int_0^1 E\left[d_1(\mathbf{0}; \mathcal{H}_u)^\alpha\right] \mathrm{d}u \\ &= \int_0^1 u^{-\alpha/d} E\left[d_1(\mathbf{0}; \mathcal{H}_1)^\alpha\right] \mathrm{d}u = \frac{d}{d-\alpha} C(d,\alpha,1), \end{aligned}$$

since we saw in the proof of Theorem 2.2.1 that $E[d_1(\mathbf{0}; \mathcal{H}_1)^\alpha] = C(d, \alpha, 1)$. □

3.3.3 Proof of Theorem 2.3.1 (ii) and (iii)

Suppose $d \in \mathbf{N}$. For $n \in \mathbf{N}$, let $Z_n(d) := \mathcal{O}^{d,1}(\mathcal{U}_n) - \mathcal{O}^{d,1}(\mathcal{U}_{n-1})$, setting $\mathcal{O}^{d,1}(\mathcal{U}_0) := 0$. That is, $Z_n(d)$ is the gain in length of the ONG on a sequence of independent uniform random points in $(0,1)^d$ on the addition of the nth point. Again we denote by v_d the volume of the unit d-ball (see (2.1)).

First we need a technical lemma.

Lemma 3.3.3 *Suppose $c \in (0,\infty)$, $\beta \in (0,\infty)$, $x \in (0,\infty)$. Then, as $cx^\beta \to \infty$*

$$\int_0^x \exp(-ct^\beta)\mathrm{d}t = \frac{c^{-1/\beta}}{\beta}\Gamma\left(\frac{1}{\beta}\right) + O\left(c^{-1/\beta}(cx^\beta)^{(1/\beta)-1}\exp(-cx^\beta)\right). \tag{3.12}$$

Proof. By the change of variable $y = ct^\beta$, we have that

$$\begin{aligned} \int_0^x \exp(-ct^\beta)\mathrm{d}t &= \frac{c^{-1/\beta}}{\beta}\int_0^{cx^\beta} y^{(1/\beta)-1}e^{-y}\mathrm{d}y \\ = \frac{c^{-1/\beta}}{\beta}\int_0^\infty y^{(1/\beta)-1}e^{-y}\mathrm{d}y &- \frac{c^{-1/\beta}}{\beta}\int_{cx^\beta}^\infty y^{(1/\beta)-1}e^{-y}\mathrm{d}y. \end{aligned} \tag{3.13}$$

This first integral in the last line of (3.13) is given by (see e.g. [1], 6.1.1)

$$\frac{c^{-1/\beta}}{\beta}\int_0^\infty y^{(1/\beta)-1}e^{-y}\mathrm{d}y = \frac{c^{-1/\beta}}{\beta}\Gamma(1/\beta).$$

For the second integral in the last line of (3.13), we have, for $\beta \geq 1$,

$$\begin{aligned} \frac{c^{-1/\beta}}{\beta}\int_{cx^\beta}^\infty y^{(1/\beta)-1}e^{-y}\mathrm{d}y &\leq \frac{c^{-1/\beta}}{\beta}(cx^\beta)^{(1/\beta)-1}\int_{cx^\beta}^\infty e^{-y}\mathrm{d}y \\ &= \frac{c^{-1/\beta}}{\beta}(cx^\beta)^{(1/\beta)-1}e^{-cx^\beta}, \end{aligned}$$

as required. Now suppose $\beta \in (0,1)$. Then $\beta \in [\frac{1}{m+1}, \frac{1}{m})$ for some $m \in \mathbf{N}$, and m-fold integration by parts yields

$$\frac{c^{-1/\beta}}{\beta} \int_{cx^\beta}^{\infty} y^{(1/\beta)-1} e^{-y} \mathrm{d}y \leq C c^{-1/\beta} (cx^\beta)^{(1/\beta)-1} e^{-cx^\beta} + O\left(c^{-1/\beta} (cx^\beta)^{(1/\beta)-2} e^{-cx^\beta} \right),$$

as $cx^\beta \to \infty$, where C is a finite constant. This completes the proof of the lemma. □

Lemma 3.3.4 *For $\alpha > 0$ and $d \in \mathbf{N}$, as $n \to \infty$,*

$$E[(Z_n(d))^\alpha] = \frac{\alpha}{d} (nv_d)^{-\alpha/d} \Gamma(\alpha/d) + o(n^{-\alpha/d}). \tag{3.14}$$

Proof. Suppose $d \in \mathbf{N}$. For $r \in (0,\infty)$ and $\mathbf{x} \in (0,1)^d$, let $A(d; \mathbf{x}; r)$ denote the volume of the region $(0,1)^d \cap B(\mathbf{x}; r)$, where $B(\mathbf{x}; r)$ is the Euclidean d-ball with centre $\mathbf{x}$ and radius r. Then, for $n = 2, 3, \ldots$,

$$P[(Z_n(d))^\alpha > r | \mathbf{U}_n] = (1 - A(d; \mathbf{U}_n; r^{1/\alpha}))^{n-1}.$$

Fix $0 < \varepsilon < \alpha/(2d)$. Let $R(d;\alpha;n)$ denote the region $[n^{\varepsilon-(\alpha/d)}, 1 - n^{\varepsilon-(\alpha/d)}]^d$. In coordinates, set $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, and for $\mathbf{x} \in (0,1)^d$ let $m(\mathbf{x}) := \min(x_1, \ldots, x_d, 1 - x_1, \ldots, 1 - x_d)$, i.e. the shortest distance from $\mathbf{x}$ to the boundary of $(0,1)^d$. First of all, consider $\mathbf{x} \in R(d;\alpha;n)$. Then, if $0 < r \leq m(\mathbf{x})$, we have that $A(d; \mathbf{x}; r^{1/\alpha}) = v_d r^{d/\alpha}$. For $r > m(\mathbf{x})$, we have $v_d m(\mathbf{x})^{d/\alpha} < A(d; \mathbf{x}; r^{1/\alpha}) < v_d r^{d/\alpha}$. So a lower bound for $E[(Z_n(d))^\alpha]$, given $\mathcal{U}_n \in R(d;\alpha;n)$, is given by

$$E[(Z_n(d))^\alpha | \mathbf{U}_n \in R(d;\alpha;n)] \geq \int_0^{m(\mathbf{U}_n)} (1 - v_d r^{d/\alpha})^{n-1} \mathrm{d}r \geq \int_0^{n^{\varepsilon-(\alpha/d)}} (1 - v_d r^{d/\alpha})^{n-1} \mathrm{d}r \tag{3.15}$$

since if $\mathbf{U}_n \in R(d;\alpha;n)$ we have $m(\mathbf{U}_n) \geq n^{\varepsilon-(\alpha/d)}$. For $x \in (0,\infty)$, Taylor's Theorem with the Lagrange form for the remainder implies that $e^{-x} = 1 - x + Cx^2$ where $C \in [0, 1/2]$, so for $r \in (0, n^{\varepsilon-(\alpha/d)})$, we have that

$$\begin{aligned}
\left(1 - v_d r^{d/\alpha}\right)^{n-1} &\geq \left(\exp(-v_d r^{d/\alpha}) - \frac{1}{2} v_d^2 r^{2d/\alpha} \right)^n \\
&= \exp(-v_d n r^{d/\alpha}) \left(1 - \frac{1}{2} v_d^2 r^{2d/\alpha} \exp(v_d r^{d/\alpha}) \right)^n \\
&\geq \exp(-v_d n r^{d/\alpha}) \left(1 + O\left(n^{(2d\varepsilon/\alpha)-2} \exp(v_d n^{(d\varepsilon/\alpha)-1})\right)\right)^n \\
&= \exp(-v_d n r^{d/\alpha})(1 + o(1)),
\end{aligned}$$

as $n \to \infty$, since $\varepsilon < \alpha/(2d)$. So from (3.15) we have that

$$E[(Z_n(d))^\alpha | \mathbf{U}_n \in R(d;\alpha;n)] \geq (1 + o(1)) \int_0^{n^{\varepsilon-(\alpha/d)}} \exp(-v_d n r^{d/\alpha}) \mathrm{d}r. \tag{3.16}$$

Using (3.12), we obtain from (3.16) that

$$E[(Z_n(d))^\alpha|\mathbf{U}_n \in R(d;\alpha;n)] \geq \frac{\alpha}{d}(nv_d)^{-\alpha/d}\Gamma(\alpha/d) + o(n^{-\alpha/d}) + O(\exp(-v_d n^{\varepsilon d/\alpha})),$$

since $\varepsilon < \alpha/(2d)$. For the upper bound, we have

$$E[(Z_n(d))^\alpha|\mathbf{U}_n \in R(d;\alpha;n)] \leq \int_0^{d^{1/2}} (1 - A(d;\mathbf{U}_n;r^{1/\alpha}))^{n-1}\mathrm{d}r$$
$$\leq \int_0^{m(\mathbf{U}_n)} \exp(-v_d n r^{d/\alpha})\mathrm{d}r + \int_{m(\mathbf{U}_n)}^{d^{1/2}} \exp(-v_d n m(\mathbf{U}_n)^{d/\alpha})\mathrm{d}r. \tag{3.17}$$

For the second term on the right hand side of (3.17), using the fact that here $m(\mathbf{U}_n) \geq n^{\varepsilon-(\alpha/d)}$, we have

$$\int_{m(\mathbf{U}_n)}^{d^{1/2}} \exp(-v_d n m(\mathbf{U}_n)^{d/\alpha})\mathrm{d}r \leq d^{1/2}\exp(-v_d n \cdot n^{(\varepsilon d/\alpha)-1}).$$

Then, using (3.12) for the first term on the right hand side of (3.17), we obtain

$$E[(Z_n(d))^\alpha|\mathbf{U}_n \in R(d;\alpha;n)] \leq \frac{\alpha}{d}(nv_d)^{-\alpha/d}\Gamma(\alpha/d) + o(n^{-\alpha/d}) + O(\exp(-v_d n^{\varepsilon d/\alpha})).$$

So combining the upper and lower bounds we obtain

$$E[(Z_n(d))^\alpha|\mathbf{U}_n \in R(d;\alpha;n)] = (nv_d)^{-\alpha/d}\frac{\alpha}{d}\Gamma(\alpha/d) + o(n^{-\alpha/d}). \tag{3.18}$$

Now consider $\mathbf{x} \in (0,1)^d \setminus R(d;\alpha;n)$. Here, for $0 \leq r \leq 2^{-\alpha}$, $(2^{-d}v_d)r^{d/\alpha} \leq A(d;\mathbf{x};r^{1/\alpha}) \leq v_d r^{d/\alpha}$. For $r \geq 2^{-\alpha}$, we have $C_1 \leq A(d;\mathbf{x};r^{1/\alpha}) \leq 1$ for some $0 < C_1 < 1$ (depending on d). Then, by similar arguments to above, we obtain

$$E[(Z_n(d))^\alpha|\mathbf{U}_n \notin R(d;\alpha;n)] = O(n^{-\alpha/d}). \tag{3.19}$$

And so, since $P[\mathbf{U}_n \notin R(d;\alpha;n)] = O(n^{\varepsilon-(\alpha/d)})$, we obtain from (3.18) and (3.19) that

$$\begin{aligned} E[(Z_n(d))^\alpha] &= E[(Z_n(d))^\alpha|\mathbf{U}_n \notin R(d;\alpha;n)]P[\mathbf{U}_n \notin R(d;\alpha;n)] \\ &\quad + E[(Z_n(d))^\alpha|\mathbf{U}_n \in R(d;\alpha;n)]P[\mathbf{U}_n \in R(d;\alpha;n)] \\ &= (nv_d)^{-\alpha/d}\frac{\alpha}{d}\Gamma(\alpha/d) + o(n^{-\alpha/d}), \end{aligned}$$

and so we have (3.14). □

Remark. By Lemma 3.3.4, for $0 < \alpha < d$, we have

$$E[\mathcal{O}^{d,\alpha}(\mathcal{U}_n)] = \frac{\alpha}{d-\alpha}v_d^{-\alpha/d}\Gamma(\alpha/d)n^{(d-\alpha)/d} + o(n^{(d-\alpha)/d}),$$

which complements the law of large numbers given in Theorem 2.3.1.

Proof of Theorem 2.3.1 (ii) and (iii). With the definition of $Z_i(d)$ in this section, let

$$W(d,\alpha) = \sum_{i=1}^{\infty} (Z_i(d))^{\alpha}.$$

The sum converges almost surely since it has non-negative terms and, by (3.14), has finite expectation for $\alpha > d$. Let $k \in \mathbf{N}$. By (3.14) and Hölder's inequality, there exists a constant $0 < C < \infty$ such that

$$\begin{aligned} E[(W(d,\alpha))^k] &= \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} E[(Z_{i_1}(d))^{\alpha}(Z_{i_2}(d))^{\alpha}\cdots(Z_{i_k}(d))^{\alpha}] \\ &\le C\sum_{i_1=1}^{\infty}\sum_{i_2=1}^{\infty}\cdots\sum_{i_k=1}^{\infty} i_1^{-\alpha/d} i_2^{-\alpha/d}\cdots i_k^{-\alpha/d} < \infty, \end{aligned}$$

since $\alpha/d > 1$. The L^k convergence then follows from the dominated convergence theorem, and we have part (ii) of Theorem 2.3.1.

Finally, for part (iii) of Theorem 2.3.1, we have, when $\alpha = d$, (3.14) implies that

$$E[\mathcal{O}^{d,d}(\mathcal{U}_n)] = v_d^{-1}\sum_{i=1}^{n} i^{-1} + O(1) = v_d^{-1}\log(n) + O(1),$$

and so we have (2.17), completing the proof of Theorem 2.3.1. □

3.3.4 Proof of Theorems 2.4.1 and 2.4.2

We first prove Theorem 2.4.1, our law of large numbers for the total weight of the random MDSF on the unit square, under the general partial order $\overset{\theta,\phi}{\preccurlyeq}$, for $0 \le \theta < 2\pi$ and $0 < \phi \le \pi$. Recall that $\mathbf{y} \overset{\theta,\phi}{\preccurlyeq} \mathbf{x}$ if $\mathbf{y} \in C_{\theta,\phi}(\mathbf{x})$, where $C_{\theta,\phi}(\mathbf{x})$ is the cone formed by the rays at θ and $\theta+\phi$ measured anticlockwise from the upwards vertical.

We consider the random point set $\mathcal{X}_n$, the point process of n independent random points on $(0,1)^2$ with common density function f. We suppose that f satisfies the condition (C1) of Section 2.1. For the general partial order given by θ, ϕ we apply Lemma 3.1.1 to obtain a law of large numbers for $\mathcal{L}^{\alpha}(\mathcal{X}_n)$. This method enables us to evaluate the limit explicitly, unlike methods based on the subadditivity of the functional which may also be applicable here (see the remark at the end of this section).

In applying Lemma 3.1.1 to the MDSF functional, we take the dimension d in the lemma to be 2. We take $\xi(\mathbf{x};\mathcal{X})$ to be $d_{\theta,\phi}(\mathbf{x};\mathcal{X})^{\alpha}$, where $d_{\theta,\phi}(\mathbf{x};\mathcal{X})$ is the distance from point $\mathbf{x}$ to its directed nearest neighbour in $\mathcal{X}$ under $\overset{\theta,\phi}{\preccurlyeq}$, if such a neighbour exists, or zero otherwise. Thus in our case $\xi(\mathbf{x};\mathcal{X}) = (d_{\theta,\phi}(\mathbf{x};\mathcal{X}))^\alpha$ where

$$d_{\theta,\phi}(\mathbf{x};\mathcal{X}) := \min\left\{\|\mathbf{x}-\mathbf{y}\| : \mathbf{y} \in \mathcal{X}\setminus\{\mathbf{x}\}, \mathbf{y} \overset{\theta,\phi}{\preccurlyeq} \mathbf{x}\right\}, \tag{3.20}$$

with the convention that $\min\{\} = 0$. We need to show this choice of ξ satisfies the conditions of Lemma 3.1.1. As before, $\mathcal{H}_1$ denotes a homogeneous Poisson process on $\mathbf{R}^d$ of intensity 1, now with $d = 2$.

Lemma 3.3.5 *With ξ as defined by (3.20), ξ is almost surely stabilizing on $\mathcal{H}_1$, in the sense of (3.1), with limit $\xi_\infty(\mathcal{H}_1) = (d_{\theta,\phi}(\mathbf{0};\mathcal{H}_1))^\alpha$.*

Proof. Let R be the (random) distance from $\mathbf{0}$ to its directed nearest neighbour in $\mathcal{H}_1$, i.e. $R = d_{\theta,\phi}(\mathbf{0};\mathcal{H}_1)$. Since $\phi > 0$ we have $0 < R < \infty$ almost surely. But then for any $\ell > R$, we have $\xi(\mathbf{0}; (\mathcal{H}_1 \cap B(\mathbf{0};\ell)) \cup \mathcal{A}) = R^\alpha$, for any finite $\mathcal{A} \subset \mathbf{R}^2 \setminus B(\mathbf{0};\ell)$. Thus ξ stabilizes on $\mathcal{H}_1$ with limit $\xi_\infty(\mathcal{H}_1) = R^\alpha$. □

Before proving that our choice of ξ satisfies the moments condition for Lemma 3.1.1, we give a geometrical lemma. For $B \subseteq \mathbf{R}^2$ with B bounded, and for $\mathbf{x} \in B$, write $\mathrm{dist}(\mathbf{x};\partial B)$ for $\sup\{r : B(\mathbf{x};r) \subseteq B\}$, and for $s > 0$, define the region

$$A_{\theta,\phi}(\mathbf{x},s;B) := B(\mathbf{x};s) \cap B \cap C_{\theta,\phi}(\mathbf{x}). \tag{3.21}$$

Lemma 3.3.6 *Let B be a convex bounded set in $\mathbf{R}^2$, and let $\mathbf{x} \in B$. If $A_{\theta,\phi}(\mathbf{x},s;B) \cap \partial B(\mathbf{x};s) \neq \emptyset$, and $s > \mathrm{dist}(\mathbf{x},\partial B)$, then*

$$|A_{\theta,\phi}(\mathbf{x},s;B)| \geq s\sin(\phi/2)\mathrm{dist}(x,\partial B)/2.$$

Proof. The condition $A_{\theta,\phi}(\mathbf{x},s;B) \cap \partial B(\mathbf{x};s) \neq \emptyset$ says that there exists $\mathbf{y} \in B \cap C_{\theta,\phi}(\mathbf{x},s)$ with $\|\mathbf{y}-\mathbf{x}\| = s$. The line segment $\mathbf{xy}$ is contained in the cone $C_{\theta,\phi}(\mathbf{x})$; take a half-line $\mathbf{h}$ starting from $\mathbf{x}$, at an angle $\phi/2$ to the line segment $\mathbf{xy}$ and such that $\mathbf{h}$ is also contained in $C_{\theta,\phi}(\mathbf{x})$. Let $\mathbf{z}$ be the point in $\mathbf{h}$ at a distance $\mathrm{dist}(\mathbf{x},\partial B)$ from $\mathbf{x}$. Then the interior of the triangle $\mathbf{xyz}$ is entirely contained in $A_{\theta,\phi}(\mathbf{x},s)$, and has area $s\sin(\phi/2)\mathrm{dist}(x,\partial B)/2$.
□

Lemma 3.3.7 *Suppose $\alpha > 0$, and f satisfies condition (C1). Then ξ given by (3.20) satisfies the moments condition (3.2) for any $p \in (1/\alpha, 2/\alpha]$.*

Proof. We give the proof in the case $f(\mathbf{x}) = 1$ for $\mathbf{x} \in (0,1)^2$ and zero otherwise. However, the result also holds (with virtually the same proof and a monotonicity argument) if f is arbitrary satisfying condition (C1). Setting $R_n := (0, n^{1/2}]^2$, we have

$$E\left[\xi(n^{1/2}\mathbf{U}_1; n^{1/2}\mathcal{U}_n)^p\right] = \int_{R_n} E\left[\xi(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1})^p\right] \frac{d\mathbf{x}}{n}. \tag{3.22}$$

For $x \in R_n$ set $m(\mathbf{x}) := \mathrm{dist}(\mathbf{x}, \partial R_n)$. Let us divide R_n into three regions

$$\begin{aligned} R_n(1) &:= \{\mathbf{x} \in R_n : m(\mathbf{x}) \le n^{-1/2}\}; \quad R_n(2) := \{\mathbf{x} \in R_n : m(\mathbf{x}) > 1\}; \\ R_n(3) &:= \{\mathbf{x} \in R_n : n^{-1/2} < m(\mathbf{x}) \le 1\}. \end{aligned}$$

For all $\mathbf{x} \in R_n$, we have $\xi(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1}) \le (2n)^{\alpha/2}$, and hence, since $R_n(1)$ has area at most 4, we can bound the contribution to (3.22) from $\mathbf{x} \in R_n(1)$ by

$$\int_{\mathbf{x}\in R_n(1)} E\left[\left(\xi(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1})\right)^p\right] \frac{d\mathbf{x}}{n} \le 4n^{-1}(2n)^{p\alpha/2} = 2^{2+p\alpha/2} n^{(p\alpha-2)/2}, \tag{3.23}$$

which is bounded provided $p\alpha \le 2$.

Now, for $\mathbf{x} \in R_n$, with $A_{\theta,\phi}(\cdot)$ defined at (3.21), we have

$$\begin{aligned} P\left[d_{\theta,\phi}(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1}) > s\right] &\le P\left[n^{1/2}\mathcal{U}_{n-1} \cap A_{\theta,\phi}(\mathbf{x}, s; R_n) = \emptyset\right] \\ &= \left(1 - \frac{|A_{\theta,\phi}(\mathbf{x}, s; R_n)|}{n}\right)^{n-1} \\ &\le \exp(1 - |A_{\theta,\phi}(\mathbf{x}, s; R_n)|), \end{aligned} \tag{3.24}$$

since $|A_{\theta,\phi}(\mathbf{x}, s; R_n)| \le n$. For $\mathbf{x} \in R_n$ and $s > m(\mathbf{x})$, by Lemma 3.3.6 we have

$$|A_{\theta,\phi}(\mathbf{x}, s; R_n)| \ge \sin(\phi/2) s m(\mathbf{x})/2 \quad \text{if} \quad A_{\theta,\phi}(\mathbf{x}, s; R_n) \cap \partial B(\mathbf{x}; s) \ne \emptyset,$$

and also

$$P[d_{\theta,\phi}(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1}) > s] = 0 \quad \text{if} \quad A_{\theta,\phi}(\mathbf{x}, s; R_n) \cap \partial B(\mathbf{x}; s) = \emptyset.$$

For $s \le m(\mathbf{x})$, we have that $|A_{\theta,\phi}(\mathbf{x}, s; R_n)| = \frac{\phi}{2}s^2 \ge \sin(\phi/2)s^2$. Combining these observations and (3.24), we obtain for all $\mathbf{x} \in R_n$ and $s > 0$ that

$$P\left[d_{\theta,\phi}(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1}) > s\right] \le \exp\left(1 - \sin(\phi/2) s \min(s, m(\mathbf{x}))/2\right), \quad \mathbf{x} \in R_n.$$

Setting $c = (1/2)\sin(\phi/2)$, we therefore have for $\mathbf{x} \in R_n$ that

$$\begin{aligned}
E\left[\xi(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1})^p\right] &= \int_0^\infty P\left[\xi(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1})^p > r\right] \mathrm{d}r \\
&= \int_0^\infty P\left[d_{\theta,\phi}(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1}) > r^{1/(\alpha p)}\right] \mathrm{d}r \\
&\leq \int_0^{m(\mathbf{x})^{\alpha p}} \mathrm{d}r \exp\left(1 - cr^{2/(\alpha p)}\right) \\
&\quad + \int_{m(\mathbf{x})^{\alpha p}}^\infty \mathrm{d}r \exp\left(1 - cm(\mathbf{x})r^{1/(\alpha p)}\right) \\
&= O(1) + \int_{m(\mathbf{x})^2}^\infty e^{1-cu}\alpha p u^{\alpha p-1} m(\mathbf{x})^{-p\alpha} \mathrm{d}u \\
&= O(1) + O(m(\mathbf{x})^{-\alpha p}). \qquad (3.25)
\end{aligned}$$

For $\mathbf{x} \in R_n(2)$, this bound is $O(1)$, and the area of $R_n(2)$ is less than n, so that the contribution to (3.22) from $R_n(2)$ satisfies

$$\limsup_{n\to\infty} \int_{R_n(2)} E\left[\left(\xi(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1})\right)^p\right] \frac{\mathrm{d}\mathbf{x}}{n} < \infty. \qquad (3.26)$$

Finally, by (3.25), there is a constant c' such that if $\alpha p > 1$, the contribution to (3.22) from $R_n(3)$ satisfies

$$\begin{aligned}
\int_{R_n(3)} E\left[\left(\xi(\mathbf{x}; n^{1/2}\mathcal{U}_{n-1})\right)^p\right] \frac{\mathrm{d}\mathbf{x}}{n} &\leq c'n^{-1/2} \int_{y=n^{-1/2}}^1 y^{-\alpha p} \mathrm{d}y \\
&\leq \left(\frac{c'n^{-1/2}}{\alpha p - 1}\right) n^{(\alpha p-1)/2}
\end{aligned}$$

which is bounded provided $\alpha p \leq 2$. Combined with the bounds in (3.23) and (3.26), this shows that the expression (3.22) is uniformly bounded, provided $1 < \alpha p \leq 2$. □

Proof of Theorem 2.4.1. Suppose $0 < \alpha < 2$. By Lemmas 3.3.5 and 3.3.7, our functional ξ, given at (3.20), satisfies the conditions of Lemma 3.1.1 with $p = 2/\alpha$ and $q = 1$, with f satisfying (C1). So by Lemma 3.1.1, we have that

$$\begin{aligned}
n^{(\alpha/2)-1}\mathcal{L}^\alpha(\mathcal{X}_n) &= n^{-1} \sum_{\mathbf{x}\in\mathcal{X}_n} \xi(n^{1/2}\mathbf{x}; n^{1/2}\mathcal{X}_n) \\
&\xrightarrow{L^1} E[\xi_\infty(\mathcal{H}_1)] \int_{\mathrm{supp}(f)} f(\mathbf{x})^{(2-\alpha)/2} \mathrm{d}\mathbf{x}. \qquad (3.27)
\end{aligned}$$

Since the disk sector $C_{\theta,\phi}(\mathbf{x}) \cap B(\mathbf{x}; r)$ has area $(\phi/2)r^2$, by Lemma 3.3.5 we have

$$P\left[\xi_\infty(\mathcal{H}_1) > s\right] = P\left[\mathcal{H}_1 \cap C_{\theta,\phi}(\mathbf{0}) \cap B(\mathbf{0}; s^{1/\alpha}) = \emptyset\right] = \exp\left(-(\phi/2)s^{2/\alpha}\right).$$

Hence, on the right hand side of (3.27)

$$E\left[\xi_\infty(\mathcal{H}_1)\right] = \int_0^\infty P\left[\xi_\infty(\mathcal{H}_1) > s\right] ds = \alpha 2^{(\alpha-2)/2}\phi^{-\alpha/2}\Gamma(\alpha/2),$$

and this gives us (2.20). Finally, in the case where $\mathcal{X}_n = \mathcal{U}_n$ and $\overset{\theta,\phi}{\preccurlyeq} = \preccurlyeq^*$, (2.21) remains true when $\mathcal{U}_n$ is replaced by $\mathcal{U}_n^0$, since

$$E[n^{(\alpha-2)/2}|\mathcal{L}^\alpha(\mathcal{U}_n^0) - \mathcal{L}^\alpha(\mathcal{U}_n)|] \leq 2^{\alpha/2}n^{(\alpha-2)/2}E[M(\mathcal{U}_n)], \tag{3.28}$$

where $M_2^*(\mathcal{U}_n)$ denotes the number of minimal elements of $\mathcal{U}_n$. By (2.26), $E[M_2^*(\mathcal{U}_n)] \leq 1 + \log n$, and hence the right hand side of (3.28) tends to 0 as $n \to \infty$ for $0 < \alpha < 2$. This gives us (2.21) with $\mathcal{U}_n^0$ under $\preccurlyeq^*$. □

Remark. A law of large numbers for Euclidean functionals of many random geometric structures can be treated by the boundary functional approach of Yukich [142]. It can be shown that the MDSF satisfies some, but possibly not all, of the appropriate conditions that would allow this approach to be successful. The MDSF functional is subadditive, its corresponding boundary functional is superadditive, and the functional and its boundary functional are sufficiently 'close in mean'. However, it is not clear that the functional is 'smooth', since the degree of the graph is not bounded.

We now prove Theorem 2.4.2. In this case, we consider $\mathbf{R}^d$, $d \in \mathbf{N}$, and take the partial order to be $\preccurlyeq^*$ or $\preccurlyeq_*$ on $\mathbf{R}^d$. Recall that for $\mathbf{x}, \mathbf{y}$ in $\mathbf{R}^d$, $\mathbf{y} \preccurlyeq^* \mathbf{x}$ if each component of $\mathbf{x} - \mathbf{y}$ is nonnegative, while $\mathbf{y} \preccurlyeq_* \mathbf{x}$ if the first component of $\mathbf{x} - \mathbf{y}$ is nonnegative.

Let $d^*(\mathbf{x}; \mathcal{X})$ and $d_*(\mathbf{x}; \mathcal{X})$ denote the distance from $\mathbf{x} \in \mathcal{X}$ to its directed nearest neighbour in $\mathcal{X}$ under partial order $\preccurlyeq^*$ and $\preccurlyeq_*$ respectively. For $\alpha > 0$, set

$$\xi^*(\mathbf{x}; \mathcal{X}) = d^*(\mathbf{x}; \mathcal{X})^\alpha, \quad \xi_*(\mathbf{x}; \mathcal{X}) = d_*(\mathbf{x}; \mathcal{X})^\alpha. \tag{3.29}$$

Lemma 3.3.8 *The functionals ξ^* and ξ_* as given by (3.29) are both almost surely stabilizing on $\mathcal{H}_1$, in the sense of (3.1), with limits $\xi^*_\infty(\mathcal{H}_1) = d^*(\mathbf{0}; \mathcal{H}_1)^\alpha$ and $\xi_{*\infty}(\mathcal{H}_1) = d_*(\mathbf{0}; \mathcal{H}_1)^\alpha$ respectively.*

Proof. Let R denote $d^*(\mathbf{0}; \mathcal{H}_1)$ or $d_*(\mathbf{0}; \mathcal{H}_1)$ as appropriate. Then, R is finite, almost surely. Let ξ denote ξ^* or ξ_* as appropriate. Then for any $\ell > R$ we have that $\xi(\mathbf{0}; (\mathcal{H}_1 \cap B(\mathbf{0}; \ell)) \cup \mathcal{A}) = R^\alpha$, for any finite $\mathcal{A} \subset \mathbf{R}^d \setminus B(\mathbf{0}; \ell)$. Thus ξ is stabilizing on $\mathcal{H}_1$ with limit $\xi_\infty(\mathcal{H}_1) = R^\alpha$. □

Proof of Theorem 2.4.2. Suppose $d \in \mathbf{N}$ and $0 < \alpha < d$. By Lemmas 3.3.8 and 3.3.9, our functionals ξ^* and ξ_*, given at (3.20), both satisfy the conditions of Lemma 3.1.1 with $p = d/\alpha$ and $q = 1$, with f satisfying (C1). So by Lemma 3.1.1, we have that

$$n^{(\alpha-d)/d}\mathcal{L}^{\alpha}(\mathcal{X}_n) = n^{-1}\sum_{\mathbf{x}\in\mathcal{X}_n} \xi(n^{1/d}\mathbf{x}; n^{1/d}\mathcal{X}_n) \xrightarrow{L^1} E[\xi_\infty(\mathcal{H}_1)] \int_{\mathrm{supp}(f)} f(\mathbf{x})^{(d-\alpha)/d} d\mathbf{x}, \quad (3.32)$$

where ξ is either ξ^* or ξ_* here. We have, for the $\preccurlyeq^*$ case,

$$\begin{aligned} P\left[\xi^*_\infty(\mathcal{H}_1) > s\right] &= P\left[\mathcal{H}_1 \cap \{\mathbf{x} \in \mathbf{R}^d : \mathbf{x} \preccurlyeq^* \mathbf{0}\} \cap B(\mathbf{0}; s^{1/\alpha}) = \emptyset\right] \\ &= \exp\left(-2^{-d} v_d s^{d/\alpha}\right). \end{aligned}$$

Hence,

$$E\left[\xi^*_\infty(\mathcal{H}_1)\right] = \int_0^\infty P\left[\xi^*_\infty(\mathcal{H}_1) > s\right] ds = 2^\alpha v_d^{-\alpha/d}\Gamma(1 + (\alpha/d)),$$

and with the $\preccurlyeq^*$ case of (3.32) this gives us (2.22). Also, for the $\preccurlyeq_*$ case,

$$\begin{aligned} P\left[\xi_{*\infty}(\mathcal{H}_1) > s\right] &= P\left[\mathcal{H}_1 \cap \{\mathbf{x} \in \mathbf{R}^d : \mathbf{x} \preccurlyeq_* \mathbf{0}\} \cap B(\mathbf{0}; s^{1/\alpha}) = \emptyset\right] \\ &= \exp\left(-(v_d/2) s^{d/\alpha}\right). \end{aligned}$$

Hence,

$$E\left[\xi_{*\infty}(\mathcal{H}_1)\right] = \int_0^\infty P\left[\xi_{*\infty}(\mathcal{H}_1) > s\right] ds = 2^{\alpha/d} v_d^{-\alpha/d}\Gamma(1 + (\alpha/d)),$$

which yields (2.24) with the $\preccurlyeq_*$ case of (3.32).

Finally, in the case where $\mathcal{X}_n = \mathcal{U}_n$ and the partial order is $\preccurlyeq^*$, (2.23) remains true when $\mathcal{U}_n$ is replaced by $\mathcal{U}_n^0$, since

$$E[n^{(\alpha-d)/d}|\mathcal{L}^{d,\alpha}(\mathcal{U}_n^0) - \mathcal{L}^{d,\alpha}(\mathcal{U}_n)|] \leq d^{\alpha/2} n^{(\alpha-d)/d} E[M_d^*(\mathcal{U}_n)], \quad (3.33)$$

where $M_d^*(\mathcal{U}_n)$ denotes the number of minimal elements of $\mathcal{U}_n$ under $\preccurlyeq^*$. By (2.27), $E[M_d^*(\mathcal{U}_n)] = O((\log n)^{d-1})$, and hence the right hand side of (3.33) tends to 0 as $n \to \infty$ for $0 < \alpha < d$. This gives us (2.23) with $\mathcal{U}_n^0$ under $\preccurlyeq^*$. □

Chapter 4

Dickman-type distributions and the MDST

4.1 Introduction

For this chapter, we consider the random minimal directed spanning tree on uniform and Poisson points in $(0,1)^2$, under the coordinate-wise partial order $\preccurlyeq^*$ (as considered by Bhatt and Roy [21]). We consider the limiting behaviour of the total length of some subsets of the edges in the MDST. In particular, we deal with the edges *incident to the origin* and the *longest edge*. Limiting distributions for these quantities are given in terms of certain Dickman-type distributions, which emerge from the Poisson-Dirichlet distribution.

The edges joined to the origin in the MDST in $(0,1)^2$ with partial order $\preccurlyeq^*$ were the principal object of analysis in [21], in which Bhatt and Roy established (amongst other things) existence of a weak limit for the total length of such edges, without fully describing that limit. We use a different method to characterize the limiting distribution (see Theorem 4.2.1) as a variant of the *Dickman distribution* which has previously arisen in such fields as probabilistic number theory, population genetics, and the theory of random search trees (see Appendix C). We also extend the result to power-weighted edges.

In addition, we derive a weak convergence result for the maximum of all edge lengths in the MDST (Theorem 4.2.3). In this case, the limiting distribution is related to the distribution of the largest component of the Poisson-Dirichlet distribution with parameter 1. The latter distribution has also sometimes been called a 'Dickman distribution' (see [7, 42]) and we shall call it the *max-Dickman* distribution. In Appendix C, we shall discuss both types of Dickman distribution in some detail; they are related, and intimately connected with the Poisson-Dirichlet distribution, which we review in Appendix C also. The material in Appendix C is therefore of a supplementary nature and may be omitted on a first reading.

We remark that as a supplementary result, we can re-derive Bhatt and Roy's result (Theorem 4.2.2) for the weak limit of the maximum length of edges *incident to the origin* using a similar method to that used in this chapter.

4.2 Main results

We work with the partial order $\preccurlyeq^*$ and weight function w_α as given by (1.2). For $\mathcal{S}$ a finite subset of $(0,1)^2$, and $\alpha > 0$, let

$$\mathcal{L}_0^\alpha(\mathcal{S}) := \sum_{\mathbf{X}\in\mathcal{S},\ \mathbf{X} \text{ minimal}} \|\mathbf{X}\|^\alpha. \tag{4.1}$$

Thus $\mathcal{L}_0^\alpha(\mathcal{S})$ is the total weight of the edges incident to the origin in the MDST on $\mathcal{S}\cup\{\mathbf{0}\}$.

We will take $\mathcal{S}$ to be $\mathcal{U}_n$, the binomial point process consisting of n independent uniform random points on $(0,1)^2$, or $\mathcal{P}_n$, the homogeneous Poisson point process with intensity n on $(0,1)^2$.

Bhatt and Roy [21] showed that, as $n \to \infty$,

$$\mathcal{L}_0^1(\mathcal{U}_n) \xrightarrow{\mathcal{D}} Y; \quad \mathcal{L}_0^1(\mathcal{P}_n) \xrightarrow{\mathcal{D}} Y,$$

where Y is a random variable with $E[Y] = 2$ and $\mathrm{Var}[Y] = 1$. As a consequence of Theorem 4.2.1 below, we will fully characterise the distribution of Y (it is what we call a 'generalized Dickman' random variable with parameter 2; see below).

Our first main result describes the limiting distribution, as $n \to \infty$, of $\mathcal{L}_0^\alpha(\mathcal{U}_n)$ or $\mathcal{L}_0^\alpha(\mathcal{P}_n)$ more fully in terms of a *Dickman distribution.* Given $\theta > 0$, we shall say a random variable X has a *generalized Dickman* distribution with shape parameter θ (or $X \sim \mathrm{GD}(\theta)$ for short) if it satisfies the distributional fixed-point identity

$$X \stackrel{\mathcal{D}}{=} U^{1/\theta}(1+X),$$

where U is uniform on $(0,1)$, and is independent of the X on the right, and where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. For further information on Dickman distributions, see Section C.

Theorem 4.2.1 *Let $\alpha > 0$. Let $Z(\theta)$ denote a random variable with the generalized Dickman distribution with shape parameter θ. Then as $n \to \infty$,*

$$\mathcal{L}_0^\alpha(\mathcal{P}_n) \xrightarrow{\mathcal{D}} Z(2/\alpha) \tag{4.2}$$

and

$$\mathcal{L}_0^\alpha(\mathcal{U}_n) \xrightarrow{\mathcal{D}} Z(2/\alpha). \tag{4.3}$$

The limiting distribution has Laplace transform

$$E[e^{-tZ(2/\alpha)}] = \exp\left(\frac{2}{\alpha}\int_0^t \frac{e^{-s}-1}{s} ds\right), \qquad t \in \mathbf{R}.$$

In the special case $\alpha = 1$, the distribution of the limiting variable $Z(2)$ has mean 2 and variance 1, and moments $m_2 = 5$, $m_3 = 44/3$, $m_4 = 293/6, \ldots$.

Remarks. Perhaps the most natural case is $\alpha = 1$ (i.e., simply take the Euclidean length of edges). By considering the more general case allowing for any $\alpha > 0$, we get the whole range of possible generalized Dickman distributions as limits.

Bhatt and Roy [21] use a different approach based on the method of moments to prove the weak convergence (4.3) (only for $\alpha = 1$). Their argument is complicated and they give only values for the first two moments of $Z(2)$, not the higher moments. Nor do they say anything about the density, distribution or moment generating functions of $Z(2)$. Thus, even for $\alpha = 1$ our approach gives a good deal of extra information beyond that provided in [21]. Conversely, since Bhatt and Roy prove convergence of all moments of $\mathcal{L}_0^1(\mathcal{U}_n)$ to the corresponding moments of $Z(2)$, this combined with our characterization of $Z(2)$ means we can identify the limit of the k-th moment of $\mathcal{L}_0^1(\mathcal{U}_n)$, for any fixed k, by computing the kth moment of $Z(2)$ recursively using the formula (C.4.9) below.

In [12], Bai, Lee and Penrose show that this two dimensional case is rather special – for $d \geq 3$ the corresponding limits for the length of the rooted edges in the MDST (under $\preccurlyeq^*$) are normally distributed. On the other hand, when $d = 1$, we simply have the first Dirichlet spacing – see Section 5.1.1.

Our second main result (Theorem 4.2.3 below) concerns the *maximum* edge length of the MDST; when considering maxima we consider only the case with $\alpha = 1$ (results on maxima for other values of α are easily deduced from results for this case).

Bhatt and Roy [21] considered the *maximum length of edges joined to the origin*, for the MDST on $\mathcal{U}_n \cup \{\mathbf{0}\}$. We denote the maximum length of the edges joined to the origin in the MDST on $\mathcal{S} \cup \{\mathbf{0}\}$ by $\mathcal{M}_0(\mathcal{S})$, so that

$$\mathcal{M}_0(\mathcal{S}) = \max_{\mathbf{X} \in \mathcal{S},\ \mathbf{X}\ \text{minimal}} \|\mathbf{X}\|.$$

Bhatt and Roy [21] proved the following result:

Theorem 4.2.2 *As $n \to \infty$,*

$$\mathcal{M}_0(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \max\{U_1, U_2\} \stackrel{\mathcal{D}}{=} U_1^{1/2}; \quad \mathcal{M}_0(\mathcal{P}_n) \xrightarrow{\mathcal{D}} U_1^{1/2}, \tag{4.4}$$

where U_1, U_2 are independent uniform random variables on $(0,1)$.

Here, we consider in addition the global maximum of *all* Euclidean edge lengths in the MDST on $\mathcal{S} \cup \{\mathbf{0}\}$, not just those joined to the origin. Denote this maximal edge length by $\mathcal{M}(\mathcal{S})$.

The limit variable for maximum edge length is given in terms of what we shall call the *max-Dickman* distribution. We define this to be the (unique) distribution of a random variable M which satisfies the distributional identity

$$M \stackrel{\mathcal{D}}{=} \max\{1 - U, UM\} \tag{4.5}$$

where U is uniformly distributed on $(0,1)$ and independent of the M on the right.

Theorem 4.2.3 *Suppose M and M' are independent max-Dickman random variables. As $n \to \infty$,*

$$\mathcal{M}(\mathcal{P}_n) \xrightarrow{\mathcal{D}} \max\{M, M'\} \tag{4.6}$$

and

$$\mathcal{M}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \max\{M, M'\}. \tag{4.7}$$

We prove Theorem 4.2.3 in section 4.4.

The generalized Dickman GD(1) and max-Dickman distributions are more closely related than might at first be apparent. In probabilistic terms, they can both be expressed in terms of a Poisson point process on $(0,1)$ with mean measure μ given by $\mathrm{d}\mu = (1/x)\mathrm{d}x$. Suppose the points of this Poisson point process are listed in decreasing order as $Y_1, Y_2, \ldots$. Then the sum $\sum_i Y_i$ has the GD(1) distribution, while the maximum spacing $\max\{1 - Y_1, Y_1 - Y_2, Y_2 - Y_3, \ldots\}$ has the max-Dickman distribution. The latter is also the distribution of the largest component of the Poisson-Dirichlet distribution; see Section C.7.

In more analytical terms, both the GD(1) and the max-Dickman probability density functions are defined in terms of the *Dickman function*, which appeared in the 1930 paper of K. Dickman [41] on large prime factors of large integers (for a more recent reference, see Donnelly and Grimmett [42]). In Appendix C, the Dickman function and the generalized Dickman and max-Dickman distributions are described in more detail.

4.3 Proof of Theorem 4.2.1

The intuition behind Theorem 4.2.1 goes as follows. If there exists a minimal point of $\mathcal{P}_n$ (or $\mathcal{U}_n$) near to the origin, then there is no minimal point lying to the north-east of that point. Hence, the minimal points are likely to all lie near to either the x-axis or the y-axis, and the contributions from these two axes are nearly independent. Near the x-axis, the x-coordinates of successive minimal points (taken in order of increasing y-coordinate) form a sequence of products of uniforms $U_1, U_1U_2, U_1U_2U_3, \ldots$ and summing these gives a Dickman distribution. Similarly for the y-axis.

In the course of the proof we shall use Slutsky's Theorem repeatedly (see Lemma A.2.1). We shall also use the following coupling lemma relating the point processes $\mathcal{U}_n$ and $\mathcal{P}_n$.

Lemma 4.3.1 *There exist point processes $\mathcal{U}'_n, \mathcal{P}'_n$ defined on the same probability space as each other for each n, such that:*

- *$\mathcal{U}'_n$ has the same distribution as $\mathcal{U}_n$.*
- *$\mathcal{P}'_n$ has the same distribution as $\mathcal{P}_n$.*
- *With probability tending to 1 as $n \to \infty$, the set of minimal elements of $\mathcal{P}'_n$ is identical to the set of minimal elements of $\mathcal{U}'_n$.*

Proof. Let $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \ldots$ be independent and uniformly distributed on $(0,1)^2$, let $N(n)$ be Poisson with parameter n and independent of $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \ldots)$, let $\mathcal{P}'_n := \{\mathbf{U}_1, \ldots, \mathbf{U}_{N(n)}\}$, and for $m \in \mathbf{N}$ set $\mathcal{U}'_m := \{\mathbf{U}_1, \ldots, \mathbf{U}_m\}$. Then $\mathcal{P}'_n \stackrel{\mathcal{D}}{=} \mathcal{P}_n$ and $\mathcal{U}'_n \stackrel{\mathcal{D}}{=} \mathcal{U}_n$.

Let A_m be the event that $\mathbf{U}_m$ is a minimal element of $\mathcal{U}'_m$, and let δ_m be the number of minimal elements of $\mathcal{U}'_m$. By exchangeability, each point $\mathbf{U}_i, i \leq m$ is equally likely to be minimal in $\mathcal{U}'_m$, so that $E[\delta_m] = mP[A_m]$. By (2.26) (see also [15], or the proof of Theorem 1.1(a) of [21]) we have $E[\delta_m] \sim \log(m)$ as $m \to \infty$. Hence, $P[A_m] \sim (\log m)/m$ as $m \to \infty$, and therefore

$$P\left[\cup_{n-n^{3/4}\leq m \leq n+n^{3/4}} A_m\right] \leq 3n^{3/4}(\log n)/n \to 0. \qquad (4.8)$$

Let E_n denote event that the set of minimal points in $\mathcal{U}'_n$ differs from the set of minimal points of $\mathcal{P}'_n$. By the coupling of $\mathcal{U}'_m$ $(m \geq 1)$ and $\mathcal{P}'_n$, E_n occurs only if A_m occurs for some m with $N(n) < m \leq n$ (if $N(n) < n$) or with $n < m \leq N(n)$ (if $N(n) > n$). Hence,

$$P[E_n] \leq P[|N(n) - n| \geq n^{3/4}] + P[\cup_{n-n^{3/4}\leq m\leq n+n^{3/4}} A_m].$$

In the right hand side, the first probability tends to zero by Chebyshev's inequality while the second tends to zero by (4.8), and hence $P[E_n] \to 0$ as asserted. □

We now work towards a proof of (4.2). Let $\mathcal{Y}_n$ be the set of minimal elements of the point set $\mathcal{P}_n$, i.e., the set of elements of $\mathcal{P}_n$ which are joined to $\mathbf{0}$ in the MDST on $\mathcal{P}_n \cup \{\mathbf{0}\}$.

Lemma 4.3.2 *As $n \to \infty$, we have* $(\log n)^{-1}\text{card}(\mathcal{Y}_n) \xrightarrow{P} 1$.

Proof. The corresponding result for the number of minimal points of binomial point process $\mathcal{U}_n$ (actually with almost sure convergence) is part (i) of Lemma 2.4.2 (or Theorem 1.1(a) of [21]). Using Lemma 4.3.1, we can deduce the result for the Poisson point process $\mathcal{P}_n$. □

Fix a constant δ lying in the range $(0, 1/2)$ but otherwise arbitrary. Define the point sets

$$\mathcal{Y}_n^x := \mathcal{Y}_n \cap \left((0,1) \times (0, n^{-\delta}]\right); \qquad \mathcal{Y}_n^y := \mathcal{Y}_n \cap \left((0, n^{-\delta}] \times (0,1)\right).$$

Fix $\alpha > 0$, as given in the statement of Theorem 4.2.1. Define the variables

$$L_n^x := \sum_{\mathbf{X}\in\mathcal{Y}_n^x} \|\mathbf{X}\|^\alpha; \qquad L_n^y := \sum_{\mathbf{X}\in\mathcal{Y}_n^y} \|\mathbf{X}\|^\alpha; \qquad (4.9)$$
$$N_n^x := \text{card}(\mathcal{Y}_n^x); \qquad N_n^y := \text{card}(\mathcal{Y}_n^y).$$

Thus, L_n^x is the total weight of α-power-weighted edges of the MDST on $\mathcal{P}_n$ which are incident to the origin and lie entirely in the horizontal strip $(0,1) \times (0, n^{-\delta}]$, while N_n^x is the number of such edges; L_n^y and N_n^y are defined analogously in terms of a vertical strip.

Proposition 4.3.1 *Let $S \sim \mathrm{GD}(1/\alpha)$, i.e. let S be a generalized Dickman random variable with parameter $\theta = 1/\alpha$. Then as $n \to \infty$,*

$$L_n^x \xrightarrow{\mathcal{D}} S, \quad \text{and} \quad L_n^y \xrightarrow{\mathcal{D}} S.$$

Proof. We give the proof only for L_n^x; the argument for L_n^y is entirely analogous.

List the minimal points $\mathcal{Y}_n^x$, in order of increasing y-coordinate, as $\mathbf{X}_1^x, \ldots, \mathbf{X}_{N_n^x}^x$. In co-ordinates we set $\mathbf{X}_j^x = (X_j^x, Y_j^x)$. Since the points $\mathbf{X}_j^x$ are minimal,

$$Y_1^x < Y_2^x < \cdots < Y_{N_n^x}^x, \quad \text{and} \quad X_1^x > X_2^x > \cdots > X_{N_n^x}^x.$$

Then $L_n^x = \sum_{j=1}^{N_n^x} \|\mathbf{X}_j^x\|^\alpha$. For each n, let S_n^x be the estimate for L_n^x obtained by counting only the projections of the edge lengths onto the x-axis, i.e., set

$$S_n^x = \sum_{j=1}^{N_n^x} (X_j^x)^\alpha.$$

If $(x, y) \in (0,1)^2$ then $\|(x,y)\| \le x + y$, and by the Mean Value Theorem,

$$\|(x,y)\|^\alpha - x^\alpha \le (x+y)^\alpha - x^\alpha \le \alpha 2^{\alpha-1} y \qquad (\alpha \ge 1)$$

whereas by the concavity of the function $t \mapsto t^\alpha$ for $\alpha < 1$,

$$\|(x,y)\|^\alpha - x^\alpha \le (x+y)^\alpha - x^\alpha \le y^\alpha \qquad (0 < \alpha < 1).$$

Hence, there is a constant $C(\alpha)$ such that with probability 1,

$$\begin{aligned} 0 \le L_n^x - S_n^x \;&\le\; C(\alpha) \sum_{j=1}^{N_n^x} (Y_j^x)^{\min(1,\alpha)} \\ &\le\; C(\alpha) n^{-\delta \min(1,\alpha)} N_n^x. \end{aligned} \tag{4.10}$$

Since $N_n^x = O(\log(n))$ in probability by Lemma 4.3.2, it follows that $n^{-\delta \min(1,\alpha)} N_n^x$ converges in probability to zero as $n \to \infty$, and hence so does $L_n^x - S_n^x$. Therefore, by Slutsky's theorem it suffices to prove that

$$S_n^x \xrightarrow{\mathcal{D}} S \ \text{ as } n \to \infty. \tag{4.11}$$

We prove this by a coupling argument in which we construct (copies of) the random variables S_n^x $(n \ge 1)$ on a common probability space.

Let $\mathcal{H}$ be a homogeneous Poisson process of unit intensity on the infinite strip $(0,1) \times (0,\infty)$. Let $\mathcal{H}_n$ be the image of $\mathcal{H}$ under the linear mapping $\tau_n : \mathbf{R}^2 \to \mathbf{R}^2$ given by

$$\tau_n((x,y)) = (x, n^{-1} y). \tag{4.12}$$

By the Mapping Theorem [84], $\mathcal{H}_n$ is a homogeneous Poisson process of intensity n on the same strip $(0,1) \times (0,\infty)$. Since we are interested only in proving a convergence in distribution result (4.11), we may assume without loss of generality that $\mathcal{P}_n$ is the restriction of the Poisson process $\mathcal{H}_n$ to the unit square $(0,1)^2$.

List the minimal elements of $\mathcal{H}$ in order of increasing y-coordinate as $\mathbf{X}_1, \mathbf{X}_2, \ldots$, with coordinate representation $\mathbf{X}_j = (X_j, Y_j)$. Then $Y_1 < Y_2 < \cdots$, and $X_1 > X_2 > \cdots$. Define $U_1 = X_1$, and set

$$U_j = \frac{X_j}{X_{j-1}}, \quad j = 2, 3, \ldots.$$

It is not hard to see that $U_1, U_2, \ldots$ are mutually independent and are each uniformly distributed over $(0,1)$. Therefore, setting

$$S := \sum_{j=1}^{\infty} X_j^{\alpha} = \sum_{j=1}^{\infty} \left(\prod_{i=1}^{j} U_i^{\alpha} \right), \tag{4.13}$$

we see from Proposition C.4.1 that S has a generalized Dickman distribution GD$(1/\alpha)$.

The set of minimal elements of a point set in $\mathbf{R}^2$ is invariant under the linear transformation $\tau_n(\cdot)$ defined at (4.12), as is the relative order of the y-coordinates of the minimal elements. Therefore, under our assumption that $\mathcal{P}_n$ is the restriction of $\tau_n(\mathcal{H})$ to the unit square, we see that $\mathbf{X}_j^x = \tau_n(\mathbf{X}_j)$ for $1 \leq j \leq N_n^x$. Hence, since the mapping τ_n leaves x-coordinates unchanged,

$$S_n^x = \sum_{j=1}^{N_n^x} X_j^{\alpha}.$$

Since N_n^x is the number of minimal elements in the restriction of $\mathcal{H}$ to the set $(0,1) \times (0, n^{1-\delta}]$, it is the case that $N_n^x \to \infty$ almost surely as $n \to \infty$. Therefore, with this particular coupled construction of the point processes $\mathcal{P}_n, n \geq 1$, the variables S_n^x converge to S as $n \to \infty$, almost surely and hence in distribution. In other words, (4.11) holds as required. □

The random variables L_n^x and L_n^y are not quite independent since they both depend on the configuration of points of $\mathcal{P}_n$ in $(0, n^{-\delta}]^2$. Our argument to deal with this fact requires some further terminology. Fix a further constant β with $0 < \beta < \delta < 1/2$, and define the following rectangular regions, as shown in Figure 4.1 below:

$$R_2^x(n) := (n^{-\beta}, 1) \times (0, n^{-\delta}]; \qquad R_2^y(n) := (0, n^{-\delta}] \times (n^{-\beta}, 1);$$

$$R_1^x(n) := (n^{-\delta}, n^{-\beta}] \times (0, n^{-\delta}]; \qquad R_1^y(n) := (0, n^{-\delta}] \times (n^{-\delta}, n^{-\beta}];$$

$$R_0(n) := (0, n^{-\delta}]^2; \qquad R_3(n) := (n^{-\delta}, 1)^2.$$

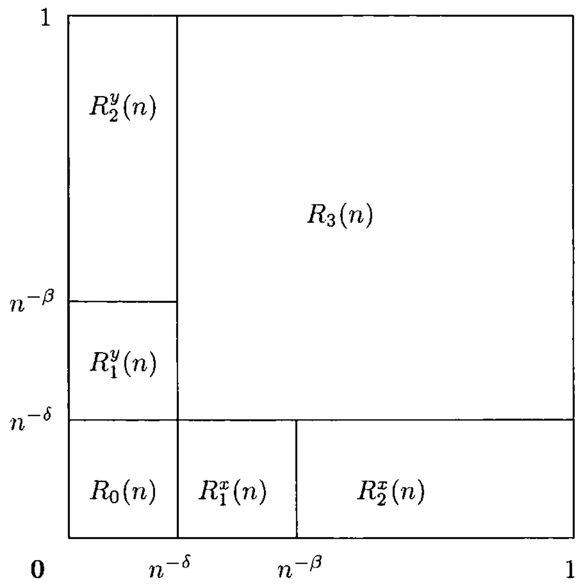


Figure 4.1: The regions of $(0,1)^2$.

Let $N_2^x(n)$, $N_2^y(n)$, $N_1^x(n)$, $N_1^y(n)$, $N_0(n)$, and $N_3(n)$ be the number of elements of $\mathcal{Y}_n$ that fall in the regions $R_2^x(n)$, $R_2^y(n)$, $R_1^x(n)$, $R_1^y(n)$, $R_0(n)$ and $R_3(n)$ respectively.

Similarly, let $L_2^x(n)$, $L_2^y(n)$, $L_1^x(n)$, $L_1^y(n)$, $L_0(n)$, and $L_3(n)$ be the total weights of edges that are incident to the origin in the MDST on $\mathcal{P}_n \cup \{\mathbf{0}\}$ and start from points that fall in the regions $R_2^x(n)$, $R_2^y(n)$, $R_1^x(n)$, $R_1^y(n)$, $R_0(n)$ and $R_3(n)$ respectively, i.e., set

$$L_2^x(n) := \sum_{\mathbf{X}\in\mathcal{Y}_n\cap R_2^x(n)} \|\mathbf{X}\|^\alpha, \quad L_1^x(n) := \sum_{\mathbf{X}\in\mathcal{Y}_n\cap R_1^x(n)} \|\mathbf{X}\|^\alpha, \quad L_0(n) := \sum_{\mathbf{X}\in\mathcal{Y}_n\cap R_0(n)} \|\mathbf{X}\|^\alpha, \tag{4.14}$$

$$L_2^y(n) := \sum_{\mathbf{X}\in\mathcal{Y}_n\cap R_2^y(n)} \|\mathbf{X}\|^\alpha, \quad L_1^y(n) := \sum_{\mathbf{X}\in\mathcal{Y}_n\cap R_1^y(n)} \|\mathbf{X}\|^\alpha, \quad L_3(n) := \sum_{\mathbf{X}\in\mathcal{Y}_n\cap R_3(n)} \|\mathbf{X}\|^\alpha. \tag{4.15}$$

Then

$$\mathcal{L}_0^\alpha(\mathcal{P}_n) = L_2^x(n) + L_2^y(n) + L_1^x(n) + L_1^y(n) + L_0(n) + L_3(n). \tag{4.16}$$

The next result shows that most of the terms in (4.16) are asymptotically negligible.

Lemma 4.3.3 *As $n \to \infty$,*

$$L_1^x(n) + L_1^y(n) + L_0(n) + L_3(n) \xrightarrow{P} 0.$$

Proof. Observe first that

$$L_1^x(n) + L_1^y(n) + L_0(n) \le (2n^{-\beta})^\alpha (N_1^x(n) + N_1^y(n) + N_0(n))$$

and since $N_1^x(n) + N_1^y(n) + N_0(n) = O(\log n)$ in probability by Lemma 4.3.2,

$$L_1^x(n) + L_1^y(n) + L_0(n) \xrightarrow{P} 0. \tag{4.17}$$

If $\text{card}(\mathcal{P}_n \cap R_0(n)) > 0$ then $L_3(n) = N_3(n) = 0$. However, $\text{card}(\mathcal{P}_n \cap R_0(n))$ is Poisson with parameter $n^{1-2\delta}$, which tends to infinity since we assume $\delta < 1/2$. Hence, $P[L_3(n) \neq 0] \to 0$, so that $L_3(n) \xrightarrow{P} 0$. Combined with (4.17), this gives us the result. □

Define $\tilde{\mathcal{P}}_n$ to be the point process $\mathcal{P}_n$ with all points in the corner region $R_0(n)$ removed, i.e., set

$$\tilde{\mathcal{P}}_n := \mathcal{P}_n \backslash R_0(n). \tag{4.18}$$

Let $\tilde{\mathcal{Y}}_n$ be the set of minimal elements of $\tilde{\mathcal{P}}_n$. Define the point sets

$$\begin{aligned} \mathcal{Z}_n^x &:= \mathcal{Y}_n \cap R_2^x(n); \quad & \tilde{\mathcal{Z}}_n^x &:= \tilde{\mathcal{Y}}_n \cap R_2^x(n); \\ \mathcal{Z}_n^y &:= \mathcal{Y}_n \cap R_2^y(n); \quad & \tilde{\mathcal{Z}}_n^y &:= \tilde{\mathcal{Y}}_n \cap R_2^y(n). \end{aligned} \tag{4.19}$$

Then $\mathcal{Z}_n^x \subseteq \tilde{\mathcal{Z}}_n^x$, since adding the points in $R_0(n)$ cannot cause any new minimal points in $R_2^x(n)$ to be created, although it can cause previously minimal points in $R_2^x(n)$ to cease to be minimal. Using the convention $\min \emptyset = +\infty$, set

$$Y_0^-(n) := \min\{Y : \mathbf{X} = (X, Y) \in \mathcal{P}_n \cap R_0(n)\},$$

which is the y-coordinate of the lowest point of $\mathcal{P}_n$ in $R_0(n)$ (or $+\infty$ if no such point exists). Let

$$Y_1^-(n) := \min\{Y : \mathbf{X} = (X, Y) \in \mathcal{P}_n \cap R_1^x(n)\},$$

which is the y-coordinate of the lowest point of $\mathcal{P}_n$ in $R_1^x(n)$ (or $+\infty$ if there are no such points).

Lemma 4.3.4 *If $Y_1^-(n) < Y_0^-(n)$, then $\mathcal{Z}_n^x = \tilde{\mathcal{Z}}_n^x$.*

Proof. If $\mathbf{X} = (X, Y) \in R_2^x(n)$ and $\mathbf{X}' = (X', Y') \in R_0(n) \cup R_1^x(n)$, then $\mathbf{X}' \preccurlyeq^* \mathbf{X}$ if and only if $Y' \leq Y$. Hence, $\tilde{\mathcal{Z}}_n^x$ consists of those minimal elements of $\mathcal{P}_n \cap R_2^x(n)$ that have a lower y-coordinate than $Y_1^-(n)$. Likewise, $\mathcal{Z}_n^x$ consists of those minimal elements of $\mathcal{P}_n \cap R_2^x(n)$ that have a lower y-coordinate than $\min(Y_1^-(n), Y_0^-(n))$. Thus, if $Y_1^-(n) < Y_0^-(n)$, then the sets $\tilde{\mathcal{Z}}_n^x$ and $\mathcal{Z}_n^x$ must be identical. □

Lemma 4.3.5 *As* $n \to \infty$, $P[Y_1^-(n) < Y_0^-(n)] \to 1$.

Proof. Assume without loss of generality that $\mathcal{P}_n$ is the restriction to $(0,1)^2$ of a Poisson process $\mathcal{H}_n$ of intensity n on $(0,1) \times (0,\infty)$. List the points of $\mathcal{H}_n \cap \big((0, n^{-\beta}] \times (0,\infty)\big)$ in order of increasing y-coordinate as $\mathbf{V}_1^n, \mathbf{V}_2^n, \mathbf{V}_3^n \ldots$. In coordinates, write $\mathbf{V}_1^n = (V_1^n, W_1^n)$. Then V_1^n is uniform on $(0, n^{-\beta}]$ and is independent of W_1^n. Also W_1^n is exponential with parameter $n^{1-\beta}$. Since $\beta < \delta$ and $\delta < 1/2 < 1-\beta$,

$$P[\{V_1 \in (n^{-\delta}, n^{-\beta}]\} \cap \{W_1 < n^{-\delta}\}] \to 1 \quad \text{as } n \to \infty.$$

However, if this event occurs then $Y_1^-(n) < Y_0^-(n)$ so the proof is complete. □

Define the random variables

$$\tilde{\mathcal{L}}_2^x(n) := \sum_{\mathbf{X} \in \tilde{\mathcal{Z}}_n^x} \|\mathbf{X}\|^\alpha, \quad \tilde{\mathcal{L}}_2^y(n) := \sum_{\mathbf{X} \in \tilde{\mathcal{Z}}_n^y} \|\mathbf{X}\|^\alpha.$$

In other words, $\tilde{\mathcal{L}}_2^x(n)$, $\tilde{\mathcal{L}}_2^y(n)$ are the total weight of edges from points in $R_2^x(n)$, $R_2^y(n)$ respectively joined to the origin in the MDST on $\tilde{\mathcal{P}}_n \cup \{\mathbf{0}\}$.

We assert that $\tilde{\mathcal{L}}_2^x(n)$ and $\tilde{\mathcal{L}}_2^y(n)$ are independent. This follows because $\tilde{\mathcal{L}}_2^x(n)$ is determined by the configuration of $\mathcal{P}_n \cap (R_1^x(n) \cup R_2^x(n))$, whereas $\tilde{\mathcal{L}}_n^y$ is determined by the configuration of $\mathcal{P}_n \cap (R_1^y(n) \cup R_2^y(n))$. Since the regions $R_1^x(n) \cup R_2^x(n)$ and $R_1^y(n) \cup R_2^y(n)$ are disjoint, the independence asserted follows from the standard spatial independence properties of the Poisson process.

Proof of Theorem 4.2.1. By the earlier definitions at (4.9) and (4.14), $L_n^x = L_2^x(n) + L_1^x(n) + L_0(n)$. Hence,

$$\tilde{\mathcal{L}}_2^x(n) = (L_n^x - L_1^x(n) - L_0(n)) + (\tilde{\mathcal{L}}_2^x(n) - L_2^x(n)).$$

By Lemma 4.3.3, $L_1^x(n) + L_0(n) \xrightarrow{P} 0$. Also, by Lemmas 4.3.4 and 4.3.5, $\tilde{\mathcal{L}}_2^x(n) - L_2^x(n) \xrightarrow{P} 0$. Hence, by Proposition 4.3.1, and Slutsky's theorem,

$$\tilde{\mathcal{L}}_2^x(n) \xrightarrow{\mathcal{D}} S,$$

where $S \sim \mathrm{GD}(1/\alpha)$, and by an analogous argument we obtain $\tilde{\mathcal{L}}_2^y(n) \xrightarrow{\mathcal{D}} S$.

