Dispersion models on a circle: universal properties and asymptotic results

Jean-François Marckert & Zoé Varin

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

Version en cours de finalisation, ne pas diffuser

Abstract

Consider some masses $m_1, m_2, \ldots$ arriving uniformly at random at some points $u_1, u_2, \ldots$ on the cycle $\mathbb{R}/\mathbb{Z}$ (or on $\mathbb{Z}/n\mathbb{Z}$, in the discrete version). On arrival, each mass undergoes a relaxation phase during which it is dispersed, possibly also at random. This models many physical phenomena (such as the diffusion of liquid in a porous medium), and in the discrete case parking models (related to additive coalescence and hashing with linear probing) in which the cars are allowed to have any random displacement distribution.

The dispersion policies considered ensure that at time k, the total covered region has Lebesgue measure $m_1 + \cdots, m_k$ (or represents a fraction $m_1 + \cdots + m_k$ of $\mathbb{Z}/n\mathbb{Z}$, in the discrete case). Moreover, throughout the relaxation phase of a given mass, the covered domain is non-decreasing, and a single covered connected component grows, continuously, except at the special times at which it merges with another covered connected component.

We show a very general exchangeability property for the sequence of covered connected components. In addition, we show a universal spacing property between these connected components, and a striking general result: if the (u_i) are independent and invariant by rotation, then the number of free (not covered) connected components has a binomial distribution (whose parameters depend only on the total mass of arrived particles), and conditional on the number of connected components, the free connected components have a simple Dirichlet distribution in the continuous case, whatever is the dispersion policy considered and the values of the individual masses.

In a second part of the paper, we study the total cost associated with these models, and make some connections with additive coalescent. We also give an asymptotic representation of the limiting covered space as the number of masses goes to $+\infty$.

1 Introduction

The dynamics we are interested in are quite general, and represent physical systems in which some matter arrives successively at random on a given surface S, and on arrival, undergoes a relaxation phase, during which this matter is dispersed on the surface until it covers a new surface proportional to its initial mass. This type of models is very common in all sciences, and is particularly common in physics, chemistry, biology, computer science, combinatorics, probability theory. At this stage, we can think of two main images to illustrate the phenomenon we want to study:

– We can imagine water drops of any random size (as in the case of rain) falling at random on a structure, and moistening the structure, according to a (possibly random) dispersion dynamic, depending on the external conditions (wind, temperature...). In this case, the space is the structure, and the connected

components (CC) of the moistened area constitute the "occupied domain"; the moistened area can be imagined as having a surface proportional to the total mass of water deposited (or equal to it, up to a change of unity).
– We can imagine some cars (all of the unit size), arriving one after the other in a parking lot. The ith car, when it arrives, chooses a place c_i at random, and parks in the nearest closest available place. In this case, the problem has a discrete nature: the space is the parking lot, and the occupied domain is the set of occupied places.

In order to fully define our models, 4 main characteristics need to be defined:
(a) the space/surface on which the matter arrives (it will be one-dimensional, discrete or continuous),
(b) the masses distribution; they can be random or not, of discrete nature or not,
(c) the arrival positions of the masses: they may be of discrete or continuous nature, but they will be random, and uniform on a well-chosen space,
(d) and the "dispersion" algorithm, which will be the dynamics, random or not, for a mass m_i arriving at position u_i, according to which it will cover a part of the space with size m_i, and remain there eventually.

We will see that our main result holds for very weak assumption on this dispersion/covering policies, while the mass model is absolutely crucial.

1.1 Content of the paper and organization

In Section 1.4 we define precisely the class of valid continuous dispersion models (CDM); continuous qualifies the space which is $\mathbb{R}/\mathbb{Z}$, while the masses can be discrete or continuous, random or not. A model will be said to be valid if it satisfies several conditions, including the fact that its definition must be invariant by rotation, and the fact that during the relaxation phase, the occupied CC which contains the point where the last mass arrival occurs, must evolve independently of the rest of the configuration (except at the collision time with other occupied components, at which, it merges with them).

The main quantities of interest, which are the occupied space $\mathbf{O}^{(k)}$ after dispersion of k masses, and the free space $\mathbf{F}^{(k)}$ are introduced. Each of them are union of intervals.

A list of examples of valid CDM is given in Section 1.5. The right diffusion at constant speed is an important model, which will play a role all along the paper: in word, when a mass m arrives at u, it is pushed to the right, and is eroded while passing on a free space at constant speed (but not eroded on occupied space). When the quantity of free space f on which it has been pushed reached the Lebesgue measure m, no mass remains to be eroded, and f becomes occupied. Other models as the infinitesimal particle like diffusion, the range of the Brownian path, and the short-sighted jam-spreader model allows to feel the extent of the class of valid CDM.

In Section 1.6, we present the main universal results on CDM. The power of these principles are easier to understand when the mass $m_0, \cdots, m_{k-1}$ are not random, but not necessarily equals (and they will be used after, for random masses).

Arguably the main theorems of the paper, shows that after k-masses $m_0, \cdots, m_{k-1}$ have been dispersed:

- the joint laws of $\left(\mathbf{O}^{(k)}, \mathbf{F}^{(k)}\right)$ does not depend on the mass arrival order (it has same distribution for $m_{\sigma_0}, \cdots, m_{\sigma_{k-1}}$ for any permutation σ) (Theorem 1.5),

- the law of $\left(\mathbf{O}^{(k)}, \mathbf{F}^{(k)}\right)$ is universal for all valid CDM chosen (Theorem 1.5). Hence, it suffices to analyze one of them to analyze all of them (and the right diffusion with constant speed is probably the simplest one),

- The law of the number $\mathbf{N}_k$ of CC of $\mathbf{F}^{(k)}$ (or of $\mathbf{O}^{(k)}$), depends only on k and on the sum of masses $\sum_{j=0}^{k-1} m_j$: Hence it does not depend on the chosen CDM, nor on the individual masses (m_j), nor on their order: $\mathbf{N}_k - 1$ follows the binomial distribution with parameters $(k-1, \sum_{j=0}^{k-1} m_j)$ (Theorem 1.5). Given $\mathbf{N}_k = n$, the joint distribution of the free interval length is also explicit: it is a Dirichlet distribution.

- The distribution of the lengths of the CC of $\mathbf{O}^{(k)}$ still does not depend on the valid CDM considered, but depends on the masses, and the formula remains explicit (Theorem 1.8).

- The law of the process $(S(t), 0 \le t \le k)$ where $S(t)$ is the multiset of sizes of the elements of $\mathbf{O}^{(t)}$ at time t, is the same for all valid CDM.

In Section 2, we provide similar results for valid discrete dispersion model (DDM), which are diffusion models on the discrete circle $\mathbb{Z}/n\mathbb{Z}$, in which, this time, the masses are integers and arrive at integer positions (and are diffused in a discrete way). In fact, to make fit the discrete case in the continuous settings (and then allow comparisons), we will prefer to work on a isomorphic version of $\mathbb{Z}/n\mathbb{Z}$ embedded in the initial circle $\mathscr{C} = \mathbb{R}/\mathbb{Z}$, so we work on the set $\mathscr{C}_n := \{k/n, 0 \le k \le n-1\}$ (as a subgroup of $\mathbb{R}/\mathbb{Z}$), and will consider that masses arrive uniformly on $\mathscr{C}_n$. The results obtained are similar to the continuous dispersion model cases (concerning the universality of the law of $\left(\mathbf{O}^{(k)}, \mathbf{F}^{(k)}\right)$ for all valid DDM considered, Theorem 2.2, or the fact that we can permute the masses without changing the law of $\left(\mathbf{O}^{(k)}, \mathbf{F}^{(k)}\right)$, the distribution of $\left(\mathbf{O}^{(k)}, \mathbf{F}^{(k)}\right)$ can be described (see Section 1.6 for more details). The distribution of the number of CC is no more binomial, but remains explicit (see Section 2.2.

In Section 3, we give the first asymptotic results, when the number of masses goes to $+\infty$ (so that their sizes goes to zero, and in the discrete settings, $n \to +\infty$). These bring difficulties that have only been partially overcome in this paper, and then few open questions are left to the interested reader.

In Section 3.2, we investigate the difference between DDM and CDM, by defining models involving masses that are multiple of $1/n$ which can be defined in both discrete and continuous settings, and for which, comparison is possible. The parking case in which all masses are multiple of $1/n$ can be very precisely studied. In this case, Chassaing & Louchard [9, Theo. 1.5] have established a phase transition when $t_n(\lambda) := n - \lambda\sqrt{n}$ cars have been parked in the discrete parking (the largest CC contains a linear number of cars at this time, and then, a macroscopic proportion in our settings). It appears that the total number of $\mathbf{N}^{(n)}_{t_n(\lambda)}/\sqrt{n}$ of $\mathbf{N}_{t_n(\lambda)}/\sqrt{n}$ and CC of $\mathbf{O}^{(t_n(\lambda))}$ in the discrete and continuous parking have different limits (in probability), so that they are very different, while large occupied CC are essentially the same in distribution (see Proposition 3.1). Generalization of this comparison is provided in Theorem 3.5, where the case of random discrete mass are treated. In Section 3.3, more general random masses are treated (some of the results of Bertoin and Miermont [6] parking problem for caravans are recalled).

Finally, in Section 4, is investigated a quite general question, on which only first answers are obtained. It may happen that one of the main questions of interest in some dispersion models depends on all its history ; this is the case, for example, for the cost of the table construction in the data structure called "hashing with linear probing" (Knuth [18, Sec. 6.4]). This cost corresponds to the total displacements of the cars in the standard parking problem (and this corresponds to the sum of $\lfloor U_i b_i \rfloor$ where the (U_i) are

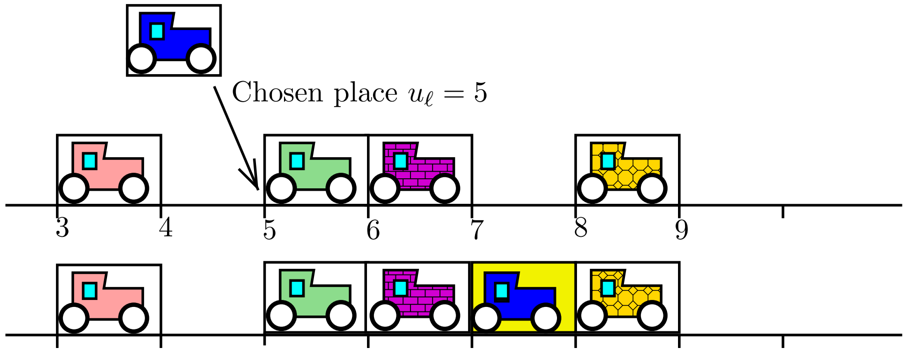

Figure 1: In the standard parking problem, a car chooses a place, here 5 and parks on the first available parking slot, which is an interval of the type $[k, k+1]$, on the right of the chosen place. Here, [7,8].

i.i.d. uniform on $[0,1]$ and b_i is the size of the occupied component that received the ith car). In general, we may imagine that the cost corresponds to some energy dissipation of the model (or some work, in the physic terminology, to spread the matter). This cost appears to be a kind of integral of a functional of the sequence, indexed by k, of the size of the CC that received the kth mass.

Mainly in the case of the parking construction, we provide limit results that describe "all large block sizes" that appears along the parking construction (Theorem 4.7). In Theorem 4.4, we prove that if the unitary cost of a single mass insertion in a block of size k satisfies some assumption, then, the limiting total cost can be expressed as a functional of the Brownian excursions. The general question is discussed in **??**, but only preliminary result are stated.

1.2 Related works / Motivation

In combinatorics, "rearranging the steps" is one of the main tools at the heart of bijections between set of paths. For example, when $N = 2n$ is even, the number of meanders with N steps (the set of paths starting at 0, with steps ± 1 and staying non-negative) coincides with the number of size N bridges (starting and ending at 0, having steps ± 1). A bijection between these sets, consists in turning over, in the bridge, the steps at which a new record minimum is reached. More involved rearrangement allows to represents paths on a graph as pairs (SAW, EC) where SAW is a self avoiding walk, and EC a heap of cycles supported by SAW (see Viennot [32], Zeilberger [33], Marchal [24], Marckert & Fredes [15]).

Rearranging the steps of structures formed by several paths is at the core of important results obtained during the last decades in probability theory: we can cite in that respect
– Propp and Wilson [30] sampling method of the uniform spanning tree,
– the result of Diaconis and Fulton [12] for a certain commutation property regarding the set of final position of random walks, when particles start successively from a sequence of vertices $s_1, \cdots, s_j$ (with possible multiplicity) and perform random walks stopped at a vertex visited for the first time (by one of the random walks): the final set of visited vertices has same distribution if the set $(s_1, \cdots, s_j)$ is permuted. This result is of the prime importance in the study of internal DLA (see e.g. Lawler & al. [22]) and in the study of loop erased random walks (see e.g. Lawler [21]).

A direct motivation of the present paper are the results obtained in Varin [31] devoted to the study of the golf process (defined in Fredes & Marckert [15] as a tool in a new proof of Aldous–Broder theorem). The golf process is defined on a graph in which some vertices contain initially a ball, some vertices contain a hole, and some vertices are neutral. Then, each ball moves according to a random walk and stops at the first hole encountered, which becomes, for subsequent balls a neutral vertex.

As for Diaconis and Fulton commutation property, the order of activation of the balls has no incidence on the distribution of the remaining holes at the end of the process. Better than that, on $\mathbb{Z}/n\mathbb{Z}$, if b balls and h holes and placed (according to a uniform distribution on the set of possible choices), then the eventual set of free holes H is the same for a large class of ball displacements policies that largely encompass random walks model. As explained in [31] (and a similar remark has been done by Tewari and Nadeau [27]), the distribution of H^1 is the same for all displacement policy under which, a ball at a given position decides the distribution of its trajectory using (at most) the data of its current position in the interval formed by the first hole at its left and at its right (this contains random walk that do not use this information, but many policies that have nothing to do with random walks).

Hence, on $\mathbb{Z}/n\mathbb{Z}$, this universal distribution property is stronger than the commutation property of Diaconis and Fulton. This result suggested that a strong property of exchangeability should hold for general diffusion policies, including continuous ones, on spaces that possesses some additional symmetries, and this leads us to the present work.

In "parking" defined on $\mathbb{Z}/n\mathbb{Z}$, even if the cars are directed to the right, the fact that the parking occupation is invariant by space symmetry $i \mapsto n-i$ is well-known (the distribution of the parking occupation appears for different regimes in [20, 29, 9, 5]...). This implies, for example, that, if the cars are directed either to the right, or to the left, with probability p and $1-p$, then, eventually, the occupation statistics would be the same for each value of p. Again, as we will prove, the fact is that even if several cars are allowed to arrive simultaneously (say n_i cars arrives at position u_i, uniform on $\mathbb{Z}/n\mathbb{Z}$, at time i), even if their policies may be complex and depends on each other choices, as long as during the evolution of the car positions, each displacement depend at most on their join positions on the current CC of park cars that contains them (or equivalently, to the distance to the interval formed by the first free place at the left and at the right), then, the distribution of the geometry of the eventual occupied places depends only on the vector $(n_1, \cdots, n_k)$, and not on the particular dispersion policy considered (and it is the same for all permutation of the n_i). This remark will be of high importance, since finally, the asymptotic behavior of "these general parking" (that will be considered as discrete dispersion models here) coincide with parking in which the cars (or caravans) are right directed, for which results are already known (Chassaing & Louchard [9], Bertoin and Miermont [6]).

The evolution of occupied blocks is made of creation of CC (when a mass arrives in a free CC), growing of blocks (during the mass deposition), and coalescence of existing blocks. For this reason, and because (discrete) additive coalescence process is related to parking models ([9, 28, 8], coalescence consideration and representation will play a role all along the paper.

1.3 The nature of general dispersion models, and restriction to "valid ones"

It is a bit difficult to define a general framework for dispersion models: in the nature, or in combinatorial/technical applications, dispersion models take very different forms. The matter that arrives on the medium can have a discrete nature or not, the medium itself can be more or less homogeneous, discrete or not, the matter can be discrete, continuous, but also can flow on the medium in a continuous way. The dispersion can be continuous in time, and after deposition/stabilization, the "layer of deposited matter" can be non-constant, non-homogeneous (as sand dunes are).

In order to take into account this wide generality of masses arrival models, we could use a process

$(A_t, t \geq 0)$, where for each t, A_t is a random Borelian measure on $\mathscr{C}$, and for a Borelian b, $A_t(b)$ would be the total mass arrived in b before time t.

And possibly, the state of the system at time t, could be described by a second measure L_t which would give the level of matter at time t (again, for a Borelian b, $L_t(b)$ would be the quantity of matters, deposited on b, at time t). But in many cases this would not be enough, since the simplest model coming from mechanics would need additional data, as the local speed, local forces, etc...

In some cases, the physical/technical phenomenon ruling the relaxation phase could be well described by partial differential equations acting on a function f where $f(t,x)$ giving the amount of matter at x at time t, but in other cases, a stochastic processes would be needed to take into account the stochastic nature of the problem under study, and maybe, in other problems, some control would be needed. In any case, the variety of models that can be imagined is huge. In many cases, an important enrichment of the probability space would be needed in order to accommodate the random physical phenomenon under consideration.

However, **the main conceptual result** of the present paper is the following: there is a huge class of dispersion models, that we will call **Valid continuous dispersion models**, that have the following characteristics:

(a) they are "discrete time processes" – the time arrivals of masses occur each slot of time,

(b) they produce a fixed "level of deposition", in words, an abscissa x in the medium is either free or occupied/covered (and once occupied it stays occupied),

(c) the masses are treated sequentially, in the sense that the stabilization/relaxation of mass m_k is done before the arrival of mass m_{k+1},

(d) during the deposition of the $k+1$th mass m_k that arrives at u_k, the connected occupied component $\mathsf{IR}(k)$ containing u_k grows continuously (except when it merges with an already existing occupied CC $O^{(k)}$ it hits) and independently from the complement of $\mathsf{IR}(k)$ in the medium: in other words, their exists a "continuous time" Interval Relaxation process $\mathsf{IR}(t)$ for $t \in [k, k+T_k]$ which describes the evolution of this CC, such that $\text{Leb}\left(\mathsf{IR}(t) \setminus O^{(k)}\right)$ is continuous, and starts at zero

(e) the model is invariant by rotation.

and all these models, after stabilization of $(m_0, \cdots, m_{k-1})$, induces the same (tractable) distribution on the eventual occupied domain $\mathbf{O}^{(k)}$.

The properties that are needed to have the universal results described above, **are not properties of the complete dispersion models, but only induced properties on occupied CC processes**: on the large probability space on which it is defined, the process $C(t)$ that encodes the behavior of the CC that received the mass can be represented as a continuous time process independent of the other components.

Moreover in the analysis (a), (b), (d) and (e) are crucial, while (c) is not, but simplifies greatly the exposition, and allows to define analyzable "cost functions" associated with the deposition process.

In Section 1.4, we add precision and formal definitions, and in Section 1.5, we provide a list of examples (that can be understood without reading Section 1.4).

1.4 Valid continuous dispersion models (valid CDM) : formal addendum

We here mainly bring some precisions on the conditions (a) – (e) discussed a bit informally in the previous subsection.

Notation. For any a, b in $\mathbb{R}/\mathbb{Z}$ denote by $\overline{a}$ and $\overline{b}$ the representatives of a and b in $[0,1)$, and $\overrightarrow{d}(a,b) = \inf\{x \geq 0 : \overline{a} + x - \overline{b} = 0 \bmod 1\}$ the distance to go from a to b, on the circle, on the direct order.

We denote by $[a,b]_{\rightarrow}$ the circular interval: for a, b in $\mathbb{R}/\mathbb{Z}$, $[a,b]_{\rightarrow} = \left\{\overline{a+x}, x \in \left[0, \overrightarrow{d}(a,b)\right]\right\}$.

Following the content of Section 1.3, we assume that "the physical dispersion phenomenon" takes place on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$),

■ **Continuous space.**

The medium on which the deposition take place is the "continuous circle" $\mathscr{C} := \mathbb{R}/\mathbb{Z}$.

■ **Model of masses.**

The sequence of masses is $(m_i, 0 \leq i \leq n)$ and n is finite or not, the mass are random or not, each mass is non-negative, and the total mass $\sum_{i=0}^{n} m_i \leq 1$; observe that the mass $m = 0$ is possible.

■ **Uniform arrival positions**

The mass m_i arrives at position u_i: the $(u_i, i \geq 0)$ are iid, independent of the masses, and taken according to $\mathbb{U} := \mathsf{Uniform}(\mathscr{C})$.

■ **Valid Continuous Dispersion Models**

A valid dispersion model will treat successively a sequence of masses arrival events (m_i, u_i) for i going from 0 to $k-1$, where u_i is the place at which the arrive a mass m_i. The set $\mathbf{O}^{(j)}$ will denote the occupied set when the j first first masses $m_0, \cdots, m_{j-1}$ have been dispersed, and $\mathbf{F}^{(j)} = \mathscr{C} \setminus \mathbf{O}^{(j)}$ the free space.

At time 0, the occupied set $\mathbf{O}^{(0)}$ is $\varnothing$, and we will add hypothesis on the diffusion process such that– among others – the set $\mathbf{O}^{(k)}$ is made of finite number of close intervals (its CC), and the sequence $(\mathbf{O}^{(k)})$ satisfies

$$\mathbf{O}^{(k)} \supseteq \mathbf{O}^{(k-1)}, \quad \text{for all } k \geq 1;$$

the event $\mathbf{O}^{(k)} = \mathbf{O}^{(k-1)}$ is possible only when a zero mass $m_{k-1} = 0$ arrived in $\mathbf{O}^{(k-1)}$ (in an already occupied point). Moreover, the Lebesgue measure of occupied space at time k corresponds to the total arrived mass before time k:

$$\mathrm{Leb}\left(\mathbf{O}^{(k)}\right) = \sum_{i=0}^{k-1} m_i. \tag{1.1}$$

As explained in the previous section, we won't define a "general dispersion model" but only fix some properties regarding the expansion of the occupied CC receiving the new mass (m_k, u_k) it has to satisfy, in other words, regarding the interval relaxation processes (IR_k) it induces.

Assume that $\mathbf{O}^{(k)}$ has been constructed (using the k elements $(u_0, m_0), \cdots, (u_{k-1}, m_{k-1})$) and has the properties described above. The hypothesis we need to make precise concerns the process $(\mathsf{IR}(t), k \leq t \leq k+1)$ that will describe the evolution of the CC that receive the new mass m_k at u_k from the time k it receive it until time $k+1$ (which includes the relaxation time interval which is $[k, k+m_k]$, see Remark 1.2 on that purpose).

To be <u>valid</u> the process

$$\mathsf{IR}_k := (\mathsf{IR}(t), k \leq t \leq k+1):$$

(i) has to be well defined a.s. for all set $O^{(k)}$ of disjoint close intervals in the support of $\mathbf{O}^{(k)}$, and for all mass arrival events (u_k, m_k) considered,

(ii) Given $(\mathbf{O}^{(k)}, u_k, m_k)$, the process IR_k can be deterministic, or random. For each $t \in [k, k+1]$, $\mathsf{IR}(t) = [a(t), b(t)]_{\rightarrow}$ is an interval of $\mathscr{C}$, and we demand that IR_k is càdlàg (and takes its values a.s. in $D([k, k+1], \mathscr{C}^2)$, where we identified the set of directed intervals on $\mathscr{C}$ with $\mathscr{C}^2$).

(iii) Define $\mathsf{IR}(k)$ as the CC of $O^{(k)}$ containing u_k, and if there are no such component, set $\mathsf{IR}(k) = \{u_k\}$. We require that during the relaxation time $[k, k+m_k]$, Leb($\mathsf{IR}(t)$) grows outside $\mathbf{O}^{(k)}$, at unit speed, that is

$$\text{Leb}\left(\mathsf{IR}(k+t)\backslash\mathbf{O}^{(k)}\right) = t, \text{for } t \in [0, m_k] \tag{1.2}$$

and can jump only when the interval $\mathsf{IR}(t)$ coalesces with one of the interval present in $\mathbf{O}^{(k)}$ (despite its appearance, this condition is not really a condition on the speed of deposition! See Remark 1.2).

(iv) For all $s \le k + m_k$, $(\mathsf{IR}(t), t \in [k, s))$ is independent from $\mathbf{O}^{(k)}\backslash\mathsf{IR}(s)$: in words, the extension of the "CC" $\mathsf{IR}(t)$ containing u_k must be independent of the rest of the state of the medium (that is, from $\mathbf{O}^{(k)} \setminus \mathsf{IR}(ts)$). From time $k + m_k$ to time $k+1$, $\mathsf{IR}(t)$ remains constant, and equal to $\mathsf{IR}(k+m_k)$.

(v) The model is invariant by rotation, meaning that the (distribution of the) diffusion processes $\mathsf{IR}(t)$ defined from time k to time $k+m_k$ does not depend on the position on the circle. Formally denote by r_α the rotation by $\alpha \in \mathbb{R}/\mathbb{Z}$ in $\mathbb{R}/\mathbb{Z}$. For all k, the diffusion process $\mathsf{IR}(t)$ defined on the time interval $[k, k+m_k]$ has a distribution which depends uniquely on $(\mathbf{O}^{(k)}, u_k, m_k)$. For all $\alpha \in \mathbb{R}/\mathbb{Z}$,

$$\mathscr{L}\left((\mathsf{IR}_k) \mid (\mathbf{O}^{(k)}, u_k, m_k)\right) = \mathscr{L}\left(r_{-\alpha}(\mathsf{IR}_k) \mid (r_\alpha(\mathbf{O}^{(k)}), r_\alpha(u_k), m_k)\right) \tag{1.3}$$

We define $\mathbf{O}^{(k+1)}$ as the close set $\mathsf{IR}(k+1) \cup \mathbf{O}^{(k)}$ (by construction $\mathsf{IR}(k+1)$ may contain zero, one or several CC of $\mathbf{O}^{(k)}$).

Definition 1.1. *A continuous dispersion model is said to be valid, if all the interval relaxation processes* (IR_k) *are valid.*

Remark 1.2 (About the relaxation time of the kth mass)**.** *A good way to understand the extent of the class of valid continuous process, is to consider that the relaxation phase of the kth mass occurs in a time frame which is independent of the discrete times enumerating the mass arrival events (m_k, u_k).*

When one wants to encode an actual physical system (e.g. the diffusion of a droplet in a random medium), it is often natural to encode the relaxation time of the kth mass by a <u>*continuous time interval*</u>*:*
(a) for example, on $[k, k+1)$ (so that, everything is done for the arrival of the next $k+1$th mass ,
(b) or on $[0, +\infty)$ to let as many time as needed for the deposition time,
(c) or on $[k, k+m_k]$, and then assume that the mass deposition is done at unit speed.

Since we observe the states after the relaxation times (to define $O^{(k)}, m_k, u_k$...), and since one can pass from any of the models $(a), \cdots, (c)$ to the others by a change of time, these points of views are equivalent for our purpose. We then choose the choice (c) to make easier the description of everything on the same clock.

When needed, we will qualify this property, to be "the constant speed deposition property".

Remark 1.3 (About the invariance by rotation)**.** *Without the invariance by rotation hypothesis (v), the diffusion process could depend on the arrival points u_i: we could design some diffusion processes that would stop their diffusion at some special places (e.g. so that to produce occupied sets $\mathbf{O}^{(k)}$ such that each CC have a (random) extremity with rational coordinate on $\mathscr{C}$). All universal results presented in Section 1.6 would not hold without condition (v).*

1.5 List of examples of valid continuous dispersion processes

In this section we provide a list of 8 valid continuous dispersion models.

