Journal of Computational and Applied Mathematics 142 (2002) 83–105

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Hydrodynamical methods for analyzing longest increasing subsequences

Piet Groeneboom[a,b]

[a] Delft University of Technology, Faculty of ITS, Mekelweg 4, 2628 CD Delft, Netherlands
[b] Free University, Faculty of Exact Sciences, de Boelelaan 1081a, 1081 HV Amsterdam, Netherlands

Received 23 October 2000; received in revised form 20 March 2001

Abstract

Let L_n be the length of the longest increasing subsequence of a random permutation of the numbers $1,\dots,n$, for the uniform distribution on the set of permutations. We discuss the "hydrodynamical approach" to the analysis of the limit behavior, which probably started with Hammersley (Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1 (1972) 345–394) and was subsequently further developed by several authors. We also give two proofs of an exact (non-asymptotic) result, announced in Rains (preprint, 2000). © 2002 Elsevier Science B.V. All rights reserved.

MSC: primary: 60C05; 60K35; secondary 60F05

Keywords: Longest increasing subsequence; Ulam's problem; Hammersley's process

1. Introduction

In recent years quite spectacular advances have been made with respect to the distribution theory of longest increasing subsequences L_n of a random permutation of the numbers $1,\dots,n$, for the uniform distribution on the set of permutations. Recent reviews of this work are given in [1,3].

However, rather than trying to give yet another review of this recent work, I will try to give a description of a different approach to the theory of the longest increasing subsequences, which in [2] is called "hydrodynamical".

As an example of a longest increasing subsequence we consider the permutation

$$\pi_n = (\pi_n(1),\dots,\pi_n(n)) = (7,2,8,1,3,4,10,6,9,5), \quad n = 10,$$

E-mail address: p.groeneboom@its.tudelft.nl (P. Groeneboom).

0377-0427/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0377-0427(01)00461-7

also used as an example in [3]. A longest increasing subsequence is

$$(1, 3, 4, 6, 9).$$

and another longest increasing subsequence is

$$(2, 3, 4, 6, 9).$$

For this example we get

$$L_n = \ell_n(\pi_n) = 5.$$

It was proved in [10] that, as $n \to \infty$,

$$L_n/\sqrt{n} \xrightarrow{\mathrm{p}} c,$$

where $\xrightarrow{\mathrm{p}}$ denotes convergence in probability, and

$$\lim_{n\to\infty} EL_n/\sqrt{n} = c,$$

for some positive constant c, where $\pi/2 \leqslant c \leqslant e$. Subsequently Kingman [12] showed that

$$1.59 < c < 2.49,$$

and later work by Logan and Shepp [14] and Vershik and Kerov [20] (expanded more fully in [21,22] showed that actually $c = 2$. The problem of proving that the limit exists and finding the value of c has been called "Ulam's problem", see, e.g., [6, p. 633].

In proving that $c = 2$, Aldous and Diaconis [2] replace the hard combinatorial work in [14,20], using Young tableaux by "hydrodynamical argument", building on certain ideas in [10], and it is this approach I will focus on in the present paper.

I will start by discussing Hammersley [10] in Section 2. Subsequently I will discuss the methods used in [2,18]. Slightly as a side-track, I will discuss an exact (non-asymptotic) result announced in [15], for which I have not seen a proof up till now, but for which I will provide a hydrodynamical proof below.

2. Hammersley's approach

The Berkeley symposium paper [10] is remarkable in several ways. The opening sentences are: "Graduate students sometimes ask, or fail to ask: "How does one do research in mathematical statistics?" It is a reasonable question because the fruits of research, lectures and published papers bear little witness to the ways and means of their germination and ripening". This beginning sets the tone for the rest of the paper, where Hammersley describes vividly the germination and ripening of his own research on the subject.

In Section 3, called "How well known is a well-known theorem?", he describes the difficulties encountered in finding a reference for a proof of the following theorem:

Theorem 2.1. *Any real sequence of at least* $mn + 1$ *terms contains either an ascending subsequence of* $m + 1$ *terms or a descending subsequence of* $n + 1$ *terms.*

This result, due to Erdös and Szekeres [9], is called the "pigeonhole principle" in Hammersley [10], a term also used by other authors. A nice description of this problem and other related problems is given in [4]. The relevance of the pigeonhole principle for the behavior of longest increasing subsequences is that one can immediately conclude from it that

$$EL_m \geqslant \tfrac{1}{2}(n+1) \quad \text{if } m = n^2 + 1, \tag{2.1}$$

As noted in [6], it is probable that Ulam, because of a long and enduring friendship with Erdös, got interested in determining the asymptotic value of EL_n and for this reason started (around 1961) a simulation study for n in the range $1 \leqslant n \leqslant 10$ ($n = 10$ being very large at the time, quoting [6]), from which he found

$$EL_n \sim 1.7\sqrt{n},$$

leading him to the conjecture that

$$\lim_{n\to\infty} \frac{EL_n}{\sqrt{n}} = c \tag{2.2}$$

exists. This is the first part of "Ulam's problem", the second part being the determination of c. Relation (2.1) then shows:

$$c \geqslant \tfrac{1}{2},$$

if we can deal with the first part of Ulam's problem (existence of the limit (2.2)).

The first part of Ulam's problem is in fact solved in [10]. It is Theorem 2.2 below (Theorem 4 on p. 352 of [10]):

Theorem 2.2. *Let $X = (X_1, X_2, \ldots)$ be an i.i.d. sequence of real-valued continuously distributed random variables, and let, respectively, L_n and L_n^* be the lengths of a longest increasing and a longest decreasing subsequence of $(X_1, \ldots, X_n)$. Then we have*

$$L_n/\sqrt{n} \xrightarrow{\text{p}} c \quad \text{and} \quad L_n^*/\sqrt{n} \xrightarrow{\text{p}} c$$

for some positive constant c, where $\xrightarrow{\text{p}}$ denotes convergence in probability. We also have convergence in the pth absolute mean of $L_n/\sqrt{n}$ and $L_n^*/\sqrt{n}$, for $0 < p < \infty$.

Note that, for a sample of n continuously distributed random variables, the vector of ranks $(R_1, \ldots, R_n)$ of the random variables $X_1, \ldots, X_n$ (for example ordered according to increasing magnitudes) has a uniform distribution over all permutations of $(1, \ldots, n)$. Because of the continuous distribution we may disregard the possibility of equal observations ("ties"), since this occurs with probability zero. So the random variable L_n, as defined in Theorem 2.2, indeed has exactly the same distribution as the length of a longest increasing subsequence of a random permutation of the numbers $1, \ldots, n$, for the uniform distribution on the set of permutations.

The key idea in [10] is to introduce a Poisson process of intensity 1 in the first quadrant of the plane and to consider longest North–East paths through points of the Poisson point process in squares $[r,s]^2$, where $0 \leqslant r < s < \infty$. A North–East path in the square $[r,s]^2$ is a sequence of points $(X_1, Y_1), \ldots, (X_k, Y_k) \in [r,s]^2$ of the Poisson process such that $X_1 < \cdots < X_k$ and $Y_1 < \cdots < Y_k$. We call k the *length of the path.*

Note that we can disregard the probability that $X_i = X_j$ or $Y_i = Y_j$ for $i \neq j$, since this happens with probability zero. A longest North–East path is a North–East path for which k is largest. Conditional on the number of points of the Poisson process in $[r,s]^2$, say n, the length of the longest North–East path has the same distribution as the longest increasing subsequence of a random permutation of $1,\dots,n$. This follows from the fact that, if $(U_1,V_1),\dots,(U_n,V_n)$ are the points of the Poisson process belonging to $[r,s]^2$, where we condition on the event that the number of points of the Poisson process in $[r,s]^2$ is equal to n, and if $V_{(1)} < \cdots < V_{(n)}$ are the order statistics of the second coordinates, then the corresponding first coordinates $U_{k_1},\dots,U_{k_n}$ behave as a sample from a Uniform distribution on $[r,s]$. A longest increasing North–East path will either consist of just one point (provided that the rectangle contains a point of the Poisson process, otherwise the length will be zero) or be a sequence of the form

$$((U_{k_i}, V_{(i)}),\dots,(U_{k_j}, V_{(j)}),$$

where $i < j$ and $U_{k_i} < \cdots < U_{k_j}$. So the length of a longest North–East path in $[r,s]^2$, conditionally on the number of points in $[r,s]^2$ being n, is distributed as the longest increasing subsequence of the sequence random variables $(U_1,\dots,U_n)$, and hence, by the remarks following Theorem 2.2 above, distributed as the length of a longest increasing subsequence of a permutation of the numbers $1,\dots,n$. If $n=0$, the length is zero and everything is trivial of course.

