

## PARTIAL MATCH QUERIES IN TWO-DIMENSIONAL QUADTREES: A PROBABILISTIC APPROACH

NICOLAS CURIEN,\* *École Normale Supérieure*

ADRIEN JOSEPH,\*\* *Université Pierre et Marie Curie*

### Abstract

We analyze the mean cost of the partial match queries in random two-dimensional quadrees. The method is based on fragmentation theory. The convergence is guaranteed by a coupling argument of Markov chains, whereas the value of the limit is computed as the fixed point of an integral equation.

*Keywords:* Quadtree; partial match query; fragmentation theory; Markov chain; coupling; integral equation

2010 Mathematics Subject Classification: Primary 60F99

Secondary 60G18; 60J05

### 1. Introduction

Introduced by Finkel and Bentley [5], the quadtree structure is a comparison-based algorithm designed for retrieving multidimensional data. It is often studied in computer science because of its numerous applications. The aim of this paper is to study the mean cost of the so-called *partial match queries* in random quadtrees. This problem was first analyzed by Flajolet *et al.* [7].

Let us briefly describe the discrete model. We choose to focus only on the two-dimensional case. Let  $P_1, \dots, P_n$  be  $n$  independent random variables uniformly distributed over  $(0, 1)^2$ . We will assume that the points have different  $x$  and  $y$  coordinates, an event that has probability 1. We construct iteratively a finite covering of  $[0, 1]^2$  composed of rectangles with disjoint interiors as follows. The first point  $P_1$  divides the original square  $[0, 1]^2$  into four closed quadrants according to the vertical and horizontal positions of  $P_1$ . By induction, a point  $P_k$  divides the quadrant in which it falls into four quadrants according to its position in this quadrant; see Figure 1. Hence, the  $n$  points  $P_1, \dots, P_n$  give rise to a covering of  $[0, 1]^2$  into  $3n + 1$  closed rectangles with disjoint interiors that we denote by  $\text{Quad}(P_1, \dots, P_n)$ .

We are interested in the partial match query. As explained by Flajolet and Sedgewick [6, Example VII.23], given  $x_0 \in [0, 1]$ , it determines the set of points  $P_i$ ,  $i \in \{1, \dots, n\}$ , with  $x$  coordinates equal to  $x_0$ , regardless of the  $y$  coordinates (that set is either empty or a singleton). Denoting the vertical segment  $[(x, 0), (x, 1)]$  by  $S_x$ , the cost of this partial match query is measured by the number  $\mathcal{N}_n(x)$  of rectangles of  $\text{Quad}(P_1, \dots, P_n)$  intersecting  $S_x$  minus 1 ( $\mathcal{N}_0(x) = 0$  by convention). We will study the cost of a *fixed* query. Our main result is as follows.

---

Received 17 September 2010; revision received 11 November 2010.

\* Postal address: Département de Mathématiques et Applications, École Normale Supérieure, 45 rue d'Ulm, 75005 Paris, France. Email address: nicolas.curien@ens.fr

\*\* Postal address: Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie, 4 place Jussieu Tour 16-26, 75005 Paris, France. Email address: adrien.joseph@upmc.fr

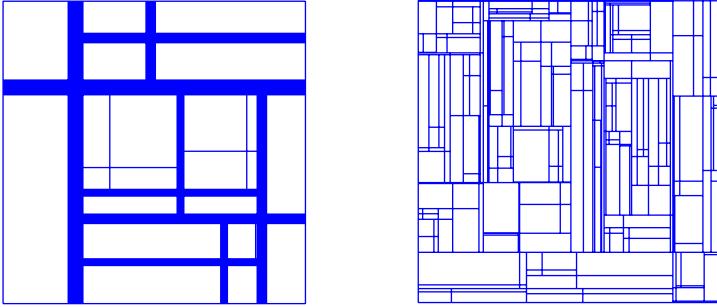


FIGURE 1: Two splittings of  $[0, 1]^2$  with 8 (left) and 100 (right) points.

**Theorem 1.** For every  $x \in [0, 1]$ , we have the following convergence:

$$n^{-\beta^*} \mathbb{E}[\mathcal{N}_n(x)] \rightarrow K_0(x(1-x))^{\beta^*/2} \quad \text{as } n \rightarrow \infty.$$

Here

$$\beta^* = \frac{\sqrt{17} - 3}{2} \quad \text{and} \quad K_0 = \frac{\Gamma(2\beta^* + 2)\Gamma(\beta^* + 2)}{2\Gamma^3(\beta^* + 1)\Gamma^2(\beta^*/2 + 1)}.$$

Flajolet *et al.* [7] proved the convergence in mean of the properly rescaled cost of partial match queries when  $x$  is *random* with the uniform law on  $[0, 1]$  and independent of  $P_1, \dots, P_n$ . See also [3] for a more precise asymptotic behavior. We will give another proof of this result using *fragmentation theory* (see Corollary 1 below). As a by-product of our techniques, we will also prove in Corollary 2 below that, when rescaled by  $n^{1-\sqrt{2}}$ ,  $\mathcal{N}_n(0)$  converges in  $\mathbb{L}^2$  (its convergence in mean was obtained in [7]).