1.5.1 Right Diffusion at Constant Speed (RDCS): an important cornerstone.

See Fig. 3 for an illustration. The Right Diffusion at Constant Speed is an important diffusion process, which will be used as a cornerstone in the rest of the paper. It is defined for any masses $(m_0, m_1, \cdots, m_k)$ as long as their sums is in $[0,1)$.

Assume that there is a mass arrival (m,u) on a $(\mathbf{O}^{(k)}, \mathbf{F}^{(k)})$. Let $\mathsf{IR}(k) = [a,b]$ be the CC of $\mathbf{O}^{(k)}$ containing u if it exists, or $\mathsf{IR}(k) = [a,b] := [u,u]$ if it does not. Now, the mass (m,u) is pushed to the right on $\mathbb{R}/\mathbb{Z}$ from u and cover the space from here: at time $t \in [k, k+m_k]$, the CC is the smallest oriented interval $\mathsf{IR}(t) = [a, b(t)]_{\rightarrow}$ on $\mathbb{R}/\mathbb{Z}$ (that grows on its right extremity) that satisfies Section 1.4 (that is $\mathrm{Leb}\left(\mathsf{IR}(k+t)\backslash\mathbf{O}^{(k)}\right) = t$, for $t \in [0, m_k]$).

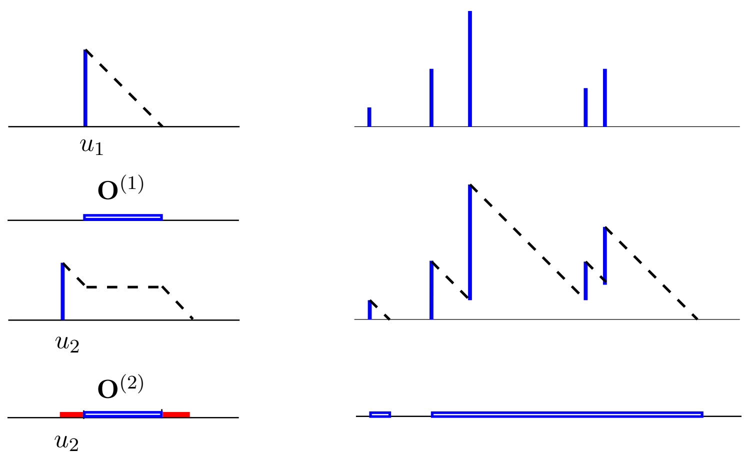

Figure 2: The cycle $\mathbb{R}/\mathbb{Z}$ is represented as a segment $[0,1)$. In the RDCS, the masses are pushed to the right. On the first column, two masses arrival events and their relaxations. A new mass is pushed to the right and covers the medium with the same speed, as if it was eroded. The third picture illustrates that the erosion occurs only on free spaces, while the mass if pushed with no loss on an occupied domain.
On the second column, a phenomenon that will appear progressively in the sequel: the eventual occupied domain does not depend on the order of appearance of the masses. It is clear in the RDCS model, because, the quantity of "non eroded mass" that passes above a given abscissa x does not depend on "the identity of the mass" which is eroded at a given place. In the RDCS, the eventual space occupation is a deterministic function of the mass arrival events (u_i, m_i).

Remark 1.4. *The illustrations of 3 appears in many papers, typically to encode paths with a deterministic slope -1, and whose randomness appears as positive jumps (notably Łukaciewisz paths are like this, and any path representing customers at which a continuous service is given). In Bertoin & Miermont [6] (see Section 3.3), the RDCS is called the "model of caravan": "masses are drop of paint brushed to the right". Their is however a subtle difference: they encode the covered space $\mathbf{O}^{(k)}$ as union as open sets, while we take them as close. To take them as close, allow to consider covered set reduced to a point, which will be proved to bring significative (global) simplifification in our analysis.*

Of course the Left Diffusion at Constant Speed (LDCS) can be defined analogously: it will be used only in the three next examples.

1.5.2 $(p, 1-p)$ proportion of the mass is diffused to the right/to the left

When a mass arrival event (m, u) occurs, first, proceed to the RDCS of (pm, u) (a proportion p of the mass is diffused to the right), then proceed to the LDCS of $((1-p)m, u)$. If we proceed to the RDCS and LDCS one after the other, the total time of diffusion is m; if we perform both task simultaneously, the eventual result is the same but occurs before, at time $\max\{p, 1-p\}m$; this is an additional instance in which we can observe that many details of the models are irrelevant when we are only interested in the final configuration.

1.5.3 With probability p performs a random diffusion, with probability $1-p$ a left one

When a mass arrival event (m, u) occurs, toss a Bernoulli (p) coin. If the result is 1, then proceed to the RDCS of (m, u) else proceed to the LDCS of (m, u).

1.5.4 Diffusion to the closest side diffusion

When a mass arrival event (m, u) occurs. If u does not belong to the current occupied domain $\mathbf{O}^{(k)}$ then proceed to a $(1/2, 1/2)$ proportion to the right/left diffusion. Else, $u \in [a, b]$ with $[a, b]$ a CC of $\mathbf{O}^{(k)}$. Compare the distance $d_{\mathbb{R}/\mathbb{Z}}(u, a)$ with $d_{\mathbb{R}/\mathbb{Z}}(u, b)$. If the first is the smaller then proceed to a LDCS of (m, u), else to a RDCS.

1.5.5 Diffusion to the closest side with constant reevaluation

Do the same thing as in "to the closest side diffusion", but each time element $\mathrm{d}t$, recompute the distance to the sides of the relaxation interval. When the sides are at equal distance, grows both side as the same speed.

1.5.6 Infinitesimal particle like diffusion

This model describes the asymptotic diffusion behavior of masses m_i composed by "infinitesimal" particle of size $1/M$ (and the limiting regime is taken for $M \to +\infty$), where these small particles perform successively independent random walks till the moment they get out of the occupied interval. Doing so, they extend the size of the occupied domain by $1/M$ for subsequent infinitesimal particles composing m_i.

In order to describe these dynamics let us assume that a mass m arrives at 0, in a block $[-a, b]_{\rightarrow}$ (with $a, b > 0$). As usual, if the arrival location 0 is in the free space, then, $[a(0), b(0)]_{\rightarrow} = \{0\}$. The dynamics being invariant by rotation, it suffices to provide a full description of this case (as we will see, we give the speed at which a and b moves, so that what we will say remains valid when CC merge: it suffices to modify a and b accordingly). For B, a Brownian motion starting at 0,

$$\mathbb{P}\big(\inf\{t : B(t) = -a\} < \inf\{t : B(t) = b\}\big) = b/(a+b).$$

At the limit over M, the speed under which a boundary moves, satisfies a simple differential equation which is the immediate limit of the preceding consideration; to understand how the interval change with time, let us work on $\mathbb{R}$ for a moment (the transfer to $\mathbb{R}/\mathbb{Z}$ is just a formality).

If $(m,0)$ is a mass arrival event, and $\mathrm{IR}(k+t)=[-a(t),b(t)]$ is the occupied CC (with $-a[t]<0<b[t]$), we have

$$b'(t)=\frac{a(t)}{a(t)+b(t)},\ \ a'(t)=\frac{b(t)}{a(t)+b(t)}.$$

We fix the deposition speed so that $a(t)+b(t)=a(0)+b(0)+t$, observe that $a'(t)+b'(t)=1$, and obtain

$$b(t)=\frac{t+a(0)+b(0)}{2}+\frac{b(0)^2-a(0)^2}{2(a(0)+b(0)+t)},\ \ a(t)=\frac{t+a(0)+b(0)}{2}+\frac{a(0)^2-b(0)^2}{2(a(0)+b(0)+t)}. \tag{1.4}$$

Of course, upon collision with another occupied block, $\mathrm{IR}(t)$ jumps, as usual.

We justify the approximation of the interval evolution by this fluid limit in Section 5.1.

1.5.7 Range of the Brownian path

Let us provide an example given by a random process: Denote by $R(u,t):=\{u+B_s \bmod 1, 0\le s\le t\}$ the range of a Brownian motion B on $\mathbb{R}/\mathbb{Z}$ (where $B_0=0$). Of course $t\mapsto R(u,t)$ is non-decreasing for the inclusion partial order. For (m_k,u_k) the $k+1$th mass arrival event, and $\mathbf{O}^{(k)}$ the occupied set at time k.

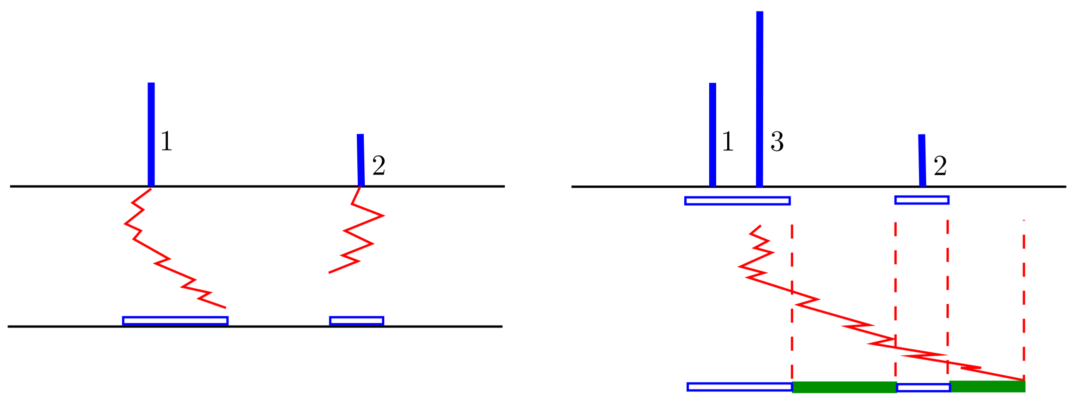

Figure 3: The range of a Brownian motion model: the first Brownian motion $B^{(0)}$ starts at u_0 and is killed at the first t for which $\mathrm{Leb}\left(\left\{B^{(0)}_s, s\le t\right\}\right)=m_0$: its track is $\mathbf{O}^{(1)}=\left\{B^{(0)}_s, s\le t\right\}$. The ith Brownian motion $B^{(i)}$ starts at u_i and is killed when $\mathrm{Leb}\left(\left\{B^{(i)}_s, s\le t\right\}\setminus \mathbf{O}^{(i-1)}\right)=m_i$, that is when the set of points it visited, but have never been visited by the previous Brownian motion, has Lebesgue measure m_i.

Start a Brownian motion at u_k, and stop it the (random) time τ when $\mathrm{Leb}(R(u_k,\tau)\setminus\mathbf{O}^{(k)})=m_k$, and then $\mathbf{O}^{(k+1)}=\mathbf{O}^{(k)}\cup R(u_k,\tau)$.

This is a case where the relaxation interval process (which grows at constant speed) can be explicitly described . Consider, for any $x\in[0,m_k]$, $\tau(x)=\inf\{s\ge 0\ :\mathrm{Leb}(R(u_k,s)\setminus\mathbf{O}^{(k)})\ge x\}$. For all $0\le x\le m_k$, define $\mathrm{IR}(k+x)$ as the CC of $\mathscr{O}k\cup\{R(u_k,\tau(x))\}$ that contains u_k (and let IR_k to be constant, as usual between time $k+m_k$ and time $k+1$).

1.5.8 The short-sighted jam-spreader model

An individual, short-sighted, wants to spread some drops of jam that falls on a donut (identified with $\mathscr{C}$) according to a classical mass arrival events $((m_i,\mathbf{u}_i), i\ge 0)$. During the spreading of the ith drop, the spreader is unable to see where the jam is lacking on $\mathscr{C}$; he therefore transforms progressively the pile of non-deposited jam, that we could represent by a Borelian measure $(M_t, t\ge i)$ on $\mathrm{IR}(t)$ whose total mass at time t is the mass of non-deposited jam. We may assume that the spreader proceeds randomly, and that his actions expand $\mathsf{IR}(t)$ continuously, till all the arrived jam is finally spread, and a new portion of $\mathscr{C}$ with Lebesgue measure equal to m_i is covered.

We assume that the "strategy" of the individual is invariant by rotation, and that the ith policies he adopt for the ith drop is independent from the current sets $\mathbf{O}^{(i)}$ and $\mathbf{F}^{(i)}$, and depends only on the occupied block in which falls this drop (and possibly on the current measure M_t he is spreading). The policies used for different i may be different.

1.6 Main universality results for valid CDM: deterministic masses case

This section is devoted to the study of the statistical properties of a valid continuous dispersion model, when the sequence of masses $m[k] := (m_i, i = 0, \cdots, k-1)$ that arrives in the system are deterministic, non-negative, and satisfies only the total weight condition

$$W(m[k]) := \sum_{i=0}^{k-1} m_i < 1. \tag{1.5}$$

We will turn to random masses in Section 3.1.

We insist on the fact that we allow the m_i to be 0 too: this will play an important role in the full understanding of our results, and on the fact that exact computation can be done. Again, a mass zero arriving in an occupied CC has no effect, while, on arriving in a free CC, it creates a new occupied CC reduced to the point $\{u\}$ at which the arrival took place.

The first aim of the paper is to describe the distribution of the occupied and free spaces $\mathbf{O}^{(k)}$ and $\mathbf{F}^{(k)}$ at time k when k masses $((m_i, \mathbf{u}_i), 0 \le i \le k-1)$ have been dispersed according to a valid continuous dispersion model; in particular, we put in bold u_i to distinguish its type "random variable" from that of the $(m_i, i \ge 0)$ simple non-negative real numbers, that are not random for the moment.

We let

$$\mathbf{N}_k = \#\mathbf{F}^{(k)} = \#\mathbf{O}^{(k)} + \mathbb{1}_{k=0}, \tag{1.6}$$

be the number of free CC when k masses have been dispersed in the system (if some of the m_i are not zeros, $\mathbf{N}_k$ is not deterministic). For $k \ge 1$, label the CC $\mathbf{O}^{(k)}_0, \cdots, \mathbf{O}^{(k)}_{\mathbf{N}_k-1}$ of $\mathbf{O}^{(k)}$, and $\mathbf{F}^{(k)}_0, \cdots, \mathbf{F}^{(k)}_{\mathbf{N}_k-1}$ those of $\mathbf{F}^{(k)}$, such that by turning around the circle $\mathbb{R}/\mathbb{Z}$, one finds successively (they are adjacent in this order),

$$\mathbf{L}^{(k)} := \mathbf{O}^{(k)}_0, \mathbf{F}^{(k)}_0, \cdots, \mathbf{O}^{(k)}_{\mathbf{N}_k-1}, \mathbf{F}^{(k)}_{\mathbf{N}_k-1}, \tag{1.7}$$

and moreover, the point

$$0 \in \mathbf{O}^{(k)}_0 \cup \mathbf{F}^{(k)}_0. \tag{1.8}$$

The point 0 of the circle is then use to decide which CC of $\mathbf{O}^{(k)}$ and $\mathbf{F}^{(k)}$ will take the name $\mathbf{O}^{(k)}_0$ and $\mathbf{F}^{(k)}_0$: this produces a bias which complicates a bit the exposition. The sequence of block lengths

$$|\mathbf{L}^{(k)}| := \left(|\mathbf{O}^{(k)}_0|, |\mathbf{F}^{(k)}_0|, \cdots, |\mathbf{O}^{(k)}_{\mathbf{N}_k-1}|, |\mathbf{F}^{(k)}_{\mathbf{N}_k-1}|\right) \tag{1.9}$$

together with the shift information

$$s_0 = -\min \mathbf{O}^{(k)}_0, \text{ and then } s_0 \in \left[0, |\mathbf{O}^{(k)}_0 \cup \mathbf{F}^{(k)}_0|\right]$$

allows to reconstitute $\mathbf{L}^{(k)}$ (each of them characterizes the other one).

1.6.1 Universal properties

The next theorem, one of the main result of the paper, gathers the properties that are universal for all valid CDM. In one slogan: the distribution of $\left(\mathbf{O}^{(k)},\mathbf{F}^{(k)}\right)$ is an invariant of all valid CDM.

Theorem 1.5. *Consider a valid* CDM *A, and some deterministic masses* $m[k]=(m_0,\cdots,m_{k-1})$*, positive or zero, satisfying the total weight condition* $W(m[k])<1$*. Denote by* $(\mathsf{IR}^A_j, 0\leq j\leq k-1)$ *the interval relaxation processes associated with A. We have*

(i) *For a fixed k, the law of* $\left(\boldsymbol{O}^{(k)},\boldsymbol{F}^{(k)}\right)$ *does not depend on A (it is the same for all valid* CDM *that are designed to treat the masses* $(m_0,\cdots,m_{k-1})$*: that is for all valid interval relaxation processes* $(\mathsf{IR}_j, 0\leq j\leq k-1)$*).*

(ii) *The law of* $\left(\boldsymbol{O}^{(k)},\boldsymbol{F}^{(k)}\right)$ *is invariant under the permutation of masses: For all permutation* $r\in\mathbf{S}k$ *(the symmetric group on* $\{0,\cdots,k-1\}$*),*

$$\mathscr{L}\left(\left(\boldsymbol{O}^{(k)},\boldsymbol{F}^{(k)}\right)\mid(m_0,\cdots,m_{k-1})\right)=\mathscr{L}\left(\left(\boldsymbol{O}^{(k)},\boldsymbol{F}^{(k)}\right)\mid(m_{r(0)},\cdots,m_{r(k-1)})\right)$$

(where we have written as a condition, the successive masses that are used).

(iii) *The number of free CC* $\mathbf{N}_k$ *has (almost) a binomial distribution:*

$$\mathbf{N}_k\overset{(d)}{=}1+B(m[k]),\ \ \textit{where}\ \ B(m[k])\sim\mathsf{Binomial}(k-1,R_k) \tag{1.10}$$

where $R_k=1-W(m[k])$ *is the final free space available after the deposition of the k first masses.*

(iv) *Take a uniform element* θ *in* $\{0,\cdots,b-1\}$ *(independently of everything). Conditional on* $\mathbf{N}_k=b$*,*

$$\left(\left|\boldsymbol{F}^{(k)}_{i+\theta\ \mathsf{mod}\ b}\right|,0\leq i\leq b-1\right)\overset{(d)}{=}|R_k|\,(\mathbf{D}_0,\cdots,\mathbf{D}_{b-1}) \tag{1.11}$$

where $(\mathbf{D}_0,\cdots,\mathbf{D}_{b-1})\sim\mathsf{Dirichlet}(1,\cdots,1)$*. Hence, the sequence* $\left(|\boldsymbol{F}^{(k)}_i|,0\leq i\leq b-1\right)$ *taken according to a uniform independent rotation, is unbiased.*

(v) *The process* $(\mathbf{N}_k,k\geq 0)$ *is a Markov chain, and its transition matrix can be computed explicitly (see Section 1.10).*

To remove the rotation in (iv) has a price, because of the bias (see Theorem 1.8 and Remark 1.10).

The free spaces possess an additional universal property that we prefer to state apart for a technical reason, appearing already in Theorem 1.5 (i).

Denote by $\overrightarrow{\mathbf{F}^{(k)}}$ for the free spaces random variable, in the right diffusion with constant speed model.

Theorem 1.6. *Under the hypothesis of Theorem 1.5, for* $k\geq 1$ *fixed,*

$$\mathscr{L}\left(|\boldsymbol{F}^{(k)}|\mid(m_0,\cdots,m_{k-1})\right)=\mathscr{L}\left(\left|\overrightarrow{\boldsymbol{F}^{(k)}}\right|\mid(W(m[k]),\underbrace{0,\cdots,0}_{k-1\ zeroes})\right).$$

Hence, even if a CDM is not made to deal with all masses, the distribution of the free spaces is the same as those for the RDCS, for which the masses m_i can be rearranged the way we want (as long as we keep their number and sum). This theorem is a Corollary of Theorem 1.5, in which by $(iii)-(v)$, we see that the total free space distribution depends only $(k,\sum_{i=0}^{k-1}m_i)$.

1.7 On the occupied blocks distribution

Theorem 1.5 tells us that it suffices to describe the distribution of $\left(\mathbf{O}^{(k)}, \mathbf{F}^{(k)}\right)$ for the RDCS. The distribution of $\mathbf{O}^{(k)}$ is more complex than that of $\mathbf{F}^{(k)}$. There are two main reasons for that.
(a) First, some combinatorics enter into play: the probability that the CC have size $s_0, \cdots, s_{b-1}$ depends on the number of partitions of the masses $(m_0, \cdots, m_{k-1})$ in disjoint subsets of total mass $s_0, \cdots, s_{b-1}$.
(b) there is a size bias when one tries to describe the occupied block in some order (see Section 1.6).

A technical tool enters into play.

Definition 1.7. *For some masses $m[k] := (m_0, \cdots, m_{k-1})$ with sum in $[0,1)$, the piling propensity*[1] *of $m[k]$ is defined by*

$$Q(m[k]) := W(m[k])^{k-1}. \tag{1.12}$$

By Theorem 1.5,

$$Q(m[k]) = \mathbb{P}(\mathbf{N}_k = 1), \tag{1.13}$$

so that $Q(m[k])$ measures the propensity of $m[k]$ to form a single occupied CC, under the action of any valid CDM. The notation $Q(m[k])$ perhaps hides a bit that Q depends on k and of $W(m[k])$ only.

Conditional on $\mathbf{N}_k = 1$, $\mathbf{O}^{(k)}$ is reduced to a single (random) interval $[A, A + W[m[k]]]_{\rightarrow}$ and by invariance by rotation A is uniform on $\mathbb{R}/\mathbb{Z}$:

$$\mathbb{P}(A \in \mathrm{d}a \mid \mathbf{N}_k = 1) = \mathrm{d}a \tag{1.14}$$

so that

$$\mathbb{P}(A \in \mathrm{d}a, \mathbf{N}_k = 1) = W(m[k])^{k-1}\,\mathrm{d}a. \tag{1.15}$$

This single block consideration integrates the fact that all m_i falls inside $[a, a + W(m[k])]$, "one" arrived in $\mathrm{d}a$, and the other ones, combine so that they cover exactly $[a, a + W(m[k])]$. This mundane observation allows to produce a multi-block formula at the double price of fixing the identity of the masses participating to each block, and to place the blocks precisely on the circle so that they don't intersect.

For a set of indices J, set

$$W(m(J)) := \sum_{j \in J} m_j.$$

Denote by $\mathbf{I}_j^{(k)}$ the indices of the mass among $m[k]$ that have been dispersed to form the occupied CC $\mathbf{O}_j^{(k)} = \left[\mathbf{A}_j^{(k)}, \mathbf{B}_j^{(k)}\right]_{\rightarrow}$. We have, as a consequence of the single block formula:

Theorem 1.8. *Take the same hypothesis as in Theorem 1.5. For any $b \in \{1, \cdots, k\}$, any partition $(J_0, \cdots, J_{b-1})$ of the set $\{0, \cdots, k-1\}$ with non-empty parts, any sequence $(a_0, a_1, \cdots, a_{b-1})$ in $\mathscr{C}^b$ such that the $a_0, a_1, \cdots, a_{b-1}$ are cyclically ordered around $\mathscr{C}$, such that turning around the circle we get $a_0 \preceq 0 \preceq a_1 \preceq \cdots \preceq a_{b-1} \preceq a_0 \preceq 0 \cdots$ and such that*

$$\operatorname{Leb}\left([a_i, a_{i+1 \bmod b}]_{\rightarrow}\right) > W(m(J_i)) \quad \textit{for } i \in \{0, \cdots, b-1\},$$

we have

$$\mathbb{P}\left(\boldsymbol{A}_j^{(k)} \in \mathrm{d}a_j, \boldsymbol{I}_j^{(k)} = J_j, 0 \le j \le b-1\right) \quad = \quad \prod_{j=0}^{b-1} Q(m(J_j))\,\mathrm{d}a_j. \tag{1.16}$$

[1] This piling propensity is "close to" the probability that all the masses arrive in a given interval of size $W(m[k])$ which is $W(m[k])^k$.

- In this formula, a bias appears under the condition $a_0 \preceq 0 \preceq a_1$.
- Notice that the event $\left\{\mathbf{A}_j^{(k)} \in \mathrm{d}a_j, \mathbf{I}_j^{(k)} = J_j, 0 \le j \le b-1\right\}$ characterizes entirely $\mathbf{O}^{(k)}$ as well as $\mathbf{F}^{(k)}$.

In order to extract the occupied block sizes, we need to "count" the number of ways to produce some given block sizes.

Theorem 1.9. *Take the same hypothesis as Theorem 1.5. Let $M_0, \cdots, M_{b-1}$ be some non-negative masses. For all $a_0 \preceq 0 \preceq a_1 \preceq \cdots \preceq a_{b-1} \preceq a_0 \preceq 0 \cdots$ and such that $\mathrm{Leb}\left([a_i, a_{i+1 \bmod b}]_{\to}\right) > M_i$ for all $i \in \{0, \cdots, b-1\}$,*

$$\mathbb{P}\left(\boldsymbol{A}_j^{(k)} \in \mathrm{d}a_j, \left|\boldsymbol{O}_j^{(k)}\right| = M_j, 0 \le j \le b-1\right) \quad = \quad \mathsf{Nb}(m[k], M[b]) \prod_{j=0}^{b-1} Q(M_j)\,\mathrm{d}a_j \tag{1.17}$$

where $\mathsf{Nb}(m[k], M[b])$ is the number of partitions $(J_0, \cdots, J_{b-1})$ in non-empty parts, such that, for all i, $W(m(J_i)) = M_i$.

Again, Theorem 1.9 characterizes the distribution of $(\mathbf{O}^{(k)}, \mathbf{F}^{(k)})$. The computation of the "marginal" $|\mathbf{O}^{(k)}|$ recording only the block length sizes is also possible: we need to compute the volume of the "translation set" $(a_0, \cdots, a_{b-1})$ described in Theorem 1.9:

$$\mathbb{P}\left(\left|\mathbf{O}_j^{(k)}\right| = M_j, 0 \le j \le b-1\right) \quad = \quad \mathsf{T}(M[b]) \times \mathsf{Nb}(m[k], M[b]) \times \prod_{j=0}^{b-1} Q(M_j) \tag{1.18}$$

where

$$\mathsf{T}(M[b]) \quad = \quad M_0 \frac{(1 - W(m[k]))^{b-1}}{(b-1)!} + \frac{(1 - W(m[k]))^{b}}{b!} \tag{1.19}$$

is the Lebesgue measure of the "Translation set" of $(a_0, \cdots, a_{b-1}) \in \mathbb{R}^b$ given in Theorem 1.9 (see Section 5.2 for a proof which relies on the fact that this set has same measure as $S := \{(u, s_0, \cdots, s_{b-2}), u \in [0, M_0 + s_0], 0 \le s_0 \le s_1 \le \cdots s_{b-2} \le 1 - W(m[j])\}$ where the variable s_j accumulates the free length spaces between the $\mathbf{O}_0^{(k)}$ and $\mathbf{O}_{j+1}^{(k)}$, and u is used to place zero in the interval $[0, M_0 + s_0]$)).

