Last-passage percolation with general weight distribution

James B. Martin
University of Oxford

December 11, 2005

Abstract

We survey the *directed last-passage percolation* model. For certain specific distributions of the weights, very precise scaling limits have been obtained, which we outline briefly. However, the main focus is on results for general weight distributions. These include: conditions under which laws of large numbers hold for the passage-times; shape theorems for associated growth models; properties concerning passage-times near to the boundary for models in two dimensions; scaling limits which apply in the case of sufficiently heavy-tailed weights. We also describe various aspects of the model of *queues in tandem*, which is one of the motivations for the two-dimensional directed last-passage percolation model with general weights. We also give simulations showing trees of optimal paths for the percolation models and displaying the evolution of the growth models.

1 Introduction

This paper is a survey of the model of *directed last-passage percolation*, concerning directed paths of maximal weight between given endpoints in the lattice $\mathbb{Z}^d$, with random weights at the vertices. This is one of many natural variations on the traditional and extensively studied model of *first-passage percolation*. But the directed last-passage model, especially in the case $d = 2$, has recently received particular attention for several reasons.

First, there are very natural correspondences between the last-passage percolation model and certain queueing networks, namely systems of *queues in tandem*.

Secondly and more recently, very precise scaling laws have been obtained for the two-dimensional model in special cases, namely when the underlying weight distribution is exponential or geometric; the asymptotics obtained are much more detailed than has been possible for analogous first-passage models. In these cases the queueing systems mentioned above correspond to Markovian interacting particle systems such as the *totally asymmetric simple exclusion process*. There are also very close links to the theory of random matrices, and to a variety of topics from combinatorics such as the longest increasing subsequence of a random permutation, the Robinson-Schensted-Knuth correspondence and certain models of non-intersecting random walks.

It is expected that the scalings obtained for the "exactly solvable" cases mentioned above – for example, limiting distributions for appropriately rescaled passage-times, and scaling laws for the shape of optimal paths – should in fact be valid for a very wide class of weight distributions; certainly for the case of any distribution with an exponential tail, and probably also for distributions with a sufficiently light polynomial tail. However, such *universality properties* currently seem to be beyond the reach of rigorous proof. Hence it has been of interest to obtain more restricted universality results which apply in the general case. Such results are the main focus of this paper.

In the next section we describe the model precisely and give some of its basic properties. In Section 3 we give a brief outline of the exactly solvable cases where precise asymptotics have been obtained, and indicate various extensions and related models where an exact solution is also available. From then on we concentrate on describing results which hold for a general underlying weight distribution.

In Section 4 we discuss conditions under which laws of large numbers hold; that is, such that the passage-time to a point grows linearly with the distance from the origin (this is an aspect which is much more significant in the last-passage case than in corresponding first-passage models). We also describe truncation results and concentration of measure estimates that are useful in several later contexts.

Just as in the first-passage case, there are *growth models* which are naturally associated with the last-passage percolation model. These are described in Section 5, and we discuss results concerning the asymptotic shape of the growing cluster under linear rescaling. In Section 6 asymptotics concerning passage-times to points near one of the boundaries in the two-dimensional case are discussed; these are closely related to the model of *Brownian directed percolation* which has recently been studied from a variety of perspectives.

The correspondence between the percolation model and the model of queues in tandem is described in detail in Section 7. The vertex weights in the percolation model correspond to service times in the queueing model; in this context, the study of general underlying distributions for the weights is particularly natural. In Section 8, the case of weight distributions with infinite dth moment is discussed. Now the tail is sufficiently heavy that, for example, the laws of large numbers and shape theorems described above must fail, and we are certainly outside the previously mentioned universality class. However, new universality classes can be described rather explicitly for the case where the tail of the distribution is *regularly varying*; these are constructed in terms of a limiting "continuous last-passage percolation" model. Finally in Section 9 we comment on a few variations and extensions of the model.

Simulations of trees of optimal paths for the percolation model are given in Figures 3.1 and 8.1, and of the evolution of growth models in two and three dimensions in Figures 5.1, 5.2 and 5.3.

2 Preliminaries

With each node $\mathbf{z} = \big(z(1), z(2), \ldots, z(d)\big) \in \mathbb{Z}^d$, associate the *weight* $X(\mathbf{z})$. We will assume that the weights $\{X(\mathbf{z}), \mathbf{z} \in \mathbb{Z}^d\}$ are i.i.d. and non-negative, with some common distribution function F. We will sometimes write X for a generic random variable with distribution F.

A *directed path* in $\mathbb{Z}^d$ is a path each step of which consists of increasing a single coordinate by 1. We use the co-ordinatewise order on $\mathbb{Z}^d$, so that $\mathbf{z}_1 \le \mathbf{z}_2$ iff $\mathbf{z}_1(i) \le \mathbf{z}_2(i)$ for $i = 1, 2, \ldots, d$. Then for $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{Z}^d$, with $\mathbf{z}_1 \le \mathbf{z}_2$, let $\Pi(\mathbf{z}_1, \mathbf{z}_2)$ be the set of directed paths from

$\mathbf{z}_1$ to $\mathbf{z}_2$. We identify a path with the set of points it contains; by convention we exclude the initial point $\mathbf{z}_1$ (but include the final point $\mathbf{z}_2$, unless $\mathbf{z}_1 = \mathbf{z}_2$). Note that all paths in $\Pi(\mathbf{z}_1, \mathbf{z}_2)$ have size (or "length") $||\mathbf{z}_2 - \mathbf{z}_1||$ (where we write $||\mathbf{z}|| = \sum_{i=1}^{d} |z(i)|$).

For $\mathbf{z}_1 \leq \mathbf{z}_2$, define $T(\mathbf{z}_1, \mathbf{z}_2)$, the *last passage time* from $\mathbf{z}_1$ to $\mathbf{z}_2$ to be the maximum weight of all directed paths from $\mathbf{z}_1$ to $\mathbf{z}_2$; that is,

$$T(\mathbf{z}_1, \mathbf{z}_2) = \max_{\pi \in \Pi(\mathbf{z}_1, \mathbf{z}_2)} \sum_{\mathbf{v} \in \pi} X(\mathbf{v}). \tag{2.1}$$

In the case $\mathbf{z}_1 = \mathbf{0}$, we write simply $\Pi(\mathbf{z}) = \Pi(\mathbf{0}, \mathbf{z})$ and

$$T(\mathbf{z}) = T(\mathbf{0}, \mathbf{z}) = \max_{\pi \in \Pi(\mathbf{z})} \sum_{\mathbf{v} \in \pi} X(\mathbf{v}); \tag{2.2}$$

thus $T(\mathbf{z})$ is the maximum weight of all directed paths from the origin to the point $\mathbf{z}$.

We immediately have the following *superadditivity* property: if $\mathbf{z}_1 \leq \mathbf{z}_2 \leq \mathbf{z}_3 \in \mathbb{Z}^d$, then

$$T(\mathbf{z}_1, \mathbf{z}_2) + T(\mathbf{z}_2, \mathbf{z}_3) \leq T(\mathbf{z}_1, \mathbf{z}_3). \tag{2.3}$$

Suppose $\mathbb{E}X < \infty$. Then also $\mathbb{E}T(\mathbf{z}) < \infty$ for all $\mathbf{z} \geq \mathbf{0}$. For $\mathbf{x} \in \mathbb{R}_+^d$, we now define

$$g(\mathbf{x}) = \sup_{n \in \mathbb{N}} \frac{1}{n} \mathbb{E}T\left(\lfloor n\mathbf{x} \rfloor\right) \tag{2.4}$$

(which may be infinite). Here and throughout we write $\mathbb{R}_+^d$ to mean the closed orthant $[0, \infty)^d$ including its boundaries. We sometimes write $g_F(\mathbf{x})$ to emphasise the dependence on the distribution F; we also write $g(x_1, \ldots, x_d)$ for $g(\mathbf{x})$ when $\mathbf{x} = (x_1, \ldots, x_d)$ in order to avoid proliferation of brackets.

The following basic properties of the function g are immediate from this definition and from the superadditivity in (2.3), using (a superadditive version of) Kingman's subadditive ergodic theorem:

Proposition 2.1

(i) For all $\mathbf{x} \in \mathbb{R}_+^d$, $\frac{1}{n} T\left(\lfloor n\mathbf{x} \rfloor\right) \to g(\mathbf{x})$ as $n \to \infty$, a.s. and (if $g(\mathbf{x}) < \infty$) in $\mathcal{L}_1$.

(ii) $g(\alpha\mathbf{x}) = \alpha g(\mathbf{x})$ for all $\alpha > 0$, $\mathbf{x} \in \mathbb{R}_+^d$.

(iii) g is invariant under permutations of the coordinates.

(iv) $g(\mathbf{x}) + g(\mathbf{y}) \leq g(\mathbf{x} + \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^d$.

In particular Proposition 2.1 gives a law of large numbers for the sequence of passage times to points "in direction $\mathbf{x}$"; if $g(\mathbf{x}) < \infty$ then these passage times scale linearly with the distance from the origin.

3 Exactly solvable cases

When $d = 2$, there are certain cases for which the form of the "shape function" g is known exactly. If F is the exponential distribution with mean 1, then

$$g(x, y) = \left(\sqrt{x} + \sqrt{y}\right)^2, \tag{3.1}$$

as was first shown by Rost [45]. If F is the geometric distribution with parameter p (meaning that $\mathbb{P}(X=k)=p(1-p)^{k-1}$ for $k=1,2,\ldots$), then [14]

$$g(x,y)=\frac{x+2\sqrt{xy(1-p)}+y}{p}.$$

In these cases much finer asymptotics are available than simply these laws of large numbers. In [26] it is shown that the distribution of

$$\frac{T(\lfloor xn\rfloor,\lfloor yn\rfloor)-ng(x,y)}{n^{1/3}} \tag{3.2}$$

converges to a non-degenerate limit as $n\to\infty$, namely the "Tracy-Widom" distribution, often denoted by F_2, which also arises as the limiting distribution of the largest eigenvalue (appropriately scaled) of a matrix from the Gaussian Unitary Ensemble.