The paper is organized as follows. In Section 2 we introduce the model embedded in continuous time and present the first properties. Section 3 is devoted to the link between quadtrees and fragmentation theory. Section 4, the most technical section, contains the proof of the convergence at a fixed point  $x$  without knowing the limit. The identification of the limit is done in Section 5 using a fixed-point argument for the integral equation.

## 2. Notation and first properties

In order to apply probabilistic techniques, we first introduce a continuous-time version of the quadtree: the points  $P_1, \dots, P_n$  are replaced by the arrival points of a Poisson point process over  $\mathbb{R}_+ \times [0, 1]^2$  with intensity  $dt \otimes dx dy$ . All the results obtained in this model can easily be translated into results for the discrete-time model.

### 2.1. The continuous-time model

Let  $\Pi$  be a Poisson point process on  $\mathbb{R}_+ \times [0, 1]^2$  with intensity  $dt \otimes dx dy$ . Let  $((\tau_i, x_i, y_i), i \geq 1)$  be the atoms of  $\Pi$  ranked in increasing order of their  $\tau$ -component. We define a process  $(Q(t))_{t \geq 0}$  with values in a finite covering of  $[0, 1]^2$  by closed rectangles with disjoint interiors as follows. We first introduce the operation **SPLIT**: for every subset  $R$  of  $[0, 1]^2$  and every  $(x, y) \in [0, 1]^2$ ,

$$\text{SPLIT}(R, x, y) = \{R \cap [0, x] \times [0, y], R \cap [0, x] \times [y, 1], R \cap [x, 1] \times [0, y], R \cap [x, 1] \times [y, 1]\}.$$

In other words, if  $R$  is a rectangle with sides parallel to the  $x$  and  $y$  axes, then  $\text{SPLIT}(R, x, y)$  is the set of four quadrants in  $R$  determined by the point  $(x, y)$ . We may now recursively define the process  $(Q(t))_{t \geq 0}$ . Let  $\tau_0 = 0$ . For every  $t \in [0, \tau_1)$ , define  $Q(t) = \{[0, 1]^2\}$ , and, for every  $t \in [\tau_i, \tau_{i+1})$ , denoting by  $R$  the only element (if any) of  $Q(\tau_{i-1})$  such that  $(x_i, y_i)$  is in the interior of the rectangle  $R$ , let

$$Q(t) = \text{SPLIT}(R, x_i, y_i) \cup Q(\tau_{i-1}) \setminus \{R\}.$$

Observe that, almost surely (a.s.), for every  $i \in \mathbb{Z}_+$ , there indeed exists a unique rectangle of  $Q(\tau_i)$  such that  $(x_{i+1}, y_{i+1})$  is in its interior; hence, the process  $(Q(t))_{t \geq 0}$  is well defined up to an event of zero probability. In the sequel we will assume that the points of  $\Pi$  always fall in the interior of some rectangle of  $(Q(t))_{t \geq 0}$ . As explained in the introduction, we are interested in the number of rectangles of  $Q(t)$  intersecting the segment  $S_x$ ; specifically, we set

$$N_t(x) = \#\{R \in Q(t) : R \cap S_x \neq \emptyset\} - 1,$$

so that  $N_t(x) = 0$  for every  $0 \leq t < \tau_1$ . Recalling that  $\tau_n$  is the arrival time of the  $n$ th point of  $\Pi$ ,  $Q(\tau_n)$  has the same distribution as the random variable  $\text{Quad}(P_1, \dots, P_n)$  of the introduction. In particular, for every  $(n, x) \in \mathbb{N} \times [0, 1]$ , we have  $N_{\tau_n}(x) = \mathcal{N}_n(x)$  in distribution.