The measure $T(M[b])$ depends only on $(M_0, W(m[k]))$: the presence of M_0 comes from the size bias. parler de la preuve de Theorem 1.5

Remark 1.10. *Theorem 1.5, says that $\boldsymbol{F}_0^{(k)}$ inherits a bias from $\boldsymbol{L}^{(k)}$, in which it is biased by $|\boldsymbol{O}_0^{(k)}| + |\boldsymbol{F}_0^{(k)}|$. By uniform re-rooting, one sees that either $\boldsymbol{F}_0^{(k)}$ gets the label 0, because the uniform mark falls in it (this happens with probability $|\boldsymbol{F}_0^{(k)}|$) or because the "occupied block" on its left received the uniform mark. Now, conditioning on $\mathbf{N}_k = b$, since these blocks are exchangeable:*
– the probability that U falls in a free block is R_k, in this case $|\boldsymbol{F}_0^{(k)}|$ is R_k times a $\beta(2, b)$ random variable (and this have density $g_1(x) = b(b+1)(x/|R_k|)(1-(x/|R_k|))^{b-1}/|R_k|\mathbb{1}_{x \in [0,|R_k|]}$) since this is the size biased distribution of the first marginal in a Dirichlet $(1, \cdots, 1)$ random variable,
– or with U falls in an occupied block is $1 - R_k$, and the block that follows is R_k times a $\beta(1, b-1)$ random variable (and this has density $g_2(x) := (b-1)(1 - x/|R_k|)^{b-2}/|R_k|\mathbb{1}_{x \in [0,|R_k|]}$.
Finally, conditional on $\mathbf{N}_k = b$, the density of $\boldsymbol{F}_0^{(k)}$ is $\mathbb{1}_{x \in [0, R_k]}\big(|R_k| g_1(x) + (1 - |R_k|) g_2(x)\big)$.

A multiset $\{\{x_1, \cdots, x_m\}\}$ is informally, a set in which elements may have an arbitrary multiplicity [2]

The last universal result we would like to state concern "the coalescence" process induced by the diffusion models we study.

[2] Define an equivalence relation between sequences: two sequences are equivalent if they have the same length, and if they are equal up to a permutation of their terms. A multiset is an equivalence class for this relation.

Theorem 1.11. *Let $(m_0, m_1, \cdots, m_{k-1})$ be some masses with sum in $[0,1)$. Consider, for $j \in \{0, \cdots, k\}$, the multiset*

$$S(j) := \left\{\left\{\left|\boldsymbol{O}_i^{(j)}\right|, 0 \le i \le \mathbf{N}_j - 1\right\}\right\}$$

which provides the sizes (with multiplicity) of the occupied CC at time j.

The distribution of the process $(S(j), 0 \le j \le k-1)$ is the same for all valid DDM.

Remark 1.12. *The free spaces have not the same property: it happen with a positive probability that using the RDCS from time t to $t+1$, a single free CC is reduced, while using the "$(p, 1-p)$ proportion to the right/left diffusion" of Section 1.5.2, it never happens.*

Remark 1.13. *Chassaing & Louchard [9], considered $S(t)$ the multiset of occupied block sizes at time t in the parking process. The blocks corresponds to a (maximal) set of consecutive occupied places together with the first free place at its right. They proved that the process $(S(t), t \ge 0)$ has the same law as in the additive coalescent process starting with n particles of mass 1 observed at successive coalescence times. This process can be encoded by coalescent forests (Pitman [28]),in which tree sizes correspond to block sizes. The joint distribution of the trees sizes is known: suitably ordered they have same distribution as iid Borel variables[3] conditioned to have a fixed sum (see e.g. Pitman [28, Prop.6], Bertoin [5, Cor. 5.8], [9, Proof of Prop. 5.1],Marckert & Wang [26]...).*

1.8 Reductio ad Right Diffusion with Constant Speed and excursion sizes

Theorem 1.5 tells us that given $m[k]$, the distribution of $(\mathbf{O}^{(k)}, \mathbf{F}^{(k)})$ for any valid CDM, can be studied using the RDCS. We recommend to have a look at Figure 4 before reading the following text.

Definition 1.14. *Define the "collecting path" process $S := (S_x, 0 \le x \le 1)$ by*

$$S_x = -x + \sum_{j=0}^{k-1} m_j \mathbb{1}_{u_j \le x}, \quad \textit{for } x \in [0,1]. \tag{1.20}$$

The extended collecting path $\overline{S} = (\overline{S}_x, x \in \mathbb{R})$ is a process indexed by $\mathbb{R}$, defined as on Figure 4 by concatenating copies of the collecting paths head to tail,

$$\overline{S_x} = S_{\{x\}} + \lfloor x \rfloor S_1, \quad \textit{for } t \in \mathbb{R}$$

where $\{x\}$ is the fractional part of x.

The collecting path is named "profile" in Bertoin & Miermont [6]. For all a level $y \in \mathbb{R}$ set

$$\tau_y = \inf\left\{t : \overline{S}_t = y\right\}$$

the first hitting time of y, and to define the excursion of $\overline{S}$ at level y, set $\tau^y = \sup\left\{x : \min\{\overline{S}_x \in [\tau_y, x] = y\}\right\}$. Set $\ell_x := \tau^y - \tau_y$ the length of the excursion above level y. Let a be the first time that S reaches its minimum on $[0,1]$:

$$a = \min \operatorname{argmin} S.$$

[3] The Borel law $\mathsf{Borel}(\lambda)$, is the discrete distribution p_λ with support $\{1,2,3,\cdots\}$ defined by $p_\lambda(k) = (\lambda k)^{k-1} \exp(-k\lambda)/k!$

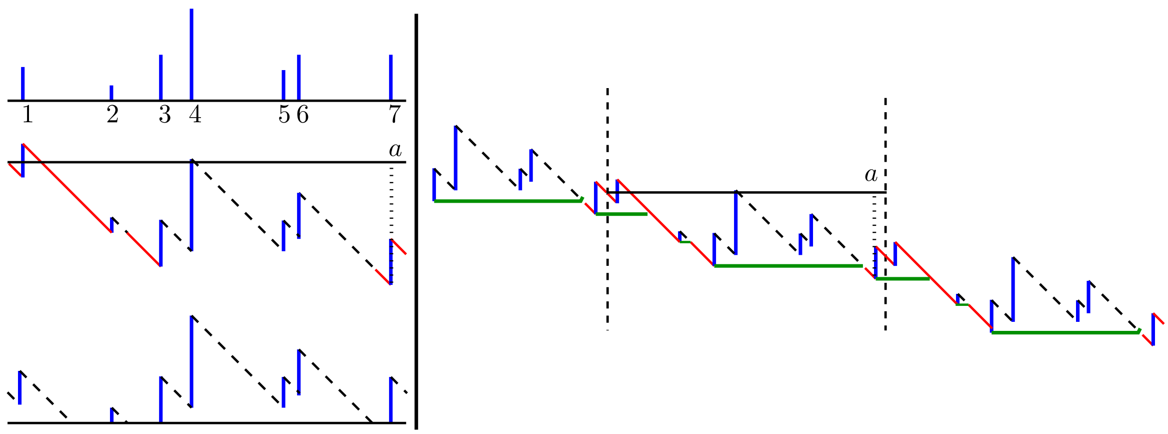

Figure 4: On the first column: By Theorem 1.5, the masses can be treated in our preferred order, and if we use the order 7,1,2,3,4,5,6 using the RDCS, we get the third picture: since we are on the cycle $\mathbb{R}/\mathbb{Z}$, there are 3 occupied CC. On the second picture, the process S is represented. To recover the excursions, there is a boundary effect: for example, the first excursion above the "minimum process" does not correspond to an occupied CC.
On the second column, one may see that by working on $\overline{S}$ obtained by pasting several trajectories of S head to tail, then, starting from the abscissa aS of the minimum of argmin S on $[0,1]$, one recovers the occupied blocs as the excursion lengths above the minimum process of $\overline{S}$ on $[a, a+1]$.
Notice also that this encoding is not sufficient to encode zero masses, since they let unchanged the collecting path, while possibly creating CC.

Lemma 1.15. *The multiset of positive occupied CC sizes in the RDCS with mass arrival events* $[(m_i, u_i), 0 \le i \le k-1]$*, corresponds to the non-zero excursion length of the collecting path* $\overline{S}$ *for* $x \in [a, a+1)$ *above the current minimum process.*

Zero masses leave unchanged the collecting paths and can create zero size CC: these CC can not be recovered $\overline{S}$.This lemma is proved in [6, Lemma 3] (and similar discrete results are present in [2], [9], ...). The study of the collecting paths is an important tool in the analysis of the cluster sizes in multiplicative and additive coalescence processes, and the link between them is given by this Lemma.

Remark 1.16. *Hence, for some deterministic masses* $m[k]$ *with sum* $W(m[k]) < 1$*, the 8 dispersions models introduced in Section 1.5 induces the same distribution on* $(\boldsymbol{O}^{(k)}, \boldsymbol{F}^{(k)})$*, and this is true also, as a simple corollary, if they use the same random masses. The number of CC of* $\boldsymbol{F}^{(k)}$ *is then always* 1 *plus a binomial random variable with parameters* $(k-1, W(m[k]))$*. The law of their sizes, etc, is the same in all cases. In particular, they can be studied using the RDCS model, and corresponds then (up to the boundary effect discussed in Figure 3), to the excursion sizes of the collections path above the current minimum.*

1.9 Proof of Theorem 1.5, Theorem 1.6 and Theorem 1.11

The proofs rely on some elementary principles.

Consider k iid uniform points $u_0, \dots, u_{k-1}$ taken on the circle $\mathscr{C} = \mathbb{R}/\mathbb{Z}$.

- **(A) A static principle.**

Then take I a measurable subset of $\mathscr{C}$ with positive Lebesgue measure, that can be random or not, but is chosen independently of the u_i.

If we condition the set $S := \{u_i, 0 \le i \le k-1\}$ to satisfy $|S \cap I| = m$ (i.e. to contains m elements of S), then $S \setminus I$, the elements of S that are not in I, has same law as a sample of $k-m$ uniform random variables on $\mathscr{C} \setminus I$. Of course, the cardinality of $|S \cap I|$ has a binomial distribution $B(k, \text{Leb}(I))$.

■ **(B) A space eating principle.**

We will "eat" the space around u_0 at constant speed. Introduce two continuous non-decreasing, non-negative functions $\alpha(t), \beta(t)$, such that $\alpha(0) = \beta(0) = 0$, and

$$\beta(t) + \alpha(t) = t, \text{ for } t \in [0,1). \tag{1.21}$$

Again, we allow $((\alpha(t), \beta(t)), t \ge 0)$ to be a random process, independent of the u_i. For this, define $s[t]$, "the swallowed space at time t around u_0", as the length t interval on the circle,

$$s[t] := [u_0 - \alpha(t), u_0 + \beta[t]]_{\to}.$$

Now, let us consider the non-eaten variables u_j at time t:

$$X(t) := \{u_0, \cdots, u_{k-1}\} \setminus s[t].$$

At any time $t \in [0,1)$, the (remaining) non-eaten space $R(t) := \mathscr{C} \setminus s[t]$, satisfies

$$\text{Leb}(R(t) \setminus s[t]) = 1 - t;$$

each of the u_i is in $s[t]$ with probability t (except u_0 which is surely in $s[t]$), so that

$$|X(t)| \sim \mathsf{Binomial}(k-1, 1-t)$$

and conditional on $|X(t)| = j$, the non-eaten variables u_i are iid uniform in $R(t)$.

■ **(C) The Markovian dynamics of $|X(t)|$.**

It is immediate that

– The process $(|X(t)|, t \in [0,1])$ is a Markov process independent of $(\alpha(t), \beta(t))$,

– and the law of $(|X(t)|, t \in [0,1])$ is the same whatever is the process $(\alpha(t), \beta(t))$.

Both statements are valid provided that $\alpha(t) + \beta(t) = t$ for all t. Indeed, from time t to time $t + \mathrm{d}t$, the swallowed space, passes from a Lebesgue measure t to $t + dt$, and this is true whatever is the relative speed of $\alpha(t)$ and of $\beta(t)$.

Each element of $X(t)$ will survive up to time $t + \mathrm{d}t$ with probability $R(t + \mathrm{d}t)/R(t)$, and at time $t + \mathrm{d}t$, each surviving random variable (included in $X(t + \mathrm{d}t)$) will be uniform in $R(t + dt)$.

It is probably worth noticing that $|X(t)|$ is always a Markov process even if $(\alpha(t), \beta(t))$ is not!

– More than, that, we may allow $(\alpha(t + \mathrm{d}t), \beta(t + \mathrm{d}t))$ to depend on the number/identity of the swallowed u_j before time t (and also, of the absorption times): as long as at any time t, $(\alpha(t + \mathrm{d}t), \beta(t + \mathrm{d}t))$ is defined independently from the non-swallowed uniform points, the probability of absorption of a new point between time t and time $t + dt$ does not depend on the choice of the distribution of (α, β), which implies that $|X(t)|$ is Markovian: the distribution of $(|X(t)|, t \ge 0)$ is universal for this class of processes (α, β).

■ **(D) Relation between continuous dispersion model and space eating configurations**

In the previous paragraphs we talk about space eating model, and we let an interval $s(t)$ grows around u_0, as in the continuous diffusion processes. In a diffusion process in which the points u_1 to u_{k-1} have masses zero while u_0 is the only point having a positive mass, this mass would have diffused around, and would have covered a space $s(m_0)$ (with Lebesgue measure m_0), and covers/eats the other u_j in $s(m_0)$. The conditions we took for $(\alpha(t),\beta(t))$, namely unit speed deposition and independence with respect to $u_1,\cdots,u_{k-1}$, are the conditions for a valid continuous dispersion model. Here again, the last item of **(C)** applies again: as long as the space eaten dynamics depend only on the eaten points and on the present eaten space $s(t)$, the distribution of the final consideration say, at time m_0, does not depend on the details of the definition of $(\alpha(t),\beta(t))$.

■ **(E) On the disappearance of the space eaten.**

A topological solution to formalize the disappearance of space on $\mathscr{C}$ consists in identifying all the points of $s(t)$. This amounts to designing an equivalence relation $\sim_{s(t)}$ on $\mathscr{C}$ where $x\sim_{s(t)} y$ if $x=y$ or if x and y are both in $s(t)$. The quotient space $\mathscr{C}/\sim_{s(t)}$ is isomorphic to $\mathbb{R}/(1-t)\mathbb{Z}$. On this latter space, the non-eaten random variables are, conditionally on their number $b-1$, uniform on $\mathbb{R}/(1-t)\mathbb{Z}$, while the eaten interval $s(t)$ corresponds to a uniform point $u_0(t)$, independent of the other, a kind of scar.

We may then relabel by $v_0(t),\cdots,v_{b-1}(t)$ these b variables (with $v_0(t)=u_0(t)$, using the initial order of the labels between the surviving u_j to decide the identity 1 to $b-1$ of the vertices $v_1(t),\cdots,v_{b-1})$. If one is just interested in the joint position of the $v_j(t)$ on $\mathbb{R}/(1-t)\mathbb{Z}$, one sees that, conditionally on the event that they are b in total, they are distributed as b uniform point on $\mathbb{R}/(1-t)\mathbb{Z}$.

Letting $\ell_0,\cdots,\ell_{b-1}$ be the sequence of lengths of the intervals of $[\mathbb{R}/(1-t)\mathbb{Z}]\setminus\{v_i(t), 0\le i\le b-1\}$ formed by the $v_i(t)$, where these intervals are taken cyclically around $[\mathbb{R}/(1-t)\mathbb{Z}]$, starting from the intervals at the right [4] of $v_0(t)$, (conditionally on the fact that there are b points), then

$$\left(\ell_0,\cdots,\ell_{b-1}\right)\overset{(d)}{=}(1-t)\,[\mathbf{D}_0,\cdots,\mathbf{D}_{b-1}] \tag{1.22}$$

where $[\mathbf{D}_0,\cdots,\mathbf{D}_{b-1}]\sim\mathsf{Dirichlet}(1,\cdots,1)$.

On $\mathscr{C}$, if one interprets now $s(t)$ as a occupied interval as well as all the other $\{u_j\}$ that are not covered by $s(t)$, then, these m elements form $\mathbf{O}^{(k)}$, and the intervals in between (that are isomorphic to those in $\mathbb{R}/((1-t)\mathbb{Z})$), the free space lengths $|\mathbf{F}^{(k)}|$, and then are distributed as $(1-t)\,[\mathbf{D}_0,\cdots,\mathbf{D}_{b-1}]$.

■ **(F) Is it allowed in the CDM to place the points $(u_i, 0\le i\le k-1)$ beforehand?**

As long as the dispersion of the masses (m_i,u_i) do not depend on the presence of the points with higher index, they can be placed beforehand. Their presence allows to see clearly that when a single mass is dispersed (while the other ones are zeroes), the dispersion of this mass, and the arrivals of the others commutes, if we are only interested in the distribution of the final configuration.

This allows to disconnect somehow "the diffusion date" "i", from the arrival places u_i. The presence of points all together in the system allows to compare the effect of the diffusion of a small quantity of matter dm from one of this point or from another.

■ **(G) Space eaten around two points / CDM diffusion.**

Now assume that u_0 and u_1 have positive mass m_0 and m_1 while the other ones u_2 to u_{k-1} still have mass 0. Assume that u_0 is again equipped with two functions $(\alpha_0(t),\beta_0(t))$ satisfying $\alpha_0(t)+\beta_0(t)=t$ for all t, which defines an interval $s_0(t)=[u_0-\alpha_0(t)u_0+\alpha_0(t)]$. In the perspective "covering" u_0 has invaded $s_0(m_0)$ at time m_0 on $\mathscr{C}$, while on the space eaten perspective, $u_0(m_0)$ is now a point on $\mathbb{R}/((1-m_0)\mathbb{Z})$ and this point is maybe now identified with other u_i, including the other massive point u_1.

On the CDM perspective, when u_1 starts, its environment has changed. There are two cases: either u_1 is in $s_0(m_0)$, or it is not.

[4] we can start the labeling from any $v_i(t)$ as long as the labeling is chosen independently from the interval lengths

In both case, it will use some functions $\alpha_1(t), \beta_1(t)$, from its activation time, $t = 1$, to eat the space or to diffuse, according to the point of view. If u_1 is in $s_0(m_0)$ (this occurs with probability m_0 by the previous discussion), u_1 may use (random) processes $\alpha_1(t)$ and $\beta_1(t)$ depending on $(u_0, s_0(m))$, while if it is not in $s_0(m_0)$ (with proba. $1-m_0$), $\alpha_1(t)$ and $\beta_1(t)$ must be independent from $s_0(m_0)$ at least, before a possible collision.

Define $s_1(t) = [u_1 - \alpha_1(t), u_1 + \beta_1(t)]_{\to}$ and we assume now that $\mathrm{Leb}(s_1(t) \setminus s_0(t)) = (t-1)$ for $t \in [1, 1+m_1]$ so that s_1 grows at unit pace (out of $s_0(m_0)$).

Again, there are two points of view on the evolution of $s_1(t)$ as t grow: as a CDM on $\mathscr{C}$ already covered, on in $\mathbb{R}/((1-m_0)\mathbb{Z})$ where $u_0(m_0)$, the special point representing $s_0(m_0)$ behave as the others.

Let us now discuss the <u>free spaces</u> evolution if $s_1(t)$ evolves.

■ On the space eaten $\mathbb{R}/((1-m_0)\mathbb{Z})$,

– if u_1 is in $s_0(m_0)$, then the evolution of $s_1(t)$ after time 1 is indistinguishable <u>in distribution</u> of the evolution of s_0 after time 1 (if we would let $s_0(m_0)$ resume its eating activity after time 1).

– if u_1 is not in $s_0(m_0)$, in terms of the free spaces, it is the same! That is, eating the space around $u_0(m_0)$ or around u_1 is the same in law, since the surviving points are distributed as iid uniform on $\mathbb{R}/((1-m_0)\mathbb{Z})$!

In the perspective of the CDM, the idea is the same: when u_1 becomes active, either it is in $s_0(t)$ or it is not. The free spaces around $\mathscr{C}$ are distributed as Dirichlet $(1, \cdots, 1)$ (under the condition that there are m intervals if u_0 has covered all but $m-1$ other points). If u_1 is in $s_0(1)$, then in terms of free space, as we said before, only the size of $s_0(1+t)$ matters, not the way it grows, so that diffusing around s_0 or around s_1 is the same, in distribution for the eventual free space (at time $1+m_1$).

Conclusion

Assume that we are back to the initial models of several masses $(u_0, m_0), \cdots, (u_{k-1}, m_{k-1})$ and a valid CDM is given. Assume for a moment that we are only interested in the free spaces $\mathbf{F}^{(k)}$.

Then, on the space eaten point of view, during the relaxation phase of the ith point, the new mass extends around one of the surviving points, that are, conditional on their numbers, uniform points on $\mathbb{R}/(1-m_0 \cdots, m_{i-2})\mathbb{Z}$. The points around which the extension is done is not important, by invariance by relabeling at time i: everything would have been the same if all the space was eaten around the same point! Hence **Theorem 1.6**, is a consequence of this fact: when is only interested in free space, we can relocate the place where the relaxation is done, without any harm.

Remark 1.17. *A Corollary of this analysis is that if we would allow the relaxation phases to occur simultaneously, possibly with different speeds, as long as the procedure is invariant by rotation and independent between different occupied CC (before coalescence), then, the distribution of the free and occupied spaces would be the same (for any valid* CDM*).*

The distribution of free space would resist to even more involved generalizations, as for example, transfer of non-deposited masses – between current occupied CC.

Now, **Theorem 1.5** which deals also with occupied component need to be examined. Point (iii), (iv) and (v) can be stated as properties of the free CC only. They are consequences of the previous discussion (details for **Theorem 1.5**(v) are postponed in Section 1.10 below).

For example for point (iii), the fact that $\mathscr{L}(\mathbf{N}_k - 1) = \mathsf{Binomial}(k-1, R_k)$ comes from the fact that we can rearrange the masses as we want when we are only interested in the free CC sizes and number, and in the case in which the first $m_0 = W$, $m_1 = \cdots = m_{k-1} = 0$, the result is obvious.

Now, the distribution of the CC of the occupied set $\mathbf{O}^{(k)}$ depend clearly on the masses (in that respect having two masses 2 and 0 is different than to have two masses 0.8 and 1.2).

To prove (i) we may add to the previous discussion some recursive argument: first, the position of the CC $s_0(m_0)$ has a distribution independent of $(\alpha_0(t),\beta_0(t))$, and of $(u_1,\cdots,u_{k-1})$ (its position is uniform on $\mathscr{C}$). Now, assume that at time i^- (just before the arrival of the $i+1$th mass (u_i,m_i)), the current states $(\mathbf{O}^{(i)},\mathbf{F}^{(i)})$ has a distribution that does not depend on the choice of the valid CDM considered. There are two cases:
– either u_i is in one the CC $O\in\mathbf{O}^{(i)}$, or it is in a free space $F\in\mathbf{O}^{(i)}$. Then the relaxation process $\mathsf{IR}(t)=(\alpha_i(t),\beta_i(t))$ (with dependence already discussed) will rule the growing of $s_i(t)$ around u_i.

In both cases, we have exchangeability of the CC sizes (which is a property of their construction), and of the free space sizes.

Now, by the discussion above, the amount of space $s_i(t)$ will cover before collision with another CC does not depend on $(\alpha_i(t),\beta_i(t))$: in particular, the probability that a collision occurs before the end of the diffusion (at time $i+m_i$) is independent of $(\alpha_i(t),\beta_i(t))$; if a collision happens, then the time of the first collision, has a distribution independent of $(\alpha_i(t),\beta_i(t))$, and in this case, $s_i(t)$ will coalesce with a CC whose size has a law which does not depend on $(\alpha_i(t),\beta_i(t))$ (it has same distribution as if it was chosen uniformly among the other CC sizes!).

The arguments for the proof of (ii) have also be given: the sizes of the occupied CC depend on the masses that are diffused from each vertex, but not on the order of diffusion (see Remark 1.17).

To finish Theorem 1.11 is a consequence of the general exchangeability results obtained (universality of the probability of collision of the connected component that undergoes the relaxation phase during a time interval $\mathrm{d}t$, and in case of collision, universality of the distribution of the size of the occupied CC hit).

1.10 Proof of Theorem 1.5(v)

In the proof, we write $\mathsf{Binomial}(n,p)$ for a binomial r.v. with parameters n and p, different parameters implying independence (and then $B(p)=\mathsf{Binomial}(1,p)$ is a Bernoulli r.v.).
Markovian property of $(\mathbf{N}_j)$: We will prove that

$$\mathscr{L}\left(\mathbf{F}^{(i)}\mid(\mathbf{N}_j,0\le j\le i-1)=(n_j,0\le j\le i-1)\right) \quad=\quad \mathscr{L}\left(\mathbf{F}^{(i)}\mid \mathbf{N}_{i-i}=n_{i-1}\right)$$

which says that the complete history of $(\mathbf{N}_j,0\le j\le i-1)$ does not give additional information on $\mathbf{F}^{(i)}$ than $\mathbf{N}_{i-1}$ alone: the reason is that the free spaces are exchangeable (conditional on their number, as long as we start to label them independently from their sizes), and Dirichlet $(1,\cdots,1)$ distributed: this last property is preserved at each time step, whatever is the value $\mathbf{N}_i-\mathbf{N}_{i-1}$.

We know that $\mathscr{L}(\mathbf{N}_k-1)=\mathsf{Binomial}(k-1,1-M)$, conditioned on $\mathbf{N}_k=b$, the uniformly rotated free spaces $(|\mathbf{F}^{(k)}_{i+\theta \bmod b}|,0\le i\le b-1)$ where θ is uniform on $\{0,\cdots,b-1\}$, independently of the $|\mathbf{F}^{(k)}_i|$) are distributed as $R_k(\mathbf{D}_0,\cdots,\mathbf{D}_{b-1})$ (by Theorem 1.5(iv)).

Now, assume that $\mathbf{N}_k=n$, and $m_0,\cdots,m_k$ are given, with $W=W(m[k])=m_0+\cdots+m_{k-1}$, $W'=W(m[k+1])=m_0+\cdots+m_k$, and set $R=R_k=1-W$, $R'=R_{k+1}=1-W'$.

Now, for a Dirichlet vector $\mathbf{D}^{(b)}=(\mathbf{D}^{(b)}_1,\cdots,\mathbf{D}^{(b)}_b)\sim\mathsf{Dirichlet}(1,\cdots,1)$ denote by $\mathbf{S}^{(b)}_j=\mathbf{D}^{(b)}_1+\cdots+\mathbf{D}^{(b)}_j$ (so that $\mathbf{S}^{(b)}_0=0$ and $\mathbf{S}^{(b)}_b=1$). Observe that, for $x\in[0,1]$,

$$\mathbb{P}\big(\mathbf{S}^{(b)}_j\le x\le\mathbf{S}^{(b)}_{j+1}\big)=\binom{b-1}{j}x^j(1-x)^{b-1-j},$$

because the Dirichlet random variable with these parameters, can be seen as the distribution of the n intervals between the order statistics of $\hat{u}_1, \cdots, \hat{u}_{b-1}$ uniform random variables on $[0,1]$. We claim that

$$\mathbb{P}(\mathbf{N}_{k+1} = n+1 \mid \mathbf{N}_k = n) \quad = \quad R_k \mathbb{P}\big(\mathbf{S}_1^{(n+1)} \leq m_{k+1}/R_k\big)$$

and for $j \in \{0, \cdots, n-1\}$,

$$\begin{aligned} \mathbb{P}(\mathbf{N}_{k+1} = n-j \mid \mathbf{N}_k = n) &= R_k \mathbb{P}\left(\mathbf{S}_{j+1}^{(n+1)} \leq m_{k+1}/R_k \leq \mathbf{S}_{j+2}^{(n+1)}\right) + (1-R_k)\mathbb{P}\left(\mathbf{S}_j^{(n)} \leq m_{k+1}/R_k \leq \mathbf{S}_{j+1}^{(n)}\right) \\ &= R_k \binom{n}{j+1} \left(\frac{m_k}{R_k}\right)^{j+1} \left(1 - \frac{m_k}{R_k}\right)^{n-1-j} + (1-R_k)\binom{n-1}{j} \left(\frac{m_k}{R_k}\right)^{j} \left(1 - \frac{m_k}{R_k}\right)^{n-1-j}. \end{aligned}$$

2 Valid discrete dispersion models

We won't detail everything, but just insist on the difference with CDM. Here, the masses will arrive on the space $\mathscr{C}_n := \{0/n, \cdots, (n-1)/n\} \subset \mathscr{C}$ considered as a subgroup of $\mathbb{R}/\mathbb{Z}$ (equipped with the addition modulo 1). The masses are now positive multiple of $1/n$, and zero masses are still allowed.