For geometric distributions, results have also been proved concerning the fluctuations of the optimal path; essentially, that the deviation of the optimal path joining $(0,0)$ and $(\lfloor xn\rfloor,\lfloor yn\rfloor)$ from the straight line joining the two points is on the order of $n^{2/3}$ as $n\to\infty$ ([9], completing a proof outlined in [27]).

In the geometric case, a *process* limit for vectors of passage times to suitable sets of points has also been obtained; in [29], it is shown that

$$\left\{n^{-1/3}\left(T(n+tn^{2/3},n-tn^{2/3})-ng(1,1)\right),t\in\mathbb{R}\right\} \tag{3.3}$$

converges as $n\to\infty$ to a stationary process known as the *Airy process* [44] (whose marginal distribution at fixed times is of course the Tracy-Widom distribution F_2 mentioned above).

By considering alternative definitions of an increasing path, one can obtain other related models which are also "exactly solvable". In our definition of increasing path in the previous section (used throughout the paper) we allowed each coordinate to increase weakly. Suppose instead that both coordinates must increase *strictly*, so that we consider paths $\mathbf{w}^{(1)}$, $\mathbf{w}^{(2)},\ldots,\mathbf{w}^{(k)}$ such that $w^{(1)}(i)<w^{(2)}(i)<\cdots<w^{(k)}(i)$ for $i=1,2$. The shape function for this "strict-strict" model in the case of a Bernoulli weight distribution was found by Seppäläinen [46]. One can also consider a mixed model in which one coordinate must increase strictly while the other may increase weakly. The shape function for this strict-weak model in the case of Bernoulli weights is given in [49], and fluctuation results analogous to (3.2) are given in [28] and [22]; for the case of geometric weights, the shpae function is given in [42].

Another closely related model, rather fundamental within the class of these exactly solvable models, is that of the *longest increasing subsequence of a random permutation*, which can be described in terms closer to the last-passage percolation model as follows. Consider a Poisson process of rate 1 in the plane, and let L_n be the maximal number of Poisson points that can be included in an increasing path from $(0,0)$ to (n,n) consisting of line-segments between the Poisson points. By superadditivity arguments, one has that $L_n/n\to c$ a.s. and in $\mathcal{L}_1$ for some c; Vershik and Kerov [55] showed that $c=2$ (see also [34]). Baik, Deift and Johansson [8] showed that the distribution of $(L_n-2n)/n^{1/3}$ converges to the Tracy-Widom distribution F_2; this is of course analogous to (3.2), and in fact such fluctuation results were proved first for the Poisson model before being transferred (with considerable further effort) to the various lattice models. Just as the totally asymmetric exclusion process corresponds

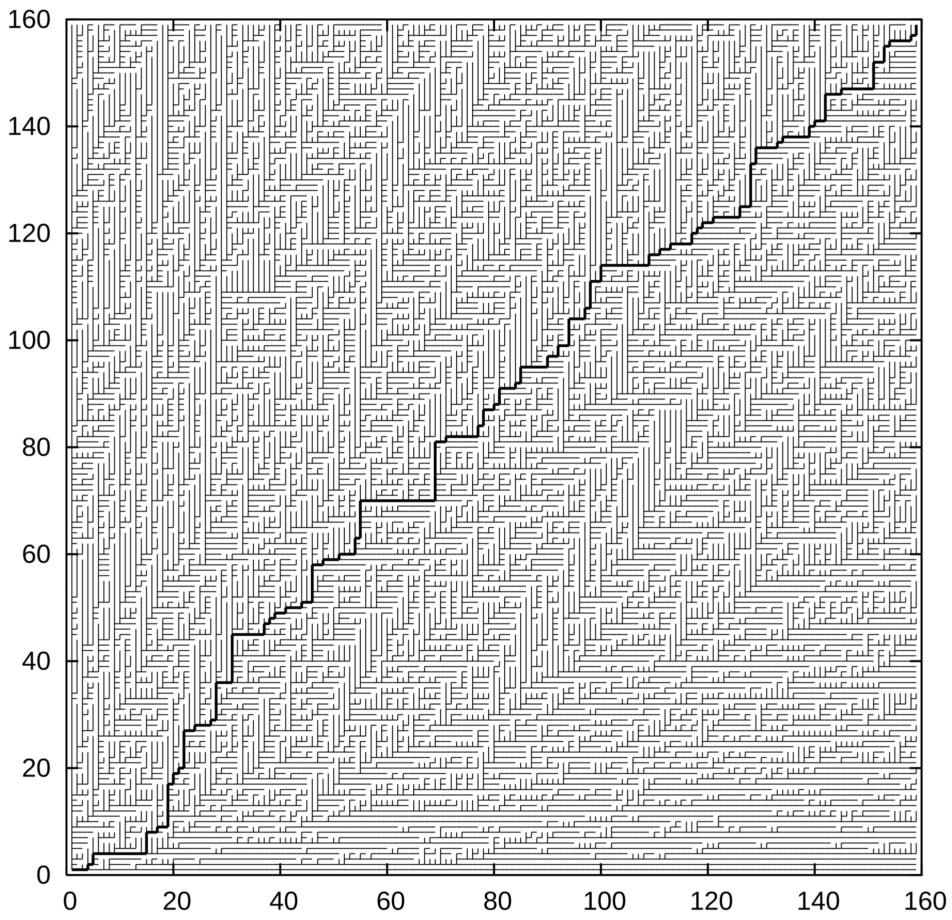

Figure 3.1: A simulation of the last-passage percolation model with F given by an exponential distribution with index 1. The "tree" consisting of optimal paths from $(1,1)$ to (i,j), for all $1 \leq i,j \leq 159$ is displayed; the thickened path is the optimal path from $(1,1)$ to $(159,159)$. The deviation from the diagonal of the optimal path to (n,n) is expected to be of order $n^{2/3}$.

to the last-passage percolation model with exponential weights, there is also an interacting particle system corresponding to the Poisson model, namely the *Hammersley process*, first introduced precisely by Aldous and Diaconis [1], who used it to give a rather simple proof that $c = 2$.

Note that this Poisson model can also be obtained as a scaling limit of various lattice models – for example, by taking the underlying weight distribution to be Bernoulli or geometric and allowing its mean to tend to 0.

More careful descriptions of all these models, and details of the proof of fluctuation reuslts such as (3.2), are beyond the limits of this article. Surveys from a variety of viewpoints, including links to topics in combinatorics and statistical physics, can be found for example in [2, 7, 17, 41]

From now on we concentrate specifically on the last-passage percolation model on the lattice, as defined in the previous section. It is expected that many of the properties described above for the case of exponential and geometric weights should remain valid for a much wider class of weight distributions. In particular, one expects "universality" of the scaling $n^{1/3}$ for the passage time fluctuations and the scaling $n^{2/3}$ for the transversal fluctuations of optimal paths, and indeed of the Tracy-Widom distribution and the Airy process which arise in the limit. The corresponding growth models (discussed in Section 5) should belong to the *Kardar-Parisi-Zhang (KPZ) universality class.* However, rigorous proofs of such universality properties seem out of reach for the moment, and in the rest of the paper we survey the weaker partial results that have been obtained in that direction.

4 Laws of large numbers

We first discuss conditions under which laws of large numbers for the passage times hold; that is, under which the "shape function" g is finite.

Glynn and Whitt [21] showed that g is finite in two dimensions when the weight distribution has an exponential tail, using large deviations estimates.

In [16, 20] a stronger result is obtained, namely that g is finite whenever

$$\mathbb{E}X^d (\log_+ X)^{d+\epsilon} < \infty \tag{4.1}$$

for some $\epsilon > 0$. (In fact, an analogous result is shown for the larger class of *lattice animals* rather than just for directed lattice paths).

The following result from [38] improves marginally on (4.1):

Theorem 4.1 *If*

$$\int_0^\infty \left(1 - F(s)\right)^{1/d} ds < \infty, \tag{4.2}$$

then $g_F(\mathbf{x}) < \infty$ *for all* $\mathbf{x} \in \mathbb{R}_+^d$.

The condition (4.2) is almost optimal, but a small gap still remains between this and the strongest known necessary condition for the finiteness of g, which is that $\mathbb{E}X^d < \infty$. When $\mathbb{E}X^d = \infty$, one has that g is infinite on all the interior of $\mathbb{R}_+^d$; indeed it's a fairly simple calculation (see for example [16]) to show that, more strongly, $\limsup M^{(n)}/n = \infty$ a.s., where $M^{(n)}$ is the largest weight in the box $[0, n]^d$. This case is explored further in Section 8.

Here is a brief outline of the proof of Theorem 4.1. From the superadditivity properties in Proposition 2.1, it is enough to show that $g_F(\mathbf{1}) = \sup_n \mathbb{E}T(n\mathbf{1})/n < \infty$ (where $\mathbf{1} = (1, 1, \ldots, 1)$).