Following [10], we denote the length of a longest North–East path in the square $[r,s]^2$ by $W_{r,s}$, and for the collection of random variables $\{W_{r,s}\colon 0 \leqslant r < s < \infty\}$ we obviously have the so-called *superadditivity property*:

$$W_{r,t} \geqslant W_{r,s} + W_{s,t}, \quad 0 \leqslant r < s < t < \infty,$$

meaning that $-W_{r,t}$ has the *subadditivity property*:

$$-W_{r,t} \leqslant -W_{r,s} - W_{s,t}, \quad 0 \leqslant r < s < t < \infty.$$

Furthermore, we clearly have, for each $r > 0$, that $\{W_{nr,(n+1)r};\ n = 1,2,\dots\}$ is an i.i.d. sequence of random variables, since $W_{nr,(n+1)r}$ is a function of the Poisson point process restricted to the square $[nr,(n+1)r]^2$, and since the restrictions of the Poisson point process to the squares $[nr,(n+1)r]^2$ are i.i.d. For the same type of reason, the distribution of $W_{r,r+k}$ does not depend on $r \in (0,\infty)$. Finally $\max\{0,-W_{0,n}\} = 0$ for each n, and it will be shown below (see (2.11)) that

$$E(-W_{0,n}) \geqslant -K \cdot n \quad \text{for each } n, \tag{2.3}$$

for a finite constant $K > 0$. So we are in a position to apply Liggett's version of Kingman's subadditive ergodic theorem,

$$\frac{W_{0,r}}{r} \stackrel{\text{a.s.}}{\to} c = \sup_{r>0} \frac{EW_{0,r}}{r}, \quad r \to \infty,$$

and also

$$E\left(\frac{W_{0,r}}{r}\right) \to c = \sup_{r>0} \frac{EW_{0,r}}{r}, \quad r \to \infty.$$

Hammersley next defines $t(n)$ as the smallest real number such that $[0,t(n)]^2$ contains exactly n points of the Poisson process. Then it is clear from the properties of the Poisson point process in

$\mathbb{R}_+^2$ that

$$\frac{t(n)}{\sqrt{n}} \stackrel{\text{a.s.}}{\longrightarrow} 1, \quad n \to \infty,$$

and hence that

$$\frac{W_{0,t(n)}}{\sqrt{n}} \stackrel{\text{a.s.}}{\longrightarrow} c, \quad n \to \infty, \tag{2.4}$$

and

$$E\left(\frac{W_{0,t(n)}}{\sqrt{n}}\right) \to c, \quad n \to \infty, \tag{2.5}$$

where the constant c is the same in (2.4) and (2.5). But since $W_{0,t(n)}$ has the same distribution as L_n, we obtain from (2.4)

$$\frac{L_n}{\sqrt{n}} \stackrel{\text{p}}{\longrightarrow} c, \quad n \to \infty, \tag{2.6}$$

where $\stackrel{\text{p}}{\rightarrow}$ denotes convergence in probability, and from (2.5),

$$E\left(\frac{L_n}{\sqrt{n}}\right) \to c, \quad n \to \infty. \tag{2.7}$$

Remark. Note that we went from the *almost sure* relation (2.4) to the convergence *in probability* (2.6) for the original longest increasing subsequence, based on a random permutation of the numbers $1,\ldots,n$. It is possible, however, also to deduce the *almost sure* convergence of $L_n/\sqrt{n}$ from (2.4), using an extra tool, as was noticed by H. Kesten in his discussion of Kingman [12] (see p. 903 of [12]). I want to thank a referee for giving the latter reference, setting "the record straight" for this issue that was still bothering Hammersley in [10].

From (2.1) we now immediately obtain

$$c \geqslant \tfrac{1}{2} \tag{2.8}$$

(see also the remark below (2.2)), but we still have to prove (2.3). This problem is dealt with by Theorem 2.3 below (Theorem 6 on p. 355 of [10]):

Theorem 2.3. *Let, for $x \in \mathbb{R}$, $\lceil x \rceil$ be the smallest integer $\geqslant x$, and let, for a fixed $t \geqslant 0$ and each positive integer N, $n = \lceil e\sqrt{N} + t \rceil$. Moreover, let P_N denote the uniform distribution on the set of permutations of the sequence $(1,\ldots,N)$ with corresponding expectation E_N, let ℓ_N^* be the length of a longest monotone (decreasing or increasing) subsequence of a permutation of the numbers $1,\ldots,N$, and let $m_{n,N}$ the number of monotone subsequences of length n under the probability measure P_N. Then we have*

$$P_N\{\ell_N^* \geqslant n\} \leqslant E_N m_{n,N} = \frac{2}{n!}\binom{N}{n} \leqslant \frac{\mathrm{e}^{-2t}}{\pi\sqrt{N}}. \tag{2.9}$$

Proof. This is proved by an application of Stirling's formula (for details of this computation which is not difficult, see [10, pp. 355–356]). □

Elementary calculations show that Theorem 2.3 implies:

$$E_N\left(\frac{\ell_N^*}{\sqrt{N}}\right) \leqslant e + \mathcal{O}\left(N^{-1/2}\right), \quad N \to \infty.$$

Since $\ell_N \leqslant \ell_N^* = \max\{\ell_N, \ell'_N\}$, we obtain

$$E\ell_N \leqslant K\sqrt{N} \tag{2.10}$$

for a constant $K > 0$. Returning to the situation of longest increasing North–East paths in the plane, we obtain from (2.10)

$$EW_{0,n} \leqslant \sum_{i=1}^{\infty} (E\ell_i)\frac{e^{-n^2}n^{2i}}{i!} \leqslant K\sum_{i=1}^{\infty}\frac{e^{-n^2}n^{2i}\sqrt{i}}{i!} = \mathcal{O}(n), \quad n \to \infty, \tag{2.11}$$

which proves (2.3). Moreover, we have

$$\limsup_{n\to\infty}\frac{EW_{0,n}}{n} = \limsup_{n\to\infty}\sum_{i=1}^{\infty}(E\ell_i)\frac{e^{-n^2}n^{2i}}{i!} \leqslant e,$$

using Theorem 2.3.

Combining this result with (2.6) and (2.7), we obtain

$$c \leqslant e,$$

and, in particular, it is proved that $c \in (0,\infty)$, since also $c \geqslant \frac{1}{2}$, by (2.8).

So the first part of Ulam's problem is now solved, and we know that the constant c in the limit (2.2) is a number between 1/2 and e. As noted above, Hammersley improved the lower bound to $\pi/2$ and in [12] the bounds were tightened further to

$$1.59 < c < 2.49.$$

However, Hammersley was in fact quite convinced that $c = 2$ (see p. 372 of [10], where he says: "However, I should be very surprised if (12.12) is false", (12.12) being the statement $c = 2$ in his paper). Hammersley [10] contains three "attacks on c" of which we will discuss the third attack, since the ideas of the third attack sparked other research, like the development in [2]. Somewhat amusingly, his second attack yielded $c \approx 1.961$, but apparently he did not believe too much in that attack, in view of the remark ("I should be very surprised") quoted above.

"Hammersley's process" is introduced in Section 9 of [10]. In this section on Monte Carlo methods he introduces a kind of interacting particle process in discrete time. The particles of this process all live on the interval $[0,1]$ and are at "time n" a subset of a Uniform(0,1) sample $X_1,\ldots,X_n$. We now give a description of this process.

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of Uniform(0,1) random variables, and let, for each n, $\mathbf{X}_n$ be defined by

$$\mathbf{X}_n = (X_1,\ldots,X_n).$$

Moreover let, for $x \in (0,1)$, $\mathbf{X}_n^x$ be defined as the subsequence of $\mathbf{X}_n$ obtained by omitting all $X_i > x$. As before, $\ell(\mathbf{X}_n)$ is the length of the longest increasing subsequence of $\mathbf{X}_n$. In a similar way, $\ell(\mathbf{X}_n^x)$

is the length of the longest increasing subsequence of $\mathbf{X}_n^x$. Hammersley now notes that $\ell(\mathbf{X}_n^x)$ is an integer-valued step function of x, satisfying the recurrence relation

$$\ell(\mathbf{X}_{n+1}^x) = \begin{cases} \ell(\mathbf{X}_n^x), & 0 \leqslant x < X_{n+1}, \\ \max\{\ell(\mathbf{X}_n^x), 1 + \ell(\mathbf{X}_n^{X_{n+1}})\}, & X_{n+1} \leqslant x < 1, \end{cases} \tag{2.12}$$

and that, for a simulation study, one only needs to keep the values of the X_i's, where the function $x \mapsto \ell(\mathbf{X}_n^x)$, $x \in (0,1)$, makes a jump. The recurrence relation starts from

$$\ell(\mathbf{X}_1^x) = \begin{cases} 0, & 0 \leqslant x < X_1, \\ 1, & X_1 \leqslant x < 1. \end{cases} \tag{2.13}$$

Suppose that the jumps of the function $x \mapsto \ell(\mathbf{X}_n^x)$ occur at the points $Y_{i,n}$, $i = 1, \ldots, I(n)$, where

$$Y_{1,n} < \cdots < Y_{I(n),n}. \tag{2.14}$$

Then it is clear from (2.12) that the jumps of the function $x \mapsto \ell(\mathbf{X}_{n+1}^x)$ occur at the points

$$Y_{1,n+1} < \cdots < Y_{I(n+1),n+1},$$

which is obtained from points (2.14) by adding $Y_{I(n)+1,n+1} = X_{n+1}$ to points (2.14), if $x_{n+1} > Y_{I(n),n}$, and otherwise by replacing the smallest value $y_{i,n} > X_{n+1}$ by X_{n+1}. Note that $Y_{1,1} = X_1$. Defining

$$\mathbf{Y}_n = (Y_{1,n}, \ldots, Y_{I(n),n}), \tag{2.15}$$

we call the particle process $n \mapsto \mathbf{Y}_n$ *Hammersley's discrete time interacting particle process.*

In [11] the following simple example is given, clarifying the way in which this process evolves. Let $n = 6$ and

$$\mathbf{X}_n = (X_1, \ldots, X_n) = (0.23, 0.47, 0.14, 0.22, 0.96, 0.83).$$

Then the sequence $(\mathbf{Y}_1, \ldots, \mathbf{Y}_6)$ is represented by the sequence of states

$$0.23 \to (0.23, 0.47) \to (0.14, 0.47) \to (0.14, 0.22) \to (0.14, 0.22, 0.96) \to (0.14, 0.22, 0.83).$$

So either a new point is added (which happens, e.g., at the second step in the example above), or the "incoming point" replaces an existing point that is immediately to the right of this incoming point (this happens, e.g., at the third step in the example above).