To be valid a discrete dispersion model must satisfy the same conditions as the continuous ones, except that:

– The invariance by rotation condition (1.3) must hold only for the rotation by $1/n$ in $\mathscr{C}_n$,

– We add the condition, that eventually, after each interval relaxation, the occupied set $\mathbf{O}^{(k)}$ is still formed by close intervals (some of them having possibly length zero), with extremities on $\mathscr{C}_n$.

The notion of occupations is still the same as in the continuous case, and the right diffusion with constant speed is still allowed.

For example, if there are three masses arrival events $(m_0, u_0) = (3/n, 0)$, $(m_1, u_1) = (1/n, 5/n)$, $(m_2, u_2) = (0, 3/n)$, $(m_3, u_3) = (0, 7/n)$, then, using the RDCS: we have $\mathbf{O}^{(1)} = \big\{[0, 3/n]\}$, $\mathbf{O}^{(2)} = \{[0, 3/n], [5/n, 6/n]\}$, $\mathbf{O}^{(3)} = \mathbf{O}^{(2)}$, $\mathbf{O}^{(4)} = \{[0, 3/n], [5/n, 6/n], \{7/n\}\big\}$.

Again, the addition of a zero mass can leave the occupied set unchanged, or can create the apparition of a point (as stated in (1.15) and below). The distance between occupied CC is at least $1/n$. Notice also, that in the example we gave, the number of points in $\mathscr{C}_n$ in each component does not correspond (up to a factor n) to the Lebesgue measure of the CC, since there are some boundary effect.

2.1 Main universality result for valid discrete dispersion models

In order to state the analogue of Theorem 1.5 for valid discrete dispersion models (DDM), we need to define a discrete analogue to the Dirichlet distribution. If the total free space is R/n for some positive integer R, then there can be at most $m \leq R$ free CC, each of them with a positive length multiple of $1/n$.

Let (k, R) be two positive integers with $k \leq R$. We call composition of R in k parts a sequence $(s_1, \cdots, s_k)$ with positive integer coordinates, summing to R. Let $\mathsf{Comp}(R, k)$ be the set of these compositions. We have

$$|\mathsf{Comp}(R,k)| = \binom{R-1}{k-1} \tag{2.1}$$

since we can associate with each increasing sequence $(x_1, \cdots, x_{k-1})$ made of elements of $\{1, \cdots, R-1\}$, the sequence $(x_1, x_2 - x_1, \cdots, x_{k-1} - x_{k-2}, R - x_{k-1})$ which is an element of $\mathsf{Comp}(R, k)$ (and this correspondence is bijective).

Definition 2.1. *For all (k,R) integers such that $1 \le k \le R$, we call Discrete Dirichlet distribution of R in k parts (and write)* $\mathsf{DDirichlet}(R,k)$ *the uniform distribution on* $\mathsf{Comp}(R,k)$.

The following analogue of Theorem 1.5 holds in the discrete settings:

Theorem 2.2. *Let $n \ge 1$ be an integer. Consider a valid DDM A, and some deterministic masses $m[k] = (m_0, \cdots, m_{k-1})$ all non-negative multiples of $1/n$, satisfying the total weight condition $W(m[k]) < 1$. Let $R_k = 1 - W(m[k])$.*

- *(i) The law of $\left(\boldsymbol{O}^{(k)}, \boldsymbol{F}^{(k)}\right)$ is the same for all valid DDM A.*
- *(ii) The law $\left(\boldsymbol{O}^{(k)}, \boldsymbol{F}^{(k)}\right)$ is invariant under the permutation of masses.*
- *(iii) The number of free CC $\mathbf{N}_k^{(n)}$ has an explicit distribution $\nu(R_k, k)$ (see Section 2.2, where can also be found an analogue of Theorem 1.5 (v)).*
- *(iv) Take a uniform element θ in $\{0, \cdots, b-1\}$ (independently of everything). Conditional on $\mathbf{N}_k^{(n)} = b$, the (unbiased) free spaces sizes*
$$\left(\left|\boldsymbol{F}^{(k)}_{i+\theta \bmod b}\right|, 0 \le i \le b-1\right) \overset{(d)}{=} C[b] := (C_0, \cdots, C_{b-1}) \tag{2.2}$$
where $nC[b]$ follows the $\mathsf{DDirichlet}(nR_k, b)$ *distribution.*

The proof is similar to that of Theorem 1.5 (and can be proved using the same kind of principles). Again, an important feature allowing one to make precise computation is the following theorem :

Theorem 2.3. *Under the hypothesis of Theorem 2.2, the distribution of $\boldsymbol{F}^{(k)}$ only depends on k and the total mass $W(m[k])$:*
$$\mathscr{L}\left(\boldsymbol{F}^{(k)} \mid (m_0, \cdots, m_{k-1})\right) = \mathscr{L}\left(\overrightarrow{\boldsymbol{F}^{(k)}} \mid \big(W(m[k], \underbrace{0, \cdots, 0}_{k-1 \text{ zeroes}})\big)\right).$$

Discrete piling propensity

Lemma 2.4. *Let A be a valid DDM process on $\mathscr{C}_n$. Then, the number of occupied components when the masses $(m_0, \cdots, m_{k-1})$ are non-negative multiple of $1/n$ and satisfy $W(m[k]) < 1$, then, the discrete piling propensity is*
$$Q^{(n)}(m[k]) = \mathbb{P}(\mathbf{N}_k^{(n)} = 1) = \left(W(m[k]) + \frac{1}{n}\right)^{k-1}. \tag{2.3}$$
Moreover, for any $\ell \in \mathscr{C}_n$, by letting A_j be the starting point of the block j, we have,
$$\mathbb{P}\left(\mathbf{N}_k^{(n)} = 1, A_0 = \ell\right) = Q^{(n)}(m[k])/n. \tag{2.4}$$

Proof. Using the rotation invariance, Formula (2.4) and (2.3) are equivalent. We prove (2.4) which is a bit easier, from a combinatorial perspective.

We remove the normalization by n, and consider the masses $nm[k]$ instead, on $\mathbb{Z}/n\mathbb{Z}$ and also, first on $\mathbb{Z}/(M+1)\mathbb{Z}$ where $M = nW(m[k])$. If we use the RDCS on $\mathbb{Z}/(M+1)\mathbb{Z}$ with these masses, a single vertex remains free at the end, and with probability $1/(M+1)$, the last vertex M is free. This means if mass m_i is put at p_i in $\mathbb{Z}/(M+1)\mathbb{Z}$, among the $(M+1)^k$ such map p, a fraction $1/(M+1)$ of them leaves the last place empty. Hence, the number of ways to put the k masses on $\mathbb{Z}/n\mathbb{Z}$ such that, after relaxation (in the discrete RDCS model) the set (of classes) $\ell, \cdots, \ell + M - 1$ is occupied in $\mathbb{Z}/n\mathbb{Z}$ is also $(M+1)^{k-1}$. It suffices to multiply by $1/n^k$ which is the probability of each map p. $\square$

There exists also an analogous of Theorem 1.8 and Theorem 1.9:

Theorem 2.5. *Consider $m_0,\cdots,m_{k-1}$ some masses, non negative multiple of $1/n$, such that $W(m[k])<1$. For any $b\in\{1,\cdots,k\}$, any partition $(J_0,\cdots,J_{b-1})$ of the set $\{0,\cdots,k-1\}$ with non-empty parts, any sequence $(a_0,a_1,\cdots,a_{b-1})$ in $\mathscr{C}_n^b$ such that the $a_0,a_1,\cdots,a_{b-1}$ are cyclically ordered around $\mathscr{C}_n$, such that turning around the circle we get $a_0\preceq 0\preceq a_1\preceq\cdots\preceq a_{b-1}\preceq a_0\preceq 0\cdots$ and such that*

$$\mathrm{Leb}\left([a_i,a_{i+1 \bmod b}]_{\to}\right)>W(m(J_i))\quad \text{for } i\in\{0,\cdots,b-1\}.$$

For a valid DDM, we have

$$\mathbb{P}\left(\boldsymbol{A}_j^{(k)}=a_j,\boldsymbol{I}_j^{(k)}=J_j,0\le j\le b-1\right) = \prod_{j=0}^{b-1}\frac{Q^{(n)}(m(J_j))}{b}. \tag{2.5}$$

and for $M_0,\cdots,M_{b-1}$ be some non-negative masses.

$$\mathbb{P}\left(\boldsymbol{A}_j^{(k)}\in \mathrm{d}a_j,\left|\boldsymbol{O}_j^{(k)}\right|=M_j,0\le j\le b-1\right) = \mathsf{Nb}(m[k],M[b])\prod_{j=0}^{b-1}Q(M_j)\,\mathrm{d}a_j \tag{2.6}$$

where $\mathsf{Nb}(m[k],M[b])$ is the number of partitions $(J_0,\cdots,J_{b-1})$ in non-empty parts, such that, for all i, $W(m(J_i))=M_i$.

We have in the discrete case the analogue of Theorem 1.11:

Theorem 2.6. *Let $(m_0,m_1,\cdots,m_{k-1})$ be masses multiple of $1/n$ with sum in $[0,1)$. Then, the distribution of the process $(S(t),t\in\{0,\cdots,k\})$ define by $S(t):=\left\{\left\{\left|\boldsymbol{O}_m^{(t)}\right|,0\le m\le \mathbf{N}_j-1\right\}\right\}$ is the same for all valid DDM.*

Remark 2.7 (Reductio ad Right Diffusion with Constant Speed and excursion sizes)**.** *The same phenomenon occurs here: the law of $(\boldsymbol{O}^{(k)},\boldsymbol{F}^{(k)})$, for a fixed k is the same for all valid DDM, and then, can be computed for the simplest model.*

2.2 Distribution of the number of CC $\mathbf{N}_k^{(n)}$.

The formula for the distribution of $\mathbf{N}_k^{(n)}$ is more complex than in the continuous case (in which is was just a binomial), but still there is a simple representation, that we provide first.

Consider $u_0,\cdots,u_{k-1}$ iid uniform on $\mathscr{C}_n$ and consider the set

$$Z_k:=\{u_0,\cdots,u_{k-1}\}$$

of arrival places. Note, that in the case in which all the masses $m_0,\cdots,m_{k-1}$ equal zero, $\mathbf{O}^{(k)}=Z_k$.The support of the cardinality $|Z_k|$ is $\{1,\cdots,\max\{k,n\}\}$ and the distribution of set Z_k is computable, since for z a subset of $\mathscr{C}_n$ with cardinality $|z|>0$

$$\mathbb{P}(Z_k=z)=S(k,|z|)n^{-k}$$

where $S(a,b):=\sum_{j=0}^{b}(-1)^{b-j}\binom{b}{j}j^a$ is the number of surjections from $\{1,\ldots,a\}$ to $\{1,\ldots,b\}$. Hence, the law of $|Z_k|$ is also explicit

$$\mathbb{P}(|Z_k|=m)=\binom{n}{m}S(k,m)n^{-k},$$

and conditional on $\{|Z_k| = m\}$ (for $m \in \{1, \cdots, \max\{k, n\}\}$), the points in Z_k is a uniform subset of $\mathscr{C}_n$ with cardinality m.

Using Theorem 2.3, in order to describe $\mathbf{N}_k^{(n)}$ when the masses $m_0, \cdots, m_{k-1}$ have been treated with a valid DDM using the RDCS: it suffices to diffuse the mass $W(m[k])$ from u_0 and keep the other elements of Z_k unchanged (and observe how many of them are not covered). The diffusion from u_0 will reach u_0 and the $M := nW(m[k])$ next points of $\mathscr{C}_n$ on its right, so that

$$\mathbf{N}_k^{(n)} = 1 + |Z_k \backslash [u_0, u_0 + W(m[k])]_{\rightarrow}|.$$

and then, a random subset V of $Z_k \setminus \{u_0\}$ with size c remains free (not covered) with probability

$$\mathbb{P}(\mathbf{N}_k^{(n)} = 1 + c) = \sum_{m=1+c}^{k} \mathbb{P}(|Z_k| = m) \frac{\binom{n-1-M}{c}\binom{M}{m-1-c}}{\binom{n-1}{m-1}} \tag{2.7}$$

where, here, the classical factor $\frac{\binom{n-1-M}{c}\binom{M}{m-1-c}}{\binom{n-1}{m-1}}$ is the probability that a subset of $m-1$ elements in a set of cardinality $n-1$, c of them fall among into a prescribed subset of size $n-1-M$.

In order to describe the transition matrix, we again take $\mathbf{N}_k^{(n)} = b$, assume that $W = W(m[k])$.

Conditional on $\mathbf{N}_k^{(n)} = b$, the occupied connected are then constituted with b components, with total mass $W(m[k])$ and then contains a proportion

$$q_b := (nW(m[k]) + b)/n = W(m[k]) + b/n$$

of the elements of $\mathscr{C}_n$; denote by $p_b := 1 - q_b$ the proportion of $\mathscr{C}_n$ still free.

We add now a mass arrival event (m_k, u_k), and will proceed to the RDCS of this mass from u_k. First u_k belongs to $\mathbf{O}^{(k)}$ with probability q_b, and in $\mathbf{F}^{(k)}$ with probability p_b.

The free spaces up to a uniform rotation, after the insertion of u_k forms a uniform partition of $R_k := n(1-W(m[k]))-b$ in $\kappa := b+X$ parts, where $X = 1$ is a $\mathsf{Bernoulli}(p)$ random variable. Denote by $(\mathbf{D}_1^{\ell,R}, \cdots, \mathbf{D}_\ell^{\ell,R})$ be a $\mathsf{DDirichlet}(R, \kappa)$ random variable. Denote again $\mathbf{S}_j^{\ell,R} = \mathbf{D}_1^{\ell,R} + \cdots + \mathbf{D}_j^{\ell,R}$. We have

$$\mathbb{P}(\mathbf{N}_{k+1} = b+1 \mid \mathbf{N}_k = b) = p_b \mathbb{P}(\mathbf{S}_1^{b+1,R} < nm_k) \tag{2.8}$$

and

$$\mathbb{P}(\mathbf{N}_{k+1} = b-j \mid \mathbf{N}_k = b) = p_b \mathbb{P}(\mathbf{S}_{j+1}^{b+1,R} \leq nm_k < \mathbf{S}_{j+2}^{b+1,R}) + q_b \mathbb{P}(\mathbf{S}_j^{b,R} \leq nm_k < \mathbf{S}_{j+1}^{b,R}), \tag{2.9}$$

and the computation can be done since

$$\mathbb{P}(S_j^{\ell,R} \leq x < S_{j+1}^{\ell,R}) = \frac{\binom{x}{j-1}\binom{R-1-x}{\ell-j}}{\binom{R-1}{\ell-1}}$$

2.3 Examples of DDMs.

- **Parking model (or hashing with linear probing):** when masses all have weight $1/n$, if the dispersion policy is such that "all the mass is spread to the right" then the model is exactly the classical parking model (introduced in Konweiss and Weiss [20],studied for example by Chassaing and Louchard [9] (see also [10]). We will come back on this model in Section 3.2; the cost analysis of the hashing with linear probing (in term of total car displacement from their chosen place to their eventually park place), due the Flajolet & a. [14], and also to [9] are discussed in Section 4.1. However, our analysis yields that the distribution of the occupied CC (at any time k), of the parking does not depend on the valid CDM used. (Janson [17] noticed the same property for three different policies in which a car may eject an already parked car, this ejected car, searching a place on its right: for these three policies, if a car arrives in a block $[a,b]$, after insertion, the new block becomes $[a,b+1/n]$ (possibly merged with the next block).

- **Particle dispersion** The masses m_i have weights proportional to $1/n$, and are composed by particles of size $1/n$. The particles do successively independent random walks till they exit from the occupied interval in which they are.
- **Discrete caravan type model:** Masses are proportional to $1/n$ and undergo the RDCS so that the fill in progressively the empty slots at the right of the arrival position.

3 Asymptotics

3.1 Discussion regarding random masses results

Theorems 1.5 and Theorem 1.6 hold for any valid CDM, for any fixed masses $(m_0,\cdots,m_{k-1})$, as long as $\sum m_i < 1$. Let us assume here that the masses $\mathbf{m}[k] = (\mathbf{m}_0,\cdots,\mathbf{m}_{k-1})$ are now random and taken according to a distribution $\mu^{(k)}$ on $[0,1)^k$, such that

$$\mathbb{P}\big(W(\mathbf{m}[k]) < 1\big) = 1. \tag{3.1}$$

Take again some iid arrival positions $(\mathbf{u}_0,\cdots,\mathbf{u}_{k-1})$, independent from the masses.

Since conditional on the event $\{\mathbf{m}_i = m_i, 0 \le i \le k-1\}$, the results of Theorems 1.5 hold, in particular, $\mathbf{N}_k \stackrel{(d)}{=} 1+B$ where $\mathscr{L}(B \mid W(\mathbf{m}[k]) = W) = \mathsf{Binomial}(k-1, 1-W)$. The occupied and free block results of Theorem 1.9 can be extended to random block size by integration (with respect to $\mu^{(k)}$). This general statement is not sufficient to understand or compute the typical behavior of the block sizes for a given distribution $\mu^{(k)}$ when k become large; important "partially tractable" examples correspond to the case for which the combinatorial terms can be understood. Other important examples correspond to the case where the right diffusion with constant speed representations allows to bypass exact representations.

However, we will need additional hypothesis on the masses distribution to design limit theorems. We will come back on particular cases in Section 3.3.

The difference and similitude between discrete and continuous models will help us to design tractable models.

3.2 Discrete masses: comparison between continuous and discrete spaces

The aim of this section is to consider the cases where discrete dispersion models and continuous dispersion models meet: when masses are multiple of $1/n$. In this case, it is possible to define the right diffusion with constant speed on $\mathscr{C}$ and $\mathscr{C}_n$ respectively, and compare the statistics of the space occupation in both cases. As we will see, discrete dispersion model and continuous dispersion model defined using the same masses have some similarities, and some discrepencies.

An interesting case, is that of continuous model of parking, that we will call the **lazy bulldozer parking model**.

In the usual parking problem defined on the discrete parking $\mathscr{C}_n$ (see [9] or Section 4.1), some cars (with size $1/n$) arrive successively, and the ith car choose a uniform place c_i in $\mathscr{C}_n$. It then parked at the first available place among $c_i, c_i + 1/n \bmod 1, c_i + 2/n \bmod 1$, etc (the place where it parks is then not available for the subsequent cars). In terms of standard occupations, it occupies the interval $[\ell, \ell + 1/n]_{\rightarrow}$ where ℓ is the first empty place discovered.

The lazy bulldozer parking model is defined on the continuous parking $\mathscr{C}$. The bulldozers have length $1/n$ and choose uniform places (u_i) on $\mathscr{C}$ (where the u_i are iid). The ith bulldozer arrives when bulldozers 0 to $i-1$ are already parked. It observes the parking and locates the first free point at the right of u_i (and y may belongs to a free space smaller than $1/n$ as on Figure 5). Bulldozer then parks on $[y, y+1/n]_{\rightarrow}$, pushing if needed, to the right the bulldozers already on $[y, y+1/n]_{\rightarrow}$ if any, as on Figure 5. More precisely, if the interval $[y, y+1/n]_{\rightarrow}$ intersects an occupied block O, then, this components is translated on its right (and then O will possibly push the next occupied component, and this recursively, if needed), so that the length of O is preserves, and after this action by the bulldozer, O is at the right of $[y, y+1/n]_{\rightarrow}$ and adjacent to it.

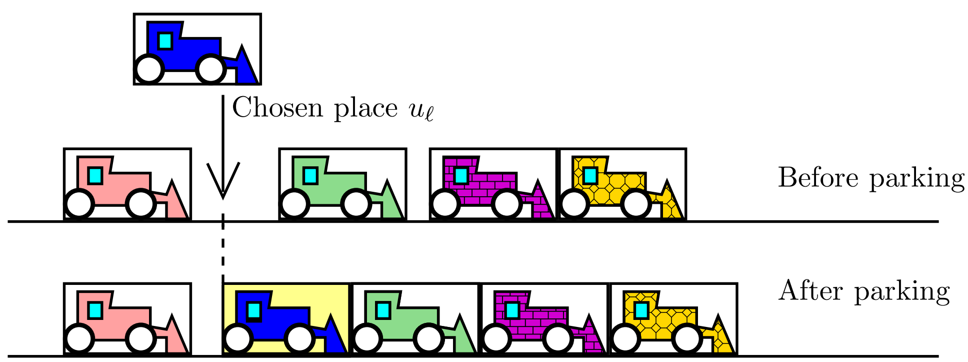

Figure 5: The bulldozer parks at the first free space after the chosen place and pushs what is present to so that to make free the needed place (if the chosen place u_ℓ is already occupied, it searches for the first free point y at the right of u_ℓ and park on $[y, y+1/n]_{\rightarrow}$, pushing what is needed to clear the place it wants.

– Consider the number $\mathbf{N}_k$ of blocks in the (continuous state) bulldozer model, and denote by $\mathbf{O}^{(k)}, \mathbf{F}^{(k)}$ the state at time k (with k cars of size $1/n$ treated)
– Consider the number $\mathbf{N}_k^{(n)}$ of blocks in the (discrete state) parking problems, and denote by $\mathbf{O}^{(n,k)}, \mathbf{F}^{(n,k)}$ the state at time k, with an additional exponent n, (with k cars of size $1/n$ treated).

By (1.12) and (2.3), if all the m_i are multiple of $1/n$, the discrete versus continuous piling propensities are very similar:

$$Q(m[n]) = \mathbb{P}(\mathbf{N}_k = 1) = W(m[k])^{k-1} \text{ and } Q^{(n)}(m[n]) = \mathbb{P}\left(\mathbf{N}_k^{(n)} = 1\right) = (1/n + W(m[k]))^{k-1}.$$

This similarity, as well as the resemblance between the bloc sizes Theorems 1.9 and 2.5 may give the intuition that discrete and continuous diffusion processes should behave very similarly... as we will see, this is only partially true.

A first discrepancy is that the support of $\mathbf{N}_k$ and $\mathbf{N}_k^{(n)}$ are different for large k. Indeed, if $n-k$ cars (masses) have parked (in the parking problem), the remaining space has size k/n. Since every free CC has size at least $1/n$, there can be at most k occupied CC in $\mathbf{O}^{(n,n-k)}$: hence the support of $\mathbf{N}^{(n)}_{n-k}$ is $[\![1,\min\{k,n-k\}]\!]$ (the number of CC is also smaller than the number of cars $n-k$).

In the bulldozer case, the free CC may have arbitrary small sizes, and then, there is a positive probability that every bulldozer is isolated, so the support of $\mathbf{N}_{n-k}$ is easily seen to coincide with $[\![1,n-k]\!]$.

We will give thinner comparison results in the rest of this section: the two models have huge similarities, and huge discrepancies, depending on the observed statistics.

- The similarities concern large occupied CC,
- The discrepancies concern small occupied CC, and also, the number of occupied CC.

While comparing the two models, we focus on the phase transition that occurs at time

$$t_n(\lambda) := \lfloor n - \lambda\sqrt{n} \rfloor,$$

for some $\lambda > 0$, at which t_n vehicles/masses have been treated (this phase transition has been identified in [9]). For any $k \geq 1$, and any integer $t \in \{0,\cdots,n\}$, set

$$\begin{aligned} L^{(n),\downarrow}[t,k] &:= (L_i^{(n)}, 1 \leq i \leq k), \\ L^{\downarrow}[t,k] &:= (L_i, 1 \leq i \leq k) \end{aligned}$$

respectively, the Lebesgue measures of the k largest occupied CC of $\mathbf{O}^{(n,t)}$ in the discrete parking and of $\mathbf{O}^{(t)}$, in the continuous parking, both lists being sorted in decreasing order, and completed by zeroes, if there are less than k CC.

Proposition 3.1. *In the discrete parking,*

$$\frac{\mathbf{N}^{(n)}_{t_n}}{\sqrt{n}} \xrightarrow[n]{(proba.)} \lambda(1-e^{-1}) \tag{3.2}$$

while, in the continuous parking

$$\frac{\mathbf{N}_{t_n}}{\sqrt{n}} \xrightarrow[n]{(proba.)} \lambda. \tag{3.3}$$

For any k, the following convergence in distribution holds in $\mathbb{R}^k$,

$$n^{-1}\,L^{(n),\downarrow}[t_n(\lambda),k] \xrightarrow[n]{(d)} (\ell^{\downarrow}_i(\lambda,\mathsf{e}), 1 \leq i \leq k) \tag{3.4}$$

$$n^{-1}\,L^{\downarrow}[t_n(\lambda),k] \xrightarrow[n]{(d)} (\ell^{\downarrow}_i(\lambda,\mathsf{e}), 1 \leq i \leq k) \tag{3.5}$$

where $(\ell^{\downarrow}_i(\lambda,\mathsf{e}), 1 \leq i \leq k)$ is the (random) vector formed by the k largest excursion sizes of the process $t \mapsto (\mathsf{e}_t - \lambda_t) - \min_{s\leq t}(\mathsf{e}_s - \lambda s)$, sorted in descending order (where e is a normalised Brownian excursion).

Remark 3.2. *In Bertoin [5, Lemma 5.11], the limiting fragment sizes sorted in decreasing order are described as the ranked atoms $(a_1,a_2,\cdots)$ of a Poisson point measure on $(0,+\infty)$ with intensity $\frac{\lambda}{\sqrt{2\pi a^3}}da$, conditioned by $\sum_{i\geq 1} a_i = 1$. In Aldous & Pitman [3, Theo.4, Cor. 5], Chassaing & Louchard [9, Theo. 1.5], description of the the law of the blocks when sorted according to a size bias order are given.*

Remark 3.3. *Since there can be only a finite number of blocks of size $\geq xn$ for any given x, the convergence stated in the theorem is equivalent to the vague convergence on $(0,1]$ of the measure $\sum_{O\in\mathbf{O}^{((k))}} \delta_{\mathrm{Leb}(O)/n}$ to the sum of the Dirac masses at the excursion lengths of $(\mathsf{e}_t - \lambda_t) - \min_{s\leq t}(\mathsf{e}_s - \lambda s)$.*

Remark 3.4. *The discrepancy between the distribution of the occupied domain in the discrete and continuous cases comes from the combine effect of the piling propensity and the "spacing propensity" which is a measure of the set of non-intersecting positioning of CC around the circle:*

▪ *in the continuous case, the measure of the space of the possible distance between b occupied blocks when the remaining space is* R*, is*

$$S(b,R) = R^{b-1}/(b-1),$$

independently from the masses, since this is the measure of the simplex $\{(x_1,\cdots,x_b)\in[0,1]^b, \sum x_i = R\}$*,*

▪ *in the discrete case, the number of ways of sharing k (unit) remaining spaces (each of them being of Lebesgue measure* $1/n$*), into b non-empty (integer) blocks is (as explained in (2.1))* $S^n(b,R) = |\mathsf{Comp}(nR,b)| = \binom{nR-1}{b-1}$*.Of course, the two formulas of* $S(b,R)$ *and* $S^n(b,R)$ *have to be included in a larger picture – this is the goal of Theorems 1.9 and 2.5 – but the fact that theses quantities do not have the same support, and are not (close to) 'proportional" is the main message here.*

Proof of Proposition 3.1. Proof of (3.3): We have $\mathbf{N}_{t_n(\lambda)}-1$ follows the binomial distribution with parameter $(t_n(\lambda)-1, 1-t_n(\lambda)/n) = (t_n(\lambda)-1, \lambda/\sqrt{n})$. Hence $\mathbf{N}_{t_n(\lambda)}/\sqrt{n}$ has mean $\lambda + o(1)$ and variance $\lambda O(1/\sqrt{n})$, so that (3.2) follows by Bienaymé-Tchebichev inequality.

