



# The Ulam-Hammersley problem for multiset permutations

Lucas Gerin

## ► To cite this version:

| Lucas Gerin. The Ulam-Hammersley problem for multiset permutations. 2023. hal-03920072v4

**HAL Id: hal-03920072**

**<https://hal.science/hal-03920072v4>**

Preprint submitted on 20 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Ulam-Hammersley problem for multiset permutations

LUCAS GERIN

March 20, 2024

## Abstract

We obtain the asymptotic behaviour of the longest increasing/non-decreasing subsequences in a random uniform multiset permutation in which each element in  $\{1, \dots, n\}$  occurs  $k$  times, where  $k$  may depend on  $n$ . This generalizes the famous Ulam-Hammersley problem of the case  $k = 1$ . The proof relies on poissonization and on a careful non-asymptotic analysis of variants of the Hammersley-Aldous-Diaconis particle system.

**Keywords:** Combinatorial probability, longest increasing subsequence, interacting particle systems, Hammersley process.

## 1 Introduction

A  $k$ -multiset permutation of size  $n$  is a word with letters in  $\{1, 2, \dots, n\}$  such that each letter appears exactly  $k$  times. When this is convenient we identify a multiset permutation  $s = (s(1), \dots, s(kn))$  and the set of points  $\{(i, s(i)), 1 \leq i \leq kn\}$ . We introduce two partial orders over the quarter-plane  $[0, \infty)^2$ :

$$\begin{aligned} (x, y) &\prec (x', y') \text{ if } x < x' \text{ and } y < y', \\ (x, y) &\preceq (x', y') \text{ if } x < x' \text{ and } y \leq y'. \end{aligned}$$

For a finite set  $\mathcal{P}$  of points in the quarter-plane we put

$$\begin{aligned} \mathcal{L}_{<}(\mathcal{P}) &= \max \{L; \text{ there exists } P_1 \prec P_2 \prec \dots \prec P_L, \text{ where each } P_i \in \mathcal{P}\}, \\ \mathcal{L}_{\leq}(\mathcal{P}) &= \max \{L; \text{ there exists } P_1 \preceq P_2 \preceq \dots \preceq P_L, \text{ where each } P_i \in \mathcal{P}\}. \end{aligned}$$

In words the integer  $\mathcal{L}_{<}(\mathcal{P})$  (resp.  $\mathcal{L}_{\leq}(\mathcal{P})$ ) is the length of the longest increasing (resp. non-decreasing) subsequence of  $\mathcal{P}$ .

Let  $S_{k;n}$  be a  $k$ -multiset permutation of size  $n$  drawn uniformly among the  $\frac{(kn)!}{k!^n}$  possibilities. In the case  $k = 1$  the word  $S_{1;n}$  is simply a uniform permutation and estimating  $\mathcal{L}_{<}(S_{1;n}) =$

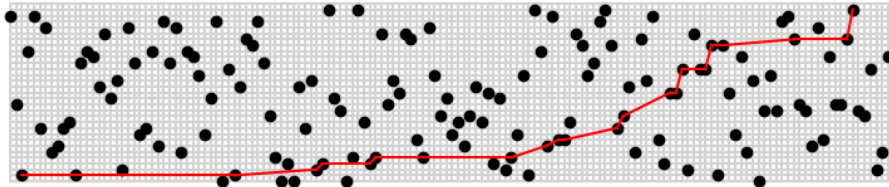


Figure 1: A uniform 5-multiset permutation  $S_{5,30}$  of size  $n = 30$  and one of its longest non-decreasing subsequences.

$\mathcal{L}_{\leq}(S_{1;n})$  is known as the Hammersley or Ulam-Hammersley problem. The first order was solved by Veršik and Kerov [VK77] and simultaneously by Logan and Shepp:

$$\mathbb{E}[\mathcal{L}_{<}(S_{1;n})] \stackrel{n \rightarrow +\infty}{\sim} 2\sqrt{n}.$$

Note that the above limit also holds in probability<sup>1</sup>:  $\mathcal{L}_{<}(S_{1;n}) = 2\sqrt{n} + o_{\mathbb{P}}(\sqrt{n})$ . This problem has a long history and has revealed deep and unexpected connections between combinatorics, interacting particle systems, calculus of variations, random matrix theory, representation theory. We refer to Romik [Rom15] for a very nice description of this problem and some of its ramifications.

In the context of card guessing games it is asked in [CDH<sup>+</sup>22, Question 4.3] the behaviour of  $\mathcal{L}_{<}(S_{k;n})$  for a fixed  $k$ . Using the Veršik-Kerov Theorem we can make an educated guess. The intuition is that, for fixed  $k$ , it is quite unlikely that many points at the same height contribute to the same longest increasing/non-decreasing subsequence. Thus at the first order everything should happen as if the  $kn$  points had distinct heights and we expect that

$$\mathcal{L}_{<}(S_{k;n}) \approx \mathcal{L}_{\leq}(S_{k;n}) \approx \mathcal{L}_{<}(S_{1;kn}) \approx 2\sqrt{kn}.$$

The original motivation of the present paper was to make this approximation rigorous. We actually adress this question in the case where  $k$  depends on  $n$ .

**Theorem 1** (Longest increasing subsequences). *Let  $(k_n)$  be a sequence of integers such that  $k_n \leq n$  for all  $n$ . Then<sup>2</sup>*

$$\mathbb{E}[\mathcal{L}_{<}(S_{k_n;n})] = 2\sqrt{nk_n} - k_n + o(\sqrt{nk_n}). \quad (1)$$

(Of course if  $k_n = o(n)$  then the RHS of (1) reduces to  $2\sqrt{nk_n} + o(\sqrt{nk_n})$ .)

**Theorem 2** (Longest non-decreasing subsequences). *Let  $(k_n)$  be an arbitrary sequence of integers. Then*

$$\mathbb{E}[\mathcal{L}_{\leq}(S_{k_n;n})] = 2\sqrt{nk_n} + k_n + o(\sqrt{nk_n}). \quad (2)$$

**Strategy of proof and organization of the paper.** In Section 2 we first provide the proof of Theorems 1 and 2 in the case of a constant or slowly growing sequence  $(k_n)$ . The proof is elementary (assuming known the Veršik-Kerov Theorem).

For the general case we first borrow a few tools in the literature. In particular we introduce and analyze poissonized versions of  $\mathcal{L}_{<}(S_{k_n;n})$ ,  $\mathcal{L}_{\leq}(S_{k_n;n})$ . As already suggested by Hammersley ([Ham72], Sec.9) and achieved by Aldous-Diaconis [AD95] the case  $k = 1$  can be tackled by considering an interacting particle system which is now known as the Hammersley or Hammersley-Aldous-Diaconis (HAD) process.

In Section 3 we introduce and analyze the two variants of the Hammersley process adapted to multiset permutations. The first one is the discrete-time HAD process [Fer96, FM06], the second one had recently appeared in [Boy22] with a connection to the O'Connell-Yor Brownian polymer. The standard path to analyze Hammersley-like processes consists in using subadditivity

<sup>1</sup>We use the shortcut  $o_{\mathbb{P}}(a_n)$  for a sequence of random variables  $(X_n)$  such that  $X_n/a_n \rightarrow 0$  in probability.

<sup>2</sup>If  $k_n \geq n$  for some  $n$  then the following greedy strategy shows that  $\mathbb{E}[\mathcal{L}_{<}(S_{k_n;n})] = n - o(n)$  so the picture is complete.

Indeed, first choose the leftmost point  $(x_1, 1)$  in  $S_{k_n;n}$  with height 1. Then recursively define  $(x_\ell, \ell)$  at the leftmost point (if any) in  $S_{k_n;n}$  with height  $\ell$  such that  $x_\ell > x_{\ell-1}$ , and so on until you are stuck (either because  $\ell = n$  or because there is no point in  $S_{k_n;n} \cap (x_{\ell-1}, kn] \times \{\ell\}$ ). A few elementary computations show that this strategy defines an increasing path of length  $n - o(n)$  with probability tending to one. As  $\mathcal{L}_{<}(S_{k_n;n}) \leq n$  a.s. this yields  $\mathbb{E}[\mathcal{L}_{<}(S_{k_n;n})] = n - o(n)$ .

to prove the existence of a limiting shape and then proving that this limiting shape satisfies a variational problem. Typically this variational problem is solved either using convex duality [Sep97, CG19] or through the analysis of *second class particles* [CG06, CG19]. The issue here is that since we allow  $k_n$  to have different scales we cannot use this approach and we need to derive non-asymptotic bounds for both processes. This is the purpose of Theorem 9 whose proof is the most technical part of the paper. In Section 4 we detail the multivariate de-poissonization procedure in order to conclude the proof of Theorem 1. De-poissonization is more convoluted for non-decreasing subsequences: see Section 5.

**Beyond expectation.** In the course of the proof we actually obtain results beyond the estimation of the expectation. We obtain concentration inequalities for the poissonized version of  $\mathcal{L}_<(S_{k_n;n}), \mathcal{L}_\leq(S_{k_n;n})$ : see Theorem 9 and also the discussion in Section 6. We also obtain the convergence in probability, unfortunately for some technical reasons we miss a small range of scales of  $(k_n)$ 's.

**Proposition 3.** *Let  $(k_n)$  be either a small or a large sequence. Then*

$$\frac{\mathcal{L}_<(S_{k_n;n})}{2\sqrt{nk_n - k_n}} \xrightarrow{\text{prob.}} 1, \quad \frac{\mathcal{L}_\leq(S_{k_n;n})}{2\sqrt{nk_n + k_n}} \xrightarrow{\text{prob.}} 1.$$

We refer to (3),(31) below for the formal definitions of small/large sequences. Let us just say that sequences such that  $k_n = \mathcal{O}((\log n)^{1-\varepsilon})$  for some  $\varepsilon > 0$  are small while sequences such that  $(\log n)^{1+\varepsilon} = \mathcal{O}(k_n)$  are large. Sequences in-between are neither small nor large so in Proposition 3 we miss scales like  $k_n \approx \log(n)$ .

Regarding fluctuations a famous result by Baik, Deift and Johansson [BDJ99, Th.1.1] states that

$$\frac{\mathcal{L}_\leq(S_{1;n}) - 2\sqrt{n}}{n^{1/6}} \xrightarrow{(d)} \text{TW}$$

where TW is the Tracy-Widom distribution. The intuition given by the comparison with the Hammersley process would suggest that the fluctuations of  $\mathcal{L}_<(S_{k_n;n}), \mathcal{L}_\leq(S_{k_n;n})$  might be of order  $(k_n n)^{1/6}$  as long as  $(k_n)$  does not grow too fast. A natural question to explore for furthering this work would involve understanding for which  $(k_n)$  the model preserves KPZ scaling exponents. The non-asymptotic estimates of Section 3 could serve as a first step in this direction.

**Comparison with previous works.** There are only few random sets  $\mathcal{P}$  for which the asymptotics of  $\mathcal{L}_<(\mathcal{P}), \mathcal{L}_\leq(\mathcal{P})$  are known:

- As already mentioned, the case of a uniform permutation (and its poissonized version) is very well understood, via different approaches. For proofs close to the spirit of the present paper, we refer to [AD95] and [CG05].
- The case where  $\mathcal{P}$  is given by a field of i.i.d. Bernoulli random variables on the square grid has been solved by Seppäläinen in [Sep97] for  $\mathcal{L}_<$  and in [Sep98] for  $\mathcal{L}_\leq$ . (See [BEGG16] also for an elementary proof of both results).

We are not aware of previous results for multiset permutations. However Theorems 1 and 2 in the linear regime  $k_n \sim \text{constant} \times n$  should be compared to a result by Biane ([Bia01, Theorem 3]).

We need a few notations to describe his result. Let  $\mathcal{W}_{q_N;N}$  be the random word given by  $q_N$  i.i.d. uniform letters in  $\{1, 2, \dots, N\}$ . The word  $\mathcal{W}_{q_N;N}$  is not a multiset permutation but since for large  $N$  there are in average  $q_N/N$  points on each horizontal line of  $\mathcal{W}_{q_N;N}$  we expect that  $\mathcal{L}_<(\mathcal{W}_{q_N;N}) \approx \mathcal{L}_<(S_{q_N/N;N})$  and  $\mathcal{L}_\leq(\mathcal{W}_{q_N;N}) \approx \mathcal{L}_\leq(S_{q_N/N;N})$ .

Biane obtains the exact limiting shape of the random Young Tableau induced through the RSK correspondence by  $\mathcal{W}_{q_N;N}$  in the regime where  $\sqrt{q_N}/N \rightarrow c$  for some constant  $c > 0$ . As the length of the first row (resp. the number of rows) in the Young Tableau corresponds to the length of the longest non-decreasing subsequence in  $\mathcal{W}_{k;n}$  (resp. the length of the longest decreasing sequence) a consequence of ([Bia01, Theorem 3]) is that, in probability,

$$\liminf \frac{1}{\sqrt{q_N}} \mathcal{L}_{<}(\mathcal{W}_{q_N;N}) \geq (2 - c), \quad \limsup \frac{1}{\sqrt{q_N}} \mathcal{L}_{\leq}(\mathcal{W}_{q_N;N}) \leq (2 + c).$$

For that regime our Theorems 1 and 2 respectively suggest:

$$\begin{aligned} \mathcal{L}_{<}(\mathcal{W}_{q_N;N}) &\approx \mathcal{L}_{<}(S_{q_N/N;N}) \approx \mathcal{L}_{<}(S_{c^2N;N}) \sim 2Nc - c^2N \sim (2 - c)\sqrt{q_N}, \\ \mathcal{L}_{\leq}(\mathcal{W}_{q_N;N}) &\approx \mathcal{L}_{\leq}(S_{q_N/N;N}) \approx \mathcal{L}_{\leq}(S_{c^2N;N}) \sim 2Nc + c^2N \sim (2 + c)\sqrt{q_N}, \end{aligned}$$

which is indeed consistent with Biane's result.