Let S and S' be independent $\mathrm{GD}(1/\alpha)$ variables and let $Z(2/\alpha) \sim GD(2/\alpha)$. Since $\tilde{\mathcal{L}}_2^x(n)$ and $\tilde{\mathcal{L}}_2^y(n)$ are independent, we obtain

$$\tilde{\mathcal{L}}_2^x(n) + \tilde{\mathcal{L}}_2^y(n) \xrightarrow{\mathcal{D}} S + S' \stackrel{\mathcal{D}}{=} Z(2/\alpha), \tag{4.20}$$

where the last distributional equality follows from Proposition C.4.2 (b). By (4.16),

$$\mathcal{L}_0^\alpha(\mathcal{P}_n) - (\tilde{\mathcal{L}}_2^x(n) + \tilde{\mathcal{L}}_2^y(n)) = (L_2^x(n) - \tilde{\mathcal{L}}_2^x(n)) + (L_2^y(n) - \tilde{\mathcal{L}}_2^y(n)) \\ + L_1^x(n) + L_1^y(n) + L_0(n) + L_3(n).$$

In this expression, the right hand side tends to zero in probability by Lemmas 4.3.3, 4.3.4 and 4.3.5. Hence, by (4.20) and Slutsky's theorem we obtain (4.2).

Next we prove (4.3). To do this we use the coupled copies $\mathcal{U}_n'$ and $\mathcal{P}_n'$ of $\mathcal{U}_n$, $\mathcal{P}_n$ respectively, given by Lemma 4.3.1. That result shows that

$$\mathcal{L}_0^\alpha(\mathcal{P}_n') - \mathcal{L}_0^\alpha(\mathcal{U}_n') \xrightarrow{P} 0, \text{ as } n \to \infty. \tag{4.21}$$

Since $\mathcal{P}_n' \stackrel{\mathcal{D}}{=} \mathcal{P}_n$, we see from (4.2) that $\mathcal{L}_0^\alpha(\mathcal{P}_n')$ converges in distribution to the GD($2/\alpha$) variable $Z(2/\alpha)$. By (4.21) and Slutsky's theorem, the same is true of $\mathcal{L}_0^\alpha(\mathcal{U}_n')$, and (4.3) follows since $\mathcal{U}_n' \stackrel{\mathcal{D}}{=} \mathcal{U}_n$. □

In the case $\alpha = 1$, the limiting variable $Z(2)$ is GD(2); its moments and moment generating function are obtained by application of Proposition C.4.2. □

4.4 Proof of Theorem 4.2.3

The intuition behind Theorem 4.2.3 is that the longest edge is likely to be near either the x-axis or y-axis. Near the x-axis, the x-coordinates of the points of $\mathcal{P}_n$ (or $\mathcal{U}_n$), taken in order of increasing y-coordinate, form a sequence of uniforms with each uniform joined to its nearest predecessor lying to its left. Similarly for the y-coordinate.

The proof of Theorem 4.2.3 follows similar lines to that of Theorem 4.2.1 (see Section 4.3). Fix a constant $\delta \in (1/2, 1)$. Define the point sets

$$\mathcal{P}_n^x := \mathcal{P}_n \cap \left((0,1) \times (0, n^{-\delta}]\right); \qquad \mathcal{P}_n^y := \mathcal{P}_n \cap \left((0, n^{-\delta}] \times (0,1)\right). \tag{4.22}$$

For $\mathbf{X} \in \mathcal{P}_n$, if $\mathbf{X}'$ is the directed nearest neighbour of $\mathbf{X}$ in $\mathcal{P}_n$, write $d(\mathbf{X})$ for the length of the edge from $\mathbf{X}$ in the MDST, i.e. $d(\mathbf{X}) = \|\mathbf{X} - \mathbf{X}'\|$. Define

$$M_n^x := \max_{\mathbf{X} \in \mathcal{P}_n^x} d(\mathbf{X}); \qquad M_n^y := \max_{\mathbf{X} \in \mathcal{P}_n^y} d(\mathbf{X}). \tag{4.23}$$

Thus, M_n^x is the length of the longest edge in the MDST on $\mathcal{P}_n$ from points in the horizontal strip $(0,1) \times (0, n^{-\delta}]$; M_n^y is defined analogously in terms of a vertical strip.

Proposition 4.4.1 *Let M have the max-Dickman distribution given by (C.7.29). Then as $n \to \infty$,*

$$M_n^x \xrightarrow{\mathcal{D}} M, \quad \text{and} \quad M_n^y \xrightarrow{\mathcal{D}} M.$$

Proof. We give the proof only for M_n^x; the argument for M_n^y is entirely analogous.

Define the random variable $\nu(n) := \text{card}(\mathcal{P}_n^x)$. List the points of $\mathcal{P}_n^x$, in order of increasing y-coordinate, as $\mathbf{X}_1^x, \mathbf{X}_2^x, \mathbf{X}_3^x, \ldots, \mathbf{X}_{\nu(n)}^x$. In co-ordinates we set $\mathbf{X}_j^x = (X_j^x, Y_j^x)$. Then $Y_1^x < Y_2^x < \cdots < Y_{\nu(n)}^x$.

For each n, let ξ_n^x be the estimate for M_n^x obtained by considering only the projections of the edge lengths onto the x-axis, i.e., set

$$\xi_n^x = \max_{1 \le i \le \nu(n)} \left\{ X_i^x - \max_{0 \le j < i} \left(X_j^x \mathbf{1}_{\{X_j^x < X_i^x\}} \right) \right\}. \tag{4.24}$$

where we set $X_0^x := 0$.

By construction of the MDST and the triangle inequality, with probability 1,

$$0 \le M_n^x - \xi_n^x \le n^{-\delta},$$

so that $M_n^x - \xi_n^x$ converges to 0 almost surely. Therefore, by Slutsky's theorem it suffices to prove that

$$\xi_n^x \xrightarrow{\mathcal{D}} M \quad \text{as } n \to \infty. \tag{4.25}$$

As in the proof of Proposition 4.3.1, let $\mathcal{H}$ be a homogeneous Poisson process of unit intensity on the infinite strip $(0,1) \times (0,\infty)$, and let $\mathcal{H}_n$ be the image of $\mathcal{H}$ under the linear mapping τ_n defined at (4.12). Again, we may assume without loss of generality that $\mathcal{P}_n$ is the restriction of the Poisson process $\mathcal{H}_n$ to the unit square $(0,1)^2$.

List the elements of $\mathcal{H}$ in order of increasing y-coordinate as $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \ldots$, with coordinate representation $\mathbf{X}_j = (X_j, Y_j)$. Since the linear mapping τ_n preserves x-coordinates and the relative order of y-coordinates, our coupling of $\mathcal{P}_n$ to $\mathcal{H}$ means that the sequence $X_1^x, \ldots X_{\nu(n)}^x$ is identical to the first $\nu(n)$ terms in the infinite sequence $(X_1, X_2, \ldots)$.

An *upper record value* (compare Section 2.4.3) in the sequence $X_1, X_2, X_3, \ldots$ is a value X_i which exceeds $\max\{X_1, \ldots, X_{i-1}\}$ (the first value X_1 is also included as a record value). Let $j(1), j(2), j(3), \ldots$ be the values of $i \in \{1, 2, 3, \ldots\}$ such that X_i is a record value, arranged in increasing order so that $1 = j(1) < j(2) < j(3) < \cdots$. Let $R_n := \max\{k : j(k) \le \nu(n)\}$ be the number of record values in the finite sequence $(X_1, X_2 \ldots, X_{\nu(n)})$.

Since each non-record X_i lies in an interval between preceding record values, the first maximum in the definition at (4.24) is achieved at a record value, so that

$$\xi_n^x = \max_{1 \le i \le R_n} \left\{ X_{j(i)} - X_{j(i-1)} \right\}, \tag{4.26}$$

where we set $j(0) = 0$ and $X_0 = 0$. Define $U_1 = 1 - X_1$, and set

$$U_i = \frac{1 - X_{j(i)}}{1 - X_{j(i-1)}}, \quad i = 2, 3, \ldots.$$

It is not hard to see that $U_1, U_2, \ldots$ are mutually independent and are each uniformly distributed over $(0,1)$. Therefore, setting

$$M := \max\{1 - U_1, U_1(1 - U_2), U_1U_2(1 - U_3), U_1U_2U_3(1 - U_4), \ldots\}, \tag{4.27}$$

we see that M indeed has the max-Dickman distribution as described in Proposition C.7.1 (b). Furthermore,

$$(1 - U_k) \prod_{i=1}^{k-1} U_i = \frac{X_{j(k)} - X_{j(k-1)}}{1 - X_{j(k-1)}} \prod_{i=1}^{k-1} \left(\frac{1 - X_{j(i)}}{1 - X_{j(i-1)}} \right) = X_{j(k)} - X_{j(k-1)}, \tag{4.28}$$

for $k = 2, 3, \ldots$.

With our chosen coupling of $\mathcal{P}_n$ to $\mathcal{H}$, $\nu(n) := \text{card}(\mathcal{P}_n^x)$ is the number of points in the restriction of $\mathcal{H}$ to the set $(0,1) \times (0, n^{1-\delta}]$, so that $\nu(n) \to \infty$ almost surely as $n \to \infty$. Therefore, since there are almost surely infinitely many records, $R_n \to \infty$ almost surely as $n \to \infty$. Hence by (4.26), (4.27) and (4.28), $\xi_n^x \to M$ as $n \to \infty$, almost surely with this coupling. Hence, (4.25) holds as required. □

Let $M_3(n)$ denote the maximum edge length of edges of the MDST on $\mathcal{P}_n$ starting in $(n^{-\delta}, 1)^2$, i.e., set

$$M_3(n) := \max\{\|d(\mathbf{X})\| : \mathbf{X} \in \mathcal{P}_n \cap (n^{-\delta}, 1)^2\}. \tag{4.29}$$

Lemma 4.4.1 *It is the case that* $M_3(n) \xrightarrow{P} 0$ *as* $n \to \infty$.

Proof. Recall that $(1/2) < \delta < 1$. Choose a second constant $\varepsilon \in (0, 1 - \delta)$. Consider a collection of overlapping horizontal and vertical rectangles of the form

$$((i-1)n^{-\varepsilon}, in^{-\varepsilon}] \times ((j-1)n^{-\delta}, jn^{-\delta}], \qquad (i,j) \in \mathbf{N} \times \mathbf{N}, i \le \lfloor n^{\varepsilon} \rfloor, j \le \lfloor n^{\delta} \rfloor,$$
$$((i-1)n^{-\delta}, in^{-\delta}] \times ((j-1)n^{-\varepsilon}, jn^{-\varepsilon}], \qquad (i,j) \in \mathbf{N} \times \mathbf{N}, i \le \lfloor n^{\delta} \rfloor, j \le \lfloor n^{\varepsilon} \rfloor.$$

For each rectangle, the number of points of $\mathcal{P}_n$ in the rectangle is Poisson with parameter $n^{1-\delta-\varepsilon}$, so that the probability that at least one subsquare contains no point of $\mathcal{P}_n$ is bounded by

$$2n^{\delta+\varepsilon}\exp(-n^{1-\delta-\varepsilon}) \to 0.$$

However, if each rectangle contains at least one point of $\mathcal{P}_n$ then $M_3(n)$ is bounded by $3n^{-\varepsilon}$, and the result follows. □

Proof of Theorem 4.2.3 It is a little easier to deal with the non-independence of M_n^x and M_n^y than with the corresponding problem in the proof of Theorem 4.2.1. Define $\tilde{M}_n^x$ to be the maximal edge-length of edges starting in $(0,1)\times(0,n^{-\delta}]$ for the MDST on the point set

$$(\mathcal{P}_n \cap ((n^{-\delta},1)\times(0,n^{-\delta}])) \cup \{\mathbf{0}\}.$$

In other words, $\tilde{M}_n^x$ is the same as M_n^x except that Poisson points in $(0,n^{-\delta}]^2$ are ignored in defining $\tilde{M}_n^x$. By independence properties of the Poisson process, $\tilde{M}_n^x$ is independent of M_n^y.

It is not hard to see that

$$|M_n^x - \tilde{M}_n^x| \leq 2n^{-\delta}, \qquad \text{almost surely.} \tag{4.30}$$

Let M, M' be independent random variables both having the max-Dickman distribution. By Proposition 4.4.1, equation (4.30), and Slutsky's theorem,

$$\tilde{M}_n^x \xrightarrow{\mathcal{D}} M \quad \text{and} \quad M_n^y \xrightarrow{\mathcal{D}} M,$$

and since $\tilde{M}_n^x$ and M_n^y are independent,

$$\max(\tilde{M}_n^x, M_n^y) \xrightarrow{\mathcal{D}} \max(M, M'). \tag{4.31}$$

By (4.30), with probability 1,

$$|\max(M_n^x, M_n^y) - \max(\tilde{M}_n^x, M_n^y)| \leq 2n^{-\delta},$$

so by (4.31) and Slutsky's theorem,

$$\max(M_n^x, M_n^y) \xrightarrow{\mathcal{D}} \max(M, M'). \tag{4.32}$$

Also,

$$\mathcal{M}(\mathcal{P}_n) = \max(M_n^x, M_n^y, M_3(n)),$$

so that

$$0 \leq \mathcal{M}(\mathcal{P}_n) - \max(M_n^x, M_n^y) \leq M_3(n),$$

which tends to zero in probability by Lemma 4.4.1. Hence, a further application of Slutsky's theorem to (4.32) shows that $\mathcal{M}(\mathcal{P}_n) \xrightarrow{\mathcal{D}} \max(M, M')$, i.e., (4.6) holds.

To deduce (4.7) from (4.6), consider the coupled point processes $\mathcal{U}'_n$ and $\mathcal{P}'_n$ described in Lemma 4.3.1, given in terms of a sequence of independent uniform points $\mathbf{U}_i$ in $(0,1)^2$ and an independent Poisson variable $N(n)$ as given in the proof of Lemma 4.3.1. Let B_n be the event that at least one point of the symmetric difference $\mathcal{U}'_n \triangle \mathcal{P}'_n$ lies in $(0,1)^2 \setminus (n^{-\delta}, 1)^2$. Then

$$\begin{aligned} P[B_n] &\leq P[|N(n) - n| > n^{(1/4)+(\delta/2)}] + 2n^{(1/4)+(\delta/2)} P[\mathbf{U}_1 \in (0,1)^2 \setminus (n^{-\delta}, 1)^2] \\ &\to 0, \text{ as } n \to \infty, \end{aligned} \tag{4.33}$$

where the convergence follows by Chebyshev's inequality and the fact that we took $\delta > 1/2$.

Recall that $M_3(n)$ denotes the maximum length for edges of the MDST on $\mathcal{P}_n$ starting in $(n^{-\delta}, 1)^2$; similarly, let $M'_3(n)$, respectively $\tilde{M}_3(n)$, denote the maximum edge length for edges of the MDST on $\mathcal{P}'_n$, respectively on $\mathcal{U}'_n$, starting in $(n^{-\delta}, 1)^2$. Then $M'_3(n) \xrightarrow{P} 0$ by Lemma 4.4.1, and a similar proof shows that $\tilde{M}_3(n) \xrightarrow{P} 0$ as well. Using also (4.33) we obtain

$$|\mathcal{M}(\mathcal{U}'_n) - \mathcal{M}(\mathcal{P}'_n)| \leq 2\mathbf{1}_{B_n} + M'_3(n) + \tilde{M}_3(n) \xrightarrow{P} 0,$$

and since $\mathcal{M}(\mathcal{U}_n) \stackrel{\mathcal{D}}{=} \mathcal{M}(\mathcal{U}'_n)$ and $\mathcal{M}(\mathcal{P}_n) \stackrel{\mathcal{D}}{=} \mathcal{M}(\mathcal{P}'_n)$, equation (4.7) follows from (4.6) by yet another application of Slutsky's theorem. □

Chapter 5

One dimensional nearest-neighbour type graphs

In this chapter we concentrate on the one dimensional case of our nearest-neighbour type graphs. Further, we restrict ourselves to uniform random points. Thus our graphs are defined on uniform random points in the unit interval $[0, 1]$. The one dimensional case has some special features which allow a rather more detailed analysis than in higher dimensions. In particular, these are the connection to the well-known theory of Dirichlet spacings, and, especially in the case of the on-line graphs, a form of *self-similarity* which enables one to use a 'divide-and-conquer' approach.

In Section 5.1 we describe the theoretical background we will need. Then in the subsequent sections we deal with each of our four graphs in turn: the nearest-neighbour (directed) graph, the on-line nearest-neighbour graph, and the so-called 'directed linear tree', which is an on-line version of the MDST.

We endeavour to make this Chapter largely self-contained. However, for the main results on the MDST presented in Chapter 6, we only require the results presented in Section 5.2. The preliminary material presented in Section 5.1, and the subsequent analysis in Sections 5.3, 5.4 and 5.5, may therefore be omitted on a first reading.

5.1 Preliminaries

5.1.1 Spacings

All of the one-dimensional models in this chapter can be viewed in terms of the *spacings* of points in the unit interval. The theory of so-called Dirichlet spacings will be useful in the analysis of the one-dimensional graphs considered in the sequel. For some general references on spacings, see for example [116]. A large number of statistical tests are based on spacings, see e.g. [38] for a few examples.

Recall that $\mathcal{U}_n$ denotes the binomial point process consisting of n independent uniform random variables on $(0,1)$, $U_1, U_2, \ldots, U_n$. Given $\{U_1, \ldots, U_n\} \subseteq (0,1)$, denote the order statistics of $U_1, \ldots, U_n$, taken in increasing order, as $U^n_{(1)}, U^n_{(2)}, \ldots, U^n_{(n)}$. Thus $(U^n_{(1)}, \ldots, U^n_{(n)})$ is a nondecreasing sequence, forming a permutation of the original $(U_1, \ldots, U_n)$.

The points $U_1, \ldots, U_n$ divide $[0,1]$ into $n+1$ intervals. Denote the intervals between points by $I^n_j := (U^n_{(j-1)}, U^n_{(j)})$ for $j = 1, 2, \ldots, n+1$, where we set $U^n_{(0)} := 0$ and $U^n_{(n+1)} := 1$. Let the widths of these intervals (the spacings) be

$$S^n_j := |I^n_j| = U^n_{(j)} - U^n_{(j-1)},$$

for $j = 1, 2, \ldots, n+1$. Recall that for $n \in \mathbf{N}$, $\Delta_n \subset \mathbf{R}^n$ is the n-dimensional simplex, as given by (C.1.1). By the definition of S^n_j, we have that $S^n_j \geq 0$ for $j = 1, \ldots, n+1$ and $\sum_{j=1}^{n+1} S^n_j = 1$. So we see that the vector $(S^n_1, S^n_2, \ldots, S^n_{n+1})$ is completely specified by any n of its $n+1$ components, and any such n-vector belongs to the simplex Δ_n. Any such n-vector is, in fact, uniformly distributed over the simplex. To see this, observe that $(U^n_{(1)}, \ldots, U^n_{(n)})$ is uniformly distributed over

$$\{(x_1, \ldots, x_n) : 0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1\}.$$

Now

$$\begin{pmatrix} S^n_1 \\ S^n_2 \\ S^n_3 \\ \vdots \\ S^n_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix} \begin{pmatrix} U^n_{(1)} \\ U^n_{(2)} \\ U^n_{(3)} \\ \vdots \\ U^n_{(n)} \end{pmatrix}.$$

The n by n matrix here has determinant 1. Hence $(S^n_1, \ldots, S^n_n)$ is uniform over the simplex Δ_n, and $S^n_{n+1} = 1 - \sum_{i=1}^n S^n_i$.

Thus $(S_1^n, S_2^n, \dots, S_{n+1}^n)$ has the symmetric *Dirichlet distribution* with parameter 1 (see Section C.1, or [24], p. 246), and any n-vector of the S_j^n has the Dirichlet density

$$f(x_1, \dots, x_n) = n!, \quad (x_1, \dots, x_n) \in \Delta_n. \tag{5.1}$$

In particular, the spacings S_j^n, $j = 1, \dots, n+1$ are *exchangeable* – the distribution of $(S_1^n, S_2^n, \dots, S_{n+1}^n)$ is invariant under any permutation of its components.

By integrating out over the simplex, from (5.1) one can readily obtain the marginal distributions for the spacings. Thus, for $n \geq 1$, a single spacing has density

$$f(x_1) = n(1 - x_1)^{n-1}, \quad 0 \leq x_1 \leq 1, \tag{5.2}$$

while for $n \geq 2$, any two spacings have joint density

$$f(x_1, x_2) = n(n-1)(1 - x_1 - x_2)^{n-2}, \quad (x_1, x_2) \in \Delta_2, \tag{5.3}$$

and for $n \geq 3$ any three spacings have joint density

$$f(x_1, x_2, x_3) = n(n-1)(n-2)(1 - x_1 - x_2 - x_3)^{n-3}, \quad (x_1, x_2, x_3) \in \Delta_3. \tag{5.4}$$

It then follows from (5.2) that, for $\beta > 0$, $n \geq 1$

$$E\left[(S_1^n)^\beta\right] = \frac{\Gamma(n+1)\Gamma(\beta+1)}{\Gamma(n+\beta+1)}, \tag{5.5}$$

and from (5.3) that for $\beta > 0$, $n \geq 2$

$$E\left[(S_1^n)^\beta (S_2^n)^\beta\right] = \frac{\Gamma(n+1)\Gamma(\beta+1)^2}{\Gamma(n+2\beta+1)}. \tag{5.6}$$

When considering nearest-neighbour graphs, we will encounter the *minimum* of two (or more) spacings. The following results will be useful.

Lemma 5.1.1 *For $n \geq 1$,*

$$\min\{S_1^n, S_2^n\} \stackrel{\mathcal{D}}{=} S_1^n/2. \tag{5.7}$$

For $n \geq 2$,

$$(S_1^n, \min\{S_2^n, S_3^n\}) \stackrel{\mathcal{D}}{=} (S_1^n, S_2^n/2). \tag{5.8}$$

Finally, for $n \geq 3$

$$(\min\{S_1^n, S_2^n\}, \min\{S_3^n, S_4^n\}) \stackrel{\mathcal{D}}{=} (S_1^n/2, S_2^n/2), \tag{5.9}$$

and

$$\min\{S_1^n, S_2^n, S_3^n\} \stackrel{\mathcal{D}}{=} S_1^n/3. \tag{5.10}$$

Proof. We give the proof of (5.7). The other results follow by very similar calculations based on (5.3) and (5.4). Suppose $n \geq 2$. From (5.3), we have, for $0 \leq r \leq 1/2$

$$\begin{aligned} P[\min\{S_1^n, S_2^n\} > r] &= P[S_1^n > r,\ S_2^n > r] \\ &= n(n-1)\int_r^{1-r} \mathrm{d}x_1 \int_r^{1-x_1} (1 - x_1 - x_2)^{n-2}\mathrm{d}x_2 \\ &= (1-2r)^n = P[S_1^n > 2r], \end{aligned}$$

and so we have (5.7). □

5.1.2 The contraction method for distributional recurrences

Suppose that, for $d \in \mathbf{N}$, a sequence of random d-vectors Y_n, $n \in \mathbf{N}$, satisfies

$$Y_n \stackrel{\mathcal{D}}{=} \sum_{r=1}^{k} A_r(n) Y^{\{r\}}_{I_r(n)} + B(n), \tag{5.11}$$

where $k \in \mathbf{N}$, $Y_n^{\{r\}}$, $r = 1, \ldots, k$, are independent copies of the random vector Y_n, $I_r(n) \in \{0, \ldots, n\}$ for $r = 1, \ldots, k$ are random cardinalities, $A_1(n), \ldots, A_k(n)$ are random $d \times d$ matrices, $B(n)$ is a random d-vector, and the random vector $(A_1(n), \ldots, A_k(n), B(n), I_1(n), \ldots, I_k(n))$ is independent of $(Y_n^{\{1\}}, \ldots, Y_n^{\{k\}})$.

The equation (5.11) is an example of a *recursive distributional equation* (see [2]) or *divide-and-conquer* recurrence. Examples of such equations appear in random recursive structures, such as trees or divide-and-conquer algorithms, where the random variable of interest can be decomposed into a sum of copies of itself – that is, some sort of *self-similarity* is present.

In recent years, there has been considerable interest in equations like (5.11), in particular, convergence in distribution results for Y_n. That is, to prove that $Y_n \xrightarrow{\mathcal{D}} Y$, where Y satisfies some fixed-point equation of the form

$$Y \stackrel{\mathcal{D}}{=} \sum_{r=1}^{k} A_r Y^{\{r\}} + B, \tag{5.12}$$

where $k \in \mathbf{N}$, Y and B are random d-vectors, $A_1, \ldots, A_k$ are random $d \times d$ matrices, $Y^{\{r\}}$, $r = 1, \ldots, k$, are independent copies of the random vector Y, and $(A_1, \ldots, A_k, B)$ is a random vector, independent of $(Y^{\{1\}}, \ldots, Y^{\{k\}})$.

One method for analysing the conditions on (5.11) such that some limit of the type (5.12) exists, now developed in considerable generality, is the so-called contraction method. The name of the method comes from the fact that it relies on the property that, under some appropriate choice of metric on probability distributions, the map from the law of some Y to the law of $\sum_{r=1}^{k} A_r Y^{\{r\}} + B$ is a contraction. Hence there exists some fixed-point distribution, and under some appropriate conditions that fixed-point is unique.

We appeal to a general result of Neininger and Rüschendorf [102] for our results. For further background on the contraction method and divide-and-conquer recurrences, see also the references in [102], and in particular [40, 124, 125] and Chapter 9 of [118].

The metrics that turn out to be most useful here are the Zolotarev ζ_s metrics, $s > 0$. See [144], and also [102, 117, 118]. For our purposes, we need only $s = 2$ (however, $s = 3$ seems to be required for Conjecture 5.2.1 below). The metric ζ_2 is defined for d-dimensional random vectors X, Y by

$$\zeta_2(X, Y) := \sup_{f \in \mathcal{F}_2} |E[f(X) - f(Y)]|, \tag{5.13}$$

where for $\mathcal{F}_2$ is the set of all continuous once-differentiable functions $f : \mathbf{R}^d \to \mathbf{R}$ such that $||f'(x) - f'(y)|| \leq ||x - y||$ for all $x, y \in \mathbf{R}^d$. For some properties of ζ_s, see [102, 117]. In particular (see [117] Chapter 14), convergence in ζ_s implies convergence is distribution.

Suppose the sequence of random variables (Y_n) satisfies (5.11) as described. Suppose $d \in \{1, 2\}$. Define the normalised quantities, for $n \geq 0$

$$\tilde{Y}_n := Y_n - E[Y_n]. \tag{5.14}$$

Then $(\tilde{Y}_n)$ satisfies the modified recurrence

$$\tilde{Y}_n \stackrel{\mathcal{D}}{=} \sum_{r=1}^{k} A_r(n) \tilde{Y}^{\{r\}}_{I_r(n)} + \tilde{B}(n), \tag{5.15}$$

where, given $I_1(n), \ldots, I_k(n)$ and $A_1(n), \ldots, A_k(n)$,

$$\tilde{B}(n) = \left(B(n) - E[Y_n] + \sum_{r=1}^{k} A_r(n) E[Y_{I_r(n)}] \right).$$

Let $|| \cdot ||_{\text{op}}$ denote the d-dimensional operator norm, that is for a $d \times d$ matrix A, v a d-vector and $|| \cdot ||$ the Euclidean norm,

$$||A||_{\text{op}} := \sup_{v : ||v|| = 1} ||Av||.$$

When $d = 1$, we have $||A||_{\text{op}} = ||A||$.

The following result is contained in Theorem 4.1 of [102].

Lemma 5.1.2 *[102] Let $(\tilde{Y}_n)$ be given by (5.14), and suppose $E[|\tilde{Y}_n|^2] < \infty$. Suppose the following conditions on the random variables in (5.15) hold:*

$$(A_1(n), \ldots, A_k(n), \tilde{B}(n)) \xrightarrow{L^2} (A_1, \ldots, A_k, B), \tag{5.16}$$

$$E \sum_{r=1}^{k} \|A_r\|_{\mathrm{op}}^2 < 1, \tag{5.17}$$

$$E\left[\mathbf{1}_{\{I_r(n) \leq \ell\} \cup \{I_r(n)=n\}} \|A_r(n)\|_{\mathrm{op}}^2\right] \to 0, \quad \text{for all } \ell \in \mathbf{N}, r = 1, \ldots, k, \tag{5.18}$$

as $n \to \infty$. Then $(\tilde{Y}_n)$ converges to a limit $\tilde{Y}$,

$$\zeta_2(\tilde{Y}_n, \tilde{Y}) \to 0,$$

as $n \to \infty$, where $\mathcal{L}(\tilde{Y})$ is the unique law for twice-integrable random d-vectors with zero mean satisfying the fixed point equation

$$\tilde{Y} \stackrel{\mathcal{D}}{=} \sum_{r=1}^{k} A_r \tilde{Y}^{\{r\}} + B, \tag{5.19}$$

where $(A_1, \ldots, A_k, B)$ is independent of $(\tilde{Y}^{\{1\}}, \ldots, \tilde{Y}^{\{k\}})$ and $\tilde{Y}^{\{i\}}$, $i = 1, \ldots, k$ are independent copies of $\tilde{Y}$.

Note that this result says that, under the conditions of the lemma, the fixed-point solution to (5.19) is *unique*. This is essentially a consequence of the contraction mapping theorem; see Theorem 3 of Rösler [124] (proved using the contraction mapping theorem in the case $s = 2$; see also [102, 125]). This result will guarantee uniqueness of solutions to all the distributional fixed-point equalities considered in the sequel.

Some of our convergence in distribution results in Sections 5.2.2 and 5.2.3 are stated in terms of distributions that are (unique) solutions to fixed-point equations of the type (5.19). We define these random variables in the appropriate sections.

5.2 Results

Here we state our results for the one-dimensional nearest-neighbour type graphs we consider (see Chapter 2 for definitions and notation). We prove these results in the subsequent sections of this chapter.

5.2.1 The nearest-neighbour graph in one dimension: results

Our next result gives exact expressions for the expectation and variance of the total weight $\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)$ of the nearest neighbour (directed) graph on n independent uniform random points in the unit interval. Let ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ denote the (Gauss) hypergeometric function. For $n \in \{2,3,\ldots\}$, $\alpha > 0$, set

$$J_{n,\alpha} := 6^{-\alpha-1} \frac{\Gamma(n+1)\Gamma(2+2\alpha)}{(\alpha+1)\Gamma(2\alpha+n+1)} \, {}_2F_1(-\alpha, 1+\alpha; \alpha+2; 1/3). \tag{5.20}$$

Also, for $\alpha > 0$, set

$$j_\alpha := 8 \lim_{n\to\infty} (n^{2\alpha} J_{n,\alpha}) = 8 \cdot \frac{6^{-\alpha-1}\Gamma(2+2\alpha)}{(1+\alpha)} \, {}_2F_1(-\alpha, 1+\alpha; \alpha+2; 1/3). \tag{5.21}$$

Theorem 5.2.1 *For $n \in \{2,3,4,\ldots\}$ and $\alpha > 0$*

$$E[\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)] = ((n-2)2^{-\alpha}+2)\frac{\Gamma(n+1)\Gamma(\alpha+1)}{\Gamma(n+\alpha+1)} \sim 2^{-\alpha}\Gamma(\alpha+1)n^{1-\alpha}, \tag{5.22}$$

as $n \to \infty$. Also, for $n \in \{4,5,6,\ldots\}$ and $\alpha > 0$

$$\begin{aligned} \mathrm{Var}[\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)] &= \frac{\Gamma(n+1)}{\Gamma(n+2\alpha+1)} \left[\Gamma(2\alpha+1)(2 - 2\cdot 3^{-2\alpha} + 4^{-\alpha}n + 2\cdot 3^{-1-2\alpha}n)\right. \\ &\quad \left. +\Gamma(\alpha+1)^2(4 + 12\cdot 4^{-\alpha} - 12\cdot 2^{-\alpha} + 2^{2-\alpha}n - 7\cdot 4^{-\alpha}n + 4^{-\alpha}n^2)\right] \\ &\quad - \left(E[\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)]\right)^2 + 8(n-3)J_{n,\alpha}, \end{aligned} \tag{5.23}$$

where $E[\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)]$ is given by (5.22) and $J_{n,\alpha}$ is given by (5.20). Further, as $n \to \infty$, we have, for $\alpha > 0$

$$n^{2\alpha-1}\mathrm{Var}[\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)] \to V_\alpha, \tag{5.24}$$

where

$$V_\alpha := (4^{-\alpha} + 2\cdot 3^{-1-2\alpha})\Gamma(2\alpha+1) - 4^{-\alpha}(3+\alpha^2)\Gamma(\alpha+1)^2 + j_\alpha, \tag{5.25}$$

with j_α given by (5.21).

Using (5.23), with (5.20), one can obtain, for instance

$$\mathrm{Var}[\mathcal{N}_1^{1,1}(\mathcal{U}_n)] = \frac{2n^2+17n+12}{12(n+1)^2(n+2)} = \frac{1}{6}n^{-1} + O(n^{-2}),$$

and

$$\mathrm{Var}[\mathcal{N}_1^{1,2}(\mathcal{U}_n)] = \frac{85n^3+3645n^2+7154n-456}{108(n+1)^2(n+2)^2(n+3)(n+4)} = \frac{85}{108}n^{-3} + O(n^{-4}).$$

Also, the limiting constants j_α can be evaluated explicitly, so that one can obtain values for V_α as given by (5.25). Table 5.1 below gives some values of V_α. We prove Theorem 5.2.1 in Section 5.3.1. In Section 5.3.2 we give a brief discussion on the values of the corresponding limiting variances in the Poisson case, and give a "Poissonized" version of

α	$1/2$	1	2	3	4
V_α	$\frac{1}{2}+\sqrt{2}\arcsin(1/\sqrt{3})-\frac{13\pi}{32}\approx 0.094148$	$\frac{1}{6}$	$\frac{85}{108}$	$\frac{149}{18}$	$\frac{135793}{972}$

Table 5.1: Some values of V_α.

(5.24) (see (5.58)).

Remark. One can obtain similar results to those given for the one-dimensional NNG for the one-dimensional MDST (in fact, the MDST is simpler). We do not go into detail here.

5.2.2 The on-line nearest-neighbour graph in one dimension: results

In Theorem 5.2.2 below, we present results on the total weight of ONG in $d = 1$. These complement the $d = 1$ cases of the results in Section 2.3 (Theorem 2.3.1). Our results include convergence in distribution results for the total weight (suitably centred and scaled, in some cases) of the graph. The limiting distributions are of different types depending on the value of α. For $0 < \alpha \leq 1/2$, we believe that the limits are normal (see Conjecture 5.2.1); for $\alpha > 1/2$ we define these limiting distributions in Theorem 5.2.2, in terms of distributional fixed-point equations (see Section 5.1.2).

Define the random variable $\tilde{G}_1$ to have the unique distribution that is the solution to the distributional fixed-point equation

$$\tilde{G}_1 \stackrel{\mathcal{D}}{=} \min\{U, 1-U\} + U\tilde{G}_1^{\{1\}} + (1-U)\tilde{G}_1^{\{2\}} + \frac{U}{2}\log U + \frac{1-U}{2}\log(1-U), \quad (5.26)$$

where U is uniform on $(0,1)$ and independent of the other variables on the right. We shall see later (Proposition 5.4.2) that $E[\tilde{G}_1] = 0$.

For $\alpha > 1/2$, $\alpha \neq 1$, let $\tilde{G}_\alpha$ denote a random variable with distribution characterized by the fixed-point equation

$$\tilde{G}_\alpha \stackrel{\mathcal{D}}{=} U^\alpha \tilde{G}_\alpha^{\{1\}} + (1-U)^\alpha \tilde{G}_\alpha^{\{2\}} + \min\{U, 1-U\}^\alpha + \frac{2^{-\alpha}}{\alpha - 1}\left(U^\alpha + (1-U)^\alpha - 1\right), \quad (5.27)$$

where again U is uniform on $(0,1)$ and independent of the other variables on the right. We shall see later that, for $\alpha > 1$, the $\tilde{G}_\alpha$ arise as centred versions of the random variables G_α, satisfying the slightly simpler fixed-point equation (5.28) below, and $E[\tilde{G}_\alpha] = 0$ (see

Propositions 5.4.3 and 5.4.4). For $\alpha > 1$, we have

$$G_\alpha \stackrel{\mathcal{D}}{=} U^\alpha G_\alpha{}^{\{1\}} + (1-U)^\alpha G_\alpha{}^{\{2\}} + \min\{U, 1-U\}^\alpha, \tag{5.28}$$

where again U is uniform on $(0,1)$ and independent of the other variables on the right. The expectation of G_α is given in Proposition 5.4.4.

Define the random variable $\tilde{H}_1$ to have the unique distribution that is the solution to the distributional fixed-point equation

$$\tilde{H}_1 \stackrel{\mathcal{D}}{=} U\tilde{G}_1 + (1-U)\tilde{H}_1 + \frac{U}{2} + \frac{1}{2}U\log U + \frac{1}{2}(1-U)\log(1-U), \tag{5.29}$$

where $\tilde{G}_1$ has the distribution given by (5.26), the $\tilde{G}_1$ and $\tilde{H}_1$ on the right are independent, and U is uniform on $(0,1)$ and independent of the other variables on the right. We shall see later (Proposition 5.4.7) that $E[\tilde{H}_1] = 0$.

For $\alpha > 1/2$, $\alpha \neq 1$, let $\tilde{H}_\alpha$ denote a random variable with distribution characterized by the fixed-point equation

$$\tilde{H}_\alpha \stackrel{\mathcal{D}}{=} U^\alpha \tilde{G}_\alpha + (1-U)^\alpha \tilde{H}_\alpha + U^\alpha\left(1 + \frac{2^{-\alpha}}{\alpha-1}\right) + ((1-U)^\alpha - 1)\left(\frac{1}{\alpha} + \frac{2^{-\alpha}}{\alpha(\alpha-1)}\right), \tag{5.30}$$

where $\tilde{G}_\alpha$ has the distribution given by (5.27) and U is uniform on $(0,1)$ and independent of the other variables on the right. We shall see later that, for $\alpha > 1$, the $\tilde{H}_\alpha$ arise as centred versions of the random variables H_α, satisfying the slightly simpler fixed-point equation (5.31) below, and $E[\tilde{H}_\alpha] = 0$ (see Propositions 5.4.6 and 5.4.8). For $\alpha > 1$, we have

$$H_\alpha \stackrel{\mathcal{D}}{=} U^\alpha + U^\alpha G_\alpha + (1-U)^\alpha H_\alpha, \tag{5.31}$$

where again U is uniform on $(0,1)$ and independent of the other variables on the right. The expectation of H_α is given in Proposition 5.4.8.

Theorem 5.2.2 below gives our main results for the ONG$(\mathcal{U}_n)$ in one dimension for $\alpha > 1/2$. In Section 5.4.1 we present similar results for the slightly modified graphs ONG$(\mathcal{U}_n^{0,1})$ and ONG$(\mathcal{U}_n^0)$, in Theorems 5.4.1 and 5.4.2 respectively. Let $\tilde{\mathcal{O}}^{d,\alpha}(\mathcal{U}_n) := \mathcal{O}^{d,\alpha}(\mathcal{U}_n) - E[\mathcal{O}^{d,\alpha}(\mathcal{U}_n)]$. We write $\mathcal{N}(0,\sigma^2)$ for the normal distribution with mean zero and variance σ^2. The following conjecture is a central limit theorem in the case $0 < \alpha \leq 1/2$.

Conjecture 5.2.1 *Suppose $0 < \alpha < 1/2$. Then there exists s_α with $0 < s_\alpha < \infty$, such that, as $n \to \infty$,*

$$n^{\alpha-(1/2)}\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n) \stackrel{\mathcal{D}}{\longrightarrow} \mathcal{N}(0, s_\alpha^2). \tag{5.32}$$

Also, there exists $s_{1/2}$ with $0 < s_{1/2} < \infty$, such that, as $n \to \infty$,

$$(\log n)^{-1/2}\tilde{\mathcal{O}}^{1,1/2}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_{1/2}^2). \tag{5.33}$$

It should be possible to prove this conjecture using the contraction method of [102] (as discussed in Section 5.1.2). In this case, however, unlike the other cases considered here, it seems that we need to use $s = 3$ rather than $s = 2$, and this in turn requires calculation of the variance of the quantity of interest. We hope to address this in future work.

Recall from (C.3.5) that $\gamma \approx 0.57721566$ denotes Euler's constant.

Theorem 5.2.2 *(i) For $1/2 < \alpha < 1$, we have that, as $n \to \infty$,*

$$\begin{aligned}\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}}\ & U^\alpha \tilde{H}_\alpha^{\{1\}} + (1-U)^\alpha \tilde{H}_\alpha^{\{2\}} \\ & + \left(U^\alpha + (1-U)^\alpha - \frac{2}{1+\alpha}\right)\left(\frac{1}{\alpha} - \frac{2^{-\alpha}}{\alpha(1-\alpha)}\right),\end{aligned} \tag{5.34}$$

where $\tilde{H}_\alpha^{\{1\}}, \tilde{H}_\alpha^{\{2\}}$ are independent with the distribution given by the fixed-point equation (5.30).

(ii) For $\alpha = 1$, we have that, as $n \to \infty$,

$$\begin{aligned}\mathcal{O}^{1,1}(\mathcal{U}_n) - \frac{1}{2}(\gamma + \log n) \xrightarrow{\mathcal{D}}\ & U\tilde{H}_1^{\{1\}} + (1-U)\tilde{H}_1^{\{2\}} \\ & + \frac{U}{2}\log U + \frac{1-U}{2}\log(1-U),\end{aligned} \tag{5.35}$$

where γ is given by (C.3.5) and $\tilde{H}_1^{\{1\}}, \tilde{H}_1^{\{2\}}$ are independent with distribution given by the fixed-point equation (5.29). Also, (5.35) holds in the sense of convergence of expectations.

(iii) For $\alpha > 1$, the distribution of the limit $W(1,\alpha)$ of (2.16) is given by

$$W(1,\alpha) \stackrel{\mathcal{D}}{=} U^\alpha H_\alpha{}^{\{1\}} + (1-U)^\alpha H_\alpha{}^{\{2\}},$$

where $H_\alpha{}^{\{1\}}, H_\alpha{}^{\{2\}}$ are independent with the distribution given by the fixed-point equation (5.31).

We prove Theorem 5.2.2 in Section 5.4.

Remarks. (a) In Theorem 3.6 of [106], a central limit theorem is obtained for the case $0 < \alpha < d/4$. In the context of Theorem 2.3.1, the result of [106] implies that, provided

$0 < \alpha < d/4$, as $n \to \infty$, $n^{(\alpha/d)-(1/2)}\tilde{\mathcal{O}}^{d,\alpha}(\mathcal{U}_n)$ is asymptotically normal. In [106], it is remarked that it should be possible to extend the result to the case $0 < \alpha < d/2$. Equation (5.32) would be such a result when $d = 1$.

On the other hand, when $\alpha = d/2$, (5.33) shows that $(\log n)^{-1/2}\tilde{\mathcal{O}}^{d,d/2}(\mathcal{U}_n)$ is asymptotically normal for $d = 1$. It would be of interest to determine whether the same result holds for $d \geq 2$.

(b) Of interest is the limit behaviour of $\mathcal{O}^{d,d}(\mathcal{U}_n)$ (i.e., when $\alpha = d$). When $d = 1$, we have that $\mathcal{O}^{1,1}(\mathcal{U}_n) - E[\mathcal{O}^{1,1}(\mathcal{U}_n)]$ converges in distribution to a non-normal limiting random variable (see Theorem 5.2.2 (i)). It would be interesting to determine whether $\mathcal{O}^{d,d}(\mathcal{U}_n) - E[\mathcal{O}^{d,d}(\mathcal{U}_n)]$ converges in distribution to a nondegenerate random variable for general $d = 2, 3, 4, \ldots$, and whether or not this distribution is normal.

(c) Figure 5.1 is a plot of the estimated probability density function of the limit on the right hand side of (5.35). This was obtained by performing 10^5 repeated simulations of the ONG on a sequence of 10^3 uniform (simulated) random points on $(0, 1)$. For each simulation, the expected value of $\mathcal{O}^{1,1}(\mathcal{U}_{10^3})$ was subtracted from the total length of the simulated ONG to give an approximate realization of the distributional limit. The density function was then estimated from the sample of 10^5 approximate realizations, using a window width of 0.0025. The simulated sample from which the density estimate was taken had sample mean $\approx 3 \times 10^{-3}$ and sample variance ≈ 0.0425, which are reasonably close to the expectation and variance of the limit on the right hand side of (5.35).

(d) In addition to the total weight result, we can also obtain convergence in distribution results for quantities such as the total weight of the edges joined to 0, or the length of the longest edge incident to the origin, both for the ONG on $\mathcal{U}_n$ but with initial points at 0 and 1. These limiting distributions admit characterisations in terms of fixed-point equations and properties such as moments and probability density functions can be obtained in some cases. However, since these quantities are not of direct interest to us here, we omit such results for reasons of space.

5.2.3 The directed linear forest and tree: results

The directed linear forest (DLF) and directed linear tree (DLT) are similar to the ONG in $d = 1$, with the difference that edges can only run to the left of a newly added point. Once more, we are mainly concerned with establishing second order results, i.e., weak convergence results for the distribution of the total length, suitably centred and scaled.

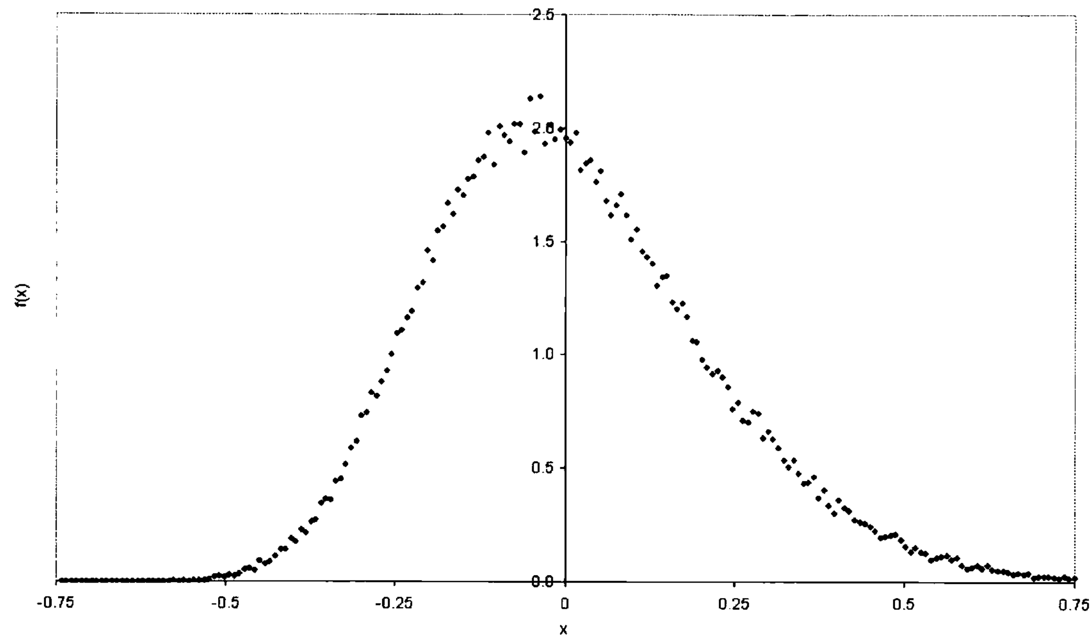


Figure 5.1: Estimated probability density function for the limit on the right hand side of (5.35).

The DLF will also be important for analysis of the MDSF in $(0,1)^2$ under $\preccurlyeq^*$ – near the boundary the MDSF can be approximated by the DLF. In the present section we derive the properties of the DLF that we need (in particular, Theorem 5.2.3); subsequently, in Theorem 6.3.1, we shall see that the total weight of edges from the points near the boundaries, as $n \to \infty$, converges in distribution to the limit of the total weight of the DLF.

In the DLF, each point in a sequence of independent uniform random points in an interval is joined to its nearest neighbour to the left, amongst those points arriving earlier in the sequence. Thus the DLF can be seen as a directed variant of the ONG, or as an on-line version of the MDSF under $\leq$.

The DLT is also of some intrinsic interest. It is constructed via a fragmentation process similar to those seen in, for example, [20] and references therein; the tree provides a historical representation of the fragmentation process.

For any finite sequence $\mathcal{T}_n = (x_1, x_2, \ldots, x_n) \in (0,1)^n$, we construct the directed linear forest (DLF) as follows. We start with the unit interval $(0,1)$ and insert the points x_i in order, one at a time, starting with $i = 1$. At the insertion of each point, we join the new point to its nearest neighbour among those points already present that lie to the *left* of the point (provided that such a point exists). In other words, for each point x_i, $i \geq 2$, we join x_i by a directed edge to the point $\max\{x_j : 1 \leq j < i,\ x_j < x_i\}$. If $\{x_j : 1 \leq j < i,\ x_j < x_i\}$ is empty, we do not add any directed edge from x_i. In this way we construct a 'directed linear forest', which we denote by $\mathrm{DLF}(\mathcal{T}_n)$. We denote the total weight (under weight function with exponent α) of $\mathrm{DLF}(\mathcal{T}_n)$ by $D^\alpha(\mathcal{T}_n)$, that is, we set

$$D^\alpha(\mathcal{T}_n) := \sum_{i=2}^{n} (x_i - \max\{x_j : 1 \leq j < i, x_j < x_i\})^\alpha \mathbf{1}_{\{\min\{x_j : 1 \leq j < i\} < x_i\}}.$$

Further, given $\mathcal{T}_n$, let $\mathcal{T}_n^0$ be the sequence $(x_0, x_1, \ldots, x_n)$ where the initial term is $x_0 := 0$. Then the DLF on $\mathcal{T}_n^0$ is constructed in the same way, where now for each $i \geq 1$, we join x_i by an edge to the point $\max\{x_j : 0 \leq j < i,\ x_j < x_i\}$. But now we see that x_1 will always be joined to $x_0 = 0$, and x_2 will be joined either to x_1 (if $x_2 > x_1$) or to x_0, and so on. In this way we construct a 'directed linear tree' (DLT) on vertex set $\{x_0, x_1, \ldots, x_n\}$ with n edges. Denote the total weight of this tree with weight exponent α by $D^\alpha(\mathcal{T}_n^0)$; that is, set

$$D^\alpha\left(\mathcal{T}_n^0\right) := \sum_{i=1}^{n} (x_i - \max\{x_j : 0 \leq j < i, x_j < x_i\})^\alpha.$$

We shall be mainly interested in the case where $\mathcal{T}_n$ is a random vector in $(0,1)^n$. In this case, set $\tilde{D}^\alpha(\mathcal{T}_n) := D^\alpha(\mathcal{T}_n) - E[D^\alpha(\mathcal{T}_n)]$ the centred total weight of the DLF, and $\tilde{D}^\alpha(\mathcal{T}_n^0) = D^\alpha(\mathcal{T}_n^0) - E[D^\alpha(\mathcal{T}_n^0)]$ the centred total weight of the DLT.

We take $\mathcal{T}_n$ to be a vector of uniform variables. Let $(U_1, U_2, U_3, \ldots)$ be a sequence of independent uniformly distributed random variables in $(0,1)$, and for $n \in \mathbf{N}$ set $\mathcal{U}_n := (U_1, U_2, \ldots, U_n)$. We consider $D^\alpha(\mathcal{U}_n)$ and $D^\alpha(\mathcal{U}_n^0)$. For these variables, we establish asymptotic behaviour of the mean value in Propositions 5.5.1 and 5.5.2, along with the following convergence results (Theorems 5.2.3 and 5.2.4), which are the principal results of this section.

Some of our weak convergence results are given in terms of distributional fixed-point equations of the form (5.19). Here we collect here all the fixed-point distributions that appear in this section.

Define the random variable $\tilde{D}_1$, to have the distribution that is the unique solution to the distributional fixed-point equation

$$\tilde{D}_1 \stackrel{\mathcal{D}}{=} U\tilde{D}_1^{\{1\}} + (1-U)\tilde{D}_1^{\{2\}} + U\log U + (1-U)\log(1-U) + U, \tag{5.36}$$

where U is uniform on $(0,1)$ and independent of the other variables on the right. We shall see later (in Propositions 5.5.7 and 5.5.8) that $E[\tilde{D}_1] = 0$ and $\mathrm{Var}[\tilde{D}_1] = 2 - \pi^2/6$; higher order moments are given recursively by (5.45).

For $\alpha > 1/2$, $\alpha \neq 1$, let $\tilde{D}_\alpha$ denote a random variable with distribution characterized by the fixed-point equation

$$\tilde{D}_\alpha \stackrel{\mathcal{D}}{=} U^\alpha \tilde{D}_\alpha^{\{1\}} + (1-U)^\alpha \tilde{D}_\alpha^{\{2\}} + \frac{\alpha}{\alpha-1} U^\alpha + \frac{1}{\alpha-1}(1-U)^\alpha - \frac{1}{\alpha-1}, \tag{5.37}$$

where again U is uniform on $(0,1)$ and independent of the other variables on the right.

For $\alpha > 1/2$, $\alpha \neq 1$, let $\tilde{F}_\alpha$ denote a random variable with distribution characterized by the fixed-point equation

$$\tilde{F}_\alpha \stackrel{\mathcal{D}}{=} U^\alpha \tilde{F}_\alpha + (1-U)^\alpha \tilde{D}_\alpha + \frac{U^\alpha}{\alpha(\alpha-1)} + \frac{(1-U)^\alpha}{\alpha-1} - \frac{1}{\alpha(\alpha-1)}, \tag{5.38}$$

where U is uniform on $(0,1)$, $\tilde{D}_\alpha$ has the distribution given by (5.37), and the U, $\tilde{D}_\alpha$ and $\tilde{F}_\alpha$ on the right are independent.

In Section 5.5 we shall see that for $\alpha > 1$ the random variables $\tilde{D}_\alpha$, $\tilde{F}_\alpha$ arise as centred versions of random variables (denoted D_α, F_α respectively) satisfying somewhat simpler fixed point equations (see below). Thus $\tilde{D}_\alpha$ and $\tilde{F}_\alpha$ both have mean zero; their variances are given by (5.151) and (5.152) below.

For $\alpha > 1$, let D_α denote a random variable with distribution characterized by the fixed-point equation

$$D_\alpha \stackrel{\mathcal{D}}{=} U^\alpha D_\alpha^{\{1\}} + (1-U)^\alpha D_\alpha^{\{2\}} + U^\alpha, \tag{5.39}$$

where U is uniform on $(0,1)$ and independent of the other variables on the right. Also for $\alpha > 1$, let F_α denote a random variable with distribution characterized by the fixed-point equation

$$F_\alpha \stackrel{\mathcal{D}}{=} U^\alpha F_\alpha + (1-U)^\alpha D_\alpha, \tag{5.40}$$

where U is uniform on $(0,1)$, D_α has the distribution given by (5.39), and the U, D_α and F_α on the right are independent. The corresponding centred random variables $\tilde{D}_\alpha := D_\alpha - E[D_\alpha]$ and $\tilde{F}_\alpha := F_\alpha - E[F_\alpha]$ satisfy the fixed-point equations (5.37) and (5.38) respectively. The solutions to equation (5.37) and equation (5.38) are unique by Lemma 5.1.2, and hence the solutions to equation (5.39) and equation (5.40) are also unique.

We conjecture that a central limit theorem analogous to Conjecture 5.2.1 holds for $\tilde{D}^\alpha(\mathcal{U}_n^0)$ and $\tilde{D}^\alpha(\mathcal{U}_n)$ in the case $0 < \alpha \leq 1/2$.

Theorem 5.2.3 *(i) For $1/2 < \alpha < 1$, we have that, as $n \to \infty$,*

$$\tilde{D}^\alpha(\mathcal{U}_n^0) \xrightarrow{\mathcal{D}} \tilde{D}_\alpha, \tag{5.41}$$

where $\tilde{D}_\alpha$ has the distribution given by the fixed-point equation (5.37).

(ii) *For $\alpha = 1$, we that, as $n \to \infty$,*

$$\tilde{D}^1(\mathcal{U}_n^0) \xrightarrow{L^2} \tilde{D}_1, \tag{5.42}$$

where $\tilde{D}_1$ has the distribution given by the fixed-point equation (5.36). Also, the variance of $\tilde{D}_1$ is $2 - \pi^2/6 \approx 0.355066$.

(iii) *For $\alpha > 1$, as $n \to \infty$ we have $D^\alpha(\mathcal{U}_n^0) \to D_\alpha$, almost surely and in L^2, where the distribution of D_α is given by the fixed-point equation (5.39). Also, $E[D_\alpha] = (\alpha - 1)^{-1}$ and $\mathrm{Var}(D_\alpha)$ is given by (5.151).*

Proof. Part (i) of Theorem 5.2.3 follows from Proposition 5.5.3. Part (ii) follows from Propositions 5.5.7 (i) and 5.5.8. Part (iii) follows from Proposition 5.5.5. □

Theorem 5.2.4 (i) *For $1/2 < \alpha < 1$, we have that, as $n \to \infty$,*

$$\tilde{D}^\alpha(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha, \tag{5.43}$$

where $\tilde{F}_\alpha$ has the distribution given by the fixed-point equation (5.38).

(ii) *For $\alpha = 1$, we that, as $n \to \infty$,*

$$\tilde{D}^1(\mathcal{U}_n) \xrightarrow{L^2} \tilde{F}_1, \tag{5.44}$$

where $\tilde{F}_1$ has the distribution of $\tilde{D}_1$ as given by the fixed-point equation (5.36). Thus, the variance of $\tilde{F}_1$ is $2 - \pi^2/6 \approx 0.355066$. Further, with $\tilde{D}_1$ the limit in (5.42), we have that $\mathrm{Cov}(\tilde{D}_1, \tilde{F}_1) = (7/4) - \pi^2/6 \approx 0.105066$.

(iii) *For $\alpha > 1$, as $n \to \infty$ we have $D^\alpha(\mathcal{U}_n) \to F_\alpha$, almost surely and in L^2, where the distribution of F_α is given by the fixed-point equation (5.40). Also, $E[F_\alpha] = (\alpha(\alpha - 1))^{-1}$ and $\mathrm{Var}(F_\alpha)$ is given by (5.152).*

Proof. Part (i) of Theorem 5.2.4 follows from Proposition 5.5.4. Part (ii) follows from Propositions 5.5.7 (ii) and 5.5.9. Part (iii) follows from Proposition 5.5.6. □

Of particular interest is the distribution of the variable $\tilde{D}_1$ given by (5.36), which appears in Theorems 5.2.3, 5.2.4, and also in Theorem 6.1.1. In Section 5.5.5, we give a plot (Figure 5.2) of the probability density function of this distribution, estimated by simulation. Also, we can use the fixed-point equation (5.36) to calculate the moments of $\tilde{D}_1$ recursively. Writing

$$f(U) := U \log U + (1 - U) \log(1 - U) + U,$$

and setting $m_k := E[\tilde{D}_1^k]$, we obtain

$$m_k = E[(f(U))^k] + \sum_{i=2}^{k} \binom{k}{i} \sum_{j=0}^{i} \binom{i}{j} E[(f(U))^{k-i} U^j (1-U)^{i-j}] m_j m_{i=j}. \tag{5.45}$$

The fact that $m_1 = 0$ simplifies things a little, and we can rewrite this as

$$m_k = E[(f(U))^k] + \sum_{i=1}^{k} \binom{k}{i} \Bigg[m_i E[(f(U))^{k-i}(U^i + (1-U)^i)] + \sum_{j=2}^{i-2} \binom{i}{j} E[(f(U))^{k-i} U^j (1-U)^{i-j}] m_j m_{i-j} \Bigg].$$

So, for example, when $k = 3$ we obtain $m_3 \approx 0.15411$, which shows $\tilde{D}_1$ is not Gaussian and is consistent with the skewness of the plot in Figure 5.1.

An interesting property of the DLT, which we use in establishing fixed-point equations for limit distributions, is its *self-similarity* (scaling property). In terms of the total weight, this says that for any $t \in (0,1)$, if $Y_1, \ldots, Y_n$ are independent and uniformly distributed on $(0,t]$, then the distribution of $D^\alpha(Y_1, \ldots, Y_n)$ is the same as that of $t^\alpha D^\alpha(U_1, \ldots, U_n)$.

5.3 The NNG in one dimension: analysis

5.3.1 Proof of Theorem 5.2.1

We make use of the theory of Dirichlet spacings as discussed in Section 5.1.1.

Proof of Theorem 5.3.1. Since the nearest-neighbour (directed) graph joins each vertex (which sits at the endpoint of each spacing apart from the points 0 and 1) to its nearest neighbour, we have, for $n \geq 3$

$$\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n) = (S_2^n)^\alpha + (S_n^n)^\alpha + \sum_{i=2}^{n-1} \left(\min\{S_i^n, S_{i+1}^n\}\right)^\alpha. \tag{5.46}$$

Now, from (5.46), using exchangeability we have that

$$E[\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)] = 2E[(S_1^n)^\alpha] + (n-2)E[(\min\{S_1^n, S_2^n\})^\alpha],$$

where, from (5.7) and (5.5) we have

$$E\left[(\min\{S_1^n, S_2^n\})^\alpha\right] = 2^{-\alpha} E[(S_1^n)^\alpha] = 2^{-\alpha} \frac{\Gamma(\alpha+1)\Gamma(n+1)}{\Gamma(n+\alpha+1)}. \tag{5.47}$$

Then (5.22) follows. We now prove (5.23). Squaring both sides of (5.46) and taking expectations, we have

$$\begin{aligned} & E\left[\left(\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)\right)^2\right] \\ = & \sum_{i=2}^{n-1} E\left[\left(\min\{S_i^n, S_{i+1}^n\}\right)^{2\alpha}\right] + 2\sum_{i=3}^{n-1}\sum_{j=2}^{i-1} E\left[\left(\min\{S_i^n, S_{i+1}^n\}\right)^{\alpha}\left(\min\{S_j^n, S_{j+1}^n\}\right)^{\alpha}\right] \\ & + E[(S_2^n)^{2\alpha}] + E[(S_n^n)^{2\alpha}] + 2\sum_{i=2}^{n-1} E[(S_2^n)^{\alpha}(\min\{S_i^n, S_{i+1}^n\})^{\alpha}] \\ & + 2\sum_{i=2}^{n-1} E[(S_n^n)^{\alpha}(\min\{S_i^n, S_{i+1}^n\})^{\alpha}] + 2E[(S_2^n)^{\alpha}(S_n^n)^{\alpha}]. \end{aligned}$$

Then, using exchangeability,

$$\begin{aligned} & E\left[\left(\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)\right)^2\right] \\ = & (n-2)E\left[(\min\{S_1^n, S_2^n\})^{2\alpha}\right] + 2E[(S_1^n S_2^n)^{\alpha}] \\ & +(n-3)(n-4)E\left[(\min\{S_1^n, S_2^n\})^{\alpha}\left(\min\{S_3^n, S_4^n\}\right)^{\alpha}\right] \\ & +2(n-3)E\left[(\min\{S_1^n, S_2^n\})^{\alpha}\left(\min\{S_2^n, S_3^n\}\right)^{\alpha}\right] + 2E[(S_1^n)^{2\alpha}] \\ & +4(n-3)E[(S_1^n)^{\alpha}(\min\{S_2^n, S_3^n\})^{\alpha}] + 4E[(S_1^n)^{\alpha}(\min\{S_1^n, S_2^n\})^{\alpha}]. \end{aligned} \tag{5.48}$$

Now, by (5.6) and (5.8) we have

$$E[(S_1^n)^{\alpha}(\min\{S_2^n, S_3^n\})^{\alpha}] = 2^{-\alpha}\frac{\Gamma(n+1)\Gamma(1+\alpha)^2}{\Gamma(n+1+2\alpha)},$$

and, using (5.6) this time with (5.9) we obtain

$$E[(\min\{S_1^n, S_2^n\})^{\alpha}(\min\{S_3^n, S_4^n\})^{\alpha}] = 2^{-2\alpha}\frac{\Gamma(n+1)\Gamma(1+\alpha)^2}{\Gamma(n+1+2\alpha)}.$$

Also we have that

$$\begin{aligned} E[(S_1^n)^{\alpha}(\min\{S_1^n, S_2^n\})^{\alpha}] &= E[(S_1^n)^{2\alpha}\mathbf{1}_{\{S_1^n<S_2^n\}}] + E[(S_1^n)^{\alpha}(S_2^n)^{\alpha}\mathbf{1}_{\{S_1^n>S_2^n\}}] \\ &= \frac{1}{2}E[(\min\{S_1^n, S_2^n\})^{2\alpha}] + \frac{1}{2}E[(S_1^n)^{\alpha}(S_2^n)^{\alpha}]. \end{aligned}$$

Hence from (5.47) and (5.6) we obtain

$$E[(S_1^n)^{\alpha}(\min\{S_1^n, S_2^n\})^{\alpha}] = \frac{1}{2}\left(2^{-2\alpha}\Gamma(1+2\alpha) + \Gamma(1+\alpha)^2\right)\frac{\Gamma(n+1)}{\Gamma(n+1+2\alpha)}.$$

The final term on the right hand side of (5.48) that we need to evaluate is

$$\begin{aligned} E[(\min\{S_1^n, S_2^n\})^{\alpha}(\min\{S_2^n, S_3^n\})^{\alpha}] &= E[(S_2^n)^{2\alpha}\mathbf{1}_{\{S_2^n<S_1^n,\ S_2^n<S_3^n\}}] \\ &\quad +4E[(S_1^n)^{\alpha}(S_2^n)^{\alpha}\mathbf{1}_{\{S_1^n<S_2^n<S_3^n\}}]. \end{aligned} \tag{5.49}$$

Now, for the first term on the right of (5.49), we have

$$\begin{aligned} E[(S_2^n)^{2\alpha}\mathbf{1}_{\{S_2^n<S_1^n,\ S_2^n<S_3^n\}}] &= \frac{1}{3}E[(\min\{S_1^n,S_2^n,S_3^n\})^{2\alpha}] \\ &= 3^{-1-2\alpha}\frac{\Gamma(1+2\alpha)\Gamma(n+1)}{\Gamma(n+1+2\alpha)}, \end{aligned}$$

the last equality following from (5.10). Now consider the second term on the right of (5.49). By a direct computation, we have

$$\begin{aligned} & E[(S_1^n)^\alpha(S_2^n)^\alpha\mathbf{1}_{\{S_1^n<S_2^n<S_3^n\}}] \\ = {} & n(n-1)(n-2)\int_0^{1/3}\mathrm{d}y\int_y^{(1-y)/2}\mathrm{d}x\int_x^{1-x-y}\mathrm{d}z x^\alpha y^\alpha(1-x-y-z)^{n-3} \\ = {} & n(n-1)\int_0^{1/3}\mathrm{d}y\int_y^{(1-y)/2}x^\alpha y^\alpha(1-y-2x)^{n-2}\mathrm{d}x, \end{aligned}$$

which, via the change of variables $w=y+2x$ and Fubini's theorem is the same as

$$n(n-1)2^{-\alpha-1}\int_0^1\mathrm{d}w(1-w)^{n-2}\int_0^{w/3}y^\alpha(w-y)^\alpha\mathrm{d}y,$$

which yields the expression for $J_{n,\alpha}$ as given by (5.20). Then, by (5.48) and the subsequent calculations, we obtain (5.23).