### 2.2. Main equations

Let  $x \in [0, 1]$ . We denote by  $\mathcal{A}$  the set of words over the alphabet  $\{0, 1\}$ , i.e.

$$\mathcal{A} = \bigcup_{n \geq 0} \{0, 1\}^n,$$

where, by convention,  $\{0, 1\}^0 = \{\emptyset\}$ . Thus, if  $u \in \mathcal{A}$ ,  $u$  is either  $\emptyset$  or a finite sequence of 0 and 1. If  $u$  and  $v$  are elements of  $\mathcal{A}$ , then  $uv$  denotes the concatenation of the two words  $u$  and  $v$ . We label the rectangles appearing in  $(Q(t))_{t \geq 0}$  whose intersection with the segment  $S_x$  is nonempty by elements of  $\mathcal{A}$  according to the following rule. By convention,  $R_\emptyset(x)$  is the unit square  $[0, 1]^2$ . The first point  $(\tau_1, x_1, y_1)$  of  $\Pi$  splits  $[0, 1]^2$  into four rectangles, where, a.s., only two of them intersect  $S_x$ ; we denote the bottom rectangle by  $R_0(x)$  and the top rectangle by  $R_1(x)$ . Inductively, for every  $u \in \mathcal{A}$ , a point of  $\Pi$  eventually falls into  $R_u(x)$ , dividing it into four rectangles. Almost surely, only two of them intersect  $S_x$ ; denote the bottom rectangle by  $R_{u0}(x)$  and the top rectangle by  $R_{u1}(x)$ . For  $u \in \mathcal{A}$ , we denote the minimal and maximal horizontal coordinates of  $R_u(x)$  by  $G_u(x)$  and  $D_u(x)$ , respectively, and define the *place* of  $x$  in  $R_u(x)$  to be

$$X_u(x) = \frac{x - G_u(x)}{D_u(x) - G_u(x)}.$$

If  $u \neq \emptyset$ , we denote the parent of  $u$  by  $\overleftarrow{u}$ , which is the word  $u$  without its last letter. We write  $M_u(x)$  for the ratio of the (two-dimensional) Lebesgue measure  $\text{Leb}(R_u(x))$  of  $R_u(x)$  and the measure of  $R_{\overleftarrow{u}}(x)$ , i.e.

$$M_u(x) = \frac{\text{Leb}(R_u(x))}{\text{Leb}(R_{\overleftarrow{u}}(x))}.$$

We also set, for all  $x \in [0, 1]$ ,  $M_\emptyset(x) = 1$ . For  $u \in \{0, 1\}$  and  $t \geq 0$ , we introduce the ‘subquadtrees’  $Q_{u,x}(t) = \{R \in Q(t + \tau_1) : R \subset R_u(x)\}$ . Then, for every  $t \geq 0$ , we have

$$N_t(x) = \mathbf{1}_{\{t \geq \tau_1\}} + \mathbf{1}_{\{t \geq \tau_1\}} \sum_{u \in \{0,1\}} (\#\{R \in Q_{u,x}(t - \tau_1) : R \cap S_x \neq \emptyset\} - 1). \tag{1}$$

If  $R$  is a rectangle with sides parallel to the  $x$  and  $y$  axes, we denote by  $\Phi_R: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  the only affine transformation that maps the bottom-left vertex of  $R$  to  $(0, 0)$ , the bottom-right vertex of  $R$  to  $(1, 0)$ , and the upper-left vertex of  $R$  to  $(0, 1)$ . It should be plain from properties of Poisson point measures that, conditionally on  $(M_u(x), X_u(x), R_u(x))$ , the process  $(\Phi_{R_u(x)}(Q_{u,x}(t)))_{t \geq 0}$  has the same distribution as the process  $(\tilde{Q}(M_u(x)t))_{t \geq 0}$ , where  $\tilde{Q}$  is an independent copy of  $Q$ . In particular, conditionally on  $(M_u(x), X_u(x))$ , the number of rectangles in  $Q_{u,x}$  that intersect  $S_x$  (minus 1), viewed as a process of  $t$ , has the same distribution as the process  $(\tilde{N}_{M_u(x)t}(X_u(x)))_{t \geq 0}$ , where  $\tilde{N}$  is defined from  $\tilde{Q}$  in the same way as  $N$  is defined from  $Q$ . Since  $M_0(x)$  and  $M_1(x)$  have the same distribution, (1) yields

$$E[N_t(x)] = P(t \geq \tau_1) + 2E[\tilde{N}_{M_0(x)(t-\tau_1)}(X_0(x))], \tag{2}$$

with the convention that  $\tilde{N}_t(x) = 0$  whenever  $t < 0$ . More generally, if we write  $\mathfrak{z}_k \in \mathcal{A}$  for  $\mathfrak{z}_k = 0 \cdots 0$  repeated  $k$  times, then, for every positive integer  $k$ ,

$$E[N_t(x)] = g_k(t) + 2^k E[\tilde{N}_{M_{\mathfrak{z}_1}(x) \cdots M_{\mathfrak{z}_k}(x)t - F_k}(X_{\mathfrak{z}_k}(x))], \tag{3}$$

where  $g_k$  is a function such that  $0 \leq g_k \leq 2^k - 1$  and  $F_k$  is a nonnegative random variable defined by

$$F_k = \sum_{i=1}^k \tilde{\tau}_i \prod_{j=i}^k M_{\mathfrak{z}_j}(x),$$

with  $(\tilde{\tau}_i)_{i \geq 1}$  a sequence of independent exponential variables with parameter 1.