## 2 Preliminaries: the case of small $k_n$

We first prove Theorems 1 and 2 in the case of a *small* sequence  $(k_n)$ . We say that a sequence  $(k_n)$  of integers is *small* if

$$k_n^2(k_n)! = o(\sqrt{n}). \quad (3)$$

Note that a sequence of the form  $k_n = (\log n)^{1-\varepsilon}$  is small while  $k_n = \log n$  is not small.

*Proof of Theorems 1 and 2 in the case of a small sequence  $(k_n)$ .* (In order to lighten notation we skip the dependence in  $n$  and write  $k = k_n$ .)

Let  $\sigma_{kn}$  be a random uniform permutation of size  $kn$ . We can associate to  $\sigma_{kn}$  a  $k$ -multiset permutation  $S_{k;n}$  in the following way. For every  $1 \leq i \leq kn$  we put

$$S_{k;n}(i) = \lceil \sigma(i)/k \rceil.$$

It is clear that  $S_{k;n}$  is uniform and we have

$$\mathcal{L}_{<}(S_{k;n}) \leq \mathcal{L}_{\leq}(\sigma_{kn}) \leq \mathcal{L}_{\leq}(S_{k;n}). \quad (4)$$

The Vershik-Kerov Theorem says that the middle term in the above inequality grows like  $2\sqrt{kn}$ . Hence we need to show that if  $(k_n)$  is small then

$$\mathcal{L}_{\leq}(S_{k;n}) = \mathcal{L}_{<}(S_{k;n}) + o_{\mathbb{P}}(\sqrt{kn}),$$

which proves the small case of Proposition 3 and Theorems 1 and 2. For this purpose we introduce for every  $\delta > 0$  the event

$$\mathcal{E}_{\delta} := \{ \mathcal{L}_{\leq}(S_{k;n}) \geq \mathcal{L}_{<}(S_{k;n}) + \delta\sqrt{n} \}.$$

If  $\mathcal{E}_{\delta}$  occurs then in particular there exists a non-decreasing subsequence with  $\delta\sqrt{n}$  ties, *i.e.* points of  $S_{k;n}$  which are at the same height as their predecessor in the subsequence. These ties have distinct heights  $1 \leq i_1 < \dots < i_{\ell} \leq n$  for some  $\delta\sqrt{n}/k \leq \ell \leq \delta\sqrt{n}$ . Fix

- Integers  $m_1, \dots, m_{\ell} \geq 2$  such that  $(m_1 - 1) + \dots + (m_{\ell} - 1) = \delta\sqrt{n}$ ;
- Column indices  $r_{1,1} < \dots < r_{1,m_1} < r_{2,1} < r_{2,m_1} < \dots < r_{\ell,1} < \dots < r_{\ell,m_{\ell}}$ .

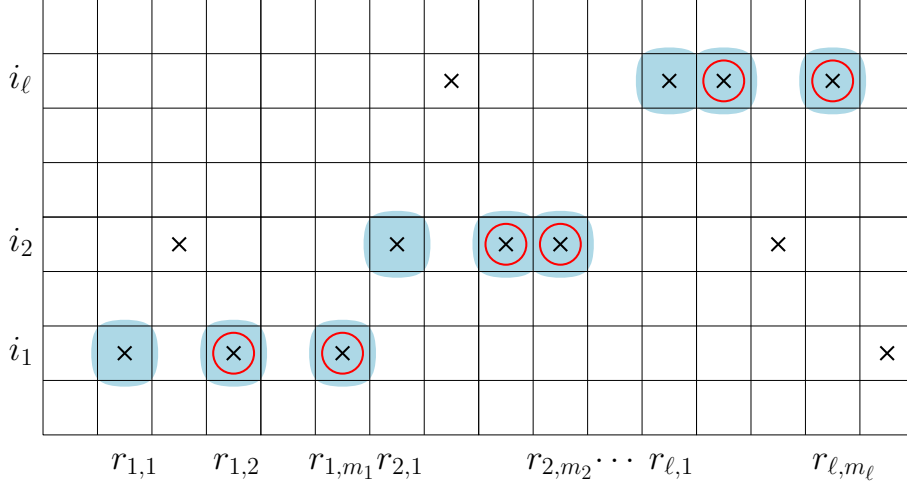


Figure 2: The event  $F$ . (Ties are surrounded in red. Points with blue background represent the subsequence with  $\delta\sqrt{n}$  ties.)

We then introduce the event

$$F = F((i_\ell)_\ell, (r_{i,j})_{i \leq \ell, j \leq m_i}) \\ = \{S(r_{1,1}) = \dots = S(r_{1,m_1}) = i_1, S(r_{2,1}) = \dots = S(r_{2,m_2}) = i_2, \dots, S(r_{\ell,1}) = \dots = S(r_{\ell,m_\ell}) = i_\ell\}.$$

By the union bound (we skip the integer parts)

$$\mathbb{P}(\mathcal{E}_\delta) \leq \sum_{\delta\sqrt{n}/k \leq \ell \leq \delta\sqrt{n}} \sum_{1 \leq i_1 < \dots < i_\ell \leq n} \sum_{(r_{i,j})_{i \leq \ell, j \leq m_i}} \mathbb{P}(F((i_\ell)_\ell, (r_{i,j})_{i \leq \ell, j \leq m_i})).$$

Using that

$$\text{card} \left\{ \sum m_i = \delta\sqrt{n} + \ell; \text{ each } m_i \geq 2 \right\} = \text{card} \left\{ \sum p_i = \delta\sqrt{n}; \text{ each } p_i \geq 1 \right\} = \binom{\delta\sqrt{n} - 1}{\ell - 1}$$

we obtain

$$\sum_{(r_{i,j})_{i \leq \ell, j \leq m_i}} \mathbb{P}(F) = \frac{1}{\binom{kn}{k \dots k}} \underbrace{\binom{nk}{\sum m_i}}_{\text{choices of } r\text{'s}} \underbrace{\binom{\delta\sqrt{n} - 1}{\ell - 1}}_{\text{choices of } m_i\text{'s}} \underbrace{\binom{kn - \sum m_i}{(k - m_1)(k - m_2) \dots (k - m_\ell)k \dots k}}_{\text{choices of } kn - \sum m_i \text{ remaining points}} \\ = \frac{(k!)^\ell (\delta\sqrt{n} - 1)!}{(\delta\sqrt{n} + \ell)! (\delta\sqrt{n} - \ell)! (\ell - 1)! (k - m_1)! (k - m_2)! \times \dots \times (k - m_\ell)!}.$$

Bounding each factor  $(k - m_i)!$  by 1 we get

$$\sum_{(r_{i,j})_{i \leq \ell, j \leq m_i}} \mathbb{P}(F) \leq \frac{(k!)^\ell}{(\delta\sqrt{n})^{\ell+1} (\delta\sqrt{n} - \ell)! (\ell - 1)!}.$$

We now sum over  $1 \leq i_1 < \dots < i_\ell \leq n$  and then sum over  $\ell$ :

$$\mathbb{P}(\mathcal{E}_\delta) \leq \sum_{\ell = \delta\sqrt{n}/k}^{\delta\sqrt{n}} \binom{n}{\ell} \frac{(k!)^\ell}{(\delta\sqrt{n})^{\ell+1} (\delta\sqrt{n} - \ell)! (\ell - 1)!} \\ \leq \sum_{\ell = \delta\sqrt{n}/k}^{\delta\sqrt{n}-3} \binom{n}{\ell} \frac{(k!)^\ell}{(\delta\sqrt{n})^{\ell+1} (\delta\sqrt{n} - \ell)! (\ell - 1)!} + 3 \binom{n}{\delta\sqrt{n}} \frac{(k!)^{\delta\sqrt{n}}}{(\delta\sqrt{n})^{\delta\sqrt{n}-2} (\delta\sqrt{n} - 3)!} \quad (5)$$

Using the two following inequalities valid for every  $j \leq m$  (see *e.g.* [CLRS09, eq.(C.5)])

$$\binom{m}{j} \leq \left(\frac{me}{j}\right)^j, \quad m! \geq m^m \exp(-m)$$

we first obtain that if  $k_n! = o(\sqrt{n})$  (which is the case if  $(k_n)$  is small) then the last term of (5) tends to zero. Regarding the sum we write

$$\begin{aligned} \mathbb{P}(\mathcal{E}_\delta) &\leq \sum_{\ell=\delta\sqrt{n}/k}^{\delta\sqrt{n}-3} \left(\frac{ne}{\ell}\right)^\ell \frac{(k!)^\ell}{(\delta\sqrt{n})^{\ell+1}(\delta\sqrt{n}-\ell)^{\delta\sqrt{n}-\ell}e^{-\delta\sqrt{n}+\ell}(\ell-1)^{\ell-1}e^{-\ell+1}} + o(1) \\ &\leq \sum_{\ell=\delta\sqrt{n}/k}^{\delta\sqrt{n}-3} \left(\frac{nek!(\delta\sqrt{n}-\ell)}{\delta\sqrt{n}\ell(\ell-1)}\right)^\ell \underbrace{\frac{(\ell-1)e^{-1}}{\delta\sqrt{n}}}_{\leq 1} \underbrace{\left(\frac{e}{\delta\sqrt{n}-\ell}\right)^{\delta\sqrt{n}}}_{\leq e/3 < 1} + o(1) \\ &\leq \sum_{\ell=\delta\sqrt{n}/k}^{\delta\sqrt{n}-3} \left(\frac{\sqrt{ne}k!(\delta\sqrt{n}-\ell)}{\delta\ell(\ell-1)}\right)^\ell \left(\frac{e}{\delta\sqrt{n}-\ell}\right)^\ell + o(1) \\ &\leq \sum_{\ell=\delta\sqrt{n}/k}^{\delta\sqrt{n}-3} \left(\frac{\sqrt{ne}e^2k!}{\delta\ell(\ell-1)}\right)^\ell + o(1) \leq \sum_{\ell=\delta\sqrt{n}/k}^{\delta\sqrt{n}-3} \left(\frac{e^2k^2k!}{\delta^3\sqrt{n}}\right)^\ell + o(1) \end{aligned} \tag{6}$$

which tends to zero for every  $\delta > 0$ , as long as  $(k_n)$  satisfies (3). This proves that  $\mathcal{L}_\leq(S_{k;n}) = \mathcal{L}_<(S_{k;n}) + o_{\mathbb{P}}(\sqrt{nk_n})$ . Combining this with (4), this proves that

$$\frac{\mathcal{L}_<(S_{k;n})}{2\sqrt{nk_n}} \xrightarrow{\text{prob.}} 1, \quad \frac{\mathcal{L}_\leq(S_{k;n})}{2\sqrt{nk_n}} \xrightarrow{\text{prob.}} 1,$$

which is the "small" case of Proposition 3 since  $k_n = o(\sqrt{nk_n})$ .

To conclude the proof of small cases of Theorems 1 and 2 we observe that we have the crude bounds  $\mathcal{L}_<(S_{k;n}) \leq n$  and  $\mathcal{L}_\leq(S_{k;n}) \leq nk_n$ . This allows us to write

$$\mathbb{E}[|\mathcal{L}_\leq(S_{k;n}) - \mathcal{L}_<(S_{k;n})|] \leq \delta\sqrt{n} + nk_n \times \mathbb{P}(\text{not } \mathcal{E}_\delta)$$

Together with eq.(6) this implies that

$$\mathbb{E}[\mathcal{L}_\leq(S_{k;n})] = \mathbb{E}[\mathcal{L}_<(S_{k;n})] + o(\sqrt{nk_n}).$$

We use again Veršik-Kerov and (4) to deduce that both sides are  $2\sqrt{nk_n} + o(\sqrt{nk_n})$ .  $\square$

### 3 Poissonization: variants of the Hammersley process

In this section we define formally and analyze two *semi-discrete* variants of the Hammersley process.

**Remark.** In the sequel,  $\text{Poisson}(\mu)$  (resp.  $\text{Binomial}(n, q)$ ) stand for generic random variables with Poisson distribution with mean  $\mu$  (resp. Binomial distribution with parameters  $n, q$ ).

Notation  $\text{Geometric}_{\geq 0}(1-\beta)$  stands for a geometric random variable with the convention  $\mathbb{P}(\text{Geometric}_{\geq 0}(1-\beta) = k) = (1-\beta)\beta^k$  for  $k \geq 0$ . In particular  $\mathbb{E}[\text{Geometric}_{\geq 0}(1-\beta)] = \frac{\beta}{1-\beta}$ .