Finally, (5.24) follows from (5.23) by some routine asymptotic calculations involving Stirling's formula, using the fact that, for any $\beta>0$, as $n\to\infty$,

$$\frac{\Gamma(n+1)}{\Gamma(n+1+\beta)}=n^{-\beta}-\frac{1}{2}\beta(\beta+1)n^{-\beta-1}+O(n^{-\beta-2}). \tag{5.50}$$

This completes the proof of the theorem. □

5.3.2 NNG variances: the Poisson case

Let $\mathcal{P}_n$ denote the homogeneous Poisson point process of intensity n on $(0,1)$. In addition to the variance of $\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)$, as covered in Theorem 5.2.1, we may also be interested in the variance of $\mathcal{N}_1^{1,\alpha}(\mathcal{P}_n)$. Here we provide some partial results in this direction, using the results in [107].

One can verify that $\mathcal{N}_1^{1,\alpha}(\mathcal{P}_n)$ satisfies the conditions of Theorem 2.2 of [107]. From Theorem 2.2 in [107], we have that

$$\lim_{n\to\infty}n^{2\alpha-1}\mathrm{Var}[\mathcal{N}_1^{1,\alpha}(\mathcal{P}_n)]=C_\alpha, \tag{5.51}$$

for some constant C_α given in terms of expectations of nearest-neighbour distances in a homogeneous Poisson point process of unit intensity on $\mathbf{R}$. These integrals are in general difficult (even in one dimension; the higher dimensional analogues seem harder still), but progress can be made in some cases.

Recall that $d_1(x; \mathcal{X})$ denotes the distance from x to its nearest neighbour in $\mathcal{X}$. Let $\mathcal{H}_1$ denote the homogeneous Poisson process of unit intensity on $\mathbf{R}$, and for $y \in \mathbf{R}$ write $\mathcal{H}_1^y$ for $\mathcal{H}_1 \cup \{y\}$. From (2.15) of [107], we have that

$$C_\alpha = E[d_1(0; \mathcal{H}_1)^{2\alpha}] + \int_{\mathbf{R}} \left[E[d_1(0; \mathcal{H}_1^z)^\alpha d_1(z; \mathcal{H}_1^0)^\alpha] - (E[d_1(0; \mathcal{H}_1)^\alpha])^2 \right] \mathrm{d}z. \tag{5.52}$$

We have that, for $\beta > 0$,

$$E[d_1(0; \mathcal{H}_1)^\beta] = 2^{-\beta}\Gamma(1+\beta). \tag{5.53}$$

We now need to calculate $E[d_1(0; \mathcal{H}_1^z)^\alpha d_1(z; \mathcal{H}_1^0)^\alpha]$. Consider points at 0 and z, where $z > 0$. Let R_0, L_0 and R_z, L_z be the right and left nearest-neighbour distances of 0 in $\mathcal{H}_1^z$ and z in $\mathcal{H}_1^0$ respectively. Then L_0 and R_z are independent and exponentially distributed with parameter 1. Conditioning on the position of the first point to the right of 0 in $\mathcal{H}_1$, we have

$$\begin{aligned} & E[d_1(0; \mathcal{H}_1^z)^\alpha d_1(z; \mathcal{H}_1^0)^\alpha] \\ = & \int_0^z e^{-r} E[(\min\{L_0, r\})^\alpha] \times \\ & \left(\int_0^{z-r} e^{-s} E[(\min\{s, R_z\})^\alpha] \mathrm{d}s + e^{r-z} E[(\min\{z-r, R_z\})^\alpha] \right) \mathrm{d}r \\ & + e^{-z} E[(\min\{L_0, z\})^\alpha] E[(\min\{R_z, z\})^\alpha]. \end{aligned} \tag{5.54}$$

Now, for H an exponential random variable with parameter 1, and $h > 0$, for $\alpha > 0$

$$\begin{aligned} E[(\min\{H, h\})^\alpha] &= \int_0^h r^\alpha e^{-r} \mathrm{d}r + h^\alpha e^{-h} \\ &= \Gamma(1+\alpha) - \Gamma(1+\alpha, h) + h^\alpha e^{-h}, \end{aligned} \tag{5.55}$$

where $\Gamma(\cdot, \cdot)$ is the incomplete Gamma function. Therefore, one can in theory compute the right hand side of (5.54). However, the required integrals are rather hard, and only appear reasonably tractable (with the help of Mathematica) for positive integer values of α. In particular,

$$E[\min\{H, h\}] = 1 - e^{-h}.$$

So, using (5.54), we have

$$E[d_1(0; \mathcal{H}_1^z) d_1(z; \mathcal{H}_1^0)] = \frac{1}{4} + \left(\frac{z}{2} - \frac{5}{4} \right) e^{-2z} + e^{-3z}.$$

Hence from (5.52) and (5.53) we obtain

$$C_1 = \frac{1}{2} + 2\int_0^\infty \left[\left(\frac{z}{2} - \frac{5}{4}\right)e^{-2z} + e^{-3z}\right]\mathrm{d}z = \frac{1}{6}.$$

In a similar way, one obtains $C_2 = 28/27$ and $C_3 = 379/36$.

In order to "de-Poissonize" these results to obtain the corresponding limits in the binomial case, where we have $\mathcal{U}_n$ rather than $\mathcal{P}_n$, we apply a further result from [107]. According to Theorem 2.4 of [107], we have that

$$\lim_{n\to\infty} n^{2\alpha-1}\mathrm{Var}[\mathcal{N}_1^{1,\alpha}(\mathcal{U}_n)] = C_\alpha - \delta_\alpha^2,$$

with C_α as given here by (5.52), and in this case δ_α is given by

$$\delta_\alpha = E[d_1(0;\mathcal{H}_1)^\alpha] + \int_{\mathbf{R}} E[d_1(0;\mathcal{H}_1^y)^\alpha - d_1(0;\mathcal{H}_1)^\alpha]\mathrm{d}y. \tag{5.56}$$

We have that

$$\begin{aligned} E[d_1(0;\mathcal{H}_1^y)^\alpha - d_1(0;\mathcal{H}_1)^\alpha] &= \int_{|y|}^\infty 2(|y|^\alpha - r^\alpha)\exp(-2r)\mathrm{d}r \\ &= |y|^\alpha\exp(-2|y|) - 2^{-\alpha}\Gamma(1+\alpha, 2|y|), \end{aligned}$$

so, with (5.53), we have

$$\begin{aligned} \delta_\alpha &= 2^{-\alpha}\Gamma(1+\alpha) + 2\int_0^\infty \left(y^\alpha\exp(-2y) - 2^{-\alpha}\Gamma(1+\alpha,2y)\right)\mathrm{d}y \\ &= 2^{-\alpha}\left(2\Gamma(1+\alpha) - \Gamma(2+\alpha)\right) = 2^{-\alpha}(1-\alpha)\Gamma(1+\alpha). \end{aligned} \tag{5.57}$$

Thus $\delta_1 = 0$, $\delta_2 = -1/2$ and $\delta_3 = -3/2$. Thus we obtain

$$C_\alpha - (\delta_\alpha)^2 = \begin{cases} 1/6 & \alpha = 1 \\ 85/108 & \alpha = 2 \\ 149/18 & \alpha = 3 \end{cases}$$

in agreement with Theorem 5.2.1. Note that we can also, therefore, obtain "Poissonized" results from Theorem 5.2.1. In particular, we have that as $n \to \infty$

$$n^{2\alpha-1}\mathrm{Var}[\mathcal{N}_1^{1,\alpha}(\mathcal{P}_n)] \to V_\alpha + \delta_\alpha^2, \tag{5.58}$$

where V_α is given by (5.25) and δ_α is given by (5.57). Table 5.2 below gives some values of $V_\alpha + \delta_\alpha^2$.

α	1/2	1	2	3	4
$V_\alpha + \delta_\alpha^2$	$\frac{1}{2} + \sqrt{2}\arcsin(1/\sqrt{3}) - \frac{3\pi}{8} \approx 0.192323$	$\frac{1}{6}$	$\frac{28}{27}$	$\frac{379}{36}$	$\frac{38869}{243}$

Table 5.2: Some values of $V_\alpha + \delta_\alpha^2$.

5.4 The ONG in $d = 1$: analysis

5.4.1 Notation and results

In this section we analyse the ONG on a random sequence of points in the interval $(0, 1)$. Our final aim is to prove Theorem 5.2.2. To work in this direction, we study two slightly different models for the ONG. In the first, we choose our first two points to be 0 and 1, and take the rest to be random. In the second, we choose our first point to be 0, and take the rest to be random. By studying these two simpler cases, where the root of the tree is fixed, we will later (in Section 5.4.5) be able to prove Theorem 5.2.2. For this section, our main results are Theorems 5.4.1 and 5.4.2 below, which give results for the two special cases of the one-dimensional ONG that we study in the remainder of this section.

For any finite sequence of points $\mathcal{T}_n = (x_1, x_2, \ldots, x_n) \subset (0, 1)^n$ with distinct inter-point distances, we construct the ONG as follows. We start with the unit interval $(0, 1)$ and insert the points x_i in order, one at a time, starting with $i = 1$. At the insertion of each point, we join the new point to its nearest neighbour among those already present, provided that such a point exists. In other words, for each point x_i, $i \geq 2$, we join x_i by an edge to the point of $\{x_j : 1 \leq j < i\}$ that minimises $|x_i - x_j|$. In this way we construct a tree rooted at x_1, which we denote by ONG($\mathcal{T}_n$). Denote the total weight (under weight function w_α, $\alpha > 0$) of ONG($\mathcal{T}_n$) by $\mathcal{O}^{1,\alpha}(\mathcal{T}_n)$, to be consistent with our previous notation.

Also, given $\mathcal{T}_n$, let $\mathcal{T}_n^0$ be the sequence $(x_0, x_1, \ldots, x_n)$ where the initial term is $x_0 = 0$. Then the ONG on $\mathcal{T}_n^0$ is constructed in a similar manner as before, where now for each $i \geq 1$ we join x_i by an edge to its nearest neighbour in $\{x_j : 0 \leq j < i\}$. In this case we see that x_1 will always be joined to $x_0 = 0$, and our tree is rooted at 0. Let $\mathcal{O}^{1,\alpha}(\mathcal{T}_n^0)$ denote the total weight (under weight function w_α, $\alpha > 0$) of the ONG on $\mathcal{T}_n^0$.

Further, given $\mathcal{T}_n$, let $\mathcal{T}_n^{0,1}$ be the sequence $(x_{-1}, x_0, x_1, \ldots, x_n)$ where we set $x_{-1} = 0$ and $x_0 = 1$. Then the ONG on $\mathcal{T}_n^{0,1}$ is constructed as before. This time we connect x_0 to x_{-1} by an edge (of length 1), and for each $i \geq 1$ we join x_i by an edge to its nearest neighbour in $\{x_j : -1 \leq j < i\}$. Now, x_1 will be joined to x_0 or to x_{-1}, and once more we have a tree rooted at 0. Let $\mathcal{O}^{1,\alpha}(\mathcal{T}_n^{0,1})$ denote the total weight (under weight function w_α, $\alpha > 0$) of the ONG on $\mathcal{T}_n^{0,1}$.

For what follows, our main interest is the case in which $\mathcal{T}_n$ is a random vector in $(0,1)^n$. In this case, set $\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{T}_n) := \mathcal{O}^{1,\alpha}(\mathcal{T}_n) - E[\mathcal{O}^{1,\alpha}(\mathcal{T}_n)]$, the centred total weight of the ONG on $\mathcal{T}_n$. Define $\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{T}_n^0)$ and $\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{T}_n^{0,1})$ similarly. Let $(U_1, U_2, U_3, \ldots)$ be a sequence of independent uniformly distributed random variables in $(0,1)$, and for $n \in \mathbf{N}$ set $\mathcal{U}_n := (U_1, U_2, \ldots, U_n)$, so that $\mathcal{U}_n^0 = (0, U_1, \ldots, U_n)$ and $\mathcal{U}_n^{0,1} = (0, 1, U_1, \ldots, U_n)$. We consider ONG$(\mathcal{U}_n^{0,1})$ and ONG$(\mathcal{U}_n^0)$.

The main results of this section are the two theorems below.

Theorem 5.4.1 *(i) For $1/2 < \alpha < 1$, we have that, as $n \to \infty$,*

$$\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n^{0,1}) \xrightarrow{\mathcal{D}} \tilde{G}_\alpha, \tag{5.59}$$

where $\tilde{G}_\alpha$ has distribution given by the fixed-point equation (5.27).

(ii) For $\alpha = 1$, we have that, as $n \to \infty$,

$$\mathcal{O}^{1,1}(\mathcal{U}_n^{0,1}) - \frac{1}{2}(\log n + \gamma + 1) \xrightarrow{\mathcal{D}} \tilde{G}_1, \tag{5.60}$$

where $\tilde{G}_1$ has the distribution given by the fixed-point equation (5.26). Also, $E[\tilde{G}_1] = 0$, $\mathrm{Var}[\tilde{G}_1] = ((1 + \log 2)/4) - (\pi^2/24) \approx 0.012053$, and $E[\tilde{G}_1^3] \approx -0.00005732546$. Also, (5.60) holds in the sense of convergence of expectations.

(iii) For $\alpha > 1$, we have that, as $n \to \infty$, $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1}) \to 1 + G_\alpha$, almost surely and in L^2, where the distribution of G_α is given by the fixed-point equation (5.28). Also, $E[G_\alpha] = 2^{-\alpha}(\alpha - 1)^{-1}$.

Theorem 5.4.2 *(i) For $1/2 < \alpha < 1$, we have that, as $n \to \infty$,*

$$\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n^0) \xrightarrow{\mathcal{D}} \tilde{H}_\alpha, \tag{5.61}$$

where $\tilde{H}_\alpha$ has distribution given by the fixed-point equation (5.30).

(ii) For $\alpha = 1$, we have that, as $n \to \infty$,

$$\mathcal{O}^{1,1}(\mathcal{U}_n^0) - \frac{1}{2}(\log n + \gamma) \xrightarrow{\mathcal{D}} \tilde{H}_1, \tag{5.62}$$

where $\tilde{H}_1$ has the distribution given by the fixed-point equation (5.29). Also, $E[\tilde{H}_1] = 0$, $\mathrm{Var}[\tilde{H}_1] = ((3 + \log 2)/8) - (\pi^2/24) \approx 0.050410$ and $E[\tilde{H}_1^3] \approx 0.00323456$.

(iii) For $\alpha > 1$, we have that, as $n \to \infty$, $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0) \to H_\alpha$, almost surely and in L^2, where the distribution of H_α is given by the fixed-point equation (5.30). Also, $E[H_\alpha] = (1/\alpha) + 2^{-\alpha}\alpha^{-1}(\alpha-1)^{-1}$.

An interesting property of the ONG, which we use in establishing fixed-point equations for limit distributions, is its *self-similarity* (scaling property). In terms of the total weight, this says that for any $t \in (0,1)$, if $Y_1, \ldots, Y_n$ are independent and uniformly distributed on $(0,t)$, then the distribution of $\mathcal{O}^{1,\alpha}(Y_1,\ldots,Y_n)$ is the same as that of $t^\alpha \mathcal{O}^{1,\alpha}(U_1,\ldots,U_n)$.

In Section 5.4.2, we study the total weight of the ONG on $\mathcal{U}_n^{0,1}$, and give a proof of Theorem 5.4.1. We deal with the ONG on $\mathcal{U}_n^0$ and the proof of Theorem 5.4.2 in Section 5.4.4.

5.4.2 The total weight of ONG($\mathcal{U}_n^{0,1}$)

For $n = 1, 2, 3, \ldots$ denote by Z_n the random variable given by the gain in length of the tree on the addition of one point (U_n) to an existing $n-1$ points in the ONG on a sequence of uniform random variables $\mathcal{U}_{n-1}^{0,1}$, i.e. with the convention $\mathcal{O}^{1,1}(\mathcal{U}_0^{0,1}) = 1$ we set

$$Z_n := \mathcal{O}^{1,1}(\mathcal{U}_n^{0,1}) - \mathcal{O}^{1,1}(\mathcal{U}_{n-1}^{0,1}). \tag{5.63}$$

Thus, with weight exponent α, the nth edge to be added has weight Z_n^α.

Lemma 5.4.1 *(i) Z_n has distribution function F_n given by $F_n(t) = 0$ for $t < 0$, $F_n(t) = 1$ for $t \geq 1/2$, and $F_n(t) = 1 - (1-2t)^n$ for $0 \leq t \leq 1/2$.*

(ii) For $\beta > 0$,

$$E[Z_n^\beta] = 2^{-\beta}\frac{\Gamma(n+1)\Gamma(\beta+1)}{\Gamma(n+\beta+1)}. \tag{5.64}$$

In particular, for $k \in \mathbf{N}$,

$$E[Z_n^k] = \frac{k!n!}{2^k(n+k)!}; \quad E[Z_n] = \frac{1}{2(n+1)}; \quad \mathrm{Var}[Z_n] = \frac{n}{4(n+1)^2(n+2)}. \tag{5.65}$$

(iii) For $\beta > 0$, as $n \to \infty$

$$E[Z_n^\beta] = 2^{-\beta}\Gamma(\beta+1)n^{-\beta} + O(n^{-\beta-1}). \tag{5.66}$$

(iv) As $n \to \infty$,

$$2nZ_n \xrightarrow{\mathcal{D}} \mathrm{Exp}(1),$$

where Exp(1) *is an exponential random variable with parameter 1.*

Proof. For (i), we have

$$P[Z_n > t] = P\left[\text{none of } 1, 0, U_1, \ldots, U_{n-1} \text{ within } t \text{ of } U_n\right].$$

This probability is zero for $t > 1/2$, and 1 for $t < 0$. Suppose $0 \leq t \leq 1/2$. Then we have

$$\begin{aligned} P[Z_n > t] &= P[\{U_1, \ldots, U_{n-1} \notin (U_n - t, U_n + t)\} \cap \{t < U_n < 1 - t\}] \\ &= (1 - 2t)^n. \end{aligned}$$

Thus we have proved (i). For (ii), we now have

$$\begin{aligned} E[Z_n^\beta] &= \int_0^{2^{-\beta}} P[Z_n > t^{1/\beta}]\mathrm{d}t = \int_0^{2^{-\beta}} (1 - 2t^{1/\beta})^n \mathrm{d}t \\ &= 2^{-\beta} \frac{\Gamma(n+1)\Gamma(\beta+1)}{\Gamma(n+\beta+1)}. \end{aligned} \tag{5.67}$$

Part (iii) then follows from (5.67) by Stirling's formula. For (iv), we have that, for $t \in [0, \infty)$, and n large enough so that $t/(2n) \leq 1/2$,

$$P[2nZ_n > t] = P[Z_n > t/(2n)] = (1 - (t/n))^n \to e^{-t},$$

as $n \to \infty$, but $1 - e^{-t}$, $t \geq 0$ is the distribution function of an exponential random variable with parameter 1. □

Recall that $\gamma \approx 0.57721566$ is Euler's constant, defined at (C.3.5).

Proposition 5.4.1 *As $n \to \infty$, the expected total weight of* $\mathrm{ONG}(\mathcal{U}_n^{0,1})$ *under weight function w_α, $\alpha > 0$, satisfies*

$$E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1})] = \frac{\Gamma(\alpha+1)}{1-\alpha} 2^{-\alpha} n^{1-\alpha} + 1 - \frac{2^{-\alpha}}{1-\alpha} + O(n^{-\alpha}); \quad (0 < \alpha < 1) \tag{5.68}$$

$$E[\mathcal{O}^{1,1}(\mathcal{U}_n^{0,1})] - \frac{1}{2}\log n = \frac{1}{2}(\gamma + 1) + O(n^{-1}); \tag{5.69}$$

$$E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1})] = 1 + \frac{2^{-\alpha}}{\alpha - 1} + O(n^{1-\alpha}) \ (\alpha > 1) \tag{5.70}$$

Proof. Counting the first edge from 1 to 0, we have

$$E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1})] = 1 + \sum_{i=1}^n \left(E[\mathcal{O}^{1,\alpha}(\mathcal{U}_i^{0,1})] - E[\mathcal{O}^{1,\alpha}(\mathcal{U}_{i-1}^{0,1})]\right) = 1 + \sum_{i=1}^n E[Z_i^\alpha].$$

In the case where $\alpha = 1$, $E[Z_i] = (2(i+1))^{-1}$ by (5.65), and (5.69) follows by (C.3.5). For general $\alpha > 0$, $\alpha \neq 1$, from (5.64) we have that

$$\begin{aligned} E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1})] &= 1 + 2^{-\alpha}\Gamma(1+\alpha) \sum_{i=1}^n \frac{\Gamma(i+1)}{\Gamma(1+\alpha+i)} \\ &= 1 + \frac{2^{-\alpha}}{\alpha - 1} - \frac{2^{-\alpha}\Gamma(1+\alpha)\Gamma(n+2)}{(\alpha-1)\Gamma(n+1+\alpha)}. \end{aligned} \tag{5.71}$$

By Stirling's formula, the last term satisfies

$$-\frac{2^{-\alpha}\Gamma(1+\alpha)\Gamma(n+2)}{(\alpha-1)\Gamma(n+1+\alpha)} = -2^{-\alpha}\frac{\Gamma(1+\alpha)}{\alpha-1}n^{1-\alpha}(1+O(n^{-1})), \tag{5.72}$$

which tends to zero as $n \to \infty$ for $\alpha > 1$, to give us (5.70). For $\alpha < 1$, we have (5.68) from (5.71) and (5.72). □

5.4.3 Proof of Theorem 5.4.1

We now address the proof of Theorem 5.4.1, by first proving the following lemma and propositions. We make use of the theory of Dirichlet spacings (see Section 5.1.1).

We will make use of the following discussion for the proof of Lemma 5.4.2 below, and also for Lemma 5.4.6 later, and so we consider general $\alpha > 0$. For $n = 1, 2, 3, \ldots$ let Z_n and H_n denote the random variable given by the gain in length, on the addition of the point U_n, of the ONG on $\mathcal{U}_n^{0,1}$ and $\mathcal{U}_n^0$ respectively. Then, for $\alpha > 0$,

$$\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1}) - \mathcal{O}^{1,\alpha}(\mathcal{U}_n^0) = 1 + \sum_{i=1}^{n} (Z_i^\alpha - H_i^\alpha). \tag{5.73}$$

Consider the arrival of the point U_i. For any i, Z_i and H_i are the same unless the point U_i falls in the right hand half of the rightmost spacing S_i^{i-1}. Denote this latter event by E_i. Given S_i^{i-1}, the probability of E_i is $S_i^{i-1}/2$. Given S_i^{i-1}, and given that E_i occurs, the value of Z_i is given by $(1-V_i)S_i^{i-1}/2$ and the value of H_i by $(1+V_i)S_i^{i-1}/2$, where $V_i = 1 + 2(U_i - 1)/S_i^{i-1}$ is uniform on $(0,1)$ given E_i. So we have that

$$H_i^\alpha - Z_i^\alpha = \mathbf{1}_{E_i}\left(\frac{S_i^{i-1}}{2}\right)^\alpha ((1+U_i)^\alpha - (1-U_i)^\alpha), \tag{5.74}$$

where E_i is an event with probability $S_i^{i-1}/2$.

Remark. In the context of Dirichlet spacings, we see that the variable Z_n as defined at (5.63) satisfies, for $n \geq 1$, $Z_n \stackrel{\mathcal{D}}{=} \min\{S_1^n, S_2^n\}$. Thus, from (5.7), $Z_n \stackrel{\mathcal{D}}{=} S_1^n/2$. This fact also follows from the equivalence of the distribution functions (see Lemma 5.4.1).

Lemma 5.4.2 *For a sequence of uniform random variables $\mathcal{U}_n$, we have that*

$$E[\mathcal{O}^{1,1}(\mathcal{U}_n^{0,1}) - \mathcal{O}^{1,1}(\mathcal{U}_n^0)] = \frac{1}{2} + \frac{1}{2(n+1)}. \tag{5.75}$$

Proof. From the $\alpha = 1$ case of (5.74), we have that

$$E[H_i - Z_i] = \frac{1}{4}E[(S_i^{i-1})^2].$$

By (5.5), we have that $E[(S_i^{i-1})^2] = 2i^{-1}(i+1)^{-1}$. Thus, from (5.73),

$$E[\mathcal{O}^{1,1}(\mathcal{U}_n^{0,1}) - \mathcal{O}^{1,1}(\mathcal{U}_n^0)] = 1 + E\sum_{i=1}^n (Z_i - H_i) = 1 - \frac{1}{2}\sum_{i=1}^n \left(\frac{1}{i} - \frac{1}{1+i}\right),$$

which gives (5.75). □

From (5.74) and (5.75), we see that $\sum_{i=1}^n (Z_i - H_i)$ converges almost surely to $\sum_{i=1}^\infty (Z_i - H_i)$. Proposition 5.4.2 below studies this in some more detail. First, we need the following result. We use the notation $\log^+ x := \max\{\log x, 0\}$ for $x \geq 0$.

Lemma 5.4.3 *Let U be uniform on $(0,1)$ and, given U, let $N(n) \sim \mathrm{Bin}(n-1, U)$. Then, as $n \to \infty$,*

$$U(\log^+ N(n) - \log n) \xrightarrow{L^2} U \log U; \tag{5.76}$$

$$(1-U)(\log^+(n-1-N(n)) - \log n) \xrightarrow{L^2} (1-U)\log(1-U). \tag{5.77}$$

Proof. For $n \in \mathbf{N}$, let $M_n := \log^+ N(n) - \log n - \log U$. First, suppose $N(n) \geq nU/2$. We have that

$$-\log 2 \leq M_n \mathbf{1}_{\{N(n) \geq nU/2\}} \mathbf{1}_{\{nU \geq 2\}} \leq -\log U.$$

Hence

$$U^2 M_n^2 \mathbf{1}_{\{N(n) \geq nU/2\}} \mathbf{1}_{\{nU \geq 2\}} \leq U^2 \max\{(\log 2)^2, (\log U)^2\}. \tag{5.78}$$

The expected value of the right hand side of (5.78) is finite. Also, $U^2 M_n^2 \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$, by continuity and the strong law of large numbers for $N(n)$. Hence, by the dominated convergence theorem,

$$E[U^2 M_n^2 \mathbf{1}_{\{N(n) \geq nU/2\}} \mathbf{1}_{\{nU \geq 2\}}] \to 0. \tag{5.79}$$

Also, we have $0 \leq \log^+ N(n) \leq \log n$, so that

$$-\log n \leq M_n \leq -\log U.$$

Hence

$$U^4 M_n^4 \leq (\log n)^4 + (\log U)^4, \tag{5.80}$$

so that $E[U^4 M_n^4] = O((\log n)^4)$. Since $P[nU < 2] = 2n^{-1}$, we then obtain, by Cauchy-Schwarz, that there exists a finite positive constant C such that

$$E[U^2 M_n^2 \mathbf{1}_{\{N(n) \geq nU/2\}} \mathbf{1}_{\{nU < 2\}}] \leq C(\log n)^2 n^{-1/2} \to 0, \tag{5.81}$$

as $n \to \infty$. Now, suppose $0 \leq N(n) < nU/2$. In this case, from (5.80), and Cauchy-Schwarz again, for some finite positive constant C

$$E[U^2 M_n^2 \mathbf{1}_{\{N(n)<nU/2\}}] \leq C(\log n)^2 (P[N(n) < nU/2])^{1/2} \to 0, \tag{5.82}$$

as $n \to \infty$, since

$$P[N(n) < nU/2] \leq P[U < n^{-1/2}] + P[U > n^{-1/2},\ N(n) < nU/2],$$

which is $o((\log n)^4)$ as $n \to \infty$, using standard bounds for the tail of a binomial distribution (see, e.g., Lemma 1.1 in [104]) for the final probability. The results (5.79), (5.81), and (5.82) then give (5.76). The argument for (5.77) is similar. □

Proposition 5.4.2 *As $n \to \infty$,*

$$\begin{pmatrix} \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0,1}) \\ \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0,1}) - \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0}) \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \tilde{G}_1 \\ \tilde{Q} \end{pmatrix}, \tag{5.83}$$

where $(\tilde{G}_1, \tilde{Q})$ satisfies the fixed-point equation

$$\begin{pmatrix} \tilde{G}_1 \\ \tilde{Q} \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} U & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{G}_1^{\{1\}} \\ \tilde{Q}^{\{1\}} \end{pmatrix} + \begin{pmatrix} 1-U & 0 \\ 0 & 1-U \end{pmatrix} \begin{pmatrix} \tilde{G}_1^{\{2\}} \\ \tilde{Q}^{\{2\}} \end{pmatrix}$$
$$+ \begin{pmatrix} \frac{1}{2}U\log U + \frac{1}{2}(1-U)\log(1-U) + \min\{U, 1-U\} \\ (1-2U)\mathbf{1}_{\{U>1/2\}} + \frac{1}{2}U \end{pmatrix}. \tag{5.84}$$

In particular, $\tilde{G}_1$ satisfies the fixed-point equation (5.26), and $\tilde{Q}$ satisfies

$$\tilde{Q} \stackrel{\mathcal{D}}{=} (1-U)\tilde{Q} + (1-2U)\mathbf{1}_{\{U>1/2\}} + \frac{1}{2}U. \tag{5.85}$$

Further, $E[\tilde{G}_1] = E[\tilde{Q}] = 0$,

$$\mathrm{Var}[\tilde{G}_1] = \frac{1}{4}(1 + \log 2) - \frac{\pi^2}{24} \approx 0.012053, \tag{5.86}$$

and $E[\tilde{G}_1^3] \approx -0.00005732546$.

Proof. We make use of Theorem 4.1 of [102], which is a general result for 'divide-and-conquer' type recurrences. For ease of notation, write $Y_n := \mathcal{O}^{1,1}(\mathcal{U}_n^{0,1}) - 1$, where we subtract 1 so that Y_n does not include the length of the edge from 1 to 0, and let

$$Q_n := \mathcal{O}^{1,1}(\mathcal{U}_n^{0,1}) - 1 - \mathcal{O}^{1,1}(\mathcal{U}_n^{0}),$$

again discounting the edge from 1 to 0. Write $U = U_1$ for the position of the first arrival. Given U, let $N(n) \sim \text{Bin}(n-1, U)$ be the number of points of $U_2, U_3, \ldots, U_n$ that arrive to the left of $U_1 = U$. Using the self-similarity and scaling properties of the ONG, we have that (Y_n, Q_n) satisfies

$$\begin{pmatrix} Y_n \\ Q_n \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} U & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Y^{\{1\}}_{N(n)} \\ Q^{\{1\}}_{N(n)} \end{pmatrix} + \begin{pmatrix} 1-U & 0 \\ 0 & 1-U \end{pmatrix} \begin{pmatrix} Y^{\{2\}}_{n-1-N(n)} \\ Q^{\{2\}}_{n-1-N(n)} \end{pmatrix} + \begin{pmatrix} \min\{U, 1-U\} \\ (1-2U)\mathbf{1}_{\{U>1/2\}} \end{pmatrix}. \tag{5.87}$$

where, given U and $N(n)$, $Y^{\{1\}}_{N(n)}$, $Y^{\{2\}}_{n-1-N(n)}$ are independent copies of $Y_{N(n)}$, $Y_{n-1-N(n)}$ respectively, and similarly for the Qs.

This equation is of the form of (21) in [102]. We now renormalise (5.87) by taking $(\tilde{Y}_n, \tilde{Q}_n) := (Y_n - E[Y_n], Q_n - E[Q_n])$ (in the notation of [102], we take $C_n \equiv 1$). By (5.69) we have

$$E[Y_n] = E[\mathcal{O}^{1,1}(\mathcal{U}^{0,1}_n)] - 1 = \frac{1}{2}\log n + (\gamma - 1)/2 + h(n),$$

where $h(n) = o(1)$, while by (5.75) $E[Q_n] = -(1/2) + k(n)$, where $k(n) = O(n^{-1})$. Then by (5.87)

$$\begin{pmatrix} \tilde{Y}_n \\ \tilde{Q}_n \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} U & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{Y}^{\{1\}}_{N(n)} \\ \tilde{Q}^{\{1\}}_{N(n)} \end{pmatrix} + \begin{pmatrix} 1-U & 0 \\ 0 & 1-U \end{pmatrix} \begin{pmatrix} \tilde{Y}^{\{2\}}_{n-1-N(n)} \\ \tilde{Q}^{\{2\}}_{n-1-N(n)} \end{pmatrix} + \begin{pmatrix} A_n \\ B_n \end{pmatrix} \tag{5.88}$$

where

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} = \begin{pmatrix} \min\{U, 1-U\} + \frac{1}{2}\left(U(\log^+ N(n) - \log n) + (1-U)(\log^+(n-1-N(n)) - \log n)\right) \\ (1-2U)\mathbf{1}_{\{U>1/2\}} + \frac{1}{2}U \end{pmatrix} + \begin{pmatrix} Uh(N(n)) + (1-U)h(n-1-N(n)) - h(n) \\ (1-U)k(n-1-N(n)) - k(n) \end{pmatrix}.$$

In order to apply Theorem 4.1 of [102], we need to verify the conditions (24), (25) and (26) there. Writing $\|\cdot\|_{\text{op}}$ for the operator norm, we have that, for condition (24) in [102], it follows from Lemma 5.4.3 that, as $n \to \infty$,

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} \xrightarrow{L^2} \begin{pmatrix} \frac{1}{2}U\log U + \frac{1}{2}(1-U)\log(1-U) + \min\{U, 1-U\} \\ (1-2U)\mathbf{1}_{\{U>1/2\}} + \frac{1}{2}U \end{pmatrix}. \tag{5.89}$$

Also, for condition (25) in [102],

$$E\left[\left\|\begin{pmatrix} U & 0 \\ 0 & 0 \end{pmatrix}\right\|_{\mathrm{op}}^2 + \left\|\begin{pmatrix} 1-U & 0 \\ 0 & 1-U \end{pmatrix}\right\|_{\mathrm{op}}^2\right] = 2/3 < 1. \tag{5.90}$$

Finally, for condition (26) in [102], we have that for any $\ell \in \mathbf{N}$, as $n \to \infty$,

$$E\left[\mathbf{1}_{\{N(n)\leq \ell\}\cup\{N(n)=n\}}U^2\right] \to 0; \ E\left[\mathbf{1}_{\{n-1-N(n)\leq \ell\}\cup\{n-1-N(n)=n\}}(1-U)^2\right] \to 0. \tag{5.91}$$

Taking $s = 2$ in Theorem 4.1 of [102], applied to the equation (5.88), with the conditions (5.90), (5.89) and (5.91), implies that $(\tilde{Y}_n, \tilde{Q}_n)$ converges in the Zolotarev ζ_2 metric (which implies convergence in distribution; see e.g. Chapter 14 of [117]) to $(\tilde{Y}, \tilde{Q})$, where $E[\tilde{Y}] = E[\tilde{Q}] = 0$ and the distribution of $(\tilde{Y}, \tilde{Q})$ is characterized by the fixed-point equation

$$\begin{pmatrix} \tilde{Y} \\ \tilde{Q} \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} U & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \tilde{Y}^{\{1\}} \\ \tilde{Q}^{\{1\}} \end{pmatrix} + \begin{pmatrix} 1-U & 0 \\ 0 & 1-U \end{pmatrix}\begin{pmatrix} \tilde{Y}^{\{2\}} \\ \tilde{Q}^{\{2\}} \end{pmatrix}$$
$$+ \begin{pmatrix} \frac{1}{2}U\log U + \frac{1}{2}(1-U)\log(1-U) + \min\{U, 1-U\} \\ (1-2U)\mathbf{1}_{\{U>1/2\}} + \frac{1}{2}U \end{pmatrix}. \tag{5.92}$$

That is, $\tilde{Y}$ satisfies (5.26), so that $\tilde{Y}$ has the distribution of $\tilde{G}_1$, and $\tilde{Q}$ satisfies (5.85). Then setting $\tilde{Y} = \tilde{G}_1$ in (5.92) gives (5.84). Since $\tilde{Y}_n = \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0,1})$ and $\tilde{Q}_n = \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0,1}) - \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0})$, we have (5.83).

It remains to prove the results for the higher moments of $\tilde{G}_1$. For the variance of $\tilde{G}_1$, squaring both sides of (5.26), taking expectations, and using independence and the fact that $E[\tilde{G}_1] = 0$, we obtain

$$\begin{aligned} E[\tilde{G}_1^2] &= \frac{2}{3}E[\tilde{G}_1^2] + E[\min\{U, 1-U\}^2] + \frac{1}{2}E[U^2(\log U)^2] \\ &\quad + \frac{1}{2}E[U(1-U)\log U \log(1-U)] + 2E[U\log U \min\{U, 1-U\}]. \end{aligned}$$

The integrals required for the expectations are standard, and we find that $E[\tilde{G}_1^2] = ((1+\log 2)/4) - (\pi^2/24)$, which yields (5.86). Similarly, we obtain the third moment $E[\tilde{G}_1^3] = -0.00005732546\ldots$ from (5.26), although in this case numerical methods are required for some of the integrals. □

Let U be uniform on $(0, 1)$, and given U, let $N(n) \sim \mathrm{Bin}(n-1, U)$. Set

$$B_\alpha(n) := (n-1)^{1/2}\left(U^\alpha\left(\frac{N(n)}{n-1}\right)^{1-\alpha} + (1-U)^\alpha\left(\frac{n-1-N(n)}{n-1}\right)^{1-\alpha} - 1\right). \tag{5.93}$$

Lemma 5.4.4 *Suppose* $0 \leq \alpha \leq 1$. *Then, as* $n \to \infty$,

$$B_\alpha(n) \xrightarrow{L^3} 0. \tag{5.94}$$

Proof. The result is trivial when $\alpha = 1$ or $\alpha = 0$. Suppose $0 < \alpha < 1$. Suppose $n > 1$. To ease notation, for the duration of this proof, set $m = n - 1$. Then we have that for any $U \in (0,1)$ and $0 \leq N(n) \leq m$,

$$-1 \leq U^\alpha \left(\frac{N(n)}{m}\right)^{1-\alpha} + (1-U)^\alpha \left(\frac{m - N(n)}{m}\right)^{1-\alpha} - 1 \leq 0, \tag{5.95}$$

so that in particular $|B_\alpha(n)| \leq n^{1/2}$ almost surely for $0 \leq \alpha \leq 1$. Let

$$W_n := \frac{N(n) - mU}{\sqrt{mU(1-U)}},$$

so that $E[W_n] = 0$, $E[W_n^2] = 1$, and

$$\frac{N(n)}{mU} = 1 + W_n \sqrt{\frac{1-U}{mU}}; \quad \frac{m - N(n)}{m(1-U)} = 1 - W_n \sqrt{\frac{U}{m(1-U)}}.$$

Then, by Taylor's theorem,

$$\begin{aligned} U^\alpha \left(\frac{N(n)}{m}\right)^{1-\alpha} &= U \left(1 + (1-\alpha) W_n \sqrt{\frac{1-U}{mU}} - R_1(n) W_n^2 \frac{1-U}{mU}\right) &(5.96) \\ &= U \left(1 + R_2(n) W_n \sqrt{\frac{1-U}{mU}}\right), &(5.97) \end{aligned}$$

for remainder terms $R_1(n)$, $R_2(n)$ (which depend on W_n and U). Similarly, we have

$$\begin{aligned} &(1-U)^\alpha \left(\frac{m - N(n)}{m}\right)^{1-\alpha} \\ &= (1-U) \left(1 - (1-\alpha) W_n \sqrt{\frac{U}{m(1-U)}} - R_3(n) W_n^2 \frac{U}{m(1-U)}\right) &(5.98) \\ &= (1-U) \left(1 - R_4(n) W_n \sqrt{\frac{U}{m(1-U)}}\right). &(5.99) \end{aligned}$$

By the Lagrange form of the remainder in Taylor's theorem and a continuity argument at $x = 0$ there exists a constant $B \in (0, \infty)$ such that for$\beta = 1 - \alpha$,

$$0 \geq \frac{(1+x)^\beta - 1 - \beta x}{x^2} \geq -B, \text{ and } 0 \leq \frac{(1+x)^\beta - 1}{x} \leq B,$$

for all $x \geq -1$. Thus we we have, for $i \in \{1, 2, 3, 4\}$,

$$0 \leq R_i(n) < C, \tag{5.100}$$

for a finite positive constant C.

For $n > 1$, $m = n - 1$, let E_n denote the event $m^{-3/4} < U < 1 - m^{-3/4}$. From (5.96) and (5.98) we obtain

$$\begin{aligned} |B_\alpha(n)\mathbf{1}_{E_n}| &= \left|-R_1(n)W_n^2(1-U)m^{-1/2} - R_3(n)W_n^2 U m^{-1/2}\right) \mathbf{1}_{D_n^3}\mathbf{1}_{E_n} \\ &\leq Cm^{-1/2}W_n^2\mathbf{1}_{E_n}, \end{aligned}$$

for some $0 < C < \infty$. By a standard moment generating function calculation,

$$\begin{aligned} E[(N(n) - mU)^6|U] &= mU(1-U)\left[15m^2U^2(1-U)^2 - 130mU^2(1-U)^2\right. \\ &\quad \left.+25mU(1-U) - 30U(1-U)(1-2U)^2 + 1\right] \\ &\leq mU(1-U)(15m^2U^2(1-U)^2 + 25mU(1-U) + 1)(5.101) \end{aligned}$$

By (5.101) we have that

$$E[W_n^6\mathbf{1}_{E_n}] \leq E[(N(n) - mU)^6 m^{-3}U^{-3}(1-U)^{-3}|E_n] = O(1),$$

as $n \to \infty$, so from (5.102) we have that

$$B_\alpha(n)\mathbf{1}_{E_n} \xrightarrow{L^3} 0. \qquad (5.102)$$

Also, from (5.97) and (5.99) we have,

$$\left|B_\alpha(n)\mathbf{1}_{E_n^c}\right| = \left|(R_2(n) - R_4(n))W_nU^{1/2}(1-U)^{1/2}\right|\mathbf{1}_{E_n^c},$$

and so using (5.100) we have

$$\left|B_\alpha(n)\mathbf{1}_{E_n^c}\right| \leq C|W_n||U^{1/2}(1-U)^{1/2}\mathbf{1}_{E_n^c}. \qquad (5.103)$$

Now, from (5.101) we have that

$$E[(W_nU^{1/2}(1-U)^{1/2})^6] = m^{-3}E[(N(n) - mU)^6] = O(1),$$

as $n \to \infty$, so by Cauchy-Schwarz and the fact that $P[E_n^c] = O(n^{-3/4})$ we obtain from (5.103) that as $n \to \infty$

$$E\left[\left|B_\alpha(n)\mathbf{1}_{E_n^c}\right|^3\right] \to 0. \qquad (5.104)$$

So (5.102) and (5.104) complete the proof. □

For the next few results, we will make use of the following. Suppose $\alpha > 0$. For ease of notation, denote $Y_n = \mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1}) - 1$, where by subtracting 1 we discount the length of the edge from 1 to 0. Then, writing $U = U_1$ for the position of the first arrival of $\mathcal{U}_n$, and using the self-similarity of the ONG, we have

$$Y_n \stackrel{\mathcal{D}}{=} (\min\{U, 1-U\})^\alpha + U^\alpha Y^{\{1\}}_{N(n)} + (1-U)^\alpha Y^{\{2\}}_{n-1-N(n)}, \qquad (5.105)$$

where, given U, $N(n) \sim \mathrm{Bin}(n-1, U)$, and, given U and $N(n)$, $Y^{\{1\}}_{N(n)}$ and $Y^{\{2\}}_{n-1-N(n)}$ are independent with the distribution of $Y_{N(n)}$ and $Y_{n-1-N(n)}$, respectively.

Proposition 5.4.3 *Suppose* $1/2 < \alpha < 1$. *Then, as* $n \to \infty$,

$$\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n^{0,1}) \stackrel{\mathcal{D}}{\longrightarrow} \tilde{G}_\alpha,$$

where $\tilde{G}_\alpha$ *satisfies the fixed-point equation (5.27), and* $E[\tilde{G}_\alpha] = 0$.

Proof. Suppose $1/2 < \alpha < 1$. For $n > 0$, let

$$\tilde{Y}_n := Y_n - E[Y_n],$$

and

$$\begin{aligned} \tilde{Y}_{N(n)} &:= Y_{N(n)} - E\left[Y_{N(n)} | N(n)\right], \\ \tilde{Y}_{n-1-N(n)} &:= Y_{n-1-N(n)} - E\left[Y_{n-1-N(n)} | N(n)\right]. \end{aligned}$$

Then, using (5.68), we can rewrite the $1/2 < \alpha < 1$ case of (5.105) as

$$\begin{aligned} \tilde{Y}_n \stackrel{\mathcal{D}}{=} & U^\alpha \tilde{Y}^{\{1\}}_{N(n)} + (1-U)^\alpha \tilde{Y}^{\{2\}}_{n-1-N(n)} + (\min\{U, 1-U\})^\alpha + C n^{(1/2)-\alpha} B_\alpha(n) \\ & + (U^\alpha + (1-U)^\alpha - 1)\frac{2^{-\alpha}}{1-\alpha} + U^\alpha h(N(n)) + (1-U)^\alpha h(n-1-N(n)), \end{aligned} \qquad (5.106)$$

where $h(n) \to 0$ as $n \to \infty$ and $B_\alpha(n)$ is as defined by (5.93), and C is a constant. From Lemma 5.4.4, for $1/2 < \alpha < 1$, $n^{(1/2)-\alpha} B_\alpha(n)$ tends to 0 in L^2. Also

$$E\left[U^{2\alpha} + (1-U)^{2\alpha}\right] = \frac{2}{2\alpha + 1} < 1,$$

for $\alpha > 1/2$. So we can apply Theorem 4.1 of [102] to (5.106), with $s = 2$, to obtain that $\tilde{Y}_n \to \tilde{G}_\alpha$ in the Zolotarev ζ_2 metric, and hence in distribution, where $\tilde{G}_\alpha$ satisfies (5.27).
□

Proposition 5.4.4 *Let* $\alpha > 1$. *Then there exists a random variable* G_α *such that as* $n \to \infty$ *we have* $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1}) \to 1 + G_\alpha$ *almost surely and in* L^2. *Also, the random variable* G_α *satisfies the distributional fixed-point equality (5.28). Further,* $E[G_\alpha] = 2^{-\alpha}/(\alpha - 1)$.

Proof. Let Z_i be the length of the ith edge of the ONG on $\mathcal{U}_n^{0,1}$, as defined at (5.63). Let $G_\alpha := \sum_{i=1}^{\infty} Z_i^\alpha$. The sum converges almost surely since it has non-negative terms and, by (5.66), has finite expectation for $\alpha > 1$. Then

$$E[{G_\alpha}^2] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E[Z_i^\alpha Z_j^\alpha]. \tag{5.107}$$

By (5.107), (5.66) and Cauchy-Schwarz, there exists a constant $0 < C < \infty$ such that

$$E[{G_\alpha}^2] \leq C \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} i^{-\alpha} j^{-\alpha} < \infty,$$

since $\alpha > 1$. The L^2 convergence then follows from the dominated convergence theorem.

Once again, we have (5.105), this time for $\alpha > 1$. As $n \to \infty$, $N(n)$ and $n - N(n)$ both tend to infinity almost surely, and so, by taking $n \to \infty$ in (5.105), we obtain the fixed-point equation (5.28).

The identity $E[G_\alpha] = 2^{-\alpha}(\alpha - 1)^{-1}$ is obtained either from (5.70), or by taking expectations in (5.28). Next, if we set $\tilde{G}_\alpha = G_\alpha - E[G_\alpha]$, (5.28) yields (5.27). □

Proof of Theorem 5.4.1. Part (i) of the theorem follows from Proposition 5.4.3. Part (ii) follows from Proposition 5.4.2 with (5.69), and part (iii) follows from Proposition 5.4.4. □

5.4.4 The total weight of ONG($\mathcal{U}_n^0$)

As well as considering the ONG on $\mathcal{U}_n^{0,1}$, we consider the ONG on $\mathcal{U}_n^0$, since our final results on $\mathcal{O}^{1,\alpha}(\mathcal{U}_n)$ (in Theorem 5.2.2) are described in terms of distributional limits of $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0)$.

For $n = 1, 2, 3, \ldots$ denote by H_n the random variable given by the gain in length of the tree on the addition of one point (U_n) to an existing $n - 1$ points in the ONG on a sequence of uniform random variables $\mathcal{U}_{n-1}^0$, i.e. with the convention $\mathcal{O}^{1,1}(\mathcal{U}_0^0) = 0$ we set

$$H_n := \mathcal{O}^{1,1}(\mathcal{U}_n^0) - \mathcal{O}^{1,1}(\mathcal{U}_{n-1}^0). \tag{5.108}$$

Thus, with weight exponent α, the nth edge to be added has weight H_n^α.

Lemma 5.4.5 *(i) H_n has distribution function G_n given by $G_n(t) = 0$ for $t < 0$, $G_n(t) = 1$ for $t \geq 1$, and*

$$G_n(t) = 1 - \frac{1}{n}(1-t)^n - \frac{n-1}{n}(1-2t)^n \quad (0 \leq t \leq 1/2); \tag{5.109}$$

$$G_n(t) = 1 - \frac{1}{n}(1-t)^n \quad (1/2 \leq t \leq 1). \tag{5.110}$$

(ii) For $\beta > 0$,

$$E[H_n^\beta] = (1 + 2^{-\beta}(n-1))\frac{\Gamma(n)\Gamma(\beta+1)}{\Gamma(n+\beta+1)}. \quad (5.111)$$

In particular,

$$E[H_n] = \frac{1}{2n}; \quad \mathrm{Var}[H_n] = \frac{n+3}{4n(n+1)(n+2)}. \quad (5.112)$$

(iii) For $\beta > 0$, as $n \to \infty$

$$E[H_n^\beta] = 2^{-\beta}\Gamma(\beta+1)n^{-\beta} + O(n^{-\beta-1}). \quad (5.113)$$

(iv) As $n \to \infty$,

$$2nH_n \xrightarrow{\mathcal{D}} \mathrm{Exp}(1),$$

where Exp(1) *is an exponential random variable with parameter 1.*

Proof. First we prove (i). With Z_n as defined at (5.63), we have $H_n \geq Z_n$ with equality except in some cases where U_n is the rightmost point of $\mathcal{U}_n$. So for $0 \leq t \leq 1$,

$$P[H_n > t] = P[Z_n > t] + P[\{(1 - U_n) < t\} \cap \{H_n > t\}]. \quad (5.114)$$

When $0 \leq t \leq 1/2$, using part (i) of Lemma 5.4.1 for the first term on the right hand side of (5.114), and by conditioning on $1 - U_n$ for the second, we obtain

$$P[H_n > t] = (1-2t)^n + \int_0^t (1-t-s)^{n-1}\mathrm{d}s,$$

which yields (5.109). On the other hand, for $1/2 \leq t \leq 1$, we have $P[Z_n > t] = 0$ and, by conditioning on U_n,

$$P[H_n > t] = \int_t^1 (s-t)^{n-1}\mathrm{d}s,$$

which yields (5.110), completing the proof of (i). For (ii), we have

$$\begin{aligned} E[H_n^\beta] &= \int_0^1 P[H_n > t^{1/\beta}]\mathrm{d}t \\ &= \int_0^1 \frac{1}{n}\left(1 - t^{1/\beta}\right)^n \mathrm{d}t + \int_0^{2^{-\beta}} \frac{n-1}{n}\left(1 - 2t^{1/\beta}\right)^n \mathrm{d}t. \end{aligned} \quad (5.115)$$

Using the substitutions $y = t^{1/\beta}$ in the first integral in (5.115) and $y = 2t^{1/\beta}$ in the second yields

$$E[H_n^\beta] = \frac{\beta}{n}\left(1 + 2^{-\beta}(n-1)\right)\int_0^1 (1-y)^n y^{\beta-1}\mathrm{d}y,$$

and the fact that

$$\int_0^1 (1-y)^n y^{\beta-1} dy = \frac{\Gamma(n+1)\Gamma(\beta)}{\Gamma(n+1+\beta)}$$

then gives the result (5.111), from which (5.112) follows by taking $\beta = 1, 2$. Part (iii) follows from (5.111) by Stirling's formula. Part (iv) follows from (iv) of Lemma 5.4.1, along with the fact that $P[Z_n \neq H_n] \to 0$ as $n \to \infty$. □

Proposition 5.4.5 *As $n \to \infty$, the expected total weight of* ONG$(\mathcal{U}_n^0)$ *under weight function w_α, $\alpha > 0$, satisfies*

$$E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0)] = \frac{\Gamma(\alpha+1)}{1-\alpha} 2^{-\alpha} n^{1-\alpha} + \frac{1}{\alpha} - \frac{2^{-\alpha}}{\alpha(1-\alpha)} + O(n^{-\alpha}); \; (0 < \alpha(5.116)$$

$$E[\mathcal{O}^{1,1}(\mathcal{U}_n^0)] - \frac{1}{2}\log n = \frac{1}{2}\gamma + O(n^{-1}); \quad (5.117)$$

$$E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0)] = \frac{1}{\alpha} + \frac{2^{-\alpha}}{\alpha(\alpha-1)} + O(n^{1-\alpha}) \; (\alpha > 1) \quad (5.118)$$

Proof. We have

$$E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0)] = \sum_{i=1}^{n} \left(E[\mathcal{O}^{1,\alpha}(\mathcal{U}_i^0)] - E[\mathcal{O}^{1,\alpha}(\mathcal{U}_{i-1}^0)]\right) = \sum_{i=1}^{n} E[H_i^\alpha].$$

In the case where $\alpha = 1$, $E[H_i] = (2i)^{-1}$ by (5.112), and (5.117) follows by (C.3.5). Suppose $\alpha > 0$. Then from (5.111) we obtain for $\alpha > 0$, $\alpha \neq 1$,

$$\sum_{i=1}^{n} E[H_i^\alpha] = -\frac{\Gamma(n+1)\Gamma(\alpha-1)}{\Gamma(n+\alpha+1)} \left[(\alpha - 1 + 2^{-\alpha}) + \alpha n 2^{-\alpha}\right] + \frac{1}{\alpha} + \frac{2^{-\alpha}}{\alpha(\alpha-1)}. \quad (5.119)$$

Using the fact that $\alpha(\alpha-1)\Gamma(\alpha-1) = \Gamma(1+\alpha)$, and that, by Stirling's formula

$$\frac{\Gamma(n+1)}{\Gamma(n+\alpha+1)} = n^{-\alpha} + O(n^{-1-\alpha}),$$

we obtain (5.116) and (5.118) from (5.119). □

The next result will prove useful in relating our results on $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1})$ to $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0)$. We make use of the discussion above Lemma 5.4.2.

Lemma 5.4.6 *For a sequence of uniform random variables $\mathcal{U}_n$, we have that, for $0 < \alpha < 1$, as $n \to \infty$*

$$\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1}) - \mathcal{O}^{1,\alpha}(\mathcal{U}_n^0) \longrightarrow Q_\alpha, \quad (5.120)$$

where the convergence is in L^1 and almost sure, and Q_α is a nonnegative finite random variable.

Proof. Recall that we can decompose $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1})$ and $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0)$ into the increments Z_i and H_i respectively, for $i = 1, \ldots, n$, as described at (5.73). From the $0 < \alpha < 1$ case of (5.74), we have

$$E\left[H_i^\alpha - Z_i^\alpha \,|S_i^{i-1}\right] = \left(\frac{S_i^{i-1}}{2}\right)^{1+\alpha} E\left[(1+U)^\alpha - (1-U)^\alpha\right].$$

By (5.5), $E[(S_i^{i-1})^k] = O(i^{-k})$ for $k > 0$. So we have that, for $0 < \alpha < 1$, $H_i^\alpha - Z_i^\alpha$ are nonnegative random variables with expectation $O(i^{-1-\alpha})$. Thus we have that

$$0 \leq \mathcal{O}^{1,\alpha}(\mathcal{U}_n^0) - \mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1}) = \sum_{i=1}^{n}(H_i^\alpha - Z_i^\alpha) \to \sum_{i=1}^{\infty}(H_i^\alpha - Z_i^\alpha) < \infty,$$

as $n \to \infty$, where the convergence is almost sure and in L^1, and so we have (5.120). □

For the next few results, we will make use of the following fact. Writing $U = U_1$ for the position of the first arrival, by the self-similarity of the ONG we have that

$$\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0) \stackrel{\mathcal{D}}{=} U^\alpha \mathcal{O}^{1,\alpha}(\mathcal{U}_{N(n)}^{0,1}) + (1-U)^\alpha \mathcal{O}^{1,\alpha}(\mathcal{U}_{n-1-N(n)}^0), \tag{5.121}$$

where, given U, $N(n) \sim \text{Bin}(n-1, U)$ gives the number of points of $U_2, U_3, \ldots, U_n$ that arrive to the left of $U_1 = U$, and, given U and $N(n)$, the terms on the right are independent. Note that the term $U^\alpha \mathcal{O}^{1,\alpha}(\mathcal{U}_{N(n)}^{0,1})$ on the right hand side of (5.121) includes the edge of weight U^α from U_1 to 0.

Proposition 5.4.6 *Suppose* $1/2 < \alpha < 1$. *Then, as* $n \to \infty$,

$$\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n^0) \xrightarrow{\mathcal{D}} \tilde{H}_\alpha, \tag{5.122}$$

where $\tilde{H}_\alpha$ *satisfies the fixed-point equation (5.30), and* $E[\tilde{H}_\alpha] = 0$.

Proof. By Lemma 5.4.6, and the fact that the L^1 convergence in (5.120) implies

$$E\left[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^{0,1})\right] - E\left[\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0)\right] \to 0,$$

as $n \to \infty$, we have from Proposition 5.4.3 that for $1/2 < \alpha < 1$, there exists a random variable $\tilde{H}_\alpha$ such that as $n \to \infty$

$$\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n^0) \xrightarrow{\mathcal{D}} \tilde{H}_\alpha. \tag{5.123}$$

We now need to demonstrate that $\tilde{H}_\alpha$ satisfies the fixed-point equation (5.30). Renormalizing the $1/2 < \alpha < 1$ case of (5.121) by subtracting off expectations, using (5.69) and

(5.116), we obtain

$$\begin{aligned}\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n^0) &\stackrel{\mathcal{D}}{=} U^\alpha \tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_{N(n)}^{0,1}) + (1-U)^\alpha \tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_{n-1-N(n)}^0) + Cn^{(1/2)-\alpha} B_\alpha(n) \\ &+U^\alpha \left(1 - \frac{2^{-\alpha}}{1-\alpha}\right) + ((1-U)^\alpha - 1)\left(\frac{1}{\alpha} - \frac{2^{-\alpha}}{\alpha(1-\alpha)}\right) \\ &+U^\alpha h(N(n)) + (1-U)^\alpha h(n-1-N(n)) - h(n), \end{aligned} \tag{5.124}$$

where $h(n) \to 0$ as $n \to \infty$, C is a constant, and $B_\alpha(n)$ is as defined by (5.93). By Lemma 5.4.4, for $1/2 < \alpha < 1$, $n^{(1/2)-\alpha} B_\alpha(n)$ tends to 0 in L^2. Given U, we have that $N(n)$ and $n-1-N(n)$ tend to infinity almost surely. So, by Proposition 5.4.3, we have that $\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_{N(n)}^{0,1})$ converges in distribution to $\tilde{G}_\alpha$ as $n \to \infty$. Since by (5.123) $\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_{n-1-N(n)}^0)$ converges in distribution to $\tilde{H}_\alpha$, we deduce from (5.124) that $\tilde{H}_\alpha$ does indeed satisfy the fixed-point equation (5.30). Finally, the fact that $E[\tilde{H}_\alpha] = 0$ follows by taking expectations in (5.30). □

Proposition 5.4.7 *As $n \to \infty$, $\tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^0)$ converges in distribution to a random variable $\tilde{H}_1$, where $\tilde{H}_1$ satisfies the fixed-point equation (5.29). Further, $E[\tilde{H}_1] = 0$,*

$$\mathrm{Var}[\tilde{H}_1] = \frac{1}{8}(3 + \log 2) - \frac{\pi^2}{24} \approx 0.050410, \tag{5.125}$$

and $E[\tilde{H}_1^3] \approx 0.00323456$.

Proof. The fact that $\tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^0)$ converges in distribution follows from writing

$$\tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^0) = \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0,1}) - \left(\tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0,1}) - \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^0)\right), \tag{5.126}$$

and using Proposition 5.4.2. From Proposition 5.4.2 and (5.126), we have that as $n \to \infty$,

$$\tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^0) = (1,-1)\begin{pmatrix} \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0,1}) \\ \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^{0,1}) - \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n^0) \end{pmatrix} \xrightarrow{\mathcal{D}} (1,-1)\begin{pmatrix} \tilde{G}_1 \\ \tilde{Q} \end{pmatrix} = \tilde{G}_1 - \tilde{Q}.$$

By (5.84) we have that the term on the right of the above expression is equal in distribution to

$$\begin{aligned} &(1,-1)\begin{pmatrix} U & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \tilde{G}_1^{\{1\}} \\ \tilde{Q}^{\{1\}} \end{pmatrix} + (1,-1)\begin{pmatrix} 1-U & 0 \\ 0 & 1-U \end{pmatrix}\begin{pmatrix} \tilde{G}_1^{\{2\}} \\ \tilde{Q}^{\{2\}} \end{pmatrix} \\ &+(1,-1)\begin{pmatrix} \frac{1}{2}U\log U + \frac{1}{2}(1-U)\log(1-U) + \min\{U, 1-U\} \\ (1-2U)\mathbf{1}_{\{U>1/2\}} + \frac{1}{2}U \end{pmatrix} \\ = \; &U\tilde{G}_1^{\{1\}} + (1-U)(\tilde{G}_1^{\{2\}} - \tilde{Q}^{\{2\}}) + \frac{1}{2}U\log U + \frac{1}{2}(1-U)\log(1-U) + \frac{U}{2}. \end{aligned}$$

That is, the distribution of the limit $\tilde{H}_1 := \tilde{G}_1 - \tilde{Q}$ satisfies the fixed-point equation (5.29).

The fact that $E[\tilde{H}_1] = 0$ follows by taking expectations in (5.29). For the variance, we square both sides of (5.29) and take expectations, using independence and the fact that $E[\tilde{G}_1] = E[\tilde{H}_1] = 0$ to give

$$\begin{aligned} E[\tilde{H}_1^2] &= \frac{1}{3}E[\tilde{G}_1^2] + \frac{1}{3}E[\tilde{H}_1^2] + E[U^2/4] + \frac{1}{2}E[U^2(\log U)^2] + \frac{1}{2}E[U^2 \log U] \\ &+ \frac{1}{2}E[U(1-U)\log(1-U)] + \frac{1}{2}E[U(1-U)\log U \log(1-U)]. \end{aligned}$$

The integrals required for the expectations are standard, yielding (5.125). Similarly, we obtain the third moment $E[\tilde{H}_1^3] = 0.00323456\ldots$, although in this case numerical methods are required for some of the integrals, using the results for $E[\tilde{G}_1^2]$ and $E[\tilde{G}_1^3]$ in Theorem 5.4.1. □

Proposition 5.4.8 *Let $\alpha > 1$. Then there exists a random variable H_α such that as $n \to \infty$ we have $\mathcal{O}^{1,\alpha}(\mathcal{U}_n^0) \to H_\alpha$ almost surely and in L^2. Also, the random variable H_α satisfies the distributional fixed-point equality (5.31). Further, $E[H_\alpha] = (1/\alpha) + 2^{-\alpha}/(\alpha(\alpha-1))$.*

Proof. The convergence almost surely and in L^2 follows by a similar argument to that in the proof the corresponding result of Proposition 5.4.4, but using H_i rather than Z_i and with (5.113) in place of (5.66).

Consider the $\alpha > 1$ case of (5.121). As $n \to \infty$, $N(n)$ and $n - N(n)$ both tend to infinity almost surely, and so, by taking $n \to \infty$ in (5.121), and using the fact that $\mathcal{O}^{1,\alpha}(\mathcal{U}_{N(n)}^{0,1})$ converges almost surely to $1 + G_\alpha$ (see Proposition 5.4.4), and that $\mathcal{O}^{1,\alpha}(\mathcal{U}_{n-1-N(n)}^0)$ converges almost surely to H_α (by the argument above) we obtain the fixed-point equation (5.31).

The identity $E[H_\alpha] = \alpha^{-1} + 2^{-\alpha}\alpha^{-1}(\alpha-1)^{-1}$ is obtained either from (5.118), or by taking expectations in (5.31). Next, if we set $\tilde{H}_\alpha = H_\alpha - E[H_\alpha]$, (5.31) yields (5.30). □

Proof of Theorem 5.4.2. Part (i) of the theorem follows from Proposition 5.4.6. Part (ii) follows from Proposition 5.4.7 with (5.117), and part (iii) follows from Proposition 5.4.8. □

5.4.5 Proof of Theorem 5.2.2

In order to prove Theorem 5.2.2, we make use of our results from Section 5.4, in particular Theorems 5.4.1 and 5.4.2.