Proof of (3.2): By (2.3), the number of occupied CC would be the same if we would place a "big" mass of size $t_n(\lambda)/n = (1-\lambda/\sqrt{n})$ (which by invariance by rotation can be placed wherever we want, say on $[0, t_n(\lambda)/n]$, and $(t_n(\lambda)-1)$ masses of size 0. In this case $\mathbf{N}^{(n)}_{t_n}$ would be 1 (big block) plus, the number of sites in $S_n := [t_n(\lambda)/n + 1/n, \cdots, (n-1)/n] \cap \mathscr{C}_n$, receiving at least one mass. A given site in $\mathscr{S}_n$ receives a binomial $B(t_n(\lambda), 1/n)$ number of zero masses (this converges to a Poisson(1) r.v., in distribution), from what we deduce easily that since $|S_n| \sim \lambda\sqrt{n}$, $\mathbb{E}(S_n) \sim \lambda\sqrt{b}(1-\exp(-1))$, and is approximately a Poisson distributed r.v.. To get (3.2), a concentration argument (relying on a second moment methods, suffices to conclude).

We give just some elements on (3.4), (3.5), for which variants exist in the literature. First, to prove the asymptotic proximity between both largest occupied CC, the idea is to construct on the probability space on which is defined the continuous space parking, with the data $((1/n, \mathbf{U}_i), 0 \le i \le t_n(\lambda))$ of mass arrival events, a discrete counterpart, by taking

$$\mathbf{U}^{(n)}_i := \lceil n\mathbf{U}_i \rceil / n, \tag{3.6}$$

as location of the ith arrival. Hence $\mathbf{U}^{(n)}_i$ is a bit on the right of the continuous one $\mathbf{U}_i$, but, on interval of the form $(k/n, (k+1)/n]$ for some k, the number of arrivals in the discrete and continuous parking coincides perfectly. We construct the collecting paths (already seen in (1.20)) associated with both parkings: Set, for two discrete time t,

$$S^{(n)}(t,x) = -x + \sum_{i\le t} \mathbb{1}_{\mathbf{U}^{(n)}_i \le x}/n, \tag{3.7}$$

$$S(t,x) = -x + \sum_{i\le t} \mathbb{1}_{\mathbf{U}_i \le x}/n. \tag{3.8}$$

The term $-x$ encodes the deposition on $[0,x]$ while $\sum_{i\le t} \mathbb{1}_{\mathbf{U}_i\le x}/n$, encodes the total mass arrived on the same interval. The erosion is encoded even when there are not mass to erode; this method has the advantage to make appear under the form of excursions above the current minimum, the occupied components. These collecting paths [5] encode the occupation of the parking (the idea of this coupling is bor-

[5] The continuous version $S(t,x)$ has been represented on Figure 3, and the other is defined similarly, up to the displacement of the masses u_i to $\lceil nu_i \rceil/n$, that is, on the next element of $\mathscr{C}_n$ at the right of u_i

rowed to Chassaing & Marckert [10] and can be found also in Bertoin & Miermont [6, Section 5.]). Now, there are essentially two methods to conclude:

• either we prove that $\sqrt{n}S^{(n)}(t_n(\lambda),.)$ converges in distribution in $C[0,1]$ to b_λ a Brownian bridge conditioned to end at $b_\lambda(1) = -\lambda$ (in terms of a classical linear Brownian motion, $(b_\lambda(t), t \in [0,1]) \overset{(d)}{=} (B_t - tB_1 - \lambda t, t \in [0,1])$, and we recover the length sizes of the occupied CC by taking into account Figure 3: in words, we need to define a rotate version of $S^{(n)}$ with respect to $a_n = \min \operatorname{argmin}(S^{(n)}(t_n(\lambda),.)$ (which amounts in Figure 3 to put the new origin at $(a_n, m_n(\lambda))$ (where $m_n(\lambda) = \min\{S^{(n)}(t_n(\lambda), y), y \in [0,1])$ and consider in this new basis, the obtained length 1 trajectory. Technically, this corresponds to set $\overline{S}^{(n)}(t_n(\lambda),.) = \mathsf{Rot}(a_n, S^{(n)}(t_n(\lambda),.)$ where the rotation is defined more generally, for all function f and $a \in [0,1]$ by

$$\begin{aligned}\mathsf{Rot}(a,f)(x) &= f(a+x) - (\min f) && \text{for } x \in [0,1-a],\\ &= f(1) - (\min(f)) + f(x-(1-a)) && \text{for } x \in [1-a,1].\end{aligned}$$

Then to complete the proof by the Skhorokhod representation theorem, it is possible to find a probability space, containing copies of $S^{(n)}(t_n(\lambda),.)$ and b_λ (still denoted by the same names in the sequel), such that $\sqrt{n}\overline{S}^{(n)} \xrightarrow[n]{(as.)} b_\lambda$, and since b_λ reaches is minimum a.s. once (says in a), and then, one has also on this space $a_n \xrightarrow[n]{(as.)} a$, and then $\sqrt{n}\overline{S}^{(n)}(t_n(\lambda),.) \xrightarrow[n]{(d)} \mathsf{Rot}(a, b_\lambda)$. Now the fact that $\mathsf{Rot}(a, b_\lambda)$ is distributed as $(e_x - \lambda x, x \in [0,1])$ is known (see e.g. Varin [31, Sec. 2.6]), and the convergence of the excursion sizes above the current minimum toward those of the limit is due to Aldous [2, Lemma 7].

• Or we may use the method of [9] which consists in working on the complete parking conditioned to have the last vertex free when $n-1$ cars are parked (which has no incidence on the distribution of the occupied CC sizes at time $t_n(\lambda)$). In this case the collecting path $S^{(n)}(n-1,.)$ converges in distribution to the Brownian excursion e and,

$$\left(\sqrt{n}\,S^{(n)}(t_n(\lambda),x), 0 \le x \le 1\right) \xrightarrow[n]{(d)} (\mathsf{e}(x) - \lambda x, 0 \le x \le 1) \tag{3.9}$$

where the convergence holds in $C[0,1]$, as well as the joint law for a finite number of λ_j (the proof is [9], but other proofs can be given, e.g. using the argument in [25], and the convergence, as a function of (t,x) has also been given in [8], as recalled further in the paper, in (5.6)). From here, again by Aldous [2, Lemma 7], we get (3.4)

Now, it remains to explain why in continuous parking version, (3.5) holds, that is, why the result are the same in discrete and continuous parking, regarding large excursion sizes. We have, for a fixed λ,

$$\sqrt{n}\left(S^{(n)}(t_n(\lambda),.), S(t_n(\lambda),.)\right) \xrightarrow[n]{(d)} (b_\lambda, b_\lambda) \tag{3.10}$$

for the Skorokhod topology on $D([0,1],\mathbb{R}^2)$.We have already explained the convergence of the first marginal, so it suffice to explain why the uniform distance between the two processes goes to zero in probability (and this has already been proved in in Chassaing & Marckert [10] and is simple: the two processes $(S(t,x), x \in [0,1])$ and $(S^{(n)}(t,.), x \in [0,1])$ coincides at the points $x = k/n$ (with k integer). Let $R(k,t) = \{i : i \le t, U_i \in [k/n,(k+1)/n)\}$ be the binomial distributed number of cars arrived in $[k/n,(k+1)/n)$ at time t, and clearly

$$\sup_{x\in[0,1]} \left|S^{(n)}(t,x) - S(t,x)\right| \le \max_{0\le k\le n-1} R(k,t)/n \le \max_{0\le k\le n-1} R(k,n)/n,$$

which is then the corresponding value in the completely full parking. Now, it is well known that there exists a constant $C > 0$ such that

$$\mathbb{P}\left(\max_{0\le k\le n-1} R(k,n) \ge C\log n\right) \le n\mathbb{P}\left(R(1,n) \ge C\log n\right) \le 1/n$$

for C large enough.)

Again, Aldous [2, Lemma 7] allows to deduce the convergence of the second marginal in (3.10).

□

3.2.1 Extension to more general discrete masses model

The general message, here, is that if we are given a model of masses $(\mathbf{m}_i, 0 \le i \le t_n(\lambda))$, where the $\mathbf{m}_i$ are multiple of $1/n$ (the $\mathbf{m}_i$ being random or not), taken at some time $t_n(\lambda)$, random or not[6]: again the collecting paths $S^{(n)}(t,.)$ and $S(t,.)$ associated with the discrete and continuous models as defined in (3.7) are some tool to consider first.

For all models for which the range of $S^{(n)}$ is typically to be of larger order than $1/n$ (globally), the large excursions of $S^{(n)}(t,.)$ and $S(t,.)$ above their current minima will be asymptotically identical (and this can be proved using the arguments in the proof of Proposition 3.1 together with Aldous [2, Lemma 7]):

Theorem 3.5. *Consider a sequence of DDM such that, for some sequences of times* $(t_n, n \ge 0)$

$$\left(\alpha_n S^{(n)}(t_n, x), x \in [0,1]\right) \xrightarrow[n]{(d)} \left(S^{\infty}(x), x \in [0,1]\right) \tag{3.11}$$

where $\alpha_n/n \to 0$ *and* $\alpha_n \to +\infty$*, with* $S^{\infty}(1) < 0$ *and where the convergence holds in* $C[0,1]$*. If* S^{∞} *reaches its minimum a.s. once on* $[0,1]$ *at some* $a \in [0,1]$*, then for* $S = \mathsf{Rot}(a, S^{\infty})$*, the sequence formed by the* k *largest clusters converges in distribution to the length of the excursions of* S *above its current minimum*

$$n^{-1} L^{(n),\downarrow}[t_n, k] \xrightarrow[n]{(d)} (\ell_i^{\downarrow}(S), 1 \le i \le k) \tag{3.12}$$

$$n^{-1} L^{\downarrow}[t_n, k] \xrightarrow[n]{(d)} (\ell_i^{\downarrow}(S), 1 \le i \le k) \tag{3.13}$$

• We could have stated convergence for the vague topology instead.

Both trajectories $S^{(n)}(t,.)$ and $S(t,.)$ coincides at the discrete points $(k/n, 0 \le k \le n)$, and moreover the evolution of the $-x$ term is $-1/n$ on such an interval. A small figure is sufficient to see that, if $\alpha_n S^{(n)}(t,.)$ converges in distribution to a limit in $D[0,1]$, then, if $\alpha_n = o(n)$, the largest excursion lengths of $S^{(n)}$ and of S above their current minimum should be close asymptotically. The condition $\alpha_n = o(n)$ is needed, because if $S^{(n)}$ has order $1/n$, then the block sizes are expected to be very small (of order $1/n$), and the asymptotic behavior of $\alpha_n S^{(n)}(t,.)$ would not be the right tool to study them. Again, as exemplified in the parking case, statistics concerning small occupied CC are expected to be different in DDM and CDM.

3.3 Random models (and caravans)

By Theorem 1.5, Theorem 1.8, Theorem 1.11, Theorem 2.2, Theorem 2.5, and Theorem 2.6 the study of the occupied CC of all valid continuous or discrete diffusion processes reduced to that of right diffusion with constant speed, which are called "caravan" in Bertoin & Miermont [6] (up to the open/close representation of the occupied space, and the processing of zero masses, see Remark 1.4).

[6] when the masses are taken as iid random variables, to stop the diffusion process just before the space overflow, that is, at time $\tau - 1$ with $\tau = \inf\{t : m_0 + \cdots + m_{t-1} > 1\}$, provides a random time as which it is natural to stop the diffusion process

3.3.1 Bertoin & Miermont results about caravans

In order to study the asymptotic behavior of valid CDM or DDM when the number of masses go to $+\infty$, some hypothesis on the masses need to be done. The case in which all masses have the same size correspond to parking [7] (of cars, or bulldozers, according to the fact that the space is continuous or discrete) and was discussed in Section 3.2.

When the number t_n of masses go to $+\infty$ (with sum in $[0,1)$), and the masses are random, then the collecting path $W(t,.)$ will have a limiting distribution under regularity assumptions on the masses (for example, the fact that they are iid, have some moments, or a regular tail).

Bertoin & Miermont [6] consider masses $(m_i, i \geq 0)$ that comes from some normalized iid random variables $(l_i, i \geq 0)$ with finite expectation $\mathbb{E}(\ell) = \mu_1 < +\infty$ and either finite second moment ($\mu_2 = \mathbb{E}(\ell^2)$, in this case, ℓ is said to be in $\mathscr{D}_2$), or have a regular tail:

$$\mathbb{P}(\ell > c) \underset{x \to +\infty}{\sim} c x^{-\alpha} \tag{3.14}$$

for some $c \in (0,+\infty)$ and $\alpha \in (1,2)$. In these cases, ℓ is in the domain of attraction of spectrally positive stable distribution with index α, and said to be in $\mathscr{D}_\alpha$). Then, they consider for (small) $\varepsilon > 0$,

$$\tau_{1/\varepsilon} = \inf\{i : \ell_0 + \cdots + \ell_i > 1/\varepsilon\} \tag{3.15}$$

then, they work on a circle of size $1/\varepsilon$, with these masses, but in terms of the present paper, this amounts to taking $k = 1 + \tau_{1/\varepsilon}$ masses

$$m_i = \varepsilon \ell_i, \text{ for } i < \tau_{1/\varepsilon}$$

and with the last mass taken to complete "1", that is $m_{\tau_{1/\varepsilon}} = 1 - W(m[\tau_{1/\varepsilon}])$. They then prove the convergence of the associated collecting path $(1/\varepsilon)^{1/\alpha}(S_t, 0 \leq t \leq 1)$ toward an analogous of the Brownian bridge with a linear drift for stable processes, the standard stable loop: see definition in Bertoin & Miermont [6], in formula (2) and around. They deduce from this the finite dimensional convergence in distribution, of the sizes of the largest occupied CC in their Theorem 1, in a critical time window defined in terms of α).

Additional discussion concerning discrete masses arriving on the discrete circle $\mathbb{Z}/n\mathbb{Z}$ are provided in their Section 5, notably a coupling equivalent to (3.7) is introduced for the description of the limiting discrete parking in terms of the continuous one.

3.3.2 Other conditioned tractable models

We can imagine two other classes of models of masses that lead to asymptotic tractable models:
(1). Iid random masses (m_i) conditioned by their sums. Their numbers K_n can be random or not.
(2). Prescribed deterministic masses (m_i).

In case (2), in order to obtain convergence of the collecting paths, some regularities are needed. For example, one can fix in the nth model, the proportion $p_j(n)$ of masses equal to $w_j(n)$, and then demands some regularities for $p_j(n)$ and $w_j(n)$ (for example, require that $p_j(n) \to p_j$ for each j, where $\sum p_j = 1$, $\sum j p_j = 1$, and $n w_j(n)$ converges for j fixed). In words, we prescribe the proportion of masses of each type, and require some convergence (in general additional assumption are needed to get a limiting behavior, if the number of possible masses type "j"goes to $+\infty$), see e.g. [7].

[7] even if the sizes of the cars is $\varepsilon > 0$ and is not of the form of $1/n$, one may take a discretization of $\lfloor 1/\varepsilon \rfloor$ parking places with an additional "incomplete place" which clearly, can be ignored in the asymptotic range of Section 3.2

To illustrate the case (1) without developing too much these considerations, we may for example, construct the masses using iid random variables (X_i) taking their values in $\mathbb{N}$ (this simplifies the condition), with common distribution μ, mean $m = 1$ and variance $\sigma^2 \in [0, +\infty)$, and take as mass

$$m_i = X_i/n;$$

for this choice, the space is naturally totally occupied at time K_n close to n (with $\sqrt{n}$ fluctuations), and we may then require that, for example at time $K_n = n - a\sqrt{n}$ the free space is $b/\sqrt{n}$.

We propose instead to require that it is totally full at time n exactly, which will allows to let the time go backward in **??** without getting lost in the change of time/variable, without modifying the nature of the phenomenon in play. We will then condition by the event

$$\mathcal{E}_n := \left\{\sum_{i=0}^{n-1} m_i = 1\right\} = \left\{\sum_{i=0}^{n-1} X_i = n\right\}.$$

When one writes the final collecting path

$$S_n^{(n)}(x) = -x + \sum_{i=0}^{n-1} (X_i/n)\, \mathbb{1}_{u_i \le x}$$

one sees that there is a binomial number $B(n, x)$ masses that falls in $[0, x]$ (and if one is interested in the multidimensional distributions $S(x_i) - S(x_{i-1})$ where $x_0 = 0$ and $x_k = 1$, then one notices that

$$\begin{aligned} \big(|\{i : u_i \in (x_{j-1}, x_j]\}|, 1 \le j \le k\big) &=: (\Delta_j(n), 1 \le j \le k) &(3.16)\\ &\sim \mathsf{Multinomial}\big(n, (x_j - x_{j-1}, 1 \le j \le k)\big). &(3.17)\end{aligned}$$

This means that, in distribution, since the X_i are exchangeable, if one defines

$$W_n(k) = \sum_{j=1}^{k} X_j/n, \quad k \in \{0, \cdots, n\}, \tag{3.18}$$

one has

$$\begin{aligned} \sqrt{n}\big(S_n(x_j), 1 \le j \le k\big) &\overset{(d)}{=} \sqrt{n}\big(-x_j + W_{\Delta_1(n)+\cdots+\Delta_j(n)}, 1 \le j \le k\big) &(3.19)\\ &= \sqrt{n}\left(-x_j + \frac{\Delta_1(n)+\cdots+\Delta_j(n)}{n}, 1 \le j \le k\right) &(3.20)\\ &\quad + \sqrt{n}\left(W_{\Delta_1(n)+\cdots+\Delta_j(n)} - \frac{\Delta_1(n)+\cdots+\Delta_j(n)}{n}, 1 \le j \le k\right). &(3.21)\end{aligned}$$

We have
– If $\sigma^2 = 0$, then $X_i = 1$ a.s. (this is the parking case), $W_n(k) = k$, so that the term in (3.21) vanish, and $\sqrt{n} S_n(.)$ converges to the Brownian bridge in distribution.
– If $\sigma^2 > 0$, under the hypothesis we have on X_i, $(\sqrt{n}(W_n(nt) - t)/\sigma, t \in [0,1])$ converges in distribution in $C[0,1]$ to the Brownian bridge (we assume here that W_n is interpolated linearly between integer positions). Since $\frac{\Delta_1(n)+\cdots+\Delta_j(n)}{n} \xrightarrow[n]{(proba.)} x_j$, Refs?

$$\sqrt{n}\left(W_{\Delta_1(n)+\cdots+\Delta_j(n)} - \frac{\Delta_1(n)+\cdots+\Delta_j(n)}{n}, 1 \le j \le k\right) \xrightarrow[n]{(d)} \sigma(b_{x_1}, \cdots, b_{x_k}).$$

while, by standard limit theorem on multinomial random variables

$$\sqrt{n}\left(-x_j+W_{\Delta_1(n)+\cdots+\Delta_j(n)},1\leq j\leq k\right)\xrightarrow[n]{(d)}\left(\tilde{b}_{x_j},1\leq j\leq k\right)$$

where b and $\tilde{b}$ are two independent Brownian bridges. Hence,

$$\sqrt{n}S_n(.)\xrightarrow[n]{(d)}\sqrt{1+\sigma^2}b$$

in $C[0,1]$ [8].

Under this condition, it is possible to go back in time to see the appearance of the giant components: in words, the last $\lambda\sqrt{n}$ masses m_i have a deterministic effect on the collecting path, because their contribution to the total mass is $\lambda/\sqrt{n}$, and the repartition on $[0,1]$ is "almost uniform", because of the concentration of the binomial distribution, so that

Proposition 3.6. *For all compact interval* $\Lambda\subset[0,+\infty)$

$$\left(\sqrt{n}S_{n-\lambda\sqrt{n}}(x)-\sqrt{n}S_n(x),0\leq x\leq 1,\lambda\in\Lambda\right)\xrightarrow[n]{(d)}(-\lambda x,0\leq x\leq 1,\lambda\in\Lambda) \tag{3.22}$$

where the convergence holds in $C([0,1]\times\Lambda,\mathbb{R}^+)$*, and*

$$\left(\sqrt{n}S_{n-\lambda\sqrt{n}}(x),0\leq x\leq 1,\lambda\in\Lambda\right)\xrightarrow[n]{(d)}(\sqrt{1+\sigma^2}b_x-\lambda x,0\leq x\leq 1,\lambda\in\Lambda) \tag{3.23}$$

for the same topology.

Proof. The proof of (3.22) here is simple because the convergence of the FDD is a consequence the concentration of a sum of random variables, and the tightness comes from the fact that $(t,x)\mapsto\sqrt{n}S_{n-t\sqrt{n}}(x)$ is monotonous in λ and in x and the limit is continuous (in the deterministic case, point-wise convergence toward a continuous functions is uniform (on compact sets) as soon as the functions are monotonous. [More details are given in [25], for "a similar correction phenomenon", and additional elements on topology are given in the Appendix there]. □

In Aldous [4, p.168-169], general criteria for tightness of "auto-normalized" processes with exchangeable increments (as those of the two previous examples) are given under the hypothesis that the sum of the increments are zero and sum of their square is 1. This has to be adapted a bit here since we want limiting trajectories to end at a negative height (while regarding this height as the rescaled limit of the number of occupied components).

In terms of excursion length above the current minima, the limit are then the same as for the parking process.

4 Energy dissipation and other cost associated with dispersion model

This section attacks a natural question related to dissipation models: Assume that a mass m_k arrives at some point u_k in an already partially occupied space $\mathscr{C}$. In a physical system, or in an abstract data

[8] The Brownian bridge is a centered Gaussian process with covariance function $\mathrm{cov}(b_s,b_t)=s(1-t)$ for $0\leq s\leq t\leq 1$. Hence, the sum $\overline{b}:=c_1b+c_2\tilde{b}$ for two independent Brownian bridges b and $\tilde{b}$, and two constants c_1,c_2 gives as a covariance function $\mathrm{cov}(\overline{b}_s,\overline{b}_t)=(c_1^2+c_2^2)(s(1-t))$

structure (as are parking models, or hash tables), the mass – which can be composed of smaller particles – while it undergoes its dissipation process (until occupying a new space $\mathbf{O}^{(k)}\backslash\mathbf{O}^{(k-1)}$), have(s) moved or flow, or maybe has loss some potential energy, or been pushed along a given distance, or be submitted to some other physical process during a particular time. In a lot of situations, the main quantities of interest in a diffusion process are related to these additional considerations.

Let us assume that one can associate a "unitary cost" (a real number) to the dissipation of (m_k, u_k). The "global cost" at time t will then be defined as the sum of the unitary costs of the first t masses. For example, in the RDCS, a mass m can be seen as composed of tiny masses $\mathrm{d}m$, and if a mass m arrives at a point u inside $\mathbf{O}^{(k)}$, if one assumes that transporting this mass $\mathrm{d}m$ at distance x has a cost $x\mathrm{d}m$, then the resulting cost of the displacement is

$$\int_{\mathbf{O}^{(k+1)}\backslash\mathbf{O}^{(k)}} \overrightarrow{d}(u,y)\mathrm{d}y$$

(recall notation given in Section 1.4).

The problem we face is that the arrival of a single mass, on a CC O of $\mathbf{O}^{(k)}$ can result in the coalescence of many occupied CC; in general it is quite intricate to study, and even to define such a cost model, since we allow CDM that depends on the current occupied interval during the dispersion.

The simplest models are the DDM in which the only masses have size $1/n$ on $\mathscr{C}_n$ (for any valid DDM since when the mass is fixed, the law of the block sizes process is fixed too (Theorem 1.11), so that, only the unitary cost will be important). These models are simpler because the dispersion of the kth mass stops when it gets out of the CC of $\mathbf{O}^{(k)}$ in which it arrived (on its left or on its right), so that the global cost is a function of the collections of arrival block sizes $(|B_j|, j \leq t)$.

In this case, the cost is entirely determine by the "unitary cost model" which specifies the unitary cost distribution, for a single car arriving in an interval of length ℓ, for all ℓ. These kind of models have been studied for "simple cost models", for analysis of algorithm purpose mainly.

We first discuss in length this case, and will come back for partial results, in the general case, in Section 4.3

4.1 Cost of parking construction for general unitary cost function

Before turning to general cost, let us discuss the standard parking cost functions. In order to fit with standard representation, we work on $\mathbb{Z}/n\mathbb{Z}$ and will come back to $\mathscr{C}_n$ later on. So, let us write $\mathbf{c}_0, \cdots, \mathbf{c}_{k-1}$ the arrival places of the cars, that are now iid uniform on $\mathbb{Z}/n\mathbb{Z}$ (the letter "c" for, the choice). In the standard parking problem, the car i parks at the first available place among $\mathbf{c}_i, \mathbf{c}_i + 1 \bmod n, \cdots$, say at place $\mathbf{c}_i + d_i \bmod n$ (with $d_i \in \{0, \cdots, n-1\}$). Hence, d_i is the distance car i has to do from its choice c_i to its eventual place.

Define

$$\mathsf{Cost}(k) = \sum_{i=0}^{k-1} d_i. \tag{4.1}$$

Remark 4.1. *In computer science, the hashing with linear probing is an algorithm devoted to store some data $(x_i, i \geq 0)$, taken in a set $\mathscr{D}$ from more or less any type, in an array whose entries are labeled by $\mathbb{Z}/n\mathbb{Z}$. A hash function $h : \mathscr{D} \to \mathbb{Z}/n\mathbb{Z}$ is applied, and data x_i is stored in the first available place among $h(x_i) + k \bmod \mathbb{Z}/n\mathbb{Z}$, for k ranging from 0 to $n-1$. Up to the vocabulary, this is is the parking model under the hypothesis that the hash function produces (almost) uniform random variables. Under this hypothesis,*

$\mathsf{Cost}(k)$ *corresponds to the number of cell availability tests in the array needed for the storage of the k first* x_i*, and it is therefore an important parameter in the analysis of algorithm perspective.*

Let us consider again the phase transition that occurs at time

$$t_n(\lambda) := \lfloor n - \lambda\sqrt{n} \rfloor,$$

Proposition 4.2. *We have the following convergences in distribution, [Flajolet & Viola and Poblete [14], Janson [16], Chassaing & Marckert [10]],*

$$\mathsf{Cost}(n)/n^{3/2} \xrightarrow[n]{(d)} \int_0^1 e(t)\mathrm{d}t$$

and [Chassaing & Louchard[9] *[9]]*

$$\mathsf{Cost}(t_n(\lambda))/n^{3/2} \xrightarrow[n]{(d)} \int_0^1 \mathsf{e}_\lambda(s) - \underline{\mathsf{e}_\lambda}(s)\mathrm{d}s. \tag{4.2}$$

where e *is again, the Brownian excursion, and*

$$\mathsf{e}_\lambda(s) = \mathsf{e}(s) - \lambda s, \textit{ and } \underline{\mathsf{e}_\lambda}(s) = \min_{t\le s} \mathsf{e}_\lambda(t). \tag{4.3}$$

The "cluster size insertion sequence".