### 3.1 Definitions of the processes $L_{<}(t)$ and $L_{\leq}(t)$

For a parameter  $\lambda > 0$  let  $\Pi^{(\lambda)}$  be the random set  $\Pi^{(\lambda)} = \cup_i \Pi_i^{(\lambda)}$  where  $\Pi_i^{(\lambda)}$ 's are independent and each  $\Pi_i^{(\lambda)}$  is a homogeneous Poisson Point Process (PPP) with intensity  $\lambda$  on  $(0, \infty) \times \{i\}$ . For simplicity set

$$\Pi_{x,t}^{(\lambda)} = \Pi^{(\lambda)} \cap ([0, x] \times \{1, \dots, t\}).$$

The goal of the present section is to obtain non-asymptotic bounds for  $\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)})$  and  $\mathcal{L}_{\leq}(\Pi_{x,t}^{(\lambda)})$ . Indeed if we then choose

$$\lambda_n \approx \frac{1}{n}, \quad x = kn, \quad t = n$$

then there are  $kn + \mathcal{O}(\sqrt{kn})$  points on each line of a  $\Pi_{x,t}^{(\lambda)}$  and we expect that

$$\mathcal{L}_{<}(\Pi_{kn,n}^{(\lambda_n)}) \approx \mathcal{L}_{<}(S_{k;n}), \quad \mathcal{L}_{\leq}(\Pi_{kn,n}^{(\lambda_n)}) \approx \mathcal{L}_{\leq}(S_{k;n})$$

Fix  $x > 1$  throughout the section. For every  $t \in \{0, 1, 2, \dots\}$  the function  $y \in [0, x] \mapsto \mathcal{L}_{<}(y, t)$  (resp.  $\mathcal{L}_{\leq}(y, t)$ ) is a non-decreasing integer-valued function whose all steps are equal to  $+1$ . Therefore this function is completely determined by the finite set

$$L_{<}(t) := \{y \leq x, \mathcal{L}_{<}(y, t) = \mathcal{L}_{<}(y, t^-) + 1\}.$$

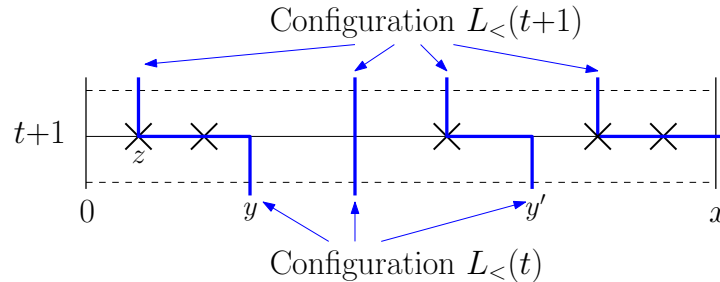
(Respectively:

$$L_{\leq}(t) := \{y \leq x, \mathcal{L}_{\leq}(y, t) = \mathcal{L}_{\leq}(y, t^-) + 1\}.)$$

Sets  $L_{<}(t)$  and  $L_{\leq}(t)$  are finite subsets of  $[0, x]$  whose elements are considered as particles. It is easy to see that for fixed  $x > 0$  both processes  $(L_{<}(t))_t$  and  $(L_{\leq}(t))_t$  are Markov processes taking their values in the family of point processes of  $[0, x]$ .

Exactly the same way as for the classical Hammersley process ([Ham72, Sec.9], [AD95]) the individual dynamic of particles is very easy to describe:

- **The process  $L_{<}$ .** We put  $L_{<}(0) = \emptyset$ . In order to define  $L_{<}(t+1)$  from  $L_{<}(t)$  we consider particles from left to right. A particle at  $y$  in  $L_{<}(t)$  moves at time  $t+1$  at the location of the leftmost available point  $z$  in  $\Pi_{t+1}^{(\lambda)} \cap (0, y)$  (if any, otherwise it stays at  $y$ ). This point  $z$  is not available anymore for subsequent particles, as well as every other point of  $\Pi_{t+1}^{(\lambda)} \cap (0, y)$ .



If there is a point in  $\Pi_{t+1}^{(\lambda)}$  which is on the right of  $y' := \max\{L_{<}(t)\}$  then a new particle is created in  $L_{<}(t+1)$ , located at the leftmost point in  $\Pi_{t+1}^{(\lambda)} \cap (y', x)$ . (In pictures this new particle comes from the right.)

A realization of  $L_{<}$  is shown on top-left of Fig.3.



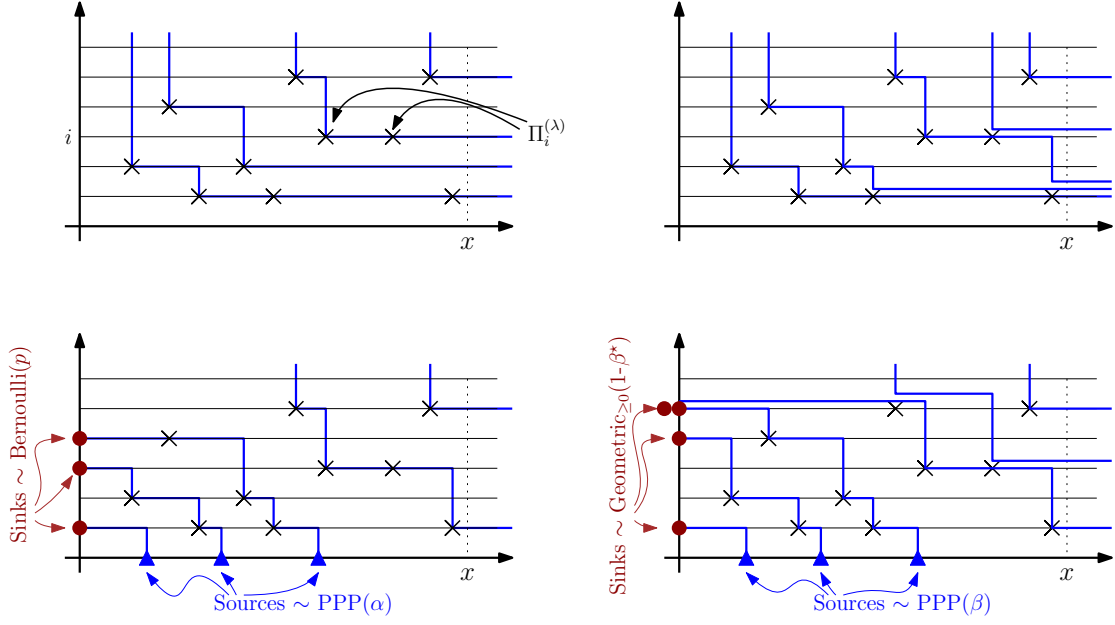
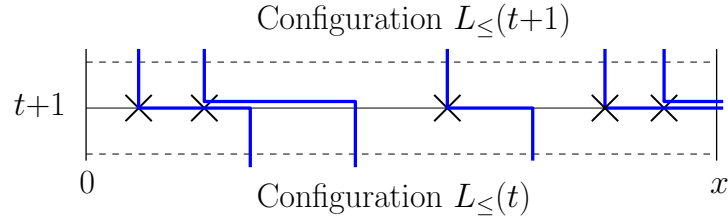


Figure 3: Our four variants of the Hammersley process (time goes from bottom to top, trajectories of particles are indicated in blue). Top left: The process  $L_{<}(t)$ . Top right: The process  $L_{\leq}(t)$ . Bottom left: The process  $L_{\leq}^{(\alpha,p)}(t)$ . Bottom right: The process  $L_{\leq}^{(\beta,\beta^*)}(t)$ .

- **The process  $L_{\leq}$ .** We put  $L_{\leq}(0) = \emptyset$ . In order to define  $L_{\leq}(t+1)$  from  $L_{\leq}(t)$  we also consider particles from left to right. A particle at  $y$  in  $L_{\leq}(t)$  moves at time  $t+1$  at the location of the leftmost available point  $z$  in  $\Pi_{t+1}^{(\lambda)} \cap (0, y)$ . This point  $z$  is not available anymore for subsequent particles, **other points in  $(z, y)$  remain available**.



If there is a point in  $\Pi_{t+1}^{(\lambda)}$  which is on the right of  $y' := \max\{L_{<}(t)\}$  then new particles are created in  $L_{<}(t+1)$ , one for each point in  $\Pi_{t+1}^{(\lambda)} \cap (y', x)$ .  
A realization of  $L_{\leq}$  is shown on top-right of Fig.3.

Processes  $L_{<}(t)$  and  $L_{\leq}(t)$  are designed in such a way that they record the length of longest increasing/non-decreasing paths in  $\Pi$ . In fact particles trajectories correspond to the level sets of the functions  $(x, t) \mapsto \mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)})$ ,  $(x, t) \mapsto \mathcal{L}_{\leq}(\Pi_{x,t}^{(\lambda)})$ .

**Proposition 4.** For every  $x$ ,

$$\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)}) = \text{card}(L_{<}(t)), \quad \mathcal{L}_{\leq}(\Pi_{x,t}^{(\lambda)}) = \text{card}(L_{\leq}(t)),$$

where on each right-hand side we consider the particle system on  $[0, x]$ .

*Proof.* We are merely restating the original construction from Hammersley ([Ham72], Sec.9). We only do the case of  $L_<(t)$ .

Let us call each particle trajectory a *Hammersley line*. By construction each Hammersley line is a broken line starting from the right of the box  $[0, x] \times [0, t]$  and is formed by a succession of north/west line segments. Because of this, two distinct points in a given longest increasing subsequence of  $\Pi_{x,t}^{(\lambda)}$  cannot belong to the same Hammersley line. Since there are  $L_<(t)$  Hammersley's lines this gives  $\mathcal{L}_<(\Pi_{x,t}^{(\lambda)}) \leq \text{card}(L_<(t))$ .

In order to prove the converse inequality we build from this graphical construction a longest increcreasing subsequence of  $\Pi_{x,t}^{(\lambda)}$  with exactly one point on each Hammersley line. To do so, we order Hammersley's lines from bottom-left to top-right, and we build our path starting from the top-right corner. We first choose any point of  $\Pi_{x,t}^{(\lambda)}$  belonging to the last Hammersley line. We then proceed by induction: we choose the next point among the points of  $\Pi_{x,t}^{(\lambda)}$  lying on the previous Hammersley line such that the subsequence remains increasing. (This is possible since Hammersley's lines only have North/West line segments.) This proves  $\mathcal{L}_<(\Pi_{x,t}^{(\lambda)}) \geq \text{card}(L_<(t))$ .  $\square$

### 3.2 Sources and sinks: stationarity

Proposition 4 tells us that in our way to prove Theorem 1 and Theorem 2 we need to understand the asymptotic behaviour of processes  $L_<, L_\leq$ .

It is proved in [FM06] that the homogeneous PPP with intensity  $\alpha$  on  $\mathbb{R}$  is stationary for  $(L_<(t))_t$ . However we need non-asymptotic estimates for  $(L_<(t))_t$  (and  $(L_\leq(t))_t$ ) on a given interval  $(0, x)$ . To solve this issue we use the trick of *sources/sinks* introduced formally and exploited by Cator and Groeneboom [CG05] for the continuous HAD process:

- *Sources* form a finite subset of  $[0, x] \times \{0\}$  which plays the role of the initial configuration  $L_<(0), L_\leq(0)$ .
- *Sinks* are points of  $\{0\} \times [1, t]$  which add up to  $\Pi^{(\lambda)}$  when one defines the dynamics of  $L_<(t), L_\leq(t)$ . For  $L_\leq(t)$  it makes sense to add several sinks at the same location  $(0, i)$  so sinks may have a multiplicity.

Examples of dynamics of  $L_<, L_\leq$  under the influence of sources/sinks is illustrated at the bottom of Fig.3.

Here is the discrete-time analogous of [CG05, Th.3.1.]:

**Lemma 5.** *For every  $\lambda, \alpha > 0$  let  $L_<^{(\alpha, p)}(t)$  be the Hammersley process defined as  $L_<(t)$  with:*

- *sources distributed according to a homogeneous PPP with intensity  $\alpha$  on  $[0, x] \times \{0\}$  ;*
- *sinks distributed according to i.i.d. Bernoulli( $p$ ) with*

$$\frac{\lambda}{\lambda + \alpha} = p. \quad (7)$$

*If sources, sinks, and  $\Pi^{(\lambda)}$  are independent then the process  $(L_<^{(\alpha, p)}(t))_{t \geq 0}$  is stationary.*

**Lemma 6.** *For every  $\beta > \lambda > 0$  let  $L_\leq^{(\beta, \beta^*)}(t)$  be the Hammersley process defined as  $L_\leq(t)$  with:*

- *sources distributed according to a homogeneous PPP with intensity  $\beta$  on  $[0, x] \times \{0\}$  ;*

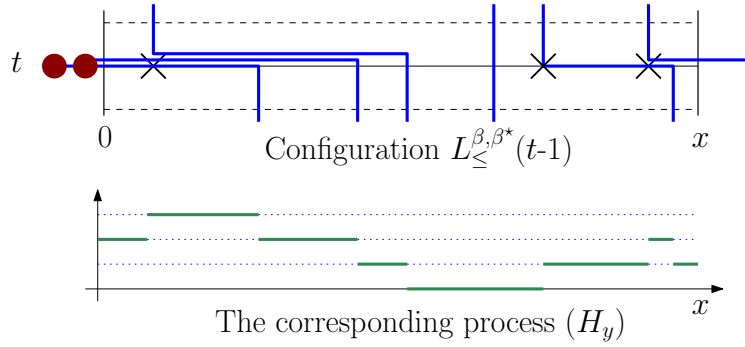
- sinks distributed according to i.i.d.  $\text{Geometric}_{\geq 0}(1 - \beta^*)$  with

$$\beta^* \beta = \lambda. \quad (8)$$

If sources, sinks, and  $\Pi^{(\lambda)}$  are independent then the process  $(L_{\leq}^{(\beta, \beta^*)}(t))_{t \geq 0}$  is stationary.