The first step is to consider the case where F is a Bernoulli distribution, taking value 1 with probability p and 0 with probability $1 - p$; the idea is to show that there exists some $c = c(d)$ such that

$$\mathbb{E}_{\mathrm{Ber}(p)} T(n\mathbf{1}) \leq cnp^{1/d} \tag{4.3}$$

for all n and p. This is done using a renormalisation argument. Suppose p is small. We group the lattice sites into boxes of side length approximately $p^{-1/d}$, and so of volume $V \approx p^{-1}$. The distribution of the sum of the weights in such a box is Binomial(V, p) which is approximately

Poisson with mean 1. Any directed path on the original lattice induces a directed path on the "coarser" lattice (the new path consists of the set of boxes which intersect with the original path). Thus one can find an upper bound on the expected path weight for the original problem in terms of a last-passage problem on the coarse lattice with weights which are (approximately) Poisson with mean 1. One obtains

$$\begin{aligned}\mathbb{E}_{\mathrm{Ber}(p)}\frac{T(n\mathbf{1})}{n} &\lessapprox \mathbb{E}_{\mathrm{Poisson}(1)}\frac{T(\lceil np^{1/d}\rceil\mathbf{1})}{n}\\ &\approx p^{1/d}g_{\mathrm{Poisson}(1)}(\mathbf{1}).\end{aligned}$$

Since the Poisson distribution has a (faster than) exponentially decaying tail, large deviations estimates show that g is finite in this case, and we obtain (4.3) as desired.

The result now follows from an exchange of a maximum and an integral:

$$\begin{aligned}\mathbb{E}_F T(n\mathbf{1}) &= \mathbb{E}\max_{\pi\in\Pi(n\mathbf{1})}\sum_{v\in\pi}X(v)\\ &= \mathbb{E}\max_{\pi\in\Pi(n\mathbf{1})}\sum_{v\in\pi}\int_0^\infty I(X(v)>x)dx\\ &\le \int_0^\infty \mathbb{E}\Big[\max_{\pi\in\Pi(n\mathbf{1})}\sum_{v\in\pi}I(X(v)>x)\Big]dx\\ &= \int_0^\infty \mathbb{E}_{\mathrm{Ber}(1-F(x))}T(n\mathbf{1})dx\\ &\le cn\int_0^\infty (1-F(x))^{1/d}dx.\end{aligned}$$

Dividing through by n we obtain that $g(\mathbf{1})$ is finite as desired.

In fact, we have obtained an explicit bound on $g(\mathbf{1})$ (and hence also on $g(\mathbf{x})$ for any $\mathbf{x}\le\mathbf{1}$). It turns out to be very helpful to have such an explicit bound, rather than simply knowing that the function g is finite. For example, for $L>0$, define $T^{(L)}(\mathbf{z})$ and $g^{(L)}(\mathbf{z})$ just as $T(\mathbf{z})$ and $g(\mathbf{x})$, but with each weight $X(\mathbf{z})$ replaced by $X(\mathbf{z})\wedge L$. An argument analogous to the one outlined for Theorem 4.1 establishes the following property:

Lemma 4.2 *There exists $c=c(d)<\infty$ such that for all $L>0$ and all $\mathbf{x}<\mathbf{1}$,*

$$g(\mathbf{x})-g^{(L)}(\mathbf{x})\le c\int_L^\infty (1-F(x))^{1/d}dx.$$

This sort of *truncation* bound is frequently very useful, often allowing much work to be carried out in the case of bounded weights, where tools such as the following *concentration of measure* result are available. This is a corollary of Theorem 8.1 of Talagrand [54], proved in this form as Lemma 5.1 of [38].

Lemma 4.3 *Let $Y_i, i\in I$ be a finite collection of independent random variables, such that*

$$\mathbb{P}(|Y_i|\le L)=1$$

for all $i\in I$. Let $\mathcal{C}$ be a set of subsets of I such that

$$\max_{C\in\mathcal{C}}|C|\le R,$$

and let

$$Z = \max_{C \in \mathcal{C}} \sum_{i \in C} Y_i.$$

Then for any $u > 0$,

$$\mathbb{P}\big(|Z - \mathbb{E}Z| \geq u\big) \leq \exp\left(-\frac{u^2}{64RL^2} + 64\right).$$

In our context, the variables Y_i represent weights, and the sets $C \in \mathcal{C}$ represent, say, directed paths between two given points. Then Z represents the last-passage time between the two points, and we have strong control over the deviation of this passage time from its expectation.

A first example of a situation where this recipe of "truncate, then apply concentration of measure" can be applied is the following continuity result. From the concavity property in Proposition 2.1(iv), it is immediate that g is continuous on the interior $\mathbb{R}_+^d$. However, it's not obvious that g must be continuous at the boundaries.

Theorem 4.4 *If F satisfies (4.2), then the function g is continuous on all of $\mathbb{R}_+^d$, including at the boundaries.*

Theorem 4.4 is proved in [39], using Lemmas 4.2 and 4.3. It would be neater to have a result stating that g was continuous whenever it was finite. However, were there to exist a case in which g was finite while F did not satisfy (4.2) – we don't know whether such a case can occur – then Lemma 4.2 would be useless and so some other ingredient would be necessary.

Finally we remark that there are still several important open questions regarding the form of g; especially concerning *smoothness* and *strict concavity.*

One can construct examples in which g is linear in certain regions; it is sufficient that the distribution of X is bounded and that X attains the maximum possible value with sufficiently high probability (larger than the critical probability for oriented percolation in $\mathbb{Z}^d$). This is essentially the same property observed for the case of first-passage percolation by Durrett and Liggett [19]. Are there any other ways in which g can fail to be strictly concave?

Similary, can g ever fail to be smooth? The only cases where the smoothness of g has been settled are those where g is known exactly (exponential or geometric distributions for $d = 2$). This question is significant in the context of the model of tandem queues; see Section 7.2. Analogous questions for first-passage percolation models have proved notoriously hard to settle, however.

5 Growth models and shape theorems

To obtain the *growth model* associated to the last-passage percolation model, we define $B(t)$, the *last-passage shape at time t*, by

$$B(t) = \left\{\mathbf{x} \in \mathbb{R}_+^d : T\left(\lfloor \mathbf{x} \rfloor\right) \leq t\right\}$$

for all $t \geq 0$.

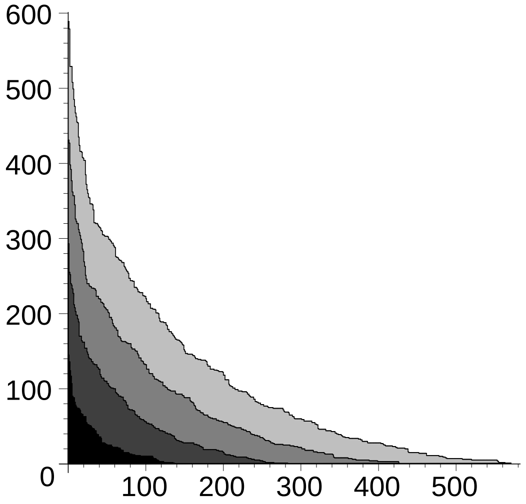

Figure 5.1: Simulation of the last-passage process for $d = 2$ and F exponential with mean 1. The sets $B(t)$ are shown for $t = 150$ (darkest), $300, 450, 600$ (lightest). Here the asymptotic shape B is known to be $\{(x, y) \in \mathbb{R}^2_+ : \sqrt{x} + \sqrt{y} \leq 1\}$.

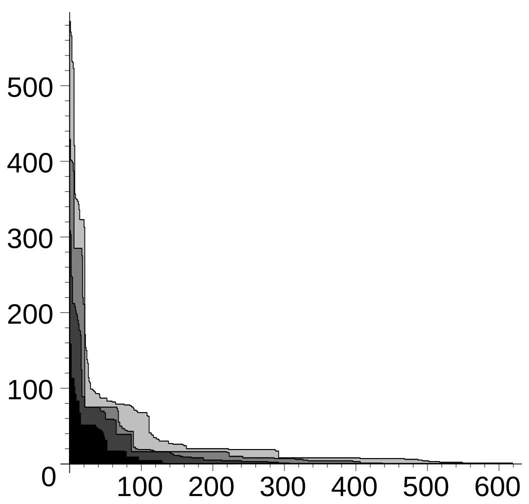

Figure 5.2: As Fig. 5.1, but now for the Pareto distribution $F(x) = \min\left(0, 1 - (3x)^{-3/2}\right)$. Since $\mathbb{E}X^2 = \infty$, the shape theorem no longer applies; the equivalent of the asymptotic shape B consists only of two lines, between the origin and (0,1) and between the origin and (1,0).

$B(t)$ is increasing in t, in the sense that for $0 \leq t_1 \leq t_2$, $B(t_1) \subseteq B(t_2)$. In addition, for all t, $B(t)$ is a decreasing subset of $\mathbb{R}^d_+$: if $\mathbf{x}_1 \leq \mathbf{x}_2$ and $\mathbf{x}_2 \in B(t)$, then also $\mathbf{x}_1 \in B(t)$.

$B(t)$ may be called the "covered set at time t". The model can alternatively be described in terms of the local growth rules governed by the weights $X(\mathbf{z})$. The square $\mathbf{z} \in \mathbb{Z}^2_+$ (or, more precisely, the square $\mathbf{z} + [0,1)^2 \in \mathbb{R}^2_+$) is "added" to the covered set at time $T(\mathbf{z})$. Before this can happen, the squares to the left of $\mathbf{z}$ and below $\mathbf{z}$ must already have been added; once this condition is fulfilled, namely at time $\max\left\{T\left(\mathbf{z} - (0,1), \mathbf{z} - (1,0)\right)\right\}$, the square waits a random time $X(\mathbf{z})$ until it joins the set.