Proof of Theorem 5.2.2. Write $U = U_1$ for the position of the first arrival. There is no edge from U_1 (as it is the first point), so using the self-similarity of the ONG, we have that

$$\mathcal{O}^{1,\alpha}(\mathcal{U}_n) \stackrel{\mathcal{D}}{=} U^\alpha \mathcal{O}^{1,\alpha}_{\{1\}}(\mathcal{U}^0_{N(n)}) + (1-U)^\alpha \mathcal{O}^{1,\alpha}_{\{2\}}(\mathcal{U}^0_{n-1-N(n)}), \tag{5.127}$$

where, given U, $N(n) \sim \text{Bin}(n-1, U)$ gives the number of points of $U_2, U_3, \ldots, U_n$ that arrive to the left of $U_1 = U$, and, given U and $N(n)$, $\mathcal{O}^{1,\alpha}_{\{1\}}(\mathcal{U}^0_{N(n)})$ and $\mathcal{O}^{1,\alpha}_{\{2\}}(\mathcal{U}^0_{n-1-N(n)})$ are independent copies of $\mathcal{O}^{1,\alpha}(\mathcal{U}^0_{N(n)})$ and $\mathcal{O}^{1,\alpha}(\mathcal{U}^0_{n-1-N(n)})$ respectively. We will use the notation

$$\begin{aligned}
\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}^0_{N(n)}) &:= \left(\mathcal{O}^{1,\alpha}(\mathcal{U}^0_{N(n)}) - E[\mathcal{O}^{1,\alpha}(\mathcal{U}^0_{N(n)})|N(n)]\right)\mathbf{1}_{\{N(n)>0\}}, &(5.128)\\
\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}^0_{n-1-N(n)}) &:= \left(\mathcal{O}^{1,\alpha}(\mathcal{U}^0_{n-1-N(n)}) - E[\mathcal{O}^{1,\alpha}(\mathcal{U}^0_{n-1-N(n)})|N(n)]\right)\mathbf{1}_{\{N(n)>0\}}.
\end{aligned}$$

Now, we prove part (i) of the theorem. Suppose $1/2 < \alpha < 1$. Then, taking expectations in (5.127), from (5.116) we have, recalling the definition of $B_\alpha(n)$ from (5.93), for $1/2 < \alpha < 1$

$$\begin{aligned}
E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n)|U, N(n)] &= \frac{\Gamma(\alpha+1)}{1-\alpha}2^{-\alpha}(n-1)^{(1/2)-\alpha}B_\alpha(n)\\
&\quad + (U^\alpha + (1-U)^\alpha)\left(\frac{1}{\alpha} - \frac{2^{-\alpha}}{\alpha(1-\alpha)}\right)\\
&\quad + U^\alpha h(N(n)) + (1-U)^\alpha h(n-1-N(n)) + o(1),
\end{aligned}$$

where $h(n) = o(1)$ as $n \to \infty$. Now, by Lemma 5.4.4 $E[(n-1)^{(1/2)-\alpha}B_\alpha(n)] \to 0$ as $n \to \infty$ for $1/2 < \alpha < 1$, so we have

$$E[\mathcal{O}^{1,\alpha}(\mathcal{U}_n)] = \frac{2}{1+\alpha}\left(\frac{1}{\alpha} - \frac{2^{-\alpha}}{\alpha(1-\alpha)}\right) + o(1). \tag{5.129}$$

Then, subtracting off expectations in (5.127), using (5.129) and (5.116), with the notation of (5.128), gives for $1/2 < \alpha < 1$

$$\begin{aligned}
\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}_n) &\stackrel{\mathcal{D}}{=} U^\alpha \tilde{\mathcal{O}}^{1,\alpha}_{\{1\}}(\mathcal{U}^0_{N(n)}) + (1-U)^\alpha \tilde{\mathcal{O}}^{1,\alpha}_{\{2\}}(\mathcal{U}^0_{n-1-N(n)}) + \frac{\Gamma(\alpha+1)}{1-\alpha}2^{-\alpha}n^{(1/2)-\alpha}B_\alpha(n)\\
&\quad + \left(U^\alpha + (1-U)^\alpha - \frac{2}{1+\alpha}\right)\left(\frac{1}{\alpha} - \frac{2^{-\alpha}}{\alpha(1-\alpha)}\right)\\
&\quad + U^\alpha h(N(n)) + (1-U)^\alpha h(n-1-N(n)) + o(1),
\end{aligned}$$

where again $h(n) = o(1)$ as $n \to \infty$. Now, letting $n \to \infty$ and using Lemma 5.4.4 yields (5.34).

Now, to prove part (ii) of the theorem, suppose $\alpha = 1$. From the $\alpha = 1$ case of (5.127), using (5.117), we have

$$\begin{aligned} E[\mathcal{O}^{1,1}(\mathcal{U}_n)|U, N(n)] &= \frac{1}{2}\log n + \frac{1}{2}\gamma + \frac{1}{2}U(\log N(n) - \log n + h(N(n))) + o(1) \\ &\quad + \frac{1}{2}(1-U)(\log(n-1-N(n)) - \log n + h(n-1-N(n))), \end{aligned}$$

where $h(n) \to 0$ as $n \to \infty$. Thus, using Lemma 5.4.3 we have

$$\begin{aligned} E[\mathcal{O}^{1,1}(\mathcal{U}_n)] - \frac{1}{2}\log n &\to \frac{1}{2}\gamma + \frac{1}{2}E[U\log U] + \frac{1}{2}E[(1-U)\log(1-U)] \\ &= \frac{1}{2}\gamma - \frac{1}{4}. \end{aligned} \tag{5.130}$$

Now, subtracting off the expectations from both sides of the $\alpha = 1$ case of (5.127), using (5.130) and (5.117), with the notation of (5.128), gives

$$\begin{aligned} \tilde{\mathcal{O}}^{1,1}(\mathcal{U}_n) &\stackrel{\mathcal{D}}{=} U\tilde{\mathcal{O}}^{1,1}_{\{1\}}(\mathcal{U}^0_{N(n)}) + (1-U)\tilde{\mathcal{O}}^{1,1}_{\{2\}}(\mathcal{U}^0_{n-1-N(n)}) + \frac{1}{4} + o(1) \\ &\quad + \frac{U}{2}(\log^+ N(n) - \log n) + \frac{1-U}{2}(\log^+(n-1-N(n)) - \log n) \\ &\quad + Uh(N(n)) + (1-U)h(n-1-N(n)), \end{aligned} \tag{5.131}$$

with $h(n) \to 0$ as $n \to \infty$. We now use the fact that $N(n)$ and $n - N(n)$ tend to infinity almost surely, the independence given U and $N(n)$, Lemma 5.4.3, and the convergence in distribution of $\tilde{\mathcal{O}}^{1,1}(\mathcal{U}^0_n)$ (Proposition 5.4.7) to obtain (5.35). The convergence of expectations version also follows from (5.130).

Finally we prove part (iii) of the theorem. Suppose $\alpha > 1$. Consider the $\alpha > 1$ case of (5.127). Now we use the fact that $N(n)$ and $n - N(n)$ tend to infinity almost surely, the independence given U and $N(n)$, and the convergence in L^2 and almost surely of $\tilde{\mathcal{O}}^{1,\alpha}(\mathcal{U}^0_n)$ (for $\alpha > 1$) to obtain the stated result. This completes the proof of the theorem. □

5.5 The DLF and DLT: analysis

5.5.1 The mean total weight of the DLF and DLT

First we consider the rooted case, i.e. the DLT on $\mathcal{U}^0_n$. For $n \in \mathbf{N}$ denote by Z_n the random variable given by the gain in length of the tree on the addition of one point (U_n) to an existing $n-1$ points in the DLT on a sequence of uniform random variables $\mathcal{U}^0_{n-1}$, i.e. with the conventions $D^1(\mathcal{U}^0_0) = 0$ and $U_0 = 0$, we set

$$Z_n := D^1(\mathcal{U}^0_n) - D^1(\mathcal{U}^0_{n-1}) = U_n - \max\{U_j : 0 \le j < n, U_j < U_n\}. \tag{5.132}$$

Thus, with weight exponent α, the nth edge to be added has weight Z_n^α.

Lemma 5.5.1 *(i) Z_n has distribution function F_n given by $F_n(t) = 0$ for $t < 0$, $F_n(t) = 1$ for $t > 1$, and $F_n(t) = 1 - (1-t)^n$ for $0 \leq t \leq 1$.*

(ii) For $\beta > 0$, Z_n has moments

$$E[Z_n^\beta] = \frac{\Gamma(n+1)\Gamma(1+\beta)}{\Gamma(1+\beta+n)}. \tag{5.133}$$

In particular,

$$E[Z_n] = \frac{1}{n+1}; \quad \mathrm{Var}[Z_n] = \frac{n}{(n+1)^2(n+2)}. \tag{5.134}$$

(iii) For $\beta > 0$, as $n \to \infty$ we have

$$E[Z_n^\beta] \sim \Gamma(\beta+1)n^{-\beta}. \tag{5.135}$$

(iv) As $n \to \infty$, nZ_n converges in distribution to an exponential random variable with parameter 1.

Proof. For $0 \leq t \leq 1$ we have

$$P[Z_n > t] = P[U_n > t \text{ and none of } U_1, \ldots, U_{n-1} \text{ lies in } (U_n - t, U_n)] = (1-t)^n,$$

and (i) follows. We then obtain (ii) since for any $\beta > 0$

$$E[Z_n^\beta] = \int_0^1 P[Z_n > t^{1/\beta}]\mathrm{d}t = \int_0^1 (1 - t^{1/\beta})^n \mathrm{d}t.$$

Then (iii) follows by Stirling's formula. For (iv), we have from (i) that, for $t \in [0, \infty)$, and n large enough so that $(t/n) \leq 1$,

$$P[nZ_n \leq t] = F_n\left(\frac{t}{n}\right) = 1 - \left(1 - \frac{t}{n}\right)^n \to 1 - e^{-t}, \text{ as } n \to \infty.$$

But $1 - e^{-t}$, $t \geq 0$ is the exponential distribution function with parameter 1. $\square$

Remark. Note that Z_n has the same distribution as the spacing S_1^n (see Section 5.1.1). In Lemma 5.5.3 we get some further insight into this.

The following result gives the asymptotic behaviour of the expected total weight of the DLT. Recall that γ denotes Euler's constant, as at (C.3.5).

Proposition 5.5.1 *As $n \to \infty$ the expected total weight of the DLT under α-power weighting on $\mathcal{U}_n^0$ satisfies*

$$E[D^\alpha(\mathcal{U}_n^0)] \sim \frac{\Gamma(\alpha+1)}{1-\alpha} n^{1-\alpha} \quad (0 < \alpha < 1); \tag{5.136}$$

$$E[D^1(\mathcal{U}_n^0)] - \log n \to \gamma - 1; \tag{5.137}$$

$$E[D^\alpha(\mathcal{U}_n^0)] = \frac{1}{\alpha - 1} + O(n^{1-\alpha}) \quad (\alpha > 1). \tag{5.138}$$

Proof. We have

$$E[D^\alpha(\mathcal{U}_n^0)] = \sum_{i=1}^{n} \left(E[D^\alpha(\mathcal{U}_i^0)] - E[D^\alpha(\mathcal{U}_{i-1}^0)]\right) = \sum_{i=1}^{n} E[Z_i^\alpha].$$

In the case where $\alpha = 1$, $E[Z_i] = (i+1)^{-1}$ by (5.134), and (5.137) follows by (C.3.5). For general $\alpha > 0$, $\alpha \neq 1$, from (5.133) we have that

$$E[D^\alpha(\mathcal{U}_n^0)] = \Gamma(1+\alpha) \sum_{i=1}^{n} \frac{\Gamma(i+1)}{\Gamma(1+\alpha+i)} = \frac{1}{\alpha-1} - \frac{\Gamma(1+\alpha)\Gamma(n+2)}{(\alpha-1)\Gamma(n+1+\alpha)}. \tag{5.139}$$

By Stirling's formula, the last term satisfies

$$-\frac{\Gamma(1+\alpha)\Gamma(n+2)}{(\alpha-1)\Gamma(n+1+\alpha)} = -\frac{\Gamma(1+\alpha)}{\alpha-1} n^{1-\alpha}(1 + O(n^{-1})), \tag{5.140}$$

which tends to zero as $n \to \infty$ for $\alpha > 1$, to give us (5.138). For $\alpha < 1$, we have (5.136) from (5.139) and (5.140). □

Now consider the unrooted case, i.e., the directed linear forest. For $\mathcal{U}_n$ as above the total weight of the DLF is denoted $D^\alpha(\mathcal{U}_n)$, and the centred total weight is $\tilde{D}^\alpha(\mathcal{U}_n) := D^\alpha(\mathcal{U}_n) - E[D^\alpha(\mathcal{U}_n)]$.

Let $D_0^\alpha(\mathcal{U}_n^0)$ denote the total weight of the edges incident to 0 in the DLT on $\mathcal{U}_n^0$. Then, given $\mathcal{U}_n$, by the construction of the DLF and DLT we have that

$$D^\alpha(\mathcal{U}_n^0) = D^\alpha(\mathcal{U}_n) + D_0^\alpha(\mathcal{U}_n^0). \tag{5.141}$$

The following lemma says that $D_0^\alpha(\mathcal{U}_n^0)$ converges to a random variable that has the generalized Dickman distribution with parameter $1/\alpha$ (see Chapter 4, in particular Propositions C.4.1 and C.4.2).

Lemma 5.5.2 *Let $\alpha > 0$. There is a random variable D_0^α with the generalized Dickman distribution with parameter $1/\alpha$, such that as $n \to \infty$, we have that $D_0^\alpha(\mathcal{U}_n^0) \to D_0^\alpha$, almost surely and in L^2.*

Proof. Let $\delta_D(\mathcal{U}_n^0)$ denote the degree of the origin in the directed linear tree on $\mathcal{U}_n^0$, so that $\delta_D(\mathcal{U}_n^0)$ is the number of lower records in the sequence $(U_1, \ldots, U_n)$. Then

$$D_0^\alpha(\mathcal{U}_n^0) = V_1^\alpha + (V_1V_2)^\alpha + \cdots + (V_1 \cdots V_{\delta_D(\mathcal{U}_n^0)})^\alpha, \tag{5.142}$$

where $(V_1, V_2, \ldots)$ is a certain sequence of independent uniform random variables on $(0,1)$, namely the ratios between successive lower records of the sequence (U_n). The sum $V_1^\alpha + (V_1V_2)^\alpha + (V_1V_2V_3)^\alpha + \cdots$ has nonnegative terms and finite expectation, so it converges almost surely to a limit which we denote D_0^α. Then D_0^α has the generalized Dickman distribution with parameter $1/\alpha$ (see Proposition C.4.1).

Since $\delta_D(\mathcal{U}_n^0)$ tends to infinity almost surely as $n \to \infty$, we have $D_0^\alpha(\mathcal{U}_n^0) \to D_0^\alpha$ almost surely. Also, $E[(D_0^\alpha)^2] < \infty$, by (C.4.9), and $(D_0^\alpha - D_0^\alpha(\mathcal{U}_n^0))^2 \leq (D_0^\alpha)^2$ for all n. Thus $E[(D_0^\alpha(\mathcal{U}_n^0) - D_0^\alpha)^2] \to 0$ by the dominated convergence theorem, and so we have the L^2 convergence as well. □

Proposition 5.5.2 *As $n \to \infty$ the expected total weight of the DLF under α-power weighting on $\mathcal{U}_n$ satisfies*

$$E[D^\alpha(\mathcal{U}_n)] \sim \frac{\Gamma(\alpha+1)}{1-\alpha} n^{1-\alpha} \quad (0 < \alpha < 1); \tag{5.143}$$

$$E[D^1(\mathcal{U}_n)] - \log n \to \gamma - 2; \tag{5.144}$$

$$E[D^\alpha(\mathcal{U}_n)] \to \frac{1}{\alpha(\alpha-1)} \quad (\alpha > 1). \tag{5.145}$$

Proof. By (5.141) we have $E[D^\alpha(\mathcal{U}_n)] = E[D^\alpha(\mathcal{U}_n^0)] - E[D_0^\alpha(\mathcal{U}_n^0)]$. By Lemma 5.5.2 and (C.4.9),

$$E[D_0^\alpha(\mathcal{U}_n^0)] \longrightarrow E[D_0^\alpha] = 1/\alpha.$$

We then obtain (5.143), (5.144) and (5.145) from Proposition 5.5.1. □

In the following sections we analyse the limiting behaviour of the total weight of the DLT and DLF. In some cases, we follow the contraction method for establishing fixed-point limits as employed in the analysis of the ONG in Section 5.4. In the particular case $\alpha = 1$, a more direct approach is enabled by the special covariance structure of the DLT (see Section 5.5.3). In all cases, the self-similarity of the DLF and DLT will be of central importance.

Taking $U = U_1$ (the first arrival) here, by the self-similarity of the DLT we have that, for $n \in \mathbf{N}$,

$$D^\alpha(\mathcal{U}_n^0) \stackrel{\mathcal{D}}{=} U^\alpha D_{\{1\}}^\alpha(\mathcal{U}_{N(n)}^0) + (1-U)^\alpha D_{\{2\}}^\alpha(\mathcal{U}_{n-1-N(n)}^0) + U^\alpha, \tag{5.146}$$

where $N(n) \sim \text{Bin}(n-1, U)$, given U, and, given U and $N(n)$, $D^{\alpha}_{\{1\}}(\mathcal{U}^0_{N(n)})$ and $D^{\alpha}_{\{2\}}(\mathcal{U}^0_{n-1-N(n)})$ are independent with the distribution of $D^{\alpha}(\mathcal{U}^0_{N(n)})$ and $D^{\alpha}(\mathcal{U}^0_{n-1-N(n)})$, respectively.

5.5.2 Limit behaviour for $1/2 < \alpha < 1$

Proposition 5.5.3 *For $1/2 < \alpha < 1$, we have that, as $n \to \infty$,*

$$\tilde{D}^{\alpha}(\mathcal{U}^0_n) \xrightarrow{\mathcal{D}} \tilde{D}_{\alpha},$$

where $\tilde{D}_{\alpha}$ has the distribution given by the fixed-point equation (5.37), and $E[\tilde{D}_{\alpha}] = 0$.

The proof of Proposition 5.5.3 follows similar lines to that of Proposition 5.4.3; we do not give the details this time round.

Proposition 5.5.4 *For $1/2 < \alpha < 1$, we have that, as $n \to \infty$,*

$$\tilde{D}^{\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \tilde{F}_{\alpha},$$

where $\tilde{F}_{\alpha}$ has the distribution given by the fixed-point equation (5.38), and $E[\tilde{F}_{\alpha}] = 0$.

Proof. This follows from Proposition 5.5.3 and Lemma 5.5.2, in the same manner as the proof of Proposition 5.4.6. □

5.5.3 Orthogonal increments for $\alpha = 1$

In this section we shall show (in Lemma 5.5.5) that when $\alpha = 1$, the variables $Z_i, i \geq 1$ are mutually orthogonal, in the sense of having zero covariances, which will be used later on to establish convergence of the (centred) total length of the DLT. We use the spacings notation of Section 5.1.1.

We can arrange the spacings themselves $(S^n_j, 1 \leq j \leq n+1)$ into increasing order to give $S^n_{(1)}, S^n_{(2)}, \ldots, S^n_{(n+1)}$. Then let $\mathcal{F}^n_S$ denote the sigma field generated by these ordered spacings, so that

$$\mathcal{F}^n_S = \sigma\left(S^n_{(1)}, \ldots, S^n_{(n+1)}\right). \tag{5.147}$$

The following interpretation of $\mathcal{F}^n_S$ may be helpful. The set $(0,1) \setminus \{U_1, \ldots, U_n\}$ consists almost surely of $n+1$ connected components ('fragments') of total length 1, and $\mathcal{F}^n_S$ is the σ-field generated by the collection of lengths of these fragments, ignoring the order in which they appear.

By definition, the value of Z_n must be one of the (ordered) spacings $S^n_{(1)}, \dots, S^n_{(n+1)}$. The next result says that, given the values of these spacings, each of the possible values for Z_n is equally likely.

Lemma 5.5.3 *For $n \geq 1$ we have*

$$P\left[Z_n = S^n_{(i)} \middle| \mathcal{F}^n_S\right] = \frac{1}{n+1} \quad \text{a.s., } \textit{for } i = 1, \dots, n+1. \tag{5.148}$$

Hence,

$$E\left[Z_n \left| \mathcal{F}^n_S\right.\right] = \frac{1}{n+1} \sum_{i=1}^{n+1} S^n_{(i)} = \frac{1}{n+1}. \tag{5.149}$$

Proof. We have (see Section 5.1.1) that $(S^n_1, \dots, S^n_{n+1})$ has the symmetric Dirichlet distribution with parameter 1. In particular, the S^n_j are exchangeable. Thus given $S^n_{(1)}, \dots, S^n_{(n+1)}$, i.e. $\mathcal{F}^n_S$, the actual values of $S^n_1, \dots, S^n_{n+1}$ are equally likely to be any permutation of $S^n_{(1)}, \dots, S^n_{(n+1)}$, and given $S^n_1, \dots, S^n_{n+1}$ the value of Z_n is equally likely to be any of $S^n_1, \dots, S^n_n$ (but cannot be S^n_{n+1}).

Hence, given $S^n_{(1)}, \dots, S^n_{(n+1)}$ the probability that $Z_n = S^n_{(i)}$ is $(1/n) \times n/(n+1) = 1/(n+1)$, i.e. we have (5.148), and then (5.149) follows since $\sum_{j=1}^{n+1} S^n_{(j)} = 1$. □

Lemma 5.5.4 *Let $1 \leq n < \ell$. Given $\mathcal{F}^n_S$, Z_ℓ and Z_n are conditionally independent.*

Proof. Given $\mathcal{F}^n_S$, we have $S^n_{(1)}, \dots, S^n_{(n+1)}$, and by (5.148), the (conditional) distribution of Z_n is uniform on $\{S^n_{(1)}, \dots, S^n_{(n+1)}\}$. The conditional distribution of Z_ℓ, $\ell > n$, given $\mathcal{F}^n_S$, depends only on $S^n_{(1)}, \dots, S^n_{(n+1)}$ and not which one of them Z_n happens to be. Hence Z_n and Z_ℓ are conditionally independent. □

Lemma 5.5.5 *For $1 \leq n < \ell$, the random variables Z_n, Z_ℓ satisfy* $\mathrm{Cov}\,[Z_n, Z_\ell] = 0$.

Proof. From Lemmas 5.5.4 and 5.5.3,

$$E\left[Z_n Z_\ell | \mathcal{F}^n_S\right] \;=\; E\left[Z_n | \mathcal{F}^n_S\right] E\left[Z_\ell | \mathcal{F}^n_S\right] = \frac{1}{n+1} E\left[Z_\ell | \mathcal{F}^n_S\right],$$

and by taking expectations we obtain

$$E\left[Z_n Z_\ell\right] \;=\; \frac{1}{n+1} E\left[Z_\ell\right] = \frac{1}{n+1} \cdot \frac{1}{\ell+1} = E[Z_n] \cdot E[Z_\ell].$$

Hence the covariance of Z_n and Z_ℓ is zero. □

Remarks. (a) Calculations yield, for example, that $E[D^1(\mathcal{U}_1^0)] = E[Z_1] = 1/2$, $E[D^1(\mathcal{U}_2^0)] = 5/6$, and $\mathrm{Var}[Z_1] = 1/12$, $\mathrm{Var}[Z_2] = 1/18$, $\mathrm{Var}[D^1(\mathcal{U}_2^0)] = 5/36$.

(b) The orthogonality structure of the Z_n^α is unique to the $\alpha = 1$ case. For example, it can be shown that, for $\alpha > 0$,

$$E[Z_1^\alpha]E[Z_2^\alpha] = \frac{2}{(1+\alpha)^2(2+\alpha)}, \text{ and } E[Z_1^\alpha Z_2^\alpha] = \frac{1}{2(1+\alpha)^2}\left(1 + \frac{2\Gamma(\alpha+2)^2}{\Gamma(2\alpha+3)}\right).$$

Then

$$\mathrm{Cov}[Z_1^\alpha, Z_2^\alpha] = \frac{(\alpha-2)\Gamma(2\alpha+3) + 2(\alpha+2)\Gamma(\alpha+2)^2}{2(\alpha+1)^2(\alpha+2)\Gamma(2\alpha+3)},$$

and this quantity is zero only if $\alpha = 1$; it is positive for $\alpha > 1$ and negative for $0 < \alpha < 1$.

5.5.4 Limit behaviour for $\alpha > 1$

We now consider the limit distribution of the total weight of the DLT and DLF. In the present section we consider the case of α-power weighted edges with $\alpha > 1$; that is, we work towards part (iv) of Theorem 5.2.3. To describe the moments of the limiting distribution of $D^\alpha(\mathcal{U}_n^0)$ and $D^\alpha(\mathcal{U}_n)$, we introduce the notation

$$J(\alpha) := \int_0^1 u^\alpha(1-u)^\alpha \mathrm{d}u = 2^{-1-2\alpha}\sqrt{\pi}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+3/2)}. \tag{5.150}$$

We start with the rooted case ($D^\alpha(\mathcal{U}_n^0)$), and subsequently consider the unrooted case ($D^\alpha(\mathcal{U}_n)$).

Proposition 5.5.5 *Let $\alpha > 1$. Then there exists a random variable D_α such that as $n \to \infty$ we have $D^\alpha(\mathcal{U}_n^0) \to D_\alpha$ almost surely and in L^2. Also, the random variable D_α satisfies the distributional fixed-point equality (5.39). Further, $E[D_\alpha] = 1/(\alpha-1)$ and*

$$\mathrm{Var}[D_\alpha] = \frac{\alpha\,(\alpha - 2 + 2(2\alpha+1)J(\alpha))}{(\alpha-1)^2(2\alpha-1)}. \tag{5.151}$$

Proof. Let Z_i be the length of the ith edge of the DLT, as defined at (5.132). Let $D_\alpha := \sum_{i=1}^\infty Z_i^\alpha$. The sum converges almost surely since it has non-negative terms and, by (5.135), has finite expectation for $\alpha > 1$. By (5.135) and Cauchy-Schwarz, there exists a constant $0 < C < \infty$ such that

$$E[D_\alpha^2] = \sum_{i=1}^\infty \sum_{j=1}^\infty E[Z_i^\alpha Z_j^\alpha] \leq C \sum_{i=1}^\infty \sum_{j=1}^\infty i^{-\alpha} j^{-\alpha} < \infty,$$

since $\alpha > 1$. The L^2 convergence then follows from the dominated convergence theorem.

Consider the $\alpha > 1$ case of (5.146). As $n \to \infty$, $N(n)$ and $n - N(n)$ both tend to infinity almost surely, and so, by taking $n \to \infty$ in (5.146), we obtain the fixed-point equation (5.39).

The identity $E[D_\alpha] = (\alpha - 1)^{-1}$ is obtained either from (5.138) of Proposition 5.5.1, or by taking expectations in (5.39). Next, if we set $\tilde{D}_\alpha = D_\alpha - E[D_\alpha]$, (5.39) yields (5.37). Then, using the definition (5.150) of $J(\alpha)$, the fact that $E[\tilde{D}_\alpha] = 0$, and independence, we obtain from (5.37) that

$$E[\tilde{D}_\alpha^2] \quad = \quad \frac{2E[\tilde{D}_\alpha^2]}{2\alpha+1} + \frac{\alpha^2+1}{(\alpha-1)^2(2\alpha+1)} + \frac{2\alpha J(\alpha)}{(\alpha-1)^2} - \frac{1}{(\alpha-1)^2},$$

and rearranging this gives (5.151). □

Recall from Lemma 5.5.2 that D_0^α is the limiting weight of edges attached to the origin in the DLT on uniform points. Combining this fact with Proposition 5.5.5, we obtain a similar result to the latter for the unrooted case as follows:

Proposition 5.5.6 *Let $\alpha > 1$. There is a random variable F_α, satisfying the distributional fixed-point equality (5.40), such that $D^\alpha(\mathcal{U}_n) \to F_\alpha$, as $n \to \infty$, almost surely and in L^2. Further, $E[F_\alpha] = 1/(\alpha(\alpha-1))$, and*

$$\mathrm{Var}[F_\alpha] = \frac{1}{2\alpha}\mathrm{Var}[D_\alpha] + \frac{\alpha + 2(2\alpha+1)J(\alpha) - 2}{2\alpha^2(\alpha-1)^2}, \tag{5.152}$$

where $J(\alpha)$ is given by (5.150) and $\mathrm{Var}[D_\alpha]$ *by (5.151).*

Proof. By Lemma 5.5.2 and Proposition 5.5.5, there are random variables D_α and D_0^α such that as $n \to \infty$ we have $D^\alpha(\mathcal{U}_n^0) \xrightarrow{L^2} D_\alpha$ and $D_0^\alpha(\mathcal{U}_n^0) \xrightarrow{L^2} D_0^\alpha$, also with almost sure convergence in both cases. Hence, setting $F_\alpha := D_\alpha - D_0^\alpha$, we have by (5.141) that

$$D^\alpha(\mathcal{U}_n) = D^\alpha(\mathcal{U}_n^0) - D_0^\alpha(\mathcal{U}_n^0) \to F_\alpha, \qquad \text{a.s. and in } L^2. \tag{5.153}$$

Next, we show that F_α satisfies the distributional fixed-point equality (5.40). The self-similarity of the DLT implies that

$$D^\alpha(\mathcal{U}_n) \stackrel{\mathcal{D}}{=} U^\alpha D^\alpha(\mathcal{U}_{N(n)}) + (1-U)^\alpha D^\alpha(\mathcal{U}_{n-1-N(n)}^0), \tag{5.154}$$

where $N(n) \sim \mathrm{Bin}(n-1, U)$, given U, and $D^\alpha(\mathcal{U}_{N(n)})$ and $D^\alpha(\mathcal{U}_{n-1-N(n)}^0)$ are independent, given U and $N(n)$. As $n \to \infty$, $N(n)$ and $n - N(n)$ both tend to infinity almost surely, so taking $n \to \infty$ in (5.154), using Proposition 5.5.5 and (5.153), we obtain the fixed-point equation (5.40).

The identity $E[F_\alpha] = \alpha^{-1}(\alpha-1)^{-1}$ is obtained either by (5.145), or by taking expectations in (5.40) and using the formula for $E[D_\alpha]$ in Proposition 5.5.5. Then with $\tilde{F}_\alpha := F_\alpha - E[F_\alpha]$, we obtain (5.38) from (5.40), and using independence and the fact that $E[\tilde{F}_\alpha] = E[\tilde{D}_\alpha] = 0$ we obtain

$$\frac{2\alpha}{2\alpha+1}E[\tilde{F}_\alpha^2] = \frac{E[\tilde{D}_\alpha^2]}{2\alpha+1} + \frac{2\alpha J(\alpha)-1}{\alpha^2(\alpha-1)^2} + \frac{\alpha^2+1}{\alpha^2(\alpha-1)^2(2\alpha+1)},$$

which yields (5.152). □

Examples. When $\alpha = 2$ we have that $E[D_2] = 1$ and $J(2) = 1/30$, so that $\mathrm{Var}[D_2] = 2/9$. Also, $E[F_2] = 1/2$ and $\mathrm{Var}[F_2] = 7/72 \approx 0.0972$.

5.5.5 Limit behaviour for $\alpha = 1$

Unlike in the case $\alpha > 1$, for $\alpha = 1$ the mean of the total weight $D^1(\mathcal{U}_n^0)$ diverges as $n \to \infty$ (see Proposition 5.5.1), so clearly there is no limiting distribution for $D^1(\mathcal{U}_n^0)$. Nevertheless, by using the orthogonality of the increments of the sequence $(D^1(\mathcal{U}_n^0), n \geq 1)$, we are able to show that the *centred* total weight $\tilde{D}^1(\mathcal{U}_n^0)$ does converge in distribution (in fact, in L^2) to a limiting random variable, and likewise for the unrooted case; this is our next result.

Subsequently, we shall characterize the distribution of the limiting random variable (for both the rooted and unrooted cases) by a fixed-point identity, and thereby complete the proof of Theorem 5.2.3 (i).

Proposition 5.5.7 *(i) As $n \to \infty$, the random variable $\tilde{D}^1(\mathcal{U}_n^0)$ converges in L^2 to a limiting random variable $\tilde{D}_1$, with $E[\tilde{D}_1] = 0$ and $\mathrm{Var}[\tilde{D}_1] = 2 - \pi^2/6$. In particular, $\mathrm{Var}\,[D^1(\mathcal{U}_n^0)] \to 2 - \pi^2/6$ as $n \to \infty$.*

(ii) As $n \to \infty$, $\tilde{D}^1(\mathcal{U}_n)$ converges in L^2 to the limiting random variable $\tilde{F}_1 := \tilde{D}_1 - D_0^1 + 1$.

Proof. Adopt the convention $D^1(\mathcal{U}_0^0) = 0$. By the orthogonality of the Z_j (Lemma 5.5.5) and (5.134), for $0 \leq \ell < n$,

$$\begin{aligned}
\mathrm{Var}\left[\tilde{D}^1(\mathcal{U}_n^0) - \tilde{D}^1(\mathcal{U}_\ell^0)\right] &= \mathrm{Var} \sum_{j=\ell+1}^{n} (Z_j - E[Z_j]) \\
&= \sum_{j=\ell+1}^{n} \frac{j}{(j+1)^2(j+2)} \longrightarrow 0 \text{ as } n, \ell \to \infty.
\end{aligned}$$

Hence $\tilde{D}_1(\mathcal{U}_n^0)$ is a Cauchy sequence in L^2, and so converges in L^2 to a limiting random variable, which we denote $\tilde{D}_1$. Then $E[\tilde{D}_1] = \lim_{n\to\infty} E[\tilde{D}_1(\mathcal{U}_n^0)] = 0$, and

$$\mathrm{Var}[\tilde{D}_1] = \lim_{n\to\infty} \mathrm{Var}\left[\tilde{D}^1(\mathcal{U}_n^0)\right] = \sum_{j=1}^{\infty} \frac{j}{(j+1)^2(j+2)}$$

$$= \sum_{j=1}^{\infty}\left[\frac{2}{j+1} - \frac{2}{j+2}\right] - \sum_{j=1}^{\infty}\frac{1}{(j+1)^2} = 1 - \left(\frac{\pi^2}{6} - 1\right) = 2 - \frac{\pi^2}{6}.$$

It remains to prove part (ii), the convergence for the centred total length of the DLF $\tilde{D}^1(\mathcal{U}_n)$. We have by (5.141) that

$$\tilde{D}^1(\mathcal{U}_n) = \tilde{D}^1(\mathcal{U}_n^0) - D_0^1(\mathcal{U}_n^0) + E[D_0^1(\mathcal{U}_n^0)] \xrightarrow{L^2} \tilde{D}_1 - D_0^1 + 1,$$

where the convergence follows by Lemma 5.5.2 and part (i). Thus $\tilde{D}^1(\mathcal{U}_n)$ converges in L^2 as $n \to \infty$. □

For the next few results (which we will need, in particular, in Chapter 6) it is more convenient to consider the DLF defined on a Poisson number of points. Let $(U_1, U_2, \ldots)$ be a sequence of independent uniformly distributed random variables in $(0,1)$, and let $(N(t), t \geq 0)$ be the counting process of a homogeneous Poisson process of unit rate in $(0,\infty)$, independent of $(U_1, U_2, \ldots)$. Thus $N(t)$ is a Poisson variable with parameter t. As before, let $\mathcal{U}_n = (U_1, \ldots, U_n)$, and (for this section only) let $\mathcal{P}_t := \mathcal{U}_{N(t)}$. Let $\mathcal{P}_t^0 := \mathcal{U}_{N(t)}^0$, so that $\mathcal{P}_t^0 = (0, U_1, U_2, \ldots, U_{N(t)})$.

We construct the DLF and DLT on $U_1, U_2, \ldots, U_{N(t)}$ as before. Let $\tilde{D}^1(\mathcal{P}_t^0) = D^1(\mathcal{P}_t^0) - E\left[D^1(\mathcal{P}_t^0)\right]$ and $\tilde{D}^1(\mathcal{P}_t) = D^1(\mathcal{P}_t) - E\left[D^1(\mathcal{P}_t)\right]$. We aim to show that the limit distribution for $\tilde{D}^1(\mathcal{P}_t^0)$ is the same as for $\tilde{D}^1(\mathcal{U}_n^0)$, and likewise in the unrooted case. We shall need the following result.

Lemma 5.5.6 *As $t \to \infty$,*

$$\frac{\mathrm{d}}{\mathrm{d}t}E[D^1(\mathcal{P}_t)] = \frac{1}{t} + O(t^{-2}); \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}E[D^1(\mathcal{P}_t^0)] = \frac{1}{t} + O(t^{-2}). \tag{5.155}$$

Proof. The point set $\{U_1, \ldots, U_{N(t)}\}$ is a homogeneous Poisson point process in $(0,1)$, so we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}E[D^1(\mathcal{P}_t)] &= E[\text{length of new arrival}]\\ &= \int_0^1 \mathrm{d}uE[\text{dist. to next pt. to the left of } u \text{ in } \mathcal{P}_t]\\ &= \int_0^1 \mathrm{d}u \int_0^u ste^{-ts}\mathrm{d}s = \frac{1}{t} + \frac{2}{t^2}\left(e^{-t}-1\right) + \frac{e^{-t}}{t}\\ &= \frac{1}{t} + O\left(t^{-2}\right).\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}E[D^1(\mathcal{P}_t^0)] &= \int_0^1 \mathrm{d}uE[\text{dist. to next pt. to the left of } u \text{ in } \mathcal{P}_t \cup \{0\}]\\ &= \int_0^1 \mathrm{d}u \int_0^u P[\text{dist. to next pt. to the left} > s]\mathrm{d}s\\ &= \int_0^1 \mathrm{d}u \int_0^u e^{-ts}\mathrm{d}s = \frac{1}{t} + \frac{e^{-t}-1}{t^2}\\ &= \frac{1}{t} + O\left(t^{-2}\right). \square\end{aligned}$$

Lemma 5.5.7 *(i) As $t \to \infty$, $\tilde{D}^1(\mathcal{P}_t^0)$ converges in distribution to $\tilde{D}_1$, the L^2 large-n limit of $\tilde{D}^1(\mathcal{U}_n^0)$.*

(ii) As $t \to \infty$, $\tilde{D}^1(\mathcal{P}_t)$ converges in distribution to $\tilde{F}_1$, the L^2 large-n limit of $\tilde{D}^1(\mathcal{U}_n)$.

Proof. (i) From Proposition 5.5.7, we have $\tilde{D}^1(\mathcal{U}_n^0) \stackrel{L^2}{\longrightarrow} \tilde{D}_1$ as $n \to \infty$. Let $a_t := E[D^1(\mathcal{P}_t^0)]$ and $\mu_n := E[D^1(\mathcal{U}_n^0)]$. Since $\mu_n = E\sum_{i=1}^n Z_i = \sum_{i=1}^n (1+i)^{-1}$ by (5.134), for any positive integers ℓ, n we have

$$|\mu_n - \mu_\ell| = \sum_{j=\min(n,\ell)+1}^{\max(n,\ell)} \frac{1}{j+1} \leq \log\left(\frac{\max(n,\ell)+1}{\min(n,\ell)+1}\right) = \left|\log\left(\frac{n+1}{\ell+1}\right)\right|. \tag{5.156}$$

Note the distributional equalities

$$\begin{aligned}\mathcal{L}\left(D^1(\mathcal{P}_t^0)|N(t)=n\right) &= \mathcal{L}\left(D^1(\mathcal{U}_n^0)\right);\\ \mathcal{L}\left(D^1(\mathcal{P}_t^0) - \mu_{N(t)}|N(t)=n\right) &= \mathcal{L}\left(\tilde{D}^1(\mathcal{U}_n^0)\right).\end{aligned} \tag{5.157}$$

First we aim to show that $a_t - \mu_{\lfloor t \rfloor} \to 0$ as $t \to \infty$. Set $p_m(t) := e^{-t}\frac{t^m}{m!}$. Then we can write

$$\begin{aligned}a_t - \mu_{\lfloor t \rfloor} &= \sum_{m=0}^{\infty} p_m(t)(\mu_m - \mu_{\lfloor t \rfloor})\\ &= \sum_{|m-\lfloor t \rfloor| \leq t^{3/4}} p_m(t)(\mu_m - \mu_{\lfloor t \rfloor}) + \sum_{|m-\lfloor t \rfloor| > t^{3/4}} p_m(t)(\mu_m - \mu_{\lfloor t \rfloor}).\end{aligned} \tag{5.158}$$

We examine these two sums separately. First consider the sum for $|m - \lfloor t \rfloor| \le t^{3/4}$. By (5.156), we have

$$\sup_{m:|m-\lfloor t\rfloor|\le t^{3/4}} \left|\mu_m - \mu_{\lfloor t \rfloor}\right| \le \max\left(\log\left(\frac{\lfloor t\rfloor + 1 + t^{3/4}}{\lfloor t \rfloor + 1}\right), \log\left(\frac{\lfloor t \rfloor + 1}{\lfloor t \rfloor + 1 - t^{3/4}}\right)\right)$$
$$= O\left(t^{-1/4}\right) \to 0 \text{ as } t \to \infty.$$

Hence the first sum in (5.158) tends to zero as $t \to \infty$. To estimate the second sum, observe that

$$\begin{aligned}\sum_{|m-\lfloor t\rfloor|>t^{3/4}} p_m(t)(\mu_m - \mu_{\lfloor t \rfloor}) &\le \sum_{|m-\lfloor t\rfloor|>t^{3/4}} p_m(t)(m+t) \\ &= E\left[(N(t)+t)\mathbf{1}_{\{|N(t)-\lfloor t\rfloor|>t^{3/4}\}}\right] \\ &\le \left(E\left[(N(t)+t)^2\right] P\left[|N(t) - \lfloor t \rfloor| > t^{3/4}\right]\right)^{1/2}. \qquad (5.159)\end{aligned}$$

By Chernoff bounds on the tail probabilities of a Poisson random variable (e.g. Lemma 1.4 of [104]), the expression (5.159) is $O(t\exp(-t^2/18))$ and so tends to zero. Hence the second sum in (5.158) tends to zero, and thus

$$a_t - \mu_{\lfloor t \rfloor} \to 0 \text{ as } t \to \infty. \qquad (5.160)$$

Now we show that $\tilde{D}^1(\mathcal{P}_t^0) \xrightarrow{\mathcal{D}} \tilde{D}_1$ as $t \to \infty$. We have

$$\tilde{D}^1(\mathcal{P}_t^0) = \left(D^1(\mathcal{P}_t^0) - \mu_{N(t)}\right) + \left(\mu_{N(t)} - \mu_{\lfloor t \rfloor}\right) + \left(\mu_{\lfloor t \rfloor} - a_t\right). \qquad (5.161)$$

The final bracket tends to zero, by (5.160). Also, by (5.157) and the fact that $N(t) \to \infty$ a.s. as $t \to \infty$, we have

$$D^1(\mathcal{P}_t^0) - \mu_{N(t)} \xrightarrow{\mathcal{D}} \tilde{D}_1.$$

Finally, using (5.156), we have

$$\left|\mu_{N(t)} - \mu_{\lfloor t \rfloor}\right| \le \left|\log \frac{N(t)+1}{\lfloor t \rfloor + 1}\right| \xrightarrow{P} 0,$$

as $t \to \infty$, since $N(t)/\lfloor t \rfloor \xrightarrow{P} 1$. So Slutsky's theorem applied to (5.161) yields $\tilde{D}^1(\mathcal{P}_t^0) \xrightarrow{\mathcal{D}} \tilde{D}_1$ as $t \to \infty$, completing the proof of (i)

The proof of (ii) follows in the same way as that of (i), except that in (5.156) the first equals sign is replaced by an inequality $\le$. This does not affect the rest of the proof. □

Proposition 5.5.8 *The limiting random variable $\tilde{D}_1$ of Proposition 5.5.7 (i) satisfies the fixed-point equation (5.36).*

Proof. For integer $n > 0$, let $T_n := \min\{s : N(s) \geq n\}$, the nth arrival time of the Poisson process with counting process $N(\cdot)$. Set $T := T_1$, and set $U := U_1$ (which is uniform on $(0,1)$).

By the Marking Theorem for Poisson processes [84], the two-dimensional point process $\mathcal{Q} := \{(U_n, T_n) : n \geq 1\}$ is a homogeneous Poisson process of unit intensity on $(0,1) \times (0,\infty)$. Given the value of (U,T), the restriction of $\mathcal{Q}$ to $(0,U] \times (T,\infty)$ and the restriction of $\mathcal{Q}$ to $(U,1) \times (T,\infty)$ are independent homogeneous Poisson processes on these regions. Hence, by scaling properties of the Poisson process (see the Mapping Theorem in [84]) and of the DLT, writing $D^1_{\{i\}}(\cdot)$, $i = 1,2$ for independent copies of $D^1(\cdot)$, we have

$$D^1(\mathcal{P}^0_t) \stackrel{\mathcal{D}}{=} \left(UD^1_{\{1\}}(\mathcal{P}^0_{U(t-T)}) + (1-U)D^1_{\{2\}}(\mathcal{P}^0_{(1-U)(t-T)}) + U\right) \mathbf{1}_{\{t>T\}}. \tag{5.162}$$

Let $a_s = 0$ for $s \leq 0$, and $a_s = E[D^1(\mathcal{P}^0_s)]$ for $s > 0$. Then $\tilde{D}^1(\mathcal{P}^0_t) = D^1(\mathcal{P}^0_t) - a_t$, so that by (5.162),

$$\begin{aligned}\tilde{D}^1(\mathcal{P}^0_t) &\stackrel{\mathcal{D}}{=} \left(U\tilde{D}^1_{\{1\}}(\mathcal{P}^0_{U(t-T)}) + (1-U)\tilde{D}^1_{\{2\}}(\mathcal{P}^0_{(1-U)(t-T)}) + U\right) \mathbf{1}_{\{t>T\}} \\ &\quad + U\left(a_{U(t-T)} - a_t\right) + (1-U)\left(a_{(1-U)(t-T)} - a_t\right). \end{aligned} \tag{5.163}$$

From Lemma 5.5.6 we have $\frac{da_t}{dt} = \frac{1}{t} + O(t^{-2})$. Hence, if $T < t$, then

$$a_t - a_{U(t-T)} = \int_{U(t-T)}^t \frac{da_s}{ds} ds = \log t - \log\{U(t-T)\} + O\left((U(t-T))^{-1}\right),$$

and hence as $t \to \infty$,

$$a_t - a_{U(t-T)} \to -\log U, \quad \text{a.s.}. \tag{5.164}$$

Since $P[T < t]$ tends to 1, by making $t \to \infty$ in (5.163) and using Slutsky's theorem we obtain (5.36). □

Proposition 5.5.9 *The limiting random variable $\tilde{F}_1$ of Proposition 5.5.7 (ii) satisfies the fixed-point equation (5.36), and so has the same distribution as $\tilde{D}_1$. Also,* $\mathrm{Cov}(\tilde{F}_1, \tilde{D}_1) = (7/4) - \pi^2/6$.

Proof. The proof follows similar lines to that of Proposition 5.5.8. Once more let $a_s = E[D^1(\mathcal{P}^0_s)]$, for $s \geq 0$, and $a_s = 0$ for $s < 0$. Let $b_s = E[D^1(\mathcal{P}_s)]$ for $s > 0$, and $b_s = 0$ for $s \leq 0$, and let $T := \min\{t : N(t) \geq 1\}$, Then

$$D^1(\mathcal{P}_t) \stackrel{\mathcal{D}}{=} \left(UD^1_{\{1\}}(\mathcal{P}_{U(t-T)}) + (1-U)D^1_{\{2\}}(\mathcal{P}^0_{(1-U)(t-T)})\right) \mathbf{1}_{\{t>T\}}, \tag{5.165}$$

where $D^1_{\{1\}}(\cdot)$ and $D^1_{\{2\}}(\cdot)$ are independent copies of $D^1(\cdot)$. Then $\tilde{D}^1(\mathcal{P}_t) = D^1(\mathcal{P}_t) - b_t$ and $\tilde{D}^1(\mathcal{P}^0_t) = D^1(\mathcal{P}^0_t) - a_t$, so that (5.165) yields

$$\begin{aligned}\tilde{D}^1(\mathcal{P}_t) \overset{\mathcal{D}}{=} & \left(U\tilde{D}^1_{\{1\}}(\mathcal{P}_{U(t-T)}) + (1-U)\tilde{D}^1_{\{2\}}(\mathcal{P}^0_{(1-U)(t-T)})\right)\mathbf{1}_{\{t>T\}} \\ & + U\left(b_{U(t-T)} - b_t\right) + (1-U)\left(a_{(1-U)(t-T)} - b_t\right). \end{aligned} \qquad (5.166)$$

From Lemma 5.5.6 we have $\frac{db_t}{dt} = \frac{1}{t} + O(t^{-2})$. Hence, by the same argument as used at (5.164),

$$b_t - b_{U(t-T)} \to -\log U \quad \text{a.s.}$$

Also, $a_t - b_t = E[D^1_0(\mathcal{P}^0_t)]$ by (5.141), so that $\lim_{t\to\infty}(a_t - b_t) = 1$, by Lemma 5.5.2 and the fact that $E[D^1_0] = 1$, by (C.4.9). Using also (5.164) we find that as $t \to \infty$,

$$a_{(1-U)(t-T)} - b_t = (a_{(1-U)(t-T)} - a_t) + (a_t - b_t) \to 1 + \log(1-U), \qquad \text{a.s.}$$

Taking $t \to \infty$ in (5.166), and using Slutsky's theorem, we obtain

$$\tilde{F}_1 \overset{\mathcal{D}}{=} U\tilde{F}_1 + (1-U)\tilde{D}_1 + U\log U + (1-U)\log(1-U) + (1-U). \qquad (5.167)$$

The change of variable $(1-U) \mapsto U$ then shows that $\tilde{D}_1$ as defined at (5.36) satisfies (5.167), and so by the uniqueness of solution, $\tilde{F}_1$ has the same distribution as $\tilde{D}_1$ and satisfies (5.36).

To obtain the covariance of $\tilde{F}_1$ and $\tilde{D}_1$, observe from Proposition 5.5.7 (ii) that $D^1_0 = \tilde{D}_1 - \tilde{F}_1 + 1$, and therefore by (C.4.9), we have that

$$1/2 = \text{Var}[D^1_0] = \text{Var}[\tilde{D}_1] + \text{Var}[\tilde{F}_1] - 2\text{Cov}(\tilde{D}_1, \tilde{F}_1). \qquad (5.168)$$

Since $\text{Var}[\tilde{F}_1] = \text{Var}[\tilde{D}_1] = 2 - \pi^2/6$ by Proposition 5.5.7 (i), rearranging (5.168) we find that $\text{Cov}(\tilde{D}_1, \tilde{F}_1) = (7/4) - \pi^2/6$. □

Remark. One may obtain convergence in distribution results along the lines of Propositions 5.5.8 and 5.5.9 using the contraction method as in Proposition 5.4.2. In this way, one can also obtain joint convergence results for $(\tilde{D}^\alpha(\mathcal{U}^0_n), \tilde{D}^\alpha(\mathcal{U}_n))$.

Remark. Figure 5.2 is a plot of the estimated probability density function of $\tilde{D}_1$. This was obtained by performing 10^6 repeated simulations of the DLT on a sequence of 10^3 uniform (simulated) random points on $(0,1)$. For each simulation, the expected value of $D^1(\mathcal{U}_{10^3})$ (which is precisely $(1/2)+(1/3)+\cdots(1/1001)$ by Lemma 5.5.1) was subtracted from the total length of the simulated DLT to give an approximate realization of $\tilde{D}_1$.

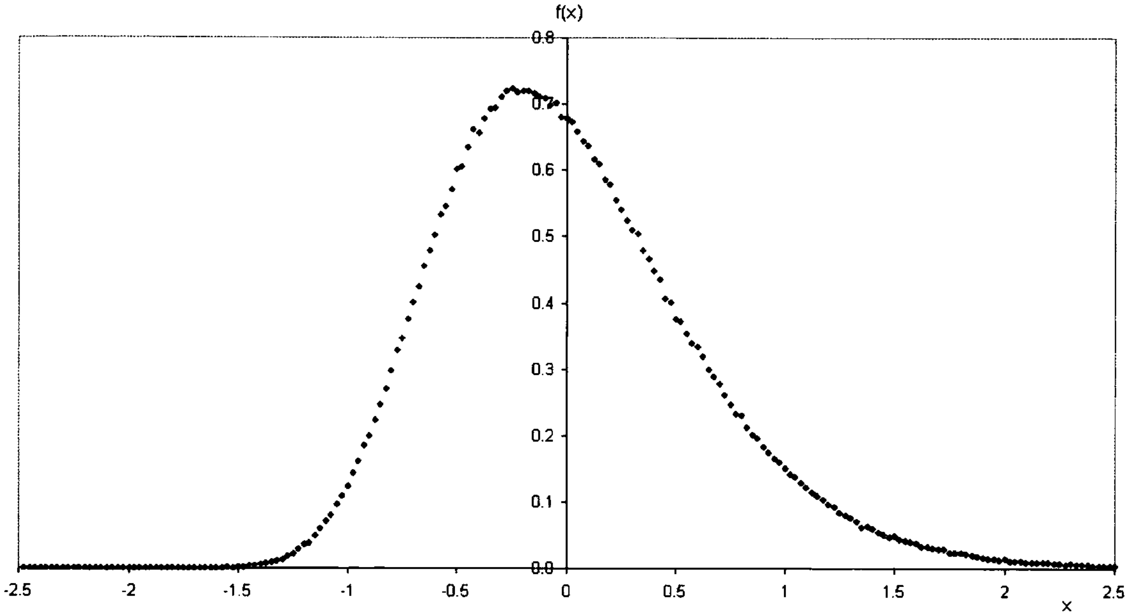


Figure 5.2: Estimated probability density function for $\tilde{D}_1$.

The density function was then estimated from the sample of 10^6 approximate realizations of $\tilde{D}_1$, using a window width of 0.0025. The simulated sample from which the density estimate for $\tilde{D}_1$ was taken had sample mean $\approx -2 \times 10^{-4}$ and sample variance ≈ 0.3543, which are reasonably close to the expectation and variance of $\tilde{D}_1$.

Chapter 6

The total length of the minimal directed spanning tree

6.1 Introduction and main results

In this chapter we give weak convergence results for the total weight of the minimal directed spanning tree/forest (suitably centred and scaled) on random points in $(0,1)^2$. Our results concern the two special partial orders $\preccurlyeq^*$ and $\preccurlyeq_*$, as given in Section 2.4.1.

Recall that in the MDST introduced by Bhatt and Roy in [21], the partial order is $\preccurlyeq^*$. Here, each point $\mathbf{x}$ of a finite (random) subset $\mathcal{S}$ of $(0,1)^2$ is connected by a directed edge to the nearest $\mathbf{y} \in \mathcal{S} \cup \{\mathbf{0}\}$ such that $\mathbf{y} \neq \mathbf{x}$ and $\mathbf{y} \preccurlyeq^* \mathbf{x}$, where $\mathbf{y} \preccurlyeq^* \mathbf{x}$ means that each component of $\mathbf{x} - \mathbf{y}$ is nonnegative. We refer to $\preccurlyeq^*$ as the "coordinate-wise" partial order. The second partial order we consider, $\preccurlyeq_*$, is the so-called "one coordinate" partial order. In $d = 2$, we take this to be such that $\mathbf{y} \preccurlyeq_* \mathbf{x}$ means that the *second* (i.e. vertical) component of $\mathbf{x} - \mathbf{y}$ is nonnegative.

In [21], Bhatt and Roy mention that the total length of the MDST under $\preccurlyeq^*$ is an object of considerable interest, although they restrict their analysis to the length of the edges joined to the origin (see also Chapter 4). A first order result for the total length of the MDST or MDSF is a law of large numbers; see Theorem 2.4.1 for a LLN for a family of MDSFs indexed by partial orderings on $\mathbf{R}^2$, which include $\preccurlyeq^*$ and $\preccurlyeq_*$ as special cases.

For the length of edges from points in the region away from the boundary, we prove a central limit theorem. The boundary effects are significant, and near the boundary the MDST can be described in terms of a one-dimensional, on-line version of the MDST; thus we will encounter the on-line nearest-neighbour graph (ONG) and directed linear tree

(DLT) which we studied in Chapter 5.

We consider power-weighted edges. Our weak convergence results (Theorem 6.1.1) demonstrate that, depending on the value chosen for the weight exponent of the edges, there are two regimes in which either the boundary effects dominate or those edges away from the boundary are dominant, and that there is a critical value, or phase transition, (when we take simple Euclidean length as the weight) for which neither effect dominates.

Our main result (Theorem 6.1.1) presents convergence in distribution for the cases where the partial order is $\preccurlyeq^*$ or $\preccurlyeq_*$; the limiting distributions are of a different type in the three cases $\alpha = 1$, $0 < \alpha < 1$, and $\alpha > 1$. We define these limiting distributions in Theorem 6.1.1, in terms of distributional fixed-point equations (see Section 5.1.2).

Recall from Section 2.4.2 that $\mathcal{L}^{d,\alpha}(\mathcal{S})$ denotes the total weight of the MDSF on $\mathcal{S} \subset (0,1)^d$ under weight function w_α, $\alpha > 0$, as given by (1.2), and that $\tilde{\mathcal{L}}^{d,\alpha}(\mathcal{S})$ denotes the centred version. For the remainder of this chapter, we take $d = 2$ and suppress this notation by setting $\mathcal{L}^{\alpha}(\cdot) := \mathcal{L}^{2,\alpha}(\cdot)$ and $\tilde{\mathcal{L}}^{\alpha}(\cdot) := \tilde{\mathcal{L}}^{2,\alpha}(\cdot)$.

For ease of notation, we define the random variables $\tilde{Z}_\alpha$, $\alpha \geq 1$, as follows. Set

$$\tilde{Z}_1 \stackrel{\mathcal{D}}{=} U\tilde{H}_1^{\{1\}} + (1-U)\tilde{H}_1^{\{2\}} + \frac{1}{2}U\log U + \frac{1}{2}(1-U)\log(1-U), \tag{6.1}$$

where U is uniform on $(0,1)$, and $\tilde{H}_1^{\{1\}}$, $\tilde{H}_1^{\{2\}}$, independent of U, are independent copies of the random variable $\tilde{H}_1$ with distribution given by (5.29). Also, for $\alpha > 1$, let

$$\tilde{Z}_\alpha \stackrel{\mathcal{D}}{=} U^\alpha\tilde{H}_\alpha^{\{1\}} + (1-U)^\alpha\tilde{H}_\alpha^{\{2\}} + \left(U^\alpha + (1-U)^\alpha - \frac{2}{1+\alpha}\right)\left(\frac{1}{\alpha} - \frac{2^{-\alpha}}{\alpha(1-\alpha)}\right), \tag{6.2}$$

where U is uniform on $(0,1)$, and $\tilde{H}_\alpha^{\{1\}}$, $\tilde{H}_\alpha^{\{2\}}$, independent of U, are independent copies of the random variable $\tilde{H}_\alpha$ with distribution given by (5.30).

Note that from the properties of $\tilde{H}_\alpha$, $\alpha \geq 1$ (see Section 5.4) it follows that $E[\tilde{Z}_\alpha] = 0$ for $\alpha \geq 1$ and

$$\mathrm{Var}[\tilde{Z}_1] = \frac{19}{48} + \frac{\log 2}{12} - \frac{\pi^2}{24} \approx 0.042362.$$

The main result of this chapter is as follows.

Theorem 6.1.1 *Suppose the weight exponent is $\alpha > 0$.*

(a) Suppose that the partial order is $\preccurlyeq^*$. Then there exist constants $0 < t_\alpha^2 \leq s_\alpha^2$ such that, for normal random variables $Y_\alpha \sim \mathcal{N}(0, s_\alpha^2)$ and $W_\alpha \sim \mathcal{N}(0, t_\alpha^2)$: as $n \to \infty$,

$$n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n^0) \xrightarrow{\mathcal{D}} Y_\alpha \quad \text{and} \quad n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n^0) \xrightarrow{\mathcal{D}} W_\alpha \ (0<\alpha<1); \tag{6.3}$$

$$\tilde{\mathcal{L}}^1(\mathcal{P}_n^0) \xrightarrow{\mathcal{D}} \tilde{D}_1^{\{1\}} + \tilde{D}_1^{\{2\}} + Y_1 \quad \text{and} \quad \tilde{\mathcal{L}}^1(\mathcal{U}_n^0) \xrightarrow{\mathcal{D}} \tilde{D}_1^{\{1\}} + \tilde{D}_1^{\{2\}} + W_1; \tag{6.4}$$

$$\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n^0) \xrightarrow{\mathcal{D}} \tilde{D}_\alpha^{\{1\}} + \tilde{D}_\alpha^{\{2\}} \quad \text{and} \quad \tilde{\mathcal{L}}^\alpha(\mathcal{U}_n^0) \xrightarrow{\mathcal{D}} \tilde{D}_\alpha^{\{1\}} + \tilde{D}_\alpha^{\{2\}} \ (\alpha > 1). \tag{6.5}$$

Here all the random variables in the limits are independent, and $\tilde{D}_\alpha^{\{i\}}$, $i = 1, 2$ are independent copies of the random variable $\tilde{D}_\alpha$ defined at (5.36) for $\alpha = 1$ and (5.37) for $\alpha > 1$. Also, as $n \to \infty$,

$$n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n) \xrightarrow{\mathcal{D}} Y_\alpha \quad \text{and} \quad n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n) \xrightarrow{\mathcal{D}} W_\alpha \ (0 < \alpha < 1); \tag{6.6}$$

$$\tilde{\mathcal{L}}^1(\mathcal{P}_n) \xrightarrow{\mathcal{D}} \tilde{D}_1^{\{1\}} + \tilde{D}_1^{\{2\}} + Y_1 \quad \text{and} \quad \tilde{\mathcal{L}}^1(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \tilde{D}_1^{\{1\}} + \tilde{D}_1^{\{2\}} + W_1; \tag{6.7}$$

$$\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha^{\{1\}} + \tilde{F}_\alpha^{\{2\}} \quad \text{and} \quad \tilde{\mathcal{L}}^\alpha(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha^{\{1\}} + \tilde{F}_\alpha^{\{2\}} \ (\alpha > 1) \ . \tag{6.8}$$

Here all the random variables in the limits are independent, and $\tilde{D}_1^{\{i\}}$, $i = 1, 2$, are independent copies of $\tilde{D}_1$ with distribution defined at (5.36), and for $\alpha > 1$, $\tilde{F}_\alpha^{\{i\}}$, $i = 1, 2$, are independent copies of $\tilde{F}_\alpha$ with distribution defined at (5.38).

(b) *Suppose that the partial order is $\preccurlyeq_*$. Then there exist constants $0 < t_\alpha^2 \leq s_\alpha^2$ such that, for normal random variables $Y_\alpha \sim \mathcal{N}(0, s_\alpha^2)$ and $W_\alpha \sim \mathcal{N}(0, t_\alpha^2)$: as $n \to \infty$,*

$$n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n) \xrightarrow{\mathcal{D}} Y_\alpha \quad \text{and} \quad n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n) \xrightarrow{\mathcal{D}} W_\alpha \ (0 < \alpha < 1); \tag{6.9}$$

$$\tilde{\mathcal{L}}^1(\mathcal{P}_n) \xrightarrow{\mathcal{D}} \tilde{Z}_1 + Y_1 \quad \text{and} \quad \tilde{\mathcal{L}}^1(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \tilde{Z}_1 + W_1; \tag{6.10}$$

$$\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n) \xrightarrow{\mathcal{D}} \tilde{Z}_\alpha \quad \text{and} \quad \tilde{\mathcal{L}}^\alpha(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \tilde{Z}_\alpha \ (\alpha > 1). \tag{6.11}$$

Here all the random variables in the limits are independent, and $\tilde{Z}_\alpha$, $\alpha \geq 1$ has the distribution given by (6.1) for $\alpha = 1$ and (6.2) for $\alpha > 1$.

Remarks. The normal random variables Y_α or W_α arise from the edges away from the boundary (see Section 6.2). The non-normal variables (the $\tilde{D}$s, $\tilde{F}$s and $\tilde{Z}$s) arise from the edges very close to the boundary, where the MDSF is asymptotically close to the directed linear forest (in the $\preccurlyeq^*$ case) or the on-line nearest-neighbour graph (in the $\preccurlyeq_*$ case), as discussed in Chapter 5. See Section 6.3.

Theorem 6.1.1 indicates a phase transition in the character of the limit law as α increases. The normal contribution (from the points away from the boundary) dominates for $0 < \alpha < 1$, while the boundary contributions dominate for $\alpha > 1$. In the critical case $\alpha = 1$, neither effect dominates and both terms contribute significantly to the asymptotic behaviour.

Noteworthy, under $\preccurlyeq^*$, in the case $\alpha = 1$ is the fact that by (6.4) and (6.7), the limiting distribution is the same for $\tilde{\mathcal{L}}^1(\mathcal{P}_n)$ as for $\tilde{\mathcal{L}}^1(\mathcal{P}_n^0)$, and the same for $\tilde{\mathcal{L}}^1(\mathcal{U}_n)$ as for $\tilde{\mathcal{L}}^1(\mathcal{U}_n^0)$. Note, however, that the difference $\tilde{\mathcal{L}}^1(\mathcal{P}_n) - \tilde{\mathcal{L}}^1(\mathcal{P}_n^0)$ is the (centred) total length of edges incident to the origin, which is not negligible, but itself converges in distribution (see Theorem 4.2.1, or [108]) to a non-degenerate random variable, namely a centred generalized

Dickman random variable with parameter 2 (see Chapter 4). As an extension of Theorem 6.1.1, it should be possible to show that the joint distribution of $(\tilde{\mathcal{L}}^1(\mathcal{P}_n), \tilde{\mathcal{L}}^1(\mathcal{P}_n^0))$ converges to that of two coupled random variables, both having the distribution of $\tilde{D}_1$, whose difference has the centred generalized Dickman distribution with parameter 2. Likewise for the joint distribution of $(\tilde{\mathcal{L}}^1(\mathcal{U}_n), \tilde{\mathcal{L}}^1(\mathcal{U}_n^0))$.

The remainder of this chapter is organized as follows. The proof of Theorem 6.1.1 is prepared in Sections 6.2 and 6.3, and completed in Section 6.4. In these proofs, we repeatedly use Slutsky's theorem (see Lemma A.2.1). The material in Section 6.2 draws on Chapter 3, while that in Section 6.3 draws on Chapter 5.

6.2 Central limit theorem away from the boundary

For each $n \in \mathbf{N}$, define the region $S_{0,n} := (n^{\varepsilon-1/2}, 1)^2$, where $\varepsilon \in (0, 1/2)$ is a small constant to be chosen later. Similarly, set $S'_{0,n} := (0,1) \times (n^{\varepsilon-1/2}, 1)$.

In this section, we use the general central limit theorems of Section 3.2 to demonstrate a central limit theorem for the contribution to the total weight of the MDSF, under $\preccurlyeq^*$, from edges away from the boundary, that is from points in the region $S_{0,n}$. We state the analogous result for $\preccurlyeq_*$ and the region $S'_{0,n}$, but do not go into details with the proof – it is very similar to the $\preccurlyeq^*$ case.

Given $\alpha > 0$, consider the MDSF total weight functional $H = \mathcal{L}^\alpha$ on point sets in $\mathbf{R}^2$. For $\mathbf{x} \in \mathcal{X}$, let the directed nearest neighbour distance $d(\mathbf{x}; \mathcal{X})$ and the corresponding α-weighted functional $\xi(\mathbf{x}; \mathcal{X})$ be given by (3.20), where now we take $\overset{\theta,\phi}{\preccurlyeq}$ to be $\preccurlyeq^*$. For $R \subseteq \mathbf{R}^2$, set

$$\mathcal{L}^\alpha(\mathcal{X}; R) = \sum_{\mathbf{x} \in \mathcal{X} \cap R} \xi(\mathbf{x}; \mathcal{X}), \tag{6.12}$$

and set $\mathcal{L}^\alpha(\mathcal{X}) := \mathcal{L}^\alpha(\mathcal{X}; \mathbf{R}^2)$. The main result of this section is the following.

Theorem 6.2.1 *Suppose that $\alpha > 0$.*

(a) Suppose that the partial order is $\preccurlyeq^*$. Then there exist constants $0 < t_\alpha \leq s_\alpha$, not depending on the choice of ε, such that, as $n \to \infty$,

(i) $n^{\alpha-1}\mathrm{Var}\left[\mathcal{L}^\alpha\left(\mathcal{U}_n; S_{0,n}\right)\right] \to t_\alpha^2$;

(ii) $n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha\left(\mathcal{U}_n; S_{0,n}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, t_\alpha^2\right)$;

(iii) $n^{\alpha-1}\mathrm{Var}\left[\mathcal{L}^{\alpha}\left(\mathcal{P}_n; S_{0,n}\right)\right] \to s_{\alpha}^2$;

(iv) $n^{(\alpha-1)/2}\tilde{\mathcal{L}}^{\alpha}\left(\mathcal{P}_n; S_{0,n}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, s_{\alpha}^2\right)$.