*Proof of Lemmas 5 and 6.* Lemma 6 could be obtained from minor adjustments of [Boy22, Chap.3, Lemma 3.2]. (Be aware that we have to switch  $x \leftrightarrow t$  and sources  $\leftrightarrow$  sinks in [Boy22] in order to fit our setup.) For the sake of the reader we however propose the following alternative proof which explains where (8) come from.

Consider for some fixed  $t \geq 1$  the process  $(H_y)_{0 \leq y \leq x}$  given by the number of Hammersley lines passing through the point  $(y, t)$ .



The initial value  $H_0$  is the number of sinks at  $(0, t)$ , which is distributed as a  $\text{Geometric}_{\geq 0}(1 - \beta^*)$ . The process  $(H_y)$  is a random walk (reflected at zero) with '+1 rate' equal to  $\lambda$  and '-1 rate' equal to  $\beta$ . (Jumps of  $(H_y)$  are independent from sinks as sinks are independent from  $\Pi^{(\lambda)}$ .) The  $\text{Geometric}_{\geq 0}(1 - \beta^*)$  distribution is stationary for this random walk exactly when (8) holds. The set of points of  $L_{\leq}^{(\beta, \beta^*)}(t)$  is given by the union of  $\Pi_t^{(\lambda)}$  and the points of  $L_{\leq}^{(\beta, \beta^*)}(t)$  that do not correspond to a '-1' jump. Computations given in Appendix B show that this is distributed as a homogeneous PPP with intensity  $\beta$ .

Lemma 5 is proved exactly in the same way, calculations are even easier. In this case the corresponding process  $(H_y)_{0 \leq y \leq x}$  takes its values in  $\{0, 1\}$  and its stationary distribution is the Bernoulli distribution with mean  $\lambda/(\alpha + \lambda)$ , hence (7).  $\square$

### 3.3 Processes $L_{<}(t)$ and $L_{\leq}(t)$ : non-asymptotic bounds

From Lemmas 5 and 6 it is straightforward to derive non-asymptotic upper bounds for  $L_{<}(t), L_{\leq}(t)$ .

For  $y \leq x$  let  $\text{So}_x^{(\alpha)}$  be the random set of sources with intensity  $\alpha$  and for  $s \leq t$  let  $\text{Si}_t^{(p)}$  the random set of sinks with intensity  $p$ . In particular,

$$\text{card}(\text{So}_x^{(\alpha)}) \stackrel{(d)}{=} \text{Poisson}(\alpha x), \quad \text{card}(\text{Si}_t^{(p)}) \stackrel{(d)}{=} \text{Binomial}(t, p).$$

It is convenient to use the notation  $\mathcal{L}_{=<}(\mathcal{P})$  which is, as before, the length of the longest increasing path taking points in  $\mathcal{P}$  but when the path is also allowed to go through several sources (which have however the same  $y$ -coordinate) or several sinks (which have the same  $x$ -coordinate). Formally,

$$\mathcal{L}_{=<}(\mathcal{P}) = \max \{L; \text{ there exists } P_1 =< P_2 =< \dots =< P_L, \text{ where each } P_i \in \mathcal{P} \},$$

where

$$(x, y) = \prec (x', y') \text{ if } \begin{cases} x < x' \text{ and } y < y', \\ \text{or } x = x' = 0 \text{ and } y < y', \\ \text{or } x < x' \text{ and } y = y' = 0. \end{cases}$$

Proposition 4 generalizes easily to the settings of sinks and sources.

**Claim.**

$$\mathcal{L}_{=<} \left( \Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\alpha)} \cup \text{Si}_t^{(p)} \right) = L_{<}^{(\alpha,p)}(t) + \text{card}(\text{Si}_t^{(p)}). \quad (9)$$

*Proof of the Claim.* By the same reasoning as in the proof of Proposition 4 the LHS is exactly the number of broken lines in the box  $[0, x] \times [0, t]$ . Each such line escapes the box either through the left (it thus corresponds to a sink) or through the top (and is thus counted by  $L_{<}^{(\alpha,p)}(t)$ ).  $\square$

**Lemma 7** (Domination for  $\mathcal{L}_{<}$ ). *For every  $\alpha, p \in (0, 1)$  such that (7) holds, there is a stochastic domination of the form:*

$$\mathcal{L}_{<} \left( \Pi_{x,t}^{(\lambda)} \right) \preceq \text{Poisson}(x\alpha) + \text{Binomial}(t, p). \quad (10)$$

(The Poisson and Binomial random variables involved in (10) are not independent.)

*Proof.* Adding sources and sinks may not decrease longest increasing paths. Thus,

$$\begin{aligned} \mathcal{L}_{<} \left( \Pi_{x,t}^{(\lambda)} \right) &\preceq \mathcal{L}_{=<} \left( \Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\alpha)} \cup \text{Si}_t^{(p)} \right) \\ &= L_{<}^{(\alpha,p)}(t) + \text{card}(\text{Si}_t^{(p)}) \text{ (using (9))} \\ &\stackrel{(d)}{=} L_{<}^{(\alpha,p)}(0) + \text{card}(\text{Si}_t^{(p)}) \text{ (using stationarity: Lemma 5)} \\ &\stackrel{(d)}{=} \text{Poisson}(x\alpha) + \text{Binomial}(t, p). \end{aligned} \quad \square$$

Taking expectations in (10) we obtain

$$\mathbb{E} \left[ \mathcal{L}_{<} \left( \Pi_{x,t}^{(\lambda)} \right) \right] \leq x\alpha + tp.$$

The LHS in the above equation does not depend on  $\alpha, p$  so the idea is to apply (10) with the minimizing choice

$$\bar{\alpha}, \bar{p} := \operatorname{argmin}_{\alpha, p \text{ satisfying (7)}} \{x\alpha + tp\},$$

i.e.

$$\bar{\alpha} = \sqrt{\frac{t\lambda}{x}} - \lambda, \quad \bar{p} = \sqrt{\frac{x\lambda}{t}}, \quad x\bar{\alpha} + t\bar{p} = 2\sqrt{xt\lambda} - x\lambda. \quad (11)$$

We have proved

$$\mathbb{E} \left[ \mathcal{L}_{<} \left( \Pi_{x,t}^{(\lambda)} \right) \right] \leq 2\sqrt{xt\lambda} - x\lambda.$$

(Compare with (1).) We have a similar statement for non-decreasing subsequences:

**Lemma 8** (Domination for  $\mathcal{L}_{\leq}$ ). *For every  $\beta, \beta^* \in (0, 1)$  such that (8) holds, there is a stochastic domination of the form:*

$$\mathcal{L}_{\leq} \left( \Pi_{x,t}^{(\lambda)} \right) \preceq \text{Poisson}(x\beta) + \mathcal{G}_1^{(\beta^*)} + \cdots + \mathcal{G}_t^{(\beta^*)}, \quad (12)$$

where  $\mathcal{G}_i^{(\beta^*)}$ 's are i.i.d.  $\text{Geometric}_{\geq 0}(1 - \beta^*)$ .

We put

$$\bar{\beta}, \bar{\beta}^* := \operatorname{argmin}_{\beta, \beta^* \text{ satisfying (8)}} \left\{ x\beta + t \left( \frac{\beta^*}{1 - \beta^*} \right) \right\}, \quad (13)$$

*i.e.*

$$\bar{\beta} = \sqrt{\frac{t\lambda}{x}} + \lambda, \quad \bar{\beta}^* = \frac{1}{1 + \sqrt{t/x\lambda}}, \quad x\bar{\beta} + t \left( \frac{\bar{\beta}^*}{1 - \bar{\beta}^*} \right) = 2\sqrt{xt\lambda} + x\lambda. \quad (14)$$

(In particular  $\bar{\beta} > \lambda$ , as required in Lemma 6.) Eq.(12) yields

$$\mathbb{E} \left[ \mathcal{L}_{\leq} \left( \Pi_{x,t}^{(\lambda)} \right) \right] \leq 2\sqrt{xt\lambda} + x\lambda. \quad (15)$$

(Compare with (2).)

**Theorem 9** (Concentration for  $\mathcal{L}_{<}$ ,  $\mathcal{L}_{\leq}$ ). *There exist strictly positive functions  $g, h$  such that for all  $\varepsilon > 0$  and for every  $x, t \geq 1$ ,  $\lambda > 0$  such that  $t \geq x\lambda$*

$$\mathbb{P}(\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)}) > (1 + \varepsilon)(2\sqrt{xt\lambda} - x\lambda)) \leq \exp(-g(\varepsilon)(\sqrt{xt\lambda} - x\lambda)), \quad (16)$$

$$\mathbb{P}(\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)}) < (1 - \varepsilon)(2\sqrt{xt\lambda} - x\lambda)) \leq \exp(-h(\varepsilon)(\sqrt{xt\lambda} - x\lambda)). \quad (17)$$

*Similarly:*

$$\mathbb{P}(\mathcal{L}_{\leq}(\Pi_{x,t}^{(\lambda)}) > (1 + \varepsilon)(2\sqrt{xt\lambda} + x\lambda)) \leq \exp(-g(\varepsilon)\sqrt{xt\lambda}), \quad (18)$$

$$\mathbb{P}(\mathcal{L}_{\leq}(\Pi_{x,t}^{(\lambda)}) < (1 - \varepsilon)(2\sqrt{xt\lambda} + x\lambda)) \leq \exp(-h(\varepsilon)\sqrt{xt\lambda}). \quad (19)$$

For the proof of Theorem 9 we will focus on the case of  $\mathcal{L}_{<}$ , *i.e.* eq.(16), (17). When necessary we will give the slight modification needed to prove eq.(18) and (19). The beginning of the proof mimics Lemmas 4.1 and 4.2 in [BEGG16].

We first prove similar bounds for the stationary processes with minimizing sources and sinks.

**Lemma 10** (Concentration for  $\mathcal{L}_{<}$  with sources and sinks). *Let  $\bar{\alpha}, \bar{p}$  be defined by (11). There exists a strictly positive function  $g_1$  such that for all  $\varepsilon > 0$  and for every  $x, t \geq 1$ ,  $\lambda > 0$  such that  $t \geq x\lambda$*

$$\mathbb{P}(\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) > (1 + \varepsilon)(2\sqrt{xt\lambda} - x\lambda)) \leq 2 \exp(-g_1(\varepsilon)(\sqrt{xt\lambda} - x\lambda)) \quad (20)$$

$$\mathbb{P}(\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) < (1 - \varepsilon)(2\sqrt{xt\lambda} - x\lambda)) \leq 2 \exp(-g_1(\varepsilon)(\sqrt{xt\lambda} - x\lambda)). \quad (21)$$

*Proof of Lemma 10.* By stationarity (Lemma 5) we have

$$\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) \stackrel{(d)}{=} \text{Poisson}(x\bar{\alpha}) + \text{Binomial}(t, \bar{p}).$$

Then

$$\begin{aligned} \mathbb{P}(\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) > (1 + \varepsilon)(2\sqrt{xt\lambda} - x\lambda)) &\leq \mathbb{P} \left( \text{Poisson}(x\bar{\alpha}) > (1 + \frac{\varepsilon}{2})(\sqrt{xt\lambda} - x\lambda) \right) \\ &\quad + \mathbb{P} \left( \text{Binomial}(t, \bar{p}) > (1 + \frac{\varepsilon}{2})\sqrt{xt\lambda} \right). \end{aligned}$$

Recall that  $x\bar{\alpha} = \sqrt{xt\lambda} - x\lambda$ ,  $t\bar{p} = \sqrt{xt\lambda}$ . Using the tail inequality for the Poisson distribution (Lemma 15):

$$\mathbb{P} \left( \text{Poisson}(x\bar{\alpha}) > (1 + \frac{\varepsilon}{2})(\sqrt{xt\lambda} - x\lambda) \right) \leq \exp \left( -(\sqrt{xt\lambda} - x\lambda)\varepsilon^2/4 \right)$$

Using the tail inequality for the binomial (Lemma 16) we get

$$\mathbb{P} \left( \text{Binomial}(t, \bar{p}) > (1 + \frac{\varepsilon}{2})\sqrt{xt\lambda} \right) \leq \exp(-\frac{1}{12}\varepsilon^2\sqrt{xt\lambda}) \leq \exp(-\frac{1}{12}\varepsilon^2(\sqrt{xt\lambda} - x\lambda)) \quad (22)$$

The proof of (21) is identical. This shows Lemma 10 with  $g_1(\varepsilon) = \varepsilon^2/12$ .  $\square$

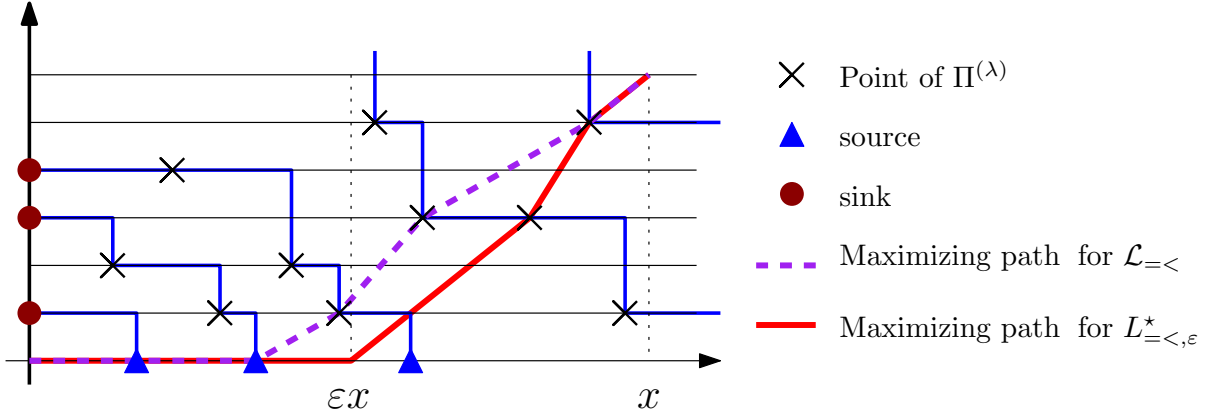


Figure 4: A sample of  $\Pi_{x,t}^{(\lambda)}$ , sources, sinks, and the corresponding trajectories of particles (in blue). Here  $\mathcal{L}_{=<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) = 5$  (pink path) and  $L_{<}^{(\alpha,p)}(t) = 2$  (two remaining particles at the top of the box).