We note in passing that the first stage of Viennot's geometric construction of the Robinson-Schensted correspondence, given in [23], is in fact Hammersley's discrete time interacting particle process on a lattice. A nice exposition of this construction is given in [17].

We can now discuss the "third attack on c" (Section 12 in [10]). The argument at the bottom of p. 372 and top of p. 373 is close to Eq. (9) of Aldous and Diaconis [2], (as pointed out to me by Aldous [1]), and is the *hydrodynamical argument* that inspired the approach in [2]. The argument (called "treacherous" by Hammersley) runs as follows.

Let $X_1, X_2, \ldots$ be i.i.d. Uniform(0,1) random variables, and let $Y_{1,n} < \cdots < Y_{I(n),n}$ be the points of Hammersley's discrete time interacting particle process at time n, associated with the sample $X_1, \ldots, X_n$, as described above. Moreover, let $\ell(\mathbf{X}_n)$ be the length of a longest increasing subsequence of $X_1, \ldots, X_n$. Then we have

$$E\ell(\mathbf{X}_{n+1}) - E\ell(\mathbf{X}_n) = E(1 - Y_{I(n),n}), \tag{2.16}$$

since $\ell(\mathbf{X}_{n+1}) - \ell(\mathbf{X}_n) = 1$, if $X_{n+1} \in (Y_{I(n),n}, 1)$, and otherwise $\ell(\mathbf{X}_{n+1}) - \ell(\mathbf{X}_n) = 0$. Since $E\ell(\mathbf{X}_n) \sim c\sqrt{n}$, as $n \to \infty$, we have (quoting [10, bottom of p. 372]) that "the left side of (2.16) is the result of differencing $c\sqrt{n} + o(\sqrt{n})$ with respect to n, and ought to be about $\frac{1}{2}c/\sqrt{n}$ if the error term is smooth". Continuing quoting [10] (top of p. 373) we get that "the right side of (2.16) is the displacement in x near $x = 1$, just sufficient to ensure unit increase in $\ell(\mathbf{X}_n^x)$, and should be the reciprocal of $(\partial/\partial x)\ell(\mathbf{X}_n^x)$ at $x = 1$, namely $2/(c\sqrt{n})$". The last statement is referring to relation (12.1) on p. 370 of [10], which is

$$\ell(\mathbf{X}_n^x) = c\sqrt{nx} + o(\sqrt{n}). \tag{2.17}$$

There is of course some difficulty in interpreting the equality sign in (2.17), does it mean "in probability", "almost surely" or is an asymptotic relation for expectations meant? Let us give (2.17) the latter interpretation. Then we would get from (2.16), following Hammersley's line of argument:

$$E\ell(\mathbf{X}_{n+1}) - E\ell(\mathbf{X}_n) \sim \frac{c}{2\sqrt{n}} \sim \frac{1}{(\partial/\partial x)E\ell(\mathbf{X}_n^x)|_{x=1}} \sim \frac{2}{c\sqrt{n}}. \tag{2.18}$$

This would yield $c = 2$. Following Hammersley's daring practice (in Section 12 of [10]) of differentiating w.r.t. a discrete parameter (in this case n), we can rewrite (2.18) in the form

$$\frac{\partial}{\partial n}E\ell(\mathbf{X}_n) \sim \frac{c}{2\sqrt{n}} \sim \frac{1}{(\partial/\partial x)E\ell(\mathbf{X}_n^x)|_{x=1}} \sim \frac{2}{c\sqrt{n}} \tag{2.19}$$

or

$$\frac{\partial}{\partial n}E\ell(\mathbf{X}_n) \cdot \left.\frac{\partial}{\partial x}E\ell(\mathbf{X}_n^x)\right|_{x=1} \to 1, \quad n \to \infty. \tag{2.20}$$

We shall return to Eq. (2.20) in the next sections, where it will be seen that it can be given different interpretations, corresponding to the different approaches in [2] (and perhaps more implicitly in [18]).

3. The Hammersley–Aldous–Diaconis interacting particle process

As I see it, one major step forward, made in [2], is to turn Hammersley's discrete time interacting particle process, as described in Section 2, into a continuous time process. One of the difficulties in interpreting relation (2.20) is the differentiation w.r.t. the discrete time parameter n, and this difficulty would be removed if we can differentiate with respect to a continuous time parameter (but other difficulties remain!).

The following gives an intuitive description of the extension of Hammersley's process on $\mathbb{R}_+$ to a continuous time process, developing according to the rules specified in [2]. Start with a Poisson point process of intensity 1 on $\mathbb{R}_+^2$. Now shift the positive x-axis vertically through (a realization of) this point process, and, each time a point is caught, shift to this point the previously caught point that is immediately to the right.

Alternatively, imagine, for each $x > 0$, an interval $[0, x]$, moving vertically through the Poisson point process. If this interval catches a point that is to the right of the points caught before, a new extra point is created in $[0, x]$, otherwise we have a shift to this point of the previously caught point that is immediately to the right and belongs to $[0, x]$ (note that this mechanism is exactly the same

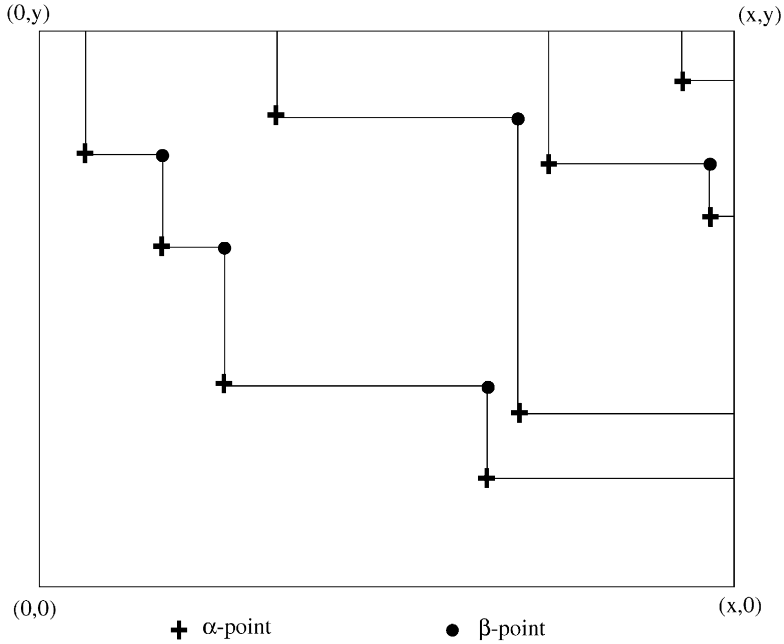

Fig. 1. Space–time curves of the Hammersley–Aldous–Diaconis process, contained in $[0,x] \times [0,y]$.

as the mechanism by which the points $Y_{i,n}$ of Hammersley's discrete time process are created in Section 2). The number of points, resulting from this "catch and shift" procedure at time y on the interval $[0,x]$, is denoted in [2] by

$$N^{+}(x,y), \quad x,y \geqslant 0.$$

So the process evolves in time according to "Rule 1" in [2], which is repeated here for ease of reference:

Rule 1. *At times of a Poisson* (*rate* x) *process in time*, *a point* U *is chosen uniformly on* $[0,x]$, *independent of the past*, *and the particle nearest to the right of* U *is moved to* U, *with a new particle created at* U *if no such particle exists in* $[0,x]$.

We shall call this process the *Hammersley–Aldous–Diaconis interacting particle process*. For a picture of the space–time curves of this process, see Fig. 1; an "α-point" is an added point and a "β-point" is a deleted point for this continuous time process (time is running along the vertical axis).

In Hammersley's "third attack on c", one of the crucial assumptions he was not able to prove was the assumption that the distribution of the points $Y_{i,n}$ was locally homogeneous, so, actually, is locally behaving as a homogeneous Poisson process; he calls this "assumption α" (p. 371 of [10]). This key assumption is in fact what is proved in [2] for the Hammersley–Aldous–Diaconis interacting particle process in Theorem 5 on p. 204 (which is the central result of their paper):

Theorem 3.1. (a) $c = 2$.

(b) *For each fixed* $a > 0$, *the random particle configuration with counting process*

$$\{N^+(at + y, t) - N^+(at, t):\ y \in (-\infty, \infty)\},$$

converges in distribution, as $t \to \infty$, *to a homogeneous Poisson process with intensity* a^{-1}.

After stating this theorem, they give the following heuristic argument. *Suppose* the spatial process around position x at time t approximates a Poisson process of some rate $\lambda(x,t)$. Then clearly

$$\frac{\partial}{\partial t} EN^+(x,t) = ED_{x,t},$$

where $D_{x,t}$ is the distance from x to the nearest particle to the left of x. For a Poisson process, $ED_{x,t}$ would be 1/(spatial rate), so

$$ED_{x,t} \approx \frac{1}{\lambda(x,t)} \approx \frac{1}{(\partial/\partial x)EN^+(x,t)}.$$

In other words, $w(x,t) = EN^+(x,t)$ satisfies approximately the partial differential equation

$$\frac{\partial w}{\partial x}\frac{\partial w}{\partial t} = 1, \quad w(x,0) = w(0,t) = 0, \ \ x,t \geqslant 0, \tag{3.1}$$

whose solution is $w(x,t) = w(t,x) = 2\sqrt{tx}$. Note that (3.1) is very close to (2.20) above, but that we got rid of the differentiation w.r.t. a discrete time parameter!