We further define a subset B of $\mathbb{R}^d_+$ by

$$B = \{\mathbf{x} : g(\mathbf{x}) \leq 1\}.$$

By Proposition 2.1(iv), B is concave. B is the *asymptotic shape* for the process $B(t)$, in the sense of the following *shape theorem*, which is analogous to well-known results for undirected first-passage percolation models (see e.g. Cox and Durrett [15] and Kesten [30]).

Theorem 5.1 *Suppose F satisfies (4.2). Then for any $\epsilon > 0$,*

$$(1 - \epsilon)B \subseteq \frac{B(t)}{t} \subseteq (1 + \epsilon)B$$

for all sufficiently large t, with probability 1.

Essentially, Theorem 5.1 says that the convergence in the Law of Large Numbers for direction $\mathbf{x}$ given by Proposition 2.1(i) occurs uniformly in the direction, with probability 1.

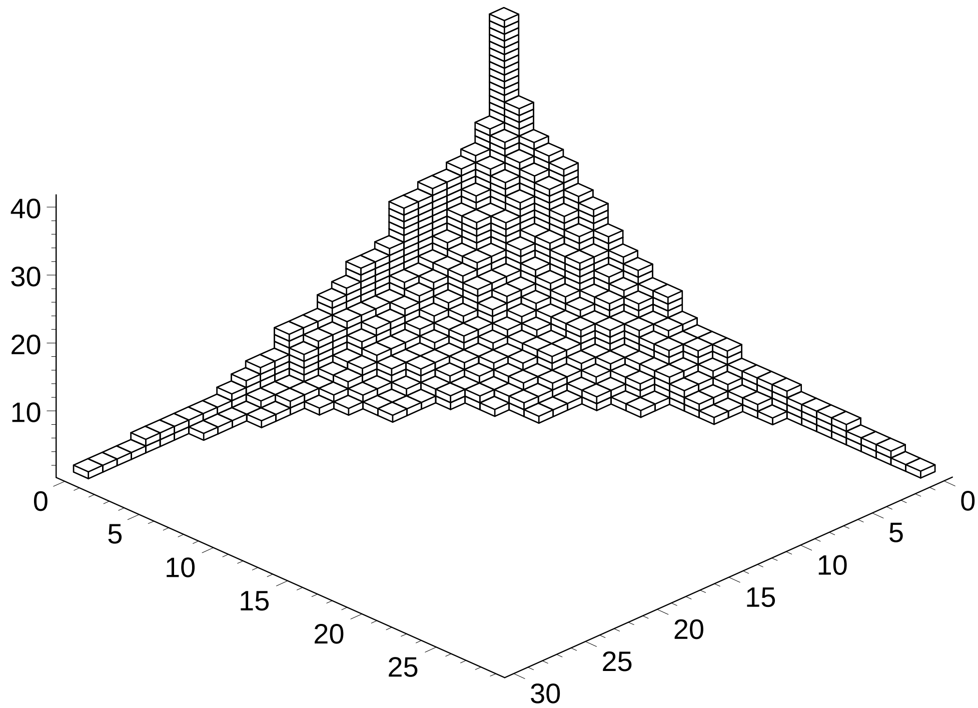

Figure 5.3: A simulation showing $B(30)$ for $d = 3$ and F exponential with mean 1.

Theorem 5.1 is proved in [39]. Again, the approach is to truncate the weights using Lemma 4.2, and apply the concentration of measure result of Lemma 4.3 to bound the probability that $T(\mathbf{z})$ deviates by more than $\epsilon||\mathbf{z}||$ from its expectation. Since the bound obtained decays exponentially in $||\mathbf{z}||$, one can then sum over all $\mathbf{z} \in \mathbb{Z}_+^d$ and obtain a finite quantity; applying the Borel-Cantelli Lemma then gives the uniform convergence desired.

Note that if the weight distribution is geometric or exponential, then the process $B(t)$ is Markov (in discrete or continuous time respectively). In $d = 2$, the function g is known in these cases (see Section 3) and so B is also known: for example, in the case of exponential weights with mean 1, we have $B = \{(x_1, x_2) \in \mathbb{R}_+^2 : \sqrt{x_1} + \sqrt{x_2} \leq 1\}$. For example, in the exponential case, each square outside the covered set whose neighbours below and to the left are already included is added at rate 1. In fact, the process $B(t)$ is also Markov when the weight distribution is Bernoulli, taking values 0 and 1; in this case, however, it has not yet proved possible to obtain detailed information about the function g.

Simulations of these growth models can be found in Figures 5.1, 5.2 and 5.3.

6 Asymptotics at the boundary for $d = 2$

In Theorem 4.4 we saw that, under condition (4.2), the asymptotic shape is continuous at the boundaries of the orthant $\mathbb{Z}_+^d$.

In two dimensions, more extensive universality results are available concerning the behaviour of paths near to the boundary of $\mathbb{Z}_+^2$.

The intuition behind these results is the following. Consider a path from the origin to the point (n, k) where n is large and k/n is small. Such a path must consist mostly of long "horizontal sections" where the second coordinate remains constant. By the central limit theorem, the sum of the weights along such a horizontal section follows approximately a normal distribution, with parameters which depend only on the mean and variance of the underlying weight distribution.

In [39] the following result concerning the behaviour of g near the boundary is proved:

Theorem 6.1 *Let $d = 2$. Suppose that F has mean μ and variance σ^2, and satisfies (4.2). Then as $\alpha \downarrow 0$,*

$$g(1, \alpha) = \mu + 2\sigma\sqrt{\alpha} + o\big(\sqrt{\alpha}\big). \tag{6.1}$$

Theorem 6.1 concerns the case where the ratio k/n is small but fixed as $n \to \infty$. Another regime that has been considered by several authors is that where k behaves as a power n^a for some $a < 1$. Glynn and Whitt showed in [21] that if the weight distribution has mean μ and variance σ^2 and has an exponentially decaying tail, then

$$\frac{T\big(n, \lfloor n^a \rfloor\big) - n\mu}{\sigma n^{(1+a)/2}} \to \alpha \tag{6.2}$$

in probability, as $n \to \infty$, for some constant α independent of the distribution and of the value of a. They conjectured that $\alpha = 2$, and this was first proved by Seppäläinen in [47], by developing a variational coupling between the model with exponential weight distribution and a collection of totally asymmetric exclusion processes with suitably chosen initial states.

The convergence in (6.2) can be extended to give a universality result for the rescaled asymptotic distribution itself:

Theorem 6.2 *Suppose that $\mathbb{E}X^p < \infty$ for some $p > 2$ and let $\mu = \mathbb{E}X$ and $\sigma^2 = \operatorname{Var} X$. Then for all a with $0 < a < \frac{6}{7}\left(\frac{1}{2} - \frac{1}{p}\right)$,*

$$\frac{T\left(n, \lfloor n^a \rfloor\right) - n\mu - 2\sigma n^{\frac{1+a}{2}}}{\sigma n^{\frac{1}{2} - \frac{a}{6}}} \to F_2$$

in distribution as $n \to \infty$ (where F_2 is the GUE Tracy-Widom distribution discussed in Section 3).

Theorem 6.2 is proved by Bodineau and Martin in [12], by coupling the discrete model with a *Brownian directed percolation* model which arises as a natural limiting object in the context of the heuristics described above. The Brownian directed percolation model is defined as follows. Let $B^{(r)}_t$, $t \geq 0$, $r = 0, 1, 2, \ldots$, be a sequence of independent Brownian motions, and for each $t > 0$, $n = 1, 2, \ldots$, define the last-passage time random variable $L(n, t)$ by

$$L(n, t) = \sup_{0 = u_0 < u_1 < \cdots < u_n = t} \sum_{r=0}^{n-1} \left(B^{(r)}_{u_r+1} - B^{(r)}_{u_r}\right).$$

This model has been much studied recently, from a variety of points of view; see e.g. [11, 22, 25]. The connection to random matrix theory is particularly explicit. For example, for any n, the process $L(n, t), t \geq 0$ has the same distribution as the process of the largest eigenvalue in a *Hermitian Brownian motion*; this is a matrix-valued process whose entries perform Brownian motions, and whose marginal at any fixed time is a random matrix from the GUE. See for example O'Connell [41] for a survey. In particular, as $n \to \infty$, the distribution of $L(n,t)$, appropriately scaled, converges to the Tracy-Widom GUE distribution. (It is not important how t varies with n; indeed, by Brownian scaling, one has simply that $L(n,t) \stackrel{d}{=} \sqrt{t}L(n,1)$ for all t).

The coupling used to prove Theorem 6.2 is the *strong coupling* of Komlós, Major and Tusnády (see e.g. [31] – the same strong coupling to a Brownian model had already been used by Glynn and Whitt to establish (6.2)). For a given "level" r, consider the process of partial sums of weights along a given level: $S_i^{(r)} = \sum_{l=0}^{i-1} X(l,r)$. For each r and n, the KMT strong coupling makes it possible to couple the first n weights on level r with a Brownian motion $B^{(r)}$ in such a way that

$$\mathbb{P}\Big(\sup_{0\le t\le n}\Big|B_t^{(r)} - S_{\lfloor t\rfloor}^{(r)}\Big| > x\Big) \le Cnx^{-p} \tag{6.3}$$

for $x \in [n^{1/p}, n^{1/2}]$.

Letting $k = n^{\lfloor a\rfloor}$ and carrying out this coupling for each level $0 \le r < k$, one can obtain a coupling of $T(n,k)$ to a Brownian last-passage time random variable $L(k,t)$. For a and p within an appropriate range, one can then use (6.3) to show that the error in the approximation of the discrete passage-time by the appropriately scaled Brownian passage-time is of smaller order than the fluctuations of the Brownian passage-time itself. From this it follows that the limiting distribution of the two quantities is the same (namely F_2).