For longest non-decreasing subsequences we have a statement similar to Lemma 10. The only modification in the proof is that in order to estimate the number of sinks one has to replace Lemma 16 (tail inequality for the Binomial) by Lemma 17 (tail inequality for a sum of geometric random variables<sup>3</sup>). During the proof we need to bound  $\sqrt{xt\lambda} + x\lambda$  by  $\sqrt{xt\lambda}$ , this explains the form of the right-hand side in eq.(18) and (19).

*Proof of Theorem 9.* Adding sources/sinks may not decrease  $\mathcal{L}_{<}$  so

$$\mathcal{L}_{=<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) \geq \mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)}),$$

thus the upper bound (16) is a direct consequence of Lemma 10.

Let us now prove the lower bound. We consider the length of a maximizing path among those using sources from 0 to  $\varepsilon x$  and then only increasing points of  $\Pi_{x,t}^{(\lambda)} \cap ([\varepsilon x, x] \times [0, t])$  (see Fig.4). Formally we set

$$\begin{aligned} L_{=<,\varepsilon}^* &:= \text{card} \left( \text{So}_{\varepsilon x}^{(\bar{\alpha})} \right) + \mathcal{L}_{<} \left( \Pi_{x,t}^{(\lambda)} \cap ([\varepsilon x, x] \times [0, t]) \right) \\ &\stackrel{(d)}{=} \text{Poisson}(\varepsilon x \bar{\alpha}) + \mathcal{L}_{<} \left( \Pi_{x,t}^{(\lambda)} \cap ([\varepsilon x, x] \times [0, t]) \right). \end{aligned} \quad (23)$$

The idea is that for any fixed  $\varepsilon$  the paths contributing to  $L_{=<,\varepsilon}^*$  will typically not contribute to  $\mathcal{L}_{=<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) = L_{<}^{(\bar{\alpha},p)}(t) + \text{card}(\text{Si}_t^{(\bar{p})})$ . Indeed eq.(23) suggests that for large  $x, t$

$$\begin{aligned} L_{=<,\varepsilon}^* &\approx \mathbb{E}[\text{Poisson}(\varepsilon x \bar{\alpha})] + \mathbb{E} \left[ \mathcal{L}_{<} \left( \Pi_{x,t}^{(\lambda)} \cap ([\varepsilon x, x] \times [0, t]) \right) \right] \\ &\approx x\varepsilon \bar{\alpha} + 2\sqrt{x(1-\varepsilon)\lambda t} - x(1-\varepsilon)\lambda \\ &= 2\sqrt{x\lambda t} - x\lambda - \sqrt{xt\lambda}\delta(\varepsilon), \end{aligned}$$

where  $\delta(\varepsilon) = 2 - \varepsilon - 2\sqrt{1-\varepsilon}$  is positive and increasing. In order to make the above approximation rigorous we first write

$$2\sqrt{x\lambda t} - x\lambda - \frac{1}{2}\sqrt{xt\lambda}\delta(\varepsilon) = x\varepsilon \bar{\alpha} + \frac{1}{4}\sqrt{xt\lambda}\delta(\varepsilon) + 2\sqrt{x(1-\varepsilon)\lambda t} - x(1-\varepsilon)\lambda + \frac{1}{4}\sqrt{xt\lambda}\delta(\varepsilon). \quad (24)$$

<sup>3</sup>Note that it is only stated for  $0 < \varepsilon < 1$  but this is enough for our purpose since the left-hand side of (22) is non-increasing in  $\varepsilon$ .

Combining (23) and (24) gives

$$\mathbb{P}\left(L_{=<,\varepsilon}^* \geq 2\sqrt{x\lambda t} - x\lambda - \frac{1}{2}\sqrt{xt\lambda}\delta(\varepsilon)\right) \leq \mathbb{P}_1 + \mathbb{P}_2$$

where

$$\begin{aligned}\mathbb{P}_1 &= \mathbb{P}\left(\text{Poisson}(x\varepsilon\bar{\alpha}) \geq x\varepsilon\bar{\alpha} + \frac{1}{4}\sqrt{xt\lambda}\delta(\varepsilon)\right), \\ \mathbb{P}_2 &= \mathbb{P}\left(\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)} \cap ([\varepsilon x, x] \times [0, t])) \geq 2\sqrt{x(1-\varepsilon)\lambda t} - x(1-\varepsilon)\lambda + \frac{1}{4}\sqrt{xt\lambda}\delta(\varepsilon)\right).\end{aligned}$$

Using the tail inequality for the Poisson distribution (see Lemma 15) we have that

$$\mathbb{P}_1 \leq \exp\left(-\frac{xt\lambda\delta(\varepsilon)^2}{16 \times 4\varepsilon^2(\sqrt{xt\lambda} - x\lambda)}\right) \leq \exp\left(-\sqrt{xt\lambda}\delta(\varepsilon)^2/64\varepsilon^2\right).$$

Besides

$$\begin{aligned}\mathbb{P}_2 &\leq \mathbb{P}\left(\mathcal{L}_{<}(\Pi_{x,t}^{(\lambda)} \cap ([\varepsilon x, x] \times [0, t])) \geq \left(2\sqrt{x(1-\varepsilon)\lambda t} - x(1-\varepsilon)\lambda\right) \times \left(1 + \frac{1}{8}\delta(\varepsilon)\right)\right) \\ &\leq \exp\left(-g(\delta(\varepsilon)/8)(\sqrt{x(1-\varepsilon)t\lambda} - x(1-\varepsilon)\lambda)\right) \text{ (using the upper bound (16)).}\end{aligned}$$

Finally we can find some positive  $h$  such that

$$\mathbb{P}\left(L_{=<,\varepsilon}^* \geq 2\sqrt{x\lambda t} - x\lambda - \frac{1}{2}\sqrt{xt\lambda}\delta(\varepsilon)\right) \leq \exp\left(-h(\varepsilon)(\sqrt{xt\lambda} - x\lambda)\right). \quad (25)$$

One proves exactly in the same way a similar bound for the length of a maximizing path among those using sinks in  $\{0\} \times [0, \varepsilon t]$  and then only increasing points of  $\Pi_{x,t}^{(\lambda)} \cap ([0, x] \times [\varepsilon t, t])$ .

Choose now one of the maximizing paths  $\mathcal{P}$  for  $\mathcal{L}_{=<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})})$  (if there are many of them, choose one arbitrarily in a deterministic way: the lowest, say). Denote by  $\text{sources}(\mathcal{P})$  and  $\text{sinks}(\mathcal{P})$  the number of sources and sinks in the path  $\mathcal{P}$ :

$$\text{sources}(\mathcal{P}) = \text{card}\{0 \leq y \leq x \text{ such that } (y, 0) \in \mathcal{P}\}.$$

In Fig.4 the path  $\mathcal{P}$  is sketched in pink, in that example  $\text{sources}(\mathcal{P}) = 2$ ,  $\text{sinks}(\mathcal{P}) = 0$ .

**Lemma 11.** *There exists a positive function  $\psi$  such that for all real  $\eta > 0$*

$$\mathbb{P}\left(\text{sources}(\mathcal{P}) + \text{sinks}(\mathcal{P}) \geq \eta\sqrt{x\lambda t}\right) \leq 2\exp(-\psi(\eta)(\sqrt{x\lambda t} - x\lambda)).$$

*Proof of Lemma 11.* (As the left-hand side is non-increasing in  $\eta$  it is enough to prove the lemma for  $\eta < 1$ .)

If the event  $\{\text{sources}(\mathcal{P}) \geq \eta\sqrt{x\lambda t}\}$  holds then there exists a (random)  $\varepsilon$  such that the two following events occur:

- $\text{So}_{\varepsilon x} \geq \eta\sqrt{x\lambda t}$  ;
- $L_{=<,\varepsilon}^* = \mathcal{L}_{=<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) = L_{<}^{(\bar{\alpha}, \bar{p})}(t) + \text{card}(\text{Si}_t^{(\bar{p})}).$

This implies that this random  $\varepsilon$  is larger than  $\eta/2 > 0$  unless the number of sources in  $[0, x\eta/2]$  is improbably high:

$$\begin{aligned}
\mathbb{P}(\text{sources}(\mathcal{P}) \geq \eta\sqrt{x\lambda t}) &\leq \mathbb{P}(\text{So}_{\eta x/2} \geq \eta\sqrt{x\lambda t}) + \mathbb{P}(\text{sources}(\mathcal{P}) \geq \eta\sqrt{x\lambda t}; \text{So}_{\eta x/2} < \eta\sqrt{x\lambda t}) \\
&\leq \mathbb{P}(\text{So}_{\eta x/2} \geq \eta\sqrt{x\lambda t}) &=: \mathbb{P}_3 \\
&+ \mathbb{P}(L_{<}^{(\bar{\alpha}, \bar{p})}(t) \leq \sqrt{x\lambda t} - x\lambda - \frac{1}{4}\delta(\eta/3)\sqrt{x\lambda t}) &=: \mathbb{P}_4 \\
&+ \mathbb{P}(\text{card}(\text{Si}_t^{(\bar{p})}) \leq \sqrt{x\lambda t} - \frac{1}{4}\delta(\eta/3)\sqrt{x\lambda t}) &=: \mathbb{P}_5 \\
&+ \mathbb{P}(L_{<}^* \geq 2\sqrt{x\lambda t} - x\lambda - \frac{1}{2}\delta(\eta/3)\sqrt{x\lambda t} \text{ for some } \eta/2 \leq \varepsilon \leq 1) &=: \mathbb{P}_6.
\end{aligned}$$

From previous calculations, the three first terms in the above display are less than  $\exp(-\phi(\eta)(\sqrt{x\lambda t} - x\lambda))$  for some positive function  $\phi$ . To see why:

- We bound  $\mathbb{P}_3$  with Lemma 15 again (recall  $\text{So}_{\eta x/2}$  is a Poisson random variable);
- The term  $\mathbb{P}_4$  is bounded thanks to Lemma 10 (recall also (9));
- We bound  $\mathbb{P}_5$  with Lemma 16 (recall that  $\text{Si}_t^{(\bar{p})}$  is a Binomial).

To conclude the proof it remains to bound  $\mathbb{P}_6$ . Let  $K$  be an integer larger than  $144/\eta^3$ , by definition of  $L_{<}^*$  we have for every  $1 \leq k \leq \lceil xK \rceil$  and every  $\varepsilon \in [\frac{k}{K}, \frac{k+1}{K}]$

$$L_{<}^* \leq L_{<, k/K}^* + \text{card}(\text{So}_x^{(\bar{\alpha})} \cap [\frac{k}{K}, \frac{k+1}{K}]).$$

Thus

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{\eta/2 \leq \varepsilon \leq 1} \left\{L_{<}^* > 2\sqrt{x\lambda t} - x\lambda - \frac{1}{2}\delta(\eta/3)\sqrt{x\lambda t}\right\}\right) \\
&\leq \sum_{k \geq \lceil \eta K/2 \rceil} \mathbb{P}\left(L_{<, k/K}^* > 2\sqrt{x\lambda t} - x\lambda - \delta(\eta/3)\sqrt{x\lambda t}\right) \\
&+ \sum_{k \geq \lceil \eta K/2 \rceil} \mathbb{P}\left(\text{card}(\text{So}_x^{(\bar{\alpha})} \cap [\frac{k}{K}, \frac{k+1}{K}]) > \frac{1}{2}\delta(\eta/3)\sqrt{x\lambda t}\right) \\
&\leq \sum_{k \geq \lceil \eta K/2 \rceil} \mathbb{P}\left(L_{<, k/K}^* > 2\sqrt{x\lambda t} - x\lambda - \delta(k/K)\sqrt{x\lambda t}\right) \quad (\text{since } K > \frac{144}{\eta^3} > \frac{6}{\eta} \text{ and } \delta \text{ is increasing}) \\
&+ \sum_{k \geq \lceil \eta K/2 \rceil} \mathbb{P}\left(\text{card}(\text{So}_x^{(\bar{\alpha})} \cap [\frac{k}{K}, \frac{k+1}{K}]) > \frac{1}{2}\delta(\eta/3)\sqrt{x\lambda t}\right) \\
&\leq \sum_{k \geq \lceil \eta K/2 \rceil} \exp(-h(k/K)(\sqrt{x\lambda t} - x\lambda)) \quad (\text{using (25)}) \\
&+ K \times \mathbb{P}\left(\text{Poisson}(\bar{\alpha}/K) > \frac{1}{2}\delta(\eta/3)\sqrt{x\lambda t}\right) \\
&\leq K \exp(-h(\eta/3)(\sqrt{x\lambda t} - x\lambda)) \\
&+ K \times \mathbb{P}\left(\text{Poisson}(\bar{\alpha}/K) > \frac{1}{2}\delta(\eta/3)\sqrt{x\lambda t}\right). \tag{26}
\end{aligned}$$