Let us call occupied clusters, a successive set of occupied places (as discussed above, an occupied CC $[a,b]$ on $\mathbb{R}/n\mathbb{Z}$, corresponds to the occupied cluster $\{a, a+1 \bmod n, \cdots, b-1 \bmod n\}$ in the standard parking terminology.

When the kth car arrives in the parking there are two possibilities. Either:
– the choice $\mathbf{c}_k$ is an element of an occupied cluster $C = \{a, a+1 \bmod n, \cdots a+s-1 \bmod n\}$ of size $\mathbf{s}_k := s \ge 1$, and this occurs with probability s/n for an occupied cluster of size s; in this case conditional on the fact that $\mathbf{c}_k \in C$, then $\mathbf{c}_k$ is uniform in C,
– or it arrives on a free site f this arrival cluster is then seen to be of size $\mathbf{s}_k = 0$, and park there. Each free place at time k has probability $1/n$ to be the choice of car $\mathbf{c}_k$.

Denote by $\mathsf{UCost}^{(s)}$ the unitary cost corresponding to the cost arrival in a block of size s (knowing that if the arrival takes place in such a block, the position is uniform in it). In general, we have

$$\mathsf{Cost}_n(t) = \sum_{k\le t} \mathsf{UCost}_k^{(\mathbf{s}_k)} \tag{4.4}$$

where again, $\mathbf{s}_k$ is the size of the occupied cluster in which arrives the kth car, and the index k at $\mathsf{UCost}_k^{(\mathbf{s}_k)}$ is here to express that we take different and independent copies of $\mathsf{UCost}^{(s)}$ if several block of the same size s are considered (which is the case in general). There are two main actors that enter into play for the analysis of the global cost Cost_n: the sequence of distributions of the unitary cost random variables $(\mathsf{UCost}^{(s)}, s \ge 0)$, and $(\mathbf{s}_k)$ the cluster size insertion sequence.

[9] In [9], (4.2) is not stated, but it is a consequence of their Lemma 3.7, in which they prove the convergence in distribution of the Lukaciewisz path encoding the parking at time $n - \lambda\sqrt{n}$, and normalized by $n^{1/2}$ and suitably change of origin, to $(e_\lambda(s) - \underline{e_\lambda}(s), s \in [0,1])$. The cost being an integral of the Lukaciewisz path, the convergence stated in our Prop. 4.2 follows (this tool is used in a lot of places in the literature, including [10], [8], [26]).

Asymptotic behavior of the cluster size insertion sequence.

We already have some information on the cluster size sequence (see Proposition 3.1), and additional can be found in Pittel [29], Chassaing & Louchard [9], and in other references already cited. At the time $t_n(\lambda)$, only $\lambda\sqrt{n}$ cars are lacking, and the sequence $(\mathbf{s}_k, k \le t_n(\lambda))$ is composed by $t_n(\lambda)$ elements, most of them being very small (say $\le C\log n$, by Pittel), but a lot of them are large too. We first establish a limit theorem for the empirical measure associated with the "large blocks" in the cluster size insertion sequence.

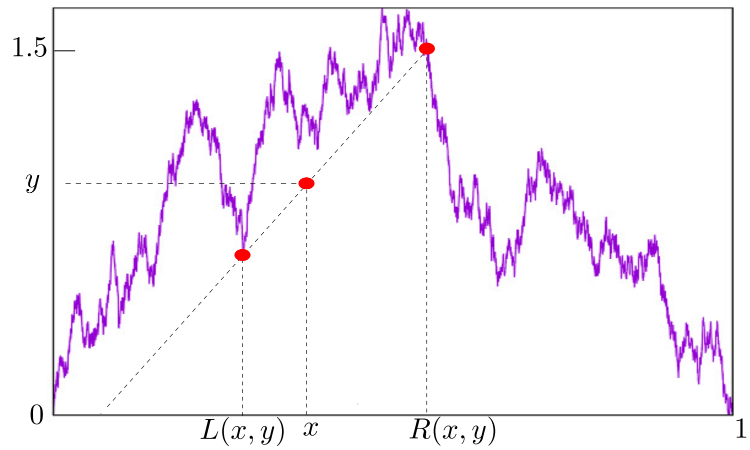

Figure 6: Illustration of a Brownian excursion, the pair $(L(x,y), R(x,y))$ associated with a point $(x,y) \in$ under(e).

Again, $\mathsf{e} = (\mathsf{e}_x, 0 \le x \le 1)$ is the Brownian excursion. Let $\mathscr{L}(\mathsf{e})$ be the set of abscissa of local minima of e and let under(e) $= \{(x,y) : 0 < x < 1, 0 \le y \le \mathsf{e}_x\}$ be the set of points under the Brownian excursion curve. We say that e is above a linear map $\ell : [0,1] \to \mathbb{R}$ on $[a,b]$, if, for any $x \in [a,b]$, $\mathsf{e}_x \ge \ell(x)$ (see Figure 6). For any $(x,y) \in$ under(e) and $u \in \mathscr{L}(\mathsf{e})$ such that $u < x$, define the linear map $\ell_{u,x} : [0,1] \to \mathbb{R}$ of the line passing by (u, e_u) and (x,y) (that is $\ell_{u,x}(u) = \mathsf{e}_u$ and $\ell_{u,x}(x) = y$). Now, let $L(x,y)$ be the largest $u \in \mathscr{L}(\mathsf{e})$ such that $u < x$ and such that e is above $\ell_{u,x}$ on $[0,x]$. Then, set $R(x,y)$ be the abscissa of the first intersection of $\ell_{u,x}$ with the graph of e, on $[x,1]$. Let $S(x,y)$ be the slope of $\ell_x := \ell_{u,x}$.

It is easy to see that the map $(x,y) \to (L(x,y), R(x,y), S(x,y))$ is well measurable

Lemma 4.3. *For any continuous function $f : [0,1] \to \mathbb{R}$ such that $|f(x)| \le Cx$ for some constant C, set*

$$\langle f, M_\lambda(\mathsf{e})\rangle = \int_0^1 \int_0^{\mathsf{e}_x} \frac{2}{R(x,y) - L(x,y)} f\Big(R(x,y) - L(x,y)\Big) \mathbb{1}_{S(x,y)\ge\lambda} \mathrm{d}y\mathrm{d}x. \tag{4.5}$$

▪ *Almost surely (on e), for f continuous $f : [0,1] \to \mathbb{R}$, $\langle f, M_\lambda(\mathsf{e})\rangle$ is well defined, finite, and linear: $M_\lambda(\mathsf{e})$ is a positive (random) measure on $[0,1]$ (equipped with the Borelian σ-field).*
▪ *The measure M_λ has a.s. infinite total mass.*

In the sequel we will often write M_λ instead of $M_\lambda(\mathsf{e})$.

Proof. Notice that for $f := x \mapsto x$ (we will write sometimes $f = \mathsf{Id}$)

$$\langle f, M_\lambda(\mathsf{e})\rangle = 2\int_0^1 \int_0^{\mathsf{e}_x} \mathbb{1}_{S(x,y)\ge\lambda} \mathrm{d}y\mathrm{d}x = 2\int_0^1 \mathsf{e}_\lambda(x) - \underline{\mathsf{e}}_\lambda(x)\,\mathrm{d}x < +\infty \text{ a.s.}. \tag{4.6}$$

This last property can be seen by a simple picture (the first integral is the Lebesgue measure of the points above some chords in e, and, in the graphical representation of $\mathsf{e}_\lambda - \underline{\mathsf{e}}_\lambda$, the corresponding chords, correspond to the intervals that supports the excursion above zeros).
The fact that M_λ has a.s. infinite total mass is a consequence of (4.21) (that will appear later on). $\square$

Theorem 4.4. *Suppose that for each s, $\mathsf{UCost}^{(s)}$ is in L^2, and set*

$$\Psi(s) = \mathbb{E}\big(\mathsf{UCost}^{(s)}\big), \;\; V(k) = Var\big(\mathsf{UCost}^{(s)}\big).$$

Assume that there exists a sequence (α_n) with $\alpha_n \to +\infty$ such that the following conditions hold:
(i) for all $t > 0$, $\Psi(\{nt\})/\alpha_n \to \Psi_\infty(t)$ uniformly on $[a,1]$ for all $a > 0$, where Ψ_∞ is continuous on $[0,1]$ and satisfies $|\Psi_\infty(t)| \le Ct$ for some constant $C \ge 0$, in a neighborhood of zero.
(ii) For $\varepsilon > 0$

$$\limsup_n \frac{1}{\alpha_n n^{1/2}} \left[n\varepsilon^{1/2} \sum_{k=1}^{n\varepsilon} \frac{\Psi(k)}{k^{3/2}} + n \sum_{k=1}^{n\varepsilon} \frac{\Psi(k)}{k^2} \right] = o(\varepsilon). \tag{4.7}$$

(iii)For $\varepsilon > 0$,

$$\limsup_n \frac{1}{\alpha_n^2 n} \left[n\varepsilon^{1/2} \sum_{k=1}^{n\varepsilon} \frac{V(k)}{k^{3/2}} + n \sum_{k=1}^{n\varepsilon} \frac{V(k)}{k^2} \right] = o(\varepsilon). \tag{4.8}$$

(iv) For all $\varepsilon > 0$

$$\limsup_n \sup_{x \ge \varepsilon} \frac{1}{\sqrt{n}\alpha_n^2} V(\lfloor nx \rfloor) = 0. \tag{4.9}$$

Under these four conditions, for all $\lambda \ge 0$,

$$\frac{\mathsf{Cost}_n(\lfloor n - \lambda\sqrt{n} \rfloor)}{\sqrt{n}\alpha_n} \xrightarrow[n]{(d)} \langle \Psi_\infty, M_\lambda \rangle. \tag{4.10}$$

Remark 4.5. *Some conditions, in the theorem, are probably stronger than needed. For example, the condition on Ψ_∞ near zero is sufficient to have $\langle 1_{[0,\varepsilon]} |\Psi_\infty|, M_\lambda \rangle \to 0$ in probability as ε goes not 0, because*

$$\langle \mathbb{1}_{[0,\varepsilon]} |\Psi_\infty|, M_\lambda \rangle \le C \int_0^1 \int_0^{\mathsf{e}_x} \mathbb{1}_{R(x,y) - L(x,y) \le \varepsilon} dy dx \xrightarrow[\varepsilon \to 0]{(a.s.)} 0. \tag{4.11}$$

4.1.1 Applications of Theorem 4.4

In each of the following case, the expectation of the limit can be computed thanks to the formulas of Section 5.3.

1-Standard parking: $\mathsf{UCost}^{(k)}$ is $1 + U_k$ where U_k is uniform on $\{0, \cdots, k-1\}$ (in general, the cost is defined as the set of "tries" a car needs to park, but this +1 can be dropped, if one prefers, it does not change the asymptotic results). We then have

$$\Psi(k) = (k+1)/2, \;\; V_k = (k^2 - 1)/12.$$

It follows the hypothesis of Theorem 4.4 holds, since :
– (i) with $\alpha_n = n$, $\Psi_\infty(t) = t/2$,

– (ii) since $\sum_{k=1}^{n\varepsilon}(k/2)/k^{3/2} = O((n\varepsilon)^{1/2})$,$\sum_{k=1}^{n\varepsilon}(k/2)/k^2 = O(\log(n\varepsilon))$,
– (iii) since $\sum_{k=1}^{n\varepsilon}((k^2-1)/12)/k^{3/2} = O((n\varepsilon)^{3/2})$, $\sum_{k=1}^{n\varepsilon}((k^2-1)/12)/k^2 = O(n\varepsilon)$,
– (iv), immediate.
Notice that (ii) is tight in this case, "there is no margin". We then have

$$\frac{\mathsf{Cost}_n(\lfloor n-\lambda\sqrt{n}\rfloor)}{n^{3/2}} \xrightarrow[n]{(d)} \langle \Psi_\infty, M_\lambda\rangle. \tag{4.12}$$

and we recover by (4.6), the Chassaing & Louchard results recalled in Proposition 4.2.

2-Parking with random direction: If upon arrival the drivers choose to go to the right with probability p, and to the left with probability $1-p$, the unitary cost has the same distribution as for the previous point: the same asymptotic result holds.

3-Choosing the closest place strategy: This strategy provides a cost $1+\min\{U_k, k-U_k\}$ for U_k a uniform random variable on $\{0,\cdots,k\}$. It is a simple exercise to see that $\Psi(k)/k \to 1/3$ and $V(k)/k^2 \to 3/80$ (since $(1+\min\{U_k, k-U_k\})/k$ is a sequence of uniformly bounded random variable converging in distribution to a $\beta(1,3)$ random variable, its mean and variance converges to those of the $\beta(1,3)$. Besides $V(k) = O(k^2)$ (the maximum variance for a random variable with support $[\![1, k+1]\!]$. Hence, the standard parking (1) analysis above applies, with $\Psi_\infty(t) = t/3$ instead (so, asymptotically, a factor $(1/3)/(1/2) = 2/3$ of the usual parking cost).

4-Making a p-random walk to find a place, with $p > 1/2$: $\mathsf{UCost}^{(k)}$ is the exit time of an interval of size k by a p-random walk, starting from a uniform point. We skip the details since the asymptotic are the same for two costs that differs from a constant. In this case $\mathsf{UCost}^{(k)}/k$ converges in distribution to $U/(2p-1)$ (the asymptotic speed of a p random walk is $1/(2p-1)$), and U is uniform on $[0,1]$. The convergence of $\Psi(k)/k$ and $V(k)/k^2$ to $\mathbb{E}(U)/(2p-1) = 1/(2(2p-1))$ and to $\mathsf{V}(U)/((2p-1)^2) = 1/(12(2p-1)^2)$ can be proved using the Hoeffding inequality which ensures exponentially rare deviation for $\mathsf{UCost}^{(k)}/k$ from its expected values (at a smaller scale). Hence,

$$\mathsf{Cost}(n-\lambda\sqrt{n})/n^{3/2} \xrightarrow[n]{(d)} \int_0^1 \frac{e_\lambda(s) - \underline{e_\lambda}(s)}{2p-1}\,ds \tag{4.13}$$

5-Making a p-random walk to find a place, with $p = 1/2$: Here $\mathsf{UCost}^{(k)}$ is the time $T^{(k)}$ needed for a centered random walk to exit an interval of time k, starting from a uniform position. It can be proved that

$$\Psi(k)/k^2 \to 1/6, V(k)/k^4 \to 19/180.$$

There are two main methods to prove this result. The first one consists in noticing that $\mathsf{UCost}^{(k)}/k^2 \xrightarrow[n]{(d)} T_U$ where T_U is the needed time needed for a Brownian motion $B^{(U)}$ starting from U uniform on $[0,1]$ to exit the interval $[0,1]$, which is a consequence of Donsker. Since by Lépingle [23], for a Brownian motion $B^{(u)}$ starting at a fixed point $u \in [0,1]$, the exit time T_u from $[0,1]$, satisfies

$$\mathbb{E}\left[\exp\left(\alpha^2 T_u/2\right)\right] = \frac{\cos(\alpha(1-2u)/2)}{\cos(\alpha/2)}$$

for $0 \le \alpha \le \pi$. Follows that,

$$\mathbb{E}\left[\exp\left(\alpha^2 T_U/2\right)\right] = \frac{2\sin(\alpha/2)}{\alpha\cos(\alpha/2)}, \tag{4.14}$$

and then T_U admits all its moments, and $\mathbb{E}(T_U) = 1/6$ as well as $\mathbb{E}(T_U^2) = 2/15$ can be obtained by expansion of $\mathbb{E}(\exp(xT_U))$ (set $\alpha^2/2 = x$ in (4.14)). Now, to prove that the moments of $T^{(k)}/k^2$ converges to those of T_U some uniform integrability argument can be used, for example, Komlos, Major and Tusnády [19] strong convergence theorem (with speed), or again, the second part of Lépingle [23] paper, who computed the Laplace transform of the exit time of a random walk from an interval, which ensures the convergence of the Laplace transform of $T^{(k)}/k^2$ to that of T_U (for positive arguments, in a neighborhood of zero), from which convergence of moments are immediate. A second proof, totally different, would rely on martingale argument: it is well known that the hitting time $T_{a,b}$ of $\{a,b\}$ for a random walk starting at 0, were $a < 0 < b$, satisfies $\mathbb{E}(T_{a,b}) = -ab$, and this can be proved by the martingale stopping time theorem, applied to the martingale $(W_k^2 - k, k \ge 0)$ where W_k is the simple symmetric random walk. The other moments of $T_{a,b}$ can be computed using the family of martingales,indexed by c, $Z_n^{(c)} = e^{cW_n}(\cosh c)^{-n}$ (see also Lépingle [23], and Chung [11, Sec. II.3] for Brownian motion techniques that can be adapted to random walks). Now take $\alpha_n = n^2$. Since $\Psi(k)/k^2 \to a := 1/6$ and $V(k)/k^4 \to b := 19/180$

$$\Psi(nt)/n^2 \to at^2, \limsup_n \sup_{x\ge\varepsilon} \frac{V(\lfloor nx\rfloor)}{\sqrt{n}\alpha_n^2} \to 0.$$

Since $\Psi(k)/k^2 \le C$ and $V(k)/k^4 \le C$ by convergence, for a well chosen constant C, one can check that (ii), (iii) and (iv) hold. (The convergence in (i) follows the fact that $\mathbb{E}(T_{a,b}) = -ab$ from which the results follows. We then have

$$\mathsf{Cost}(n-\lambda\sqrt{n}, p)/n^{5/2} \xrightarrow[n]{(d)} \langle g, M_\lambda\rangle = \frac{1}{3}\int_0^1\int_0^{\mathsf{e}_x}(R(x,y)-L(x,y))\mathbb{1}_{S(x,y)\ge\lambda}\,dy\,dx \tag{4.15}$$

where $g(x) = x^2/6$.

4.2 Proof of Theorem 4.4

This section is constituted by three subsections. First, we talk about some properties of M_λ, and provide another representation of this measure. Then, we introduce a measure $\Theta_n^{(\lambda)}$ that encodes the insertion length sequence $(\mathbf{s}_k, 0 \le k \le t_n(\lambda))$, then establish its convergence to M_λ after normalisation, and then conclude by the proof of Theorem 4.4.

4.2.1 Some distribution identities concerning the measures M_λ

Take a continuous function $f : [0,1] \to \mathbb{R}$ such that $|f(x)| \le \mathsf{Cste}.x$, and $\lambda > 0$.
Consider the null set $Z_\lambda = \{x : \mathsf{e}_\lambda(x) - \underline{\mathsf{e}}_\lambda(x) = 0\}$. The complement of Z_λ in $[0,1]$ is an open set L_λ, whose CC are countably many open intervals $L_\lambda = \cup_{i\ge1} L_\lambda(i)$. The null set Z_λ has Lebesgue measure zero, a.s., so that a.s., for $\ell_k(\lambda) = \mathrm{Leb}(L_\lambda(i))$, we have

$$\sum_{k\ge1} \ell_k(\lambda) = 1,$$

and then, for any, for

$$\rho_\lambda = \sum_{k\ge1} \delta_{\ell_k(\lambda)}, \tag{4.16}$$

we have

$$\langle \rho_\lambda, f\rangle = \sum_{k\geq 1} f(\ell_k(\lambda))$$

is well-defined, by Fubini, and finite. In particular, this is the case for $f(x) = x^\beta$ for $\beta \geq 1$.

A.s. the map $\lambda \mapsto \langle \rho_\lambda, f\rangle$ is Lebesgue measurable, and Lebesgue integrable on any compact of $\mathbb{R}^+$. To prove this, we take f non-negative (which is sufficient),and take a time interval $[t,t'] \subset [0,+\infty)$. Hence, for $\varepsilon > 0$ there exists N such that the N largest intervals $\left(\widehat{\ell}_{t'}(1),\cdots,\widehat{\ell}_{t'}(N)\right)$, sorted in decreasing order, satisfy $\sum_{k=1}^N \widehat{\ell}_{t'}(k) \geq 1-\varepsilon$. Since the intervals in $L_{t'}$ are obtained from those of L_t by fragmentation, for any $t \leq \lambda \leq t'$, $\sum_{k=1}^N \widehat{\ell}_\lambda(k) \geq 1-\varepsilon$. Since $|f(x)| \leq \mathsf{Cste}.x$, up to $\varepsilon > 0$, $\langle \rho_\lambda, f\rangle$ is determined by the N largest intervals of L_λ (because $\sum_{k>N} f(\widehat{\ell}_\lambda(k)) \leq C\sum_{k>N}\widehat{\ell}_\lambda(k) \leq C\varepsilon$).

Let $\overline{L}_\lambda(k)$ be the interval, at time λ, that contains the kth largest interval $L_{t'}(n_k)$ at time t' (whose length is $\widehat{\ell}_{t'}(k)$) (it may happen that $\overline{L}_\lambda(k) = \overline{L}_\lambda(j)$ for some $k \neq j$). If it is the case, then keep only the ones with smaller index, and keep $(\overline{L}_\lambda(k), k \in I(\lambda))$ where $I(\lambda)$ is the corresponding set of indices.

For the corresponding interval lengths $\left(\overline{\ell}_\lambda(k), k \in I(\lambda)\right)$, we have also $\left|\sum_{k\in I(\lambda)} f(\overline{\ell}_\lambda(k)) - \langle \rho_\lambda, f\rangle\right| < \varepsilon$. Since the maps $\lambda \mapsto \overline{\ell}_\lambda(k)$ as λ goes from t' to t, are non decreasing (and may possibly vanish in case of coalescence with another interval of this same family with lower index). It follows that, for fixed k, $\lambda \mapsto f(\overline{L}_\lambda(k))$ is measurable (by taking the value zero in case of disappearance). The integrability is clear since $\langle \rho_\lambda, f\rangle \leq C$. Thus

$$\int_t^{t'} \langle \rho_\lambda, f\rangle \mathrm{d}\lambda = \int_t^{t'} \sum_k f(\ell_\lambda(k))\mathrm{d}\lambda, \tag{4.17}$$

is well defined and finite. This well-definiteness can be extended to Lebesgue measurable functions f satisfying $\sup\{|f(x)/x|, x \in [0,1]\} < +\infty$, by density.

Theorem 4.6. *For all bounded and Lebesgue measurable function $g : [0,1] \to \mathbb{R}^+$,*

$$\langle g, M_\lambda(\mathsf{e})\rangle \quad = \quad \int_\lambda^{+\infty} \sum_{k\geq 0} g(\ell_k(t))\,\ell_k(t)\,\mathrm{d}t < +\infty, \quad a.s. \tag{4.18}$$

In other words, for $G := x \mapsto xg(x)$, $\langle g, M_t(\mathsf{e})\rangle = \int_t^{+\infty}\langle \rho_\lambda, G\rangle d\lambda$. In particular, for $f = \mathsf{Id}/2$, we have:

$$\langle f, M_\lambda(\mathsf{e})\rangle \quad = \quad \int_0^1 \mathsf{e}_\lambda(s) - \underline{\mathsf{e}}_\lambda(s)\,\mathrm{d}x = \frac{1}{2}\int_\lambda^{+\infty}\sum_{k=1}^{+\infty} \ell_k(t)^2\mathrm{d}t. \tag{4.19}$$

The first equality is discussed above, see (4.6). See Section 5.3 for close formulas for $\mathbb{E}\left(\langle x \mapsto x^k, M_\lambda\rangle\right)$.

Proof. **Decomposition of $\langle f, M_\lambda(\mathsf{e})\rangle$ as a sum over the abscissa of local minima of e for general f.**
We may partition $\mathsf{under}(\mathsf{e}) = \cup_{m\in\mathscr{L}(\mathsf{e})}\mathscr{P}(m)$ where

$$\mathscr{P}(m) = \{(x,y) \in \mathsf{under}(\mathsf{e}) \ : \ L(x,y) = m\}.$$

The sets $(\mathscr{P}(m), m \in \mathscr{L}(\mathsf{e}))$ form a partition of $\mathsf{under}(\mathsf{e})$, and then any integral on the set $\mathsf{under}(\mathsf{e})$ can be written as a sum over $\mathscr{L}(\mathsf{e})$ (under usual conditions, as Fubini's theorem). We will show that :

$$\langle g, M_\lambda(\mathsf{e})\rangle \quad = \quad \int_\lambda^{+\infty} \sum_{m\in\mathscr{L}(\mathsf{e})} \ell(m,t) g(\ell(m,t))\mathrm{d}t, \tag{4.20}$$

where $\ell(m,t)$ is the size of the segment with slope t, starting at $(m,\mathsf{e}(m))$ and stopped when it passes above the graph of e, or, in other words, the length of the excursion of $x \to \mathsf{e}_t(x) - \underline{\mathsf{e}}_t(x)$ starting at m.

For any Borelian function $f : [0,1] \to \mathbb{R}^+$,

$$\begin{aligned}\langle f, M_\lambda(\mathsf{e})\rangle &= \int_0^1 \int_0^{\mathsf{e}_x} \frac{2f\Big(R(x,y)-L(x,y)\Big)}{R(x,y)-L(x,y)} \mathbb{1}_{S(x,y)\geq\lambda} \mathrm{d}y\mathrm{d}x \\ &= \sum_{m\in\mathscr{L}(e)} \int_0^1 \int_0^{\mathsf{e}_x} \frac{2f\Big(R(x,y)-m\Big)}{R(x,y)-m} \mathbb{1}_{S(x,y)\geq\lambda} \mathbb{1}_{(x,y)\in\mathscr{P}(m)} \mathrm{d}y\mathrm{d}x.\end{aligned}$$

For any $(x,y)\in\mathscr{P}(m)$, by letting t be the slope $t = S(x,y)$, one has $y = \mathsf{e}(m) + t(x-m)$. We may then proceed to a change of variable $(x,y) \to (x,t)$. For this choice, $\mathbb{1}_{S(x,y)\geq\lambda}$ will be transferred to the integration domain, $\{t : t \geq \lambda\}$. We obtain

$$\langle f, M_\lambda(\mathsf{e})\rangle = \sum_{m\in\mathscr{L}(e)} \int_0^1 \int_\lambda^{+\infty} \frac{2f\Big(R(x,\mathsf{e}(m)+t(x-m))-m\Big)}{R(x,\mathsf{e}(m)+t(x-m))-m} \mathbb{1}_{(x,y)\in\mathscr{P}(m)}(x-m)\mathrm{d}t\mathrm{d}x \tag{4.21}$$

$$= \sum_{m\in\mathscr{L}(e)} \int_\lambda^{+\infty} \frac{2f\Big(\ell(m,t)\Big)}{\ell(m,t)} \int_m^{m+\ell(m,t)} (x-m)\,dx\mathrm{d}t \tag{4.22}$$

$$= \int_\lambda^{+\infty} \sum_{m\in\mathscr{L}(\mathsf{e})} \ell(m,t) f(\ell(m,t))\mathrm{d}t. \tag{4.23}$$

□

4.2.2 A measure encoding the cluster size insertion sequence

We store in a measure $\mathsf{Mes}_n^{(t)}$ the elements of the cluster size insertion sequence $(\mathbf{s}_k, k \geq 0)$ viewed before time t (with multiplicity) :

$$\mathsf{Mes}_n^{(t)} = \sum_{j\leq t} \delta_{\mathbf{s}_j}. \tag{4.24}$$

A consequence of (4.4), is that when some unitary cost laws (the laws of UCost^s) are fixed, then the total cost can be expressed as

$$\mathsf{Cost}_n(t) = \sum_k \sum_{j=1}^{\mathsf{Mes}_n^{(t)}(\{k\})} \mathsf{UCost}_j^{(k)}. \tag{4.25}$$

Foreword. In view of (4.24) and (4.25), we may expect that under some regularity assumption on the unitary cost laws, a limiting behavior for $\mathsf{Mes}_n^{(t)}$,after rescaling, when $n \to +\infty$, and for $\mathsf{UCost}^{(k)}$ should lead to the convergence result for the global cost. This reasoning is somehow valid, but in order to capture the asymptotic behavior of $\mathsf{Mes}_n^{(t)}$, we are forced to apply a normalization under which clusters with sublinear sizes vanished at the limit, and those with linear sizes provide a limit random measure with infinite total mass. The difficulty that follows is that, for some costs functions, small clusters can play a role, while for others, only large one will matter: this is apparent in Theorem 4.4, where the hypothesis we made imply that only large clusters will characterize the limiting behavior of the cost.