Appealing as the above argument may seem, a closer examination of [2] learns that their proof not really proceeds along the lines of this heuristic argument. In fact, they prove separately that $c \leqslant 2$ and that $c \geqslant 2$, by arguments that do not seem to use the differential equation (3.1). The proofs of $c \leqslant 2$ and $\geqslant 2$ are based on coupling arguments, where the Hammersley–Aldous–Diaconis process is coupled to a stationary version of this process, starting with a non-empty configuration, and living on $\mathbb{R}$ instead of $\mathbb{R}_+$. This process evolves according to the following rule (p. 205 of [2]).

Rule 2. *The restriction of the process to the interval* $[x_1, x_2]$ *satisfies*:

- (i) *There is some arbitrary set of times at which the leftmost point* (*if any*) *in the interval is removed.*
- (ii) *At times of a rate* $x_2 - x_1$ *Poisson process in time, a point* U *is chosen uniformly on* $[x_1, x_2]$, *independent of the past, and the particle nearest to the right of* U *is moved to* U, *with a new particle created at* U *if no such particle exists in* $[x_1, x_2]$.

To avoid a possible misunderstanding, rule 2 above is not the same a "rule 2" in [2]. The existence of such a process is ensured by the following lemma (Lemma 6 in [2]).

Lemma 3.1. *Suppose an initial configuration* $N(\cdot, 0)$ *of particles in* $\mathbb{R}$ *satisfies*

$$\liminf_{x \to -\infty} \frac{N(x,0)}{x} > 0, \quad a.s. \tag{3.2}$$

Let $\mathscr{P}$ *be a Poisson point process of intensity* 1 *in the plane, and let* $L^{\nearrow}((z,0),(x,t))$ *be the maximal number of Poisson points on a North–East path from* $(z,0)$ *to* (x,t), *in the* *sense that it is piecewise linear with vertices* $(z,0)$, (x_i,y_i), $1\leqslant i\leqslant k$ *and* (x,t) *such that* $z\leqslant x_1<\cdots<x_k\leqslant x$, $0\leqslant y_1<\cdots<y_k\leqslant t$, *for points* (x_i,y_i) *belonging a realization of* $\mathscr{P}$, *and such that* k *is largest. Then the process, defined by*

$$N(x,t)=\sup_{-\infty<z\leqslant x}\{N(z,0)+L^{\nearrow}((z,0),(x,t))\},\quad x\in\mathbb{R},\ t\geqslant 0, \tag{3.3}$$

evolves according to Rule 2.

A process of this type is called "Hammersley's process on $\mathbb{R}$" in [2], but we will call this the *Hammersley–Aldous–Diaconis process on* $\mathbb{R}$. As an example of such a process, consider an initial configuration, corresponding to a Poisson point process $\mathscr{P}_\lambda$ of intensity $\lambda>0$. The initial configuration will be invariant for the process; that is: $N(\cdot,t)$ will have the same distribution as $\mathscr{P}_\lambda$, for each $t>0$. The following key lemma (Lemma 7 in [2]) characterizes the invariant distributions in the process on $\mathbb{R}$.

Lemma 3.2. *A finite intensity distribution is invariant and translation-invariant for the Hammersley–Aldous–Diaconis process on* $\mathbb{R}$ *if and only if it is a mixture of Poisson point processes* $\mathscr{P}_\lambda$.

Other key properties are given in the next lemma (part (i) is Lemma 8, part (ii) and (iii) are Lemma 3, and part (iv) is Lemma 11 in [2]).

Lemma 3.3. (i) *Let* N *be a Hammersley–Aldous–Diaconis process on* $\mathbb{R}$, *with the invariant distribution* $\mathscr{P}_\lambda$, *run for time* $t\in\mathbb{R}$. *Let, for* $x\geqslant 0$ *and* $t\in\mathbb{R}$, $\tilde N(x,t)$ *be defined by*

$\tilde N(x,t)=$ *number of particles of* $\{N(s,\cdot)\colon s\in(t,\infty)\}$ *which exit during the time interval* $[0,x]$.

Then $x\to\tilde N(x,\cdot)$, $x\geqslant 0$, *is also a Hammersley–Aldous–Diaconis process on* $\mathbb{R}$, *with invariant distribution* $\mathscr{P}_{1/\lambda}$.

(ii) (*Space–time interchange property*) *Let*

$$\tilde L^{\nearrow}((x_1,y_1),(x_2,y_2))=L^{\nearrow}((x_1,y_1),(y_2,x_2)),\quad \max\{x_1,y_1\}<\min\{x_2,y_2\}.$$

Then

$$\tilde L^{\nearrow}((x_1,y_1),(x_2,y_2))\stackrel{\mathrm{d}}{=}L^{\nearrow}((x_1,y_1),(x_2,y_2)).$$

(iii) (*Scaling property*) *For all* $x,y,k>0$,

$$L^{\nearrow}((0,0),(x,y))\stackrel{\mathrm{d}}{=}L^{\nearrow}((0,0),(kx,y/k)).$$

(iv) *For fixed* $x,y\in\mathbb{R}$ *we have.*

$$\lim_{t\to\infty}\{EN^+(t+x,t+y)-EN^+(t,t)\}=\tfrac{1}{2}c(x+y).$$

Parts (ii) and (iii) are immediately clear from the fact that the distribution of $L^{\nearrow}((x_1,y_1),(x_2,y_2))$ only depends on the area of the rectangle $[x_1,y_1]\times[x_2,y_2]$, since the expected number of points of the Poisson point process only depends on the area of the rectangle, and since the shape of the rectangle has no influence on the distribution of $L^{\nearrow}((x_1,y_1),(x_2,y_2))$. The proofs of (i) and (iv) rely on somewhat involved coupling arguments, and we refer for that to [2].

The argument for $c \geqslant 2$ now runs as follows. The processes

$$t \mapsto N^+(t+\cdot,t) \quad \text{and} \quad t \mapsto N^+(t,t+\cdot)$$

have subsequential limits, which have to be translation-invariant and invariant for the Hammersley–Aldous–Diaconis process. By part (ii) of Lemma 3.3 (the space–time interchange property) these limit processes must have the same distribution. By Lemma 3.2 and the invariance properties the limit process must have, these processes must be mixtures of Poisson processes. Suppose the subsequential limit of the process $t \mapsto N(t+\cdot,t)$ is such a mixture of Poisson processes with random intensity Λ. Then, according to part (i) of Lemma 3.3, the subsequential limit of the process $t \mapsto N(t,t+\cdot)$ must be a mixture of Poisson processes with random intensity Λ^{-1}. We have

$$E\Lambda\Lambda^{-1} = 1,$$

and, by Jensen's inequality, $E\Lambda^{-1} \geqslant 1/E\Lambda$, implying

$$E\Lambda E\Lambda^{-1} \geqslant 1.$$

But, since the limit processes must have the same distribution, we also have $E\Lambda = E\Lambda^{-1}$, and hence

$$(E\Lambda)^2 = E\Lambda E\Lambda^{-1} \geqslant 1,$$

implying $E\Lambda \geqslant 1$. Using this fact, in combination with Fatou's lemma and part (iv) of Lemma 3.3, we get

$$2 \leqslant 2E\lambda \leqslant \liminf_{n\to\infty} E\{N^+(t_n+2,t_n) - N^+(t_n,t_n)\} = c,$$

if (t_n) is the sequence for which we have the subsequential limit. Hence $c \geqslant 2$.

Proving $c \leqslant 2$ is easier; it is proved by a rather straightforward coupling argument in [2] and it also follows from the result in Section 4 below (see the last paragraph of Section 4). So the conclusion is that $c = 2$. Moreover, since $\Lambda \stackrel{d}{=} \Lambda^{-1}$ and since the covariance of Λ and Λ^{-1} is equal to zero, we can only have: $\Lambda = 1$ almost surely. This proves part (b) of Theorem 3.1 for the case $a = 1$, and the case $a \neq 1$ then follows from part (iii) of Lemma 3.3.

4. A nonasymptotic result for longest North–East paths in the unit square

The purpose of the present section is to give a proof of the following result.

Theorem 4.1. *Let $\mathscr{P}_1$ be a Poisson process of intensity λ_1 on the lower edge of the unit square $[0,1]^2$, $\mathscr{P}_2$ a Poisson process of intensity λ_2 on the left edge of the unit square, and $\mathscr{P}$ a Poisson process of intensity $\lambda_1\lambda_2$ in the interior of the unit square. Then the expected length of a longest North–East path from $(0,0)$ to $(1,1)$, where horizontal or vertical parts are allowed at the start of the path, is equal to $\lambda_1+\lambda_2$.*

Here, as before, the length of the path is defined as the number of points of the point process, "picked up" along the path. However, in the present situation it is allowed to pick up points from the boundary, and, moreover, there are Poisson point processes on both the left and the lower boundary. The exact result about the expectation of the length of the longest North–East path (for the case $\lambda_1 = \lambda_2$) is announced in [15], which refers to a manuscript in preparation of Prähofer and Spohn (which I have not seen).