We now compute the joint distribution of  $(M_0(x), X_0(x))$ , which will be of great use throughout this work. If  $f$  is a nonnegative measurable function, easy calculations yield

$$\begin{aligned} E[f(M_0(x), X_0(x))] &= \int_0^1 du \int_0^1 dv \left( \mathbf{1}_{\{x < u\}} f\left(uv, \frac{x}{u}\right) + \mathbf{1}_{\{x > u\}} f\left((1-u)v, \frac{x-u}{1-u}\right) \right) \\ &= \int_x^1 \frac{dy}{y} \int_0^{x/y} dm f(m, y) + \int_0^x \frac{dy}{1-y} \int_0^{(1-x)/(1-y)} dm f(m, y) \end{aligned} \tag{4}$$

$$\begin{aligned} &= \int_0^x dm \int_x^1 \frac{dy}{y} f(m, y) + \int_x^1 dm \int_x^{x/m} \frac{dy}{y} f(m, y) \\ &\quad + \int_0^{1-x} dm \int_0^x \frac{dy}{1-y} f(m, y) + \int_{1-x}^1 dm \int_{1-(1-x)/m}^x \frac{dy}{1-y} f(m, y). \end{aligned} \tag{5}$$

### 2.3. Depoissonization

The following lemma contains a large deviations argument that will enable us to shift results from the continuous-time model to the discrete-time model.

**Lemma 1.** *For every  $\varepsilon > 0$ , we have*

$$E \left[ \sup_{x \in [0, 1]} |N_{\tau_n}(x) - N_n(x)|^2 \mathbf{1}_{\{\tau_n \notin [n(1-\varepsilon), n(1+\varepsilon)]\}} \right] \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

*Proof.* Note that, for every  $x \in [0, 1]$ ,  $t \mapsto N_t(x)$  is nondecreasing and that  $N_t(x)$  is at most the number of points fallen so far:  $N_t(x) \leq \max\{i \in \mathbb{Z}_+ : \tau_i \leq t\}$ . In particular,  $N_{\tau_n}(x) \leq n$ ; thus, we have

$$\sup_{x \in [0, 1]} |N_{\tau_n}(x) - N_n(x)|^2 \mathbf{1}_{\{\tau_n > n(1+\varepsilon)\}} \leq n^2 \mathbf{1}_{\{\tau_n > n(1+\varepsilon)\}}.$$

A large deviations argument ensures that  $n^2 \mathbb{P}(\tau_n > n(1 + \varepsilon))$  tends to 0 as  $n \rightarrow \infty$ . On the other hand, applying the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} & \mathbb{E} \left[ \sup_{x \in [0, 1]} |N_{\tau_n}(x) - N_n(x)|^2 \mathbf{1}_{\{\tau_n < n(1-\varepsilon)\}} \right] \\ & \leq \sqrt{\mathbb{E}[(\max\{i \in \mathbb{Z}_+ : \tau_i \leq n\})^4]} \sqrt{\mathbb{P}(\tau_n < n(1 - \varepsilon))}. \end{aligned}$$

As  $\mathbb{E}[(\max\{i \in \mathbb{Z}_+ : \tau_i \leq n\})^4] = O(n^4)$ , large deviations ensure that the quantity on the right-hand side tends to 0 as  $n \rightarrow \infty$ . This completes the proof.

### 3. Particular cases and fragmentation theory

We give below the definition for a particular case of fragmentation process. For more details, we refer the reader to [1, pp. 6–65]. Let  $\nu$  be a probability measure on  $\{(s_1, s_2) : s_1 \geq s_2 > 0 \text{ and } s_1 + s_2 \leq 1\}$ . A self-similar fragmentation  $(\mathfrak{F}_t)_{t \geq 0}$  with dislocation measure  $\nu$  and index of self-similarity 1 is a Markov process with values in the set  $\mathcal{S}^\downarrow = \{(s_1, s_2, \dots) : s_1 \geq s_2 \geq \dots \geq 0 \text{ and } \sum_i s_i \leq 1\}$  describing the evolution of the masses of particles that undergo fragmentation. The process is informally characterized as follows: if at time  $t$  we have  $\mathfrak{F}(t) = (s_1(t), s_2(t), \dots)$  then, for every  $i \geq 1$ , the  $i$ th ‘particle’ of mass  $s_i(t)$  lives an exponential time with parameter  $s_i(t)$  before splitting into two particles of masses  $r_1 s_i(t)$  and  $r_2 s_i(t)$ , where  $(r_1, r_2)$  has been sampled from  $\nu$  independently of the past and of the other particles. In other words, each particle undergoes a self-similar fragmentation with time rescaled by its mass. In the next section we establish a link between fragmentation theory and the process  $N_t(U)$ , where  $U$  is a random variable uniformly distributed over  $[0, 1]$  and independent of  $(Q(t))_{t \geq 0}$ . This connection will provide a new proof of a result of [3] and [7]. See also [4] for another recent application of fragmentation theory to a combinatorial problem where the exponent  $(\sqrt{17} - 3)/2$  appears.