Essentially the same result was obtained independently by Baik and Suidan [10], again by means of a coupling of the discrete model to the Brownian model, but using a Skorokhod embedding argument instead of the KMT coupling.

In passing, one should mention that the bound $a < 3/7$ in Theorem 6.2 is merely an artefact of the coupling method described above and seems extremely unlikely to have any real significance. One would expect the result to hold for all $a < 1$, certainly for distributions with an exponential tail, and perhaps even for all $p > 2$. See [12] for a further discussion.

7 Queues in tandem

Consider a series of single-server queues, labelled by integers. Each customer moves through the queues in order, moving to queue $k+1$ upon finishing service at queue k. The customers are also labelled by integers; the queues are first-in-first-out so that customer n precedes customer $n+1$ at each queue.

Let $X(n,k)$ be the service time required by customer n at queue k. Let $D(n,k)$ be the time at which customer n finishes service at queue k. The time at which customer n can begin service at queue k is then $\max\big(D(n,k-1), D(n-1,k)\big)$. (This is because the following two events must occur before customer n starts service at queue k: customer n must complete service at queue $k-1$, and customer $n-1$ must complete service at queue k). Thus we have the following set of recurrences for the variables D in terms of the variables X:

$$D(n,k) = X(n,k) + \max\big(D(n,k-1), D(n-1,k)\big). \tag{7.1}$$

Consider for the moment the following boundary conditions:

$$D(0,0) = 0 \text{ and } D(n,k) = 0 \text{ whenever } n < 0 \text{ or } k < 0. \tag{7.2}$$

The interpretation is as follows: the customers and queues are both labelled starting from 0. Initially, every queue is empty except queue 0 which contains all the customers. The first event in the system occurs at time 0: customer 0 moves from queue 0 into queue 1.

It is easy to see that (7.1) and (7.2) together uniquely determine the collection $\{D(n,k), (n,k) \in \mathbb{Z}_+^2\}$ in terms of the collection $X(n,k), (n,k) \in \mathbb{Z}_+^2$.

In fact, one has precisely that $D(n,k) = T(n,k)$, the maximal weight of a directed path from $(0,0)$ to (n,k), as defined by equation (2.2). This follows easily by induction from (7.1); any directed path ending at (n,k) is the union of the point (n,k) itself and a directed path ending either at $(n-1,k)$ or at $(n,k-1)$. Hence we have a direct correspondence between the model of queues in tandem and the directed last-passage percolation model for $d = 2$. This correspondence was first observed by Muth [40]. By extension, the dynamics of the queueing system correspond precisely to the dynamics of the growth model described in Section 5. Of course, the asymptotic results proved earlier for the percolation model translate directly to the queueing model. In particular, if the service times $X(n,k)$ are i.i.d. with a common distribution F satisfying (4.2), then $n^{-1}D(n, \lfloor xn \rfloor) \to g_F(1,x)$ a.s. as $n \to \infty$, for any $x > 0$. This is a simple form of *hydrodynamic limit* for the queueing model.

It is clear that the percolation model is invariant under the exchange of n and k. In the queueing context, this corresponds to a duality between the sequence of customers and the sequence of queues. In Section 6 we considered the asymptotics of the passage-times to the point (n,k) where n grows much faster than k. In the queueing context, this corresponds to considering the asymptotic behaviour of the first few queues, or, by duality, of the "front of the wave" of customers progressing through the system. Various aspects of these "heavy-traffic" and "light-traffic" regimes are studied by Glynn and Whitt [21]; for example, they obtain the Brownian directed percolation model described in Section 6 as the limit in a functional central limit theorem for the evolution of the front of the wave.

If $X(n,k)$ are i.i.d. exponential, then the state of the queueing system evolves as a Markov process. In the language of queueing theory, we speak of "$./M/1$ queues in tandem". There is a close correspondence between this case and the "totally asymmetric simple exclusion process", or TASEP. In the TASEP, each site of $\mathbb{Z}$ contains either a particle or a hole. Each particle tries to jump to the right at rate 1; a jump succeeds if the site to the right is empty (i.e. contains a hole); in this case the hole and the particle "swap positions".

Consider numbering the particles in the system sequentially, decreasing from left to right, and the holes similarly but increasing from left to right. Let $X(n,k)$ be the total amount of time for which particle n is one site to the left of hole k. Then $X(n,k)$ are indeed i.i.d. exponential. Now let $D(n,k)$ be the time at which particle n swaps positions with hole k. The dynamics of the system are again given by (7.1). In effect, the particles in the TASEP model correspond to customers in the tandem queueing model, and the holes correspond to the queues.

The boundary conditions (7.2) then correspond to the initial condition for the TASEP considered by Rost [45], in which every site $i \leq 0$ contains a particle and every site $i > 0$ is empty. More general initial conditions for the TASEP can be obtained by setting boundary conditions $D(n,k)=0, (n,k)\in\mathcal{R}$ for more general sets $\mathcal{R}$. See for example [48] for an extensive development of this approach to obtain a variety of scaling limits and large deviations properties for the TASEP.

7.1 External arrival processes

In the previous section we considered systems with a given initial condition at time 0. An alternative class of boundary conditions for the recurrences in (7.1) was introduced by Baccelli, Borovkov and Mairesse [4] in order to model the stationary behaviour of a system of queues in tandem, fed by an external arrival process.

We again consider a half-infinite series of queues, indexed by $k=1,2,3,\dots$, but now the stream of customers will be infinite in both directions. Queue 1 is fed by an arrival process, which we represent by a sequence of interarrival times $A(n), n\in\mathbb{Z}$. Customer 0 arrives at queue 1 at time 0, and the time between the arrival of customer n and the arrival of customer $n+1$ is $A(n)$ for all $n\in\mathbb{Z}$.

Thus we set $D(n,0)=\sum_{i=0}^{n-1}A(i)$ for $n\geq 0$ and $D(n,0)=-\sum_{i=n}^{-1}A(i)$ for $n<0$, and wish to derive $D(n,k), k\in\mathbb{N}, n\in\mathbb{Z}$ using the recurrences in (7.1).

By Loynes' construction [35] for a single queue, we can obtain $\{D(n,1)\}$, the departure process from queue 1, in terms of the arrival process $\{D(n,0)\}$ and the service process $\{X(n,1\}$ by

$$D(n,1)=\sup_{n'\leq n} D(n',0)+\sum_{i=n'}^{n}X(i,1);$$

see for example [5] for a detailed exposition. Applying this construction recursively to queues $1,2,\dots,k$ we obtain

$$\begin{aligned} D(n,k) &= \sup_{n'_0\leq n'_1\leq\cdots\leq n'_k=n}\left\{D(n'_0,0)+\sum_{j=1}^{k}\sum_{i=n'_{j-1}}^{n'_j}X(i,j)\right\} \\ &= \sup_{n'\leq n}\left\{D(n',0)+X(n',1)+T\big((n',1),(n,k)\big)\right\} \\ &= D(n,0)+\sup_{n'\leq n}\left\{X(n',1)+T\big((n',1),(n,k)\big)-\sum_{i=n'}^{n-1}A(i)\right\}. \end{aligned} \tag{7.3}$$

That is, $D(n,k)$ is equal to the maximum, over all points $(n',0)$ on the n-axis, of the sum of the arrival time $D(n',0)$ of customer n' at queue 1 and the maximum weight of a path from $(n',1)$ to (n,k), (now including the weight of its initial point – hence the extra term $X(n',1)$ which occurs).

This construction was carried out in [4] and is closely related to a previous formulation by Szczotka and Kelly [53]. In general it need not be the case that the formula for $D(n,k)$ in (7.3) gives a finite answer. However, if any finite solution to the recurrences (7.1) exists with the given boundary conditions, then (7.3) gives the *minimal* solution, and under various natural sets of assumptions, it is in fact the unique solution.

This construction can be used to model the "stable" state of the queueing system, if such a state exists. If the distribution of $\{A(n),X(n,k),k\geq 1,n\in\mathbb{Z}\}$ is stationary in n, then so is the distribution of the collection $\{D(n,k+1)-D(n,k), D(n,k)-D(n-1,k), k\geq 0, n\in\mathbb{Z}\}$ derived from (7.3). Then we can follow the progress of a "typical customer", say customer 0: the quantity $D(0,k)$ represents the time taken between arrival in the system and the end of service at queue k.

The following result gives conditions under which the collection $D(n,k)$ is indeed well-defined, and a law of large numbers for the speed of a typical customer through the system. As in previous sections, we now assume that all the service times (weights) are i.i.d. The result was proved in [37], slightly extending a result of [4].

Theorem 7.1 *Suppose that $X(n,k)$ are i.i.d. with common distribution F satisfying (4.2). Suppose also that the sequence $\{A(n), n\in\mathbb{Z}\}$ is ergodic with $\mathbb{E}A(n)=\alpha>\mathbb{E}X$. Let $D(n,k)$ be defined by (7.3). Then with probability 1, $D(n,k)<\infty$ for all n, k, and*

$$\frac{D(0,k)}{k}\to\sup_{x\geq 0}\{g_F(1,x)-\alpha x\} \tag{7.4}$$

a.s. as $k\to\infty$.

Thus the system is stable as long as the expected interarrival time is larger than the expected service time; the system admits a positive *throughput*. Of course, the system containing infinitely many queues is an abstraction, but the result implies lower bounds for the throughput of finite systems which are uniform in the size of the system.