The idea of the proof is to show that if we start with the (possibly empty) configuration of points on the lower edge, and let the point process develop according to the rules of the Hammersley–Aldous–Diaconis process, where we let the leftmost point "escape" at time t, if the left edge of the unit square contains a point of the Poisson process of intensity λ_2 at $(0,t)$, the process will be stationary. This means that the expected number of points of the process will be equal to λ_1 at each time t, so in particular at time $t = 1$. Since the expected number of points on the left edge is λ_2, the expected number of space–time curves of the Hammersley–Aldous–Diaconis process (with "left escapes") will be $\lambda_1 + \lambda_2$. This implies the result, since the length of a longest North–East path is equal to the number of space–time curves (note that such a longest North–East path "picks up" one point from each space–time curve).

A proof of Theorem 4.1 in the spirit of the methods used in [2,18] would run as follows. Start with a Poisson process ξ_0 of intensity λ_1 on $\mathbb{R}$ and a Poisson process of intensity $\lambda_1\lambda_2$ in the upper half plane. Now let the Poisson process ξ_0 develop according to the rules of the Hammersley–Aldous–Diaconis process on the whole real line, and let ξ_t denote the process at time t. Then the process $\{\xi_t\colon t \geqslant 0\}$ will be invariant in the sense that it will be distributed as a Poisson process of intensity λ_1 at any positive time. The restriction of ξ_t to the interval $[0,1]$ will be a Poisson process of intensity λ_1 on this interval. Since (by the "bus stop paradox") the distribution of the distance of the rightmost point in the interval $(-\infty,0)$ to 0 will have an exponential distribution with (scale) parameter $1/\lambda_1$, the leftmost points in the interval $[0,1]$ will escape at rate λ_2, because an escape will happen if a new point is "caught" in the interval between the rightmost point of the process in $(-\infty,0)$ and 0, and because the intensity of the Poisson process in the upper half plane is $\lambda_1\lambda_2$. So the point process on $[0,1]$, induced by the stationary process $\{\xi_t\colon t \geqslant 0\}$ on $\mathbb{R}$, develops exactly along the rules of the Hammersley–Aldous–Diaconis process "with left escapes", described above, and the desired stationarity property follows.

The proof of Theorem 4.1 below uses an "infinitesimal generator" approach. It is meant to draw attention to yet another method that could be used in this context and this is the justification of presenting it here, in spite of the fact that it is much longer than the proof we just gave (but most of the work is setting up the right notation and introducing the right spaces). Also, conversely, the proof below can be used to prove the property that a Poisson process on $\mathbb{R}$ is invariant for the Hammersley–Aldous–Diaconis process; this property is a key to the proofs in [2].

Let ξ denote a point process on $[0,1]$. That is, ξ is a random (Radon) measure on $[0,1]$, with realizations of the form

$$\xi(f) \stackrel{\text{def}}{=} \int_0^1 f(x)\,\mathrm{d}\xi(x) = \sum_{i=1}^{N} f(\tau_i), \tag{4.1}$$

where $\tau_1,\ldots,\tau_N$ are the points of the point process ξ and f is a bounded measurable function $f:(0,1) \to \mathbb{R}_+$. If $N = 0$, we define the right side of (4.1) to be zero.

We can consider the random measure ξ as a random sum of Dirac measures:

$$\xi = \delta_{\tau_1} + \cdots + \delta_{\tau_N}, \tag{4.2}$$

and hence

$$\xi(B) = \sum_{i=1}^{N} \delta_{\tau_i}(B) = \sum_{i=1}^{N} 1_B(\tau_i),$$

for Borel sets $B \subset (0,1)$. So $\xi(B)$ is just the number of points of the point process ξ, contained in B, where the sum is defined to be zero if $N = 0$. The realizations of a point process, applied on Borel subsets of $[0,1]$, take values in $\mathbb{Z}_+$ and belong to a strict subset of the Radon measures on $[0,1]$. We will denote this subset, corresponding to the point processes, by $\mathscr{N}$, and endow it with the vague topology of measures on $[0,1]$, see, e.g., [11, p. 32]. For this topology, $\mathscr{N}$ is a (separable) *Polish space* and a closed subset of the set of Radon measures on $[0,1]$, see Propositions 15.7.7 and 15.7.4, pp. 169–170 of [11]. Note that, by the compactness of the interval $[0,1]$, the *vague topology* coincides with the *weak topology*, since all continuous functions $f : [0,1] \to \mathbb{R}$ have compact support, contained in the compact interval $[0,1]$. For this reason we will denote the topology on $\mathscr{N}$ by the *weak* topology instead of the *vague* topology in the sequel. Note that the space $\mathscr{N}$ is in fact *locally compact* for the weak topology.

In our case we have point processes ξ_t, for each time $t \geqslant 0$, of the form

$$\xi_t = \delta_{\tau_1} + \cdots + \delta_{\tau_{N_t}},$$

where N_t denotes the number of points at time t, and where $0 \leqslant \tau_1 \leqslant \cdots \leqslant \tau_{N_t} < 1$; ξ_t is defined to be the zero measure on $[0,1]$, if $N_t = 0$. Note that, if we start with a Poisson process of intensity $\lambda_1 > 0$, the initial configuration ξ_0 will with probability one be either the zero measure or be of the form

$$\sum_{i=1}^{n} \delta_{\tau_i}, \quad 0 < \tau_1 < \cdots < \tau_n < 1 \tag{4.3}$$

for some $n > 0$, but since we want to consider the space of bounded continuous functions on $\mathscr{N}$, it is advantageous to allow configurations where some τ_i's will be equal. We also allow the τ_i's to take values at 0 or 1. If we have a "stack" of τ_i's at the same location in $(0,1]$, we move one point ("the point on top") from the stack to the left, if a point immediately to the left of the location of the stack appears, leaving the other points at the original location. Likewise, if a stack of τ_i's is located at 0, we remove the point on top of the stack at time t if the Poisson point process on the left lower boundary has a point at $(0,t)$.

Now let $\mathscr{F}_c$ be the Banach space of continuous bounded functions $\phi : \mathscr{N} \to \mathbb{R}$ with the supremum norm. For $\phi \in \mathscr{N}$ and $t > 0$ we define the function $P_t\phi : \mathscr{N} \to \mathbb{R}$ by

$$[P_t\phi](\xi) = E\{\phi(\xi_t) \mid \xi_0 = \xi\}.$$

We want to show that the operator P_t is a mapping from $\mathscr{F}_c$ into itself.

Boundedness of $P_t\phi$ is clear if $\phi : \mathscr{N} \to \mathbb{R}$ is bounded and continuous, so we only must prove the continuity of $P_t\phi$, if ϕ is a bounded continuous function $\phi : \mathscr{N} \to \mathbb{R}$. If ξ is the zero measure and (ξ_n) is a sequence of measures in $\mathscr{N}$, converging weakly to ξ, we must have

$$\lim_{n\to\infty} \xi_n([0,1]) = 0,$$

and hence $\xi_n([0,1]) = 0$, for all large n. This implies that

$$E\{\phi(\xi_t) \mid \xi_0 = \xi\} = E\{\phi(\xi_t) \mid \xi_0 = \xi_n\}$$

for all large n. If $\xi = \sum_{i=1}^{m} \delta_{\tau_i}$, with $m > 0$, and (ξ_n) is a sequence of measures in $\mathcal{N}$, converging weakly to ξ, we must have $\xi_n([0,1]) = m = \xi([0,1])$ for all large n, so ξ_n is of the form

$$\xi_n = \sum_{i=1}^{m} \delta_{\tau_{n,i}}$$

for all large n. Moreover, the ordered $\tau_{n,i}$'s have to converge to the ordered τ_i's in the Euclidean topology. Since the x-coordinates of a realization of the Poisson process of intensity $\lambda_1\lambda_2$ in $(0,1) \times (0,t]$ will with probability one be different from the τ_i's, sample paths of the processes $\{\xi_t\colon t \geqslant 0\}$, either starting from ξ or from ξ_n, will develop in the same way, if n is sufficiently large, for such a realization of the Poisson process in $(0,1) \times (0,t]$. Hence

$$\lim_{n\to\infty} E\{\phi(\xi_t) \mid \xi_0 = \xi_n\} = E\{\phi(\xi_t) \mid \xi_0 = \xi\}.$$

So we have

$$\lim_{n\to\infty} [P_t\phi](\xi_n) = [P_t\phi](\xi)$$

if the sequence (ξ_n) converges weakly to ξ, implying the continuity of $P_t\phi$.