We finally bound the last display. First recall from our notation that

$$\bar{\alpha} < \sqrt{t\lambda/x}, \quad x \geq 1, \quad \delta(\varepsilon) = 2 - \varepsilon - 2\sqrt{1 - \varepsilon} \geq \varepsilon^2/4.$$



Then:

$$\begin{aligned}
\mathbb{P}\left(\text{Poisson}(\bar{\alpha}/K) > \frac{1}{2}\delta(\eta/3)\sqrt{x\lambda t}\right) &= \mathbb{P}\left(\text{Poisson}(\bar{\alpha}/K) > \bar{\alpha}/K - \bar{\alpha}/K + \frac{1}{2}\delta(\eta/3)\sqrt{x\lambda t}\right) \\
&\leq \mathbb{P}\left(\text{Poisson}(\bar{\alpha}/K) > \bar{\alpha}/K + \sqrt{x\lambda t}\left(-\frac{1}{xK} + \frac{1}{2}\delta(\eta/3)\right)\right) \\
&\leq \mathbb{P}\left(\text{Poisson}(\bar{\alpha}/K) > \bar{\alpha}/K + \sqrt{x\lambda t}\left(-\frac{\eta^3}{144} + \frac{\eta^2}{72}\right)\right).
\end{aligned} \tag{27}$$

We can find a positive function  $\varphi$  such that (26) and (27) are both less than  $\frac{144}{\eta}e^{-\varphi(\eta)(\sqrt{xt\lambda}-x\lambda)}$ . We then choose a positive function  $\psi$  such that

$$\min\left\{1, \frac{288}{\eta}e^{-\varphi(\eta)(\sqrt{xt\lambda}-x\lambda)} + 3e^{-\phi(\eta)(\sqrt{x\lambda t}-x\lambda)}\right\} \leq 2e^{-\psi(\eta)(\sqrt{xt\lambda}-x\lambda)}$$

and thus  $\mathbb{P}(\text{sources}(\mathcal{P}) \geq \eta\sqrt{x\lambda t}) \leq \exp(-\psi(\eta)(\sqrt{xt\lambda}-x\lambda))$ . With minor modifications one proves the same bound for sinks (possibly by changing  $\psi$ ):  $\mathbb{P}(\text{sinks}(\mathcal{P}) \geq \eta\sqrt{x\lambda t}) \leq \exp(-\psi(\eta)(\sqrt{xt\lambda}-x\lambda))$  and Lemma 11 is proved.  $\square$

We can conclude the proof of the lower bound in Theorem 9. Let us write

$$L_{<}(t) \geq \mathcal{L}_{=<}(\Pi_{x,t}^{(\lambda)} \cup \text{So}_x^{(\bar{\alpha})} \cup \text{Si}_t^{(\bar{p})}) - \text{sources}(\mathcal{P}) - \text{sinks}(\mathcal{P}),$$

we bound the right-hand side using Lemmas 10 and 11.  $\square$

## 4 Proof of Theorem 1 when $k_n \rightarrow +\infty$ : de-Poissonization

In order to conclude the proof of Theorem 1 it remains to de-Poissonize Theorem 9. We need a few notation. For any integers  $i_1, \dots, i_n$  let  $\mathcal{S}_{i_1, \dots, i_n}$  be the random set of points given by  $i_\ell$  uniform points on each horizontal line:

$$\mathcal{S}_{i_1, \dots, i_n} = \cup_{\ell=1}^n \cup_{r=1}^{i_\ell} \{U_{\ell,r}\} \times \{\ell\},$$

where  $(U_{\ell,r})_{\ell,r}$  is an array of i.i.d. uniform random variables in  $[0, 1]$ . Set also  $e_{i_1, \dots, i_n} = \mathbb{E}[\mathcal{L}_{<}(\mathcal{S}_{i_1, \dots, i_n})]$ . By uniformity of  $U$ 's we have the identity  $\mathbb{E}[\mathcal{L}_{<}(S_{k;n})] = e_{k, \dots, k}$  and therefore our problem reduces to estimating  $e_{k, \dots, k}$ . On the other hand if  $X_1, \dots, X_n$  are i.i.d. Poisson random variables with mean  $k$  then

$$\mathbb{E}[e_{X_1, \dots, X_n}] = \mathbb{E}\left[\mathcal{L}_{<}(\Pi_{nk_n, n}^{(1/n)})\right] = 2\sqrt{nk_n} - k_n + o(\sqrt{nk_n}). \tag{28}$$

The last equality is obtained by combining Theorem 9 for

$$x = nk_n, \quad t = n, \quad \lambda_n = \frac{1}{n}$$

with the trivial bound  $\mathcal{L}_{<}(\Pi_{nk_n, n}^{(1/n)}) \leq n$ . In order to exploit (28) we need the following smoothness estimate.

**Lemma 12.** *For every  $i_1, \dots, i_n$  and  $j_1, \dots, j_n$*

$$|e_{i_1, \dots, i_n} - e_{j_1, \dots, j_n}| \leq 6\sqrt{\sum_{\ell=1}^n |i_\ell - j_\ell|}.$$

*Proof.* Let  $\mathcal{S} = \mathcal{S}_{i_1, \dots, i_n}$  be as above. If we replace in  $\mathcal{S}$  the  $y$ -coordinate of each point of the form  $(x, \ell)$  by a new  $y$ -coordinate uniform in the interval  $(\ell, \ell + 1)$  (independent from anything else) then this defines a uniform permutation  $\sigma_{i_1 + \dots + i_n}$  of size  $i_1 + \dots + i_n$ . The longest increasing subsequence in  $\mathcal{S}$  is mapped onto an increasing subsequence in  $\sigma_{i_1 + \dots + i_n}$  and thus this construction shows the stochastic domination  $\mathcal{L}_{<}(\mathcal{S}_{i_1, \dots, i_n}) \preceq \mathcal{L}_{<}(\sigma_{i_1 + \dots + i_n})$ . Thus for every  $i_1, \dots, i_n$ ,

$$e_{i_1, \dots, i_n} \leq \mathbb{E}[\mathcal{L}_{<}(\sigma_{i_1 + \dots + i_n})] \leq 6\sqrt{i_1 + \dots + i_n}. \quad (29)$$

(The second inequality follows for example from [Ste97, Lemma 1.4.1].) Besides, consider for two  $n$ -tuples  $i_1, \dots, i_n$  and  $j_1, \dots, j_n$  two independent sets of points  $\mathcal{S}_{i_1, \dots, i_n}, \tilde{\mathcal{S}}_{j_1, \dots, j_n}$  then

$$\mathcal{L}_{<}(\mathcal{S}_{i_1, \dots, i_n}) \leq \mathcal{L}_{<}(\mathcal{S}_{i_1, \dots, i_n} \cup \tilde{\mathcal{S}}_{j_1, \dots, j_n}) \leq \mathcal{L}_{<}(\mathcal{S}_{i_1, \dots, i_n}) + \mathcal{L}_{<}(\tilde{\mathcal{S}}_{j_1, \dots, j_n}).$$

This proves that

$$e_{i_1, \dots, i_n} \leq e_{i_1 + j_1, \dots, i_n + j_n} \leq e_{i_1, \dots, i_n} + e_{j_1, \dots, j_n}.$$

(In particular  $(i_1, \dots, i_n) \mapsto e_{i_1, \dots, i_n}$  is non-decreasing with respect to any of its coordinate.) Therefore

$$\begin{aligned} e_{i_1, \dots, i_n} &\leq e_{(i_1 - j_1)^+, \dots, (i_n - j_n)^+} + e_{j_1 - (i_1 - j_1)^-, \dots, j_n - (i_n - j_n)^-} \\ &\leq e_{|i_1 - j_1|, \dots, |i_n - j_n|} + e_{j_1, \dots, j_n}. \end{aligned}$$

By switching the role of  $i$ 's and  $j$ 's:

$$|e_{i_1, \dots, i_n} - e_{j_1, \dots, j_n}| \leq e_{|i_1 - j_1|, \dots, |i_n - j_n|} \leq 6\sqrt{\sum_{\ell=1}^n |i_\ell - j_\ell|},$$

using (29). □

*Proof of Theorem 1 for any sequence  $(k_n) \rightarrow +\infty$ .* Using smoothness we write

$$|e_{k, \dots, k} - \mathbb{E}[e_{X_1, \dots, X_n}]| \leq \mathbb{E}[|e_{k, \dots, k} - e_{X_1, \dots, X_n}|] \leq 6 \times \mathbb{E} \left[ \left( \sum_{\ell=1}^n |X_\ell - k| \right)^{1/2} \right]. \quad (30)$$

Using twice the Cauchy-Schwarz inequality:

$$\begin{aligned} \mathbb{E} \left[ \left( \sum_{\ell=1}^n |X_\ell - k| \right)^{1/2} \right] &\leq \sqrt{\mathbb{E} \left[ \sum_{\ell=1}^n |X_\ell - k| \right]} \\ &\leq \sqrt{n \mathbb{E}[|X_1 - k|]} \\ &\leq \sqrt{n \mathbb{E}[|X_1 - k|^2]^{1/2}} = \sqrt{n \sqrt{\text{Var}(X_1)}} = \sqrt{n \sqrt{k}}. \end{aligned}$$

If  $k = k_n \rightarrow \infty$  then the last display is a  $o(\sqrt{nk_n})$  and eq.(30) and (28) show that

$$e_{k, \dots, k} = \mathbb{E}[\mathcal{L}_{<}(S_{k;n})] = 2\sqrt{nk_n} - k_n + o(\sqrt{nk_n}).$$

□

## 5 Proof of Theorem 2

### 5.1 Proof for large $(k_n)$

We now prove Theorem 2 for a *large* sequence  $(k_n)$ . We say that  $(k_n)$  is *large* if

$$n^2 k_n \exp(-(k_n)^\alpha) = o(\sqrt{n k_n}) \quad (31)$$

for some  $\alpha \in (0, 1)$ . Recall that  $k_n = \log n$  is not large while  $k_n = (\log n)^{1+\varepsilon}$  is large.

We first observe that de-Poissonization cannot be applied as in the previous section. We lack smoothness as, for instance,  $\mathbb{E}[\mathcal{L}_\leq(\mathcal{S}_{i_1,0,0,\dots,0})] = i_1 \neq \mathcal{O}(\sqrt{\sum i_\ell})$ . The strategy is to apply Theorem 9 with

$$x = n k_n, \quad t = n, \quad \lambda_n \approx \frac{1}{n}.$$

(The exact value of  $\lambda_n$  will be different for the proofs of the lower and upper bounds.)

**Proof of the upper bound of (2) for large  $(k_n)$ .**

Choose  $\alpha$  such that  $n^2 k_n \exp(-k_n^\alpha) = o(\sqrt{n k_n})$ . Put

$$\lambda_n = \frac{1}{n} + \frac{\delta_n}{n}, \quad \text{with } \delta_n = k_n^{-(1-\alpha)/2}.$$

Let  $E_n^{\lambda_n}$  be the event

$$E_n^{\lambda_n} = \left\{ \text{there are at least } k_n \text{ points in each row of } \Pi_{n k_n, n}^{(\lambda_n)} \right\}.$$

The event  $E_n$  occurs with large probability. Indeed,

$$\begin{aligned} 1 - \mathbb{P}(E_n^{\lambda_n}) &\leq n \mathbb{P}(\text{Poisson}(n k_n \lambda_n) \leq k_n) \\ &\leq n \mathbb{P}(\text{Poisson}(n k_n \lambda_n) \leq n k_n \lambda_n + k_n - n k_n \lambda_n) \\ &\leq n \mathbb{P}(\text{Poisson}(n k_n \lambda_n) \leq n k_n \lambda_n - k_n \delta_n) \\ &\leq n \exp\left(-\frac{k_n^2 \delta_n^2}{4 n k_n \lambda_n}\right) \leq n \exp\left(-\frac{1}{8} k_n \delta_n^2\right) = n \exp\left(-\frac{1}{8} k_n^\alpha\right). \end{aligned} \quad (32)$$

At the last line we used Lemma 15. The latter probability tends to 0 as  $(k_n)$  is large.