There are **several normalizations possible** for $\mathsf{Mes}_n^{(t)}$, depending on the values of t we want to study, and the unitary cost model. The normalization is mainly three folds, and it is natural to consider

$$\frac{1}{\alpha_n}\mathsf{Mes}_n^{(t_n)}(\,.\,/\beta_n)$$

and α_n addresses the total mass normalization, while β_n is used to target a given block size, and t_n is the time.

<u>For example,</u> if $\mathsf{UCost}^{(k)}=1$ for all $k\geq 0$, the total cost is $\mathsf{Cost}_n(t)=t$, it is not needed to search for a limit of $\mathsf{Mes}_n^{(t)}$ to see that! Nevertheless when one searches for a limit of $\mathsf{Mes}_n^{(t_n)}$, for the time $t=t_n:=an$ (and $a\in(0,1)$, the maximum $\mathbf{s}_j$ for $j\leq an$ has order $O(\log n)$ in probability by [29]). It means that, taking a limit over n,

$$\frac{1}{n}\mathsf{Mes}_n^{(an)}(\,.\,/n)\xrightarrow[n]{(d)} a\delta_0,$$

so that the block size information essentially vanishes. This exemplifies a more general phenomenon: it could seem natural to take $\alpha_n=t_n$ to keep a measure with total mass one, but since most of the blocks are very small (at least for t_n close to n) this leads to the (degenerated) Dirac mass at the limit. We must then change the normalization in order to analyze much of the model of costs.

In order to describe the large blocks, α_n must be different from the total mass: the number of clusters that have a linear size has order $\sqrt{n}$ (for the critical time $t_n(\lambda)=\lfloor n-\lambda\sqrt{n}\rfloor$). We then consider the random Borelian measure $\Theta_n^{(\lambda)}$ defined, for all Borelian $A\subset[0,1]$, by

$$\begin{aligned}
\Theta_n^{(\lambda)}(A) &:= \frac{1}{\sqrt{n}}\mathsf{Mes}_n^{(t_n(\lambda))}(nA) && (4.26)\\
&= \sum_{k:k\in nA}\frac{1}{\sqrt{n}}\mathsf{Mes}_n^{(t_n(\lambda))}(\{k\}) && (4.27)\\
&= \int_0^n \frac{1}{\sqrt{n}}\mathsf{Mes}_n^{(t_n(\lambda))}(\lfloor x\rfloor)\mathbb{1}_{x\in nA}dx && (4.28)\\
&= n\int_0^1 \frac{1}{\sqrt{n}}\mathsf{Mes}_n^{(t_n(\lambda))}(\lfloor ny\rfloor)\mathbb{1}_{x\in A}dy && (4.29)
\end{aligned}$$

these 4 representations help to understand the normalisation into play: for example

$$\langle f,\Theta_n^{(\lambda)}\rangle = n\int_0^1\frac{1}{\sqrt{n}}\mathsf{Mes}_n^{(t_n(\lambda))}(\lfloor nx\rfloor)f(x)dx. \tag{4.30}$$

The measure $\Theta_n^{(\lambda)}$ has total mass $\sqrt{n}-\lambda\to+\infty$ with, much of this mass close to the point zero. Let us consider the vague topology on the set of Borel measures on $\mathscr{M}(0,1)$ (which, prosaically, allows to ignore this degeneracy at zero, since the test functions for these topologies are those with compact support strictly included in $(0,1)$).

Theorem 4.7. *For any $\lambda>0$, $\Theta_n^{(\lambda)}\xrightarrow[n]{(d)} M_\lambda$ for the vague convergence topology on $\mathscr{M}(0,1)$ (and the joint convergence $(\Theta_n^{(\lambda_i)},1\leq i\leq\kappa)\xrightarrow[n]{(d)}(M_{\lambda_i},1\leq i\leq\kappa)$ holds too, for any times $0<\lambda_1<\cdots<\lambda_\kappa$ and any κ).*

Proof. We borrow to Chassaing–Louchard [9] some elements of their analysis.

Consider a random walk $S[n]:=(S_0=0,S_1,S_2,\cdots)$ with iid increments $(X_i-1,i\geq 1)$, where the X_i are Poisson(1) random variable. Denote by $\tau_{-1}(S)=\inf\{j:S_j=-1\}$.

The collecting path (recall Section 1.8) of the classical parking problem on a size n parking, stopped when $n-1$ cars are parked, and conditional on the fact that the last place is empty, is distributed as $S[n]$ conditioned by $\tau_{-1}(S)=\inf\{j: S_j=-1\}=n$. Moreover, the vector $(Y_1,\cdots,Y_n)$ of cars that have chosen place 1 to n, satisfies $\mathscr{L}(Y_1,\cdots,Y_n)=\mathscr{L}\left((X_1,\cdots,X_n)\mid \tau_{-1}(S)=n\right)$. The corresponding state of this random walk at time $t_n(\lambda)$, that is when $\lambda\sqrt{n}$ cars are lacking, is denoted $S^{(t_n(\lambda)}$. Let

$$\left(s_t^{n,(\lambda)}, t\in[0,1]\right):=\left(n^{-1/2}\,S_{nt}^{(t_n(\lambda))}, t\in[0,1]\right) \tag{4.31}$$

where $S^{(t_n(\lambda))}$ is seen as a continuous process, interpolated linearly between integer points. In Broutin & Marckert [8, Theo.8] is has been shown that

$$\left(s_t^{n,(\lambda)}, t\in[0,1],\lambda\geq 0\right)\xrightarrow[n]{(d)}\left((\mathsf{e}_t-\lambda t)-\min_{s\leq t}(\mathsf{e}_s-\lambda s), t\in[0,1],\lambda\geq 0\right) \tag{4.32}$$

in $D(\mathbb{R}^+,(C[0,1],\mathbb{R}))$ when the convergence of FDD were already present in [9] (we see $s^{n,(\lambda)}$ as a process indexed by λ, taking, for each λ its values in $C([0,1],\mathbb{R})$).

By the Skhorokhod representation theorem, there exists a probability space on which exists some copies $\left(\widetilde{s}_t^{n,(\lambda)},\lambda\geq 0, t\in[0,1]\right)$ of $\left(s_t^{n,(\lambda)},\lambda\geq 0,t\in[0,1]\right)$ (for each n), and a copie $\widetilde{\mathsf{e}}$ of e, such that

$$\left(\widetilde{s}_t^{n,(\lambda)}-\left[(\widetilde{\mathsf{e}}_t-\lambda t)-\min_{s\leq t}(\widetilde{\mathsf{e}}_s-\lambda s)\right], t\in[0,1],\lambda\geq 0\right)\xrightarrow[n]{(as.)}0$$

for the same topology.

Consider $\mathscr{E}$ the two following conjunction of events:
(i) $\mathscr{L}(\mathsf{e})$ is countable and each $x\in(0,1)\cap\mathscr{L}(\mathsf{e})$ is a global minimum on a non-empty open interval,
(ii) if for j from 1 to 3, the $p_j:=(x_j,\mathsf{e}_{x_j})$ are points on the curve of e, such that the x_i are distinct elements of $\mathscr{L}(\mathsf{e})$, then the p_j are not aligned.
The measure $\Theta_n^{(\lambda)}$ records only the insertion block sizes, but can be enriched to record also the position of these cluster (on $\widetilde{s}_t^{n,(\lambda)}$): for B a Borelian of $\mathbb{R}^2$, set

$$\Xi_n^{(\lambda)}(B)=\frac{1}{\sqrt{n}}\sum_{k\leq t_n(\lambda)}\delta_{[a_k/n,b_k/n]\in B} \tag{4.33}$$

where $[a_k,b_k]$ is the cluster, which size $\mathbf{s}_k$, which contains the arrival position u_k (and which was then an excursion position in the kth collecting path $S^{(k)}$). Now, recall the computation made in the proof of Theorem 4.6, and in particular the decomposition of the measure M_λ as a sum over the abscissa of local minima of e, that is $M_\lambda=\sum_{m\in\mathscr{L}(e)}M_\lambda^{(m)}$ and

$$\langle f,M_\lambda^{(m)}\rangle=\int_0^1\int_0^{\mathsf{e}_x}\frac{2f\Big(R(x,y)-m\Big)}{R(x,y)-m}\mathbb{1}_{S(x,y)\geq\lambda}\mathbb{1}_{(x,y)\in\mathscr{P}(m)}\mathrm{d}y\mathrm{d}x, \tag{4.34}$$

then, with this decomposition M_λ can also be seen as a measure on $\mathbb{R}^2$ of the limiting block positions, and in $M_\lambda^{(m)}(A)$, the number m is the left hand side and A measure the position of the right and side. Further, for $\lambda'>\lambda$,

$$M_\lambda^{(m)}(A)=\int_0^1\int_0^{\mathsf{e}_x}\frac{2}{R(x,y)-m}\mathbb{1}_{S(x,y)\geq\lambda}\mathbb{1}_{(x,y)\in\mathscr{P}(m)}\mathbb{1}_{R_{x,y}\in m+A}\mathrm{d}y\mathrm{d}x, \tag{4.35}$$

By the same computation as in (4.21), for $\lambda'\geq\lambda$,

$$\langle f,M_\lambda^m(\mathsf{e})-M_{\lambda'}^m(\mathsf{e})\rangle=\int_\lambda^{\lambda'}\ell(m,t)f(\ell(m,t))\mathrm{d}t \tag{4.36}$$

so that for $f = \mathbb{1}_{[a+m,b+m]}$,

$$\langle f, M^m_\lambda(\mathrm{e}) - M^m_{\lambda'}(\mathrm{e})\rangle = \int_\lambda^{\lambda'} \ell(m,t)\mathbb{1}_{\ell(m,t)\in[a,b]}\mathrm{d}t. \tag{4.37}$$

Let us see that this gives the limit number of insertions in clusters located "approximately" in $[a+m, b+m]$, normalized by $\sqrt{n}$ between the times $t_n(\lambda')$ and $t_n(\lambda)$. For this, we already now that $\widetilde{s}^{n,(\lambda)}_t \to e_t - \lambda t$ uniformly for $(\lambda, t) \in [\Lambda_1, \Lambda_2]\times[0,1]$ (because the limit is a.s. continuous), and that there exists an interval $[m_a, m_b]$ containing m in its interior on which m is a global minima (a.s.). Let m_n be the value of t on which $\widetilde{s}^{n,(\lambda)}_t$ takes its minimum on $[m_a, m_b]$ for the last time. By defining the analogous $R_n(x,y)$, $S_n(x,y)$, $\ell_n(m_n,t)$ of $R(x,y)$ and $S(x,y)$, $\ell(m,t)$ one sees that

$$\langle f, M^m_\lambda(\mathrm{e}) - M^m_{\lambda'}(\mathrm{e})\rangle = \lim_n \int_\lambda^{\lambda'} \ell_n(m_n,t)\mathbb{1}_{\ell_n(m_n,t)\in[a,b]}\mathrm{d}t, \tag{4.38}$$

for almost all a, b, λ, λ', a.s.. It remains to see that this quantity gives also the limit number of insertions during the time interval $[t_n(\lambda'), t_n(\lambda)]$, in an occupied interval starting in nm_n and ending in $n[a,b]$, divided by $\sqrt{n}$ (call this value $\xi_n(\lambda,\lambda',m,a,b)$, and this is a sort of equivalent of $\Theta^{(\lambda)}_n - \Theta^{(\lambda)}_n$, enriched, to take into account the minimum m, and the start and end of insertion intervals). For some time $k \in [t_n(\lambda'), t_n(\lambda)]$, the probability for u_k to arrive in such an interval, conditional on $S^{(k)}$ is either 0 if there is no excursion starting in m_n and ending in $n[a,b]$, or $\ell_n(m_n,t)/n$ if there is an excursion with length $\ell_n(m_n,t)$ at time k, with $t \in [a,b]$. Now, by a concentration argument for martingales with bounded increments (as Azuma's inequality), we have

$$\xi_n(\lambda,\lambda',m,a,b) - \int_\lambda^{\lambda'} \ell_n(m_n,t)\mathbb{1}_{\ell_n(m_n,t)\in[a,b]}\mathrm{d}t,$$

goes to zero in probability. This result implies the announced result. □

Lemma 4.8. *(i) For all $\lambda \geq 0$,*

$$X^{(\lambda)}_n := \max_{j\leq t_n(\lambda)} \mathbf{s}_j/n \xrightarrow[n]{(d)} X^{(\lambda)}$$

where $X^{(\lambda)}$ is a.s. in $(0,1)$.

(ii) We have $X^{(\lambda)} \xrightarrow[n]{(proba.)} 0$ when $\lambda \to +\infty$.

(iii) For all $\varepsilon > 0$, there exists Λ such that for all $\lambda \geq \Lambda$, for all n large enough

$$\mathbb{P}\left(X^{(\lambda)}_n \geq \varepsilon\right) \leq \varepsilon. \tag{4.39}$$

Proof. (i) This is a consequence of [9, Section 3.3] (which itself relies on Aldous [2, Lemma], which allows to deduce the convergence of the excursion size above the current minima of a process, which converges in distribution, to those of the limit under some natural assumptions.

(ii) We borrow the argument to [31] (many other proofs are possible but this one does not depend on the statistical properties of the process $e_t - \lambda_t$): consider any continuous function $f : [0,1] \to \mathbb{R}$ such that $f(0) = f(1) = 0$, and consider the maximal length L_λ of an excursion of $\Psi_\lambda(f) : x \mapsto f(x) - \lambda x$ above its current minimum. As a matter of fact the map $\Psi_\lambda(f)$ lies between the two lines with equations $y = -\lambda_x + \max f$ (above) and $y = -\lambda_x + \min f$ (below). A simple picture shows that the excursions above the current minima of a function in this band is $O(1/\lambda)$

(iii) Observe that $\lambda \mapsto X^{(\lambda)}_n$ is non-increasing. Take Λ large enough such that $\mathbb{P}(X^{(\lambda)} \geq \varepsilon/2) \leq \varepsilon$; by (ii), for n large enough, $|\mathbb{P}(X^{(\lambda)} \geq \varepsilon/2) - \mathbb{P}(X^{(\lambda)}_n \geq \varepsilon/2)| \leq \varepsilon/2$ which allows to conclude (we used also the fact that $X^{(\lambda)}$ has an atomless distribution. □

4.2.3 Proof of Theorem 4.4

Theorem 4.4 is a consequence of the three following Lemmas, and of Remark 4.5 which explains that for all $\lambda > 0$ fixed, all $\delta, \delta' > 0$, for all small enough $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\langle 1_{[0,\varepsilon]}\Psi_\infty, M_\lambda\rangle\right| \geq \delta\right) \leq \delta'.$$

Recall that $t_n(\lambda) := \lfloor n - \lambda\sqrt{n}\rfloor$, and from (4.25) (or(4.4)),

$$C_n(\lambda) := \mathsf{Cost}_n(t_n(\lambda)) = \sum_k \sum_{j=1}^{\mathsf{Mes}_n^{(t_n(\lambda))}(\{k\})} \mathsf{UCost}_j^{(k)}.$$

A strategy consist in this sum, to use the convergence of $\Theta_n^{(\lambda)}$, and then, to work conditionally on $\mathsf{Mes}_n^{(t_n(\lambda))}(\{k\})$, somehow. The convergence of $\Theta_n^{(\lambda)}$ captures only the linear size blocks. By decomposing

$$\mathsf{UCost}_j^{(k)} = \Psi(k) + \left(\mathsf{UCost}_j^{(k)} - \Psi(k)\right)$$

along its mean and its fluctuation around its mean, we naturally, write $C_n(\lambda)$ has a sum of 4 terms.

For $\varepsilon \in (0,1)$, $\lambda \geq 0$ fixed, set

$$C_n(\lambda) := L_n(\lambda,\varepsilon) + S_n(\lambda,\varepsilon) + LV_n(\lambda,\varepsilon) + SV_n(\lambda,\varepsilon) \tag{4.40}$$

where

$$\begin{cases} L_n(\lambda,\varepsilon) &= \displaystyle\sum_{k\geq n\varepsilon} \Psi(k)\mathsf{Mes}_n^{(t_n(\lambda))}(\{k\}), \\ S_n(\lambda,\varepsilon) &= \displaystyle\sum_{k< n\varepsilon} \Psi(k)\mathsf{Mes}_n^{(t_n(\lambda))}(\{k\}), \\ SF_n(\lambda,\varepsilon) &= \displaystyle\sum_{k< n\varepsilon} \sum_{j=1}^{\mathsf{Mes}_n^{(t_n(\lambda))}} \left(\mathsf{UCost}_j^{(k)} - \Psi(k)\right), \\ LF_n(\lambda,\varepsilon) &= \displaystyle\sum_{k\geq n\varepsilon} \sum_{j=1}^{\mathsf{Mes}_n^{(t_n(\lambda))}} \left(\mathsf{UCost}_j^{(k)} - \Psi(k)\right) \end{cases} \tag{4.41}$$

(symbols L and S are chosen to represent "large" and "small" blocks contributions, and the F for the "fluctuation terms", conditional on $\mathsf{Mes}_n^{(t_n(\lambda))}$).

Using Theorem 4.7, by the Skhorokhod representation theorem, there exists a probability space $(\overline{\Omega}, \mathscr{A}, \mathbb{P})$, on which are defined some copies $\left(\overline{\Theta}_n^{(\lambda)}, n \geq 0\right)$ of $\left(\Theta_n^{(\lambda)}, n \geq 0\right)$ that converges a.s. to a copy $\overline{M}_\lambda$ of M_λ (for the vague topology).

Lemma 4.9 (Analysis of $L_n(\lambda,\varepsilon)$)**.** *Under the hypothesis of Theorem 4.4, for $\varepsilon > 0$, $\lambda \geq 0$ fixed, on the probability space $(\overline{\Omega}, \mathscr{A}, \mathbb{P})$,*

$$\frac{L_n(\lambda,\varepsilon)}{\alpha_n n^{1/2}} \xrightarrow[n]{(proba.)} \langle \mathbb{1}_{[\varepsilon,1]}\Psi_\infty, \overline{M}_\lambda\rangle.$$

Proof. Observe that since Ψ_∞ is continuous on $(0,1]$,

$$\|\Psi_\infty\|_{\infty,\varepsilon} = \max_{x\in[\varepsilon,1]} |\Psi_\infty(x)| < +\infty.$$

Therefore,

$$\langle 1_{[\varepsilon,1]}|\Psi_\infty|, M_\lambda\rangle = \int_0^1\int_0^{e_x} \frac{1_{R(x,y)-L(x,y)\geq\varepsilon}}{R(x,y)-L(x,y)}|\Psi_\infty(R(x,y)-L(x,y))|dydx \tag{4.42}$$

$$\leq \|\Psi_\infty\|_{\infty,\varepsilon}\int_0^1\int_0^{e_x} 1/\varepsilon dydx \tag{4.43}$$

$$\leq \|\Psi_\infty\|_{\infty,\varepsilon}\int_0^1 \mathsf{e}(s)/\varepsilon ds < +\infty \text{ a.s..} \tag{4.44}$$

With the same argument (taking somehow $\Psi_\infty = 1$), one sees that we have also

$$\langle 1_{[\varepsilon,1]}, M_\lambda\rangle < +\infty \text{ a.s..} \tag{4.45}$$

Now, using that $\mathsf{Mes}_n^{(t_n(\lambda))}(nx)$ is zero when nx is not an integer, we have (by (4.30)),

$$\frac{L_n(\lambda,\varepsilon)}{n^{1/2}\alpha_n} = \sum_{k\geq n\varepsilon}\frac{\Psi(k)}{\alpha_n}\frac{1}{\sqrt{n}}\mathsf{Mes}_n^{(t_n(\lambda))}(k) \tag{4.46}$$

$$= \left\langle 1_{[\varepsilon,1]}\Psi_\infty, \Theta_n^{(\lambda)}\right\rangle + \left\langle 1_{[\varepsilon,1]}\left(\frac{\Psi(\lfloor nx\rfloor)}{\alpha_n} - \Psi_\infty(x)\right), \Theta_n^{(\lambda)}\right\rangle. \tag{4.47}$$

By the a.s. convergence of $\Theta_n^{(\lambda)}$ to M_λ (for the vague topology on $(0,1)$), the fact that Ψ_∞ is continuous on $(0,1]$ (and that M_λ has a.s. no atom at ε, and by (4.44)),

$$\left\langle 1_{[\varepsilon,1]}\Psi_\infty, \Theta_n^{(\lambda)}\right\rangle \xrightarrow[n]{(as.)} \left\langle 1_{[\varepsilon,1]}\Psi_\infty, \overline{M}_\lambda\right\rangle.$$

Since $\frac{\Psi(\lfloor nx\rfloor)}{\alpha_n}\to\Psi_\infty(x)$ uniformly on $[\varepsilon,1]$ and by (4.45),

$$\left\langle 1_{[\varepsilon,1]}\left(\frac{\Psi(\lfloor nx\rfloor)}{\alpha_n} - \Psi_\infty(x)\right), \Theta_n^{(\lambda)}\right\rangle \xrightarrow[n]{(proba.)} 0. \tag{4.48}$$

□

Lemma 4.10. *[Analysis of $S_n(\lambda,\varepsilon)$] Let $\lambda\geq 0$ and $\eta,\delta>0$ be fixed. If ε is small enough, for all n large enough,*

$$\mathbb{P}\left(\frac{S_n(\lambda,\varepsilon)}{\alpha_n n^{1/2}}\geq\eta\right)\leq\delta.$$

Proof. It suffices to prove the Lemma for $\lambda=0$ (since $\lambda\mapsto S_n(\lambda,\varepsilon)$ is non-increasing). Recall that $\mathbf{s}_m$ is the size of cluster receiving the $m+1$th car: it is of size 0 if the chosen place $\mathbf{c}_{m+1}$ is free. The length of the cluster is observed before the insertion of this $m+1$th car. We have

$$\mathbb{P}(\mathbf{s}_m = k) = \mathbb{1}_{k=0}\frac{n-m}{n} + \mathbb{1}_{0<k\leq m}\binom{m}{k}n(k+1)^{k-1}(n-k-1)^{m-k}\frac{n-1-m}{n-k-1}\frac{1}{n^m}\frac{k}{n} \tag{4.49}$$

$$= \mathbb{1}_{k=0}\frac{n-m}{n} + \mathbb{1}_{0<k\leq m}\mathbb{P}\left(\mathsf{Bin}\left(m,\frac{k+1}{n}\right)=k\right)\frac{n-1-m}{n-k-1}\frac{k}{k+1} \tag{4.50}$$

where $\mathsf{Bin}(m,\frac{k+1}{n})$ is a binomial random variable with parameters m and $(k+1)/n$.
The proof of the first formula is mainly combinatorial (while the passage to the other is algebraic), the case $k=0$ being clear, let us see the case $k>0$. We will add the total weights of the parking histories leading to the event $\{\mathbf{s}_m = k\}$.

– The identities of the k cars of the cluster of size k is any subset of $\{1,\cdots,m\}$ of size k :this the factor $\binom{m}{k}$.
– Now, let us place this cluster somewhere on $\mathbb{Z}/n\mathbb{Z}$ by fixing the empty place e preceding it: n choices.
– The k cars we just talk about, must have fallen between $(e+1) \bmod n, \cdots e+k \bmod n$, and all of them must be finally parked on the same places: the number of such parking schemes is known to be $(k+1)^{k-1}$.
– The $(m-k)$ other cars must have fallen on $e+k+2 \bmod n, \cdots, e+1 \bmod n$ (in total $n-(k+1)$ places), and leave the last place free. To count the number of compatible positions of arrivals of cars, we use the (already used) trick to work on $\mathbb{Z}/(n-k-1)\mathbb{Z}$ instead and demand position zero to be empty: $(n-k-1)^{n-m}$ is the number of ways to put $n-m$ cars on $n-k-1$ places, and $\frac{n-1-m}{n-k-1}$ is the proportion of those letting zero free (by invariance by rotation).

Now, it remains to introduce the weight: $\frac{1}{n^m}$ is the probability of each story of the m first cars, and k/n is the probability that the $m+1$th car arrives in the block of size k, we were talking about all along.

We have, by bounding $k/(k+1) \leq 1$, and $1/(n-1-k) \leq 1/(n(1-\varepsilon))$,

$$\begin{aligned}
\mathbb{E}(S_n(1,\varepsilon)) &= \sum_{k<n\varepsilon} \sum_{m=1}^{n} \Psi(k)\mathbb{P}(\mathbf{s}_m = k) && (4.51)\\
&\leq \sum_{m=1}^{n} \frac{n-m}{n}\Psi(0) + \frac{1}{n(1-\varepsilon)} \sum_{k=1}^{n\varepsilon} \Psi(k) \sum_{m=1}^{n} \mathbb{P}\left(\mathsf{Bin}\left(m, \frac{k+1}{n}\right) = k\right)(n-m) && (4.52)
\end{aligned}$$

Observe that $\sum_{m\geq 0} \mathbb{P}\left(\mathsf{Bin}\left(m, \frac{k+1}{n}\right) = k\right) = 1 + n/(k+1)$ since this is the mean time, a random walk $(S_j^{(k)}, j \geq 0)$ with Bernoulli increments with parameter $(k+1)/n$ passes at level k: The first time $\tau_k^{(k)}$ at which such a random walk reaches k is around time m such that $m(k+1)/n$ is close to k, so, when m is around n, so that the factor $n-m-1$ in the rhs of (4.52) can not be bounded by n, otherwise, the upperbound would be much too large, and useless.