Since $\{\xi_t\colon t \geqslant 0\}$ is a Markov process with respect to the natural filtration $\{\mathcal{F}_t\colon t \geqslant 0\}$, generated by this process, we have the semi-group property

$$P_{s+t}\phi = [P_t \circ P_s]\phi$$

for bounded continuous functions $\phi : \mathcal{N} \to \mathbb{R}$. Moreover, we can define the *generator* $\mathcal{G}$ of the process (ξ_t), working on the bounded continuous functions $\phi : \mathcal{N} \to \mathbb{R}$ by

$$[\mathcal{G}\phi](\xi) = \lambda_1\lambda_2 \int_0^1 \{\phi(\delta_x) - 1\}\,\mathrm{d}x, \tag{4.4}$$

if ξ is the zero measure on $(0,1)$, and by

$$[\mathcal{G}\phi](\xi) = \lambda_1\lambda_2 \sum_{i=1}^{n+1} \int_{\tau_{i-1}}^{\tau_i} \{\phi(\xi^x) - \phi(\xi)\}\,\mathrm{d}x + \lambda_2 \left\{ \phi\left(\sum_{i=2}^{n} \delta_{\tau_i} \right) - \phi(\xi) \right\} \tag{4.5}$$

if $\xi = \sum_{i=1}^{n} \delta_{\tau_i}$, where $\tau_0 = 0$, $\tau_{n+1} = 1$, and where ξ^x is defined by

$$\xi^x = \begin{cases} \delta_x + \sum_{i=2}^{n} \delta_{\tau_i} & \text{if } 0 < x < \tau_1, \\ \sum_{i=1}^{j-1} \delta_{\tau_i} + \delta_x + \sum_{i=j+1}^{n} \delta_{\tau_i} & \text{if } \tau_{j-1} < x < \tau_j,\ 1 < j < n, \\ \sum_{i=1}^{n} \delta_{\tau_i} + \delta_x & \text{if } \tau_n < x < 1. \end{cases} \tag{4.6}$$

The first term on the right hand side of (4.5) corresponds to the insertion of a new point in one of the intervals (τ_{i-1}, τ_i) and the shift of τ_i to this new point if the new point is not in the rightmost interval, and the second term on the right of (4.5) corresponds to an "escape on the left". Note that $\mathcal{G}\phi(\xi)$ is computed by evaluating

$$\lim_{h\downarrow 0} \left[\frac{[P_h\phi](\xi) - \phi(\xi)}{h} \right].$$

The definition of $\mathscr{G}$ can be continuously extended to cover the configurations

$$\sum_{i=1}^{n} \delta_{\tau_i}, \quad 0 \leqslant \tau_1 \leqslant \cdots \leqslant \tau_n \leqslant 1, \tag{4.7}$$

working with the extended definition of P_t, described above.

So we have a semigroup of operators P_t, working on the Banach space of bounded continuous functions $\phi : \mathscr{N} \to \mathbb{R}$, with generator $\mathscr{G}$. It now follows from Theorem 13.35 in Rudin (1991) that we have the following lemma.

Lemma 4.1. *Let $\mathscr{N}$ be endowed with the weak topology and let $\phi : \mathscr{N} \to \mathbb{R}$ be a bounded continuous function. Then we have, for each $t > 0$,*

$$\frac{\mathrm{d}}{\mathrm{d}t}[P_t\phi] = [\mathscr{G}P_t\phi] = [P_t\mathscr{G}\phi].$$

Proof. It is clear that conditions (a)–(c) of Definition 13.34 in Rudin [16] are satisfied, and the statement then immediately follows. □

We will also need the following lemma (this is the real "heart" of the proof).

Lemma 4.2. *Let for a continuous function $f : [0,1] \to \mathbb{R}_+$, the function $L_f : \mathscr{N} \to \mathbb{R}_+$ be defined by*

$$L_f(\xi) = \exp\{-\xi(f)\}. \tag{4.8}$$

Then:

$$E[\mathscr{G}L_f](\xi_0) = 0, \; \textit{for all continuous } f : [0,1] \to \mathbb{R}_+. \tag{4.9}$$

Proof. We first consider the value of $\mathscr{G}L_f(\xi_0)$ for the case where ξ_0 is the zero measure, i.e., the interval $[0,1]$ contains no points of the point process ξ_0. By (4.4) we then have

$$\mathscr{G}L_f(\xi_0) = \lambda_1\lambda_2 \int_0^1 \{\mathrm{e}^{-f(x)} - 1\}\,\mathrm{d}x. \tag{4.10}$$

If $\xi_0 = \delta_x$, for some $x \in (0,1)$, we get

$$\mathscr{G}L_f(\xi_0) = \lambda_2\{1 - \mathrm{e}^{-f(x)}\} + \lambda_1\lambda_2 \int_0^x \{\mathrm{e}^{-f(u)} - \mathrm{e}^{-f(x)}\}\,\mathrm{d}u + \lambda_1\lambda_2 \int_x^1 \{\mathrm{e}^{-f(x)-f(u)} - \mathrm{e}^{-f(x)}\}\,\mathrm{d}u.$$

Hence

$$\begin{aligned}
&E\mathscr{G}L_f(\xi_0)1_{\{\xi_0([0,1]) \leqslant 1\}} \\
&\quad = \lambda_1\lambda_2\mathrm{e}^{-\lambda_1} \int_0^1 \{\mathrm{e}^{-f(x)} - 1\}\,\mathrm{d}x + \lambda_1\lambda_2\mathrm{e}^{-\lambda_1} \int_0^1 \{1 - \mathrm{e}^{-f(x)}\}\,\mathrm{d}x \\
&\qquad - \lambda_1^2\lambda_2\mathrm{e}^{-\lambda_1} \int_0^1 u\mathrm{e}^{-f(u)}\,\mathrm{d}u + \lambda_1^2\lambda_2\mathrm{e}^{-\lambda_1} \int\int_{0<x<u<1} \mathrm{e}^{-f(x)-f(u)}\,\mathrm{d}x\,\mathrm{d}u \\
&\quad = -\lambda_1^2\lambda_2\mathrm{e}^{-\lambda_1} \int_0^1 u\mathrm{e}^{-f(u)}\,\mathrm{d}u + \lambda_1^2\lambda_2\mathrm{e}^{-\lambda_1} \int\int_{0<x<u<1} \mathrm{e}^{-f(x)-f(u)}\,\mathrm{d}x\,\mathrm{d}u.
\end{aligned}$$

Now generally suppose that, for $n > 1$,

$$E\mathscr{G}L_f(\xi_0)1_{\{\xi_0([0,1])\leqslant n-1\}}$$

$$= -\lambda_1^n\lambda_2 e^{-\lambda_1}\int_{0<x_1<\cdots<x_{n-1}<1} x_1 \exp\left\{-\sum_{i=1}^{n-1} f(x_i)\right\} dx_1 \ldots dx_{n-1} \tag{4.11}$$

$$+\lambda_1^n\lambda_2 e^{-\lambda_1}\int_{0<x_1<\cdots<x_n<1} \exp\left\{-\sum_{i=1}^{n} f(x_i)\right\} dx_1 \cdots dx_n. \tag{4.12}$$

Then, by a completely similar computation, it follows that

$$E\mathscr{G}L_f(\xi_0)1_{\{\xi_0([0,1])\leqslant n\}} = E\mathscr{G}L_f(\xi_0)1_{\{\xi_0([0,1])\leqslant n-1\}} + E\mathscr{G}L_f(\xi_0)1_{\{\xi_0([0,1])=n\}}$$

$$= -\lambda_1^{n+1}\lambda_2 e^{-\lambda_1}\int_{0<x_1<\cdots<x_n<1} x_1 \exp\left\{-\sum_{i=1}^{n} f(x_i)\right\} dx_1 \ldots dx_n$$

$$+\lambda_1^{n+1}\lambda_2 e^{-\lambda_1}\int_{0<x_1<\cdots<x_{n+1}<1} \exp\left\{-\sum_{i=1}^{n+1} f(x_i)\right\} dx_1 \ldots dx_{n+1}.$$

So we get

$$E\mathscr{G}L_f(\xi_0) = \lim_{n\to\infty} E\mathscr{G}L_f(\xi_0)1_{\{\xi_0([0,1])\leqslant n\}} = 0,$$

since

$$\int_{0<x_1<\cdots<x_n<1} x_1 \exp\left\{-\sum_{i=1}^{n} f(x_i)\right\} dx_1 \ldots dx_n < \frac{1}{n!},$$

and similarly

$$\int_{0<x_1<\cdots<x_{n+1}<1} \exp\left\{-\sum_{i=1}^{n+1} f(x_i)\right\} dx_1 \ldots dx_{n+1} \leqslant \frac{1}{(n+1)!}. \qquad \square$$

We now have the following corollary.

Corollary 4.1. *Let $\phi : \mathscr{N} \to \mathbb{R}$ be a continuous function with compact support in $\mathscr{N}$. Then*:

$$E[\mathscr{G}\phi](\xi_0) = 0. \tag{4.13}$$

Proof. Let C be the compact support of ϕ in $\mathscr{N}$. The functions L_f, where f is a continuous function $f : [0,1] \to \mathbb{R}_+$, are closed under multiplication and hence linear combinations of these functions, restricted to C, form an *algebra*. Since the constant functions also belong to this algebra and the functions L_f separate points of C, the Stone–Weierstrass theorem implies that ϕ can be uniformly approximated by functions from this algebra, see, e.g., [8, (7.3.1), p. 137]. The result now follows from Lemma 4.2, since $\mathscr{G}$ is clearly a bounded continuous operator on the Banach space of continuous functions $\psi : C \to \mathbb{R}$. $\square$

Now let $\phi:\mathcal{N}\to\mathbb{R}$ be a continuous function with compact support in $\mathcal{N}$. Then $P_t\circ\phi$ is also a continuous function with compact support in $\mathcal{N}$, for each $t>0$. By Corollary 4.1 we have

$$E[\mathcal{G}P_t\phi](\xi_0)=0.$$

Hence, by Lemma 4.1,

$$E[P_t\phi](\xi_0)-E\phi(\xi_0)=\int_0^t E[\mathcal{G}P_s\phi](\xi_0)\,\mathrm{d}s=0,\quad t>0,$$

implying

$$E\phi(\xi_t)=E[P_t\phi](\xi_0)=E\phi(\xi_0),$$

for each continuous function $\phi:\mathcal{N}\to\mathbb{R}$ with compact support in $\mathcal{N}$. But since $\mathcal{N}$ is a Polish space, every probability measure on $\mathcal{N}$ is "tight", and hence ξ_t has the same distribution as ξ_0 for every $t>0$ (here we could also use the fact that $\mathcal{N}$ is in fact locally compact for the weak topology). Theorem 4.1 now follows as before.