**Lemma 13.** *Random sets  $S_{k_n;n}$  and  $\Pi_{n k_n, n}^{(\lambda_n)}$  can be defined on the same probability space in such a way that*

$$\mathcal{L}_\leq(S_{k_n;n}) \leq \mathcal{L}_\leq(\Pi_{n k_n, n}^{(\lambda_n)}) + n k_n (1 - \mathbf{1}_{E_n^{\lambda_n}}). \quad (33)$$

*Proof of Lemma 13.* Draw a sample of  $\Pi_{n k_n, n}^{(\lambda_n)}$  and let  $\tilde{\Pi}_{n k_n, n}^{(\lambda_n)}$  be the subset of  $\Pi_{n k_n, n}^{(\lambda_n)}$  obtained by keeping only the  $k_n$  leftmost points in each row. If  $E_n^{\lambda_n}$  occurs then the relative orders of points in  $\tilde{\Pi}_{n k_n, n}^{(\lambda_n)}$  corresponds to a uniform  $k_n$ -multiset permutation. If  $E_n^{\lambda_n}$  does not hold we bound  $\mathcal{L}_\leq(S_{k_n;n})$  by the worst case  $n k_n$ .  $\square$

Taking expectations in (33) and using the upper bound (15) yields

$$\mathbb{E}[\mathcal{L}_\leq(S_{k_n;n})] \leq 2\sqrt{n k_n (1 + \delta_n)} + k_n (1 + \delta_n) + n^2 k_n \exp\left(-\frac{1}{8} k_n^\alpha\right),$$

hence the upper bound in (2).

**Proof of the lower bound of (2) for large  $(k_n)$ .** Choose now  $\lambda_n = \frac{1}{n}(1 - \delta_n)$  with  $\delta_n = k_n^{-(1-\alpha)/2}$ . Let  $F_n$  be the event

$$F_n^{\lambda_n} = \left\{ \text{at most } k_n \text{ points in each row of } \Pi_{nk_n, n}^{(\lambda_n)} \right\}.$$

The event  $F_n^{\lambda_n}$  occurs with large probability. Indeed

$$1 - \mathbb{P}(F_n^{\lambda_n}) \leq n\mathbb{P}(\text{Poisson}(nk_n\lambda_n) \geq k_n) \leq n \exp\left(-\frac{1}{8}k_n^\alpha\right),$$

which tends to zero. Random sets  $S_{k_n; n}$  and  $\Pi_{nk_n, n}^{(\lambda_n)}$  can be defined on the same probability space in such a way that

$$\mathcal{L}_{\leq}(S_{k_n; n}) \geq \mathcal{L}_{\leq}(\Pi_{nk_n, n}^{(\lambda_n)}) \mathbf{1}_{F_n^{\lambda_n}}.$$

Therefore

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}_{\leq}(S_{k_n; n}) < (2\sqrt{nk_n(1 - \delta_n)} + k_n(1 - \delta_n))(1 - \varepsilon)\right) &\leq \\ \mathbb{P}\left(\mathcal{L}_{\leq}(\Pi_{nk_n, n}^{(\lambda_n)}) < (2\sqrt{nk_n(1 - \delta_n)} + k_n(1 - \delta_n))(1 - \varepsilon)\right) &+ \mathbb{P}(\text{not } F_n^{\lambda_n}). \end{aligned}$$

and we conclude with (19).

## 5.2 The gap between small and large $(k_n)$ : Conclusion of the proof of Theorem 2

After I circulated a preliminary version of this article, Valentin Féray came up with a simple argument for bridging the gap between small and large  $(k_n)$ . This allows to prove Theorem 2 for an arbitrary sequence  $(k_n)$ , I reproduce his argument here with his permission.

**Lemma 14.** *Let  $n, k, A$  be positive integers. Two random uniform multiset permutations  $\tilde{S}_{kA; \lfloor n/A \rfloor}$  and  $S_{k; n}$  can be built on the same probability space in such a way that*

$$\mathcal{L}_{\leq}(S_{k; n}) \leq \mathcal{L}_{\leq}(\tilde{S}_{kA; \lfloor n/A \rfloor}) + kA.$$

*Proof of Lemma 14.* Draw  $S_{k; n}$  uniformly at random, the idea is to group all points of  $S_{k; n}$  whose height is between 1 and  $A$ , to group all points whose height is between  $A + 1$  and  $2A$ , and so on.

Formally, denote by  $1 \leq i_1 < i_2 < \dots < i_{kA \lfloor n/A \rfloor}$  the indices such that  $1 \leq i_\ell \leq \lfloor n/A \rfloor$  for every  $\ell$  (see Fig.5). For  $1 \leq \ell \leq kA \lfloor n/A \rfloor$  put

$$\tilde{S}(\ell) = \lceil S(i_\ell)/k \rceil.$$

The word  $\tilde{S}$  is a uniform  $kA$ -multiset permutation of size  $\lfloor n/A \rfloor$ . A longest non-decreasing subsequence in  $S$  is mapped onto a non-decreasing subsequence in  $\tilde{S}$ , except maybe some points with height  $> A \lfloor n/A \rfloor$  (there are no more than  $kA$  such points). This shows the Lemma.  $\square$

We conclude the proof of Theorem 2 by an estimation of  $\mathbb{E}[\mathcal{L}_{\leq}(S_{k_n; n})]$  in the case where there are infinitely many  $k_n$ 's such that, say,  $(\log n)^{3/4} \leq k_n \leq (\log n)^{5/4}$ . For the lower bound the job is already done by Theorem 1 since

$$\mathbb{E}[\mathcal{L}_{\leq}(S_{k_n; n})] \geq \mathbb{E}[\mathcal{L}_{<}(S_{k_n; n})] = 2\sqrt{nk_n} - k_n + o(\sqrt{nk_n}),$$

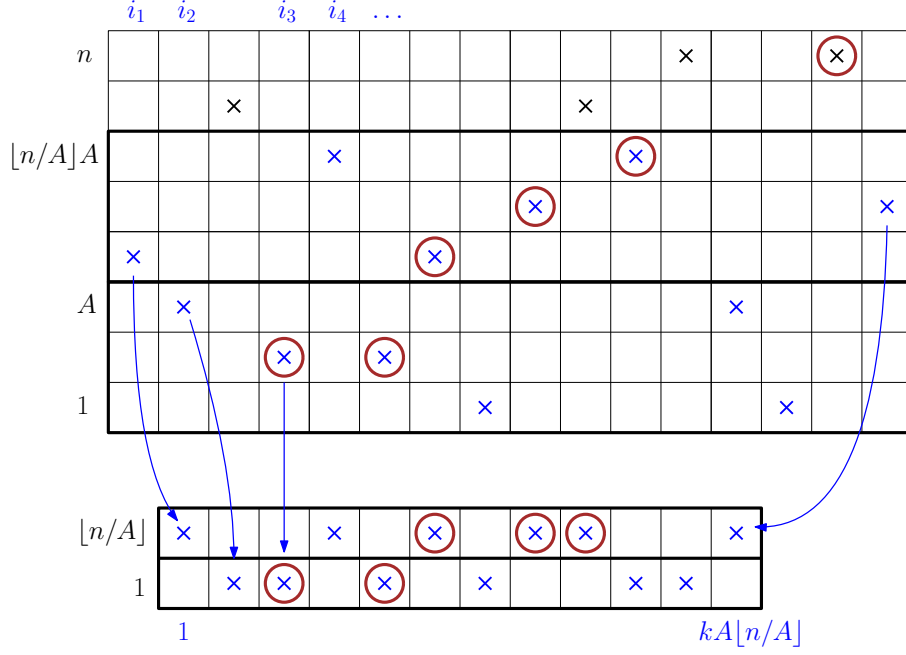


Figure 5: Illustration of the notation of Lemma 14. Top: the multiset permutation  $S_{k;n}$ . Bottom: the corresponding  $\tilde{S}$ . The longest non-decreasing subsequence in  $S_{k;n}$  (circled points) is mapped onto a non-decreasing subsequence in  $\tilde{S}$ , except one point with height  $> A\lfloor n/A \rfloor$ .

which is of course also  $2\sqrt{nk_n} + o(nk_n)$  for this range of  $(k_n)$ . For the upper bound take  $A = \lfloor \log n \rfloor$  in Lemma 14:

$$\mathbb{E}[\mathcal{L}_{\leq}(S_{k_n;n})] \leq \mathbb{E}[\mathcal{L}_{\leq}(S_{k_n \log n; \lfloor n/\lfloor \log n \rfloor \rfloor})] + k_n \log n \quad (34)$$

and we can apply the large case since

$$(n/\log n)^2 k_n \log n \exp(-(k_n \log n)^\alpha) = o(k_n \log n \times \lfloor n/\lfloor \log n \rfloor \rfloor).$$

Thus the right-hand side of (34) is also  $2\sqrt{nk_n} + o(\sqrt{nk_n})$ .

## 6 Conclusion: Proof of Proposition 3

In this short section we give the arguments needed to enhance estimates in expectation into convergences in probability. We have to prove that for every  $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}\left(L_{<}(S_{k_n;n}) > (2\sqrt{nk_n} - k_n)(1 + \varepsilon)\right) &\rightarrow 0, & \mathbb{P}\left(L_{<}(S_{k_n;n}) < (2\sqrt{nk_n} - k_n)(1 - \varepsilon)\right) &\rightarrow 0 \\ \mathbb{P}\left(L_{\leq}(S_{k_n;n}) > (2\sqrt{nk_n} + k_n)(1 + \varepsilon)\right) &\rightarrow 0, & \mathbb{P}\left(L_{\leq}(S_{k_n;n}) < (2\sqrt{nk_n} + k_n)(1 - \varepsilon)\right) &\rightarrow 0 \end{aligned}$$

We only write the details for the top-left case, as the three other ones are almost identical.

The case where  $(k_n)$  is small has been proved in Section 2 so it remains to prove the case

where  $(k_n)$  is large. We reuse the event  $E_n^{\lambda_n}$  introduced in Section 5.1.

$$\begin{aligned}
\mathbb{P}\left(L_{<}(S_{k_n;n}) > (2\sqrt{nk_n} - k_n)(1 + \varepsilon)\right) &\leq \mathbb{P}\left(E_n^{\lambda_n} \text{ does not occur}\right) \\
&+ \mathbb{P}\left(\mathcal{L}_{<}\left(\Pi_{nk_n,n}^{(\lambda_n)}\right) > (1 + \delta_n)(2\sqrt{nk_n} - k_n)\frac{1 + \varepsilon}{1 + \delta_n}\right) \\
&\leq n \exp\left(-\frac{1}{8}k_n^\alpha\right) \quad (\text{recall (32)}) \\
&+ \mathbb{P}\left(\mathcal{L}_{<}\left(\Pi_{nk_n,n}^{(\lambda_n)}\right) > (2\sqrt{nk_n(1 + \delta_n)} - k_n(1 + \delta_n))\frac{1 + \varepsilon}{1 + \delta_n}\right) \\
&\leq n \exp\left(-\frac{1}{8}k_n^\alpha\right) + \exp(-\tilde{g}(\varepsilon/2)(\sqrt{nk_n} - k_n)),
\end{aligned}$$

for large enough  $n$  and for some positive  $\tilde{g}$ , using (16). This tends to zero as desired.

The lower bound for  $L_{<}(S_{k_n;n})$  is proved in the same way. For the convergence of  $L_{\leq}(S_{k_n;n})$  we reuse the event  $F_n^{\lambda_n}$  with  $\lambda_n = \frac{1}{n}(1 + \log(n))$ .

## A Useful tail inequalities

We collect here for convenience some (non-optimal) tail inequalities.

**Lemma 15** ((See Chap.2 in [JLR00])). *Let  $\text{Poisson}(\lambda)$  be a Poisson random variable with mean  $\lambda$ . For every  $A > 0$*

$$\begin{aligned}
\mathbb{P}(\text{Poisson}(\lambda) \leq \lambda - A) &\leq \exp(-A^2/4\lambda), \\
\mathbb{P}(\text{Poisson}(\lambda) \geq \lambda + A) &\leq \exp(-A^2/4\lambda).
\end{aligned}$$

**Lemma 16** (Th.2.1 in [JLR00]). *Let  $\text{Binomial}(n, p)$  be a Binomial random variable with parameters  $(n, p)$ . For  $0 < \varepsilon < 1$ ,*

$$\begin{aligned}
\mathbb{P}(\text{Binomial}(n, p) \leq np - \varepsilon np) &\leq \exp(-\varepsilon^2 np/2), \\
\mathbb{P}(\text{Binomial}(n, p) \geq np + \varepsilon np) &\leq \exp(-\varepsilon^2 np/3).
\end{aligned}$$

**Lemma 17.** *Fix  $\alpha \in (0, 1)$  and let  $\mathcal{G}_1^{(\alpha)}, \dots, \mathcal{G}_k^{(\alpha)}$  be i.i.d. random variables with distribution  $\text{Geometric}_{\geq 0}(1 - \alpha)$ . Then  $\mathbb{E}[\mathcal{G}_1^{(\alpha)}] = \frac{\alpha}{1 - \alpha}$  and for every  $0 < \varepsilon < 1$ ,*

$$\begin{aligned}
\mathbb{P}\left(\mathcal{G}_1^{(\alpha)} + \dots + \mathcal{G}_k^{(\alpha)} \geq (1 + \varepsilon)k\frac{\alpha}{1 - \alpha}\right) &\leq \exp(-\varepsilon^2 k\alpha/20), \\
\mathbb{P}\left(\mathcal{G}_1^{(\alpha)} + \dots + \mathcal{G}_k^{(\alpha)} \leq (1 - \varepsilon)k\frac{\alpha}{1 - \alpha}\right) &\leq \exp(-\varepsilon^2 k\alpha/20).
\end{aligned}$$

*Proof of Lemma 17.* We will use the two inequalities

$$\exp(z) \leq 1 + z + z^2 \text{ for } |z| < 1, \quad \frac{1}{1 - u} \leq \exp(u + u^2) \text{ for } |u| < 1/2.$$