#### 3.1. The uniform case

We consider here the case where the point  $x$  is chosen at random uniformly over  $[0, 1]$  and independently of  $(Q(t))_{t \geq 0}$ .

**Proposition 1.** *Let  $U$  be a random variable uniformly distributed over  $[0, 1]$  and independent of the quadtree  $(Q(t))_{t \geq 0}$ . Let  $u \in \mathcal{A}$ , and denote by  $u_0 = \emptyset, u_1, \dots, u_k = u$  its ancestors. Then  $X_u(U)$  is uniform over  $[0, 1]$  and independent of  $(M_{u_1}(U), \dots, M_{u_k}(U))$ , which is a sequence of independent random variables all having density  $2(1 - m)\mathbf{1}_{\{m \in [0, 1]\}}$ .*

*Proof.* We prove Proposition 1 by induction on  $k$ . Let  $u \in \mathcal{A}$ . Denote by  $u_0 = \emptyset, u_1, \dots, u_k = u$  its ancestors. Integrating (4) for  $x \in [0, 1]$ , we deduce that, for every  $v \in \{0, 1\}$ ,  $X_v(U)$  and  $M_v(U)$  are independent and distributed according to

$$\mathbf{1}_{\{u \in [0, 1]\}} du \otimes \mathbf{1}_{\{m \in [0, 1]\}} 2(1 - m) dm. \tag{6}$$

Recalling that  $Q_{u_1, U}(t) = \{R \in Q(t + \tau_1) : R \subset R_{u_1}(U)\}$ , conditionally on  $(X_{u_1}(U), M_{u_1}(U))$ , the process  $\Phi_{R_{u_1}(U)}(Q_{u_1, U})$  has the same distribution as  $(\hat{Q}(M_{u_1}(U)t))_{t \geq 0}$ , where  $\hat{Q}$  is an

independent copy of  $Q$ . Since  $X_{u_1}(U)$  is uniform over  $[0, 1]$ , we deduce by induction on the sub-quadtree  $Q_{u_1, U}$  that  $X_u(U)$  is uniform over  $[0, 1]$  and independent of  $(M_{u_2}(U), \dots, M_{u_k}(U))$ , which is a sequence of independent random variables all having density  $2(1 - m)\mathbf{1}_{\{m \in [0, 1]\}}$ . Furthermore, it is easy to see that

$$E[(X_{u_i}(U), M_{u_i}(U))_{2 \leq i \leq k} \mid (X_{u_1}(U), M_{u_1}(U))] = E[(X_{u_i}(U), M_{u_i}(U))_{2 \leq i \leq k} \mid X_{u_1}(U)].$$

Hence, by (6),  $X_u(U)$  is also independent of  $M_{u_1}(U)$ .

Letting  $\hat{m}(t) = E[N_t(U)]$  (recall that, when  $t < 0$ ,  $N_t(x) = 0$  for all  $x \in [0, 1]$ ), (2) becomes

$$\hat{m}(t) = P(t \geq \tau_1) + 2E[\hat{m}(M(t - \tau_1))], \tag{7}$$

where  $M$  is independent of  $\tau_1$  and has density  $2(1 - m)\mathbf{1}_{\{m \in [0, 1]\}}$ .

**Proposition 2.** *Let  $U$  be uniform over  $[0, 1]$  and independent of  $(Q(t))_{t \geq 0}$ . We have the following convergence:*

$$\lim_{t \rightarrow \infty} t^{-\beta^*} E[N_t(U)] = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)}, \quad \text{where } \beta^* = \frac{\sqrt{17} - 3}{2}.$$

*Proof.* We consider an auxiliary fragmentation process  $(\mathfrak{F}_t)_{t \geq 0}$  with index of self-similarity 1 and dislocation probability measure  $\nu$  given by

$$\int \nu(ds_1, ds_2) f(s_1, s_2) = E[f(M_1(U) \vee M_0(U), M_1(U) \wedge M_0(U))].$$