(b) Suppose that the partial order is $\preccurlyeq_*$. Then there exist constants $0 < t_\alpha \leq s_\alpha$, not depending on the choice of ε, such that, as $n \to \infty$,

(i) $n^{\alpha-1}\mathrm{Var}\left[\mathcal{L}^{\alpha}\left(\mathcal{U}_n; S'_{0,n}\right)\right] \to t_{\alpha}^2$;

(ii) $n^{(\alpha-1)/2}\tilde{\mathcal{L}}^{\alpha}\left(\mathcal{U}_n; S'_{0,n}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, t_{\alpha}^2\right)$;

(iii) $n^{\alpha-1}\mathrm{Var}\left[\mathcal{L}^{\alpha}\left(\mathcal{P}_n; S'_{0,n}\right)\right] \to s_{\alpha}^2$;

(iv) $n^{(\alpha-1)/2}\tilde{\mathcal{L}}^{\alpha}\left(\mathcal{P}_n; S'_{0,n}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, s_{\alpha}^2\right)$.

The following corollary states that Theorem 6.2.1 (a) remains true in the rooted cases too, i.e. with $\mathcal{U}_n$ replaced by $\mathcal{U}_n^0$ and $\mathcal{P}_n$ replaced by $\mathcal{P}_n^0$.

Corollary 6.2.1 *Suppose that $\alpha > 0$ and the partial order is $\preccurlyeq^*$. Then, with t_α, s_α as given in Theorem 6.2.1, we have that as $n \to \infty$,*

(i) $n^{\alpha-1}\mathrm{Var}\left[\mathcal{L}^{\alpha}\left(\mathcal{U}_n^0; S_{0,n}\right)\right] \to t_{\alpha}^2$;

(ii) $n^{(\alpha-1)/2}\tilde{\mathcal{L}}^{\alpha}\left(\mathcal{U}_n^0; S_{0,n}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, t_{\alpha}^2\right)$;

(iii) $n^{\alpha-1}\mathrm{Var}\left[\mathcal{L}^{\alpha}\left(\mathcal{P}_n^0; S_{0,n}\right)\right] \to s_{\alpha}^2$;

(iv) $n^{(\alpha-1)/2}\tilde{\mathcal{L}}^{\alpha}\left(\mathcal{P}_n^0; S_{0,n}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, s_{\alpha}^2\right)$.

Proof. For each region $R \subseteq [0,1)^2$ and point set $\mathcal{S} \subset [0,1)^2$ with $\mathbf{0} \in \mathcal{S}$, let $\mathcal{L}_0^{\alpha}(\mathcal{S}; R)$ denote the total weight of the edges incident to $\mathbf{0}$ in the MDST on $\mathcal{S}$ from points in R. Then $\mathcal{L}^{\alpha}(\mathcal{P}_n^0; S_{0,n})$ equals $\mathcal{L}^{\alpha}(\mathcal{P}_n; S_{0,n}) + \mathcal{L}_0^{\alpha}(\mathcal{P}_n^0; S_{0,n})$, so that

$$\begin{aligned} & \mathrm{Var}[\mathcal{L}^{\alpha}(\mathcal{P}_n^0; S_{0,n})] - \mathrm{Var}[\mathcal{L}^{\alpha}(\mathcal{P}_n; S_{0,n})] \\ = \; & 2\mathrm{Cov}[\mathcal{L}^{\alpha}(\mathcal{P}_n; S_{0,n}), \mathcal{L}_0^{\alpha}(\mathcal{P}_n^0; S_{0,n})] + \mathrm{Var}[\mathcal{L}_0^{\alpha}(\mathcal{P}_n^0; S_{0,n})]. \end{aligned} \tag{6.13}$$

Let N_n denote the number of points of $\mathcal{P}_n$, and let E_n denote the event that at least one point of $\mathcal{P}_n \cap S_{0,n}$ is joined to $\mathbf{0}$ in the MDST on $\mathcal{P}_n^0$. Then

$$P[E_n] \leq P\left[(0, n^{\varepsilon-1/2}]^2 \cap \mathcal{P}_n = \emptyset\right] = \exp(-n^{2\varepsilon}),$$

and $\mathcal{L}_0^{\alpha}(\mathcal{P}_n^0; S_{0,n}) \leq 2^{\alpha/2} N_n \mathbf{1}_{E_n}$. Thus by the Cauchy-Schwarz inequality, for some finite constant C we have

$$\mathrm{Var}\left[\mathcal{L}_0^{\alpha}(\mathcal{P}_n^0; S_{0,n})\right] \leq E\left[\mathcal{L}_0^{\alpha}(\mathcal{P}_n^0; S_{0,n})^2\right] \leq Cn^2 \exp(-n^{2\varepsilon}/2), \tag{6.14}$$

and combining this with (6.13), Theorem 6.2.1 (iii) and the Cauchy-Schwarz inequality shows that

$$n^{\alpha-1}(\mathrm{Var}\left[\mathcal{L}^{\alpha}(\mathcal{P}_n^0; S_{0,n})\right] - \mathrm{Var}\left[\mathcal{L}^{\alpha}(\mathcal{P}_n; S_{0,n})\right]) \to 0,$$

so that from Theorem 6.2.1 (iii) we obtain the corresponding rooted result (iii). Also, since (6.14) implies $n^{\alpha-1}\mathrm{Var}\left[\mathcal{L}_0^{\alpha}(\mathcal{P}_n^0; S_{0,n})\right]$ tends to zero, from Theorem 6.2.1 (iv) and Slutsky's theorem we obtain the corresponding rooted result (iv).

The binomial results (i) and (ii) follow in the same manner as above, with slight modifications. □

As remarked above, we give the proof of Theorem 6.2.1 (a) only; the proof of part (b) is analogous (and often simpler, in fact). To prove Theorem 6.2.1, we demonstrate that our functional $\mathcal{L}^{\alpha}$ satisfies suitable versions of the conditions of Theorem 3.2.1 and Corollary 3.2.1. First, we see that $\mathcal{L}^{\alpha}$ is polynomially bounded (see (3.4)), since

$$\mathcal{L}^{\alpha}(\mathcal{X}; B) \leq (\mathrm{diam}(\mathcal{X}))^{\alpha}\,\mathrm{card}(\mathcal{X}).$$

Also, $\mathcal{L}^{\alpha}$ is homogeneous of order α.

Lemma 6.2.1 *Suppose the partial order is $\preccurlyeq^*$, and $\alpha > 0$. Then $\mathcal{L}^{\alpha}$ is strongly stabilizing, in the sense of Definition 3.2.1.*

Proof. To prove stabilization it is sufficient to show that there exists an almost surely finite random variable R, the radius of stabilization, such that the add one cost is unaffected by changes in the configuration at a distance greater than R from the added point. We show that there exists such an R.

For $s > 0$ construct eight disjoint triangles $T_j(s), 1 \leq j \leq 8$, by splitting the square $Q(\mathbf{0}; s)$ into eight triangles via drawing in the diagonals of the square and the x and y axes. Label the triangle with vertices $(0,0), (0,s), (s,s)$ as $T_1(s)$ and then label increasingly in a clockwise manner. See Figure 6.1. Note that $T_j(t) \subset T_j(s)$ for $t < s$. Let the random variable S be the minimum s such that the triangles $T_j(s), 1 \leq j \leq 8$, each contain at least one point of $\mathcal{P}$. Then S is almost surely finite.

We claim that $R = 3S$ is a radius of stabilization for $\mathcal{L}^{\alpha}$, that is any points at distance $d \geq 3S$ from the origin have no impact on the set of added or removed edges when a point is inserted at the origin.

First, $\mathbf{0}$ can have no point at a distance of at least $3S$ away as its directed nearest neighbour, since there will be points in T_5 and T_6 within a distance of at most $\sqrt{2}S$ of $\mathbf{0}$.

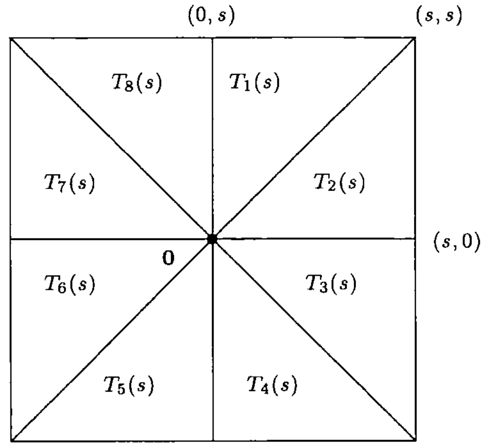


Figure 6.1: The triangles $T_1(s), \dots, T_8(s)$, $s > 0$.

We now need to show that no point at a distance at least $3S$ from $\mathbf{0}$ can have the origin as its directed nearest neighbour. Clearly, for the partial order $\preccurlyeq^*$, we need only consider points in the region $(0, \infty)^2$.

Consider a point (x, y) in the first quadrant, such that $\|(x, y)\| \geq 3S$. Consider the disk sector

$$D_{(x,y)} := B\left((x, y), \|(x, y)\|\right) \cap \{\mathbf{w} : \mathbf{w} \preccurlyeq^* (x, y)\}.$$

We aim to show that given any (x, y) of the above form, at least one of the $T_j(S)$, $j = 1, \dots, 8$, is contained in $D_{(x,y)}$, which implies that the origin cannot be the directed nearest neighbour of (x, y). To demonstrate this, we show that given such an (x, y), $D_{(x,y)}$ contains all three vertices of at least one of the $T_j(S)$.

First suppose $x > S$, $y > S$. Then we have that $T_1(S)$ and $T_2(S)$ are in $D_{(x,y)}$, since we have, for example,

$$\begin{aligned} \|(x, y) - \mathbf{0}\|^2 - \|(x, y) - (0, S)\|^2 &= \left(x^2 + y^2\right) - \left(x^2 + (y - S)^2\right) \\ &= S(2y - S) > 0. \end{aligned}$$

By symmetry, the only other situation we need consider is when $0 < x \leq S$. Then $y^2 \geq 9S^2 - x^2 \geq 8S^2$, so $y \geq 2\sqrt{2}S$. Then we have that $T_8(S)$ is in $D_{(x,y)}$, since

$$\begin{aligned} \|(x, y) - \mathbf{0}\|^2 - \|(x, y) - (-S, S)\|^2 &= \left(x^2 + y^2\right) - \left((x + S)^2 + (y - S)^2\right) \\ &= 2S(y - x - S) \geq 4S^2(\sqrt{2} - 1) > 0. \end{aligned}$$

This completes the proof. □

Lemma 6.2.2 *Suppose the partial order is $\preccurlyeq^*$ and $\alpha > 0$. Then the distribution of $\Delta(\infty)$ is non-degenerate.*

Proof. We demonstrate the existence of two configurations that occur with strictly positive probability and give rise to different values for $\Delta(\infty)$. Note that adding a point at the origin causes some new edges to be formed (namely those incident to the origin), and the possible deletion of some edges (namely the edges from points which have the origin as their directed nearest neighbour after its insertion).

Let $\eta > 0$, with $\eta < 1/3$. Later we shall impose further conditions on η. Again we refer to the construction in Figure 6.1. Let E_1 denote the event that for each i, $1 \leq i \leq 8$, there is a single point of $\mathcal{P}$, denoted $\mathbf{W}_i$, in each of $T_i(\eta)$, and that there are no other points in $[-1,1)^2$. Suppose that E_1 occurs. Then, on addition of the origin, the only edges that can possibly be removed are those from $\mathbf{W}_1$ and from $\mathbf{W}_2$ (see the proof of Lemma 6.2.1). These removed edges have length at most $\eta\sqrt{8}$, and hence

$$\Delta \geq -2(\eta\sqrt{8})^{\alpha} := \delta_1, \qquad \text{on } E_1. \tag{6.15}$$

Now let E_2 denote the event that there is a single point of $\mathcal{P}$, denoted $\mathbf{Z}_1$, in the square $(\eta, 2\eta) \times (0, \eta)$, a single point denoted $\mathbf{Z}_2$ in the square $(0, \eta) \times (\eta, 2\eta)$, a single point denoted $\mathbf{W}$ in the square $(-1-\eta, -1) \times (-\eta, 0)$, and no other point in $[-3, 3]^2$. See Figure 6.2.

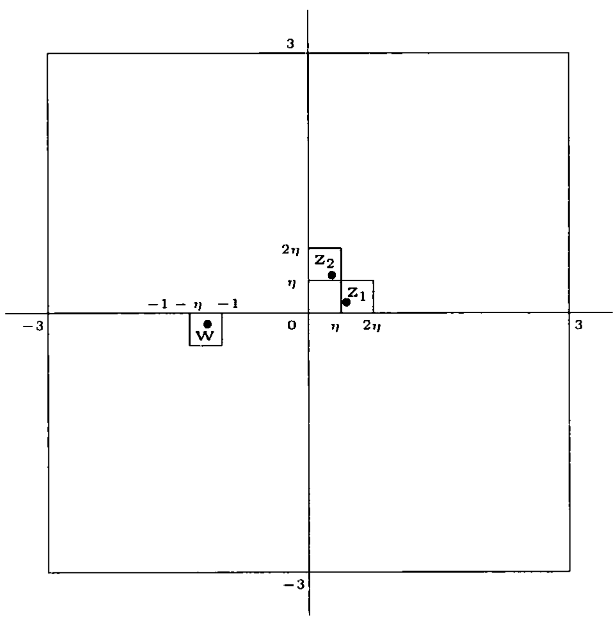


Figure 6.2: A possible configuration for event E_2.

Suppose that E_2 occurs. Now, on addition of the origin, an edge of length at most $1 + 2\eta$ is added from the origin to $\mathbf{W}$. On the other hand, for $i = 1, 2$ the edge from $\mathbf{Z}_i$ to $\mathbf{W}$ (of length at least 1) is replaced by an edge from $\mathbf{Z}_i$ to the origin (of length at most

3η). It is also possible that some other edges from points outside $[-3,3]^2$ are replaced by shorter edges from these points to the origin. Combining the effect of all these additions and replacements of edges, we find that

$$\Delta \leq (1+2\eta)^\alpha + 2((3\eta)^\alpha - 1) := \delta_2, \quad \text{on } E_2. \tag{6.16}$$

Given α, by taking η small enough we can arrange that $\delta_1 > -1/4$ and $\delta_2 < -3/4$. With such a choice of η, events E_1 and E_2 both have strictly positive probability which shows that the distribution of Δ is non-degenerate. □

For the next lemma, we set $R_0 := (0,1)^2$, recalling that $S_{0,n} := (n^{\varepsilon-1/2}, 1)^2$ throughout this section, and let $\mathcal{R}_0$ be as defined just before Corollary 3.2.1.

Lemma 6.2.3 *Suppose the partial order is $\preccurlyeq^*$ and $\alpha > 0$. Then $\mathcal{L}^\alpha$ satisfies the uniform bounded moments condition (3.5) on $\mathcal{R}_0$.*

Proof. Choose some $(A,B) \in \mathcal{R}_0$ such that $\mathbf{0} \in A$, i.e., such that for some $n \in \mathbf{N}$ the set A is a translate of $(0, n^{1/2})^2$ containing the origin and B is the corresponding translate of $n^{1/2}S_{0,n} = (n^\varepsilon, n^{1/2})^2$. Note that $|A| = n$, and choose $m \in [n/2, 3n/2]$.

Denote the m independent random vectors on A comprising $\mathcal{U}_{m,A}$ by $\mathbf{V}_1, \ldots, \mathbf{V}_m$. For contributions to $\Delta(\mathcal{U}_{m,A}; B)$ we are only interested in edges from points in the region B away from the boundary of A, although the origin can be inserted anywhere in A. Contributions to $\Delta(\mathcal{U}_{m,A}; B)$ come from the edges that are added or deleted on the addition of $\mathbf{0}$. We split $\Delta(\mathcal{U}_{m,A}; B)$ into two parts: the positive contribution from added edges, $\Delta^+(\mathcal{U}_{m,A}; B)$, and the negative contribution, $\Delta^-(\mathcal{U}_{m,A}; B)$, from removed edges.

By construction of the MDSF, the added edges are those that have $\mathbf{0}$ as an end-point after it has been inserted. Thus an upper bound on $\Delta^+(\mathcal{U}_{m,A}; B)$ is $L_{\max}^\alpha \delta(\mathbf{0}) + L_0^\alpha$, where $L_{\max}$ is the length of the longest edge from a point of $\mathcal{U}_{m,A} \cap B$ to $\mathbf{0}$, and $\delta(\mathbf{0})$ is the number of such edges (or zero if no such edge exists), and L_0 is the length of the edge from $\mathbf{0}$, or zero if no such edge exists.

For $\mathbf{w} \in A$ and $\mathbf{x} \in B$, with $\mathbf{w} \preccurlyeq^* \mathbf{x}$, define the region

$$R(\mathbf{w}, \mathbf{x}) := \{\mathbf{y} \in A : \mathbf{y} \preccurlyeq^* \mathbf{x}, \|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{w} - \mathbf{x}\|\}.$$

Since points in B are distant at least 1 from the lower or left boundary of A, by Lemma 3.3.6 there exists a constant $0 < C < \infty$ such that

$$|R(\mathbf{w}, \mathbf{x})| \geq C\|\mathbf{x} - \mathbf{w}\|, \text{ for all } \mathbf{w} \in A, \mathbf{x} \in B \text{ with } \mathbf{w} \preccurlyeq^* \mathbf{x} \text{ and } \|\mathbf{x} - \mathbf{w}\| \geq 1. \tag{6.17}$$

Suppose there is a point at $\mathbf{x}$ with $\mathbf{0} \preccurlyeq^* \mathbf{x}$. Then, the probability of the event $E(\mathbf{x})$ that $\mathbf{x}$ is joined to the origin in the MDSF on $\mathcal{U}_{m,A} \cup \{\mathbf{0}\}$ is

$$\begin{aligned} P[E(\mathbf{x})] &= P[R(\mathbf{0},\mathbf{x}) \text{ empty}] = \left(1 - \frac{|R(\mathbf{0},\mathbf{x})|}{|A|}\right)^{m-1} \\ &\leq \exp\left((1-m)\left(\frac{|R(\mathbf{0},\mathbf{x})|}{n}\right)\right) \leq \exp\left(1 - |R(\mathbf{0},\mathbf{x})|/2\right), \end{aligned} \quad (6.18)$$

since $m \geq n/2$ and $|R(\mathbf{0},\mathbf{x})| \leq n$.

We have that $L^{\alpha}_{\max}\delta(\mathbf{0}) \leq \max_{i=1,\ldots,m} W_i$, where

$$W_i = \|\mathbf{V}_i\|^{\alpha} \operatorname{card}(B(\mathbf{0};\|\mathbf{V}_i\|) \cap \mathcal{U}_{m,A} \cap \{\mathbf{y} : \mathbf{0} \preccurlyeq^* \mathbf{y}\})\ \mathbf{1}\{\mathbf{V}_i \text{ joined to } \mathbf{0} \text{ and } \mathbf{V}_i \in B\}.$$

Let $N(\mathbf{x})$ denote the number of points of $\mathcal{U}_{m-1,A}$ in $B(\mathbf{0};\|\mathbf{x}\|) \cap \{\mathbf{y} : \mathbf{0} \preccurlyeq \mathbf{y}\}$. Then we obtain

$$E[L^{4\alpha}_{\max}\delta(\mathbf{0})^4] \leq E\sum_{i=1}^{m} W_i^4 = m\int_B \|\mathbf{x}\|^{4\alpha} E[(N(\mathbf{x})+1)^4 \mathbf{1}\{E(\mathbf{x})\}]\frac{d\mathbf{x}}{|A|}.$$

By the Cauchy-Schwarz inequality and the fact that $m \leq 3|A|/2$ by assumption,

$$E[L^{4\alpha}_{\max}\delta(\mathbf{0})^4] \leq \frac{3}{2}\int_B \|\mathbf{x}\|^{4\alpha}(E[(N(\mathbf{x})+1)^8])^{1/2} P[E(\mathbf{x})]^{1/2} d\mathbf{x}. \quad (6.19)$$

The mean of $N(\mathbf{x})$ is bounded by a constant times $\|\mathbf{x}\|^2$ so $E[(N(\mathbf{x})+1)^8] = O(\max(\|\mathbf{x}\|^{16}, 1))$. This follows from the binomial moment generating function for $\mathrm{Bin}(n,p)$, from which we have for $\beta > 0$ that $E[X^{\beta}] \leq k_1(E[X])^{\beta}$ if $pn > 1$ and $E[X^{\beta}] \leq k_2 E[X]$ if $pn < 1$, for some constants $k_1, k_2 > 0$.

Combined with (6.17), (6.18) and (6.19), this shows that $E[L^{4\alpha}_{\max}\delta(\mathbf{0})^4]$ is bounded by a constant times

$$\int_{\mathbf{x}\in B:\|\mathbf{x}\|\geq 1} \|\mathbf{x}\|^{4\alpha+8} \exp(-C\|\mathbf{x}\|/4)d\mathbf{x} + \int_{\mathbf{x}\in B:\|\mathbf{x}\|\leq 1} \|\mathbf{x}\|^{4\alpha} d\mathbf{x},$$

which is bounded by a constant that does not depend on the choice of (A, B).

We need to consider L_0 only when $\mathbf{0} \in B$. For $\mathbf{x} \in \mathbf{R}^2$ with $\mathbf{x} \preccurlyeq^* \mathbf{0}$, let $E'(\mathbf{x})$ denote the event that $R(\mathbf{x},\mathbf{0})$ is empty (i.e., contains no point of $\mathcal{U}_{m-1,A}$). By (6.17) and (6.18), for $\mathbf{0} \in B$ we have

$$\begin{aligned} E[L_0^{4\alpha}] &\leq m\int_{\mathbf{x}\in A:\mathbf{x}\preccurlyeq^*\mathbf{0}} \|\mathbf{x}\|^{4\alpha} P[E'(\mathbf{x})]\frac{d\mathbf{x}}{|A|} \\ &\leq \frac{3}{2}\left[\int_{\mathbf{x}\in A:\mathbf{x}\preccurlyeq^*\mathbf{0},\|\mathbf{x}\|\geq 1} \|\mathbf{x}\|^{4\alpha}\exp(1 - C\|\mathbf{x}\|/2)d\mathbf{x} + \int_{\mathbf{x}\in A:\mathbf{x}\preccurlyeq^*\mathbf{0},\|\mathbf{x}\|\leq 1} \|\mathbf{x}\|^{4\alpha} d\mathbf{x}\right] \end{aligned}$$

which is bounded by a constant. Thus $\Delta^+(\mathcal{U}_{m,A}; B)$ has bounded fourth moment.

Now consider the set of deleted edges. As at (3.20), let $d(\mathbf{x}; \mathcal{U}_{m,A})$ denote the distance from $\mathbf{x}$ to its directed nearest neighbour in $\mathcal{U}_{m,A}$, or zero if no such point exists. Again use $E(\mathbf{x})$ for the event that $\mathbf{x}$ becomes joined to $\mathbf{0}$ on the addition of the origin, and let $E''(\mathbf{V}_i) := E(\mathbf{V}_i) \cap \{\mathbf{V}_i \in B\}$. Then

$$E[\Delta^-(\mathcal{U}_{m,A}; B)^4] = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{\ell=1}^{m} E[d(\mathbf{V}_i; \mathcal{U}_{m,A})^\alpha d(\mathbf{V}_j; \mathcal{U}_{m,A})^\alpha \times d(\mathbf{V}_k; \mathcal{U}_{m,A})^\alpha d(\mathbf{V}_\ell; \mathcal{U}_{m,A})^\alpha \mathbf{1}\{E''(\mathbf{V}_i) \cap E''(\mathbf{V}_j) \cap E''(\mathbf{V}_k) \cap E''(\mathbf{V}_\ell)\}]. \tag{6.20}$$

For i, j, k, ℓ distinct, the (i, j, k, ℓ)th term of (6.20) is bounded by

$$\int_B \int_B \int_B \int_B \frac{d\mathbf{w}}{n}\frac{d\mathbf{x}}{n}\frac{d\mathbf{y}}{n}\frac{d\mathbf{z}}{n} E[d_{m-4}(\mathbf{w})^\alpha d_{m-4}(\mathbf{x})^\alpha d_{m-4}(\mathbf{y})^\alpha d_{m-4}(\mathbf{z})^\alpha \times \mathbf{1}\{E_{m-4}(\mathbf{w}) \cap E_{m-4}(\mathbf{x}) \cap E_{m-4}(\mathbf{y}) \cap E_{m-4}(\mathbf{z})\}], \tag{6.21}$$

where $d_{m-4}(\mathbf{x}) := d(\mathbf{x}, \mathcal{U}_{m-4,A} \cup \{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\})$ (using the notation of (3.20)), and $E_{m-4}(\mathbf{x})$ is the event that $\mathbf{0}$ is the directed nearest neighbour of $\mathbf{x}$ in the set $\mathcal{U}_{m-4,A} \cup \{\mathbf{0}, \mathbf{x}\}$.

Let $I_{m-4}(\mathbf{x})$ denote the indicator variable of the event that $\mathbf{x}$ is a minimal element of $\mathcal{U}_{m-4,A} \cup \{\mathbf{x}\}$. An upper bound for $d_{m-4}(\mathbf{x})$ is provided by $d(\mathbf{x}; \mathcal{U}_{m-4,A} \cup \mathbf{x})$ except when this is zero, so that

$$d_{m-4}(\mathbf{x})^{8\alpha} \leq d(\mathbf{x}; \mathcal{U}_{m-4,A} \cup \{\mathbf{x}\})^{8\alpha} + d(\mathbf{x}; \{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\})^{8\alpha} I_{m-4}(\mathbf{x}). \tag{6.22}$$

For $\mathbf{x} \in B$, it can be shown, by a similar argument to the one used above for L_0, that there is a constant C' such that

$$E[(d(\mathbf{x}; \mathcal{U}_{m-4,A} \cup \{\mathbf{x}\}))^{8\alpha}] < C'. \tag{6.23}$$

Moreover, if $\mathbf{w} \in A$ with $\mathbf{w} \preccurlyeq \mathbf{x}$ and $\|\mathbf{x} - \mathbf{w}\| = t > 0$, then by a similar argument to that at (6.18), and (6.17), we have that

$$E[I_{m-4}(\mathbf{x})] \leq \exp(4 - |R(\mathbf{w}, \mathbf{x})|/2) \leq \exp(4 - Ct/2), \quad t \geq 1,$$

and hence, uniformly over A, B and $\{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\} \subset A$ with $\mathbf{x} \in B$, we have

$$E[d(\mathbf{x}; \{\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}\})^{8\alpha} I_{m-4}(\mathbf{x})] \leq \max\left\{\sup_{t \geq 1}\left(t^{8\alpha}\exp(4 - Ct/2)\right), 1\right\}.$$

Combining this with equation (6.23), we see from equation (6.22) that $E[d_{m-4}(\mathbf{x})^{8\alpha}]$ is bounded by a constant. Also, by a similar argument to (6.18) and (6.17), it can be shown that $P[E_{m-4}(\mathbf{x})] \leq \exp(4 - C\|\mathbf{x}\|/2)$ for $\|\mathbf{x}\| \geq 1$. Therefore, by Hölder's inequality, the expression (6.21) is bounded by a constant times

$$n^{-4} \int\int\int\int d\mathbf{w}d\mathbf{x}d\mathbf{y}d\mathbf{z} \exp(-C(\|\mathbf{w}\| + \|\mathbf{x}\| + \|\mathbf{y}\| + \|\mathbf{z}\|)/16)$$

and therefore is $O(n^{-4})$. Since the number of distinct (i, j, k, ℓ) in the summation (6.20) is bounded by m^4, and hence by $(3/2)^4 n^4$, this shows that the contribution to (6.20) from i, j, k, ℓ distinct is uniformly bounded.

Likewise, the number of terms (i, j, k, ℓ) with only three distinct values (e.g., $i = j$ with i, k, ℓ distinct) is $O(n^3)$. Such a term is bounded by an expression like (6.21) but now with a triple integral, which by a similar argument is $O(n^{-3})$. Hence the contribution to (6.20) of these terms is also bounded. Similarly, the contribution to (6.20) from (i, j, k, ℓ) with two distinct values has $O(n^2)$ terms which are $O(n^{-2})$, and so is bounded. Likewise the contribution to (6.20) from terms with $i = j = k = \ell$ is bounded. Thus the expression (6.20) is uniformly bounded.

Hence $\Delta(\mathcal{U}_{m,A}; B)$ has bounded fourth moments, uniformly in A, B, m. □

Proof of Theorem 6.2.1. By Lemmas 6.2.1, 6.2.2, 6.2.3 and the fact that $\mathcal{L}^\alpha$ is homogeneous of order α, we can apply Corollary 3.2.1, taking $R_0 := (0,1)^2$ and $S_{0,n} := (n^{\varepsilon-1/2}, 1)^2$, to obtain Theorem 6.2.1 (a). □

Remark. An alternative method for proving central limit theorems in geometrical probability is based on dependency graphs. Such a method was employed by Avram and Bertsimas [11] to give central limit theorems for nearest neighbour graphs and other random geometrical structures. A general version of this method is provided by [114]. By a similar argument to [11], one can show that, under $\preccurlyeq^*$, the total weight (for $\alpha > 2/3$) of edges in the MDST from points in the region $(\varepsilon_n, 1)^2$ satisfies a central limit theorem, where

$$\varepsilon_n = \left(\left\lfloor \sqrt{\frac{n}{c \log n}} \right\rfloor\right)^{-1}.$$

Such an approach can be suitably adapted to show that a central limit theorem also holds under the more general partial order specified by θ, ϕ, in the region $(\varepsilon_n, 1 - \varepsilon_n)^2$. The benefit of this method is that it readily yields rates of convergence bounds for the CLT. The martingale method employed has the advantage of yielding the convergence of the variance.

6.3 The edges near the boundary

Next in our analysis of the MDSF and MDST on random points in the unit square, we consider the length of the edges close to the boundary of the square. The limiting structure of the MDSF and MDST near the boundaries is described, in the $\preccurlyeq^*$ case, by the directed linear forest model discussed in Section 5.5, and in the $\preccurlyeq_*$ case, by the on-line nearest-neighbour graph discussed in Section 5.4.

In the $\preccurlyeq^*$ case we have two sub-cases: initially we consider the 'rooted' case where we insert a point at the origin. Later we analyse the multiple sink (or 'unrooted') case, where we do not insert a point at the origin, in a similar way.

Fix $\sigma \in (1/2, 2/3)$. Let B_n denote the L-shaped boundary region $(0,1)^2 \backslash (n^{-\sigma}, 1)^2$. Let B_n' denote the lower boundary region $(0,1) \times (0, n^{-\sigma}]$. Recall from (6.12) that $\mathcal{L}^\alpha(\mathcal{X}; R)$ denotes the contribution to the total weight of the MDST on $\mathcal{X}$ from edges starting at points of $\mathcal{X} \cap R$. When $\mathcal{X}$ is a random point set, set $\tilde{\mathcal{L}}^\alpha(\mathcal{X}; R) := \mathcal{L}^\alpha(\mathcal{X}; R) - E\mathcal{L}^\alpha(\mathcal{X}; R)$.

Theorem 6.3.1 *(a) Suppose the partial order is $\preccurlyeq^*$. Then as $n \to \infty$ we have*

$$\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n^0; B_n) \xrightarrow{\mathcal{D}} \tilde{D}_\alpha^{\{1\}} + \tilde{D}_\alpha^{\{2\}} \qquad (\alpha \geq 1); \tag{6.24}$$

$$\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n^0; B_n) \xrightarrow{\mathcal{D}} \tilde{D}_\alpha^{\{1\}} + \tilde{D}_\alpha^{\{2\}} \qquad (\alpha \geq 1), \tag{6.25}$$

where $\tilde{D}_\alpha^{\{1\}}$, $\tilde{D}_\alpha^{\{2\}}$ are independent random variables with the distribution of $\tilde{D}_\alpha$ given by the fixed-point equation (5.36) for $\alpha = 1$ and by (5.37) for $\alpha > 1$. Also, as $n \to \infty$,

$$\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; B_n) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha^{\{1\}} + \tilde{F}_\alpha^{\{2\}} \qquad (\alpha \geq 1); \tag{6.26}$$

$$\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n; B_n) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha^{\{1\}} + \tilde{F}_\alpha^{\{2\}} \qquad (\alpha \geq 1), \tag{6.27}$$

where $\tilde{F}_\alpha^{\{1\}}$, $\tilde{F}_\alpha^{\{2\}}$ are independent random variables with the same distribution as $\tilde{D}_1$ for $\alpha = 1$ and with the distribution given by the fixed-point equation (5.38) for $\alpha > 1$. Also, as $n \to \infty$,

$$n^{(\alpha-1)/2}\mathcal{L}^\alpha(\mathcal{P}_n; B_n) \xrightarrow{L^1} 0 \qquad (0 < \alpha < 1); \tag{6.28}$$

$$n^{(\alpha-1)/2}\mathcal{L}^\alpha(\mathcal{P}_n^0; B_n) \xrightarrow{L^1} 0 \qquad (0 < \alpha < 1). \tag{6.29}$$

(b) Suppose the partial order is $\preccurlyeq_*$. Then as $n \to \infty$ we have

$$\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; B_n') \xrightarrow{\mathcal{D}} \tilde{Z}_\alpha \qquad (\alpha \geq 1); \tag{6.30}$$

$$\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n; B_n') \xrightarrow{\mathcal{D}} \tilde{Z}_\alpha \qquad (\alpha \geq 1), \tag{6.31}$$

where $\tilde{Z}_\alpha$ is given by (6.1) for $\alpha = 1$ and by (6.2) for $\alpha > 1$. Also, as $n \to \infty$,

$$n^{(\alpha-1)/2}\mathcal{L}^\alpha(\mathcal{P}_n; B'_n) \xrightarrow{L^1} 0 \qquad (0 < \alpha < 1). \tag{6.32}$$

We present the proof of Theorem 6.3.1 (a). Part (b) is proved in a similar manner, but is rather more straightforward (since there is only one boundary contribution to consider), and we omit the details.

The idea behind the proof of Theorem 6.3.1 (a) is to show that the MDSF near each of the two boundaries is close to a DLF system defined on a sequence of uniform random variables coupled to the points of the MDSF. To do this, we produce two explicit sequences of random variables on which we construct the DLF coupled to $\mathcal{P}_n$, a Poisson process of intensity n on $(0,1)^2$, on which the MDSF is constructed.

Let B_n^x be the rectangle $(n^{-\sigma}, 1) \times (0, n^{-\sigma}]$, let B_n^y be the rectangle $(0, n^{-\sigma}] \times (n^{-\sigma}, 1)$, and let B_n^0 be the square $(0, n^{-\sigma}]^2$; see Figure 6.3. Then $B_n = B_n^0 \cup B_n^x \cup B_n^y$. Define the

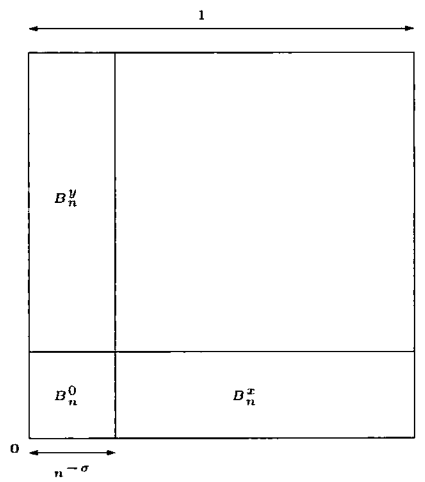


Figure 6.3: The boundary regions

point processes

$$\mathcal{V}_n^x := \mathcal{P}_n \cap (B_n^x \cup B_n^0), \quad \mathcal{V}_n^y := \mathcal{P}_n \cap (B_n^y \cup B_n^0), \text{ and } \mathcal{V}_n^0 := \mathcal{P}_n \cap B_n^0. \tag{6.33}$$

Let $N_n^x := \text{card}(\mathcal{V}_n^x)$, $N_n^y := \text{card}(\mathcal{V}_n^y)$ and $N_n^0 := \text{card}(\mathcal{V}_n^0)$. List $\mathcal{V}_n^x$ in order of increasing y-coordinate as $\mathbf{U}_i^x$, $i = 1, 2, \ldots, N_n^x$. In coordinates, set $\mathbf{U}_i^x = (X_i^x, Y_i^x)$ for each i. Similarly, list $\mathcal{V}_n^y$ in order of increasing x-coordinate as $\mathbf{U}_i^y = (X_i^y, Y_i^y)$, $i = 1, \ldots, N_n^y$. Set $\mathcal{U}_n^x = (X_i^x, i = 1, 2, \ldots, N_n^x)$ and $\mathcal{U}_n^y = (Y_i^y, i = 1, 2, \ldots, N_n^y)$. Then $\mathcal{U}_n^x$ and $\mathcal{U}_n^y$ are sequences of uniform random variables in $(0,1)$, on which we may construct a DLF. Also, we write $\mathcal{U}_n^{x,0}$ for the sequence $(0, X_1^x, X_2^x, \ldots, X_{N_n^x}^x)$, and $\mathcal{U}_n^{y,0}$ for the sequence $(0, Y_1^y, Y_2^y, \ldots, Y_{N_n^y}^y)$.

With the total DLF/DLT weight functional $D^\alpha(\cdot)$ defined in Section 5.5 for random finite sequences in $(0,1)$, the DLF weight $D^\alpha(\mathcal{U}_n^x)$ is coupled in a natural way to the MDSF contribution $\mathcal{L}^\alpha(\mathcal{V}_n^x)$, and likewise for $D^\alpha(\mathcal{U}_n^y)$ and $\mathcal{L}^\alpha(\mathcal{V}_n^y)$, for $D^\alpha(\mathcal{U}_n^{x,0})$ and $\mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\})$, and for $D^\alpha(\mathcal{U}_n^{y,0})$ and $\mathcal{L}^\alpha(\mathcal{V}_n^y \cup \{\mathbf{0}\})$.

Lemma 6.3.1 *Suppose the partial order is $\preccurlyeq^*$. For any $\alpha \geq 1$, as $n \to \infty$,*

$$\mathcal{L}^\alpha(\mathcal{V}_n^x) - D^\alpha(\mathcal{U}_n^x) \xrightarrow{L^2} 0, \text{ and } \mathcal{L}^\alpha(\mathcal{V}_n^y) - D^\alpha(\mathcal{U}_n^y) \xrightarrow{L^2} 0; \tag{6.34}$$

$$\mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\}) - D^\alpha(\mathcal{U}_n^{x,0}) \xrightarrow{L^2} 0, \text{ and } \mathcal{L}^\alpha(\mathcal{V}_n^y \cup \{\mathbf{0}\}) - D^\alpha(\mathcal{U}_n^{y,0}) \xrightarrow{L^2} 0. \tag{6.35}$$

Further, for $0 < \alpha < 1$, as $n \to \infty$,

$$E\left[|\mathcal{L}^\alpha(\mathcal{V}_n^x) - D^\alpha(\mathcal{U}_n^x)|^2\right] = O\left(n^{2-2\sigma-2\alpha\sigma}\right), \tag{6.36}$$

and the corresponding result holds for $\mathcal{V}_n^y$ and $\mathcal{U}_n^y$, and for the rooted cases (with the addition of the origin).

Proof. We approximate the MDSF in the region B_n by two DLFs, coupled to the MDSF. Consider $\mathcal{V}_n^x$; the argument for $\mathcal{V}_n^y$ is entirely analogous.

We have the set of points $\mathcal{V}_n^x = \{(X_i^x, Y_i^x), i = 1, \ldots, N_n^x\}$. We construct the MDSF on these points, and construct the DLF on the x-coordinates, $\mathcal{U}_n^x = (X_i^x, i = 1, \ldots, N_n^x)$. Consider any point (X_i^x, Y_i^x). For any single point, either an edge exists from that point in both constructions, or in neither. Suppose an edge exists, that is suppose X_i^x is joined to a point $X_{D(i)}^x$, $D(i) < i$ in the DLF model, and (X_i^x, Y_i^x) to a point $(X_{N(i)}^x, Y_{N(i)}^x)$ in the MDST (we do not necessarily have $N(i) = D(i)$). By construction, we know that $|X_i^x - X_{D(i)}^x| \leq |X_i^x - X_{N(i)}^x|$, since $N(i) < i$ by the order of our points. It then follows that

$$\|(X_i^x, Y_i^x) - (X_{N(i)}^x, Y_{N(i)}^x)\|^\alpha \geq |X_i^x - X_{N(i)}^x|^\alpha \geq |X_i^x - X_{D(i)}^x|^\alpha,$$

and so we have established that, for all $\alpha > 0$,

$$D^\alpha(\mathcal{U}_n^x) \leq \mathcal{L}^\alpha(\mathcal{V}_n^x); \text{ and } D^\alpha(\mathcal{U}_n^{x,0}) \leq \mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\}).$$

Now, by the construction of the MDST, we have that

$$\|(X_i^x, Y_i^x) - (X_{N(i)}^x, Y_{N(i)}^x)\| \leq \|(X_i^x, Y_i^x) - (X_{D(i)}^x, Y_{D(i)}^x)\|. \tag{6.37}$$

If $(x, y) \in (0,1)^2$ then $\|(x,y)\| \leq x + y$, and by the Mean Value Theorem for the function $t \mapsto t^\alpha$, for $\alpha \geq 1$,

$$\|(x,y)\|^\alpha - x^\alpha \leq (x+y)^\alpha - x^\alpha \leq \alpha 2^{\alpha-1} y \quad (\alpha \geq 1).$$

Hence, for $\alpha \geq 1$,

$$\|(X_i^x, Y_i^x) - (X_{D(i)}^x, Y_{D(i)}^x)\|^\alpha - (X_i^x - X_{D(i)}^x)^\alpha \leq \alpha 2^{\alpha-1}(Y_i^x - Y_{D(i)}^x). \tag{6.38}$$

Then (6.37) and (6.38) yield, for $\alpha \geq 1$,

$$\|(X_i^x, Y_i^x) - (X_{N(i)}^x, Y_{N(i)}^x)\|^\alpha - (X_i^x - X_{D(i)}^x)^\alpha \leq \alpha 2^{\alpha-1}(Y_i^x - Y_{D(i)}^x).$$

Hence, for $\alpha \geq 1$,

$$0 \leq \mathcal{L}^\alpha(\mathcal{V}_n^x) - D^\alpha(\mathcal{U}_n^x) \leq \alpha 2^{\alpha-1} \sum_{i=1}^{N_n^x} (Y_i^x - Y_{D(i)}^x).$$

Thus, for $\alpha \geq 1$,

$$0 \leq \mathcal{L}^\alpha(\mathcal{V}_n^x) - D^\alpha(\mathcal{U}_n^x) \leq \alpha 2^{\alpha-1} N_n^x n^{-\sigma};$$
$$\text{and} \quad 0 \leq \mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\}) - D^\alpha(\mathcal{U}_n^{x,0}) \leq \alpha 2^{\alpha-1} N_n^x n^{-\sigma}. \tag{6.39}$$

We have $N_n^x \sim \text{Po}\,(n^{1-\sigma})$, so that since $\sigma > 1/2$, we have

$$E[(\mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\}) - D^\alpha(\mathcal{U}_n^{x,0}))^2] \leq \alpha^2 2^{2\alpha-2} n^{-2\sigma} E[(N_n^x)^2] \to 0, \quad \alpha \geq 1.$$

An entirely analogous argument leads to the same statement for $\mathcal{U}_n^y$ and $\mathcal{V}_n^y$, and we obtain (6.34), and (6.35) in identical fashion.

We now consider $0 < \alpha < 1$. By the concavity of the function $t \mapsto t^\alpha$ for $\alpha < 1$, we have for $x > 0, y > 0$ that

$$\|(x, y)\|^\alpha - x^\alpha \leq (x + y)^\alpha - x^\alpha \leq y^\alpha \quad (0 < \alpha < 1).$$

Then, by a similar argument to (6.39) in the $\alpha \geq 1$ case, we obtain

$$0 \leq \mathcal{L}^\alpha(\mathcal{V}_n^x) - D^\alpha(\mathcal{U}_n^x) \leq N_n^x n^{-\alpha\sigma}.$$

Then (6.36) follows since $N_n^x \sim \text{Po}\,(n^{1-\sigma})$, and the rooted case is similar. □

Lemma 6.3.2 *Suppose the partial order is $\preccurlyeq^*$. Suppose $\tilde{D}_1$ has distribution given by (5.36), $\tilde{D}_\alpha$, $\alpha > 1$, has distribution given by (5.37), and $\tilde{F}_\alpha$, $\alpha > 1$, has distribution given by (5.38). Then as $n \to \infty$,*

$$\tilde{\mathcal{L}}^1(\mathcal{V}_n^x \cup \{\mathbf{0}\}) \xrightarrow{\mathcal{D}} \tilde{D}_1, \quad \text{and} \quad \tilde{\mathcal{L}}^1(\mathcal{V}_n^x) \xrightarrow{\mathcal{D}} \tilde{D}_1; \tag{6.40}$$

$$\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\}) \xrightarrow{\mathcal{D}} \tilde{D}_\alpha, \quad \text{and} \quad \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha \quad (\alpha > 1). \tag{6.41}$$

Moreover, (6.40) and (6.41) also hold with $\mathcal{V}_n^x$ replaced by $\mathcal{V}_n^y$.

Proof. As usual we present the argument for $\mathcal{V}_n^x$ only, since the result for $\mathcal{V}_n^y$ follows in the same manner. First consider the $\alpha > 1$ case. We have the distributional equality

$$\mathcal{L}\left(D^\alpha(\mathcal{U}_n^{x,0})\middle|\, N_n^x = m\right) = \mathcal{L}\left(D^\alpha(\mathcal{U}_m^0)\right); \qquad \mathcal{L}\left(D^\alpha(\mathcal{U}_n^x)\middle|\, N_n^x = m\right) = \mathcal{L}\left(D^\alpha(\mathcal{U}_m)\right).$$

But N_n^x is Poisson with mean $n^{1-\sigma}$, and so tends to infinity almost surely. Thus by Theorem 5.2.3 (ii), $D^\alpha(\mathcal{U}_n^{x,0}) \xrightarrow{\mathcal{D}} D_\alpha$ and $D^\alpha(\mathcal{U}_n^x) \xrightarrow{\mathcal{D}} F_\alpha$ as $n \to \infty$, and so by Lemma 6.3.1 and Slutsky's theorem, we obtain

$$\mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\}) \xrightarrow{\mathcal{D}} D_\alpha \quad \text{and} \quad \mathcal{L}^\alpha(\mathcal{V}_n^x) \xrightarrow{\mathcal{D}} F_\alpha \quad \text{as} \quad n \to \infty. \tag{6.42}$$

Also, $E[D^\alpha(\mathcal{U}_n^{x,0})] \to (\alpha - 1)^{-1}$ by (5.138), so by Lemma 6.3.1 and Proposition 5.5.5, $E[\mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\})] \to (\alpha-1)^{-1} = E[D_\alpha]$. Similarly, by (5.145), Lemma 6.3.1 and Proposition 5.5.6, $E[\mathcal{L}^\alpha(\mathcal{V}_n^x)] \to (\alpha(\alpha-1))^{-1} = E[F_\alpha]$. Hence, (6.42) still holds with the centred variables, i.e., equation (6.41) holds.

Now suppose $\alpha = 1$. Since N_n^x is Poisson with parameter $n^{1-\sigma}$, Lemma 5.5.7 (i), with $t = n^{1-\sigma}$, then shows that $\tilde{D}^1(\mathcal{U}_n^{x,0}) \xrightarrow{\mathcal{D}} \tilde{D}_1$ as $n \to \infty$. Slutsky's theorem with Lemma 6.3.1 then implies that $\tilde{\mathcal{L}}^1(\mathcal{V}_n^x \cup \{\mathbf{0}\}) \xrightarrow{\mathcal{D}} \tilde{D}_1$. In the same way we obtain $\tilde{\mathcal{L}}^1(\mathcal{V}_n^x) \xrightarrow{\mathcal{D}} \tilde{D}_1$, this time using part (ii) instead of part (i) of Lemma 5.5.7, along with Proposition 5.5.9. □

Note that $D^\alpha(\mathcal{U}_n^x)$ and $D^\alpha(\mathcal{U}_n^y)$ are not independent. To deal with this, we define

$$\tilde{\mathcal{V}}_n^x := \mathcal{P}_n \cap B_x^n, \text{ and } \tilde{\mathcal{V}}_n^y := \mathcal{P}_n \cap B_y^n.$$

Also, recall the definition of $\mathcal{V}_n^0$ at (6.33). Let $\tilde{N}_n^x := \operatorname{card}(\tilde{\mathcal{V}}_n^x)$ and $\tilde{N}_n^y := \operatorname{card}(\tilde{\mathcal{V}}_n^y)$. Since B_n^x and B_n^y are disjoint, $\mathcal{L}^\alpha(\tilde{\mathcal{V}}_n^x)$ and $\mathcal{L}^\alpha(\tilde{\mathcal{V}}_n^y)$ are independent, by the spatial independence property of the Poisson process $\mathcal{P}_n$.

Lemma 6.3.3 *Suppose the partial order is $\preccurlyeq^*$ and $\alpha > 0$. Then:*

(i) As $n \to \infty$,

$$\mathcal{L}^\alpha(\mathcal{V}_n^x) - \mathcal{L}^\alpha(\tilde{\mathcal{V}}_n^x) \xrightarrow{L^1} 0, \text{ and } \mathcal{L}^\alpha(\mathcal{V}_n^y) - \mathcal{L}^\alpha(\tilde{\mathcal{V}}_n^y) \xrightarrow{L^1} 0; \tag{6.43}$$

$$\mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\}) - \mathcal{L}^\alpha(\tilde{\mathcal{V}}_n^x \cup \{\mathbf{0}\}) \xrightarrow{L^1} 0; \; \mathcal{L}^\alpha(\mathcal{V}_n^y \cup \{\mathbf{0}\}) - \mathcal{L}^\alpha(\tilde{\mathcal{V}}_n^y \cup \{\mathbf{0}\}) \xrightarrow{L^1} 0. \tag{6.44}$$

(ii) As $n \to \infty$, we have $\mathcal{L}^\alpha(\mathcal{V}_n^0) \xrightarrow{L^1} 0$, and $\mathcal{L}^\alpha(\mathcal{V}_n^0 \cup \{\mathbf{0}\}) \xrightarrow{L^1} 0$.

Proof. We first prove (i). We give only the argument for $\mathcal{V}_n^x$; that for $\mathcal{V}_n^y$ is analogous. Set $\Delta := \mathcal{L}^\alpha(\mathcal{V}_n^x) - \mathcal{L}^\alpha(\tilde{\mathcal{V}}_n^x)$. Let $\beta = (\sigma + (1/2))/2$. Then $1/2 < \beta < \sigma$.

Assume without loss of generality that $\mathcal{P}_n$ is the restriction to $(0,1)^2$ of a homogeneous Poisson process $\mathcal{H}_n$ of intensity n on $\mathbf{R}^2$. Let $\mathbf{U}^- = (X^-, Y^-)$ be the point of $\mathcal{H}_n \cap ((0, n^{-\beta}] \times (0, \infty))$ with minimal y-coordinate. Then X^- is uniform on $(0, n^{-\beta}]$. Let E_n be the event that $X^- > 3n^{-\sigma}$; then $P[E_n^c] = 3n^{\beta-\sigma}$ for n large enough.

Let Δ_1 be the the contribution to Δ from edges starting at points in $(0, n^{-\beta}] \times (0, n^{-\sigma}]$. Then the absolute value of Δ_1 is bounded by the product of $(\sqrt{2}n^{-\beta})^\alpha$ and the number of points of $\mathcal{P}_n$ in $(0, n^{-\beta}] \times (0, n^{-\sigma}]$. Hence, for any $\alpha > 0$,

$$
\begin{aligned}
E[|\Delta_1|] &\leq (\sqrt{2}n^{-\beta})^\alpha E\left[\operatorname{card}\left(\mathcal{P}_n \cap ((0, n^{-\beta}] \times (0, n^{-\sigma}])\right)\right] \\
&= 2^{\alpha/2} n^{1-\beta-\sigma-\alpha\beta} \to 0.
\end{aligned} \tag{6.45}
$$

Let $\Delta_2 := \Delta - \Delta_1$, the contribution to Δ from edges starting at points in $(n^{-\beta}, 1) \times (0, n^{-\sigma}]$. Then by the triangle inequality, if E_n occurs then these edges are unaffected by points in B_n^0, so that Δ_2 is zero if E_n occurs. Also, only minimal elements of $\mathcal{P}_n \cap (n^{-\beta}, 1) \times (0, n^{-\sigma}]$ can possibly have their directed nearest neighbour in $(0, n^{-\sigma}] \times (0, n^{-\sigma}]$; hence, if M_n denotes the number of such minimal elements then $|\Delta_2|$ is bounded by $2^{\alpha/2} M_n$. Hence, using (2.26), we obtain

$$
E[|\Delta_2|] \leq 2^{\alpha/2} P[E_n^c] E[M_n] = O(n^{\beta-\sigma} \log n)
$$

which tends to zero. Combined with (6.45), this gives us (6.43). The same argument gives us (6.44).

For (ii), note that

$$
E\left[\mathcal{L}^\alpha(\mathcal{V}_n^0)\right] \leq (\sqrt{2}n^{-\sigma})^\alpha E[N_n^0] = 2^{\alpha/2} n^{1-2\sigma-\sigma\alpha} \to 0, \text{ as } n \to \infty,
$$

for any $\alpha > 0$. Thus $\mathcal{L}^\alpha(\mathcal{V}_n^0) \xrightarrow{L^1} 0$, and similarly $\mathcal{L}^\alpha(\mathcal{V}_n^0 \cup \{\mathbf{0}\}) \xrightarrow{L^1} 0$. □

In proving our next lemma (and again later on) we use the following elementary fact. If $N(n)$ is Poisson with parameter n, then as $n \to \infty$ we have

$$
E[|N(n) - n| \log \max(N(n), n)] = O(n^{1/2} \log n). \tag{6.46}
$$

To see this, set $Y_n := |N(n) - n| \log \max(N(n), n)$. Then $Y_n \mathbf{1}_{\{N(n) \leq 2n\}} \leq |N(n) - n| \log(2n)$, and the expectation of this is $O(n^{1/2} \log n)$ by Jensen's inequality since $\operatorname{Var}(N(n)) =$

n. On the other hand, the Cauchy-Schwarz inequality shows that $E[Y_n \mathbf{1}_{\{N(n)>2n\}}] \to 0$, and (6.46) follows.

We now state a lemma for coupling $\mathcal{U}_n$ and $\mathcal{P}_n$. The $\alpha \geq 1$ part will be used in the proof of Theorem 6.3.1. The $0 < \alpha < 1$ part will be needed later, in the proof of Theorem 6.1.1. As in Section 6.2, let $S_{0,n}$ denote the 'inner' region $(n^{\varepsilon-1/2}, 1)^2$, with $\varepsilon \in (0, 1/2)$ a constant. The boundary region B_n is disjoint from $S_{0,n}$; let C_n denote the intermediate region $(0,1)^2 \setminus (B_n \cup S_{0,n})$, so that $B_n \cup C_n = (0,1)^2 \setminus S_{0,n}$.

Lemma 6.3.4 *Suppose the partial order is $\preccurlyeq^*$. There exists a coupling of $\mathcal{U}_n$ and $\mathcal{P}_n$ such that:*

(i) For $0 < \alpha < 1$, provided $\varepsilon < (1-\alpha)/2$, we have that as $n \to \infty$,

$$n^{(\alpha-1)/2} E[|\mathcal{L}^\alpha(\mathcal{U}_n; B_n \cup C_n) - \mathcal{L}^\alpha(\mathcal{P}_n; B_n \cup C_n)|] \to 0 \tag{6.47}$$

and

$$n^{(\alpha-1)/2} E[|\mathcal{L}^\alpha(\mathcal{U}_n^0; B_n \cup C_n) - \mathcal{L}^\alpha(\mathcal{P}_n^0; B_n \cup C_n)|] \to 0. \tag{6.48}$$

(ii) For $\alpha \geq 1$, we have that as $n \to \infty$,

$$E[|\mathcal{L}^\alpha(\mathcal{U}_n; B_n) - \mathcal{L}^\alpha(\mathcal{P}_n; B_n)|] \to 0 \tag{6.49}$$

and

$$E[|\mathcal{L}^\alpha(\mathcal{U}_n^0; B_n) - \mathcal{L}^\alpha(\mathcal{P}_n^0; B_n)|] \to 0. \tag{6.50}$$

Proof. We couple $\mathcal{U}_n$ and $\mathcal{P}_n$ in the following standard way. Let $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \ldots$ be independent uniform random vectors on $(0,1)^2$, and let $N(n) \sim \text{Po}(n)$ be independent of $(\mathbf{U}_1, \mathbf{U}_2, \ldots)$. For $m \in \mathbf{N}$ (and in particular for $m = n$) set $\mathcal{U}_m := \{\mathbf{U}_1, \ldots, \mathbf{U}_m\}$; set $\mathcal{P}_n := \{\mathbf{U}_1, \ldots, \mathbf{U}_{N(n)}\}$.

For each $m \in \mathbf{N}$, let Y_m denote the in-degree of vertex $\mathbf{U}_m$ in the MDST on $\mathcal{U}_m$. Suppose $\mathbf{U}_m = \mathbf{x}$. Then an upper bound for Y_m is provided by the number of minimal elements of the restriction of $\mathcal{U}_{m-1}$ to the rectangle $\{\mathbf{y} \in (0,1)^2 : \mathbf{x} \preccurlyeq^* \mathbf{y}\}$. Hence, conditional on $\mathbf{U}_m = \mathbf{x}$ and on there being k points of $\mathcal{U}_{m-1}$ in this rectangle, the expected value of Y_m is bounded by the expected number of minimal elements in a random uniform sample of k points in this rectangle, and hence (see (2.26)) by $1 + \log k$. Hence, given the value of $\mathbf{U}_m$, the conditional expectation of Y_m is bounded by $1 + \log m$.

First we prove the statements in part (i) ($0 < \alpha < 1$). Suppose $\varepsilon < (1-\alpha)/2$. Then

$$|\mathcal{L}^\alpha(\mathcal{U}_m; B_n \cup C_n) - \mathcal{L}^\alpha(\mathcal{U}_{m-1}; B_n \cup C_n)| \leq 2^{\alpha/2}(Y_m + 1)\mathbf{1}\{\mathbf{U}_m \in B_n \cup C_n\}. \quad (6.51)$$

Since $B_n \cup C_n$ has area $2n^{\varepsilon-1/2} - n^{2\varepsilon-1}$, we obtain

$$E[(Y_m+1)\mathbf{1}\{\mathbf{U}_m \in B_n \cup C_n\}] \leq (2 + \log m)2n^{\varepsilon-1/2}.$$

Hence, by (6.51) there is a constant C such that

$$n^{(\alpha-1)/2}E[(|\mathcal{L}^\alpha(\mathcal{P}_n; B_n \cup C_n) - \mathcal{L}^\alpha(\mathcal{U}_n; B_n \cup C_n)|)|N(n)]$$
$$\leq C|N(n) - n|\log(\max(N(n), n))n^{(\alpha+2\varepsilon-2)/2},$$

and since we assume $\alpha + 2\varepsilon < 1$, by equation (6.46) the expected value of the right hand side tends to zero as $n \to \infty$, and we obtain (6.47). Likewise in the rooted case (6.48).

Now we prove part (ii). For $\alpha \geq 1$, we have

$$|\mathcal{L}^\alpha(\mathcal{U}_m; B_n) - \mathcal{L}^\alpha(\mathcal{U}_{m-1}; B_n)| \leq 2^{\alpha/2}(Y_m + 1)\mathbf{1}\{\mathbf{U}_m \in B_n\}. \quad (6.52)$$

Since B_n has area $2n^{-\sigma} - n^{-2\sigma}$, by (6.52) there is a constant C such that

$$E[(|\mathcal{L}^\alpha(\mathcal{P}_n; B_n) - \mathcal{L}^\alpha(\mathcal{U}_n; B_n)|)|N(n)] \leq C|N(n) - n|\log(\max(N(n), n))n^{-\sigma},$$

and since $\sigma > 1/2$, by equation (6.46) the expected value of the right hand side tends to zero as $n \to \infty$, and we obtain (6.49). We get (6.50) similarly. □

Proof of Theorem 6.3.1. We prove part (a). Suppose $\alpha \geq 1$. We have that

$$\tilde{\mathcal{L}}^\alpha(\tilde{\mathcal{V}}_n^x) = \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x) + (\tilde{\mathcal{L}}^\alpha(\tilde{\mathcal{V}}_n^x) - \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x)).$$

The final bracket converges to zero in probability, by Lemma 6.3.3 (i). Thus by Lemma 6.3.2 and Slutsky's theorem, we obtain $\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha$ (where we have $\tilde{F}_1 \stackrel{\mathcal{D}}{=} \tilde{D}_1$). Now

$$\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x) + \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^y) = \tilde{\mathcal{L}}^\alpha(\tilde{\mathcal{V}}_n^x) + \tilde{\mathcal{L}}^\alpha(\tilde{\mathcal{V}}_n^y) + (\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x) - \tilde{\mathcal{L}}^\alpha(\tilde{\mathcal{V}}_n^x)) + (\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^y) - \tilde{\mathcal{L}}^\alpha(\tilde{\mathcal{V}}_n^y)).$$

The last two brackets converge to zero in probability, by Lemma 6.3.3 (i). Then the independence of $\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x)$ and $\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^y)$ and another application of Slutsky's theorem yield

$$\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x) + \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^y) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha^{\{1\}} + \tilde{F}_\alpha^{\{2\}},$$

where $\tilde{F}_\alpha^{\{1\}}$ and $\tilde{F}_\alpha^{\{2\}}$ are independent copies of $\tilde{F}_\alpha$. Similarly,

$$\tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x \cup \{\mathbf{0}\}) + \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^y \cup \{\mathbf{0}\}) \xrightarrow{\mathcal{D}} \tilde{D}_\alpha^{\{1\}} + \tilde{D}_\alpha^{\{2\}}.$$

Finally, since $\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; B_n) = \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^x) + \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^y) - \tilde{\mathcal{L}}^\alpha(\mathcal{V}_n^0)$ (with a similar statement including the origin) Lemma 6.3.3 (ii) and Slutsky's theorem complete the proof of (6.24) and (6.26).

To deduce (6.25) and (6.27), assume without loss of generality that $\mathcal{U}_n$ and $\mathcal{P}_n$ are coupled in the manner of Lemma 6.3.4. Then $\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; B_n) - \tilde{\mathcal{L}}^\alpha(\mathcal{U}_n; B_n)$ tends to zero in probability by (6.49), and $\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n^0; B_n) - \tilde{\mathcal{L}}^\alpha(\mathcal{U}_n^0; B_n)$ tends to zero in probability by (6.50). Hence by Slutsky's theorem, the convergence results (6.24) and (6.26) carry through to the binomial point process case, i.e., (6.25) and (6.27) hold.

Now suppose $0 < \alpha < 1$. Then (6.36) gives us

$$E\left[\left|n^{(\alpha-1)/2}\left(\mathcal{L}^\alpha(\mathcal{V}_n^x) - D^\alpha(\mathcal{U}_n^x)\right)\right|^2\right] = O\left(n^{(\alpha+1)(1-2\sigma)}\right), \tag{6.53}$$

which tends to 0 as $n \to \infty$, since $\sigma > 1/2$. Likewise for the rooted case,

$$E\left[\left|n^{(\alpha-1)/2}\left(\mathcal{L}^\alpha(\mathcal{V}_n^x \cup \{0\}) - D^\alpha(\mathcal{U}_n^{x,0})\right)\right|^2\right] = O\left(n^{(\alpha+1)(1-2\sigma)}\right), \tag{6.54}$$

By Proposition 5.5.2 we have

$$E[n^{(\alpha-1)/2}D^\alpha(\mathcal{U}_n^x)] = O(n^{(\alpha-1)/2}E[(N_n^x)^{1-\alpha}]) = O(n^{(\alpha-1)(\sigma-1/2)}) \to 0,$$

and combined with (6.53) this completes the proof of (6.28). Similarly, by Proposition 5.5.1,

$$E[n^{(\alpha-1)/2}D^\alpha(\mathcal{U}_n^{x,0})] = O(n^{(\alpha-1)/2}E[(N_n^x)^{1-\alpha}]) = O(n^{(\alpha-1)(\sigma-1/2)}) \to 0,$$

and combined with (6.54) this gives us (6.29). □

6.4 Proof of Theorem 6.1.1 (a)

Throughout this section, we take the partial order to be $\preccurlyeq^*$. Let $\sigma \in (1/2, 2/3)$. Let $\varepsilon > 0$ with

$$\varepsilon < \min(1/2, (1-\sigma)/3, (3-4\sigma)/10, (2-3\sigma)/8). \tag{6.55}$$

In addition, if $0 < \alpha < 1$, we impose the further condition that $\varepsilon < (1-\alpha)/2$. As in Section 6.2, denote by $S_{0,n}$ the region $(n^{\varepsilon-1/2}, 1)^2$. As in Section 6.3, let B_n denote the region $(0,1)^2 \setminus (n^{-\sigma}, 1)^2$, and let C_n denote $(0,1)^2 \setminus (B_n \cup S_{0,n})$.

We know from Sections 6.2 and 6.3 that, for large n, the weight of edges starting in $S_{0,n}$ satisfies a central limit theorem, and the weight of edges starting in B_n can be approximated by the directed linear forest. We shall show in Lemmas 6.4.2 and 6.4.3 that (with a suitable scaling factor for $\alpha < 1$) the contribution to the total weight from points in C_n has variance converging to zero. To complete the proof of Theorem 6.1.1 in the Poisson case, we shall show that the lengths from B_n and $S_{0,n}$ are asymptotically independent by virtue of the fact that the configuration of points in C_n is (with probability approaching one) sufficient to ensure that the configuration of points in B_n has no effect on the edges from points in $S_{0,n}$. To extend the result to the binomial point process case, we shall use a de-Poissonization argument related to that used in [111].

First consider the region C_n. We naturally divide this into three regions. Let

$$C_n^x := (n^{\varepsilon-1/2}, 1) \times (n^{-\sigma}, n^{\varepsilon-1/2}], \quad C_n^y := (n^{-\sigma}, n^{\varepsilon-1/2}] \times (n^{\varepsilon-1/2}, 1),$$
$$C_n^0 := (n^{-\sigma}, n^{\varepsilon-1/2}]^2.$$

Also, as in Section 6.3, let

$$B_n^x := (n^{-\sigma}, 1) \times (0, n^{-\sigma}], \quad B_n^y := (0, n^{-\sigma}] \times (n^{-\sigma}, 1), \quad B_n^0 := (0, n^{-\sigma}]^2.$$

We divide the C_n and B_n into rectangular cells as follows (see Figure 6.4.) We leave C_n^0 undivided. We set

$$k_n := \lfloor n^{1-\sigma-2\varepsilon} \rfloor \tag{6.56}$$

and divide C_n^x lengthways into k_n cells. For each cell,

$$\text{width} = (1 - n^{\varepsilon-1/2})/k_n \sim n^{2\varepsilon+\sigma-1}; \qquad \text{height} = n^{\varepsilon-1/2} - n^{-\sigma} \sim n^{\varepsilon-1/2}. \tag{6.57}$$

Label these cells Γ_i^x for $i = 1, 2, \ldots, k_n$ from left to right. For each cell Γ_i^x, define the adjoining cell of B_n^x, formed by extending the vertical edges of Γ_i^x, to be β_i^x. The cells β_i^x then have width $(1 - n^{\varepsilon-1/2})/k_n \sim n^{2\varepsilon+\sigma-1}$ and height $n^{-\sigma}$.