Remark. For a general result on stationarity of interacting particle processes (but with another state space!), using an equation of type (4.13), see, e.g., [13, Proposition 6.10, p. 52].

The argument shows more generally that, if we start with a Poisson point process of intensity $\lambda_1>0$ on $\mathbb{R}_+$ and a Poisson point process of intensity $\lambda_1\lambda_2>0$ in $\mathbb{R}_+^2$, the starting distribution on $\mathbb{R}_+$ is invariant for the Hammersley–Aldous–Diaconis process, if we let points escape at zero at rate λ_2.

It is also clear that the inequality $c\leqslant 2$ follows, since the length of a longest North–East path from $(0,0)$ to a point (t,t), with $t>0$, will, in the construction above, be always at least as big as the length of a longest North–East path from $(0,0)$ to a point (t,t), if we start with the empty configuration on the x- and y-axis: we simply have more opportunities for forming a North–East path, if we allow them to pick up points from the x- or y-axis. Since, starting with a Poisson process of intensity 1 in the first quadrant, and (independently) Poisson processes of intensity 1 on the x- and y-axis, the expected length of a longest North–East path to (t,t) will be *exactly* equal to $2t$, according to what we proved above, we obtain from this $c\leqslant 2$.

5. Seppäläinen's stick process

The result $c=2$ is proved by hydrodynamical arguments in Sections 8 and 9 of [18]. I will summarize the approach below.

First of all, a counting process on $\mathbb{R}$ instead of $(0,\infty)$ is used, and for this process a number of starting configurations are considered. Note that we cannot start with the empty configuration on $\mathbb{R}$, since points would immediately be pulled to $-\infty$, as noted in [2]. For the purposes of proving $c=2$, the most important starting configurations are:

(i) a Poisson process of intensity 1 on $(-\infty,0]$ and the empty configuration on $(0,\infty)$.
(ii) a Poisson process of intensity 1 on $\mathbb{R}$.

Let $(z_k)_{k\in\mathbb{Z}}$ be an initial configuration on $\mathbb{R}$. Seppäläinen's *stick process* is defined as a process of functions $u_t:\mathbb{Z}\to\mathbb{R}$, associated with this particle process, by

$$u_t(k)=z_k(t)-z_{k-1}(t),\quad t\geqslant 0,\ k\in\mathbb{Z}.$$

Instead of $z_k(0)$ we write z_k.

We now define

$$N(x,0)=\sup\{k\colon z_k\leqslant x\},$$

and

$$N(x,y)=\sup_{-\infty<z\leqslant x}\{N(z,0)+L^{\nearrow}((z,0),(x,y))\},\tag{5.1}$$

where $L^{\nearrow}((z,0),(x,y))$ is the maximum number of points on a North–East path in $(z,x]\times(0,y]$, as in Lemma 3.1 of Section 3.

Define again

$$L^{\nearrow}((x_1,y_1),(x_2,y_2))\tag{5.2}$$

as the maximum number of points on a North–East path in $(x_1,x_2]\times(y_1,y_2]$, where $x_2>x_1$ and $y_2>y_1$. The key to the approach in [18] is to work with a kind of inverse of (5.2), defined by

$$\Gamma((x_1,y_1),(y_2,k))=\inf\{u\geqslant 0\colon L^{\nearrow}((x_1,y_1),(x_1+u,y_2))\geqslant k\},\tag{5.3}$$

in words: $\Gamma((x_1,y_1),y_2,k))$ is the minimum horizontal distance needed for building a North–East path of k points, starting at (x_1,y_1), in the "time interval" $(y_1,y_2]$. This can be seen as another way of expanding relation (2.16), which, in fact, is also a relation between the discrete time Hammersley process and its inverse.

Now, given the initial configuration $(z_k)_{k\in\mathbb{Z}}$, the position of the particle z_k at time $y>0$ is given by

$$z_k(y)=\inf_{i\leqslant k}\{z_i+\Gamma((z_i,0),y,k-i)\}.\tag{5.4}$$

Note that for each point $z_k(y)$ at time y, originating from z_k in the original configuration, there will always be a point z_j of the original configuration, with $j\leqslant k$, such that

$$\inf_{i\leqslant k}\{z_i+\Gamma((z_i,0),y,k-i)\}=z_j+\Gamma((z_j,0),y,k-j).$$

For example, if $z_{k-1}<z_k(y)<z_k$ we get a path of length 1 from z_{k-1} to $z_k(y)$, and

$$\inf_{i\leqslant k}\{z_i+\Gamma((z_i,0),y,k-i)\}=z_{k-1}+\Gamma((z_{k-1},0),y,1)=z_k(y).$$

Similarly, if $z_k(y)<z_{k-1}$, we can always construct a path from a point z_j, with $j<k$ to $z_k(y)$ through points of the Poisson point process, "picked up" by the preceding paths ("seen from $z_k(y)$", these are descending corners in the preceding paths).

These points need not be uniquely determined. Proposition 4.4 in [18] clarifies the situation. It asserts that almost surely (that is: for almost every realization of the point processes) we have that for all $y>0$ and each $k\in\mathbb{Z}$ there exist integers $i^-(k,y)$ and $i^+(k,y)$ such that $i^-(k,y)\leqslant i^+(k,y)$ and

$$z_k(y)=z_j+\Gamma((z_j,0),y,k-j)$$

holds for $j = i^-(k,y)$ and $j = i^+(k,y)$, but fails for $j \notin [i^-(k,y), i^+(k,y)]$. The proof of this Proposition 4.4 in [18] is in fact fairly subtle!

We now first consider (proceeding a little bit differently than in [18] the evolution of the initial configuration on $(-\infty, 0]$ in the case that we have a Poisson process of intensity 1 on $(-\infty, 0]$ and the empty configuration on $(0, \infty)$, using the same method as used in Section 4. Let $\mathscr{F}$ be the set of continuous functions $f : (-\infty, 0] \to \mathbb{R}_+$ with support $[a,b] \subset (-\infty, 0)$. We denote the point process of the starting configuration by ξ_0 and the configuration at time $t > 0$ by ξ_t, where we let it develop according to the rules of the Hammersley–Aldous–Diaconis process. Then, just as before, we can prove

$$\lim_{h \downarrow 0} h^{-1} E\{\exp\{-\xi_h(f)\} - \exp\{-\xi_0(f)\} \mid \xi_0\} = 0, \tag{5.5}$$

for $f \in \mathscr{F}$. So, in the case that we have a Poisson process of intensity 1 on $(-\infty, 0]$, the empty configuration on $(0, \infty)$, and a Poisson process of intensity 1 in the upper half plane, the distribution of the initial configuration is invariant for the Hammersley–Aldous–Diaconis process.

Let P_0 be the probability measure, associated with the initial configuration (i) in the beginning of this section, and let, for $i \leqslant -1$, $\eta_i = z_i - z_{i-1}$ (the length of the "stick" at location i), where we let z_{-1} be the biggest point of the initial configuration in $(-\infty, 0)$. Moreover, let the measure m_0 on the Borel sets of $\mathbb{R}$ be defined by

$$m_0([a,b]) = b \wedge 0 - a \wedge 0, \quad -\infty < a < b < \infty,$$

and let $\eta_i = 0$, for $i \geqslant 0$ (in this way η_i is defined for all $i \in \mathbb{Z}$). Then we have, for each $\varepsilon > 0$, and each interval $[a,b] \subset \mathbb{R}$:

$$\lim_{n \to \infty} P_0 \left\{ \left| n^{-1} \sum_{i=[na]}^{[nb]} \eta_i - m_0([a,b]) \right| > \varepsilon \right\} = 0,$$

which corresponds to condition (1.10) in [18]. This condition plays a similar role as condition (3.2) in Section 3 above. Here $[x]$ denotes the largest integer $\leqslant x$, for $x \in \mathbb{R}$. It is then proved that, if $x > c^2 y/4 + 2$,

$$t^{-1} \sum_{i=0}^{[tx]} \eta_i(ty) \xrightarrow{\text{p}} \int_0^x u(v,y)\,\mathrm{d}v = \frac{1}{4} c^2 y, \tag{5.6}$$

where u is a (weak) solution of the partial differential equation

$$\frac{\partial}{\partial y} u(x,y) + \frac{1}{4} c^2 \frac{\partial}{\partial x} u(x,y)^2 = 0, \tag{5.7}$$

under the initial condition $u(x,0) = 1_{(-\infty,0]}(x)$, $x \in \mathbb{R}$. But since we also have

$$t^{-1} E \sum_{i=0}^{[tx]} \eta_i(ty) = y - \frac{1}{2t} \int_0^{ty} E\eta_{[tx]}^2(s)\,\mathrm{d}s, \tag{5.8}$$

if $x > c^2 y/4 + 2$, we get $c = 2$, if we can prove

$$\lim_{t \to \infty} \frac{1}{2t} \int_0^{ty} E\eta_{[tx]}^2(s)\,\mathrm{d}s = 0.$$

This is in fact proved in [18, p. 32]. The first term on the right of (5.8) comes from $\frac{1}{2}E\eta_{-1}^2=1$, and relation (5.8) follows from

$$t^{-1}E\sum_{i=0}^{[tx]}\eta_i(ty)=-\frac{1}{2}t^{-1}\sum_{i=0}^{[tx]}\int_0^{ty}E\{\eta_i^2(s)-\eta_{i-1}^2(s)\}\,\mathrm{d}s,$$

using

$$E\{\eta_i(y+h)-\eta_i(y)\,|\,\eta_i(y),\eta_{i-1}(y)\}=-\tfrac{1}{2}h\{\eta_i^2(y)-\eta_{i-1}^2(y)\}+\mathrm{o}(h),\quad h\downarrow 0.$$