In other words, the dislocation measure is given by the law of the decreasing ordering of  $\{M_0(U), M_1(U)\}$ . More precisely  $(\mathfrak{F}_t)_{t \geq 0}$  takes its values in  $\mathcal{S}^\downarrow$  and satisfies the following equation in distribution which completely characterizes its law:

$$(\mathfrak{F}_t) \stackrel{D}{=} ((\mathbf{1}_{\{t < \tau\}}) \dot{+} (\mathbf{1}_{\{t \geq \tau\}} M_0(U) \mathfrak{F}_{M_0(U)(t-\tau)}^{(0)})_{t \geq 0} \dot{+} (\mathbf{1}_{\{t \geq \tau\}} M_1(U) \mathfrak{F}_{M_1(U)(t-\tau)}^{(1)})_{t \geq 0})^\downarrow.$$

Here  $(\mathfrak{F}_t^{(0)})_{t \geq 0}$  and  $(\mathfrak{F}_t^{(1)})_{t \geq 0}$  are two independent copies of  $(\mathfrak{F}_t)_{t \geq 0}$  also independent of  $(M_0(U), M_1(U), \tau)$ , and  $\tau$  an independent exponential variable with parameter 1. The symbol ‘ $\dot{+}$ ’ means concatenation of sequences and  $(\cdot)^\downarrow$  is the decreasing reordering (and erasing of 0s). Then, it is straightforward to see that the expectation of the number  $\#\mathfrak{F}_t$  of fragments of  $\mathfrak{F}_t$  minus 1 satisfies the same equation as  $E[N_t(U)]$ , namely letting  $m(t) = E[\#\mathfrak{F}_t - 1]$  for  $t \geq 0$  and  $m(t) = 0$  for  $t < 0$ , we have

$$m(t) = P(t \geq \tau_1) + 2E[m(M(t - \tau_1))], \tag{8}$$

where  $M$  is independent of  $\tau_1$  and has density  $2(1 - m)\mathbf{1}_{\{m \in [0, 1]\}}$ . By (7) and (8), the functions  $\hat{m}$  and  $m$  satisfy the same integral equation:

$$f(t) = 1 - e^{-t} + 2 \int_0^1 dm 2(1 - m) \int_0^t ds e^{-s} f(m(t - s)).$$

Differentiating with respect to  $t$ , we see that both  $\hat{m}$  and  $m$  are solutions of the Cauchy problem for the integro-differential equation

$$\partial_t f(t) = 1 - f(t) + \int_0^1 dm 2(1 - m) f(mt), \quad f(0) = 0.$$

The uniqueness of the solution of this kind of integro-differential equation is known; see, e.g. [9]. We deduce that, for every  $t \geq 0$ ,  $m(t) = \hat{m}(t)$ . We now focus on  $m(t)$ . Following [2, Section 3], we let, for every  $\beta > 0$ ,  $\psi(\beta) = 1 - \int \nu(ds_1, ds_2)(s_1^\beta + s_2^\beta)$ . An easy calculation yields

$$\psi(\beta) = \frac{\beta^2 + 3\beta - 2}{(\beta + 1)(\beta + 2)}.$$

In particular, the Malthusian exponent associated to  $\nu$ , which is characterized by  $\psi(\beta) = 0$  (see [1, Section 1.2.2]) is

$$\beta^* = \frac{\sqrt{17} - 3}{2}.$$

Applying [2, Theorem 1], we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} t^{-\beta^*} E[\#\mathfrak{F}_t] &= \frac{\Gamma(1 - \beta^*)}{\beta^*} \frac{4}{2\beta^* + 3} \\ &\times \prod_{k=1}^{\infty} \left(1 - \frac{\beta^*}{k}\right) \left(1 - \frac{\beta^*}{k + \sqrt{17}}\right) \left(1 + \frac{\beta^*}{k + 1}\right) \left(1 + \frac{\beta^*}{k + 2}\right). \end{aligned}$$

Finally, we use the Weierstrass identity for the gamma function: for every complex number  $z \in \mathbb{C} \setminus \mathbb{Z}_-$ ,

$$\Gamma(z + 1) = e^{-\gamma z} \prod_{k=1}^{\infty} \left(1 + \frac{z}{k}\right)^{-1} e^{z/k},$$

where  $\gamma$  is the Euler–Mascheroni constant. We conclude that

$$\begin{aligned} \lim_{t \rightarrow \infty} t^{-\beta^*} E[N_t(U)] &= \frac{4}{\beta^*(2\beta^* + 3)} \frac{\Gamma(\sqrt{17} + 1)}{\Gamma(\sqrt{17} - \beta^* + 1)} \frac{1}{\Gamma^2(\beta^* + 2)} \frac{1}{1 + \beta^*/2} \\ &= \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)}, \end{aligned}$$

which completes the proof.