In a similar way we divide C_n^y into k_n cells Γ_i^y of height $(1 - n^{\varepsilon-1/2})/k_n$ and width $n^{\varepsilon-1/2} - n^{-\sigma}$, and divide B_n^y into the corresponding cells β_i^y, $i = 1, \ldots, k_n$.

For $i = 2, \ldots, k_n$, let $E_{x,i}$ denote the event that the cell β_{i-1}^x contains at least one point of $\mathcal{P}_n$, and let $E_{y,i}$ denote the event that β_{i-1}^y contains at least one point of $\mathcal{P}_n$.

Lemma 6.4.1 *For n sufficiently large, and for $1 \leq j < i \leq k_n$ with $i - j > 3$, if $E_{x,i}$ (respectively $E_{y,i}$) occurs then no point in the cell Γ_i^x (respectively Γ_i^y) has a directed nearest neighbour in the cell Γ_j^x or β_j^x (Γ_j^y or β_j^y).*

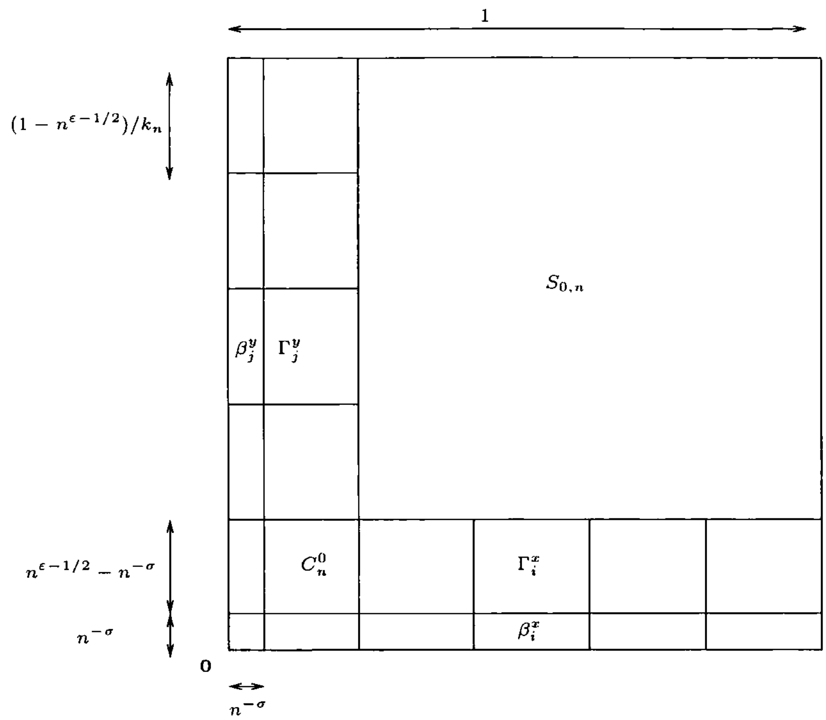


Figure 6.4: The regions of $[0,1)^2$.

Proof. Consider a point X, say, in cell Γ_i^x in C_n^x. Given $E_{x,i}$, we know that there is a point, Y say, in the cell β_{i-1}^x to the left of the β_i^x cell immediately below Γ_i^x, such that $Y \preccurlyeq^* X$, but the difference in x-coordinates between X and Y is no more than twice the width of a cell. So, by the triangle inequality, we have

$$\|X-Y\| \leq 2(1-n^{\varepsilon-1/2})/k_n + n^{\varepsilon-1/2} \sim 2n^{2\varepsilon+\sigma-1}, \tag{6.58}$$

since $\sigma > 1/2$. Now, consider a point Z in a cell Γ_j^x or β_j^x with $j \leq i-4$. In this case, the difference in x-coordinates between X and Z is at least the width of 3 cells, so that

$$\|X-Z\| \geq 3(1-n^{\varepsilon-1/2})/k_n \sim 3n^{2\varepsilon+\sigma-1}. \tag{6.59}$$

Comparing (6.58) and (6.59), we see that X is not connected to Z, which completes the proof. □

Recall from (6.12) that for a point set $\mathcal{S} \subset \mathbf{R}^2$ and a region $R \subseteq \mathbf{R}^2$, $\mathcal{L}^\alpha(\mathcal{S}; R)$ denotes the total weight of edges of the MDSF on $\mathcal{S}$ which originate in the region R.

Lemma 6.4.2 *As $n \to \infty$, we have that*

$$\mathrm{Var}[\mathcal{L}^\alpha(\mathcal{P}_n; C_n)] \to 0 \quad \text{and} \quad \mathrm{Var}[\mathcal{L}^\alpha(\mathcal{P}_n^0; C_n)] \to 0 \qquad (\alpha \geq 1); \tag{6.60}$$

$$\mathrm{Var}[n^{(\alpha-1)/2}\mathcal{L}^\alpha(\mathcal{P}_n;C_n)] \to 0 \qquad (0<\alpha<1); \tag{6.61}$$

$$\mathrm{Var}[n^{(\alpha-1)/2}\mathcal{L}^\alpha(\mathcal{P}_n^0;C_n)] \to 0 \qquad (0<\alpha<1). \tag{6.62}$$

Proof. For ease of notation, write $X_i = \mathcal{L}^\alpha(\mathcal{P}_n;\Gamma_i^x)$ and $Y_i = \mathcal{L}^\alpha(\mathcal{P}_n;\Gamma_i^y)$, for $i = 1,2,\ldots,k_n$. Also let $Z = \mathcal{L}^\alpha(\mathcal{P}_n;C_n^0)$. Then

$$\mathrm{Var}[\mathcal{L}^\alpha(\mathcal{P}_n;C_n)] \;=\; \mathrm{Var}\left[Z+\sum_{i=1}^{k_n}X_i+\sum_{i=1}^{k_n}Y_i\right]. \tag{6.63}$$

Let N_i^x, N_i^y, N_0, respectively, denote the number of points of $\mathcal{P}_n$ in Γ_i^x, Γ_i^y, C_n^0, respectively. Then by (6.57), N_i^x is Poisson with parameter asymptotic to $n^{3\varepsilon+\sigma-1/2}$, while $N_1^x+N_1^y+N_0$ is Poisson with parameter asymptotic to $2n^{3\varepsilon+\sigma-1/2}$; hence as $n\to\infty$ and we have

$$E[(N_i^x)^2] \sim n^{6\varepsilon+2\sigma-1}, \qquad E[(N_1^x+N_1^y+N_0)^2] \sim 4n^{6\varepsilon+2\sigma-1}. \tag{6.64}$$

Edges from points in $\Gamma_1^x\cup\Gamma_1^y\cup C_n^0$ are of length at most $2n^{2\varepsilon+\sigma-1}$, and hence,

$$\begin{aligned}\mathrm{Var}[X_1+Y_1+Z] &\le (2n^{2\varepsilon+\sigma-1})^{2\alpha}E[(N_1^x+N_1^y+N_0)^2]\\ &\sim 2^{2+2\alpha}n^{6\varepsilon+2\sigma-1+2\alpha(2\varepsilon+\sigma-1)}.\end{aligned} \tag{6.65}$$

For $\alpha\ge 1$, since ε is small (6.55), the expression (6.65) is $O(n^{10\varepsilon+4\sigma-3})$ and in fact tends to zero, so that

$$\mathrm{Var}(X_1+Y_1+Z)\to 0 \quad (\alpha\ge 1). \tag{6.66}$$

By Lemma 6.4.1 and (6.58), given $E_{x,i}$, an edge from a point of Γ_i^x can be of length no more than $3n^{2\varepsilon+\sigma-1}$. Thus using (6.64) we have

$$\begin{aligned}\mathrm{Var}[X_i\mathbf{1}\{E_{x,i}\}] &\le E[X_i^2\mathbf{1}\{E_{x,i}\}] \le (3n^{2\varepsilon+\sigma-1})^{2\alpha}E[(N_i^x)^2]\\ &= O(n^{6\varepsilon+2\sigma-1+2\alpha(2\varepsilon+\sigma-1)}).\end{aligned} \tag{6.67}$$

Next, observe that $\mathrm{Cov}[X_i\mathbf{1}\{E_{x,i}\},X_j\mathbf{1}\{E_{x,j}\}]=0$ for $i-j>3$, since by Lemma 6.4.1, $X_i\mathbf{1}\{E_{x,i}\}$ is determined by the restriction of $\mathcal{P}_n$ to the union of the regions $\Gamma_\ell^x\cup\beta_\ell^x, i-3\le \ell\le i$. Thus by (6.56), Cauchy-Schwarz and (6.67), we obtain

$$\begin{aligned}\mathrm{Var}\left[\sum_{i=2}^{k_n}X_i\mathbf{1}\{E_{x,i}\}\right] &= \sum_{i=2}^{k_n}\mathrm{Var}[X_i\mathbf{1}\{E_{x,i}\}]\\ &\quad+\sum_{i=2}^{k_n}\sum_{j:1\le|j-i|\le 3}\mathrm{Cov}[X_i\mathbf{1}\{E_{x,i}\},X_j\mathbf{1}\{E_{x,j}\}]\\ &= O(n^{4\varepsilon+\sigma+2\alpha(2\varepsilon+\sigma-1)}).\end{aligned} \tag{6.68}$$

For $\alpha \geq 1$, the bound in (6.68) tends to zero as $n \to \infty$, since $1/2 < \sigma < 2/3$ and ε is small (6.55).

By (6.56), the cells β_i^x, $i = 1, \ldots, k_n$, have width asymptotic to $n^{2\varepsilon+\sigma-1}$ and height $n^{-\sigma}$, so the mean number of points of $\mathcal{P}_n$ in one of these cells is asymptotic to $n^{2\varepsilon}$; hence for any cell β_i^x or β_i^y, $i = 1, \ldots, k_n$, the probability that the cell contains no point of $\mathcal{P}_n$ is given by $\exp\{-n^{2\varepsilon}(1 + o(1))\}$. Hence for n large enough, and $i = 2, \ldots, k_n$, we have $P[E_{x,i}^c] \leq \exp(-n^{\varepsilon})$, and hence by (6.64),

$$\begin{aligned} \operatorname{Var}[X_i \mathbf{1}\{E_{x,i}^c\}] \leq E[X_i^2 | E_{x,i}^c] P[E_{x,i}^c] &\leq 2^{\alpha} E[(N_i^x)^2] P[E_{x,i}^c] \\ &= O(n^{6\varepsilon+2\sigma-1} \exp(-n^{\varepsilon})). \end{aligned} \tag{6.69}$$

Hence by Cauchy-Schwarz we have

$$\begin{aligned} \operatorname{Var}\left[\sum_{i=2}^{k_n} X_i \mathbf{1}\{E_{x,i}^c\}\right] &= \sum_{i=2}^{k_n} \operatorname{Var}[X_i \mathbf{1}\{E_{x,i}^c\}] + \sum_{i \neq j} \operatorname{Cov}[X_i \mathbf{1}\{E_{x,i}^c\}, X_j \mathbf{1}\{E_{x,j}^c\}] \\ &= O\left(k_n^2 n^{6\varepsilon+2\sigma-1} \exp(-n^{\varepsilon})\right) \to 0, \end{aligned} \tag{6.70}$$

as $n \to \infty$. Then by (6.68), (6.70), and the analogous estimates for Y_i, along with the Cauchy-Schwarz inequality, we obtain for $\alpha \geq 1$ that

$$\operatorname{Var}\left[\sum_{i=2}^{k_n} X_i \mathbf{1}\{E_{x,i}\} + \sum_{i=2}^{k_n} Y_i \mathbf{1}\{E_{y,i}\} + \sum_{i=2}^{k_n} X_i \mathbf{1}\{E_{x,i}^c\} + \sum_{i=2}^{k_n} Y_i \mathbf{1}\{E_{y,i}^c\}\right] \to 0, \tag{6.71}$$

as $n \to \infty$. By (6.63) with (6.66), (6.71), and Cauchy-Schwarz again, we obtain the first part of (6.60). The argument for $\mathcal{P}_n^0$ is the same as for $\mathcal{P}_n$, so we have (6.60).

Now suppose $0 < \alpha < 1$. We obtain (6.61) and (6.62) in a similar way to (6.60), since (6.65) implies that

$$\operatorname{Var}(n^{(\alpha-1)/2}(X_1 + Y_1 + Z)) = O(n^{6\varepsilon+2\sigma-2+\alpha(4\varepsilon+2\sigma-1)})$$

and (6.68) implies

$$\operatorname{Var}\left(n^{(\alpha-1)/2} \sum_{i=2}^{k_n} X_i \mathbf{1}\{E_{x,i}\}\right) = O(n^{4\varepsilon+\sigma-1+\alpha(4\varepsilon+2\sigma-1)}),$$

and both of these bounds tend to zero when $0 < \alpha < 1$, $1/2 < \sigma < 2/3$, and ε is small (6.55). □

To prove those parts of Theorem 6.1.1 which refer to the binomial process $\mathcal{U}_n$, we need further results comparing the processes $\mathcal{U}_n$ and $\mathcal{P}_n$ when they are coupled as in Lemma 6.3.4.

Lemma 6.4.3 *Suppose $\alpha \geq 1$. With $\mathcal{U}_n$ and $\mathcal{P}_n$ coupled as in Lemma 6.3.4, we have that as $n \to \infty$*

$$\mathcal{L}^\alpha(\mathcal{U}_n; C_n) - \mathcal{L}^\alpha(\mathcal{P}_n; C_n) \xrightarrow{L^1} 0 \quad \text{and} \quad \mathcal{L}^\alpha(\mathcal{U}_n^0; C_n) - \mathcal{L}^\alpha(\mathcal{P}_n^0; C_n) \xrightarrow{L^1} 0. \tag{6.72}$$

Proof. Let $\mathcal{P}_n$ and $\mathcal{U}_m$ ($m \in \mathbf{N}$) be coupled as described in Lemma 6.3.4. Given n, for $m \in \mathbf{N}$ define the event

$$E_{m,n} := \cap_{1 \leq i \leq k_n}(\{\mathcal{U}_{m-1} \cap \beta_i^x \neq \emptyset\} \cap \{\mathcal{U}_{m-1} \cap \beta_i^y \neq \emptyset\}),$$

with the sub-cells β_i^x and β_i^y of B_n as defined near the start of Section 6.4. Then by similar arguments to those for $P[E_{x,i}^c]$ above, we have

$$P[E_{m,n}^c] = O(n^{1-\sigma-2\varepsilon}\exp(-n^\varepsilon/2)), \qquad m \geq n/2+1.$$

As in the proof of Lemma 6.3.4, let Y_m denote the in-degree of vertex $\mathbf{U}_m$ in the MDST on $\mathcal{U}_m$. Then

$$|\mathcal{L}^\alpha(\mathcal{U}_m; C_n) - \mathcal{L}^\alpha(\mathcal{U}_{m-1}; C_n)| \leq (Y_m+1)\mathbf{1}\{\mathbf{U}_m \in C_n\}\left((3n^{2\varepsilon+\sigma-1})^\alpha + 2^{\alpha/2}\mathbf{1}\{E_{m,n}^c\}\right).$$

Thus, given $N(n)$,

$$\begin{aligned} |\mathcal{L}^\alpha(\mathcal{U}_n; C_n) - \mathcal{L}^\alpha(\mathcal{P}_n; C_n)| \; &\leq \textstyle\sum_{m=\min(N(n),n)}^{\max(N(n),n)} \quad (Y_m+1)\mathbf{1}\{\mathbf{U}_m \in C_n\} \\ &\qquad \times \left(3^\alpha n^{\alpha(2\varepsilon+\sigma-1)} + 2^{\alpha/2}\mathbf{1}\{E_{m,n}^c\}\right). \end{aligned}$$

Since C_n has area less than $2n^{\varepsilon-1/2}$, by equation (2.26) there exists a constant C such that, for n sufficiently large and $N(n) \geq n/2+1$,

$$\begin{aligned} E\left[(|\mathcal{L}^\alpha(\mathcal{U}_n; C_n) - \mathcal{L}^\alpha(\mathcal{P}_n; C_n)|) | N(n)\right] &\leq 2^{\alpha/2} n \mathbf{1}_{\{N(n)<n/2+1\}} \\ +C|N(n)-n|\log(\max(N(n),n))n^{\alpha(2\varepsilon+\sigma-1)+\varepsilon-1/2}&\mathbf{1}_{\{N(n)\geq n/2+1\}}. \end{aligned} \tag{6.73}$$

By tail bounds for the Poisson distribution, we have $nP[N(n) < n/2+1] \to 0$ as $n \to \infty$, and hence, taking expectations in (6.73) and using (6.46), we obtain

$$E\left[|\mathcal{L}^\alpha(\mathcal{U}_n; C_n) - \mathcal{L}^\alpha(\mathcal{P}_n; C_n)|\right] = O(n^{\alpha(2\varepsilon+\sigma-1)+\varepsilon}\log n) + o(1),$$

which tends to zero since $\alpha \geq 1$, $1/2 < \sigma < 2/3$ and ε is small (see (6.55)). So we obtain the unrooted part of (6.72). The argument is the same in the rooted case. □

Lemma 6.4.4 *Suppose $\mathcal{U}_n$ and $\mathcal{P}_n$ are coupled as described in Lemma 6.3.4, with $N(n) :=$ card$(\mathcal{P}_n)$. Let $\Delta(\infty)$ be given by Definition 3.2.1 with $H = \mathcal{L}^1$, and set $\alpha_1 := E[\Delta(\infty)]$. Then as $n \to \infty$ we have*

$$\mathcal{L}^1(\mathcal{P}_n; S_{0,n}) - \mathcal{L}^1(\mathcal{U}_n; S_{0,n}) - n^{-1/2}\alpha_1(N(n) - n) \xrightarrow{L^2} 0; \tag{6.74}$$

$$\mathcal{L}^1(\mathcal{P}_n^0; S_{0,n}) - \mathcal{L}^1(\mathcal{U}_n^0; S_{0,n}) - n^{-1/2}\alpha_1(N(n) - n) \xrightarrow{L^2} 0. \tag{6.75}$$

Proof. The proof of the first part (6.74) follows that of equation (4.5) of [111], using our Lemma B.2.3 and the fact that the functional $\mathcal{L}^1$ is homogeneous of order 1, is strongly stabilizing by Lemma 6.2.1, and satisfies the moments condition equation (3.5) by Lemma 6.2.3.

As shown in the proof of Corollary 6.2.1 (see in particular equation (6.14)), we have that $\mathcal{L}^1(\mathcal{P}_n^0; S_{0,n}) - \mathcal{L}^1(\mathcal{P}_n; S_{0,n})$ converges to zero in L^2 and $\mathcal{L}^1(\mathcal{U}_n^0; S_{0,n}) - \mathcal{L}^1(\mathcal{U}_n; S_{0,n})$ converges to zero in L^2. Therefore the second part (6.75) follows from (6.74). □

We are now in a position to prove Theorem 6.1.1. We divide the proof into two cases: $\alpha \neq 1$ and $\alpha = 1$. In the latter case, to prove the result for the Poisson process $\mathcal{P}_n$, we need to show that $\mathcal{L}^1(\mathcal{P}_n; B_n)$ and $\mathcal{L}^1(\mathcal{P}_n; S_{0,n})$ are asymptotically independent; likewise for $\mathcal{P}_n^0$. We shall then obtain the results for the binomial process $\mathcal{U}_n$ and for $\mathcal{U}_n^0$ from those for $\mathcal{P}_n$ and $\mathcal{P}_n^0$ via the coupling described in Lemma 6.3.4.

Proof of Theorem 6.1.1 for $\alpha \neq 1$. First suppose $0 < \alpha < 1$. For the Poisson case, we have

$$n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n) = n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; S_{0,n}) + n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; B_n) + n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; C_n). \tag{6.76}$$

The first term in the right hand side of (6.76) converges in distribution to $\mathcal{N}(0, s_\alpha^2)$ by Theorem 6.2.1 (iv), and the other two terms converge in probability to 0 by (6.28) and (6.61). Thus Slutsky's theorem yields the first (Poisson) part of (6.6). To obtain the second (binomial) part of (6.6), we use the coupling of Lemma 6.3.4. We write

$$n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n) = n^{(\alpha-1)/2}\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n; S_{0,n}) + n^{(\alpha-1)/2}(\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; B_n \cup C_n)) + n^{(\alpha-1)/2}(\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n; B_n \cup C_n) - \tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; B_n \cup C_n)). \tag{6.77}$$

The first term in the right side of (6.77) is asymptotically $\mathcal{N}(0, t_\alpha^2)$ by Theorem 6.2.1 (ii). The second term tends to zero in probability by (6.28) and (6.61). The third term tends to zero in probability by (6.47). Thus we have the binomial case of (6.6).

The rooted case (6.3) is similar. Now, for the first (Poisson) part of (6.3), we use Corollary 6.2.1 (iv) with (6.29) and (6.62), and Slutsky's theorem. The second part of (6.3) follows from the analogous statement to (6.77) with the addition of the origin, using Corollary 6.2.1 (ii) with (6.29), (6.62), (6.48), and Slutsky's theorem again.

Next, suppose $\alpha > 1$. We have

$$\tilde{\mathcal{L}}^\alpha(\mathcal{P}_n) = \tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; S_{0,n}) + \tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; C_n) + \tilde{\mathcal{L}}^\alpha(\mathcal{P}_n; B_n). \tag{6.78}$$

The first term in the right hand side converges to 0 in probability, by Theorem 6.2.1 (iii). The second term also converges to 0 in probability, by the first part of (6.60). Then by (6.26) and Slutsky's theorem, we obtain the first (Poisson) part of (6.8). To obtain the rooted version, i.e. the first part of (6.5), we replace $\mathcal{P}_n$ by $\mathcal{P}_n^0$ in (6.78), and combine (6.24) with Corollary 6.2.1 (iii) and the second part of (6.60), and apply Slutsky's theorem again.

To obtain the binomial versions of the results (6.5) and (6.8), we again make use of the coupling described in Lemma 6.3.4. We have

$$\tilde{\mathcal{L}}^\alpha(\mathcal{U}_n) = \tilde{\mathcal{L}}^\alpha(\mathcal{U}_n; S_{0,n}) + \tilde{\mathcal{L}}^\alpha(\mathcal{U}_n; C_n) + \tilde{\mathcal{L}}^\alpha(\mathcal{U}_n; B_n). \tag{6.79}$$

The first term in the right hand side converges in probability to zero by Theorem 6.2.1 (i). The second term converges in probability to zero by the first part of (6.60) and the first part of (6.72). The third part converges in distribution to $\tilde{F}_\alpha^{\{1\}} + \tilde{F}_\alpha^{\{2\}}$ by by (6.27). Hence, Slutsky's theorem yields the binomial part of (6.8).

Similarly, by replacing $\mathcal{P}_n$ by $\mathcal{P}_n^0$ and $\mathcal{U}_n$ by $\mathcal{U}_n^0$ in (6.79), and using Corollary 6.2.1 (i), the second part of (6.60) and of (6.72), (6.25) and Slutsky's theorem, we obtain the binomial part of (6.5). This completes the proof for $\alpha \neq 1$.

Proof of Theorem 6.1.1 for $\alpha = 1$: the Poisson case. We now prove the first part of (6.4) and the first part of (6.7). Given n, set $q_n := 4\lfloor n^{\varepsilon+\sigma-1/2} \rfloor$. Split each cell Γ_i^x of C_n^x into $4q_n$ rectangular sub-cells, by splitting the horizontal edge into q_n segments and the vertical edge into 4 segments by a rectangular grid. Similarly, split each cell Γ_i^y by splitting the vertical edge into q_n segments and the horizontal edge into 4 segments. Finally, add a single square sub-cell in the top right-hand corner of C_n^0, of side $(1/4)n^{\varepsilon-1/2}$, and denote this "the corner sub-cell".

The total number of all such sub-cells is $1 + 8k_n q_n \sim 32n^{(1/2)-\varepsilon}$. Each of the sub-cells has width asymptotic to $(1/4)n^{\varepsilon-1/2}$ and height asymptotic to $(1/4)n^{\varepsilon-1/2}$, and so the area of each cell is asymptotic to $(1/16)n^{2\varepsilon-1}$. So for large n, for each of these sub-cells, the probability that it contains no point of $\mathcal{P}_n$ is bounded by $\exp(-n^\varepsilon)$.

Let E_n be the event that each of the sub-cells described above contains at least one point of $\mathcal{P}_n$. Then

$$P[E_n^c] = O\left(n^{(1/2)-\varepsilon}\exp(-n^\varepsilon)\right) \to 0. \tag{6.80}$$

Suppose $\mathbf{x}$ lies on the lower boundary of $S_{0,n}$. Consider the rectangular sub-cell of Γ_i^x lying just to the left of the sub-cell directly below $\mathbf{x}$ (or the corner sub-cell if that lies just to the left of the sub-cell directly below $\mathbf{x}$). All points $\mathbf{y}$ in this sub-cell satisfy $\mathbf{y} \preccurlyeq^* \mathbf{x}$, and for large n, satisfy $\|\mathbf{y}-\mathbf{x}\| < (3/4)n^{\varepsilon-1/2}$, whereas the nearest point to $\mathbf{x}$ in B_n is at a distance at least $(3/4)n^{\varepsilon-1/2}$. Arguing similarly for $\mathbf{x}$ on the left boundary of $S_{0,n}$, and using the triangle inequality, we see that if E_n occurs, no point in $S_{0,n}$ can be connected to any point in B_n, provided n is sufficiently large.

For simplicity of notation, set $X_n := \tilde{\mathcal{L}}^1(\mathcal{P}_n; B_n)$ and $Y_n := \tilde{\mathcal{L}}^1(\mathcal{P}_n; S_{0,n})$. Also, set $X := \tilde{D}_1^{\{1\}} + \tilde{D}_1^{\{2\}}$ and $Y \sim \mathcal{N}(0, s_1^2)$, independent of X, with s_1 as given in Theorem 6.2.1. We know from Theorem 6.3.1 and Theorem 6.2.1 that $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathcal{D}} Y$ as $n \to \infty$.

We need to show that $X_n + Y_n \xrightarrow{\mathcal{D}} X + Y$, where X and Y are independent random variables. We show this by convergence of the characteristic function,

$$E[\exp(it(X_n+Y_n))] \longrightarrow E[\exp(itX)]E[\exp(itY)]. \tag{6.81}$$

With ω denoting the configuration of points in C_n, we have

$$\begin{aligned} E\left[\exp(it(X_n+Y_n))\right] &= \int_{E_n} E\left[e^{itX_n}e^{itY_n}\middle|\,\omega\right] \mathrm{d}P(\omega) + E\left[e^{it(X_n+Y_n)}\mathbf{1}_{E_n^c}\right] \\ &= \int_{E_n} E\left[e^{itX_n}\right] E\left[e^{itY_n}\middle|\,\omega\right] \mathrm{d}P(\omega) + E\left[e^{it(X_n+Y_n)}\mathbf{1}_{E_n^c}\right], \end{aligned}$$

where we have used the fact that X_n and Y_n are conditionally independent, given $\omega \in E_n$, for n sufficiently large, and that X_n is independent of the configuration in C_n. Then $E[e^{it(X_n+Y_n)}\mathbf{1}_{E_n^c}] \to 0$ as $n \to \infty$, since $P[E_n^c] \to 0$. So

$$E\left[\exp(it(X_n+Y_n))\right] - E\left[e^{itX_n}\right] E\left[e^{itY_n}\mathbf{1}_{E_n}\right] \to 0,$$

and we obtain (6.81) since $E[e^{itY_n}\mathbf{1}_{E_n}] = E[e^{itY_n}] - E[e^{itY_n}\mathbf{1}_{E_n^c}]$, $E[e^{itY_n}\mathbf{1}_{E_n^c}] \to 0$, $E[e^{itX_n}] \to E[e^{itX}]$, and $E[e^{itY_n}] \to E[e^{itY}]$ as $n \to \infty$.

We can now prove the first (Poisson) part of (6.7). We have the $\alpha = 1$ case of (6.78). The contribution from C_n converges in probability to 0 by the first part of (6.60). Slutsky's theorem and (6.81) then give the first (Poisson) part of (6.7). The rooted Poisson case (6.4) follows from the rooted version of (6.78), this time applying the argument for (6.81) taking $X_n := \tilde{\mathcal{L}}^1(\mathcal{P}_n^0; B_n)$, $Y_n := \tilde{\mathcal{L}}^1(\mathcal{P}_n^0; S_{0,n})$ and X, Y as before, and then using the second part of (6.60) and Slutsky's theorem again. Thus we obtain the first (Poisson) part of (6.4).

Proof of Theorem 6.1.1 for $\alpha = 1$: the binomial case. It remains for us to prove the second part of (6.4) and the second part of (6.7). To do this, we use the coupling of Lemma 6.3.4 once more. Considering first the unrooted case, we here set $X_n := \mathcal{L}^1(\mathcal{U}_n; B_n)$ and $Y_n := \mathcal{L}^1(\mathcal{U}_n; S_{0,n})$. Set $X_n' := \mathcal{L}^1(\mathcal{P}_n; B_n)$ and $Y_n' := \mathcal{L}^1(\mathcal{P}_n; S_{0,n})$ (note that all these random variables are uncentred).

Set $Y \sim \mathcal{N}(0, s_1^2)$ with s_1 as given in Theorem 6.2.1. Set $X := \tilde{D}_1^{\{1\}} + \tilde{D}_1^{\{2\}}$, independent of Y. Then by equation (6.81) we have (in our new notation)

$$X_n' - EX_n' + Y_n' - EY_n' \xrightarrow{\mathcal{D}} X + Y. \tag{6.82}$$

By (6.49), we have $X_n - X_n' \xrightarrow{P} 0$ and $EX_n - EX_n' \to 0$. Also, with α_1 as defined in Lemma 6.4.4, equation (6.74) of that result gives us

$$Y_n' - Y_n - n^{-1/2}\alpha_1(N(n) - n) \xrightarrow{L^2} 0 \tag{6.83}$$

so that $E[Y_n'] - E[Y_n] \to 0$. Combining these observations with (6.82), and using Slutsky's theorem, we obtain

$$X_n - EX_n + Y_n - EY_n + n^{-1/2}\alpha_1(N(n) - n) \xrightarrow{\mathcal{D}} X + Y. \tag{6.84}$$

By Theorem 6.2.1 (iii) we have $\mathrm{Var}(Y_n') \to s_1^2$ as $n \to \infty$. By (6.83), and the independence of $N(n)$ and Y_n, we have

$$s_1^2 = \lim_{n\to\infty} \mathrm{Var}[Y_n + n^{-1/2}\alpha_1(N(n) - n)] = \lim_{n\to\infty} (\mathrm{Var}[Y_n] + \alpha_1^2)$$

so that $\alpha_1^2 \leq s_1^2$. Also, $n^{-1/2}\alpha_1(N(n) - n)$ is independent of $X_n + Y_n$, and asymptotically $\mathcal{N}(0, \alpha_1^2)$. Since the $\mathcal{N}(0, s^2)$ characteristic function is $\exp(-s^2t^2/2)$, for all $t \in \mathbf{R}$ we obtain from equation (6.84) that

$$E[\exp(it(X_n - EX_n + Y_n - EY_n))] \to \exp(-(s_1^2 - \alpha_1^2)t^2/2)E[\exp(itX)]$$

so that

$$X_n - EX_n + Y_n - EY_n \xrightarrow{\mathcal{D}} X + W, \tag{6.85}$$

where $W \sim \mathcal{N}(0, s_1^2 - \alpha_1^2)$, and W is independent of X.

We have the $\alpha = 1$ case of (6.79). By the first part of equation (6.60) and the first part of (6.72), the contribution from C_n tends to zero in probability. Hence by (6.85) and Slutsky's theorem, we obtain the second (binomial) part of (6.7).

For the rooted case, we apply the argument for (6.85), now taking $X_n := \mathcal{L}^1(\mathcal{U}_n^0; B_n)$, $Y_n := \mathcal{L}^1(\mathcal{U}_n^0; S_{0,n})$, with X, Y and W as before. The rooted case of (6.82) follows from the rooted case of (6.81), and now we have $X_n - X_n' \xrightarrow{P} 0$ and $EX_n - EX_n' \to 0$ by (6.50). In the rooted case (6.83) still holds by (6.75), and then we obtain the rooted case of (6.85) as before.

To obtain the second (binomial) part of (6.4), we start with the rooted version of the $\alpha = 1$ case of (6.79). By the second part of equation (6.60) and of (6.72), the contribution from C_n tends to zero in probability. Hence by the rooted version of (6.85) and Slutsky's theorem, we obtain the second part of (6.4).

This completes the proof of the $\alpha = 1$ case, and hence the proof of Theorem 6.1.1 is complete. □

Chapter 7

Conclusions and discussion

7.1 Discussion of Chapters 2-6

7.1.1 Vertex degrees

The question of vertex degress in random spatial graphs is of considerable interest with respect to models of networks, and the world wide web in particular (see, for example, [28, 44, 103]). Questions of interest include the degree distribution of the MDST, DLT and ONG. Some results on degrees in the ONG appear in [19]. Degrees of vertices in the MDST are of particular interest, for while there is no uniform upper bound on vertex degrees, simulations suggest that typically the vast majority of vertices have very small degree (less than 3).

7.1.2 The on-line nearest-neighbour graph

In this thesis, we study the total weight of the ONG. We give laws of large numbers for the ONG in $\mathbf{R}^d$ for $d \in \mathbf{N}$ (see Chapter 2), and weak convergence results for the ONG in the unit interval $(0, 1)$ (see Chapter 5). It would be of interest to obtain weak convergence results for general d. In [106] it was shown that there is a central limit theorem provided $\alpha < d/4$. Thus the question as to whether a central limit theorem holds for the total length (suitably centred and scaled) of the ONG (i.e. $\alpha = 1$) is resolved in the affirmative for $d = 5, 6, 7, \ldots$, while in $d = 1$ the limit is non-normal, by Theorem 5.2.2. As stated in [106], it is likely that a technical refinement of the methods of [106] could be achieved, leading to a CLT also in $d = 3$ and $d = 4$. The case $d = 2$ seems to need some new technique.

In $d = 1$, for $\alpha \leq 1/2$ we have Conjecture 5.2.1. The proof of this conjecture seems within reach using the 'contraction method' of Chapter 5. One might hope that the $d = 1$, $\alpha = 1/2$ result conjectured in Conjecture 5.2.1 would shed some light on the $d = 2$, $\alpha = 1$ case. So one might further conjecture that

$$(\log n)^{-1/2}\tilde{\mathcal{O}}^{2,1}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2),$$

for some $0 < \sigma < \infty$. On the other hand, the one dimensional case appears to be quite special, and simulations suggest that the correct result may instead be

$$\tilde{\mathcal{O}}^{2,1}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2),$$

for some $0 < \sigma < \infty$. The resolution of this question appears to require a new idea, which may help with other open questions in geometrical probability.

The question is also of interest in higher dimensions. For example, we might conjecture the following.

Conjecture 7.1.1 *Suppose $\alpha > 0$. Suppose $d \in \mathbf{N}$.*

(i) For $0 < \alpha < d/2$, there exists $0 < s_\alpha < \infty$ such that as $n \to \infty$

$$n^{(\alpha-d)/(2d)}\tilde{\mathcal{O}}^{d,\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0,s_\alpha^2).$$

(ii) For $\alpha > d/2$, there exists a random variable $Q(d,\alpha)$ such that as $n \to \infty$

$$\tilde{\mathcal{O}}^{d,\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} Q(d,\alpha).$$

Part (i) of Conjecture 7.1.1 has been shown to hold for $0 < \alpha < d/4$ in Theorem 3.4 of [106]. As stated for $d = 1$, this is Theorem 5.2.2 with Conjecture 5.2.1. Part (ii) holds for $d \in \mathbf{N}$ and $\alpha > d$ (see Theorem 2.3.1) with $Q(d,\alpha) = W(d,\alpha) - E[W(d,\alpha)]$. It also holds for $d = 1$ as stated (see Theorem 5.2.2 (ii), (iii), (iv)). The case $\alpha = d/2$ is unclear for general d; see the above discussion.

7.1.3 Further MDSF limit theorems

For the total weight of the MDSF on $\mathcal{U}_n$ or $\mathcal{P}_n$ in $(0,1)^2$, we have complete weak limit theorems under partial orders $\preccurlyeq^*$ and $\preccurlyeq_*$ (see Chapter 6). Our corresponding law of large numbers (Theorem 2.4.1) covers the family of partial orders $\overset{\theta,\phi}{\preccurlyeq}$ with $0 \leq \theta < 2\pi$ and

$0 < \phi \leq \pi$. It should be possible to obtain weak convergence results for these general partial orders also. We believe that in almost all cases, boundary effects are insignificant, and the limit is purely normal. For some particular partial orders (including $\preccurlyeq^*$ and $\preccurlyeq_*$, for example), boundary effects will be significant.

There are essentially four types of partial order $\overset{\theta,\phi}{\preccurlyeq}$. We classify these as follows.

Definition 7.1.1 *We classify the partial order* $\overset{\theta,\phi}{\preccurlyeq}$, $\theta \in [0, 2\pi)$, $\phi \in (0, \pi]$, *as follows.*

(a) We say the partial order is Type A *if* $\theta \in \{0, \pi/2, \pi, 3\pi/2\}$ *and* $\phi = \pi/2$.

(b) We say the partial order is Type B *if* $\theta \in \{0, \pi/2, \pi, 3\pi/2\}$ *and* $\phi = \pi$.

(c) We say the partial order is of Type C *if* $\phi \in (0, \pi/2) \cup (\pi/2, \pi)$ *and exactly one of* $\theta + \phi, \theta$ *belongs to* $\{0, \pi/2, \pi, 3\pi/2, 2\pi, 5\pi/2\}$.

(d) We say the partial order is of Type D *if it is not of Type A,B, or C.*

Note that by symmetry, it suffices to consider only particular cases of Type A, B and C partial orders as described above. Thus it suffices to consider $\preccurlyeq^*$ for Type A, $\preccurlyeq_*$ for Type B, and the case $\theta = \pi/2$ of Type C above.

Note that when the partial order is Type B (the "one-coordinate" case), there is almost surely a single minimal element of $\mathcal{U}_n$ or $\mathcal{P}_n$, so the MDSF is almost surely an MDST (see Figure 2.5 for an example).

Weak convergence results for Type A and B partial orders are covered by Theorem 6.1.1. The following conjecture gives corresponding results for Type C and D partial orders. We state the binomial case $\mathcal{U}_n$ only.

Conjecture 7.1.2 *Suppose the weight exponent is* $\alpha > 0$ *and the partial order is* $\overset{\theta,\phi}{\preccurlyeq}$.

(i) Suppose the partial order is Type C. Then there exists $t_\alpha^2 > 0$ *such that, for a normal random variable* $W_\alpha \sim \mathcal{N}(0, t_\alpha^2)$, *as* $n \to \infty$*:*

$$n^{(\alpha-1)/2} \tilde{\mathcal{L}}^{2,\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} W_\alpha \quad (0 < \alpha < 1);$$
$$\tilde{\mathcal{L}}^{2,1}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \tilde{D}_1 + W_1;$$
$$\tilde{\mathcal{L}}^{2,\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} \tilde{F}_\alpha \quad (\alpha > 1).$$

Here all the random variables in the limits are independent, $\tilde{D}_1$ *with distribution defined at (5.36), and, for* $\alpha > 1$, $\tilde{F}_\alpha$ *with distribution defined at (5.38).*

(ii) *Suppose the partial order is Type D. Then there exists $t_\alpha^2 > 0$ such that, for a normal random variable $W_\alpha \sim \mathcal{N}(0, t_\alpha^2)$, as $n \to \infty$:*

$$n^{(\alpha-1)/2} \tilde{\mathcal{L}}^{2,\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} W_\alpha \quad (\alpha > 0);.$$

The normal random variables W_α arise from the edges away from the boundary. Part (ii) says that boundary effects are insignificant under Type D partial orders. Part (i) says that significant boundary effects arise (for $\alpha \geq 1$) from close to a single boundary under Type C partial orders, where the MDSF is close to the DLT considered in Chapter 5. It should be possible to prove Conjecture 7.1.2 using the same methods as the proof of Theorem 6.1.1. We aim to address this in future work.

It is also of interest to obtain weak convergence results for the total weight of the MDSF on $\mathcal{U}_n$ or $\mathcal{P}_n$ in $(0,1)^d$, $d \geq 3$. Here, the most natural partial orders to consider are $\preccurlyeq_*$ and $\preccurlyeq^*$.

Consider $\preccurlyeq_*$. It should be possible, using the methods of Chapter 6, to show that a central limit theorem holds for edges away from the 'boundary', that now being the $d-1$ dimensional face of the unit d-cube in the 'downwards' direction according to $\preccurlyeq_*$. The MDSF close to this boundary, however, should be 'close' to a $d-1$-dimensional ONG on the boundary. Thus, if Conjecture 7.1.1 holds, we might conjecture the following (again, we state the binomial case $\mathcal{U}_n$ only).

Conjecture 7.1.3 *Suppose $\alpha > 0$, $d \in \mathbf{N}$, and the partial order is $\preccurlyeq_*$. Then there exists a constant $0 < t_\alpha^2 < \infty$ such that, for normal random variables $W_\alpha \sim \mathcal{N}(0, t_\alpha^2)$, as $n \to \infty$:*

$$n^{(2\alpha-d)/(2d)} \tilde{\mathcal{L}}^{d,\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} W_\alpha \quad (0 < \alpha < d/2);$$
$$\tilde{\mathcal{L}}^{d,d/2}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} W_1 + Q(d-1, d/2);$$
$$\tilde{\mathcal{L}}^{d,\alpha}(\mathcal{U}_n) \xrightarrow{\mathcal{D}} Q(d-1, \alpha) \quad (\alpha > d/2).$$

Here all the random variables in the limits are independent, and $Q(d, \alpha)$ is the limit arising in Conjecture 7.1.1 (ii).

We may also conjecture a similar result for the MDSF under $\preccurlyeq^*$, but in this case we would expect the boundary effects to be manifest as d independent contributions from on-line versions of the MDSF (under $\preccurlyeq^*$) on $(0,1)^{d-1}$ (which we have not studied in this thesis). We hope to address some of these questions in future work.

Chapter 8

Random walk in random environment with asymptotically zero perturbation

8.1 Introduction

In this chapter we study a problem with a classical flavour that lies in the intersection of two well-studied problems, those of random walks in one-dimensional random environments and Markov chains with asymptotically small drifts. Separately, these two problems have received considerable attention, but the problem considered in this chapter has not been analysed before. Further, our results show that the system studied here exhibits behaviour that is significantly different to that of those previously studied systems.

The random walk in random environment (or RWRE for short) was first studied by Kozlov [87] and Solomon [133], and has since received extensive attention; see for example [121] or [143] for surveys. Here we analyse the behaviour of the RWRE for which the random environment is perturbed by a vanishingly small amount.

The analysis of zero drift random walks in two or more dimensions by the method of Lyapunov functions demonstrated the importance of the investigation of one-dimensional stochastic processes with asymptotically small drifts (see [10]). For example, if (Z_t), with $t = 0, 1, 2, 3\ldots$ time is a random walk (with zero drift) in the nonnegative quarter plane, analysis of the stochastic process $\|Z_t\|$, where $\|\cdot\|$ denotes the Euclidean norm, involves the study of stochastic processes on the half-line with mean drift asymptotically zero.

Early work in this field was done by Lamperti [90, 91]. Criteria for recurrence and transience are given in [98], where the behaviour in the critical regime that Lamperti did not cover was also analysed. Passage-time moments are considered in [10]. In much of this work, Lyapunov functions play a central role.

In this chapter we demonstrate the essential difference between a nearest-neighbour random walk in a deterministic environment, perturbed from its critical (null-recurrent) regime, and a nearest-neighbour random walk in a random environment, also perturbed from its critical regime (sometimes called Sinai's regime – see below). Our results quantify the fact that in some sense the random environment is more stable, in that a much larger perturbation is required to disturb the null-recurrent situation. In particular, we give criteria for ergodicity (i.e. positive recurrence here), transience and null-recurrence for our perturbed random walk in random environment. We will show that in our (random environment) case the critical magnitude for the perturbation is of the order of $n^{-1/2}$ (see Theorem 8.2.6), where n is the distance from the origin (in fact, our more general results are much more precise than this). This compares to a critical magnitude of the order of n^{-1} in the non-random environment case (see [98], and Theorem 8.2.2 below).

Our method is based upon the theory of Lyapunov functions, a powerful tool in the classification of countable Markov chains (see [50]). Such methods have proven effective in the analysis of random walks in random environments (see e.g. [34]), in addition to Markov chains in non-random environments.

Loosely speaking, motivation for our model comes from some one-dimensional physical systems, such as a particle performing a random walk in a homogeneous random one-dimensional field, subject to some vanishing perturbation (such as the presence of another particle). Under what conditions is the perturbation sufficient to alter the character of the random walk?

We now introduce the probabilistic model that we consider. First, we need some notation. We introduce the function χ as follows, which determines our perturbation as described below. Let $\chi : [0, \infty) \to [0, \infty)$ be a function such that

$$\lim_{x \to \infty} \chi(x) = 0. \tag{8.1}$$

As we shall see below, the property (8.1) means that our perturbation is asymptotically small.

Here, we are interested in the one-dimensional RWRE on the nonnegative integers (we use the notation $\mathbf{Z}^+ := \{0, 1, 2, \ldots\}$), with reflection at the origin. One can readily obtain results for the one-dimensional RWRE on whole of $\mathbf{Z}$ in a similar manner. Formally, we define our RWRE as follows.

We define sequences of random variables ξ_i, $i = 1, 2, \ldots$ and Y_i, $i = 1, 2, \ldots$, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with the following properties.

Fix ε such that $0 < \varepsilon < 1/2$. Let ξ_i, $i = 1, 2, \ldots$, be a sequence of i.i.d. random variables such that

$$\mathbb{P}[\varepsilon \leq \xi_1 \leq 1 - \varepsilon] = 1. \tag{8.2}$$

The condition (8.2) is sometimes referred to as *uniform ellipticity.*

Let Y_i, $i = 1, 2, \ldots$, be another sequence of i.i.d. random variables taking values in $[-1, 1]$, on the same probability space as the ξ_i. We allow Y_i to depend on ξ_i, but any collections $(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k})$, $(\xi_{j_1}, \xi_{j_2}, \ldots, \xi_{j_{k'}})$ are independent if $\{i_1, \ldots, i_k\} \cap \{j_1, \ldots, j_{k'}\} = \emptyset$.

For a particular realization of the sequences $(\xi_i; i = 1, 2, \ldots)$ and $(Y_i; i = 1, 2, \ldots)$, we define the quantities p_n and q_n, $n = 1, 2, 3, \ldots$ as follows:

$$\begin{aligned} p_n &:= \begin{cases} \xi_n + Y_n \chi(n) & \text{if} \quad \varepsilon/2 \leq \xi_n + Y_n \chi(n) \leq 1 - (\varepsilon/2) \\ \varepsilon/2 & \text{if} \quad \xi_n + Y_n \chi(n) < \varepsilon/2 \\ 1 - (\varepsilon/2) & \text{if} \quad \xi_n + Y_n \chi(n) > 1 - (\varepsilon/2) \end{cases} \\ q_n &:= 1 - p_n. \end{aligned} \tag{8.3}$$

We call a particular realization of (p_n, q_n), $n = 1, 2, \ldots$, our *environment*, and we denote it by ω. A given ω is then a realization of our random environment, and is given in terms of the ξ_i and Y_i as in (8.3).

For a given environment ω, that is, a realization of (p_n, q_n), $n = 1, 2, \ldots$, we define the Markov chain $(\eta_t(\omega); t \in \mathbf{Z}^+)$ on $\mathbf{Z}^+$, starting at some point in $\mathbf{Z}^+$, defined as follows: $\eta_0(\omega) = r$ for some $r \in \mathbf{Z}^+$, and for $n = 1, 2, \ldots$,

$$\begin{aligned} P[\eta_{t+1}(\omega) = n - 1 | \eta_t(\omega) = n] &= p_n, \\ P[\eta_{t+1}(\omega) = n + 1 | \eta_t(\omega) = n] &= q_n, \end{aligned} \tag{8.4}$$

and $P[\eta_{t+1}(\omega) = 1 | \eta_t(\omega) = 0] = 1/2$, $P[\eta_{t+1}(\omega) = 0 | \eta_t(\omega) = 0] = 1/2$. The given form for the reflection at the origin ensures that the Markov chain is *aperiodic*, which eases some technical complications, but this choice is not special; it can be changed without affecting our results.

Recall that, from (8.1), $\chi(n) \to 0$ as $n \to \infty$. Thus, there exists $n_0 \in (0, \infty)$ such that, for all $n \geq n_0$, $\chi(n) < \varepsilon/2$. Hence, under condition (8.2), for $\mathbb{P}$-almost all $(\xi_n)_{n \geq n_0}$

we have $(\varepsilon/2) < \xi_n + Y_n\chi(n) < 1-(\varepsilon/2)$ (since the Y_n are bounded). Thus, for all $n \geq n_0$, (8.3) implies that, $\mathbb{P}$-almost surely,

$$p_n = \xi_n + Y_n\chi(n), \quad q_n = 1 - \xi_n - \chi(n)Y_n, \quad n \geq n_0. \tag{8.5}$$

Note that our conditions on the variables in (8.3) ensure that $(\varepsilon/2) \leq p_n \leq 1-(\varepsilon/2)$ almost surely for all n, so that for almost every (a.e.) environment, p_n and q_n are true probabilities bounded strictly away from 0 and from 1.

For $n = 1, 2, \ldots$, we set

$$\zeta_n := \log\left(\frac{\xi_n}{1-\xi_n}\right). \tag{8.6}$$

Write $\mathbb{E}$ for expectation under $\mathbb{P}$.

In our model, under (8.1), $\chi(n) \to 0$ as $n \to \infty$. Thus, from (8.5), in the limit $n \to \infty$, we approach the well-known random walk in i.i.d. random environment as studied in [87], [133] and subsequently. In addition, when $\mathbb{E}[\zeta_1] = 0$, in the limit as $n \to \infty$ we approach the critical case often referred to as *Sinai's regime* after [132]. Our results show that despite this, the behaviour of our model is, in general, very different to the behaviour of these limiting cases, depending on the nature of the perturbation χ.

In the next section we state our results. Theorems 8.2.1, 8.2.2, and 8.2.3 are special cases of the model in which some of the random variables ξ_i and Y_i are degenerate (that is, equal to a constant almost surely). In particular, Theorems 8.2.1 and 8.2.2 include some known results, when our model reduces to previously studied systems. In Theorem 8.2.4, the underlying environment is not in the 'critical regime'. Our main results, Theorems 8.2.6 and 8.2.7, deal with the main case of interest, in which the underlying environment is truly random and is, in a sense to be demonstrated, critical.

8.2 Main results

Most of our results will be formulated for almost all environments ω (in some sense, for all 'typical' environments), that is $\mathbb{P}$-almost surely over $(\Omega, \mathcal{F}, \mathbb{P})$.

If $Y_1 = 0$ $\mathbb{P}$-a.s., then our model reduces to the standard reflected one-dimensional random walk in an i.i.d. random environment. In this case $p_n = \xi_n$ and $q_n = 1 - \xi_n$, $n = 1, 2, \ldots$, and so (with the definition at (8.6)) $\zeta_n = \log(p_n/q_n)$. Criteria for recurrence of the RWRE $\eta_t(\omega)$ in this case were given by Solomon [133], for the case where $(\xi_i; i =$

$1, 2, \ldots$) is an i.i.d. sequence, and generalised by Alili [5]. For the case in which larger jumps are permitted, see, for example, [80].

The following well-known result dates back to Solomon [133].

Theorem 8.2.1 *Let* $(\eta_t(\omega); t \in \mathbb{Z}^+)$ *be the random walk in i.i.d. random environment, with* $\mathbb{P}[Y_1 = 0] = 1$. *Suppose* $\mathrm{Var}[\zeta_1] > 0$.

(i) If $\mathbb{E}[\zeta_1] < 0$, *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. transient.*

(ii) If $\mathbb{E}[\zeta_1] = 0$, *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. null-recurrent.*

(iii) If $\mathbb{E}[\zeta_1] > 0$, *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. ergodic.*

The critical (null-recurrent) regime $\mathbb{E}[\log(p_1/q_1)] = 0$ is known as *Sinai's regime*, after [132]. This regime has been extensively studied; see, for example, [35, 72, 79, 81]. For an outline proof of Theorem 8.2.1 using Lyapunov function methods, similar to those employed in this chapter, see Theorem 3.1 of [34]. In this chapter we extend the classification criteria of Theorem 8.2.1 to encompass the case in which the p_n are not i.i.d. and in which $\mathbb{E}[\log(p_n/q_n)]$ is *asymptotically* zero, as $n \to \infty$. Our results are, in some sense, a random environment analogue of those for Markov processes with asymptotically zero mean drift given in [98] (see below).

For the remainder of the chapter we suppose $\mathbb{P}[Y_1 = 0] < 1$. This includes the interesting case where $Y_1 = b$ $\mathbb{P}$-a.s., for some $b \in [-1, 1] \setminus \{0\}$. Our techniques do, however, enable us to allow Y_1 to be random.

Although not as famous as the RWRE, another system that has been well studied is the rather classical problem of a Markov chain with asymptotically zero drift. This problem was studied by Lamperti [90, 91]. General criteria for recurrence, transience and ergodicity were given by Menshikov, Asymont, and Iasnogorodskii in [98].

Theorem 8.2.2 below is a consequence of their main result, Theorem 3, applied to our problem when $\mathrm{Var}[\zeta_1] = 0$ and $\mathrm{Var}[Y_1] = 0$; that is, the distributions of ξ_1 and Y_1 are both degenerate (i.e. equal to a constant almost surely). In particular, we have a *non-random* environment ω. If, on the other hand, ξ_1 is degenerate but Y_1 is not, then we have a random (asymptotically small) perturbation on an underlying non-random environment, and we have Theorem 8.2.3 below.

We use the notation $\log_1(x) := \log(x)$ and $\log_k(x) := \log(\log_{k-1}(x))$ for $k = 2, 3, \ldots$.

Theorem 8.2.2 *Suppose* $\mathbb{P}[Y_1 = b] = 1$ *for some* $b \in [-1, 0) \cup (0, 1]$. *Suppose* $\mathbb{P}[\xi_1 = c] = 1$ *for some* $c \in (0, 1)$.

(i) If $c < 1/2$, *then* $\eta_t(\omega)$ *is transient.*

(ii) If $c > 1/2$, *then* $\eta_t(\omega)$ *is ergodic.*

(iii) Suppose $c = 1/2$. *Suppose there exist* $s \in \mathbf{Z}^+$ *and* $K \in \mathbf{N}$ *such that, for all* $n \in [K, \infty)$ *and some* $h > 1$ *the following inequality holds:*

$$b\chi(n) > \frac{1}{4n} + \frac{1}{4n \log n} + \cdots + \frac{h}{4n \prod_{i=1}^{s} \log_i n}. \tag{8.7}$$

Then $\eta_t(\omega)$ *is ergodic.*

(iv) Suppose $c = 1/2$. *Suppose there exist* $s, t \in \mathbf{Z}^+$ *and* $K \in \mathbf{N}$ *such that, for all* $n \in [K, \infty)$ *and some* $h < 1$ *the following inequality holds:*

$$\begin{aligned} -\frac{1}{4n} - \frac{1}{4n \log n} - \cdots - \frac{h}{4n \prod_{i=1}^{s} \log_i n} &\leq b\chi(n) \\ &\leq \frac{1}{4n} + \frac{1}{4n \log n} + \cdots + \frac{h}{4n \prod_{i=1}^{t} \log_i n}. \end{aligned} \tag{8.8}$$

Then $\eta_t(\omega)$ *is null-recurrent.*

(v) Suppose $c = 1/2$. *Suppose there exist* $s \in \mathbf{Z}^+$ *and* $K \in \mathbf{N}$ *such that, for all* $n \in [K, \infty)$ *and some* $h > 1$ *the following inequality holds:*

$$b\chi(n) \leq -\frac{1}{4n} - \frac{1}{4n \log n} - \cdots - \frac{h}{4n \prod_{i=1}^{s} \log_i n}. \tag{8.9}$$

Then $\eta_t(\omega)$ *is transient.*

Theorem 8.2.2 follows directly by applying Theorem 3 of [98] to our case, with $m(x) = -2\chi(x)$ and $b(x) = 1$.

Remark. In the case $c = 1/2$ the critical case in terms of the recurrence, transience and ergodicity is when the perturbation $\chi(n)$ is, ignoring logarithmic terms, of order n^{-1}; we say that the 'critical exponent' is -1. This contrasts with our results in the case where $\text{Var}[\xi_1] > 0$ (see Theorems 8.2.6 and 8.2.7), in which the critical exponent is $-1/2$.

The following result deals with the case in which the distribution of ξ_1 is degenerate, but that of Y_1 is not; in this case we have a homogeneous non-random environment subject to an asymptotically small random perturbation. In particular, parts (iii) and (iv) of the theorem deal with the case when the underlying environment is that of the simple random walk. Here, $\stackrel{\mathcal{D}}{=}$ stands for equality in distribution.

Theorem 8.2.3 *Suppose* $\mathbb{P}[\xi_1 = c] = 1$ *for some* $c \in (0,1)$, *and* $\mathrm{Var}[Y_1] > 0$.

(i) If $c < 1/2$, *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. transient.*

(ii) If $c > 1/2$, *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. ergodic.*

(iii) If $c = 1/2$ *and* $Y_1 \stackrel{\mathcal{D}}{=} -Y_1$, *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. null-recurrent.*

(iv) Suppose $c = 1/2$ *and* $\mathbb{E}[Y_1] \neq 0$. *Suppose* $\chi(n) = an^{-\beta}$ *for* $a > 0$ *and* $\beta > 0$.

(a) If $0 < \beta < 1$ *and* $\mathbb{E}[Y_1] > 0$ *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. ergodic.*

(b) If $\beta > 1$ *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. null-recurrent.*

(c) If $0 < \beta < 1$ *and* $\mathbb{E}[Y_1] < 0$ *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. transient.*

We prove Theorem 8.2.3 along with our main results in Section 8.3.

Remarks. Note that in part (iii), $Y_1 \stackrel{\mathcal{D}}{=} -Y_1$ implies that all odd moments of Y_1 are zero. By minor modifications to the proof of Theorem 8.2.3 one can obtain a more refined result, specifically that with $p := \min\{j \in \{1,3,5,\ldots\} : \mathbb{E}[Y_1^j] \neq 0\}$, for $p > 1$ we have a statement analogous to part (iv) but with $\mathbb{E}[Y_1]$ replaced by $\mathbb{E}[Y_1^p]$ and with the critical value of β being $1/(2(p-1))$ for $p > 1$, rather than 1. We do not go into further detail here.

Theorem 8.2.3 (iv) demonstrates that in the case of a randomly perturbed simple random walk, the critical exponent for the perturbation is -1, as in the case of the non-random perturbation (Theorem 8.2.2). It may be possible to refine Theorem 8.2.3 (iv) to obtain more delicate results analogous to those of Theorem 8.2.2.

For the remainder of the chapter, we ensure that the underlying environment is *random*, by supposing $\mathrm{Var}[\zeta_1] > 0$. First we consider the case $\mathbb{E}[\zeta_1] \neq 0$. Here we have the following result, which we state without proof, but which follows by similar methods to those used in [34] or later in this chapter. In this situation, the perturbation introduced by $\chi(n)Y_n$ does not affect the criteria given in (i) and (iii) of Theorem 8.2.1.

Theorem 8.2.4 *Suppose* $\mathrm{Var}[\zeta_1] > 0$, $\mathbb{E}[\zeta_1] \neq 0$, *and* $\mathbb{P}[Y_1 = 0] < 1$.

(i) *If* $\mathbb{E}[\zeta_1] < 0$, *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. transient.*

(ii) *If* $\mathbb{E}[\zeta_1] > 0$, *then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. ergodic.*

For the remainder of the chapter we consider the more interesting case where $\mathbb{E}[\zeta_1] = 0$, so that we have a random walk in a random environment that is asymptotic to Sinai's regime. We prove general results about this RWRE with asymptotically zero perturbation that are analogous to Theorem 8.2.2, but significantly different.

If $\mathbb{P}[Y_1 = 0] < 1$ (and permitting the case that $\mathbb{P}[Y_1 = c] = 1$ for some c with $0 < |c| \leq 1$) we define

$$\lambda := \mathbb{E}\left[\frac{Y_1}{\xi_1(1-\xi_1)}\right]. \tag{8.10}$$

Also, we use the notation

$$\sigma^2 := \mathrm{Var}[\zeta_1]. \tag{8.11}$$

Note that, under the condition (8.2), we have $\sigma^2 < \infty$ and, since Y_1 is bounded, $|\lambda| < \infty$. We also draw attention to the fact that, given (8.2), $\mathbb{P}$-a.s.,

$$-\frac{1}{\varepsilon^2} \leq \frac{Y_1}{\xi_1(1-\xi_1)} \leq \frac{1}{\varepsilon^2}, \tag{8.12}$$

a fact that we shall use later. For what follows, of separate interest are the two cases $\lambda = 0$ and $\lambda \neq 0$. We concentrate on the latter case for most of the results that follow (but see the remark after Theorem 8.2.7). However, our first result deals with the case in which $Y_1/\xi_1 \stackrel{\mathcal{D}}{=} -Y_1/(1-\xi_1)$. This implies $\lambda = 0$ (see (8.10)), but is a rather special case; Theorem 8.2.5 demonstrates that in this case the detailed behaviour of χ is not important: as long as $\chi(n) \to 0$ as $n \to \infty$, then $\eta_t(\omega)$ is null-recurrent for a.e. ω.

Theorem 8.2.5 *With* σ *as defined at (8.11), suppose that* $Y_1/\xi_1 \stackrel{\mathcal{D}}{=} -Y_1/(1-\xi_1)$, $\mathbb{P}[Y_1 = 0] < 1$, $\mathbb{E}[\zeta_1] = 0$, *and* $\sigma^2 > 0$. *Then* $\eta_t(\omega)$ *is* $\mathbb{P}$*-a.s. null-recurrent.*

An example of (Y_1, ξ_1) for which Theorem 8.2.5 holds is when Y_1 and ξ_1 are independent uniform random variables on $(-1, 1)$ and $(\varepsilon, 1-\varepsilon)$ respectively.

Our remaining results deal with the case $\lambda \neq 0$ (but see also the remark after Theorem 8.2.7). In our next result (Theorem 8.2.6), we give some rather specific conditions on the asymptotic behaviour of the function χ. Theorem 8.2.6 is a special case of our general result, Theorem 8.2.7.

Theorem 8.2.6 *With λ and σ defined at (8.10) and (8.11) respectively, suppose that $\lambda \neq 0$, $\mathbb{P}[Y_1 = 0] < 1$, $\mathbb{E}[\zeta_1] = 0$, and $\sigma^2 > 0$. Let $c_{\text{crit}} := \sigma 2^{-1/2}$.*

(i) If there exist constants $c > c_{\text{crit}}$ and $n_0 \in \mathbf{Z}^+$ such that $\lambda\chi(n) \geq cn^{-1/2}(\log\log n)^{1/2}$ for all $n \geq n_0$, then $\eta_t(\omega)$ is $\mathbb{P}$-a.s. ergodic.

(ii) If there exist constants $c \leq c_{\text{crit}}$ and $n_0 \in \mathbf{Z}^+$ such that $|\lambda|\chi(n) \leq cn^{-1/2}(\log\log n)^{1/2}$ for all $n \geq n_0$, then $\eta_t(\omega)$ is $\mathbb{P}$-a.s. null-recurrent.

(iii) If there exist constants $c > c_{\text{crit}}$ and $n_0 \in \mathbf{Z}^+$ such that $\lambda\chi(n) \leq -cn^{-1/2}(\log\log n)^{1/2}$ for all $n \geq n_0$, then $\eta_t(\omega)$ is $\mathbb{P}$-a.s. transient.

Remark. Theorem 8.2.6 shows that in our case the critical exponent for the perturbation is $-1/2$. This contrasts with the deterministic environment case (as in Theorem 8.2.2, and see [98], Theorem 3), in which the critical exponent is -1. When the perturbation is smaller than this critical size (as in part (ii)), it is insufficient to change the recurrence/transience characteristics of the Markov chain from those of Sinai's regime. If the perturbation is greater than the critical size, it changes the behaviour of the Markov chain from that of Sinai's regime, making it either transient or ergodic depending on the sign of the perturbation. This feature is present in our most general result, Theorem 8.2.7.

Theorem 8.2.6 will follow as a corollary to Theorem 8.2.7, below. Theorem 8.2.7 is more refined than Theorem 8.2.6. In order to formulate our deeper result, we need more precise conditions on the behaviour of the perturbation function $\chi(n)$. To achieve this, we define the notions of *k-supercritical* and *k-subcritical* below. First, we need some additional notation.

Recall the notation $\log_1(x) := \log(x)$ and $\log_k(x) := \log(\log_{k-1}(x))$ for $k = 2, 3, \ldots$. Let n_k denote the smallest positive integer such that $\log_{k+1}(n_k) \geq 0$. Let $a_k := 2$ for $k \in \mathbf{N} \setminus \{3\}$ and $a_3 := 3$. For each $k \in \mathbf{N}$ we define the $[0, \infty)$-valued function φ_k as follows (we use the given form for the φ_k due to the appearance in the sequel of the Law of the Iterated Logarithm). For $x \in [e, \infty)$ and $d \in \mathbf{R}$, let

$$\varphi_1(x; d) := ((2 + d)\log_2 x)^{1/2},$$

and for $k = 2, 3, \ldots$, with $x \in [n_k, \infty)$ and $d \in \mathbf{R}$, let

$$\varphi_k(x; d) := \left(\sum_{i=1}^{k-1} a_{i+1} \log_{i+1} x + (a_{k+1} + d) \log_{k+1} x \right)^{1/2}. \tag{8.13}$$

We shall see that the behaviour of the Markov chain $\eta_t(\omega)$ is determined by the driving function χ. By applying the Law of the Iterated Logarithm, we shall see that the critical form of χ is related to an iterated logarithm expression of the form of φ_k.