Fix  $\lambda$  such that  $|\lambda| < \min\{1, (1 - \alpha)/4\alpha\}$  so that  $\frac{\alpha}{1 - \alpha}|\lambda + \lambda^2| < 1/2$ :

$$\begin{aligned}
\mathbb{E}[e^{\lambda(\mathcal{G}_1^{(\alpha)} - \frac{\alpha}{1 - \alpha})}] &= \frac{(1 - \alpha)}{1 - \alpha e^\lambda} e^{-\lambda \frac{\alpha}{1 - \alpha}} = \frac{1}{1 - \frac{\alpha}{1 - \alpha}(e^\lambda - 1)} e^{-\lambda \frac{\alpha}{1 - \alpha}} \\
&\leq \frac{1}{1 - \frac{\alpha}{1 - \alpha}(\lambda + \lambda^2)} e^{-\lambda \frac{\alpha}{1 - \alpha}} \\
&\leq \exp\left(\frac{\alpha}{1 - \alpha}(\lambda + \lambda^2) + \left(\frac{\alpha}{1 - \alpha}\right)^2 (\lambda + \lambda^2)^2 - \lambda \frac{\alpha}{1 - \alpha}\right) \\
&\leq \exp\left(\frac{\alpha}{(1 - \alpha)^2} \lambda^2 (2 + \lambda^2 + 2\lambda)\right) \leq \exp\left(5\lambda^2 \frac{\alpha}{(1 - \alpha)^2}\right).
\end{aligned}$$

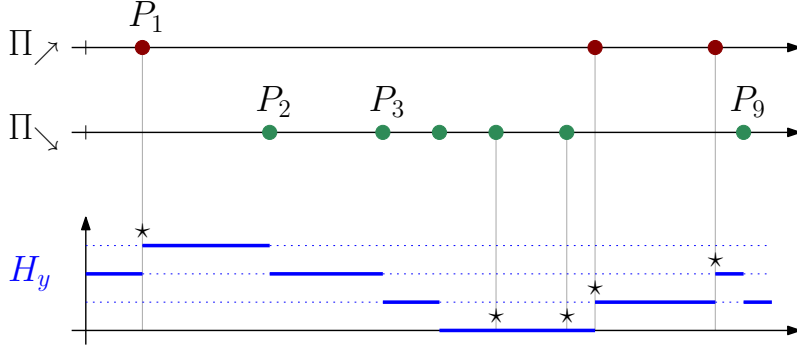


Figure 6: Notation of Lemma 18. Points of  $\bar{\Pi}$  are depicted with  $\star$ 's.

Thus, for every  $|\lambda| < \frac{1}{\beta} := \min \{1, (1 - \alpha)/4\alpha\}$  it holds that  $\mathbb{E}[e^{\lambda(\sum_{i=1}^k \mathcal{G}_i^{(\alpha)} - k \frac{\alpha}{1-\alpha})}] \leq \exp\left(\frac{\nu^2 \lambda^2}{2}\right)$  where  $\nu^2 := 10k\alpha/(1 - \alpha)^2$ .

This says that for every  $k \geq 1$  the random variable  $\mathcal{G}_1^{(\alpha)} + \dots + \mathcal{G}_k^{(\alpha)}$  is subexponential and the Chernov method applies (use *e.g.* [Wai19, Prop.2.9] with  $t = \varepsilon k\alpha/(1 - \alpha)$ ):

$$\begin{aligned} \mathbb{P}\left(\mathcal{G}_1^{(\alpha)} + \dots + \mathcal{G}_k^{(\alpha)} \geq (1 + \varepsilon)k \frac{\alpha}{1 - \alpha}\right) &\leq \exp\left(-\frac{t^2}{2\nu^2}\right) = \exp\left(-\varepsilon^2 k^2 \frac{\alpha^2}{(1 - \alpha)^2} \times \frac{(1 - \alpha)^2}{2 \times 10k\alpha}\right) \\ &= \exp\left(-\varepsilon^2 k\alpha/20\right), \end{aligned}$$

as long as

$$\varepsilon k \frac{\alpha}{1 - \alpha} \leq \frac{\nu^2}{\beta} = 10k \frac{\alpha}{(1 - \alpha)^2} \min \{1, (1 - \alpha)/4\alpha\}$$

which is always the case if  $\varepsilon < 1$ . The similar inequality holds for the left-tail bound (see [Wai19, Prop.2.9] again).  $\square$

## B An invariance property for the M/M/1 queue

To conclude we state and prove the very simple property of the recurrent M/M/1 queue which allows to prove stationarity in Lemma 6. It is very close to Burke's property of the discrete HAD process [FM06].

Let  $\beta > \lambda > 0$  be fixed parameters. Consider two independent homogeneous Poisson Point Process (PPP)  $\Pi_{\nearrow}, \Pi_{\searrow}$  over  $(0, +\infty)$  with respective intensities  $\lambda, \beta$ . Let  $(H_y)_{y \geq 0}$  be the queue whose '+1' steps (customer arrivals) are given by  $\Pi_{\nearrow}$  and '-1' steps (service times) are given by  $\Pi_{\searrow}$  and whose initial distribution  $H_0$  is drawn (independently from  $\Pi_{\nearrow}, \Pi_{\searrow}$ ) according to a  $\text{Geometric}_{\geq 0}(1 - \beta^*)$  with  $\beta^* = \lambda/\beta$ .

Let  $\Pi_0$  be the point process given by *unused* service times:

$$\Pi_0 = \{y \in \Pi_{\searrow} \text{ such that } H_y = 0\}.$$

**Lemma 18.** *The process  $\bar{\Pi} := \Pi_{\nearrow} \cup \Pi_0$  is a homogeneous PPP with intensity  $\beta$ .*

*Proof.* (The reader is invited to look at Fig.6 for notation.)

The point process  $\Pi_{\nearrow} \cup \Pi_{\searrow}$  is a homogeneous PPP with intensity  $\lambda + \beta$ , independent from  $H_0$ .

We claim that  $\bar{\Pi}$  is a subset of  $\Pi_{\nearrow} \cup \Pi_{\searrow}$  where each point in  $\Pi_{\nearrow} \cup \Pi_{\searrow}$  is taken independently with probability  $\beta/(\lambda + \beta)$ , it is therefore a homogeneous PPP with intensity  $\beta$ .

We need a few notation in order to prove the claim. Set  $P_0 = 0$  and for  $i \geq 1$  let  $P_i$  be the  $i$ -th point of  $\Pi_{\nearrow} \cup \Pi_{\searrow}$  and let  $(\tilde{H}_i)_{i \geq 0}$  be the discrete-time embedded chain associated to  $H$ , *i.e.*  $\tilde{H}_i = H_{P_i}$  for every  $i$ .

We will prove by induction that for every  $i \geq 1$ :

- the points  $P_i$  belongs to  $\bar{\Pi}$  with probability  $\beta/(\lambda + \beta)$  independently from the events  $\{P_1 \in \bar{\Pi}\}, \dots, \{P_{i-1} \in \bar{\Pi}\}$ ;
- $\tilde{H}_i$  is independent from  $\{P_1 \in \bar{\Pi}\}, \dots, \{P_i \in \bar{\Pi}\}$  and is a  $\text{Geometric}_{\geq 0}(1 - \beta^*)$ .

This implies the claim and proves the Lemma. For the base case:

$$\begin{aligned} \mathbb{P}(P_1 \in \bar{\Pi}, \tilde{H}_1 = k) &= \mathbb{P}(P_1 \in \Pi_{\nearrow}, \tilde{H}_0 = k - 1) \mathbf{1}_{k \geq 1} + \mathbb{P}(P_1 \in \Pi_{\searrow}, \tilde{H}_0 = 0) \mathbf{1}_{k=0}, \\ &= \frac{\lambda}{\lambda + \beta} \times (1 - \beta^*)(\beta^*)^{k-1} \mathbf{1}_{k \geq 1} + \frac{\beta}{\lambda + \beta} \times (1 - \beta^*) \mathbf{1}_{k=0}, \\ &= \frac{\beta}{\lambda + \beta} \times (1 - \beta^*)(\beta^*)^k \quad (\text{recall } \beta\beta^* = \lambda). \end{aligned}$$

More generally let  $E_j$  be one of the two events  $P_j \in \bar{\Pi}/P_j \notin \bar{\Pi}$ .

$$\begin{aligned} \mathbb{P}(P_i \in \bar{\Pi}, \tilde{H}_i = k \mid E_1, \dots, E_{i-1}) &= \mathbb{P}(P_i \in \Pi_{\nearrow}, \tilde{H}_{i-1} = k - 1 \mid E_1, \dots, E_{i-1}) \mathbf{1}_{k \geq 1} \\ &\quad + \mathbb{P}(P_i \in \Pi_{\searrow}, \tilde{H}_{i-1} = 0 \mid E_1, \dots, E_{i-1}) \mathbf{1}_{k=0}, \\ &= \mathbb{P}(P_i \in \Pi_{\nearrow}, \tilde{H}_{i-1} = k - 1) \mathbf{1}_{k \geq 1} \\ &\quad + \mathbb{P}(P_i \in \Pi_{\searrow}, \tilde{H}_{i-1} = 0) \mathbf{1}_{k=0}, \quad (\text{by induction hypothesis}). \\ &= \frac{\beta}{\lambda + \beta} \times (1 - \beta^*)(\beta^*)^k. \end{aligned}$$

□

**Acknowledgements.** This work started as a collaboration with Anne-Laure Basdevant, I would like to thank her very warmly. I am also extremely indebted to Valentin Féray for Lemma 14 and for having enlightened me on the links with [Bia01]. Finally, thanks to the authors of [CDH<sup>+</sup>22] for their stimulating paper and to anonymous referees for their careful readings.

## References

- [AD95] David Aldous and Persi Diaconis. Hammersley's interacting particle process and longest increasing subsequences. *Probability Theory and Related Fields*, 103(2):199–213, 1995.
- [BDJ99] Jinho Baik, Percy Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.
- [BEGG16] Anne-Laure Basdevant, Nathanaël Enriquez, Lucas Gerin, and Jean-Baptiste Gouéré. Discrete Hammersley's lines with sources and sinks. *ALEA Lat. Am. J. Probab. Math. Stat.*, 13:33–52, 2016.
- [Bia01] Philippe Biane. Approximate factorization and concentration for characters of symmetric groups. *Internat. Math. Res. Notices*, (4):179–192, 2001.



- [Boy22] Alexandre Boyer. Chapter 3 (in English) of *Stationnarité bidimensionnelle de modèles aléatoires du plan*, 2022. PhD Thesis, available at <https://tel.archives-ouvertes.fr/tel-03783603/>.
- [CDH<sup>+</sup>22] Alexander Clifton, Bishal Deb, Yifeng Huang, Sam Spiro, and Semin Yoo. Continuously increasing subsequences of random multiset permutations. *Sém. Lothar. Combin.*, 86B:Art. 4, 11, 2022. (Proceedings of FPSAC’22.).
- [CG05] Eric Cator and Piet Groeneboom. Hammersley’s process with sources and sinks. *Annals of Probability*, 33(3):879–903, 2005.
- [CG06] Eric Cator and Piet Groeneboom. Second class particles and cube root asymptotics for Hammersley’s process. *Annals of Probability*, 34(4):1273–1295, 2006.
- [CG19] Federico Ciech and Nicos Georgiou. Order of the variance in the discrete Hammersley process with boundaries. *Journal of Statistical Physics*, 176(3):591–638, 2019.
- [CLRS09] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms*. MIT Press, Cambridge, MA, third edition, 2009.
- [Fer96] Pablo A. Ferrari. Limit theorems for tagged particles. *Markov Process. Related Fields*, 2(1):17–40, 1996.
- [FM06] Pablo A. Ferrari and J. B. Martin. Multi-class processes, dual points and  $M/M/1$  queues. *Markov Process. Related Fields*, 12(2):175–201, 2006.
- [Ham72] John M. Hammersley. A few seedlings of research. In *Proceedings of the 6th Berkeley Symp. Math. Statist. and Probability*, volume 1, pages 345–394, 1972.
- [JŁR00] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
- [Rom15] Dan Romik. *The surprising mathematics of longest increasing subsequences*, volume 4 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, New York, 2015.
- [Sep97] Timo Seppäläinen. Increasing sequences of independent points on the planar lattice. *Annals of Applied Probability*, 7(4):886–898, 1997.
- [Sep98] Timo Seppäläinen. Exact limiting shape for a simplified model of first-passage percolation on the plane. *Annals of Probability*, 26(3):1232–1250, 1998.
- [Ste97] John Michael Steele. *Probability theory and combinatorial optimization*. Society for Industrial and Applied Mathematics (SIAM), 1997.
- [VK77] Anatoly M. Veršik and Sergei V. Kerov. Asymptotics of Plancherel measure of symmetrical group and limit form of Young tables. *Doklady Akademii Nauk SSSR*, 233.6:1024–1027, 1977.
- [Wai19] Martin J. Wainwright. *High-dimensional statistics*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019.

LUCAS GERIN [gerin@cmap.polytechnique.fr](mailto:gerin@cmap.polytechnique.fr)  
CMAP, CNRS, École Polytechnique,  
Institut Polytechnique de Paris,  
Route de Saclay,  
91120 Palaiseau Cedex (France).