**Remark 1.** We can derive the following equality in distribution from (1):

$$N_t(U) \stackrel{D}{=} \mathbf{1}_{\{\tau_1 \leq t\}} + N_{M_0(U)(t-\tau_1)}^{(0)}(X_0(U)) + N_{M_1(U)(t-\tau_1)}^{(1)}(X_1(U)).$$

Here  $(N_t^{(0)})_{t \geq 0}$  and  $(N_t^{(1)})_{t \geq 0}$  are independent copies of the process  $(N_t)_{t \geq 0}$ . We have already noted that  $X_0(U)$  and  $X_1(U)$  are also uniform and independent of  $(N_t^{(0)})_{t \geq 0}$ , of  $(N_t^{(1)})_{t \geq 0}$ , and of  $(M_0(U), M_1(U))$ . If  $X_0(U)$  and  $X_1(U)$  were independent then  $N_t(U)$  would satisfy the same distributional equation as  $(\#\mathfrak{F}_t - 1)_{t \geq 0}$ . However, this is not the case since we have  $X_0(U) = X_1(U)$ . This explains why we had to work with expectations.

**Corollary 1.** ([3], [7].) *We have*

$$\lim_{n \rightarrow \infty} n^{-\beta^*} E[\mathcal{N}_n(U)] = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)}.$$

*Proof.* This is a straightforward application of Lemma 1 and Proposition 2.

**Remark 2.** Observe that Chern and Hwang [3] obtained a more precise asymptotic behavior of  $E[\mathcal{N}_n(U)]$ . They proved that

$$E[\mathcal{N}_n(U)] = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)} n^{\beta^*} + O(1).$$

**3.2. The  $x = 0$  case**

As a further example of the connection with fragmentation theory, we derive asymptotics properties for  $N_t(0)$ . In this case, the sequence of the areas of the rectangles crossed by  $S_0$  is a fragmentation process, enabling us to state a convergence of  $N_t(0)$ , once rescaled, in  $\mathbb{L}^2$ . A convergence in mean has already been obtained in [7, Theorem 6] and [8].

**Theorem 2.** *The random variable*

$$\mathfrak{M}_t = \sum_{u \in \mathcal{A}} \text{Leb}(R_u(0))^{\sqrt{2}-1} \mathbf{1}_{\{R_u(0) \in Q(t)\}}, \quad t \geq 0,$$

is a uniformly integrable martingale which converges a.s. to  $\mathfrak{M}_\infty$  as  $t \rightarrow \infty$ . The distribution of  $\mathfrak{M}_\infty$  is characterized by

$$E[\mathfrak{M}_\infty] = 1 \quad \text{and} \quad \mathfrak{M}_\infty \stackrel{D}{=} M_0(0)^{\sqrt{2}-1} \mathfrak{M}_\infty^{(0)} + M_1(0)^{\sqrt{2}-1} \mathfrak{M}_\infty^{(1)}, \quad (9)$$

where  $\mathfrak{M}_\infty^{(0)}$  and  $\mathfrak{M}_\infty^{(1)}$  are two independent copies of  $\mathfrak{M}_\infty$  also independent of  $(M_0(0), M_1(0))$ . Furthermore, we have the following convergence in  $\mathbb{L}^2$ :

$$t^{1-\sqrt{2}} N_t(0) \rightarrow \frac{\Gamma(2\sqrt{2})}{\sqrt{2}\Gamma^3(\sqrt{2})} \mathfrak{M}_\infty \quad \text{as } t \rightarrow \infty.$$

*Proof.* It is easy to check from properties of Poisson measures that the rearrangement in decreasing order of the masses of the rectangles living at time  $t$  and intersecting  $S_0$ ,

$$(\text{Leb}(R_u(0)) \mathbf{1}_{\{R_u(0) \in Q(t)\}})_{t \geq 0}^\downarrow,$$

is a self-similar fragmentation with index 1 and dislocation probability measure given by the decreasing ordering of  $\{M_0(0), M_1(0)\}$ . As in the proof of Proposition 2, we introduce, for every  $\beta > 0$ ,  $\Psi(\beta) = 1 - E[M_0(0)^\beta + M_1(0)^\beta]$ , which is easily computed:

$$\Psi(\beta) = \frac{(\beta + 1)^2 - 2}{(\beta + 1)^2}.$$

Thus, the Malthusian exponent  $p^*$  of this fragmentation satisfying  $\Psi(p^*) = 0$  is

$$p^* = \sqrt{2} - 1.$$

The first two points of the theorem follow from classical results of fragmentation theory; see [1, Theorem 1.1]. We refer the reader to [10] for the characterization of the law of  $\mathfrak{M}_\infty$  via the distributional equation (9) and to [11] for some of its properties. The last point comes from [2, Corollary 6] and the Weierstrass identity for the gamma function used in a similar manner as in the proof of Proposition 2.