In order to formulate our main result we make the following definitions of *k-supercritical* and *k-subcritical.*

Definition 8.2.1 *Recall the definitions of λ and σ at (8.10) and (8.11) respectively. Suppose $\lambda \neq 0$. For $k \in \mathbf{N}$, we say χ is* k-supercritical *if there exist constants $c \in (0,\infty)$ and $n_0 \in \mathbf{Z}^+$, such that, for all $n \geq n_0$,*

$$\chi(n) \geq \frac{\sigma}{2|\lambda|} n^{-1/2} \varphi_k(n;c). \tag{8.14}$$

For $k \in \mathbf{N}$, we say χ is k-subcritical *if there exist constants $c \in (0,\infty)$ and $n_0 \in \mathbf{Z}^+$ such that, for all $n \geq n_0$,*

$$\chi(n) \leq \frac{\sigma}{2|\lambda|} n^{-1/2} \varphi_k(n;-c). \tag{8.15}$$

Remarks. Implicit in χ being k-subcritical or k-supercritical is the constant c, a fact that we make repeated use of in the proofs in Section 8.3. Whenever we consider a k-subcritical or k-supercritical function in what follows, we understand this to imply the existence of such a c, and often refer to the constant c in this context.

Also, observe that if for some $k \in \mathbf{N}$, χ is k-supercritical, with implicit constant $c \in (0,\infty)$, then for any $c' \in (0,c)$ we have that (8.14) implies

$$\chi(n) \geq \frac{\sigma}{2|\lambda|} n^{-1/2} \varphi_k(n;c) \geq \frac{\sigma}{2|\lambda|} n^{-1/2} \varphi_k(n;c').$$

Similarly if for some $k \in \mathbf{N}$, χ is k-subcritical, with implicit constant $c \in (0,\infty)$, then for any $c' \in (0,c)$ we have that (8.15) implies

$$\chi(n) \leq \frac{\sigma}{2|\lambda|} n^{-1/2} \varphi_k(n;-c) \leq \frac{\sigma}{2|\lambda|} n^{-1/2} \varphi_k(n;-c').$$

Finally, we note that Definition 8.2.1 excludes functions that oscillate significantly about the critical region $n^{-1/2}$.

Our most general result is as follows.

Theorem 8.2.7 *With λ and σ defined at (8.10) and (8.11) respectively, suppose that $\lambda \neq 0$, $\mathbb{P}[Y_1 = 0] < 1$, $\mathbb{E}[\zeta_1] = 0$ and $\sigma^2 > 0$.*

(i) If, for some $k \in \mathbf{N}$, χ is k-supercritical (8.14) and $\lambda > 0$, then $\eta_t(\omega)$ is $\mathbb{P}$-a.s. ergodic.

(ii) If, for some $k \in \mathbf{N}$, χ is k-subcritical (8.15) then $\eta_t(\omega)$ is $\mathbb{P}$-a.s. null-recurrent.

(iii) If, for some $k \in \mathbf{N}$, χ is k-supercritical (8.14) and $\lambda < 0$, then $\eta_t(\omega)$ is $\mathbb{P}$-a.s. transient.

Remark. In the general case $\lambda = 0$, it turns out that higher moments contribute, and we obtain a slightly more general form of Theorem 8.2.7. It is straightforward to modify the proof of Theorem 8.2.7 to obtain such a result. Specifically, if for $r \in \mathbf{N}$ we set

$$\lambda_r := \frac{1}{r}\mathbb{E}\left[Y_1^r\left(\frac{1}{(1-\xi_1)^r} + \frac{(-1)^{r+1}}{\xi_1^r}\right)\right],$$

and $p := \min\{j \in \mathbf{N} : \lambda_j \neq 0\}$, then for $p > 1$ a statement of the form of Theorem 8.2.7 holds but with λ replaced by λ_p and the conditions on χ being replaced by conditions on χ^p. We do not pursue the details here.

We will prove Theorem 8.2.7 in the next section. The idea behind the proof of the recurrence and transience conditions is to construct a function f of the process $\eta_t(\omega)$ such that $f(\eta_t(\omega))$ is a 'martingale' everywhere except in a finite region, and determine the cases in which this function is finite or infinite. The proof of ergodicity relies on the construction of a stationary measure and determining its properties.

8.3 Proofs of main results

Before embarking upon the proof of Theorem 8.2.7, we need some preliminary results. First we present the criteria for classification of countable Markov chains which we will require.

Let $(W_t; t \in \mathbf{Z}^+)$ be a discrete, irreducible, aperiodic, time-homogeneous Markov chain on $\mathbf{Z}^+$. We have the following classification criteria, which are consequences of those given in Chapter 2 of [50]. The following result, which we state without proof, is a consequence of Theorem 2.2.2 of [50], and is slightly more general than Proposition 2.1 of [34].

Lemma 8.3.1 *Suppose there exist a function $f : \mathbf{Z}^+ \to [0, \infty)$ which is uniformly bounded and nonconstant, and a set $A \subset \mathbf{Z}^+$ such that*

$$E[f(W_{t+1}) - f(W_t) | W_t = x] = 0, \tag{8.16}$$

for all $x \in \mathbf{Z}^+ \setminus A$, *and*

$$f(x) > \sup_{y \in A} f(y), \tag{8.17}$$

for at least one $x \in \mathbf{Z}^+ \setminus A$. *Then the Markov chain* (W_t) *is transient.*

Proof. We have $f(n) < C$ for all $n \in \mathbf{Z}^+$, where $C \in (0, \infty)$. For each n, set $g(n) = C - f(n) > 0$. For some $x \in \mathbf{Z}^+ \setminus A$, we have by (8.17) that

$$g(x) = C - f(x) < C - \sup_{y \in A} f(y) = \inf_{y \in A} g(y).$$

The function g thus satisfies the conditions of Theorem 2.2.2 of [50], proving the lemma. □

The following result is contained in Theorem 2.2.1 in [50].

Lemma 8.3.2 *Suppose that there exist a function* $f : \mathbf{Z}^+ \to [0, \infty)$ *and a finite set* $A \subset \mathbf{Z}^+$ *such that*

$$E\left[f(W_{t+1}) - f(W_t) | W_t = x\right] \leq 0, \tag{8.18}$$

for all $x \in \mathbf{Z}^+ \setminus A$, *and* $f(x) \to +\infty$ *as* $x \to \infty$. *Then the Markov chain* (W_t) *is recurrent.*

We will need Feller's refined form for the Law of the Iterated Logarithm [51]. The following result is a consequence of Theorem 7 of [51].

Lemma 8.3.3 *Let* X_i, $i = 1, 2, \ldots$, *be a sequence of independent random variables with* $E[X_i] = 0$ *for all* i, *and* $E[X_i^2] = \sigma_i^2 < \infty$ *for* $i = 1, 2, \ldots$. *Suppose the* X_i *are bounded, that is* $P[|X_i| > C] = 0$ *for all* i *and some* $0 < C < \infty$. *Let*

$$s_n^2 := \sum_{i=1}^{n} \sigma_i^2. \tag{8.19}$$

Suppose that $s_n \to \infty$ *as* $n \to \infty$. *Let* $S_n := \sum_{i=1}^{n} X_i$. *For some* $k \in \mathbf{N}$ *and* $\varepsilon \in (-\infty, \infty)$, *define* $\varphi_k(n; \varepsilon)$ *as at (8.13). Then*

$$P\left[S_n > s_n \varphi_k(s_n^2; \varepsilon) \text{ i.o.}\right] = \begin{cases} 1 & \text{if } \varepsilon < 0 \\ 0 & \text{if } \varepsilon > 0 \end{cases} \tag{8.20}$$

In particular, if the X_i *are i.i.d. and bounded random variables with* $E[X_1^2] = \sigma^2$, *we have*

$$P\left[S_n > \sigma n^{1/2} \varphi_k(n; \varepsilon) \text{ i.o.}\right] = \begin{cases} 1 & \text{if } \varepsilon < 0 \\ 0 & \text{if } \varepsilon > 0 \end{cases} \tag{8.21}$$

We will also need the following result. Recall the definition of $\varphi_k(i;d)$ at (8.13).

Lemma 8.3.4 *For $k \in \mathbf{N}$, let n_k be the smallest positive integer such that $\log_{k+1} n_k \geq 0$. For any $d \in \mathbf{R}$, we have*

$$\sum_{i=n_k}^{n} i^{-1/2}\varphi_k(i;d) = 2n^{1/2}\varphi_k(n;d) + \alpha_n, \tag{8.22}$$

where $|\alpha_n| < 6n^{1/2}$ for all n sufficiently large.

Proof. We have, for $k \in \mathbf{N}$,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x^{1/2}\varphi_k(x;d)\right) = \frac{1}{2}x^{-1/2}\varphi_k(x;d) + x^{1/2}\varphi_k'(x;d),$$

where

$$\varphi_k'(x;d) = \frac{1}{2}\left(\varphi_k(x;d)\right)^{-1}\cdot\left(\frac{2}{x\log x} + \frac{3}{x\log x\log\log x} + \cdots\right) < \frac{1}{x},$$

for x sufficiently large. Thus, for any $k \in \mathbf{N}$,

$$\begin{aligned}\int_{n_k}^{n} x^{-1/2}\varphi_k(x;d)\mathrm{d}x &= 2\left[x^{1/2}\varphi_k(x;d)\right]_{n_k}^{n} - 2\int_{n_k}^{n} x^{1/2}\varphi_k'(x;d)\mathrm{d}x \\ &= 2n^{1/2}\varphi_k(n;d) + b_n,\end{aligned} \tag{8.23}$$

where

$$|b_n| \leq 2\int_{n_k}^{n} x^{1/2}\varphi_k'(x;d)\mathrm{d}x + 2n_k^{1/2}\varphi_k(n_k;d) \leq C_k + 2\int_0^n x^{-1/2}\mathrm{d}x,$$

for some $0 < C_k < \infty$, which depends on k (and d). Thus, for each k, $|b_n| \leq 5n^{1/2}$ for all n sufficiently large. Since $x^{-1/2}\varphi_k(x;d)$ is a decreasing function for all x sufficiently large (depending on k but not d), we have that there exist finite positive constants C_k' and C_k'' such that

$$\sum_{i=n_k}^{n} i^{-1/2}\varphi_k(i;d) + C_k' \geq \int_{n_k}^{n} x^{-1/2}\varphi_k(x;d)\mathrm{d}x \geq \sum_{i=n_k+1}^{n} i^{-1/2}\varphi_k(i;d) - C_k''.$$

So we have

$$0 \leq \sum_{i=n_k}^{n} i^{-1/2}\varphi_k(i;d) - \int_{n_k}^{n} x^{-1/2}\varphi_k(x;d)\mathrm{d}x \leq n_k^{-1/2}\varphi_k(n_k;d) + C, \tag{8.24}$$

for some $0 < C < \infty$, that does not depend on n. Then from (8.24) and (8.23) we obtain (8.22). □

For a given realization ω of our random environment, with p_i and q_i, $i = 1, 2, \ldots$ defined by (8.3), let

$$D(\omega) := \sum_{i=1}^{\infty} \frac{1}{q_i} \prod_{j=1}^{i} \frac{q_j}{p_j} = \frac{1}{p_1} + \frac{q_1}{p_1 p_2} + \frac{q_1 q_2}{p_1 p_2 p_3} + \cdots . \tag{8.25}$$

Lemma 8.3.5 *If, for a given environment ω, the quantity $D(\omega)$ as defined at (8.25) is finite, then the Markov chain $\eta_t(\omega)$ is ergodic. On the other hand, if $D(\omega) = \infty$, then the Markov chain $\eta_t(\omega)$ for this ω is not ergodic.*

Proof. For fixed environment ω, i.e., given a configuration of $(p_i; i = 1, 2, \ldots)$, $\eta_t(\omega)$ is a reversible Markov chain. For this Markov chain one has the stationary measure $\mu = (\mu_0, \mu_1, \ldots)$, where

$$\mu_0 = 2, \ \mu_1 = \frac{1}{p_1}, \text{ and } \mu_n = \frac{1}{p_1} \prod_{i=1}^{n-1} \frac{q_i}{p_{i+1}} \quad n \geq 2. \tag{8.26}$$

Then, with the definition of $D(\omega)$ at (8.25), we have

$$\sum_{i=0}^{\infty} \mu_i = 2 + D(\omega).$$

Thus, if, for this ω, $D(\omega)$ is finite, then the Markov chain $\eta_t(\omega)$ is ergodic, since we can obtain a stationary distribution. On the other hand, if $D(\omega) = \infty$ for this ω, the Markov chain $\eta_t(\omega)$ is not ergodic. □

Our next useful result, Lemma 8.3.6 below, uses the Law of the Iterated Logarithm to analyse the behaviour of sums of i.i.d. random variables weighted by the function χ.

Lemma 8.3.6 *Let Z_i, $i = 1, 2, \ldots$, be a sequence of i.i.d. random variables which are bounded (so that $P[|Z_1| > B] = 0$ for some $0 < B < \infty$), such that $0 \leq E[Z_1] < \infty$. Let $\chi : [0, \infty) \to [0, \infty)$ such that $\chi(n) \to 0$ as $n \to \infty$. With λ defined at (8.10), suppose $\lambda \neq 0$.*

(a) Suppose $E[Z_1] > 0$. Suppose that, for some $k \in \mathbf{N}$, χ is k-subcritical as defined at (8.15). Then with probability one, for any $\varepsilon > 0$, for all but finitely many n,

$$-n^{\varepsilon} \leq \sum_{i=1}^{n} Z_i \chi(i) \leq \frac{\sigma E[Z_1]}{|\lambda|} n^{1/2} \varphi_k(n; -c/3). \tag{8.27}$$

(b) Suppose $E[Z_1] > 0$. Suppose that, for some $k \in \mathbf{N}$, χ is k-supercritical as defined at (8.14). Then with probability one, for all but finitely many n,

$$\sum_{i=1}^{n} Z_i\chi(i) \geq \frac{\sigma E[Z_1]}{|\lambda|} n^{1/2}\varphi_k(n; c/3). \tag{8.28}$$

(c) Suppose $E[Z_1] = 0$. Then for any $\varepsilon > 0$ with probability one, for all but finitely many n,

$$\sum_{i=1}^{n} Z_i\chi(i) \leq \varepsilon(n \log\log n)^{1/2}. \tag{8.29}$$

Remark. When we come to apply Lemma 8.3.6 later in the proofs of the theorems, the configuration $(Z_i, i \geq 1)$ that we will use will be specified by the realization of the random environment ω, so that the qualifier 'with probability one' in the lemma translates as 'for a.e. ω.'

Proof of Lemma 8.3.6. Recall the definitions of λ and σ at (8.10) and (8.11) respectively. Suppose $\lambda \neq 0$. For the proofs of parts (a) and (b), suppose that $E[Z_1] > 0$. First we prove part (a). Suppose that for some $k \in \mathbf{N}$ χ is k-subcritical. Write

$$S_n := \sum_{i=1}^{n}(Z_i - E[Z_i])\chi(i). \tag{8.30}$$

Then

$$\mathrm{Var}[S_n] = \mathrm{Var}[Z_1]\sum_{i=1}^{n}(\chi(i))^2. \tag{8.31}$$

Suppose that $\mathrm{Var}[S_n] \to \infty$ as $n \to \infty$. Then, by Lemma 8.3.3, taking $X_i = (Z_i - E[Z_i])$, we have that with probability one the configuration of $(Z_i, i \geq 1)$ is such that

$$S_n > (\mathrm{Var}[S_n])^{1/2}(3\log\log(\mathrm{Var}[S_n]))^{1/2},$$

for only finitely many n. (The constant 3 appears for the sake of simplicity, any constant strictly greater than 2 will suffice). That is, with probability one, for all but finitely many n,

$$S_n \leq (\mathrm{Var}[S_n])^{1/2}(3\log\log(\mathrm{Var}[S_n]))^{1/2} \leq (\mathrm{Var}[S_n])^{1/2}(3\log\log(n))^{1/2},$$

the second inequality following from (8.31) and (8.15). Thus, with probability one, using (8.31) and (8.15) once more, we have that for any $\varepsilon > 0$ and for all but finitely many n,

$S_n \leq n^\varepsilon$. Thus, with probability one, for all but finitely many n, since $E[Z_1] > 0$ and χ is a nonnegative function,

$$-n^\varepsilon \leq \sum_{i=1}^{n} Z_i \chi(i) \leq n^\varepsilon + E[Z_1] \sum_{i=1}^{n} \chi(i). \tag{8.32}$$

The lower bound in (8.32) establishes the lower bound in (8.27). We now need to prove the upper bound. By (8.15), we have that there exist $c \in (0, \infty)$ and $k \in \mathbf{N}$ such that for all n sufficiently large

$$\sum_{i=1}^{n} \chi(i) \leq \frac{\sigma}{2|\lambda|} \sum_{i=1}^{n} i^{-1/2} \varphi_k(i; -c/2). \tag{8.33}$$

Then from (8.33) with (8.22) we obtain, for all n sufficiently large

$$\sum_{i=1}^{n} \chi(i) \leq \frac{\sigma}{|\lambda|} n^{1/2} \varphi_k(n; -c/2) + \frac{3\sigma}{|\lambda|} n^{1/2}. \tag{8.34}$$

Hence from (8.34) and the upper bound in (8.32), we have that, with probability one, for all but finitely many n,

$$\sum_{i=1}^{n} Z_i \chi(i) \leq \frac{\sigma E[Z_1]}{|\lambda|} n^{1/2} \varphi_k(n; -c/2) + \frac{3\sigma E[Z_1]}{|\lambda|} n^{1/2} + n^\varepsilon.$$

Then we can absorb the final two terms on the right hand side to give (8.27), given that $\text{Var}[S_n] \to \infty$ as $n \to \infty$. On the other hand, suppose that $\text{Var}[S_n] \leq C$ for all n and some $C < \infty$. Then, by (8.31), we have that $\sum_{i=1}^{n} (\chi(i))^2 < C$ for some $0 < C < \infty$. So, by Jensen's inequality, and the boundedness of the Z_i, we have that for all n,

$$\sum_{i=1}^{n} Z_i \chi(i) \leq \sqrt{n \sum_{i=1}^{n} Z_i^2 (\chi(i))^2} \leq n^{1/2} B \sqrt{\sum_{i=1}^{n} (\chi(i))^2} \leq C n^{1/2}, \tag{8.35}$$

for some $0 < C < \infty$. Hence we obtain (8.27) in this case also. This proves part (a).

Now we prove part (b). Suppose that for some $k \in \mathbf{N}$ χ is k-supercritical. Again, we use the notation of (8.30). By (8.14), we have that $\text{Var}[S_n] \to \infty$ as $n \to \infty$. Then, by Lemma 8.3.3, taking $X_i = -(Z_i - E[Z_i])$, we have that, with probability one,

$$S_n < -(\text{Var}[S_n])^{1/2} (3 \log\log(\text{Var}[S_n]))^{1/2},$$

for only finitely many n. But $\chi(n) \to 0$ as $n \to \infty$, so with probability one there exists a sequence $c_1, c_2, \ldots$ such that $c_n \to \infty$ as $n \to \infty$ and $\text{Var}[S_n] < n/c_n$ for all n. Thus, with probability one,

$$S_n \geq -n^{1/2} c_n^{-1/2} (3 \log\log(n))^{1/2}, \tag{8.36}$$

for all but finitely many n. So, with probability one, for all but finitely many n,

$$\sum_{i=1}^{n} Z_i \chi(i) \geq E[Z_1] \sum_{i=1}^{n} \chi(i) - n^{1/2} c_n^{-1/2} (3 \log\log(n))^{1/2}. \tag{8.37}$$

By (8.14), we have that there exist $c \in (0, \infty)$ and $k \in \mathbf{N}$ such that for n sufficiently large

$$\sum_{i=1}^{n} \chi(i) \geq \frac{\sigma}{2|\lambda|} \sum_{i=1}^{n} i^{-1/2} \varphi_k(i; c/2). \tag{8.38}$$

Then from (8.38) with (8.22) we obtain, for all n sufficiently large

$$\sum_{i=1}^{n} \chi(i) \geq \frac{\sigma}{|\lambda|} n^{1/2} \varphi_k(n; c/2) - \frac{3\sigma}{|\lambda|} n^{1/2}. \tag{8.39}$$

Hence, with probability one, from (8.37) and (8.39) we have that, for all but finitely many n

$$\sum_{i=1}^{n} Z_i \chi(i) \geq \frac{\sigma E[Z_1]}{|\lambda|} n^{1/2} \varphi_k(n; c/2) - \frac{3\sigma E[Z_1]}{|\lambda|} n^{1/2} - n^{1/2} c_n^{-1/2} (3 \log\log(n))^{1/2},$$

which yields (8.28). Thus we have proved part (b).

Finally, we prove part (c). Suppose now that $E[Z_1] = 0$. Again use the notation of (8.30). First, suppose that $\text{Var}[S_n] \leq C$ for all n, for some $0 < C < \infty$. Then, we have that (8.35) holds. On the other hand, suppose that $\text{Var}[S_n] \to \infty$ as $n \to \infty$. But, since $\chi(n) \to 0$ as $n \to \infty$, we have that $\text{Var}[S_n] = o(n)$. Applying Lemma 8.3.3 with $X_i = Z_i \chi(i)$ then yields (8.29). Thus the proof of the lemma is complete. □

Proof of Theorem 8.2.7. First we examine the recurrence and transience criteria for $\eta_t(\omega)$. For the recurrent cases, we proceed in the second part of the proof to analyse the stationary measure given in Lemma 8.3.5, in order to distinguish between null-recurrence and ergodicity (positive recurrence). We work for a fixed environment ω, that is, a given realization of p_i and q_i for $i = 1, 2, \ldots$, as given by (8.3).

We aim to apply Lemmas 8.3.1 and 8.3.2, and so we construct a Lyapunov function f, that is, a function $f : \mathbf{Z}^+ \to \mathbf{R}^+$ such that $f(\eta_t(\omega))$ behaves as a martingale (with respect to the natural filtration) for $\eta_t(\omega) \neq 0$. To do this, we proceed as follows.

For a given environment ω, set $\Delta_1 := 1$ and for $i = 2, 3, \ldots$ let

$$\Delta_i := \prod_{j=1}^{i-1} (p_j / q_j) = \exp \sum_{j=1}^{i-1} \log(p_j / q_j), \tag{8.40}$$

and then set $f(0) := 0$ and for $n = 1, 2, 3, \ldots$ let

$$f(n) := \sum_{i=1}^{n} \Delta_i. \tag{8.41}$$

Note that $f(n) \geq 0$. Then, for fixed ω, for $t \in \mathbf{Z}^+$ and $n = 1, 2, \ldots$,

$$\begin{aligned} E[f(\eta_{t+1}(\omega)) - f(\eta_t(\omega)) | \eta_t(\omega) = n] &= p_n f(n-1) + q_n f(n+1) - f(n) \\ &= q_n \Delta_{n+1} - p_n \Delta_n = 0, \end{aligned}$$

i.e. $f(\eta_t(\omega))$ is a martingale over $1, 2, 3, \ldots$.

We need to examine the behaviour of $f(n)$ as $n \to \infty$, in order to apply Lemmas 8.3.1 and 8.3.2. Recall from (8.5) that there exists $n_0 \in \mathbf{N}$ such that for any $j > n_0$ and almost every realization of the random environment ω, $p_j = \xi_j + Y_j \chi(j)$ and $q_j = 1 - \xi_j - Y_j \chi(j)$. Then, for j sufficiently large, and a.e. ω

$$\log p_j = \log(\xi_j + Y_j \chi(j)) = \log(\xi_j) + \xi_j^{-1} Y_j \chi(j) + O\left((\chi(j))^2\right),$$

and

$$\log q_j = \log(1 - \xi_j - Y_j \chi(j)) = \log(1 - \xi_j) - (1 - \xi_j)^{-1} Y_j \chi(j) + O\left((\chi(j))^2\right),$$

so that for j sufficiently large and a.e. ω

$$\log(p_j / q_j) = \log\left(\frac{\xi_j}{1 - \xi_j}\right) + \frac{Y_j}{\xi_j (1 - \xi_j)} \chi(j) + O\left((\chi(j))^2\right). \tag{8.42}$$

Note that $\mathbb{E}[\log(p_n/q_n)] = O(\chi(n)) \to 0$ as $n \to \infty$, so that in this sense we asymptotically approach Sinai's regime.

Recall from (8.6) that for $i = 1, 2, \ldots$, $\zeta_i = \log(\xi_i / (1 - \xi_i))$. From (8.41), (8.40) and (8.42) we have, for n sufficiently large, for a.e. ω

$$f(n) = \sum_{i=1}^{n} \exp \sum_{j=1}^{i-1} \left[\zeta_j + \frac{Y_j}{\xi_j (1 - \xi_j)} \chi(j) + O((\chi(j))^2)\right]. \tag{8.43}$$

Note that for what follows the $O((\chi(j))^2)$ terms in (8.43) can be ignored, since, when $\lambda \neq 0$ (where λ is given by (8.10)), the other two terms are dominant. Thus we need to examine the behaviour of the two terms $\sum_{i=1}^{n} \zeta_i$ and $\sum_{i=1}^{n} \frac{Y_i}{\xi_i (1 - \xi_i)} \chi(i)$. This behaviour depends upon the sign of λ, and the magnitude of the perturbation χ.

First suppose that for some $k \in \mathbf{N}$ χ is k-subcritical (8.15). In this case, we show that in (8.43) the term involving the ζ_j is essentially dominant. We can apply Lemma

8.3.6 with $Z_i = Y_i\xi_i^{-1}(1-\xi_i)^{-1}$ (if $\lambda > 0$) or $Z_i = -Y_i\xi_i^{-1}(1-\xi_i)^{-1}$ (if $\lambda < 0$), and the boundedness property (8.12), so that (8.27) implies that, for any $\varepsilon > 0$, for all but finitely many n, for a.e. ω

$$-n^{\varepsilon} \leq \operatorname{sign}(\lambda) \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)} \chi(i) \leq \sigma n^{1/2} \varphi_k(n; -c/3), \tag{8.44}$$

with $c \in (0, \infty)$ as given in (8.15). Also, from the Law of the Iterated Logarithm (Lemma 8.3.3), we have that, for a.e. ω, there are infinitely many values of n for which

$$\sum_{i=1}^{n} \zeta_i \geq \sigma n^{1/2} \varphi_k(n; -c/4). \tag{8.45}$$

So from (8.44) and (8.45), we have that, for a.e. ω, there are infinitely many values of n such that, if $\lambda > 0$,

$$\sum_{i=1}^{n} \zeta_i + \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)} \chi(i) \geq \sigma n^{1/2} \varphi_k(n; -c/4) - n^{\varepsilon},$$

and if $\lambda < 0$,

$$\sum_{i=1}^{n} \zeta_i + \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)} \chi(i) \geq \sigma n^{1/2} (\varphi_k(n; -c/4) - \varphi_k(n; -c/3)).$$

Thus, by choosing ε to be small, we have that for a.e. ω, there are infinitely many values of n such that

$$\sum_{i=1}^{n} \zeta_i + \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)} \chi(i) \geq C n^{1/2}, \tag{8.46}$$

for some C with $0 < C < \infty$. Thus from (8.46), (8.40), and (8.42), there are, for a.e. ω, infinitely many values of n for which $\Delta_n > 1$, and hence as $n \to \infty$ $f(n) \to +\infty$ for a.e. ω. Thus, by Lemma 8.3.2, $\eta_t(\omega)$ is recurrent for a.e. ω.

Now suppose that for some $k \in \mathbf{N}$ χ is k-supercritical (8.14). In this case, we show that the term in (8.43) involving $Y_j\xi_j^{-1}(1-\xi_j)^{-1}$ is essentially dominant, and thus the sign of λ determines the behaviour. This time, from Lemma 8.3.6 with $Z_i = Y_i\xi_i^{-1}(1-\xi_i)^{-1}$ (if $\lambda > 0$) or $Z_i = -Y_i\xi_i^{-1}(1-\xi_i)^{-1}$ (if $\lambda < 0$), and the boundedness property (8.12), we have that (8.28) implies that, for a.e. ω, for all but finitely many n,

$$\operatorname{sign}(\lambda) \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)} \chi(i) \geq \sigma n^{1/2} \varphi_k(n; c/3). \tag{8.47}$$

Also, from the Law of the Iterated Logarithm (Lemma 8.3.3), we have that, for a.e. ω, there are only finitely many n such that

$$\sum_{i=1}^{n} \zeta_i \geq \sigma n^{1/2} \varphi_k(n; c/4). \tag{8.48}$$

If $\lambda < 0$, from (8.47) and (8.48), we have that, for a.e. ω, there are only finitely many n such that

$$\sum_{i=1}^{n} \zeta_i + \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)}\chi(i) \geq \sigma n^{1/2} \left(\varphi_k(n; c/4) - \varphi_k(n; c/3)\right). \tag{8.49}$$

So if $\lambda < 0$, from (8.49), (8.40), and (8.42), we have that for a.e. ω there are only finitely many values of n for which

$$\Delta_n \geq \exp\left(-C_1 n^{1/2}\right),$$

for some C_1, not depending on ω, with $0 < C_1 < \infty$. Thus for a.e. ω there exists a constant C_2 (depending on ω) with $0 < C_2 < \infty$ such that

$$f(n) \leq C_2 + \sum_{i=1}^{\infty} \exp\left(-C_1 i^{1/2}\right),$$

which is bounded. So in this case, by Lemma 8.3.1, we have that, for a.e. ω, $\eta_t(\omega)$ is transient.

On the other hand, if $\lambda > 0$ then Lemma 8.3.3 with (8.47) implies that for a.e. ω there are infinitely many values of n for which

$$\sum_{i=1}^{n} \zeta_i + \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)}\chi(i) \geq \sigma n^{1/2} \left(\varphi_k(n; c/3) - \varphi_k(n; c/4)\right) \geq C_1 n^{1/2}, \tag{8.50}$$

for some C_1, not depending on ω, with $0 < C_1 < \infty$. So if $\lambda > 0$, from (8.50), (8.40), and (8.42) for a.e. ω there are infinitely many values of n for which

$$\Delta_n \geq \exp\left(C_1 n^{1/2}\right).$$

Thus $f(n) \to +\infty$ $\mathbb{P}$-a.s., and in this case we have that, for a.e. ω, $\eta_t(\omega)$ is recurrent, by Lemma 8.3.2.

We now classify the recurrent cases further into ergodic (positive recurrent) and null-recurrent. To determine ergodicity, we apply Lemma 8.3.5. Given ω, and with $D(\omega)$ as defined at (8.25), we have

$$D(\omega) = \sum_{i=1}^{\infty} \frac{1}{q_i} \exp\left(-\sum_{j=1}^{i} \log\left(p_i/q_i\right)\right) = \sum_{i=1}^{\infty} \frac{1}{\Delta_{i+1} q_i},$$

where Δ_i is as defined at (8.40). By a similar argument to (8.42), we have that for n sufficiently large, for a.e. ω

$$\frac{1}{\Delta_n} = \exp\left(-\sum_{i=1}^{n-1} \zeta_i - \sum_{i=1}^{n-1} \frac{Y_i}{\xi_i(1-\xi_i)}\chi(i) + O\left(\sum_{i=1}^{n-1} (\chi(i))^2\right)\right).$$

We use similar arguments as in the proof of recurrence and transience to analyse $D(\omega)$. First suppose that for some $k \in \mathbf{N}$ χ is k-subcritical. Then, by a similar argument to (8.46), we have that for a.e. ω there are infinitely many values of i for which

$$-\sum_{i=1}^{n} \zeta_i - \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)} \chi(i) \geq C n^{1/2},$$

for $0 < C < \infty$. Thus for a.e. ω there are infinitely many values of n for which $(1/\Delta_{n+1}) > 1$ and $(1/(\Delta_{n+1} q_n)) > 1$. Hence $D(\omega) = +\infty$ for a.e. ω. So, for a.e. ω, by Lemma 8.3.5, $\eta_t(\omega)$ is not ergodic.

Now suppose that for some $k \in \mathbf{N}$ χ is k-supercritical. If $\lambda > 0$, using similar arguments to before, we have that for a.e. ω there are only finitely many n for which

$$-\sum_{i=1}^{n} \zeta_i - \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)} \chi(i) \geq \sigma n^{1/2} \left(\varphi_k(n; c/4) - \varphi_k(n; c/3) \right).$$

So for a.e. ω there are only finitely many values of n for which

$$(1/\Delta_n) \geq \exp\left(-C_1 n^{1/2}\right),$$

for some $0 < C_1 < \infty$. Thus for a.e. ω there exists a constant C_2 (depending on ω) with $0 < C_2 < \infty$ such that

$$D(\omega) \leq C_2 + \sum_{i=1}^{\infty} \exp\left(-C_1 i^{1/2}\right),$$

which is bounded. So in this case, for a.e. ω, by Lemma 8.3.5, $\eta_t(\omega)$ is ergodic.

On the other hand, if $\lambda < 0$, we have that for a.e. ω there are infinitely many n for which

$$-\sum_{i=1}^{n} \zeta_i - \sum_{i=1}^{n} \frac{Y_i}{\xi_i(1-\xi_i)} \chi(i) \geq \sigma n^{1/2} \left(\varphi_k(n; c/3) - \varphi_k(n; c/4) \right) \geq C_1 n^{1/2},$$

for some C_1, not depending on ω, with $0 < C_1 < \infty$. So for a.e. ω there are infinitely many values of n for which

$$(1/\Delta_n) \geq \exp\left(C_1 n^{1/2}\right).$$

Thus $D(\omega) = +\infty$ $\mathbb{P}$-a.s., and once again by Lemma 8.3.5, $\eta_t(\omega)$ is $\mathbb{P}$-a.s. not ergodic. This completes the proof of Theorem 8.2.7. □

Proof of Theorem 8.2.6. First we prove parts (i) and (iii). Suppose that, for all n sufficiently large, $\lambda\chi(n) \geq c n^{-1/2} (\log\log n)^{1/2}$, for some $c > c_{\text{crit}}$ where $c_{\text{crit}} = \sigma 2^{-1/2}$.

Then we see that χ is k-supercritical (8.14) for $k = 2, 3, \ldots$, since, for example

$$\begin{aligned}\frac{c}{|\lambda|}n^{-1/2}(\log\log n)^{1/2} &= \frac{c}{c_{\text{crit}}}\frac{\sigma}{2|\lambda|}n^{-1/2}(2\log\log n)^{1/2} \\ &\geq \frac{\sigma}{2|\lambda|}n^{-1/2}(2\log\log n + 4\log\log\log n)^{1/2},\end{aligned}$$

for n sufficiently large and $c > c_{\text{crit}}$. Hence (i) follows from part (i) of Theorem 8.2.7. Similarly, (iii) follows from part (iii) of Theorem 8.2.7.

For part (ii), suppose that $|\lambda|\chi(n) \leq cn^{-1/2}(\log\log n)^{1/2}$ for all n sufficiently large, $c \leq c_{\text{crit}}$. Then we see that χ is k-subcritical (8.15) for $k = 2, 3, \ldots$, since, for example

$$\begin{aligned}\frac{c}{|\lambda|}n^{-1/2}(\log\log n)^{1/2} &\leq \frac{\sigma}{2|\lambda|}n^{-1/2}(2\log\log n)^{1/2} \\ &\leq \frac{\sigma}{2|\lambda|}n^{-1/2}(2\log\log n + 2\log\log\log n)^{1/2},\end{aligned}$$

for n sufficiently large. Then part (ii) of Theorem 8.2.7 gives part (ii) of Theorem 8.2.6, and the proof of Theorem 8.2.6 is complete. □

Proof of Theorem 8.2.5. From Lemma 8.3.3, we have that for a.e. ω there are infinitely many values of n for which

$$\sum_{i=1}^{n} \zeta_i \geq \sigma n^{1/2}(\log\log(n))^{1/2}. \tag{8.51}$$

By a similar argument to (8.42), but keeping track of higher order terms in the Taylor series, we have that now

$$\log(p_i/q_i) = \zeta_i + \sum_{r=1}^{\infty}\frac{1}{r}Y_i^r\left(\frac{1}{(1-\xi_i)^r} + \frac{(-1)^{r+1}}{\xi_i^r}\right)(\chi(i))^r. \tag{8.52}$$

By the condition $Y_1/\xi_1 \stackrel{\mathcal{D}}{=} -Y_1/(1-\xi_1)$, we have that the expectation of the sum on the right of (8.52) is zero. Hence we can apply part (c) of Lemma 8.3.6 with

$$Z_i = \sum_{r=1}^{\infty}\frac{1}{r}Y_i^r\left(\frac{1}{(1-\xi_i)^r} + \frac{(-1)^{r+1}}{\xi_i^r}\right)(\chi(i))^{r-1} \tag{8.53}$$

to obtain that for all but finitely many n, for a.e. ω

$$\sum_{i=1}^{n}\sum_{r=1}^{\infty}\frac{1}{r}Y_i^r\left(\frac{1}{(1-\xi_i)^r} + \frac{(-1)^{r+1}}{\xi_i^r}\right)(\chi(i))^r \geq -\varepsilon n^{1/2}(\log\log(n))^{1/2}, \tag{8.54}$$

and by choosing ε sufficiently small we have from (8.52), (8.51) and (8.54) that, for a.e. ω there are infinitely many values of n for which

$$\sum_{i=1}^{n}\log(p_i/q_i) \geq Cn^{1/2}(\log\log n)^{1/2},$$

for $0 < C < \infty$. Thus with Δ_n defined at (8.40), we have that for a.e. ω there are infinitely many values of n for which

$$\Delta_n \geq \exp\left(Cn^{1/2}(\log\log(n))^{1/2}\right),$$

and so $f(n) \to +\infty$ $\mathbb{P}$-a.s., and so, by Lemma 8.3.2, $\eta_t(\omega)$ is recurrent for a.e. ω.

To prove null-recurrence, it remains to show that the Markov chain is not ergodic. Consider $D(\omega)$ as defined at (8.25) again. From Lemma 8.3.3, we have that for a.e. ω there are infinitely many values of n for which

$$-\sum_{i=1}^{n} \zeta_i \geq \sigma n^{1/2}(\log\log(n))^{1/2}.$$

From part (c) of Lemma 8.3.6 with Z_i as at (8.53) we have that for all but finitely many n, for a.e. ω

$$-\sum_{i=1}^{n} Z_i \chi(i) \geq -\varepsilon n^{1/2}(\log\log(n))^{1/2},$$

and by choosing ε sufficiently small we have that for a.e. ω there are infinitely many values of n for which

$$(1/\Delta_n) \geq \exp\left(Cn^{1/2}(\log\log(n))^{1/2}\right),$$

for some $0 < C < \infty$, and so $D(\omega) = +\infty$ $\mathbb{P}$-a.s. Thus, by Lemma 8.3.5, the Markov chain is $\mathbb{P}$-a.s. not ergodic. Thus, for a.e. ω, $\eta_t(\omega)$ is null-recurrent. □

Proof of Theorem 8.2.3. Parts (i) and (ii) follow easily with the methods used in the proof of Theorem 8.2.7. We prove part (iii). By a similar argument to (8.42), we have that now

$$\log(p_i/q_i) = \sum_{r=1}^{\infty} \frac{4^r}{2r-1} Y_i^{2r-1}(\chi(i))^{2r-1} = 4Y_i\chi(i) + O((\chi(i))^3). \tag{8.55}$$

Since $Y_1 \stackrel{\mathcal{D}}{=} -Y_1$, we have that all odd powers of Y_1 have zero expectation, so that the expectation of the right hand side of (8.55) is zero. Thus it is clear that for a.e. ω there are infinitely many values of n for which $\sum_{i=1}^{n} \log(p_i/q_i) \geq 0$, and hence $\Delta_n \geq 1$, and so $f(n) \to +\infty$ for a.e. ω, and we have $\mathbb{P}$-a.s. recurrence, by Lemma 8.3.2.

To prove null recurrence, it remains to show that the Markov chain is not ergodic. Once more, consider $D(\omega)$ as defined at (8.25). By a similar argument to above, for a.e. ω there are infinitely many values of n for which $\sum_{i=1}^{n} \log(p_i/q_i) \leq 0$ and hence $(1/\Delta_n) \geq 1$, and so $D(\omega) = +\infty$ for a.e. ω. Thus, by Lemma 8.3.5, the Markov chain is $\mathbb{P}$-a.s. not ergodic. This completes the proof of part (iii).

We now prove part (iv). Once again we analyse the properties of the expression (8.55).

Suppose that $\chi(n) = an^{-\beta}$ for $a > 0$, $\beta > 0$. Now suppose that $0 < \beta < 1$ and that $\mathbb{E}[Y_1] < 0$. Then from (8.55), we have that there exist $0 < C_1 < \infty$, $0 < C_2 < \infty$ such that

$$-C_1 n^{1-\beta} \leq \mathbb{E}\sum_{i=1}^{n} \log(p_i/q_i) \leq -C_2 n^{1-\beta}.$$

If $\beta \geq 1/2$, then, by the boundedness of Y_1, we have

$$\sup_n \mathbb{E}\left|\sum_{i=1}^{n} \log(p_i/q_i) - \mathbb{E}\sum_{i=1}^{n} \log(p_i/q_i)\right|^k < \infty,$$

for all $k \in \mathbf{N}$, so that $\mathbb{P}$-a.s.,

$$\left|\sum_{i=1}^{n} \log(p_i/q_i) - \mathbb{E}\sum_{i=1}^{n} \log(p_i/q_i)\right| < n^{\varepsilon},$$

for all but finitely many n, and any $\varepsilon > 0$. So, for all but finitely many n, for a.e. ω

$$\Delta_n \leq \exp\left(-Cn^{1-\beta} + n^{\varepsilon}\right),$$

for some C with $0 < C < \infty$, so that, for ε small enough,

$$f(n) = \sum_{i=1}^{n} \Delta_i$$

is bounded for a.e. ω, which implies that $\eta_t(\omega)$ is $\mathbb{P}$-a.s. transient, by Lemma 8.3.1. Also, if $0 < \beta < 1/2$, from (8.55), we have that there exist $0 < C_1 < \infty$,$0 < C_2 < \infty$ such that

$$C_1 n^{1-2\beta} \geq \mathrm{Var} \sum_{i=1}^{n} \log(p_i/q_i) \geq C_2 n^{1-2\beta} \to \infty,$$

as $n \to \infty$, and then we can apply Lemma 8.3.3 to obtain, for a.e. ω

$$\sum_{i=1}^{n} \log(p_i/q_i) \leq -C_1 n^{1-\beta} + C_2 n^{(1/2)-\beta} \log\log n,$$

for constants $0 < C_1 < \infty$, $0 < C_2 < \infty$ (depending on ω) and all but finitely many n. So once again we have $f(n)$ is $\mathbb{P}$-a.s. bounded, and so we have $\mathbb{P}$-a.s. transience by Lemma 8.3.1. This proves part (c).

To prove part (a), we apply Lemma 8.3.5. Suppose that $\mathbb{E}[Y_1] > 0$. By similar arguments to above, this time we have that for a.e. ω

$$\sum_{i=1}^{n} \log(p_i/q_i) \geq Cn^{1-\beta},$$

for some $0 < C < \infty$ and all but finitely many n. Thus, for a.e. ω, for all but finitely many n,

$$\frac{1}{\Delta_n} = \exp\left(-\sum_{i=1}^{n-1}\log(p_i/q_i)\right) \leq \exp\left(-Cn^{1-\beta}\right),$$

and so, for $D(\omega)$ as defined at (8.25), $D(\omega) < \infty$ $\mathbb{P}$-a.s., and so, by Lemma 8.3.5, the Markov chain is $\mathbb{P}$-a.s. ergodic, proving part (a).

Finally, we prove part (b). Suppose that $\beta > 1$. Now, since $-1 \leq Y_i \leq 1$ and $\chi(n) = an^{-\beta}$, we have from (8.55) that there exists a constant C_1 (not depending on ω) with $0 < C_1 < \infty$ such that, for a.e. ω,

$$\left|\sum_{i=1}^{n}\log(p_i/q_i)\right| \leq C_1\sum_{i=1}^{n} i^{-\beta} \leq C_2,$$

for finite positive C_2, not depending on ω or n, this last inequality following since $\beta > 1$. Thus for a.e. ω, for each n,

$$0 < \exp(-C_2) \leq \exp\left(\sum_{i=1}^{n}\log(p_i/q_i)\right) \leq \exp(C_2) < \infty,$$

so that for each n, Δ_n and $1/\Delta_n$ are each bounded strictly away from 0 and from ∞, so that $\mathbb{P}$-a.s. $f(n) \to +\infty$ as $n \to \infty$, and $D(\omega) = +\infty$ $\mathbb{P}$-a.s. Thus by Lemma 8.3.1 the Markov chain is $\mathbb{P}$-a.s. recurrent, and by Lemma 8.3.5 $\mathbb{P}$-a.s. not ergodic. Thus, for a.e. ω, $\eta_t(\omega)$ is null-recurrent. This completes the proof of Theorem 8.2.3. □

8.4 Discussion

Having classified completely the recurrence/transience of our RWRE, a further question would be: how far away from the origin, typically, is the random walker? More precisely, can we determine the almost sure and/or 'in probability' behaviour of $\eta_t(\omega)$ as $t \to \infty$?

In Sinai's regime for the RWRE on $\mathbf{Z}^+$, Comets, Menshikov and Popov (see [34], Theorem 3.2) give the following almost sure result:

Theorem 8.4.1 *Suppose* $\mathbb{E}[\zeta_1] = 0$, $\mathbb{P}[Y_1 = 0] = 1$, *and* $0 < \mathrm{Var}[\zeta_1] < \infty$. *Then, for almost every environment, as* $t \to \infty$,

$$\frac{\eta_t}{(\log t)^2} < (\log\log t)^{2+\varepsilon},$$

for all but finitely many t.

This result (for the RWRE on $\mathbf{Z}$) dates back to Deheuvels and Revész [39]. An exact upper limit result is given in [72]. See also [35].

The following convergence in distribution result for the RWRE on $\mathbf{Z}^+$ is due to Golosov [63]. The result is stated in terms of the so-called *annealed law* $\mathbb{Q}$ of (η_t), given by

$$\mathbb{Q}(\cdot) = \int_\Omega P(\cdot)\mathrm{d}\mathbb{P}(\omega).$$

Theorem 8.4.2 *Suppose* $\mathbb{E}[\zeta_1] = 0$, $\mathbb{P}[Y_1 = 0] = 1$, *and* $0 < \mathrm{Var}[\zeta_1] < \infty$. *Then as* $t \to \infty$,

$$\mathbb{Q}\left(\frac{\eta_t}{(\log t)^2} \le u\right) \longrightarrow F(u),$$

where $F(u)$ *is a given distribution function (see [63]).*

In future work, we aim to investigate analogues of Theorems 8.4.1 and 8.4.2 in the case of our RWRE perturbed from Sinai's regime – of interest are both the almost sure and 'in probability' (see, for example, [35, 132]) behaviour.

In work in preparation [100], we study the long-run limiting behaviour (as $t \to \infty$) of our random walk $\eta_t(\omega)$ in terms of its distance from the origin. In [100] we give almost sure results analogous to Theorem 8.4.1 (in both null-recurrent and transient cases) for our perturbed RWRE. For example, in the $\mathbb{P}$-almost sure transient case of the RWRE perturbed from Sinai's regime (that is, $\mathbb{E}[\zeta_1 = 0]$, $\mathrm{Var}[\zeta_1] > 0$, $\lambda < 0$, with $\chi(n) = n^{-\alpha}$ for some fixed $0 < \alpha < 1/2$), we have that for a.e. ω, for any $\varepsilon > 0$, as $t \to \infty$,

$$(\log\log\log t)^{-(2/\alpha)-\varepsilon} < \frac{\eta_t(\omega)}{(\log t)^{1/\alpha}} < (\log\log t)^{(2/\alpha)+\varepsilon},$$

for all but finitely many t. Thus in this case, we see that the random walk, for almost every environment, is contained in a window about $(\log t)^{1/\alpha}$. This aspect of the problem requires additional techniques, however, and we do not discuss this further in this thesis.

Appendix A

Technical background

A.1 Graph theory

Here we outline some of the fundamental notions of graph theory. For a more complete and extensive introduction, see for example [25].

A *graph* G is an ordered pair $G := (V, E)$, consisting of a *vertex set*, V, and an *edge set*, E. V is a countable set of vertices (points). In our examples, V will often be a finite subset of Euclidean space, i.e. $V = \{v_1, v_2, \ldots, v_n\} \subset \mathbf{R}^d$. We will also usually take V to be *random*. The edge set E is a collection of unordered pairs of members of V. Each unordered pair $\{u, v\} \in E$ indicates an edge between vertices $u, v \in V$. The graph $G = (V, E)$ can thus be represented diagrammatically as a set of points V connected by edges (line segments) according to E.

For our purposes, we do not permit loops – we disallow edges of the form $\{v, v\}$ for $v \in V$, and also disallow multiple edges between the same two points. Thus we assume each element of E is distinct and consists of two distinct members of V.

We say that two vertices $v_1, v_2 \in V$ are *adjacent* in G if $\{v_1, v_2\} \in E$. The graph $G = (V, E)$ is said to be *complete* (on V) if all the vertices in V are pairwise adjacent, that is E contains all distinct unordered pairs of elements of V. A graph $H = (V_H, E_H)$ is said to be a *subgraph* of $G = (V, E)$, written as $H \subseteq G$, if H is a graph and $V_H \subseteq V$, $E_H \subseteq E$. If $H \subseteq G$, $H \neq G$ and H is not the empty graph $(\emptyset, \emptyset)$, H is said to be a *proper* subgraph of G. If the graph $G = (V, E)$ is such that every element of V appears at least once in edges in E, we say the graph G *spans* the vertex set V.

A *path* in $G = (V, E)$ between $v_0 \in V$ and $v_k \in V$ is a nonempty subgraph $P(v_0, v_k) = (V_P, E_P)$ of G, with set of distinct vertices $V_P := \{v_0, v_1, v_2, \ldots, v_k\} \subseteq V$ and edge set

$E_P := \{\{v_0, v_1\}, \{v_1, v_2\}, \ldots, \{v_{k-1}, v_k\}\} \subseteq E$. A path $P(v, v) = (V_P, E_P)$ with nonempty edge set E_P in $G = (V, E)$ between $v \in V$ and itself is called a (closed) *cycle*.

A graph $G = (V, E)$ is *connected* if there exists for every pair $u_1, u_2 \in V$ a path between them in G. The *degree* of a vertex $v \in V$, is the number of edges in E that contain v, that is the number of adjacent vertices to v in $G = (V, E)$. If a vertex has degree 0, we say that it is *isolated*.

An *acyclic* graph contains no cycles and is called a *forest*. A connected forest is called a *tree*. The terminology arises since the connected components of a forest are trees. Observe that if $T = (V_T, E_T)$ is a tree, then any two vertices in V_T are linked by a unique path in T.

A *directed* graph $G = (V, E)$ consists of a vertex set V, and a set of *directed edges*, E. Now E is a collection of *ordered* pairs of members of V. Each ordered pair $(u, v) \in E$ indicates a directed edge from u to v. Most of the terminology discussed about for undirected graphs carries over into the directed case in a natural way, simply with directed edges replacing undirected edges.

A.2 Probabilistic preliminaries

Here we collect some probabilistic terminology and results that we will use throughout.

For a random variable X and a sequence of random variables X_n, we use the notation $X_n \xrightarrow{\mathcal{D}} X$, $X_n \xrightarrow{P} X$, $X_n \xrightarrow{\text{a.s.}} X$ to denote the convergence of X_n to X in distribution, in probability, and almost surely, respectively.

The following result is sometimes referred to as *Slutsky's theorem* (see for example [45], p. 72).

Lemma A.2.1 *Suppose that X_n,Y_n are sequences of random variables such that $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} 0$ as $n \to \infty$ for some random variable X. Then $X_n + Y_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.*

Proof. Suppose $x \in \mathbf{R}$ is a continuity point of F, where $F(x) := P[X \leq x]$. For any $\varepsilon > 0$, we have

$$P[X_n \leq x - \varepsilon] - P[|Y_n| > \varepsilon] \leq P[X_n + Y_n \leq x] \leq P[X_n \leq x + \varepsilon] + P[|Y_n| > \varepsilon].$$

Since x is a continuity point of F and $X_n \xrightarrow{\mathcal{D}} X$, for ε sufficiently small we have $P[X_n \leq x - \varepsilon] \to F(x - \varepsilon)$ and $P[X_n \leq x + \varepsilon] \to F(x + \varepsilon)$ as $n \to \infty$. Also, since $Y_n \xrightarrow{P} Y$, we have $P[|Y_n| > \varepsilon] \to 0$ for any $\varepsilon > 0$ as $n \to \infty$. The fact that $\varepsilon > 0$ is arbitrary and x a continuity point of F then completes the proof. □

Appendix B

Technical complements

B.1 Proof of Theorem 3.2.1: the Poisson case

Let $\mathcal{P}$ be a Poisson process of unit intensity on $\mathbf{R}^d$. We say the functional H is *weakly stabilizing* on $\mathcal{R}$ if there is a random variable $\Delta(\infty)$ such that

$$\Delta(\mathcal{P} \cap A; B) \xrightarrow{\text{a.s.}} \Delta(\infty), \tag{B.1.1}$$

as $(A, B) \to \mathbf{R}^d$ through $\mathcal{R}$, by which we mean (B.1.1) holds whenever (A, B) is an $\mathcal{R}$-valued sequence of the form $(A_n, B_n)_{n \geq 1}$, such that $\cup_{n \geq 1} \cap_{m \geq n} B_m = \mathbf{R}^d$. Note that strong stabilization of H implies weak stabilization of H.

We say the functional H satisfies the *Poisson bounded moments condition* on $\mathcal{R}$ if

$$\sup_{(A,B) \in \mathcal{R}: \mathbf{0} \in A} \left\{ E\left[\Delta(\mathcal{P} \cap A; B)^4\right] \right\} < \infty. \tag{B.1.2}$$

Theorem B.1.1 *Suppose that H is weakly stabilizing on $\mathcal{R}$ (B.1.1) and satisfies (B.1.2). Then there exists $s^2 \geq 0$ such that as $n \to \infty$, $n^{-1}\mathrm{Var}[H(\mathcal{Q}_n; S_n)] \to s^2$ and $n^{-1/2}(H(\mathcal{Q}_n; S_n) - E[H(\mathcal{Q}_n; S_n)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2)$.*

Before proving Theorem B.1.1, we require further definitions and a lemma. Let $\mathcal{P}'$ be an independent copy of the Poisson process $\mathcal{P}$. For $\mathbf{x} \in \mathbf{Z}^d$, set

$$\mathcal{P}''(\mathbf{x}) = (\mathcal{P} \setminus Q(\mathbf{x}; 1/2)) \cup (\mathcal{P}' \cap Q(\mathbf{x}; 1/2)) .$$

Then given a translation invariant functional H on point sets in $\mathbf{R}^d$, define

$$\Delta_{\mathbf{x}}(A; B) := H(\mathcal{P}''(\mathbf{x}) \cap A; B) - H(\mathcal{P} \cap A; B);$$

this is the change in $H(\mathcal{P} \cap A; B)$ when the Poisson points in $Q(\mathbf{x}; 1/2)$ are resampled.

Lemma B.1.1 *Suppose H is weakly stabilizing on $\mathcal{R}$. Then for all $\mathbf{x} \in \mathbf{Z}^d$, there is a random variable $\Delta_{\mathbf{x}}(\infty)$ such that for all $\mathbf{x} \in \mathbf{Z}^d$,*

$$\Delta_{\mathbf{x}}(A; B) \xrightarrow{\text{a.s.}} \Delta_{\mathbf{x}}(\infty), \tag{B.1.3}$$

as $(A, B) \to \mathbf{R}^d$ through $\mathcal{R}$. Moreover, if H satisfies (B.1.2), then

$$\sup_{(A,B)\in\mathcal{R}, \mathbf{x}\in\mathbf{Z}^d} E\left[(\Delta_{\mathbf{x}}(A; B))^4\right] < \infty. \tag{B.1.4}$$

Proof. Set $C_0 = Q(\mathbf{0}; 1/2)$. By translation invariance, we need only consider the case $\mathbf{x} = \mathbf{0}$, and thus it suffices to prove that the variables $H(\mathcal{P} \cap A; B) - H(\mathcal{P} \cap A \setminus C_0; B)$ converge almost surely as $(A, B) \to \mathbf{R}^d$ through $\mathcal{R}$.

The number N of points of $\mathcal{P}$ in C_0 is Poisson with parameter 1. Let $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_N$ be the points of $\mathcal{P} \cap C_0$, taken in an order chosen uniformly at random from the $N!$ possibilities. Then, provided $C_0 \subseteq A$,

$$H(\mathcal{P} \cap A; B) - H(\mathcal{P} \cap A \setminus C_0; B) = \sum_{i=0}^{N-1} \delta_i(A; B),$$

where

$$\delta_i(A; B) := H((\mathcal{P} \cap A \setminus C_0) \cup \{\mathbf{V}_1, \ldots, \mathbf{V}_{i+1}\}; B) - H((\mathcal{P} \cap A \setminus C_0) \cup \{\mathbf{V}_1, \ldots, \mathbf{V}_i\}; B).$$

Since N is a.s. finite, it suffices to prove that each $\delta_i(A; B)$ converges almost surely as $(A, B) \to \mathbf{R}^d$ through $\mathcal{R}$. Let $\mathbf{U}$ be a uniform random vector on C_0, independent of $\mathcal{P}$. The distribution of the translated point process $-\mathbf{V}_{i+1} + \{\mathbf{V}_1, \ldots, \mathbf{V}_i\} \cup (\mathcal{P} \setminus C_0)$ is the same as the conditional distribution of $\mathcal{P}$ given that the number of points in $-\mathbf{U} + C_0$ is equal to i, an event of strictly positive probability. By assumption, this satisfies weak stabilization, which proves (B.1.3).

Next we prove (B.1.4). If $Q(\mathbf{x}; 1/2) \cap A = \emptyset$ then $\Delta_{\mathbf{x}}(A; B)$ is zero with probability 1. By translation invariance, it suffices to consider the $\mathbf{x} = \mathbf{0}$ case, that is, to prove

$$\sup_{(A,B)\in\mathcal{R}: C_0\cap A\neq\emptyset} E\left[(\Delta_{\mathbf{0}}(A; B))^4\right] < \infty. \tag{B.1.5}$$

The proof of this now follows the proof of (3.4) of [111], but with $\delta_i(A)$ replaced by $\delta_i(A; B)$ everywhere. □

Proof of Theorem B.1.1. Here we can assume, without loss of generality, that $\mathcal{Q}_n = \mathcal{P} \cap R_n$. For $\mathbf{x} \in \mathbf{Z}^d$, let $\mathcal{F}_{\mathbf{x}}$ denote the σ-field generated by the points of $\mathcal{P}$ in $\cup_{\mathbf{y}\in\mathbf{Z}^d: \mathbf{y}\leq\mathbf{x}} Q(\mathbf{y}; 1/2)$, where the order in the union is the lexicographic order on $\mathbf{Z}^d$.

Let R'_n be the set of points $\mathbf{x} \in \mathbf{Z}^d$ such that $Q(\mathbf{x}; 1/2) \cap R_n \neq \emptyset$. Let $k_n = \text{card}(R'_n)$. Then we have that

$$R_n \subseteq \bigcup_{\mathbf{x} \in R'_n} Q(\mathbf{x}; 1/2) \subseteq R_n \cup \partial_1(R_n),$$

so that

$$|R_n| \leq k_n \leq |R_n| + |\partial_1(R_n)|.$$

The vanishing relative boundary condition then implies that $k_n/n \to 1$ as $n \to \infty$.

Define the filtration $(\mathcal{G}_0, \mathcal{G}_1, \ldots, \mathcal{G}_{k_n})$ as follows: let $\mathcal{G}_0$ be the trivial σ-field, label the elements of R'_n in lexicographic order as $\mathbf{x}_1, \ldots, \mathbf{x}_{k_n}$ and let $\mathcal{G}_i = \mathcal{F}_{\mathbf{x}_i}$ for $1 \leq i \leq k_n$. Then $H(\mathcal{Q}_n; S_n) - E[H(\mathcal{Q}_n; S_n)] = \sum_{i=1}^{k_n} D_i$, where we set

$$D_i = E[H(\mathcal{Q}_n; S_n)|\mathcal{G}_i] - E[H(\mathcal{Q}_n; S_n)|\mathcal{G}_{i-1}] = E[-\Delta_{\mathbf{x}_i}(R_n; S_n)|\mathcal{F}_{\mathbf{x}_i}]. \qquad \text{(B.1.6)}$$

By orthogonality of martingale differences, $\text{Var}[H(\mathcal{Q}_n; S_n)] = E\sum_{i=1}^{k_n} D_i^2$. By this fact, along with a CLT for martingale differences (Theorem 2.3 of [95] or Theorem 2.10 of [104]), it suffices to prove the conditions

$$\sup_{n \geq 1} E\left[\max_{1 \leq i \leq k_n} \{k_n^{-1/2}|D_i|\}^2\right] < \infty, \qquad \text{(B.1.7)}$$

$$k_n^{-1/2} \max_{1 \leq i \leq k_n} |D_i| \xrightarrow{P} 0, \qquad \text{(B.1.8)}$$

and for some $s^2 \geq 0$,

$$k_n^{-1} \sum_{i=1}^{k_n} D_i^2 \xrightarrow{L^1} s^2. \qquad \text{(B.1.9)}$$

Using equation (B.1.4), and the representation equation (B.1.6) for D_i, we can verify equation (B.1.7) and equation (B.1.8) in just the same manner as for the equivalent estimates (3.7) and (3.8) in [111].

We now prove (B.1.9). By (B.1.3), for each $\mathbf{x} \in \mathbf{Z}^d$ the variables $\Delta_{\mathbf{x}}(A; B)$ converge almost surely to a limit, denoted $\Delta_{\mathbf{x}}(\infty)$, as $(A, B) \to \mathbf{R}^d$ through $\mathcal{R}$. For $\mathbf{x} \in \mathbf{Z}^d$ and $(A, B) \in \mathcal{R}$, let

$$F_{\mathbf{x}}(A; B) = E[\Delta_{\mathbf{x}}(A; B)|\mathcal{F}_{\mathbf{x}}]; \quad F_{\mathbf{x}} = E[\Delta_{\mathbf{x}}(\infty)|\mathcal{F}_{\mathbf{x}}].$$

Then $(F_{\mathbf{x}}, \mathbf{x} \in \mathbf{Z}^d)$ is a stationary family of random variables. Set $s^2 = E[F_{\mathbf{0}}^2]$. We claim that the ergodic theorem implies

$$k_n^{-1} \sum_{\mathbf{x} \in R'_n} F_{\mathbf{x}}^2 \xrightarrow{L^1} s^2. \qquad \text{(B.1.10)}$$

The proof of this follows, with minor modifications, the proof of the corresponding result (3.10) in [111].

We need to show that $F_{\mathbf{x}}(R_n; S_n)^2$ approximates to $F_{\mathbf{x}}^2$. We consider $\mathbf{x}$ at the origin $\mathbf{0}$. For any $(A, B) \in \mathcal{R}$, by Cauchy-Schwarz,

$$E[|F_{\mathbf{0}}(A; B)^2 - F_{\mathbf{0}}^2|] \leq \left(E[(F_{\mathbf{0}}(A; B) + F_{\mathbf{0}})^2]\right)^{1/2} \left(E[(F_{\mathbf{0}}(A; B) - F_{\mathbf{0}})^2]\right)^{1/2}. \quad \text{(B.1.11)}$$

By the definition of $F_{\mathbf{0}}$ and the conditional Jensen inequality,

$$\begin{aligned} E[(F_{\mathbf{0}}(A; B) + F_{\mathbf{0}})^2] &= E\left[(E[\Delta_{\mathbf{0}}(A; B) + \Delta_{\mathbf{0}}(\infty)|\mathcal{F}_{\mathbf{0}}])^2\right] \\ &\leq E[E[(\Delta_{\mathbf{0}}(A; B) + \Delta_{\mathbf{0}}(\infty))^2|\mathcal{F}_{\mathbf{0}}]] \\ = E[(\Delta_{\mathbf{0}}(A; B) + \Delta_{\mathbf{0}}(\infty))^2], & \end{aligned}$$

which is uniformly bounded by (B.1.3) and (B.1.4). Similarly,

$$E[(F_{\mathbf{0}}(A; B) - F_{\mathbf{0}})^2] \leq E[(\Delta_{\mathbf{0}}(A; B) - \Delta_{\mathbf{0}}(\infty))^2], \quad \text{(B.1.12)}$$

which is also uniformly bounded by (B.1.3) and (B.1.4). For any $\mathcal{R}$-valued sequence $(A_n, B_n)_{n \geq 1}$ with $\cup_{n \geq 1} \cap_{m \geq n} B_n = \mathbf{R}^d$, the sequence $(\Delta_{\mathbf{0}}(A_n; B_n) - \Delta_{\mathbf{0}}(\infty))^2$ tends to 0 almost surely by (B.1.3), and is uniformly integrable by (B.1.4), and therefore the expression (B.1.12) tends to zero so that by (B.1.11), $E[|F_{\mathbf{0}}(A_n; B_n)^2 - F_{\mathbf{0}}^2|] \to 0$.

Returning to the given sequence (R_n, S_n), let $\varepsilon > 0$. By the vanishing relative boundary condition, we can choose K_n so that $\lim_{n \to \infty} K_n = \infty$ and $|\partial_{K_n} S_n| \leq \varepsilon n$ for all n. Let S_n' be the set of $\mathbf{x} \in \mathbf{Z}^d$ such that $Q_{1/2}(\mathbf{x})$ has non-empty intersection with $S_n \setminus \partial_{K_n}(S_n)$. Using the conclusion of the previous paragraph and translation invariance, it is not hard to deduce that

$$\lim_{n \to \infty} \sup_{\mathbf{x} \in S_n'} E[|F_{\mathbf{x}}(R_n; S_n)^2 - F_{\mathbf{x}}^2|] = 0. \quad \text{(B.1.13)}$$

Also, since we assume $|S_n| \sim n$ we have $\text{card}(S_n') \geq |S_n| - \varepsilon n \geq (1 - 2\varepsilon)n$ for large enough n. Using this with (B.1.13), the uniform boundedness of $E[|F_{\mathbf{x}}(R_n; S_n)^2 - F_{\mathbf{x}}^2|]$ and the fact that ε can be taken arbitrarily small in the above argument, it is routine to deduce that

$$k_n^{-1} \sum_{\mathbf{x} \in R_n'} (F_{\mathbf{x}}(R_n; S_n)^2 - F_{\mathbf{x}}^2) \xrightarrow{L^1} 0,$$

and therefore (B.1.10) remains true with $F_{\mathbf{x}}$ replaced by $F_{\mathbf{x}}(R_n; S_n)$; that is, (B.1.9) holds and the proof of Theorem B.1.1 is complete. □

B.2 Proof of Theorem 3.2.1: the non-Poisson case

In this section we complete the proof of Theorem 3.2.1. The first step is to show that the conditions of Theorem 3.2.1 imply those of Theorem B.1.1, as follows.

Lemma B.2.1 *If H satisfies the uniform bounded moments condition (3.5) and is polynomially bounded, then H satisfies the Poisson bounded moments condition (B.1.2).*

Proof. The proof follows, with minor modifications, that of Lemma 4.1 of [111]. □

It follows from Lemma B.2.1 that if H satisfies the conditions of Theorem 3.2.1, then Theorem B.1.1 applies and we have the Poisson parts of Theorem 3.2.1. To de-Poissonize these limits we follow [111]. Define

$$R_{m,n} := H(\mathcal{U}_{m+1,n}; B) - H(\mathcal{U}_{m,n}; B).$$

We use the following coupling lemma.