Rewrite

$$\begin{aligned}
\sum_{m=1}^{n} \mathbb{P}\left(\mathsf{Bin}\left(m, \frac{k+1}{n}\right) = k\right)(n-m) &= \mathbb{E}\left(\sum_{m=1}^{n} \mathbb{1}_{S_m^{(k)}=k}(n-m)\right) && (4.53)\\
&\leq \mathbb{E}\Big(\sum_{m=1}^{n} \mathbb{1}_{S_m^{(k)}=k}(n-\tau_k^{(k)})_+\Big) && (4.54)\\
&\leq \mathbb{E}((\tau_{k+1}^{(k)} - \tau_k^{(k)})(n-\tau_k^{(k)})_+) && (4.55)\\
&\leq \frac{n}{k+1}\mathbb{E}((n-\tau_k^{(k)})_+) && (4.56)\\
&\leq \frac{n}{k+1}\mathbb{E}(|n-\tau_k^{(k)}|), && (4.57)
\end{aligned}$$

the penultimate equality coming from the strong Markov property. Since $\tau_k^{(k)} \overset{(d)}{=} \sum_{j=1}^{k} G_j^{(k)}$ where the $G_j^{(k)}$ are geometric with parameter $(k+1)/n$; this is a negative binomial random variable with mean $\frac{kn}{k+1}$ and variance $k(1-(k+1)/n)/((k+1)/n)^2 = kn(n-k-1)/(k+1)^2$. Write, using $\mathbb{E}[|X|] \le \sqrt{\mathbb{E}\left[X^2\right]}$,

$$\begin{aligned}
\mathbb{E}\left[|n-\tau_k^{(k)}|\right] &\le \frac{n}{k+1} + \mathbb{E}\left[|n\frac{k}{k+1} - \tau_k^{(k)}|\right] \\
&\le \frac{n}{k+1} + \mathbb{E}\left(|n\frac{k}{k+1} - \tau_k^{(k)}|^2\right)^{1/2} \\
&\le \frac{n}{k+1} + \left(\frac{kn(n-k-1)}{(k+1)^2}\right)^{1/2} \\
&\le \frac{n}{k+1} + \frac{n^{1/2}(n-k-1)^{1/2}}{\sqrt{(k+1)}} \\
&\le n\varepsilon^{1/2}/\sqrt{k} + n/k,
\end{aligned}$$

this last inequality being valid since $k \le n\varepsilon$. Now, we conclude:

$$\mathbb{E}(S_n(1,\varepsilon)) \le (\log n)\Psi(0) + \frac{1}{n(1-\varepsilon)} \sum_{k=1}^{n\varepsilon} \Psi(k) \frac{n}{k+1}\left(\frac{n\varepsilon^{1/2}}{\sqrt{k}} + \frac{n}{k}\right) \tag{4.58}$$

$$\le (\log n)\Psi(0) + \frac{n\varepsilon^{1/2}}{1-\varepsilon} \sum_{k=1}^{n\varepsilon} \frac{\Psi(k)}{k^{3/2}} + \frac{n}{1-\varepsilon} \sum_{k=1}^{n\varepsilon} \frac{\Psi(k)}{k^2}. \tag{4.59}$$

Condition (4.7) is designed so that, for n large enough, this is $\lim_n \mathbb{E}(S_n/\alpha_n n^{1/2})$ is as small as wanted, up to take a small $\varepsilon > 0$; and then the Markov inequality allows to conclude the proof. $\square$

Fluctuation analysis. Since the random variables $\mathsf{UCost}_j^{(k)} - \Psi(k)$ are all centered and independent of $\mathsf{Mes}_n^{(t)}$ for all t, we have

$$\mathbb{E}(SF_n(\lambda,\varepsilon)) = \mathbb{E}(LF_n(\lambda,\varepsilon)) = 0.$$

Let us control the variance of $SF_n(\lambda,\varepsilon)$ and of $LF_n(\lambda,\varepsilon)$. We have $S_n(\lambda,\varepsilon) = \sum_{j=1}^{t_n(\lambda)} X_j^{(\mathbf{s}_j)} \mathbb{1}_{\mathbf{s}_j \le n\varepsilon}$ where the $(X_\ell^{(b)}, b \ge 0, \ell \ge 0)$ are independent random variables, and for all ℓ and b,

$$X_\ell^{(b)} \stackrel{(d)}{=} \mathsf{UCost}_1^{(b)} - \Psi(b);$$

we have

$$\mathrm{Var}(SF_n(\lambda,\varepsilon)) = \mathbb{E}\left[\left(\sum_{k=1}^{t_n(\lambda)} X_k^{(\mathbf{s}_k)} \mathbb{1}_{\mathbf{s}_k \le n\varepsilon}\right)^2\right]$$

and by conditioning by $\mathsf{Mes}_n^{(t_n(\lambda))}$, using that $\mathbb{E}\left[X_\ell^{(b)}\right] = 0, \mathrm{Var}(X_\ell^{(b)}) = \mathbb{E}\left[(X_\ell^{(b)})^2\right] = V(b)$, for all b, and by independence of the $X_i^{(j)}$, we obtain

$$\begin{cases} \mathrm{Var}(SF_n(\lambda,\varepsilon)) = \mathbb{E}\left[\sum_{k<n\varepsilon} \mathsf{Mes}_n^{(t_n(\lambda))}(\{k\}) V(k)\right], \\ \mathrm{Var}(LF_n(\lambda,\varepsilon)) = \mathbb{E}\left[\sum_{k\ge n\varepsilon} \mathsf{Mes}_n^{(t_n(\lambda))}(\{k\}) V(k)\right]. \end{cases} \tag{4.60}$$

Lemma 4.11. *Under the hypothesis of Theorem 4.4, for all $\eta, \delta > 0$ be fixed, if $\varepsilon > 0$ is small enough, for n large enough,*

$$\mathbb{P}\left(\left|\frac{SF_n(\lambda,\varepsilon)}{\alpha_n n^{1/2}}\right| \ge \eta\right) \le \delta \tag{4.61}$$

and

$$\frac{LF_n(\lambda,\varepsilon)}{\alpha_n n^{1/2}} \xrightarrow[n]{(proba.)} 0. \tag{4.62}$$

Proof. **Control of** $SF_n(\lambda,\varepsilon)$**.** Since these variables are centered, it suffices to control properly the variance. The variance $\mathrm{Var}(SF_n(\lambda,\varepsilon))$ given in (4.60) has exactly the same form as $\mathbb{E}(SF_n(\lambda,\varepsilon))$ as studied in the proof of Lemma 4.10, with $V(k)$ instead of $\Psi(k)$. By the same analysis to that of $\mathbb{E}(S_n(1,\varepsilon))$ we get

$$\mathrm{Var}(SF_n(1,\varepsilon)) \le (\log n)V(0) + n\varepsilon^{1/2}\sum_{k=1}^{n\varepsilon}\frac{V(k)}{k^{3/2}} + n\sum_{k=1}^{n\varepsilon}\frac{V(k)}{k^2};$$

Condition (4.8) is desiged so that $\limsup_n \mathrm{Var}(LF_n(\lambda,\varepsilon)/(\alpha_n n^{1/2})) = o(\varepsilon)$, and the Bienaymé-Tchebitchev inequality allows to deduce (4.61).
Control of $LF_n(\lambda,\varepsilon)$**.** Consider the variable

$$\begin{aligned} Y_{n,\varepsilon} := \mathbb{E}\left[LF_n(\lambda,\varepsilon)^2 \,|\mathsf{Mes}_n^{(t_n(\lambda))}\right] &= \mathbb{E}\left[\left(\sum_{k=1}^{t_n(\lambda)} X_k^{(\mathbf{s}_k)}\mathbb{1}_{\mathbf{s}_k\ge n\varepsilon}\right)^2 \,|\, \mathsf{Mes}_n^{(t_n(\lambda))}\right] \\ &= \sum_{k\ge n\varepsilon}\mathsf{Mes}_n^{(t_n(\lambda))}(\{k\})V(k); \end{aligned}$$

this variable is the variance of $LF_n(\lambda,\varepsilon)$ conditional on $\mathsf{Mes}_n^{(t_n(\lambda))}(\{k\})$. We then have

$$Y_{n,\varepsilon} = n\int_{x\ge\varepsilon}\mathsf{Mes}_n^{(t_n(\lambda))}(\{\lfloor nx\rfloor\})V(\lfloor nx\rfloor)dx; \tag{4.63}$$

so that,

$$\frac{Y_{n,\varepsilon}}{n\alpha_n^2} = \left\langle \frac{1}{\sqrt{n}\alpha_n^2}V(\lfloor n.\rfloor), \Theta_n^{(\lambda)}\right\rangle. \tag{4.64}$$

By hypothesis $\lim_n \sup_{x\ge\varepsilon}\frac{1}{\sqrt{n}\alpha_n^2}V(\lfloor nx\rfloor) = 0$, and by Theorem 4.7, $\Theta_n^{(\lambda)} \xrightarrow[n]{(d)} M_\lambda$ for the vague topology. This implies that $\frac{Y_{n,\varepsilon}}{n\alpha_n^2}$ converges to zero in probability, from what the conclusion follows. □

4.3 Discussion: Cost for general diffusion processes

Theorem 4.4 which is valid uniquely in the parking case (where the diffusion stops when the car get out of the initial occupied CC containing the arrival points), shows that even in this simplest case, the analysis is a bit subtle.

In general, the definition of the "total cost" of a general diffusion process, starts with the question of defining the unitary cost model, corresponding to the insertion and dispersion of a single mass. An important problem occurs: the dispersion of a single mass may result in the coalescence of many occupied CC. Hence, there are mainly two kinds of "unitary cost functions:
(a) those that depends uniquely of the size $\mathbf{s}_k$ of the occupied CC that contains the arrival position u_k,
(b) those that depends of all the occupied CC involved in the dispersion of each mass (for example, by using all the information contained in interval relaxation process IR_k).

The cost models that falls in the category (a) can be studied as for the parking model (under the same condition), if their collecting paths converge to the Brownian bridge b_λ (as stated in Proposition 3.6);the adaptation the Lemmas and theorems we used for the parking model would need to be adapted to this new model [10]. The same method could be applied to the subset of models of type (b) for which, at the first order, the unitary cost depends on $\mathbf{s}_k$ (for example, a cost that would depend on the size $\mathbf{s}'_k$ of the occupied CC that contains u_k, but after dispersion, should behave as if the size was taken before insertion, for many cost models, since very few coalescence concerns several linear size occupied CC).

Among the model of type (b), many would depend on small CC, and should not be accessible to this kind of analysis (for example, a unitary cost that would depend on the product of each of the occupied CC that merged during the kth insertion, would bring important additional problems).

5 Annexes

5.1 Justification of fluid approximation of Section 1.5.6

In order to justify the approximation of the interval evolution by this fluid limit, several method are possible: use the results on Polya urns (since the model is identical to a Polya urn starting from an urn composition $(\mathbf{a},\mathbf{b})$ (number of balls red/blue, with replacement matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$); this model is called in the literature the adverse-campaign model (Friedman) (see e.g. Section 2.2 in Flajolet & al.[13]) and it is accessible to exact computations, and in fact, it may be used as an exercise in martingale lectures. Letting $[\mathbf{a}(0),\mathbf{b}(0)]$ be given, at time t, $L(t) := \mathbf{b}(t)+\mathbf{a}(t) = t+(\mathbf{b}(0)+\mathbf{a}(0))$, and then knowing $\mathbf{b}(t)$ is sufficient to characterize the configuration at time t. Since conditional on $\mathbf{b}(t)$, $\mathbf{b}(t+1) = \mathbf{b}(t)+\mathbf{X}_t$ where $\mathbf{X}_t$ is a Bernoulli random variable with parameter $(L(t)-\mathbf{b}(t))/L(t)$ we get that

$$\mathbb{E}\left(\mathbf{b}(t+1)\ |\mathbf{b}(t)\right) = 1+\mathbf{b}(t)(1-1/L(t)). \tag{5.1}$$

From here, setting $m(t) := \mathbb{E}(\mathbf{b}(t))$, by taking the expectation in (5.1), we have

$$L(t)m(t+1) = L(t-1)m(t)+L(t) = \cdots = L(0)m(1)+\sum_{j=1}^{t} L(j) = L(0)m(1)+(t+1)(L(0))+(t+1)(t+2)/2;$$

so that

$$m(t+1) = \frac{L(0)m(1)+t(L(0))+(t+1)(t+2)/2}{t+L(0)},$$

with $m(1) = \mathbf{b}(0)+\mathbf{a}(0)/L(0)$. Now, the concentration around the mean is a standard consequence of exponential bounds for martingales with bounded increments (for example, Azuma's inequality). Now, assume that the balls have mass $1/M$ with M large, and that $a(0) = M\alpha_0$, $b(0) = M\beta_0$ corresponds to some large urns content, where $\alpha_0,\beta_0 > 0$. At time tM for some $t > 0$, $m(tM+1)/M$ is the mean position of the right extremity of the segment in this rescaled process,

$$\frac{m(tM+1)}{M} = \frac{1}{M}\frac{(\alpha_0+\beta_0)M\left(\beta_0 M+\alpha_0/(\alpha_0+\beta_0)\right)+tM(\alpha_0+\beta_0)M)+(tM+1)(tM+2)/2}{tM+(\alpha_0+\beta_0)M}$$

and we get

$$\frac{m(tM+1)}{M} \xrightarrow[M\to+\infty]{} \frac{(\alpha_0+\beta_0)\beta_0+t(\alpha_0+\beta_0))+t^2/2}{t+\alpha_0+\beta_0}$$

which can be seen to be identical to the formula giving $b(t)$ in (1.4) (in identifying (α_0,β_0) with $(a(0),b(0))$).

[10] For example, $\Theta_n^{(\lambda)}$ that encodes the insertion sequence $(\mathbf{s}_k, 0\le k\le t_n(\lambda))$ should converge toward M_λ possibly up to a constant, since the mean of the masses play a role in the time normalization

5.2 Justification of (1.19)

This says that $a_0 \in [-M_1 - a_1, 0], (a_2 - a_1) \geq M_1, \cdots, a_{n-1} - a_{n-2} \geq M_{n-2}$. We have

$$\begin{aligned} \mathsf{TL}(M[n]) &= \int_{0<y_0<\cdots<y_{n-2}<1-W(m[k])} (M_0 + y_0)\, dy_0 \cdots dy_{n-2} && (5.2) \\ &= M_0 \frac{(1-W(m[k]))^{n-1}}{(n-1)!} + \frac{(1-W(m[k]))^n}{(n)!} && (5.3) \end{aligned}$$

Consider the set $S := \{(u, s_0, \cdots, s_{n-2}), u \in [0, M_0 + s_0], 0 \leq s_0 \leq s_1 \leq \cdots s_{n-2} \leq 1 - W(m[j])\}$ (the variable s_j cumulates the free length spaces between the $\mathbf{O}_0^{(k)}$ and $\mathbf{O}_{j+1}^{(k)}$); the map

$$\begin{array}{rccl} \Psi: & S & \longrightarrow & \mathscr{C}^n \\ & (u, s_0, \cdots, s_{n-2}) & \longmapsto & (-u, -u + s_0 + M_0, \cdots, -u + s_{n-2} + (M_0 + \cdots + M_{n-2})) \end{array}$$

is linear. Its Jacobian is 1 (the linear map has determinant -1), and sends S onto the set $(a_0, \cdots, a_{n-1})$ described in Theorem 1.9.

5.3 Computation of $\mathbb{E}\left(\langle x \mapsto x^k, M_\lambda \rangle\right)$

We consider first the case when $f = \mathsf{Id}$, that is $\mathbb{E}\left(\langle x \mapsto x^1, M_\lambda \rangle\right)$; we recall that Aldous CRT usual convention, is to consider that the CRT is the continuous tree whose contour process is 2e.

According to Theorem 4.6, in this case,

$$\langle \mathsf{Id}, M_\lambda(\mathsf{e}) \rangle = \int_\lambda^{+\infty} \sum_{m \in \mathscr{L}(\mathsf{e})} (\ell(m,t))^2 \,\mathrm{d}t \qquad (5.4)$$

and we know thanks to (4.6) that the right hand side is equal to $\int_0^1 2\mathsf{e}_\lambda(x) - 2\underline{\mathsf{e}}_\lambda(x)\, dx$.

We first discuss the case $\lambda = 0$. On the one hand, we know that

$$\int_0^1 2\mathsf{e}_0(x) dx = \int_0^1 2\mathsf{e}(x) dx = \mathbb{E}_\mathsf{e}\left[2\mathsf{e}(U)\right] \overset{(d)}{=} \mathbb{E}_\mathsf{e}\left[d_{2\mathsf{e}}(U_1, U_2)\right] \qquad (5.5)$$

where U is uniform on $[0,1]$ (independent of e), and U_1 and U_2 are two uniform points taken in the CRT $T_{2\mathsf{e}}$ (in other words, using 2e as a contour process, the points U_1 and U_2 corresponds by the canonical projection $[0,1]_{\sim 2\mathsf{e}}$, to the image of independent uniform random variables u_1, u_2 on $[0,1]$). The second equality comes from the fact that the RHS can be seen as the average height of $2e$ on $[0,1]$. Moreover, if we consider a continuous random tree characterized by 2e, $2\mathsf{e}(U)$ is also the distance between a uniform point of the tree and the root. Thus, up to a uniform re-rooting of the tree (operation which preserves the continuum random tree), it is also the distance between two uniform points.

On the other hand, at some time λ, and for $t \geq \lambda$, $\sum_{m \in \mathscr{L}(2\mathsf{e})} (\ell(m,t))^2 = \sum_{m \in \mathscr{L}(\mathsf{e})} (\ell(m,t))^2$ corresponds to the probability that two points chosen uniformly at random in the CRT $T_{2\mathsf{e}}$ are in the same CC at time t in the fragmentation (corresponding to the time reversal of the additive coalescent). As proved by Aldous & Pitman [3], in order to construct this fragmentation process, one may equip the CRT $T_{2\mathsf{e}}$ with a Poisson point process with intensity 1 on the product $Skel(T_{2\mathsf{e}}) \times \mathbb{R}^+$, where $Skel(T_{2\mathsf{e}})$ is set of vertices of degree 2 of the tree (the skeleton can be seen as the spanning tree of a countable number of points taken uniformly at random in the CRT, and of its root). A point $\xi = (u,t)$ of this Poisson point process corresponds to a fragmentation event: the node u is removed at time t.

Take two uniform points U_1 and U_2 in the CRT $T_{2\mathrm{e}}$: the probability that they are still in the same component at time t is, conditional on their distance $D = d_{2\mathrm{e}}(U_1, U_2)$ in the CRT, $\mathbb{P}(\mathsf{Expo}(D) \geq t \mid D) = e^{-Dt}$ (for $t \geq 0$), where $\mathsf{Expo}(D)$ is an exponential random variable with intensity D. Hence, for any $t \geq 0$,

$$\sum_{m\in\mathscr{L}(\mathrm{e})} \ell(m,t)^2 \overset{(d)}{=} \exp(-t d_{2\mathrm{e}}(U_1,U_2)) \tag{5.6}$$

since both sides gives the probability that two uniform points are still in the same component of the fragmented CRT at time t, during the fragmentation process. As a process in t, the two sides of (5.6) are not identical in distribution, since the right hand side is a deterministic function of its value at any time $t = 1$. But the relation (5.6) is sufficient to compute the mean:

$$\mathbb{E}\Big[\int_\lambda^{+\infty} \sum_{m\in\mathscr{L}(\mathrm{e})} \ell(m,t)^2 \mathrm{d}t\Big] = \mathbb{E}\Big[\int_\lambda^{+\infty} \exp(-t d_{2\mathrm{e}}(U_1,U_2))\mathrm{d}t\Big] = \mathbb{E}\left[\frac{\exp(-\lambda d_{2\mathrm{e}}(U_1,U_2))}{d_{2\mathrm{e}}(U_1,U_2)}\right]. \tag{5.7}$$

This can be computed, because $d_{2\mathrm{e}}(U_1,U_2)$ is Rayleigh distributed (see e.g. Aldous [1, Lemma 21], [3, Section 2.1]) which is the distribution with density $xe^{-x^2/2}\mathbb{1}_{x\geq 0}$. The rhs in (5.7) is

$$\int_0^{+\infty} \frac{\exp(-\lambda x/2)}{x/2} xe^{-x^2/2}\mathrm{d}x = \sqrt{\pi/2}\mathrm{e}^{\lambda^2/2}\left(1-\mathrm{erf}\left(\lambda/\sqrt{2}\right)\right).$$

Computation of $\mathbb{E}\left(\langle x \mapsto x^m, M_\lambda(\mathrm{e})\rangle\right)$ **for** $m \geq 2$, where m is an integer. We have, by Theorem 4.6

$$\mathbb{E}\left(\langle x \mapsto x^m, M_\lambda(\mathrm{e})\rangle\right) = \mathbb{E}\left(\int_\lambda^{+\infty} \sum_{k\geq 0} \ell_k^{m+1}(t)\,\mathrm{d}t\right)$$

and by the interpretation done before, $\sum_{k\geq 0} \ell_k^{m+1}(t)\,\mathrm{d}t$ is the probability, given the process e, that $m+1$ uniform random points $U_1,\cdots,U_{m+1}$ are still in the same component of the Poisson fragmentation of the CRT at time t. When one takes $m+1 \geq 2$ random points in the CRT, the total lengths L_{m+1} of the tree spanned by these points satisfies

$$\mathbb{P}(L_{m+1} \geq y) = \mathbb{P}(N(y^2/2) < m) = \sum_{j=0}^{m-1} \exp(-y^2/2)(y^2/2)^j/j!, \tag{5.8}$$

where $N(a)$ is a Poisson random variable with parameter a (by [3, Theo.8]). Knowing L_{m+1}, the $U_1,\cdots,U_{m+1}$ and still in the same components with probability $\exp(-tL_{m+1})$ at time t. Hence

$$\mathbb{E}\left(\langle x \mapsto x^m, M_\lambda(\mathrm{e})\rangle\right) = \mathbb{E}\left(\exp(-\lambda L_{m+1})\right).$$

which can be computed thanks to (5.8), and gives the explicit expression,

$$\mathbb{E}\left(\langle x \mapsto x^m, M_\lambda(\mathrm{e})\rangle\right) = \int_0^\infty \exp(-\lambda y)\frac{\mathrm{d}}{\mathrm{d}y}(1-\mathbb{P}(L_{m+1}\geq y))dy.$$

Conclusion

The universal properties of the valid continuous dispersion model models described here are one–dimensional result. In this case, the geometry of the CC are well encoded by their lengths. In 2D (on a torus $(\mathbb{R}/\mathbb{Z})^2$, for example), there are no analogue of Theorem 1.5(i) (with the same degree of generality) which states that in 1D, the distribution of $(\mathbf{O}^{(k)},\mathbf{F}^{(k)})$ depends only on the masses, and not on the details of the valid continuous dispersion model considered: the coalescence induced by a continuous dispersion model producing Euclidean balls (say, making components grow on their boundary, with the same speed everywhere) is nothing compared to continuous dispersion model which would produce very thin hairs visiting a large part of $(\mathbb{R}/\mathbb{Z})^2$. However, we can still define, beyond the piling propensity, the cluster distribution as the distribution of the dispersion of some masses conditioned to form a single CC. The distribution of the configuration at time k can still be written as a kind of product of the clusters probability, weighted by the Lebesgue measure of the translation space (that measures the "amount" of position at which we can place these clusters while avoiding intersection). This could maybe be used to study the statistical properties of some continuous dispersion model.

Another possibility, is to stay on "more complex one dimensional spaces", as graphs or grid.

References

[1] D. Aldous. The continuum random tree. III. Ann. Probab., 21(1):248–289, 1993.

[2] D. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. Ann. Probab., 25(2):812–854, 1997.

[3] D. Aldous and J. Pitman. The standard additive coalescent. Ann. Probab., 26(4):1703–1726, 1998.

[4] D. J. Aldous. Exchangeability and related topics. In École d'été de probabilités de Saint-Flour, XIII—1983, volume 1117 of Lecture Notes in Math., pages 1–198. Springer, Berlin, 1985.

[5] J. Bertoin. Random fragmentation and coagulation processes, volume 102 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2006.

[6] J. Bertoin and G. Miermont. Asymptotics in Knuth's parking problem for caravans. Random Structures and Algorithms, 29(1):38–55, Aug. 2006. 18 pages, 2 figures.

[7] N. Broutin and J.-F. Marckert. Asymptotics of trees with a prescribed degree sequence and applications. Random Structures Algorithms, 44(3):290–316, 2014.

[8] N. Broutin and J.-F. Marckert. A new encoding of coalescent processes: applications to the additive and multiplicative cases. Probab. Theory Related Fields, 166(1-2):515–552, 2016.

[9] P. Chassaing and G. Louchard. Phase transition for parking blocks, brownian excursion and coalescence. Random Structures & Algorithms, 21(1):76–119, 2002.

[10] P. Chassaing and J.-F. Marckert. Parking functions, empirical processes, and the width of rooted labeled trees. Electron. J. Combin., 8(1):Research Paper 14, 19, 2001.

[11] K. L. Chung. Green, Brown, and probability. World Scientific Publishing Co., Inc., River Edge, NJ, 1995.

[12] P. Diaconis and W. Fulton. A growth model, a game, an algebra, Lagrange inversion, and characteristic classes. volume 49, pages 95–119. 1991. Commutative algebra and algebraic geometry, II (Italian) (Turin, 1990).

[13] P. Flajolet, P. Dumas, and V. Puyhaubert. Some exactly solvable models of urn process theory. volume AG of Discrete Math. Theor. Comput. Sci. Proc., pages 59–118. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2006.

[14] P. Flajolet, P. Poblete, and A. Viola. On the analysis of linear probing hashing. Algorithmica, 22(4):490–515, 1998. Average-case analysis of algorithms.

[15] L. Fredes and J.-F. Marckert. A combinatorial proof of Aldous-Broder theorem for general Markov chains. Random Structures Algorithms, 62(2):430–449, 2023.

[16] S. Janson. Asymptotic distribution for the cost of linear probing hashing. Random Structures Algorithms, 19(3-4):438–471, 2001. Analysis of algorithms (Krynica Morska, 2000).

[17] S. Janson. Individual displacements for linear probing hashing with different insertion policies. ACM Trans. Algorithms, 1(2):177–213, 2005.

[18] D. E. Knuth. The art of computer programming. Vol. 3. Addison-Wesley, Reading, MA, second edition, 1998. Sorting and searching.

[19] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 34(1):33–58, 1976.

[20] A. G. Konheim and B. Weiss. An occupancy discipline and applications. SIAM Journal on Applied Mathematics, 14(6):1266–1274, 1966.

[21] G. Lawler. Conformally invariant processes in the plane, volume 114 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2005.

[22] G. F. Lawler, M. Bramson, and D. Griffeath. Internal diffusion limited aggregation. Ann. Probab., 20(4):2117–2140, 1992.

[23] D. Lépingle. Une remarque sur les lois de certains temps d'atteinte. In Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French), volume 850 of Lecture Notes in Math., pages 669–670. Springer, Berlin, 1981.

[24] P. Marchal. Loop-erased random walks, spanning trees and Hamiltonian cycles. Electron. Comm. Probab., 5:39–50, 2000.

[25] J.-F. Marckert. One more approach to the convergence of the empirical process to the Brownian bridge. Electron. J. Stat., 2:118–126, 2008.

[26] J.-F. Marckert and M. Wang. A new combinatorial representation of the additive coalescent. Random Structures Algorithms, 54(2):340–370, 2019.

[27] P. Nadeau and V. Tewari. The permutahedral variety, mixed Eulerian numbers, and principal specializations of Schubert polynomials. Int. Math. Res. Not. IMRN, (5):3615–3670, 2023.

[28] J. Pitman. Coalescent random forests. J. Combin. Theory Ser. A, 85(2):165–193, 1999.

[29] B. Pittel. Linear probing: the probable largest search time grows logarithmically with the number of records. Journal of algorithms, 8(2):236–249, 1987.

[30] J. Propp and D. Wilson. How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. Journal of Algorithms, 27(2):170–217, 1998.

[31] Z. Varin. The golf model on $\mathbb{Z}/n\mathbb{Z}$ and on $\mathbb{Z}$. https://arxiv.org/abs/2401.13380, 2024.

[32] X. Viennot. Heaps of pieces, i : Basic definitions and combinatorial lemmas. In G. Labelle and P. Leroux, editors, Combinatoire énumérative, pages 321–350. Springer Berlin Heidelberg, 1986.

[33] D. Zeilberger. A combinatorial approach to matrix algebra. Discrete Mathematics, 56(1):61 – 72, 1985.

Contents