The intuition behind the proof of Theorem 7.1 is as follows. Putting $n=0$ and $n'=-\lfloor xk\rfloor$ in equation (7.3), for large k we have that $T\big((n',1),(0,k)\big)\approx kg_F(1,x)$ (by the convergence in Proposition 2.1(i)), and that $\sum_{i=n'}^{-1}A(i)\approx k\alpha x$. Dividing by k, we relate the quantity within the sup in (7.3) to the quantity within the sup in (7.4). To justify the result, we need to show that the convergence as $k\to\infty$ is uniform in x. Again, this may be done using the concentration of measure result of Lemma 4.3.

Under a further mild condition on the arrival process $\{A(n)\}$, one can also obtain convergence in $\mathcal{L}_1$ in (7.4), and show that the expected waiting time of a typical customer at queue k is finite for all k and bounded as $k\to\infty$.

7.2 Fixed points

In the last section we introduced an arrival process to queue 1, represented by the sequence $\{A(n)\}_{n\in\mathbb{Z}}$ of inter-arrival times. Suppose that this sequence is stationary and ergodic, with mean $\alpha=\mathbb{E}A(n)$. Suppose that the service times $\{X(n,k)\}_{n\in\mathbb{Z},k\in\mathbb{N}}$ are i.i.d. with mean $\beta=\mathbb{E}X<\alpha$, and that the collection of service times is independent of the process A.

By Loynes' theorem [35], the sequence of inter-departure times from queue 1, given by $\{D(n,1)\}_{n\in\mathbb{Z}}$, is also a stationary and ergodic process with mean α; then the same is also true of the departure process from any queue k, given by $\{D(n,k)\}_{n\in\mathbb{Z}}$.

If $\{A(n)\}_{n\in\mathbb{Z}}$ and $\{D(n,1)\}_{n\in\mathbb{Z}}$ have the same distribution, we say that this distribution is a *fixed point*. In this case, one of course has that $\{D(n,k)\}_{n\in\mathbb{Z}}$ has the same distribution for all $k\geq 1$; in fact, the collection $\{X(n,k), D(n,k)\}_{n\in\mathbb{Z},k\in\mathbb{N}}$ is now stationary in both n and k (rather than just in n as we already had in the previous section).

Various questions concerning such fixed points are natural both in the context of a single queue and in the context of a system of queues in tandem:

Existence: For a given distribution F of the service times $\{X(n, k)\}$ and a given $\alpha > \beta = \mathbb{E}X$, does there exist an ergodic fixed point arrival process A of mean α?

Uniqueness: If such an ergodic fixed point of mean α exists, is it unique?

Convergence: Assume that there exists a unique ergodic fixed point of mean α. If the arrival process to the first queue is given by some other ergodic process of mean α, is it necessarily the case that the law of $\{D(n, k)\}_{n\in\mathbb{Z}}$ converges weakly to the ergodic fixed point of mean α as $k \to \infty$?

The earliest result in this direction is perhaps *Burke's theorem* [13], which in this context says the following: if F is the exponential distribution with mean β, then for any $\alpha > \beta$, a fixed point is given by interarrival times $A(n)$ which are i.i.d. exponential with mean α. Anantharam [3] showed that this fixed point is unique, and Mountford and Prabhakar [36] proved the convergence property. Similar results hold in the case where F is a geometric distribution.

The distributions in the previous paragraph correspond to the "exactly solvable" cases for the percolation model described in Section 3. In the general case, Prabakhar [43] demonstrated the uniqueness of the fixed point, should it exist, and also established the convergence property.

However, a complete answer to the question of existence is not known. The strongest results to date have been proved by Mairesse and Prabakhar [36]:

Theorem 7.2 *Assume that F satisfies (4.2), and $\mathbb{E}X = \beta$. Let $\mathcal{S}$ be the set of $\alpha \in (\beta, \infty)$ such that there exists an ergodic fixed point of mean α. Then the set $\mathcal{S}$ is closed in (β, ∞), with* $\inf \mathcal{S} = \beta$ *and* $\sup \mathcal{S} = \infty$.

In fact the set $\mathcal{S}$ can be characterised as follows. Recall the shape function g defined at (2.4), and define the function γ from $\mathbb{R}_+$ to $\mathbb{R}_+$ by $\gamma(x) = g((1, x))$. (The RHS of (7.4) can be recognised as the Legendre transform of the function γ). Then $\mathcal{S}$ is equal to the image of the derivative of γ. (The function γ is concave, so its left and right derivatives exist everywhere; if they differ then one considers that they both belong to the image).

Thus there exists a fixed point of mean α for all $\alpha \in (\beta, \infty)$ if and only if the function γ is smooth. As discussed at the end of Section 4, the question of whether this smoothness holds is open except in the special cases of exponential and geometric weights, where the function is known explicitly. A natural conjecture would be that γ is always smooth, and hence that fixed points always exist; but as for analogous models of first-passage percolation, proving such a property would seem to be very difficult.

8 Heavy tails

Up until this point we have been considering weight distributions satisfying (4.2), so that one has linear growth for the passage-times as described in Theorem 4.1, and an asymptotic shape under linear rescaling for the growth model as described in Theorem 5.1.

In [24], on the other hand, cases are considered in which $\mathbb{E}X^d = \infty$; now the tail of the weight distribution is sufficiently heavy that such laws of large numbers and shape theorems fail, and the universality class described in Section 3 certainly no longer applies.

Specifically, suppose that the tail of the weight distribution is *regularly varying* with index $\alpha < d$; that is, for all $t > 0$,

$$\frac{1 - F(tx)}{1 - F(x)} \to t^{-\alpha} \text{ as } x \to \infty.$$

For example, one might have the Pareto distribution with $F(x) = 1 - x^{-\alpha}$ for $x \geq 1$. In [24] a new family of universality classes is obtained for these cases, indexed by the parameter α.

Consider for the moment the case $d = 2$, and consider $T(n, n)$, the last-passage time from the origin to the point (n, n). For $N > 0$, define also $a_N = \sup\{x : F(x) < 1 - 1/N\}$. The next result shows that rescaling $T(n, n)$ by the quantity a_{n^2}, one obtains convergence in distribution to a non-degenerate limit as $n \to \infty$. (For the Pareto distribution above, for example, one has $a_{n^2} = n^{2/\alpha}$).

Theorem 8.1 *There exists a random variable T whose distribution is concentrated on $(0, \infty)$, such that*

$$a_{n^2}^{-1} T(n, n) \to T$$

in distribution as $n \to \infty$. The distribution of T depends on α alone.

There is also a corresponding convergence result for the shape of the optimal path. (For simplicity, assume that the underlying weight distribution F is continuous so that all optimal paths are a.s. unique). Let π_n^* be the optimal path between $\mathbf{1}$ and (n, n), interpolated linearly and rescaled by a factor n so that it may be regarded as a subset of $[0, 1]^2$. We use the Hausdorff metric on subsets of $[0, 1]^2$:

$$d_H(P_1, P_2) = \sup_{x \in P_1} \inf_{y \in P_2} |x - y| + \sup_{x \in P_2} \inf_{y \in P_1} |x - y|.$$

Theorem 8.2 *There exists a "random path" $\pi^* \subset [0, 1]^2$ such that $\pi_n^* \to \pi^*$ in distribution as $n \to \infty$. The distribution of π^* is non-degenerate and depends on α alone.*

In particular, the deviations of the optimal path from the straight line between $\mathbf{1}$ and (n, n) are of order n. See Figure 8.1 for a simulation. This is to be compared with Figure 3.1: for lighter tails, the deviations from the diagonal are expected to be of order $n^{2/3}$, so that the scaling limit π^* would have a degenerate distribution concentrated on the straight line between $(0, 0)$ and $(1, 1)$.

The limits T and π^* can be constructed in terms of a "continuous last-passage percolation" model. The construction of this scaling limit is quite explicit, and we give a brief outline.

First consider the (random) collection of weights $\{X(i, j), 0 \leq i, j \leq n\}$. These are the weights which fall in the box $\{0, 1, \ldots, n\}^2$, and so are available to be used by a path from $\mathbf{1}$ to (n, n). Now rewrite these weights in decreasing order as $M_1^{(n)} \geq M_2^{(n)} \geq \cdots \geq M_{(n+1)^2}^{(n)}$. Thus $M_k^{(n)}$ is the kth largest weight in the box $\{0, 1, \ldots, n\}^2$.