We return to (5.6). Another interpretation of this relation is

$$t^{-1}\{z_{[tx]}(ty)-z_{-1}(ty)\}\xrightarrow{\mathrm{p}}U(x,y)-U(0,y),\quad t\to\infty, \tag{5.9}$$

where

$$z_{[tx]}(ty)=\inf_{j\leqslant[tx]}\{z_j+\Gamma((z_j,0),ty,[tx]-j)\},$$

and

$$U(x,y)=\inf_{z\leqslant x}\left\{z\wedge 0+\frac{(x-z)^2}{c^2y}\right\},$$

using Theorem A1 on p. 38 of [18], with $f(x)=\frac{1}{4}c^2x^2$, $x\in\mathbb{R}$. This corresponds to relation (8.8) on p. 29 of [18], which implies that

$$t^{-1}z_{[tx]}(ty)\xrightarrow{\mathrm{p}}U(x,y),\quad t\to\infty.$$

An easy calculation shows

$$U(x,y)=\left(x-\frac{1}{4}c^2y\right)\wedge 0, \tag{5.10}$$

since $U(x,y)\leqslant 0$ (note that $z\wedge 0+(x-z)^2/(c^2y)=0$ if $z=x$), and since $U(x,y)<0$ can only occur if $x-\frac{1}{4}c^2y<0$, in which case the minimizing z is given by $z=x-\frac{1}{2}c^2y$. Note that

$$t^{-1}z_{[tx]}(ty)\xrightarrow{\mathrm{p}}U(x,y)=0,$$

if $x\geqslant\frac{1}{4}c^2y$, and that the right side of (5.9) in this case is given by

$$-U(0,y)=\tfrac{1}{4}c^2y.$$

So the partial differential equation (5.7), with initial condition

$$u_0(x)=1_{(-\infty,0]}(x),\quad x\in\mathbb{R},$$

is solved by

$$u(x,y)=1_{\{x\leqslant c^2y/4\}}(x,y)=1_{(-\infty,c^2y/4]}(x),\quad x\in\mathbb{R}, \tag{5.11}$$

which, considered as a function of x, is the Radon–Nikodym derivative of the Borel measure m_y on $\mathbb{R}$, defined by

$$m_y((x_1,x_2])=U(x_2,y)-U(x_1,y),\quad -\infty<x_1<x_2<\infty.$$

Note that u only solves (5.7) in a distributional sense. That is: for any continuously differentiable "test function" $\phi : \mathbb{R} \to \mathbb{R}$ with compact support and any $y > 0$, we have

$$\int_{\mathbb{R}} \phi(x)u(x,y)\,\mathrm{d}x - \int_{\mathbb{R}} \phi(x)u_0(x)\,\mathrm{d}x = \frac{1}{4}c^2 \int_0^y \int_{\mathbb{R}} \phi'(x)u(x,t)^2\,\mathrm{d}x\,\mathrm{d}t, \tag{5.12}$$

see also (1.6) on p. 5 of [18].

We note that in Seppäläinen's interpretation of the particle process, we cannot get new points to the right of zero in the situation where we start with a Poisson point process of intensity 1 on $(-\infty,0]$ and the empty configuration on $(0,\infty)$. In this situation we have $z_i = z_{-1}$, $i \geqslant 0$, so we have infinitely many particles at location z_{-1}. This means that at each time y, where a new point of the Poisson point process in the plane occurs with x-coordinate just to the left of x_{-1}, and also to the right of points $z_i(y)$, satisfying $z_i(y) < z_{-1}$, one of the infinitely many points $z_i(y)$ that are still left at location z_{-1} shifts to the x-coordinate of this new point. Seppäläinen's interpretation with the "moving to the left" is perhaps most clearly stated in the first paragraph of Section 2 of [18].

In the interpretation of Seppäläinen's "stick process", we would have an infinite number of sticks of zero length at sites $0,1,2,\ldots$. Each time an event of the above type occurs, one of the sticks of length zero gets positive length, and the stick at the preceding site is shortened (corresponding to the shift to the left of the corresponding particle in the particle process). In this way, mass gradually shifts to the right in the sense that at each event of this type a new stick with an index higher than all indices of sticks with positive length gets itself positive length. The corresponding "macroscopic" picture is that the initial profile $u(\cdot,0) = 1_{(-\infty,0]}$ shifts to $u(\cdot,y) = 1_{(-\infty,y]}$ at time y.

The interpretation of the relation

$$t^{-1}z_{[tx]}(ty) \xrightarrow{\mathrm{p}} (x-y) \wedge 0,$$

taking $c=2$ in (5.10), is that $z_{[tx]}(0)$ travels roughly over a distance $t(y-x)$ to the left in the time interval $[0,ty]$, if $y > x \geqslant 0$ (we first need a time interval tx to get to a "stick with index $[tx]$" and length zero, then another time interval of length $t(y-x)$ to build a distance left from zero of order $t(y-x)$), and that (with high probability) $z_{[tx]}(0)$ does not travel at all during this time interval, if $x > y$ (a "stick with index $[tx]$" is not reached during this time interval).

6. Concluding remarks

In the foregoing sections I tried to explain the "hydrodynamical approach" to the theory of longest increasing subsequences of random permutations, for the uniform distribution on the set of permutations. This approach probably started with the paper [10] and the heuristics that can be found in that paper have been expanded in different directions in the papers discussed above. The hydrodynamical approach has also been used to investigate large deviation properties, see, e.g., [19], and for large deviations of the upper tail: [7]; they still treat the large deviations of the lower tail by combinatorial methods, using Young diagrams, but apparently would prefer to have a proof of a more probabilistic nature, as seems clear from their remark: "The proof based on the random Young tableau correspondence is purely combinatoric and sheds no light on the random mechanism responsible for the large deviations".

I conjecture that it is possible to push the hydrodynamical approach further for proving the asymptotic distribution results in [5], but at present these results still completely rely on an analytic representation of the probability distribution of longest increasing subsequences using Toeplitz determinants, see, e.g. [6, p. 636].

Acknowledgements

I want to thank the referees for their constructive comments.

References

[1] D. Aldous, Personal communication, 2000.
[2] D. Aldous, P. Diaconis, Hammersley's interacting particle process and longest increasing subsequences, Probab. Theory Relat. Fields 103 (1995) 199–213.
[3] D. Aldous, P. Diaconis, Longest increasing subsequences: from patience sorting to the Baik–Deift–Johansson theorem, Bull. Amer. Math. Soc 36 (1999) 413–432.
[4] M. Aigner, G.M. Ziegler, Proofs from The Book, Springer, Berlin, 1998.
[5] J. Baik, P.A. Deift, P. Johansson, On the distribution of the length of the longest increasing subsequences of random permutations, J. Amer. Math. Soc. 12 (1999) 1119–1178.
[6] P. Deift, Integrable systems and combinatorial theory, Notices AMS 47 (2000) 631–640.
[7] J.D. Deuschel, O. Zeitouni, On increasing subsequences of I.I.D. samples, Combin. Probab. Comput. 3 (2000) 247–263; 47 (2000) 631–640.
[8] J. Dieudonné, Foundations of Modern Analysis, Academic Press, New York, 1969.
[9] P. Erdös, G. Szekeres, A combinatorial problem in geometry, Compositio Math. 2 (1935) 463–470.
[10] J.M. Hammersley, A few seedlings of research, in: Proc. 6th Berkeley Symp. Mathematical Statistics and Probability, Vol. 1, 1972, pp. 345–394.
[11] O. Kallenberg, Random Measures, 4th Edition, Akademie-Verlag, Berlin, 1986.
[12] J.F.C. Kingman, Subadditive ergodic theory, Ann. Probab. 1 (1973) 883–909.
[13] T.M. Liggett, Interacting Particle Systems, Springer, New York, 1985.
[14] B.F. Logan, L.A. Shepp, A variational problem for random Young tableaux, Adv. Math. 26 (1977) 206–222.
[15] E. Rains, A mean identity for longest increasing subsequence problems. Preprint, 2000.
[16] W. Rudin, Functional Analysis, 2nd Edition, McGraw-Hill, New York, 1991.
[17] B.E. Sagan, The Symmetric Group, Wadsworth, Belmont, 1991.
[18] T. Seppäläinen, A microscopic model for the Burgers equation and longest increasing subsequences, Electronic J. Probab. 1 (1996) Paper 5, 1–51.
[19] T. Seppäläinen, Large deviations for increasing sequences on the plane, Probab. Theory Relat. Fields 112 (1998) 221–244.
[20] A.M. Vershik, S.V. Kerov, Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux. Soviet Math. Dokl. 18 (1977) 527–531. (Translation of Dokl. Acad. Nauk. SSSR 32, 1024–1027.)
[21] A.M. Vershik, S.V. Kerov, Asymptotic behavior of the maximum and generic dimensions of irreducible representations of the symmetric group, Functional Anal. Appl. 3 (1981) 246–255.
[22] A.M. Vershik, S.V. Kerov, Asymptotic theory of the characters of the symmetric group, Functional Anal. Appl. 19 (1985) 21–31.
[23] G. Viennot, Une forme géométrique de la correspondence de Robinson-Schensted, in: D. Foata (Ed.), Combinatoire et Représentation du Groupe Symétrique, Lecture Notes in Math., Vol. 579, Springer, New York, 1976.