Lemma B.2.2 *Suppose H is strongly stabilizing. Let $\varepsilon > 0$. Then there exists $\delta > 0$ and $n_0 \geq 1$ such that for all $n \geq n_0$ and all $m, m' \in [(1-\delta)n, (1+\delta)n]$ with $m < m'$, there exists a coupled family of variables D, D', R, R' with the following properties:*

(i) D and D' each have the same distribution as $\Delta(\infty)$;

(ii) D and D' are independent;

(iii) (R, R') have the same joint distribution as $(R_{m,n}, R_{m',n})$;

(iv) $P[\{D \neq R\} \cup \{D' \neq R'\}] < \varepsilon$.

Proof. Since we assume $|S_n|/|R_n| \to 1$, the probability that a random d-vector uniformly distributed over R_n lies in S_n tends to 1 as $n \to \infty$. Using this fact the proof follows, with some minor modifications, that of the corresponding result in [111], Lemma 4.2. □

Lemma B.2.3 *Suppose H is strongly stabilizing and satisfies the uniform bounded moments condition (3.5). Let $(h(n))_{n\geq 1}$ be a sequence with $n^{-1}h(n) \to 0$ as $n \to \infty$. Then*

$$\lim_{n\to\infty} \sup_{|n-m|\leq h(n)} |ER_{m,n} - E\Delta(\infty)| = 0; \tag{B.2.14}$$

$$\lim_{n\to\infty} \sup_{n-h(n)\leq m<m'\leq n+h(n)} \left|ER_{m,n}R_{m',n} - (E\Delta(\infty))^2\right| = 0; \tag{B.2.15}$$

$$\lim_{n\to\infty} \sup_{|n-m|\leq h(n)} ER_{m,n}^2 < \infty. \tag{B.2.16}$$

Proof. The proof follows that of Lemma 4.3 of [111]. □

Proof of Theorem 3.2.1 Theorem 3.2.1 now follows in the same way as Theorem 2.1 in [111], replacing $H(\,\cdot\,)$ with $H(\,\cdot\,; S_n)$. □

Appendix C

Dickman-type distributions

This section is supplementary to the main argument of this thesis, and so may be omitted on a first reading.

In this section, we review some of the properties of the distributions arising as limits in Theorems 4.2.1 and 4.2.3, before returning subsequently to the MDST. Some of these properties can be found in the literature (see, e.g., [6, 7, 24, 40, 61, 66, 71, 75, 77, 138]); we endeavour to make most of the current presentation self-contained. To begin with, we review the Dirichlet and Poisson-Dirichlet distributions.

C.1 The Dirichlet distribution

The Dirichlet distribution will be needed later on in the theory of spacings and one-dimensional nearest-neighbour type graphs, see Section 5.1.1. For the moment, the Dirichlet distribution serves as the foundation of the Poisson-Dirichlet distribution which is central to this chapter.

Here we follow Billingsley [24], p. 246. For $n \in \mathbf{N}$, let $\Delta_n \subset \mathbf{R}^n$ denote the n-dimensional simplex, that is

$$\Delta_n := \left\{ (x_1, \ldots, x_n) \in \mathbf{R}^n : x_i \geq 0,\ 1 \leq i \leq n;\ \sum_{i=1}^{n} x_i \leq 1 \right\}. \tag{C.1.1}$$

The random vector $(X_1, \ldots, X_n)$ has the *Dirichlet distribution* with parameters $\alpha_1, \ldots, \alpha_n$ if $X_n = 1 - \sum_{i=1}^{n-1} X_i$ and $(X_1, \ldots, X_{n-1})$ is distributed on the simplex Δ_{n-1} with density

$$\frac{\Gamma(\alpha_1 + \cdots + \alpha_n)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)} x_1^{\alpha_1 - 1} \cdots x_{n-1}^{\alpha_{n-1}-1} (1 - x_1 - \cdots - x_{n-1})^{\alpha_n - 1}. \tag{C.1.2}$$

If $\alpha_i = \alpha_j$ for some i, j, then from (C.1.2) on sees that the distribution of $(X_1, \ldots, X_n)$ remains the same if X_i and X_j are interchanged.

If $\alpha_i = \alpha$ for *all* i, then $(X_1, \ldots, X_n)$ is said to have the *symmetric* Dirichlet distribution (with parameter α), and in this case all the X_i are exchangeable. In particular, if $\alpha = 1$, the density in (5.37) is $(n-1)!$ – the uniform distribution over the simplex Δ_{n-1}.

C.2 The Poisson-Dirichlet distribution

Two excellent references on the Poisson-Dirichlet distribution are Chapter 9 of [84] and Section 4 of [24]. See also [7,71,83]. Here we give a brief indication of the standard way in which the Poisson-Dirichlet distribution may be arrived at; for further characterizations, see for example [7].

Suppose $(X_1, \ldots, X_n)$ is a random vector with symmetric Dirichlet distribution with parameter $\alpha = \lambda/n$, for some $\lambda > 0$. For some fixed k (less than n) let $(X_{(1)}, \ldots, X_{(k)})$ be the vector of the first k order statistics of $(X_1, \ldots, X_n)$, where

$$X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}.$$

The Poisson Dirichlet distribution emerges in the limit as $n \to \infty$ (and so $\alpha \to 0$), with $n\alpha = \lambda$ fixed. More precisely, we have (see Kingman [84], p. 94) that as $n \to \infty$, for each k,

$$(X_{(1)}, \ldots, X_{(k)}) \xrightarrow{\mathcal{D}} (Z_1, Z_2, \ldots, Z_k),$$

where the infinite sequence $(Z_1, Z_2, \ldots)$ satisfies

$$Z_1 \geq Z_2 \geq \cdots, \quad \sum_{j=1}^{\infty} Z_j = 1.$$

The distribution of $(Z_1, Z_2, \ldots)$ then depends only on λ, and is called the *Poisson-Dirichlet* distribution with parameter λ.

For what follows, we will be mostly concerned with Z_1, the first (and largest) component of the Poisson-Dirichlet distribution with parameter 1 – this is our so-called max-Dickman distribution (see Section C.7).

The Poisson-Dirichlet distribution has appeared in numerous places. The first is in relation to the distribution of largest prime factors of random integers [23] (see also [41]). For more recent proofs, see [42] and Section 4 of [24]. The literature on the Poisson-Dirichlet distribution and its place in applied probability is considerable; here we list a few examples. The Poisson-Dirichlet distribution also appears in population genetics (see e.g. [66, 78, 138]), in relation to random covering of the circle (see e.g. [54, 74]), random polynomials (see e.g. [8]), and random permutations (see e.g. [131]). The relationships between these various applications, and general structures, are discussed in, for example, [7, 9, 43, 71].

C.3 The Dickman function

Dickman's equation, which appears in analytic number theory, is the differential-difference equation:

$$u\rho'(u) + \rho(u-1) = 0 \qquad (u > 1). \tag{C.3.3}$$

The *Dickman function* is defined as the (unique) continuous solution $\rho(u)$ to (C.3.3) with $\rho(u) = 1$ for $0 < u \leq 1$ and with ρ differentiable on $(1, \infty)$. It is convenient to extend ρ over all of $\mathbf{R}$ by setting $\rho(u) = 0$ for $u \leq 0$. See [137] for analytic number theory background.

It is known (see [137]) that the Dickman function is positive and decreasing on the whole interval $(1, \infty)$; that it satisfies $\rho(u) \leq 1/\Gamma(u+1)$ for $u > 1$; and that it integrates to

$$\int_0^\infty \rho(x)\mathrm{d}x = e^\gamma, \tag{C.3.4}$$

where γ denote Euler's constant, so that $\gamma \approx 0.57721566$ and

$$\left(\sum_{i=1}^{k} \frac{1}{i}\right) - \log k = \gamma + O(k^{-1}). \tag{C.3.5}$$

Thus $e^\gamma \approx 1.78107$.

C.4 Probabilistic properties of the GD distributions

Proposition C.4.1 *Let $\theta > 0$. The following random variables X are distributionally equivalent.*

(a) A random variable X satisfying the fixed point equation

$$X \stackrel{\mathcal{D}}{=} U^{1/\theta}(1 + X), \tag{C.4.6}$$

where U is uniform on $(0, 1)$ and independent of the X on the right hand side.

(b) *A random variable X given by*

$$X = \sum_{j=1}^{\infty} \left(\prod_{i=1}^{j} U_i^{1/\theta} \right) = U_1^{1/\theta} + (U_1 U_2)^{1/\theta} + (U_1 U_2 U_3)^{1/\theta} + \cdots , \qquad \text{(C.4.7)}$$

where $U_1, U_2, U_3, \ldots$ *are independent uniform random variables on* $(0,1)$.

(c) *A random variable X given by*

$$X = \sum_{n=1}^{\infty} \exp(-T_n)$$

where $T_1, T_2, \ldots$ *are the successive arrival times of a homogeneous Poisson process of rate* θ *on the half-line* $(0, \infty)$.

(d) *A random variable X given by* $X = \sum_{n=1}^{\infty} Y_n$, *where* $Y_1, Y_2, Y_3, \cdots$ *are the points of a non-homogeneous Poisson point process on* $(0,1)$ *with mean measure* $(\theta/x)\mathrm{d}x$, *taken in decreasing order.*

We say that a random variable X given by any of the conditions (a), (b), (c) or (d) in Proposition C.4.1 has the *generalized Dickman* distribution with parameter θ (or $X \sim \mathrm{GD}(\theta)$ for short).

The term *Dickman distribution* has previously been used for the GD(1) distribution, i.e. that of a variable X satisfying $X \stackrel{\mathcal{D}}{=} U(1+X)$ (see e.g. [77]), and this is the usage we favour. The same term has also been used [40] for the distribution of a random variable Y satisfying the distributional fixed point equation $Y \stackrel{\mathcal{D}}{=} UY + 1$, where (naturally) U is uniform on $(0,1)$, and independent of Y, and also for other distributions; see [7]. It is easy to see that such a Y can be obtained by taking $Y = 1 + X$ with $X \sim \mathrm{GD}(1)$.

We shall see later (Corollary C.5.1) that if $X \sim \mathrm{GD}(1)$ then its density function satisfies Dickman's equation.

Remark. The GD(θ) distributions appear in many contexts in applied probability (particularly for $\theta = 1$ and $\theta = 2$). Examples include the limits of certain random variables in random algorithms (such as Hoare's FIND algorithm on random permutations and its variants), see e.g. [77], [94] and [33], Theorem 31). They also appear in the study of perpetuities (see [61]). For a method of simulating GD(θ) random variables, see [32]. We also discuss simulating Dickman-type random variables in Section C.8.

The natural connection to the length of the rooted edges in the MDST is via the *sums of uniform records*, see Section 6 of [6] (although note the typo γ for $1/\gamma$ after (6.9) there).

The connection between record values and rooted vertices in the MDST is discussed in Section 2.4.3; we exploit this in the proof of Theorem 4.2.1 (see Section 4.3).

Proof of Proposition C.4.1. First, suppose that X is given by the sum of the infinite random series (C.4.7), which converges almost surely because it has nonnegative terms and finite expectation. By factorizing (C.4.7),

$$X = U_1^{1/\theta}\left(1 + U_2^{1/\theta} + (U_2U_3)^{1/\theta} + (U_2U_3U_4)^{1/\theta} + \cdots\right). \tag{C.4.8}$$

The second factor in the right-hand side of (C.4.8) has the same distribution as $1 + X$, and is independent of U_1; hence, X satisfies the distributional identity (C.4.6).

Conversely, suppose that X satisfies (C.4.6). Suppose $U_1, U_2, \ldots$ are uniform on $(0,1)$, independent of X and of each other, and set $V_i := U_i^{1/\theta}$, for each i. Then X has the same distribution as $V_1(1+X) = V_1 + V_1X$, and hence the same distribution as $V_1(1 + V_2(1+X)) = V_1 + V_1V_2 + V_1V_2X$, and so on. Repeating this process, the term involving X converges in probability to zero and we see that X has the same distribution as $V_1 + V_1V_2 + V_1V_2V_3 + \cdots$.

Next, suppose that X is given by definition (c), i.e. $X = \sum_n e^{-T_n}$ where the T_n are successive arrival times of a Poisson process of rate θ on $(0,\infty)$. Set $Y_1 = T_1$ and $Y_n = T_n - T_{n-1}$ for $n \geq 2$. The inter-arrival times $Y_1, Y_2, \ldots$ are independent and exponentially distributed with parameter θ, so for each i, and for $0 < t < 1$,

$$P[e^{-Y_i} \leq t] = P[Y_i \geq -\log(t)] = e^{\theta \log t} = t^\theta$$

so that e^{-Y_i} has the same distribution as $U^{1/\theta}$, where U is uniform on $(0,1)$. Since

$$X = \sum_{n=1}^{\infty} e^{-T_n} = \sum_{n=1}^{\infty}\left(\prod_{i=1}^{n} e^{-Y_i}\right),$$

it follows that X has the same distribution as given in part (b).

Finally, definition (d) is distributionally equivalent to definition (c) by the Mapping Theorem [84], because the image of the uniform (Lebesgue) measure on $(0,\infty)$ with density θ, under the mapping $x \mapsto e^{-x}$, is the measure on $(0,1)$ with density (θ/x). □

We now collect some further properties of the generalized Dickman distribution. Most of these are scattered throughout the literature. See, for example, [24, 71, 75, 77].

Proposition C.4.2 *(a) If $X \sim \mathrm{GD}(\theta)$, $\theta > 0$, then the Laplace transform ψ of the distribution of X is given by*

$$\psi(t) = E\left[e^{-tX}\right] = \exp\left(\theta \int_0^t \frac{e^{-s}-1}{s} \mathrm{d}s\right) = \exp\left(\theta \int_0^1 \frac{e^{-tu}-1}{u} du\right), \qquad t \in \mathbf{R}.$$

(b) For $\theta, \theta' \in (0,\infty)$ if X and Y are independent random variables with $X \sim \mathrm{GD}(\theta)$ and $Y \sim \mathrm{GD}(\theta')$, then $X+Y \sim \mathrm{GD}(\theta+\theta')$.

(c) For $\theta > 0$, The $GD(\theta)$ distribution is infinitely divisible.

(d) If $X \sim GD(\theta)$, $\theta > 0$, then the k-th cumulant of X is equal to $\frac{\theta}{k}$.

(e) If $X \sim GD(\theta)$, $\theta > 0$, then the moments $m_k := E[X^k]$ satisfy $m_0 = 1$ and, for integer $k \geq 1$,

$$m_k = \frac{\theta}{k} \sum_{j=0}^{k-1} \binom{k}{j} m_j. \tag{C.4.9}$$

In particular, X has expected value θ and variance $\frac{\theta}{2}$.

(f) The moments m_k as in part (e) have the following alternative representation, for $k \in \mathbf{N}$,

$$\begin{aligned} m_k &= k! \sum_{j=k}^{\infty} \sum_{*} \prod_{i=1}^{j} \left(\left(k - \sum_{\ell=1}^{i-1} b_\ell \right) \theta + 1 \right)^{-1} (b_i!)^{-1} && \text{(C.4.10)} \\ &= k! \sum_{j=k}^{\infty} \sum_{*} \frac{(k\theta+1)^{-1}((k-b_1)\theta+1)^{-1} \cdots (b_j\theta+1)^{-1}}{(b_1!)(b_2!)\cdots(b_j!)}, && \text{(C.4.11)} \end{aligned}$$

where the sum indexed by $*$ is taken over all values of $b_1, b_2, \ldots, b_j$ such that $b_1 + b_2 + \cdots + b_j = k$ and $b_1 + 2b_2 + \cdots + jb_j = j$.

Proof. Suppose $X \sim \mathrm{GD}(\theta)$ and set $\psi(t) = E[e^{-tX}]$, the Laplace transform of the distribution of X. Then by definition, $X \stackrel{\mathcal{D}}{=} U^{1/\theta}(X+1)$ and so

$$\begin{aligned} \psi(t) &= E\left[E\left[\exp\left(-tU^{1/\theta}(X+1)\right) | U\right]\right] \\ &= \int_0^1 E\left[e^{-tu^{1/\theta}} e^{-tu^{1/\theta}X}\right] \mathrm{d}u = \int_0^1 e^{-tu^{1/\theta}} \psi(tu^{1/\theta}) \mathrm{d}u \\ &= \int_0^t e^{-w} \psi(w) \frac{\theta w^{\theta-1}}{t^\theta} \mathrm{d}w, \end{aligned}$$

from which it follows that

$$t^\theta \psi(t) = \theta \int_0^t e^{-w} \psi(w) w^{\theta-1} \mathrm{d}w,$$

and so

$$t^{\theta}\psi'(t) + \theta t^{\theta-1}\psi(t) = \theta e^{-t}\psi(t)t^{\theta-1};$$

hence

$$t\psi'(t) = \theta(e^{-t} - 1)\psi(t).$$

We have the initial condition $\psi(0) = 1$ and so

$$\log(\psi(t)) = \int_0^t \frac{\psi'(s)}{\psi(s)}\mathrm{d}s = \theta\int_0^t \frac{e^{-s}-1}{s}\mathrm{d}s = \theta\int_0^1 \frac{e^{-tu}-1}{u}\mathrm{d}u.$$

This completes the proof of part (a). Parts (b) and (c) follow at once from (a), or alternatively by a more probabilistic argument based on the Poisson process representation of X in part (d) of Proposition C.4.1.

Since the kth cumulant of X is defined to be the kth derivative of $\log\psi(-t)$, evaluated at $t = 0$, part (d) can also be deduced from (a).

To prove part (e), suppose $X \sim \mathrm{GD}(\theta)$, and write m_k for $E[X^k]$. Then by (C.4.6),

$$m_k = E[X^k] = E[U^{k/\theta}]E[(1+X)^k],$$

which implies that

$$m_k = \frac{\theta}{k+\theta}\left(m_k + \sum_{j=0}^{k-1}\binom{k}{j}m_j\right),$$

and so

$$m_k = \frac{\theta}{k}\cdot\sum_{j=0}^{k-1}\binom{k}{j}m_j.$$

Now we turn to part (f). Suppose $X \sim \mathrm{GD}(\theta)$, $\theta > 0$. Again let $m_k := E[X^k]$ be the kth moment of X. X admits the representation (C.4.7). Raising both sides of (C.4.7) to the power k and taking expectations gives (via multinomial theory)

$$E[X^k] = E\left[k!\sum_{j=k}^{\infty}\sum_{*}\frac{U_1^{\theta b_1}(U_1U_2)^{\theta b_2}\cdots(U_1U_2\cdots U_j)^{\theta b_j}}{(b_1!)(b_2!)\cdots(b_j!)}\right],$$

where the sum indexed by $*$ is taken over all values of $b_1, b_2, \ldots, b_j$ such that $b_1 + b_2 + \cdots + b_j = k$ and $b_1 + 2b_2 + \cdots + jb_j = j$. Collecting terms, and using the independence of the U_i, we have

$$m_k = k!\sum_{j=k}^{\infty}\sum_{*}\frac{E[U_1^{\theta k}]E[U_2^{\theta(k-b_1)}]E[U_3^{\theta(k-b_1-b_2)}]\cdots E[U_j^{\theta b_j}]}{(b_1!)(b_2!)\cdots(b_j!)}.$$

Then, since the U_i are uniform on $(0,1)$, we obtain (C.4.10). This completes the proof of the proposition. □

Remark. By part (b) of Proposition (C.4.2), we see that for any θ, we can represent $X \sim \mathrm{GD}(\theta)$ as the sum

$$X = Y_0 + \sum_{i=1}^{\lfloor\theta\rfloor} Y_i,$$

where $Y_i \sim \mathrm{GD}(1)$ for $i \geq 1$ and $Y_0 \sim \mathrm{GD}(\theta - \lfloor\theta\rfloor)$. It follows that, as $\theta \to \infty$,

$$\frac{X}{\theta} \xrightarrow{\text{a.s.}} 1, \quad \text{and} \quad \frac{X-\theta}{\theta^{1/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1/2).$$

C.5 GD probability density and distribution functions

In this section we derive further properties of generalized Dickman distributions, including, among other things, a partially explicit form of the probability density and distribution functions for these distributions.

We show first that the $\mathrm{GD}(\theta)$ distribution has a probability density function that is continuous except at 0, is piecewise differentiable and satisfies a certain differential-difference equation, which generalizes Dickman's equation.

Proposition C.5.1 *The generalized Dickman distribution with parameter $\theta > 0$ has a probability density function g_θ which is identically zero on $(-\infty, 0)$, is continuous on $(0,\infty)$, and is differentiable on $(0,1)\cup(1,\infty)$, satisfying the differential-difference equation*

$$t g_\theta'(t) = (\theta - 1) g_\theta(t) - \theta g_\theta(t-1). \tag{C.5.12}$$

Proof. Let $X \sim \mathrm{GD}(\theta)$. Let G_θ be the cumulative distribution function of X. By (C.4.6), we have that

$$\begin{aligned} G_\theta(t) &= P[X \leq t] = \int_0^1 P\left[u^{1/\theta}(1+X) \leq t\right] \mathrm{d}u \\ &= \int_0^1 G_\theta\left(\frac{t}{u^{1/\theta}} - 1\right) \mathrm{d}u. \end{aligned} \tag{C.5.13}$$

Make the substitution $s = \frac{t}{u^{1/\theta}} - 1$, so that $u = \left(\frac{t}{s+1}\right)^\theta$. This gives

$$G_\theta(t) = -\int_{t-1}^{\infty} G_\theta(s) \frac{\mathrm{d}u}{\mathrm{d}s} \mathrm{d}s.$$

Integrating by parts, we obtain

$$G_\theta(t) = G_\theta(t-1) + t^\theta \int_{t-1}^{\infty} (s+1)^{-\theta} \mathrm{d}G_\theta(s). \tag{C.5.14}$$

By the characterization of X in part (b) of Proposition C.4.1, $P[X > 0] = 1$; hence, $G_\theta(t) = 0$ for $t \leq 0$. By (C.5.14),

$$G_\theta(t) = \kappa_\theta t^\theta, \quad 0 \leq t \leq 1, \tag{C.5.15}$$

where $\kappa_\theta := E[(X+1)^{-\theta}]$.

By (C.5.14) and induction on n, G_θ is continuous on the interval $(-\infty, n)$ and continuously differentiable on the interval $(n-1, n)$ for $n = 1, 2, 3, \ldots$. (the case $n = 1$ is covered by (C.5.15)). Setting $g_\theta(t) = G'_\theta(t)$, for non-integer $t > 0$ we may differentiate (C.5.14) to obtain

$$g_\theta(t) = \theta t^{\theta-1} \int_{t-1}^{\infty} (s+1)^{-\theta} \mathrm{d}G_\theta(s). \tag{C.5.16}$$

Rearranging (C.5.16) and then differentiating once more yields that

$$t^{1-\theta} g_\theta(t) = \theta \int_{t-1}^{\infty} \frac{g_\theta(s)}{(s+1)^\theta} \mathrm{d}s,$$

and so

$$t^{1-\theta} g'_\theta(t) + (1-\theta) t^{-\theta} g_\theta(t) = -\theta t^{-\theta} g_\theta(t-1),$$

and further rearrangement gives us (C.5.12) for non-integer t. Finally, since probability density functions are defined only modulo a set of measure zero we may *define* the density function g_θ by (C.5.16) for integer t; with this definition we see from (C.5.16) and (C.5.12) that g_θ is continuous on the whole interval $(0, \infty)$ and differentiable on the interval $(1, \infty)$. □

Remark. From (C.5.12), we see that, for $t > 1$, $g'_\theta(t)$ is negative when $(\theta - 1) g_\theta(t) - \theta g_\theta(t-1) < 0$. This is true for all $t > 1$ if $\theta \leq 1$, and so, for $0 < \theta \leq 1$, g_θ is a decreasing function for $t > 1$. For $\theta > 1$, g_θ is eventually decreasing.

Corollary C.5.1 *The generalized Dickman distribution with parameter $\theta = 1$ has a probability density function given by*

$$g_1(x) = e^{-\gamma} \rho(x), \quad x \in \mathbf{R}, \tag{C.5.17}$$

where ρ is the Dickman function.

Proof. By the case $\theta = 1$ of Proposition C.5.1, the probability density function g_1 of the GD(1) distribution satisfies Dickman's equation (C.3.3), and since g_1 must be normalized to be a probability density function, by (C.3.4) it is given by (C.5.17), as required. □

Returning to the case of general $\theta > 0$, define the constant κ_θ by

$$\kappa_\theta := E[(1+X)^{-\theta}], \qquad X \sim \mathrm{GD}(\theta).$$

The constant κ_θ, $\theta > 0$, is actually given by

$$\kappa_\theta = \frac{e^{-\theta\gamma}}{\Gamma(\theta+1)}; \tag{C.5.18}$$

see, for example, [67] or [138]. In particular, $\kappa_1 = e^{-\gamma}$ and $\kappa_2 = e^{-2\gamma}/2$. We also note that $\kappa_\theta = \kappa_1^\theta/\Gamma(\theta+1)$.

The next result gives expressions for the GD(θ) density and distribution functions obtained piecewise on the unit intervals of the positive real line, where the piecewise components are given recursively by an integral recursion relation, which can sometimes be solved explicitly.

Proposition C.5.2 *Let g_θ and G_θ denote the probability density and cumulative distribution function, respectively, of the* GD(θ) *distribution. Then $g_\theta(t) = G_\theta(t) = 0$ for $t \le 0$, and the functions $g_\theta(t)$ and $G_\theta(t)$ can be expressed piecewise over the unit intervals $t \in [n, n+1]$ for $n \in \mathbf{N}$ as*

$$g_\theta(t) = \begin{cases} \theta\kappa_\theta t^{\theta-1} & \text{if } 0 < t \le 1 \\ \left(\frac{t}{n}\right)^{\theta-1} g_\theta(n) - \theta t^{\theta-1}\int_{n-1}^{t-1} \frac{g_\theta(s)}{(s+1)^\theta}\mathrm{d}s & \text{if } n \le t \le n+1 \ (n \in \mathbf{N}) \end{cases} \tag{C.5.19}$$

and

$$G_\theta(t) = \begin{cases} \kappa_\theta t^\theta & \text{if } 0 < t \le 1 \\ G_\theta(t-1) + \frac{t}{\theta} g_\theta(t) & \text{if } t \ge 1 \end{cases} \tag{C.5.20}$$

Proof. For both g_θ and G_θ, the case $t \le 0$ follows from Proposition C.5.1, and the case $0 < t \le 1$ follows from (C.5.15).

Suppose $n \le t \le n+1$ for $n \in \mathbf{N}$. Then equation (C.5.16) yields that

$$t^{1-\theta} g_\theta(t) - n^{1-\theta} g_\theta(n) = -\theta \int_{n-1}^{t-1} \frac{g_\theta(s)}{(s+1)^\theta}\mathrm{d}s.$$

Rearranging this gives us (C.5.19). Substituting in for the integral in equation (C.5.14) from equation (C.5.16) gives

$$\theta\left(G_\theta(t) - G_\theta(t-1)\right) = t g_\theta(t),$$

and (C.5.20) follows. □

The integrals that we are required to perform to obtain expressions for $g_\theta(t)$ and $G_\theta(t)$ with $t \in [n, n+1]$ and $n \geq 1$ get successively more complicated as n increases, and appear to be intractable for $n \geq 2$. However, one can make progress in the $n = 1$ case. By (C.5.19) we have that for $1 \leq t \leq 2$,

$$g_\theta(t) = \theta\kappa_\theta t^{\theta-1} - \theta t^{\theta-1}\int_0^{t-1}\frac{\theta\kappa_\theta s^{\theta-1}}{(s+1)^\theta}\mathrm{d}s = \theta\kappa_\theta t^{\theta-1}\left(1 - \theta\int_1^t\frac{(u-1)^{\theta-1}}{u^\theta}\mathrm{d}u\right). \quad \text{(C.5.21)}$$

In particular, for $\theta = 1$ we see that equation (C.5.21) reduces to

$$g_1(t) = \kappa_1(1 - \log t), \qquad 1 \leq t \leq 2 \quad \text{(C.5.22)}$$

and using (C.5.20) we obtain

$$G_1(t) = \kappa_1(2t - t\log t - 1), \qquad 1 \leq t \leq 2, \quad \text{(C.5.23)}$$

while for $\theta = 2$ and $1 \leq t \leq 2$ we obtain

$$\begin{aligned} g_2(t) &= 2\kappa_2 t\left(1 - 2\int_1^t\frac{u-1}{u^2}\mathrm{d}u\right) = 2\kappa_2 t\left(1 - 2\left(\log t + \frac{1}{t} - 1\right)\right) \\ &= 2\kappa_2\left(3t - 2t\log t - 2\right), \end{aligned} \quad \text{(C.5.24)}$$

and then

$$G_2(t) = \kappa_2(4t^2 - 4t - 2t^2\log t + 1), \qquad 1 \leq t \leq 2. \quad \text{(C.5.25)}$$

For general θ, we have that

$$\int\frac{s^{\theta-1}}{(s+1)^\theta}\mathrm{d}s = \frac{s^\theta}{\Gamma(\theta)}\sum_{k=0}^{\infty}\frac{\Gamma(\theta+k)(-s)^k}{(\theta+k)k!},$$

so that for $1 \leq t \leq 2$,

$$g_\theta(t) = \theta\kappa_\theta t^{\theta-1} - \theta^2\kappa_\theta t^{\theta-1}\left[\frac{(t-1)^\theta}{\Gamma(\theta)}\sum_{k=0}^{\infty}\frac{\Gamma(\theta+k)(-(t-1))^k}{(\theta+k)k!}\right]. \quad \text{(C.5.26)}$$

C.6 The generalized Dickman function

The density function g_θ also appears in connection with the Poisson-Dirichlet distribution with parameter $\theta > 0$ and with a generalization of Dickman's function; see e.g. [24, 71]. Define the function ρ_θ such that $\rho_\theta(t) = 1$ when $0 \leq t \leq 1$ and ρ_θ satisfies the differential-difference equation

$$t^\theta \rho_\theta'(t) + \theta(t-1)^{\theta-1}\rho_\theta(t-1) = 0, \quad t > 1. \tag{C.6.27}$$

Then

$$g_\theta(t) = \frac{e^{-\gamma\theta}}{\Gamma(\theta)} t^{\theta-1}\rho_\theta(t) = \theta\kappa_\theta t^{\theta-1}\rho_\theta(t), \tag{C.6.28}$$

where we can check that $g_\theta(t)$ is indeed the probability density function of our GD(θ) random variable, as it satisfies the Dickman-type equation (C.5.12). Also, notice that if we integrate (C.6.27) between 1 and ∞, making use of (C.6.28) we obtain $G_\theta(1) = \kappa_\theta$ (compare Proposition C.5.2). One can often deduce results about $\rho_\theta(x)$ by studying $g_\theta(x)$, which is often easier to handle.

As Holst remarks [71], g_θ is the density of an infinitely divisible distribution with Lévy-Khinchine measure $\theta\mathbf{1}\{0 < x < 1\}(1/x)\mathrm{d}x$. See also Section 6.3 of Goldie and Grübel [61], which is concerned with the tail behaviour of a class distributions obtained as sums of products, including the GD distributions.

In fact, the largest component of the Poisson-Dirichlet distribution with parameter θ has distribution function $\rho_\theta(1/x)$. We return to this in section C.7, where we discuss this distribution when $\theta = 1$ (which we call the max-Dickman distribution), since it turns out to describe the limiting distribution of the maximum edge length in the MDST.

C.7 The max-Dickman distribution

As in the case of the GD(θ) distributions, there are many characterizations of the max-Dickman distribution.

Proposition C.7.1 *The following random variables are distributionally equivalent.*

(a) A random variable M satisfying the fixed point equation

$$M \stackrel{\mathcal{D}}{=} \max\{1-U, UM\}, \tag{C.7.29}$$

where U is uniform on $(0,1)$ and independent of the M on the right hand side.

(b) *A random variable M given by*

$$M = \max\{1 - U_1, U_1(1 - U_2), U_1U_2(1 - U_3), U_1U_2U_3(1 - U_4), \ldots\}, \tag{C.7.30}$$

where U_i, $i = 1, 2, 3, \ldots$ are i.i.d. uniform random variables on $(0,1)$.

(c) *A random variable M given by $M = \max\{1 - Y_1, Y_1 - Y_2, Y_2 - Y_3, \ldots\}$, where $Y_1, Y_2, Y_3, \ldots$ are the points of a Poisson point process on $(0,1)$ whose intensity measure has a density $1/x$ (taken in decreasing order).*

(d) *A random variable M given by the largest (and first) component of the Poisson-Dirichlet distribution with parameter 1.*

(e) *A random variable M with distribution function $P[M \leq x] = \rho(1/x)$, where ρ is the Dickman function.*

(f) *A random variable M with the size-biased distribution of $1/(Z+1)$, where $Z \sim$ GD(1).*

We shall say that a random variable given by any of the conditions (a) – (f) in Proposition C.7.1 has the *max-Dickman distribution.* Like the $GD(\theta)$ distribution, the max-Dickman distribution on $(0,1)$ has arisen in various contexts. In particular, due to the characterization (c) above, it accompanies Poisson-Dirichlet limits, which arise in numerous combinatorial structures. A selection of examples are interval splitting problems (see e.g. [16]), the distribution of large prime factors (see e.g. [24, 41, 42, 85]), random polynomials and polynomial factorization algorithms (see e.g. [8, 53]), mathematical population genetics (see e.g. [138]), and the distribution of cycles in random permutations (see e.g. [62, 131]). See also [7, 9].

Proof of Proposition C.7.1. The proof of equivalence of (a) and (b) is similar to that given in the proof of Proposition C.4.1, and is omitted this time round.

Let $Y_1, Y_2, Y_3, \ldots$ be the points of a Poisson point process on $(0,1)$ whose intensity measure has a density $1/x$ (taken in decreasing order). We have seen in the proof of Proposition C.4.1 that the variables $Y_1, Y_2/Y_1, Y_3/Y_2, \ldots$ are independent and uniform on $(0,1)$. If we set $U_1 := Y_1$ and $U_i := Y_i/Y_{i-1}$ for $i \geq 2$, then the U_i are independent $U(0,1)$ variables, and with this definition of the U_is the definitions (b) and (c) are identical.

The equivalence of (c) and (d) follows from the fact that the vector of variables $1 - Y_1, Y_1 - Y_2, Y_2 - Y_3, \ldots$, rearranged in decreasing order, has the Poisson-Dirichlet distribution with parameter 1; see e.g. [42].

Suppose now that M is given by the definition in part (e). Then, following [71], we have for $0 \le t \le 1$ that if U is uniform on $(0,1)$ and independent of M, then

$$P[\max\{1-U, UM\} \le t] = \int_{1-t}^{1} P\left[M \le \frac{t}{u}\right] \mathrm{d}u = \int_{1-t}^{1} \rho\left(\frac{u}{t}\right) \mathrm{d}u$$
$$= t \int_{(1/t)-1}^{1/t} \rho(y)\mathrm{d}y = \rho(1/t),$$

where the last equality follows from (C.5.20) and Corollary C.5.1. Thus, M satisfies (C.7.29).

To check the equivalence of definitions (f) and (e), let $Y = (Z+1)^{-1}$ with $Z \sim \mathrm{GD}(1)$, and let f_Y denote the probability density function of Y. Then, when $0 < t \le 1$,

$$P[Y \le t] = 1 - G_1(t^{-1} - 1),$$

which implies, by Dickman's equation, that

$$f_Y(t) = t^{-2} g_1(t^{-1} - 1) = -t^{-3} g_1'(t^{-1}),$$

so that the size-biased distribution of Y has a probability density function on $(0,1)$ proportional to $-t^{-2} g_1'(t^{-1})$.

On the other hand, M given by definition (e) has probability density function $-x^{-2}\rho'(1/x)$. These two distributions are the same. □

Let h and H respectively denote the probability density and distribution functions of the max-Dickman distribution. We can obtain expressions for h and H from the GD(1) density function g_1. Again, we obtain a piecewise description of the functions, but now the intervals are $[1/(n+1), 1/n]$, $n \in \mathbf{N}$. Note that the cumulative distribution of the limiting variable in Theorem 4.2.3, namely that of the maximum of two independent max-Dickman variables, is given by $H(\cdot)^2$, so the next result provides some partial information about this distribution function.

Proposition C.7.2 *The max-Dickman density and distribution functions h and H are* *given in terms of the* GD(1) *density function* g_1 *as follows:*

$$h(x) = \begin{cases} 0 & \text{if } x \geq 1 \\ \frac{1}{x} & \text{if } \frac{1}{2} \leq x < 1 \\ \frac{1}{x} + \frac{1}{x}\log\left(\frac{x}{1-x}\right) & \text{if } \frac{1}{3} \leq x < \frac{1}{2} \\ \frac{e^\gamma}{x} g_1\left(\frac{1-x}{x}\right) & \text{if } \frac{1}{n+1} \leq x < \frac{1}{n} \end{cases} \quad \text{and } H(x) = \begin{cases} 1 & \text{if } x \geq 1 \\ 1 + \log x & \text{if } \frac{1}{2} \leq x < 1 \\ e^\gamma g_1\left(\frac{1}{x}\right) & \text{if } \frac{1}{n+1} \leq x < \frac{1}{n} \end{cases}$$

for all $n \in \mathbf{N}$, *and with* $h(x) = H(x) = 0$ *for* $x \leq 0$.

Proof. By part (e) of Proposition C.7.1, $H(x) = \rho(1/x)$. Differentiating, we obtain

$$h(x) = -\frac{1}{x^2}\rho'(1/x) = \frac{1}{x}\rho\left(\frac{1}{x} - 1\right) = \frac{e^\gamma}{x} g_1\left(\frac{1-x}{x}\right),$$

where the second equality follows from Dickman's equation. Using the fact that $g_1(x) = e^{-\gamma}$ for $0 \leq x \leq 1$ and $g_1(x) = e^{-\gamma}(1 - \log x)$ for $1 \leq x \leq 2$ then yields that

$$h(x) = \frac{1}{x} \quad (1 \geq x \geq 1/2); \; h(x) = \frac{1}{x}\left(1 - \log\frac{1-x}{x}\right), \quad (1/2 \geq x \geq 1/3),$$

and

$$H(x) = 1 - \log(1/x) = 1 + \log x, \quad (1 \geq x \geq 1/2).$$

This completes the proof. □

A graph of $h(x)$ appears on the cover of [24] (and also on p. 49 of [24]). In fact, Theorem 4.3 of [24] says (within a more general result) that, for $0 < x < 1$,

$$h(x) = \frac{1}{x} \sum_{0 \leq k < x^{-1}-1} \frac{(-1)^k}{k!} J_k\left(\frac{1-x}{x}\right). \tag{C.7.31}$$

In (C.7.31), we set $J_0(x) := 1$ and for $k \in \mathbf{N}$

$$J_k(x) := \int_{C_k(x)} t_1^{-1} \cdots t_k^{-1} \mathrm{d}t_1 \cdots \mathrm{d}t_k,$$

where $C_k(x) \subset \mathbf{R}^k$ is the region

$$C_k(x) := \left\{(t_1, \ldots, t_k) \in \mathbf{R}^k : t_1, \cdots, t_k > 1, \; \sum_{i=1}^k t_i < x\right\}.$$

One can check that (C.7.31) agrees with our Proposition C.7.2; for example, if $1/2 \leq x < 1$, (C.7.31) yields $h(x) = 1/x$.

Our form of $H(x)$ appears to disagree with Proposition 2.1 of [71], and also the discussion in Appendix D of [74]. These state (with a typo in [74] corrected) that for $0 < x < 1$,

$$H(x) = 1 + \sum_{j=1}^{\lfloor 1/x \rfloor} \frac{(\log x)^j}{j!}.$$

It appears these results are incorrect, since the range of integration in the derivations exceeds the simplex; compare Theorem 2 of [138].

Of interest (beyond the context of the MDST) is the largest component M of the Poisson-Dirichlet distribution with parameter θ, for general $\theta > 0$. See, for example, [24, 66, 71, 138]. Then $P[M \leq x] = \rho_\theta(1/x)$, where the function ρ_θ, related to g_θ, is as introduced in Section C.6.

Let $E_1(y)$ denote the exponential integral function,

$$E_1(y) = \int_y^\infty \frac{e^{-x}}{x} \mathrm{d}x = \int_1^\infty \frac{e^{-yx}}{x} \mathrm{d}x.$$

Then, Proposition 2.2 of [71] (with a minor correction to the denominator there) shows that for $k = 1, 2, 3, \ldots$,

$$E[M^k] = \frac{\Gamma(\theta)}{\Gamma(\theta + k)} \int_0^\infty y^{k-1} \exp\left(-y - \theta E_1(y)\right). \tag{C.7.32}$$

In particular, for the $\theta = 1$ case this leads to $E[M] = \int_0^\infty e^{-y-E_1(y)} \mathrm{d}y$, which can be evaluated numerically to give $E[M] \approx 0.6243299$ (see e.g. [138]). $E[M]$ is sometimes known as Golomb's constant, or the Golomb-Dickman constant (see [41] p. 9, [62]; see also [140]).

Griffiths [66] tabulates values for $P[M > x]$ for several values of θ.

Returning to the case with $\theta = 1$, we note that one can show that $E[(M + 1)^{-1}] = E[M]$, and that $E[M^{-k}] = ke^\gamma m_{k-1}$ for $k \in \mathbf{N}$, where $(m_k)_{k \geq 1}$ are the moments of the GD(1) distribution. Thus, using (C.4.9) one can recursively generate the moments of the distribution of M^{-1}, which is yet another distribution that has on occasion been given the term 'Dickman distribution' (see [7]).

C.8 Simulating Dickman-type random variables

Dickman-type distributions are simple to simulate efficiently, due to the distributional fixed-point equation representation. For some results on simulating GD(θ) random variables, see [32]. For some results on perfect simulation in this context, see [40].

Consider the following algorithm to simulate a GD(θ) random variable:

Fix $n \in \mathbb{N}$. Set $i = 1$, $X_0 := 0$. Run the following:

(1) Generate U_i, a uniform random variable on $(0, 1)$. Set $X_i = U_i^{1/\theta}(1 + X_{i-1})$.

(2) If $i = n$, stop and output X_n, else update $i \mapsto i + 1$ and return to (1).

Then, it is clear that the random variable X_n generated by the above algorithm converges almost surely as $n \to \infty$ to the random variable X where

$$X = U_1^{1/\theta} + (U_1U_2)^{1/\theta} + \cdots \sim \mathrm{GD}(\theta),$$

by (C.4.7). Further, we have that

$$X - X_n = (U_1U_2 \cdots U_n)^{1/\theta} W,$$

where

$$W = U_{n+1}^{1/\theta} + (U_{n+1}U_{n+2})^{1/\theta} + \cdots \sim \mathrm{GD}(\theta),$$

by (C.4.7). Thus using (C.4.9) we can obtain rate of convergence results such as

$$E[|X - X_n|^2] = \left(\frac{\theta}{\theta + 1}\right)^n \left(\theta^2 + \frac{\theta}{2}\right).$$

The max-Dickman distribution can be simulated equally efficiently. Consider the following algorithm to simulate a max-Dickman random variable:

Fix $n \in \mathbb{N}$. Set $i = 1$, $X_0 := 0$. Run the following:

(1) Generate U_i, a uniform random variable on $(0, 1)$. Set $X_i = \max\{1 - U_i, U_iX_{i-1}\}$.

(2) If $i = n$, stop and output X_n, else update $i \mapsto i + 1$ and return to (1).

This time, it is clear that the random variable X_n generated by the above algorithm converges almost surely as $n \to \infty$ to the random variable X where (after a relabelling)

$$X = \max\{1 - U_1, U_1(1 - U_2), \ldots\},$$

which has the max-Dickman distribution by (C.7.30).

Bibliography

[1] Abramowitz, M. and Stegun, I.A. (Eds.) (1965) Handbook of Mathematical Functions, National Bureau of Standards, Applied Mathematics Series, **55**. U.S. Government Printing Office, Washington D.C.

[2] Aldous, D.J. and Bandyopadhyay, A. (2005) A survey of max-type recursive distributional equations, *Ann. Appl. Probab.*, **15**, 1047–1110.

[3] Alexander, K.S. (1994) Rates of convergence of means for distance-minimizing subadditive Euclidean functionals, *Ann. Appl. Probab.*, **4**, 902–922.

[4] Alexander, K.S. (1996) The RSW theorem for continuum percolation and the CLT for Euclidean minimal spanning trees, *Ann. Appl. Probab.*, **6**, 466–494.

[5] Alili, S. (1999) Asymptotic behaviour for random walks in random environments, *J. Appl. Probab.*, **36**, 334–349.

[6] Arnold, B.C. and Villaseñor, J.A. (1998) The asymptotic distribution of sums of records, *Extremes*, **1**, 351–363.

[7] R. Arratia (1998) On the central role of scale invariant Poisson processes on $(0,\infty)$, in: Microsurveys in discrete probability (Princeton, NJ, 1997), pp. 21-41, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **41**, AMS, Providence, RI.

[8] Arratia, R., Barbour, A.D. and Tavaré, S. (1993) On random polynomials over finite fields, *Math. Proc. Camb. Phil. Soc.*, **114**, 347–368.

[9] Arratia, R., Barbour, A.D. and Tavaré, S. (1999) On Poisson-Dirichlet limits for random combinatorial structures, *Combin. Probab. Comput.*, **8**, 193–208.

[10] Aspandiiarov, S., Iasnogorodski, R. and Menshikov, M. (1996) Passage-time moments for nonnegative stochastic processes and an application to reflected random walks in a quadrant, *Ann. Probab.*, **24**, 932–960.

[11] Avram, F. and Bertsimas, D. (1993) On central limit theorems in geometrical probability, *Ann. Appl. Probab.*, **3**, 1033–1046.

[12] Bai, Z.-D., Lee, S. and Penrose, M.D. (2005) Rooted edges in a minimal directed spanning tree. Preprint, to appear *Adv. Appl. Probab.*

[13] Bai, Z.-D., Devroye, L., Hwang, H.-K. and Tsai, T.-H. (2005) Maxima in hypercubes, *Random Structures and Algorithms*, **27**, 290–309.

[14] Baldi, P. and Rinott, Y. (1989) On normal approximations of distributions in terms of dependency graphs, *Ann. Probab.*, **17**, 1646–1650.

[15] Barndorff-Nielsen, O. and Sobel, M. (1966) On the distribution of the number of admissible points in a vector random sample, *Theory Probab. Appl.* **11**, 249–269.

[16] Barrera, J. and Huillet, T. (2004) On random splitting of the interval, *Stat. Probab. Lett.*, **66**, 237–250.

[17] Baryshnikov, Yu. and Yukich, J.E. (2005) Gaussian limits for random measures in geometric probability, *Ann. Appl. Probab.*, **15**, 213–253.

[18] Beardwood, J., Halton, J.H. and Hammersley, J.M. (1959) The shortest path through many points, *Proc. Camb. Philos. Soc.*, **55**, 299–327.

[19] Berger, N., Bollobás, B., Borgs, C., Chayes, J., and Riordan, O. (2003) Degree distribution of the FKP model, in: Automata, Languages and Programming: 30th International Colloquium, ICALP 2003, *Lecture Notes in Computer Science*, **2719**, eds. J.C.M. Baeten, J.K. Lenstra, J. Parrow, and G.J. Woeginger, Springer, Heidelberg, 725–738.

[20] Bertoin, J. and Gnedin, A. (2004) Asymptotic laws for nonconservative selfsimilar fragmentations, *Electr. J. Probab.*, **9**, 575–593.

[21] Bhatt, A.G. and Roy, R. (2004) On a random directed spanning tree, *Adv. Appl. Probab.*, **36**, 19–42.

[22] Bickel, P.J. and Breiman, L. (1983) Sums of functions of nearest neighbour distances, moment bounds, and a goodness of fit test, *Ann. Probab.*, **11**, 185–214.

[23] Billingsley, P. (1972) On the distribution of large prime factors, *Period. Math. Hungar.*, **2**, 283–289.

[24] Billingsley, P. (1999) Convergence of Probability Measures, 2nd ed., Wiley.

[25] Bollobás, B. (1979) Graph Theory: An Introductory Course, Springer-Verlag, New York.

[26] Bollobás, B. (2001) Random Graphs, 2nd ed., *Cambridge Studies in Advanced Mathematics*, **73**, Cambridge University Press, Cambridge.

[27] Bollobás, B., Borgs, C., Chayes, J. and Riordan, O. (2003) Directed scale-free graphs, in: Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003), ACM, New York, pp. 132–139.

[28] Bollobás, B. and Riordan, O.M. (2003) Mathematical results on scale-free random graphs, in: Handbook of Graphs and Networks, Wiley-VCH, Weinheim, pp. 1–34.

[29] Bollobás, B. and Riordan, O. (2004) The diameter of a scale-free random graph, *Combinatorica*, **24**, 5–34.

[30] Bose, A., Gangopadhyay, S., Sarkar, A. and Sengupta, A. (2003) Convergence of lower records and infinite divisibility, *J. Appl. Probab.*, **40**, 865–880.

[31] Brito, M.R., Quirox, A.J. and Yukich, J.E. (2002) Graph-theoretic procedures for dimension identification, *J. Mult. Anal.*, **81**, 67–84.

[32] Chamayou, J.F. and Dunau, J.L. (2003) Random difference equations: an asymptotical result, *J. Comput. Appl. Math.*, **154**, 183–193.

[33] Chen, W.-M. and Hwang, H.-K. (2003) Analysis in distribution of two randomized algorithms for finding the maximum in a broadcast communication model, *J. Algor.*, **46**, 140–177.

[34] Comets, F., Menshikov, M. and Popov, S. (1998) Lyapunov functions for random walks and strings in random environment, *Ann. Probab.*, **26**, 1433–1445.

[35] Comets, F. and Popov, S. (2003) Limit law for transition probabilities and moderate deviations for Sinai's random walk in random environment, *Probab. Theory Related Fields*, **126**, 571-609.

[36] Cooper, C. and Frieze, A. (2003) A general model of web graphs, *Random Struct. Algor.*, **22**, 311–335.

[37] Csáki, E. (1978) On the lower limits of maxima and minima of Wiener process and partial sums, *Z. Wahrsch. Verw. Gebiete*, **43**, 205–221.

[38] Darling, D.A. (1953) On a class of problems related to the random division of an interval, *Ann. Math. Stats*, **24**, 239–253.

[39] Deheuvels, P. and Revész, P. (1986) Simple random walk on the line in random environment, *Probab. Theory Related Fields*, **72**, 215–230.

[40] Devroye, L. and Neininger, R. (2002) Density approximation and exact simulation of random variables that are solutions of fixed-point equations, *Adv. Appl. Probab.*, **34**, 441–468.

[41] Dickman, K. (1930) On the frequency of numbers containing prime factors of a certain relative magnitude, *Ark. Math. Astr. Fys.*, **22**, 1–14.

[42] Donnelly, P. and Grimmett, G. (1993) On the asymptotic distribution of large prime factors. *J. London Math. Soc.*, **47**, 395–404.

[43] Donnelly, P. and Joyce, P. (1989) Continuity and weak convergence of ranked and size-biased permutations of the infinite simplex, *Stoch. Proc. Appl.*, **31**, 89–103.

[44] Dorogovstev, S.N. and Medes, J.F.F. (2002) Evolution of networks, *Adv. Phys.*, **51**, 1079–1187.

[45] Durrett, R. (1991) Probability: Theory and Examples, Wadsworth & Brooks/Cole, Pacific Grove, CA.

[46] Dussert C., Rasigni, G., Rasigni, M., Palmari, J. and Llebaria, A. (1986) Minimal spanning tree: A new approach for studying order and disorder, *Phys. Rev. B*, **34**, 3528–3531.

[47] Dussert C., Rasigni, G., Rasigni, M., Palmari, J., Llebaria, A. and Marty, F. (1987) Minimal spanning tree analysis of biological structures, *J. Theor. Biol.*, **125**, 317–323.

[48] Evans, D. and Jones, A.J. (2002) A proof of the Gamma test, *Proc. R. Soc. Lond. A*, **458**, 2759–2799.

[49] Evans, D., Jones, A.J. and Schmidt, W.M. (2002) Asymptotic moments of near-neighbour distance distributions, *Proc. R. Soc. Lond. A*, **458**, 2839–2849.

[50] Fayolle, G., Malyshev, V.A. and Menshikov, M.V. (1995) Topics in the Constructive Theory of Countable Markov Chains, Cambridge University Press.

[51] Feller, W. (1943) The general form of the so-called law of the iterated logarithm, *Trans. Amer. Math. Soc.*, **54**, 373–402.

[52] Feller, W. (1968) An Introduction to Probability Theory and its Applications, Volume I, 3rd ed., *Wiley Series in Probability and Mathematical Statistics*, Wiley.

[53] Flajolet, P. and Goudon, X. (2001) The complete analysis of a polynomial factorization algorithm over finite fields, *J. Algor.*, **40**, 37–81.

[54] Flatto, L. and Konheim, A.G. (1962) The random division of an interval and the random covering of a circle, *SIAM Review*, **4**, 211–222.

[55] Friedman, J.H. and Rafsky, L.C. (1979) Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests, *Ann. Stat.*, **7**, 697–717.

[56] Friedman, J.H. and Rafsky, L.C. (1983) Graph-theoretic measures of multivariate association and prediction, *Ann. Stat.*, **2**, 377–391.

[57] Gangopadhyay, S., Roy, R. and Sarkar, A. (2004) Random oriented trees: a model of drainage networks, *Ann. Appl. Probab.*, **14**, 1241–1266.

[58] Gilbert, E.N. (1959) Random graphs, *Ann. Math. Stats*, **30**, 1141–1144.

[59] Gilbert, E.N. (1961) Random plane networks, *J. Soc. Indust. Appl. Math.*, **9**, 533–543.

[60] Gilbert, E.N. (1965), Random minimal trees, *J. Soc. Indust. Appl. Math.*, **13**, 376–387.

[61] Goldie, C.M. and Grübel, R. (1996) Perpetuities with thin tails, *Adv. Appl. Probab.* **28**, 463–480.

[62] Golomb, S.W. (1964) Random permutations, *Bull. Amer. Math. Soc.*, **70**, 747.

[63] Golosov, A.O. (1983) Limit distributions for random walks in random environments, *Soviet Math. Dokl.*, **28**, 18–22.

[64] Golosov, A.O. (1984) On the localization of random walk in a random medium, *Russ. Math. Surveys*, **39**, 157–158.

[65] Golosov, A.O. (1984) Localization of random walks in one-dimensional random environments, *Commun. Math. Phys.*, **92**, 491–506.

[66] Griffiths, R.C. (1988) On the distribution of points in a Poisson process, *J. Appl. Probab.*, **25**, 336–345.

[67] Hensley, D. (1986) The convolution powers of the Dickman function, *J. London Math. Soc.*, **33**, 395–406.

[68] Henze, N. (1987) On the fraction of random points with specified nearest-neighbour interrelations and degree of attraction, *Adv. Appl. Probab.*, **19**, 873–895.

[69] Henze, N. and Voigt, B. (1992) Almost sure convergence of certain slowly changing symmetric one- and multi-sample statistics, *Ann. Probab.*, **20**, 1086–1098.

[70] Hoare, C.A.R. (1961) Algorithm 64: Quicksort, *Comm. ACM*, **4**, 321.

[71] Holst, L. (2001), The Poisson-Dirichlet distribution and its relatives revisited. Preprint, available at `http://www.math.kth.se/matstat`

[72] Hu, Y. and Shi, Z. (1998) The limits of Sinai's simple random walk in random environment, *Ann. Probab.*, **26**, 1477–1521.

[73] Huang, K. (1987) Statistical Mechanics, 2nd ed., Wiley, New York.

[74] Huillet, T. (2004) On a deposition process on the circle with disorder, *Adv. Appl. Probab.*, **36**, 996–1020.

[75] Huillet, T. (2004) Random partitioning problems involving Poisson point processes on the interval. Preprint, available at `http://arxiv.org/abs/cond-mat/0412166`.

[76] Hwang, H.-K. (1998) Asymptotics of divide-and-conquer recurrences: Batcher's sorting algorithm and a minimum Euclidean matching heuristic, *Algorithmica*, **22**, 529–546.

[77] Hwang, H.-K. and Tsai, T.-H. (2002) Quickselect and Dickman function, *Combinatorics, Probability and Computing*, **11**, 353–371.

[78] Joyce, P. and Tavaré, S. (1995) The distribution of rare alleles, *J. Math. Biol.*, **33**, 602–618.

[79] Kawazu, K., Tamura, Y. and Tanaka, H. (1989) Limit theorems for one-dimensional diffusions and random walks in random environments, *Probab. Theory Related Fields*, **80**, 501–541.

[80] Key, E. S. (1984) Recurrence and transience criteria for random walk in a random environment, *Ann. Probab.*, **12**, 529–560.

[81] Kesten, H., Kozlov, M.V. and Spitzer, F. (1975) A limit law for random walk in random environment, *Compositio Mathematica*, **30**, 145–168.

[82] Kesten, H. and Lee, S. (1996) The central limit theorem for weighted minimal spanning trees on random points, *Ann. Appl. Probab.*, **6**, 495–527.

[83] Kingman, J.F.C. (1975) Random discrete distributions, *J. Roy. Stat. Soc.*, **37**, 1–22.

[84] Kingman, J.F.C. (1993) Poisson Processes, *Oxford Studies in Probability*, **3**, Oxford University Press, Oxford.

[85] Kingman, J.F.C. (2004) The Poisson-Dirichlet distribution and the frequency of large prime divisors. Preprint, available from `http://www.newton.cam.ac.uk/preprints2004.html`.

[86] Kolmogorov, A.N. and Fomin, S.V. (trans. R.A. Silverman) (1975), Introductory Real Analysis, Dover, New York.

[87] Kozlov, M.V. (1973) Random walk in one dimensional random medium, *Theory Probab. Appl.*, **18**, 387–388.

[88] Kruskal, J.B. (1956) On the shortest spanning subtree of a graph and the travelling salesman problem, *Proc. Amer. Math. Soc.*, **7**, 48–50.

[89] Kuksa, A.I. and Šor, N.Z. (1972) The method of estimating the number of conditionally optimal trajectories of discrete separable dynamic programming, (In Russian) *Kibernetika*, **6**, 37–44.

[90] Lamperti, J. (1960) Criteria for the recurrence and transience of stochastic processes I, *J. Math. Anal. Appl.*, **1**, 314–330.

[91] Lamperti, J. (1963) Criteria for stochastic processes II: passage-time moments, *J. Math. Anal. Appl.*, **7**, 127–145.

[92] Lee, S. (1997) The central limit theorem for Euclidean minimal spanning trees I, *Ann. Appl. Probab.*, **7**, 996–1020.

[93] Lee, S. (1999) The central limit theorem for Euclidean minimal spanning trees II, *Adv. Appl. Probab.*, **31**, 969–984

[94] Mahmoud, H.M. and Smythe, R.T. (1998) Probabilistic Analysis of MULTIPLE QUICK SELECT, *Algorithmica*, **22**, 569–584.

[95] McLeish, D.L. (1974) Dependent central limit theorems and invariance principles, *Ann. Probab.*, **2**, 620–628.

[96] McGivney, K. (1997) Probabilistic limit theorems for combinatorial optimization problems, Ph.D. Thesis, Lehigh University.

[97] Meester, R. and Roy, R. (1996) Continuum Percolation, Cambridge University Press, Cambridge.

[98] Menshikov, M.V., Asymont, I.M. and Iasnogorodskii, R. (1995) Markov processes with asymptotically zero drifts, *Problems of Information Transmission*, **31**, 248–261, translated from *Problemy Peredachi Informatsii*, **31**, 60–75 (in Russian).

[99] Menshikov, M.V. and Wade, A.R. (2005) Random walk in random environment with asymptotically zero perturbation. Preprint, to appear *J. Euro. Math. Soc.*

[100] Menshikov, M.V. and Wade, A.R. (2006) Logarithmic speeds for one-dimensional perturbed random walk in random environment. Preprint.

[101] Miles, R.E. (1970) On the homogeneous planar Poisson point process. *Math. Biosci.* **6**, 85–127.

[102] Neininger, R. and Rüschendorf, L. (2004) A general limit theorem for recursive algorithms and combinatorial structures, *Ann. Appl. Probab.*, **14**, 378–418.

[103] Newman, M.E.J. (2003) The structure and function of complex networks, *SIAM Rev.*, **45**, 167–256.

[104] Penrose, M. (2003) Random Geometric Graphs, *Oxford Studies in Probability*, **6**, Clarendon Press, Oxford.

[105] Penrose, M.D. (2001) A central limit theorem with applications to percolation, epidemics and Boolean models, *Ann. Probab.*, **29**, 1515–1546.

[106] Penrose, M.D. (2005) Multivariate spatial central limit theorems with applications to percolation and spatial graphs, *Ann. Probab.*, **33**, 1945–1991.

[107] Penrose, M.D. (2005) Convergence of random measures in geometric probability. Preprint available from `http://arxiv.org/abs/math.PR/0508464`.

[108] Penrose, M.D. and Wade, A.R. (2004) Random minimal directed spanning trees and Dickman-type distributions, *Adv. Appl. Probab*, **36**, 691–714.

[109] Penrose, M.D. and Wade, A.R. (2004) On the total length of the random minimal directed spanning tree. Preprint, available from `http://arxiv.org/abs/math.PR/0409201`.

[110] Penrose, M.D. and Wade, A.R. (2005) Explicit laws of large numbers for random nearest-neighbour type graphs. Preprint.

[111] Penrose, M.D. and Yukich, J.E. (2001) Central limit theorems for some graphs in computational geometry, *Ann. Appl. Probab.*, **11**, 1005–1041.

[112] Penrose, M.D. and Yukich, J.E. (2002) Limit theory for random sequential packing and deposition, *Ann. Appl. Probab.*, **12**, 272–301.

[113] Penrose, M.D. and Yukich, J.E. (2003) Weak laws of large numbers in geometric probability, *Ann. Appl. Probab.*, **13**, 277–303.

[114] Penrose, M.D. and Yukich, J.E. (2005) Normal approximation in geometric probability. To appear in *Stein's Method and Applications*, eds. Louis

H.Y. Chen, A.D. Barbour, Lecture Notes Series, Institute for Mathematical Sciences, Vol. 5, World Scientific, Singapore. Preprint available from http://arxiv.org/abs/math.PR/0409088.

[115] Percus, A.G. and Martin, O.C. (1998) Scaling universalities of kth nearest neighbor distances on closed manifolds, *Adv. Appl. Math.*, **21**, 424–436.

[116] Pyke, R. (1965) Spacings, *J. Royal. Stat. Soc. Ser. B*, **27**, 395–449.

[117] Rachev, S.T. (1991) Probability Metrics and the Stability of Stochastic Models, Wiley.

[118] Rachev, S.T. and Rüschendorf, L. (1995) Mass Transportation Problems Volume II: Applications, Springer.

[119] Ramey, D.B. (1982) A non-parametric test of bimodality with applications to cluster analysis, Ph.D. dissertation, Dept. Statistics, Yale Univ.

[120] Rényi, A. (1976) On the extreme elements of observations, in: Selected papers of Alfred Rényi, Vol. 3, Akadémiai Kiadó, pp. 50–65.

[121] Révész, P. (1990) Random Walk in Random and Nonrandom Environments, World Scientific, Teaneck, NJ.

[122] Rodriguez-Iturbe, I. and Rinaldo, A. (1997) Fractal River Basins: Chance and Self-Organization, Cambridge University Press, Cambridge.

[123] Rohlf, F.J. (1975) Generalization of the gap test for the detection of multivariate outliers, *Biometrics*, **31**, 93–101.

[124] Rösler, U. (1992) A fixed point theorem for distributions, *Stoch. Process. Appl.*, **42**, 195–214.

[125] Rösler, U. and Rüschendorf, L. (2001) The contraction method for recursive algorithms, *Algorithmica*, **29**, 3–33.

[126] Sakhanenko, A.I. (1980) On unimprovable estimates of the rate of convergence in the invariance principle, in: Colloquia Math. Soci. János Bolyai **32**: Nonparametric Statistical Inference, Budapest, Hungary, pp. 779–783.

[127] Sakhanenko, A.I. (1984) On estimates of the rate of convergence in the invariance principle, in: Advances in Probability Theory: Limit Theorems and Related Problems (A.A.Borokov, Ed.), pp. 124–135, Springer, New York.

[128] Sakhanenko, A.I. (1985) Convergence rate in the invariance principle for non-indentically distributed variables with exponential moments, in: Advances in Probability Theory: Limit Theorems for Sums of Random Variables (A.A. Borokov, Ed.), pp. 2–73, Springer, New York.

[129] Seppäläinen, T. and Yukich, J.E. (2001) Large deviation principles for Euclidean functionals and other nearly additive processes, *Probab. Theory Related Fields*, **120**, 309–345.

[130] Shao, Qi-Man (1995) Strong approximation theorems for independent random variables and their applications, *J. Mult. Anal.*, **52**, 107–130.

[131] Shepp, L.A. and Lloyd, S.P. (1966) Ordered cycle lengths in a random permutation, *Trans. Amer. Math. Soc.*, **121**, 340–357.

[132] Sinai, Ya.G. (1982) The limiting behaviour of a one-dimensional random walk in a random medium, *Theory Probab. Appl.*, **27**, 256–268.

[133] Solomon, F. (1975) Random walks in a random environment, *Ann. Probab.*, **3**, 1–31.

[134] Steele, J.M. (1990) Probabilistic and worst case analyses of classical problems of combinatorial optimization in Euclidean space, *Math. Op. Res.*, **15**, 749–770.

[135] Steele, J.M. (1993) Probability and problems in Euclidean combinatorial optimization, *Stat. Sci.*, **8**, 48–56.

[136] Steele, J.M. (1997) Probability Theory and Combinatorial Optimization, Society for Industrial and Applied Mathematics, Philadelphia.

[137] Tenenbaum, G. (1995) Introduction to Analytic and Probabilistic Number Theory, Cambridge University Press, Cambridge.

[138] Watterson, G.A. (1976) The stationary distribution of the infinitely-many alleles diffusion model, *J. Appl. Prob.*, **13**, 639–651.

[139] Watts, D.J. (1999) Small Worlds, Princeton University Press.

[140] Weisstein, E.W. (2005) Golomb-Dickman Constant. From MathWorld – A Wolfram Web Resource. `http://mathworld.wolfram.com/Golomb-DickmanConstant.html`

[141] Whitney, V.K.M. (1972) Algorithm 422: Minimal spanning tree, *Comm. ACM*, **15**, 273–274.

[142] Yukich, J.E. (1998) Probability Theory of Classical Euclidean Optimization Problems, *Lecture Notes in Mathematics*, **1675**, Springer, Berlin.

[143] Zeitouni, O. (2004) Random walks in random environments, XXXI Summer School in Probability, St. Flour (2001), in: *Lecture Notes in Math.*, **1837**, Springer, pp. 193–312.

[144] Zolotarev, V.M. (1976) Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces, *Theory Probab. Appl.*, **21**, 721–737.