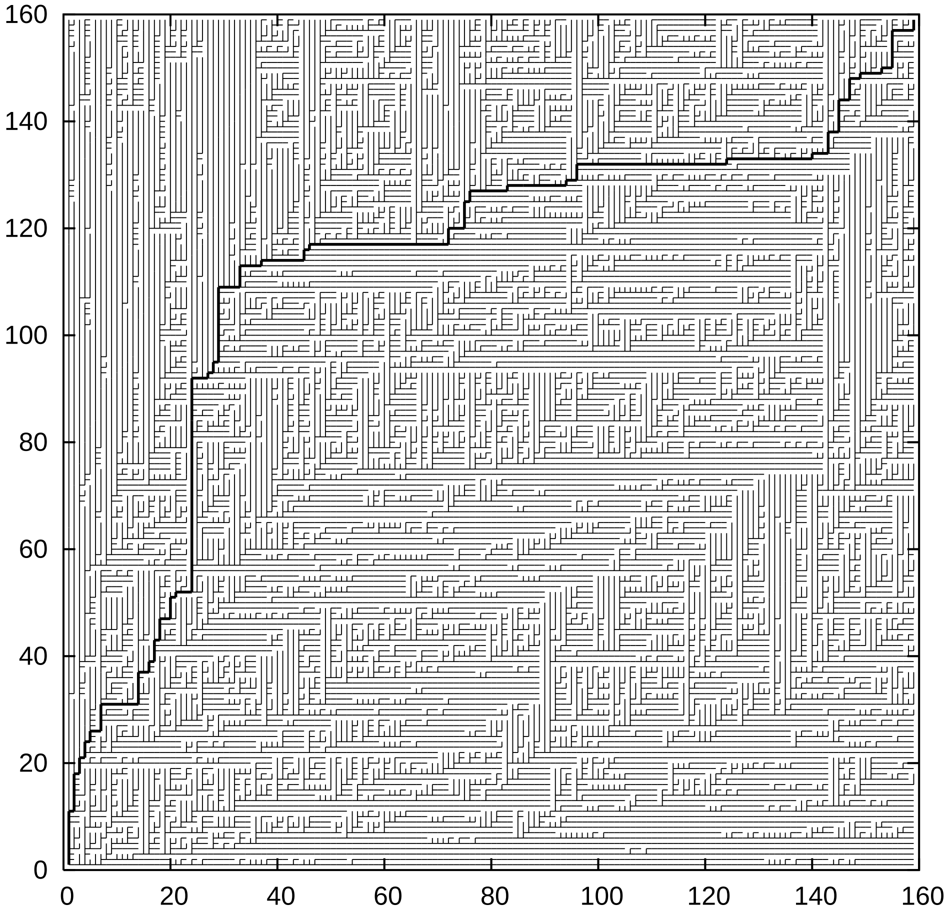

Figure 8.1: A simulation of the last-passage percolation model with F given by a Pareto distribution with index 1. The "tree" consisting of optimal paths from $(1,1)$ to (i,j), for all $1 \leq i,j \leq 159$ is displayed; the thickened path is the optimal path from $(1,1)$ to $(159,159)$. The deviation from the diagonal of the optimal path to (n,n) is now of order n; compare with the case of exponential weights in Figure 3.1.

From classical extreme value theory, one has the following distributional convergence: for any k,

$$\begin{aligned} a_{n^2}^{-1}\left(M_1^{(n)}, M_2^{(n)}, \ldots, M_k^{(n)}\right) &\xrightarrow{d} (M_1, M_2, \ldots, M_k) \\ &:= \left(E_1^{-1/\alpha}, (E_1+E_2)^{-1/\alpha}, \ldots, (E_1+\cdots+E_k)^{-1/\alpha}\right), \quad (8.1) \end{aligned}$$

where $(E_i)_{i\geq 1}$ are i.i.d. exponential random variables with mean 1. (Note already that $T(n,n) \geq M_1^{(n)}$, so (8.1) shows immediately that $a_{n^2}^{-1}$ is a lower bound for the order of $T(n,n)$; Theorem 8.1 implies that in fact $M_1^{(n)}$ and $T(n,n)$ are of the same order).

Define also $\tilde{Y}_i^{(n)}$ to be the point in $\{0,1,\ldots,n\}^2$ where the ith largest weight $M_i^{(n)}$ is located. Then $\left(\tilde{Y}_1^{(n)}, \tilde{Y}_2^{(n)}, \ldots, \tilde{Y}_{(n+1)^2}^{(n)}\right)$ is a uniformly random permutation of the set

$\{0,1,\ldots,n\}^2$. Define $Y_i^{(n)} = \tilde{Y}_i^{(n)}/n \in \{0,1/n,\ldots,1\}^2$. Then for any k it's easy to see that

$$\left(Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_k^{(n)}\right) \xrightarrow{d} (Y_1, Y_2, \ldots, Y_k) \tag{8.2}$$

as $n \to \infty$, where $(Y_i)_{i\geq 1}$ are i.i.d. uniform random variables on the box $[0,1]^2$.

In the light of the convergence in (8.1) and (8.2), the limiting continuous model is defined as follows. Let $(M_1, M_2, \ldots) = \left(E_1^{-1/\alpha}, (E_1+E_2)^{-1/\alpha}, \ldots\right)$ as in (8.1), where $(E_i)_{i\geq 1}$ are i.i.d. exp(1). Let $Y_1, Y_2, \ldots$ be i.i.d. uniform locations in the box $[0,1]^2$. Now imagine Y_i as the location of the ith weight M_i. (These locations are of course dense in $[0,1]^2$ a.s.). Define $\mathcal{C}$ to represent the set of increasing paths, by

$$\mathcal{C} = \{A \subset \mathbb{N} : Y_i \sim Y_j \text{ for all } i,j \in A\},$$

where we write $y \sim y'$ if *either* $y \leq y'$ coordinatewise *or* $y' \leq y$ coordinatewise (the locations y and y' are "compatible"; they may both be included in the same increasing path). Now define T to be the maximal weight of such an increasing path:

$$T = \sup_{A\in\mathcal{C}} \sum_{i\in A} M_i. \tag{8.3}$$

One then needs to proceed to show that $T < \infty$ a.s. Further than that, one can show that the sup in (8.3) is a.s. attained by a unique "path" $A^* \in \mathcal{C}$; to obtain the limiting path π^* of Theorem 8.2, one puts essentially $\pi^* = \bigcup_{i\in A^*} Y_i$.

The above outline illustrates an important difference between the heavy-tailed and light-tailed cases. In the heavy-tailed case, the main contribution to, say, $T(n,n)$ is made by a handful of very large weights $M_1^{(n)}$, $M_2^{(n)}, \ldots$. In the light-tailed case, on the other hand, we expect only to use weights which are located close to the diagonal. For fixed k, this becomes very unlikely to include $M_k^{(n)}$ when $n \to \infty$; now it is the "bulk" of the weight distribution and not its tail that is important.

The continuous model can be extended to cover the convergence of vectors of passage-times to different points in the plane. Namely, there exists a random field $\{\tilde{T}(x,y), x,y \in \mathbb{R}_+\}$ such that

$$\left\{a_{n^2}^{-1} T\left(\lfloor nx\rfloor, \lfloor ny\rfloor\right), x,y \in \mathbb{R}_+\right\} \to \left\{\tilde{T}(x,y), x,y \in \mathbb{R}_+\right\}$$

in the sense of convergence of finite-dimensional distributions. The field $\tilde{T}$ has a natural scaling property given by

$$\left\{\tilde{T}(\lambda x, \mu y)\right\} \stackrel{d}{=} \left\{(\lambda\mu)^{1/\alpha}\tilde{T}(x,y)\right\},$$

and one obtains that the random field defined by

$$\Theta(u,v) = \exp\left(-\frac{u+v}{\alpha}\right)\tilde{T}\left(e^u, e^v\right)$$

is *stationary* on $\mathbb{R}^2$. The convergence above can be rewritten as

$$\left\{\exp\left(-\frac{u+v}{\alpha}\right) T\left(\lfloor ne^u\rfloor, \lfloor ne^v\rfloor\right), u,v \in \mathbb{R}\right\} \to \left\{\Theta(u,v), u,v \in \mathbb{R}\right\}. \tag{8.4}$$

To remove the multiplicative factor on the LHS of (8.4), one can look at a line $u+v=\text{const}$; for example, the process $\Theta(y,-y)$. This process is stationary in $y\in\mathbb{R}$, and one can obtain the weak convergence

$$\left\{a_{n^2}^{-1}T\left(\lfloor ne^{y}\rfloor,\lfloor ne^{-y}\rfloor\right),y\in\mathbb{R}\right\}\to\left\{\Theta(y,-y),y\in\mathbb{R}\right\}.$$

This can be regarded as a heavy-tailed analogue of the "Airy process" described at (3.3) which arises in the light-tailed case.

The setting described above extends naturally to $d\geq 3$. An entirely analogous version of Theorem 8.1 holds for the convergence of the passage times. One also expects Theorem 8.2 to generalise, although for technical reasons (concerning the continuity of the optimal path for the continous model) only a weaker version has currently been shown.

9 Variations

Of course, many variations on the model presented here are natural. Replacing the max in (2.1) by a min, one obtains directed *first-passage* percolation; one can also obtain a mixture of first-passage and last-passage models by allowing the weights X to take both negative and positive values (see for example [39], and also [18] for similar developments for the related *greedy lattice animals* model).

Models in which the weights are no longer identically distributed have also been considered, mostly in the context of the exactly solvable models described in Section 3. For example, suppose that $R_n,n\in\mathbb{Z}$ are i.i.d. positive random variables, and that, conditional on $\{R_n,n\in\mathbb{Z}\}$, the variables $X(n,k),n,k\in\mathbb{Z}$ are independent with exponential distributions and with the mean of $X(n,k)$ equal to R_n. Such a model of "last-passage percolation in a random environment" corresponds to a totally asymmetric exclusion process in which different particles jump at different rates, or to a tandem queueing system in which different queues operate at different speeds, and also has applications in the modelling of traffic flow. See for example [23, 32, 52] for various aspects of these inhomogeneous models.

An intriguing open problem for a similar model concerns the "exclusion process with a slow bond". In terms of our percolation model: suppose $X(n,k)$ are all independent and each has an exponential distribution with mean 1 except for the "diagonal" weights $X(n,n)$ which are exponential with mean r. If $r=1$ then we know the "shape function" from (3.1), and for example $\lim_{n\to\infty}T(n,n)/n=g(1,1)=4$ a.s. One can show that if r is sufficiently large then this quantity is strictly larger than 4; however, it is not known whether or not this is true for all $r>1$. The problem can be described as follows: if the weights on the diagonal are typically heavier than the other weights, does the optimal path from the origin to the point (n,n) spend a positive fraction of its time on the diagonal? See for example Section III.3 of [33], and [51] for hydrodynamic properties of the model.

In Section 3 we discussed alterations to the local rules defining admissible paths. In [37], larger sets of paths are considered, which model tandem queueing systems in which the queues have finite capacity and hence various types of *blocking* occur. As in Theorem 7.1, one can obtain stability criteria for these systems; in general, however, it is no longer straightforward to calculate exactly the maximal achievable throughput. There are close links to particle systems such as the K-*exclusion process* and the model of *marching soldiers*, studied by Seppäläinen in [50]. The approach has been further developed, for example by Baccelli et al [6] to model the behaviour of the TCP protocol which is used to control the rate at which Internet sources emit data packets.

Acknowledgments

I am very grateful to the referee for several valuable suggestions, and in particular for pointing out many additional references for Sections 3 and 9.

References

[1] Aldous, D. and Diaconis, P., (1995) Hammersley's interacting particle process and longest increasing subsequences. *Probab. Theory Related Fields* **103**, 199–213.

[2] Aldous, D. and Diaconis, P., (1999) Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem. *Bull. Amer. Math. Soc. (N.S.)* **36**, 413–432.

[3] Anantharam, V., (1993) Uniqueness of stationary ergodic fixed point for a $\cdot/M/K$ node. *Ann. Appl. Probab.* **3**, 154–172.

[4] Baccelli, F., Borovkov, A. and Mairesse, J., (2000) Asymptotic results on infinite tandem queueing networks. *Probab. Theory Related Fields* **118**, 365–405.

[5] Baccelli, F. and Brémaud, P., (2003) *Elements of queueing theory*, vol. 26 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 2nd ed.

[6] Baccelli, F., Chaintreau, A., Liu, Z., Riabov, A. and Sahu, S., (2004) Scalability of reliable group communication using overlays. In *Proceedings of IEEE INFOCOM '04, Hong Kong.* INRIA research report RR-4895, `http://www.inria.fr/rrrt/rr-4895.html`.

[7] Baik, J., (2003) Limiting distribution of last passage percolation models. `math.PR/0310347`.

[8] Baik, J., Deift, P. and Johansson, K., (1999) On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12**, 1119–1178.

[9] Baik, J., Deift, P., McLaughlin, K. T.-R., Miller, P. and Zhou, X., (2001) Optimal tail estimates for directed last passage site percolation with geometric random variables. *Adv. Theor. Math. Phys.* **5**, 1207–1250.

[10] Baik, J. and Suidan, T., (2005) A GUE central limit theorem and universality of directed first and last passage site percolation. *Int. Math. Res. Not.* **2005:6**, 325–337.

[11] Baryshnikov, Y., (2001) GUEs and queues. *Probab. Theory Related Fields* **119**, 256–274.

[12] Bodineau, T. and Martin, J., (2005) A universality property for last-passage percolation paths close to the axis. *Electronic Communications in Probability* **10**, 105–112.

[13] Burke, P. J., (1956) The output of a queuing system. *Operations Res.* **4**, 699–704.

[14] Cohn, H., Elkies, N. and Propp, J., (1996) Local statistics for random domino tilings of the Aztec diamond. *Duke Math. J.* **85**, 117–166.

[15] Cox, J. T. and Durrett, R., (1981) Some limit theorems for percolation processes with necessary and sufficient conditions. *Ann. Probab.* **9**, 583–603.

[16] Cox, J. T., Gandolfi, A., Griffin, P. S. and Kesten, H., (1993) Greedy lattice animals I: upper bounds. *Ann. Appl. Probab.* **3**, 1151–1169.

[17] Deift, P., (2000) Integrable systems and combinatorial theory. *Notices Amer. Math. Soc.* **47**, 631–640.

[18] Dembo, A., Gandolfi, A. and Kesten, H., (2001) Greedy lattice animals: negative values and unconstrained maxima. *Ann. Probab.* **29**, 205–241.

[19] Durrett, R. and Liggett, T. M., (1981) The shape of the limit set in Richardson's growth model. *Ann. Probab.* **9**, 186–193.

[20] Gandolfi, A. and Kesten, H., (1994) Greedy lattice animals II: linear growth. *Ann. Appl. Probab.* **4**, 76–107.

[21] Glynn, P. W. and Whitt, W., (1991) Departures from many queues in series. *Ann. Appl. Probab.* **1**, 546–572.

[22] Gravner, J., Tracy, C. A. and Widom, H., (2001) Limit theorems for height fluctuations in a class of discrete space and time growth models. *J. Statist. Phys.* **102**, 1085–1132.

[23] Gravner, J., Tracy, C. A. and Widom, H., (2002) A growth model in a random environment. *Ann. Probab.* **30**, 1340–1368.

[24] Hambly, B. M. and Martin, J. B., (2005). Heavy tails in last-passage percolation. Preprint.

[25] Hambly, B. M., Martin, J. B. and O'Connell, N., (2002) Concentration results for a Brownian directed percolation problem. *Stochastic Process. Appl.* **102**, 207–220.

[26] Johansson, K., (2000) Shape fluctuations and random matrices. *Comm. Math. Phys.* **209**, 437–476.

[27] Johansson, K., (2000) Transversal fluctuations for increasing subsequences on the plane. *Probab. Theory Related Fields* **116**, 445–456.

[28] Johansson, K., (2001) Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)* **153**, 259–296.

[29] Johansson, K., (2003) Discrete polynuclear growth and determinantal processes. *Comm. Math. Phys.* **242**, 277–329.

[30] Kesten, H., (1986) Aspects of first passage percolation. In *École d'été de probabilités de Saint-Flour, XIV—1984*, vol. 1180 of *Lecture Notes in Math.*, pages 125–264. Springer, Berlin.

[31] Komlós, J., Major, P. and Tusnády, G., (1976) An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **34**, 33–58.

[32] Krug, J. and Ferrari, P. A., (1996) Phase transitions in driven diffusive systems with random rates. *J. Phys. A* pages L:465–471.

[33] Liggett, T. M., (1999) *Stochastic interacting systems: contact, voter and exclusion processes*, vol. 324 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin.

[34] Logan, B. F. and Shepp, L. A., (1977) A variational problem for random Young tableaux. *Advances in Math.* **26**, 206–222.

[35] Loynes, R. M., (1962) The stability of a queue with non-independent interarrival and service times. *Proc. Cambridge Philos. Soc.* **58**, 497–520.

[36] Mairesse, J. and Prabhakar, B., (2003) The existence of fixed points for the $\cdot/GI/1$ queue. *Ann. Probab.* **31**, 2216–2236.

[37] Martin, J. B., (2002) Large tandem queueing networks with blocking. *Queueing Syst. Theory Appl.* **41**, 45–72.

[38] Martin, J. B., (2002) Linear growth for greedy lattice animals. *Stochastic Process. Appl.* **98**, 43–66.

[39] Martin, J. B., (2004) Limiting shape for directed percolation models. *Ann. Probab.* **32**, 2908–2937.

[40] Muth, E. J., (1979) The reversibility property of production lines. *Management Sci.* **25**, 152–158.

[41] O'Connell, N., (2003) Random matrices, non-colliding particle systems and queues. In *Séminaire de Probabilités XXXVI*, no. 1801 in Lecture Notes in Mathematics, pages 165–182. Springer-Verlag.

[42] O'Connell, N. M., (1999) Directed percolation and tandem queues. DIAS Technical Report, DIAS-STP-99-12.

[43] Prabhakar, B., (2003) The attractiveness of fixed points for the $./GI/1$ queueing operator. *Ann. Probab.* **31**, 2237–2269.

[44] Prähofer, M. and Spohn, H., (2002) Scale invariance of the PNG droplet and the Airy process. *J. Statist. Phys.* **108**, 1071–1106.

[45] Rost, H., (1981) Nonequilibrium behaviour of a many particle process: density profile and local equilibria. *Z. Wahrsch. Verw. Gebiete* **58**, 41–53.

[46] Seppäläinen, T., (1997) Increasing sequences of independent points on the planar lattice. *Ann. Appl. Probab.* **7**, 886–898.

[47] Seppäläinen, T., (1997) A scaling limit for queues in series. *Ann. Appl. Probab.* **7**, 855–872.

[48] Seppäläinen, T., (1998) Coupling the totally asymmetric simple exclusion process with a moving interface. *Markov Process. Related Fields* **4**, 593–628. I Brazilian School in Probability (Rio de Janeiro, 1997).

[49] Seppäläinen, T., (1998) Exact limiting shape for a simplified model of first-passage percolation on the plane. *Ann. Probab.* **26**, 1232–1250.

[50] Seppäläinen, T., (1999) Existence of hydrodynamics for the totally asymmetric simple K-exclusion process. *Ann. Probab.* **27**, 361–415.

[51] Seppäläinen, T., (2001) Hydrodynamic profiles for the totally asymmetric exclusion process with a slow bond. *J. Statist. Phys.* **102**, 69–96.

[52] Seppäläinen, T. and Krug, J., (1999) Hydrodynamics and platoon formation for a totally asymmetric exclusion model with particlewise disorder. *J. Statist. Phys.* **95**, 525–567.

[53] Szczotka, W. and Kelly, F. P., (1990) Asymptotic stationarity of queues in series and the heavy traffic approximation. *Ann. Probab.* **18**, 1232–1248.

[54] Talagrand, M., (1995) Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Etudes Sci. Publ. Math.* **81**, 73–205.

[55] Vershik, A. M. and Kerov, S. V., (1977) Asymptotics of the Plancherel measure of the symmetric group and the limit form of Young tables. *Soviet Math. Dokl.* **18**, 527–531.

Department of Statistics,
University of Oxford,
1 South Parks Road,
Oxford OX1 3TG,
UK
martin@stats.ox.ac.uk