**Corollary 2.** *We have the following convergence in  $\mathbb{L}^2$ :*

$$n^{1-\sqrt{2}}N_{\tau_n}(0) \rightarrow \frac{\Gamma(2\sqrt{2})}{\sqrt{2}\Gamma^3(\sqrt{2})}\mathfrak{M}_\infty \quad \text{as } n \rightarrow \infty.$$

*Proof.* This proposition is easily derived from Lemma 1 and Theorem 2.

**Remark 3.** Observe that Corollary 2 implies the following convergence in distribution:

$$n^{1-\sqrt{2}}\mathcal{N}_n(0) \rightarrow \frac{\Gamma(2\sqrt{2})}{\sqrt{2}\Gamma^3(\sqrt{2})}\mathfrak{M}_\infty \quad \text{as } n \rightarrow \infty.$$

**Remark 4.** It is worthwhile to note that the behavior of the cost of the partial match query in the  $x = 0$  case is drastically different from its behavior in the case when  $x$  is uniform or  $x$  is fixed in  $(0, 1)$  (see Theorem 1 and Proposition 2).

**3.3. An a priori uniform bound**

This subsection is devoted to the proof of an a priori uniform bound on  $s^{-\beta^*}E[N_s(x)]$  over  $(x, s) \in (0, 1) \times (0, \infty)$  that will be useful in many places.

**Lemma 2.** *There exists a  $C < \infty$  such that*

$$\sup_{x \in (0,1)} \sup_{s>0} E[s^{-\beta^*}N_s(x)] \leq C. \tag{10}$$

*Proof.* As a first step, we start by proving that there exists a  $C_1 < \infty$  such that, for every  $x \in (0, 1)$ ,

$$\sup_{s>0} E[s^{-\beta^*}N_s(x)] \leq \frac{C_1}{x \wedge (1-x)}. \tag{11}$$

Combining (2) with the densities computed in (4), we deduce that, for every  $x \in (0, 1)$ ,

$$\begin{aligned} t^{-\beta^*}E[N_t(x)] &= t^{-\beta^*}P(t \geq \tau_1) + 2\left(\int_x^1 \frac{dy}{y} \int_0^{x/y} dm E[t^{-\beta^*}N_{m(t-\tau_1)}(y)] \right. \\ &\quad \left. + \int_0^x \frac{dy}{1-y} \int_0^{(1-x)/(1-y)} dm E[t^{-\beta^*}N_{m(t-\tau_1)}(y)]\right). \end{aligned} \tag{12}$$

By the monotonicity of  $t \mapsto N_t(x)$  we have  $E[t^{-\beta^*}N_{m(t-\tau_1)}(y)] \leq E[t^{-\beta^*}N_t(y)]$ . Furthermore, recalling that  $\beta^* < 1$ , there exists a constant  $C'$  such that, for every  $t > 0$ ,  $t^{-\beta^*}P(t \geq \tau_1) \leq C'$ . Hence,

$$\begin{aligned} t^{-\beta^*}E[N_t(x)] &\leq C' + 2\left(\int_x^1 \frac{x dy}{y^2} E[t^{-\beta^*}N_t(y)] + \int_0^x \frac{(1-x) dy}{(1-y)^2} E[t^{-\beta^*}N_t(y)]\right) \\ &\leq C' + \frac{2}{x \wedge (1-x)} \int_0^1 dy E[t^{-\beta^*}N_t(y)] \\ &= C' + \frac{2}{x \wedge (1-x)} E[t^{-\beta^*}N_t(U)]. \end{aligned}$$

It has been shown in Proposition 2 that  $E[t^{-\beta^*}N_t(U)]$  has a finite limit as  $t \rightarrow \infty$ , and, for every  $t > 0$ ,  $E[N_t(U)] \leq t$ . Thus, the quantity  $E[t^{-\beta^*}N_t(U)]$  is bounded over  $(0, \infty)$ . Inequality (11) follows from these considerations.

Introducing  $S(x) = \sup_{s>0} s^{-\beta^*} \mathbb{E}[N_s(x)]$  for every  $x \in [0, 1]$ , we have just shown that  $S(x) \leq C_1(x \wedge (1-x))^{-1}$ . Using (12), we have, for every  $x \in (\frac{1}{2}, 1)$ ,

$$\begin{aligned}
 S(x) &= \sup_{t>0} \left\{ t^{-\beta^*} \mathbb{P}(t \geq \tau_1) + 2 \left( \int_x^1 \frac{dy}{y} \int_0^{x/y} dm \mathbb{E}[t^{-\beta^*} N_{m(t-\tau_1)}(y)] \right. \right. \\
 &\quad \left. \left. + \int_0^x \frac{dy}{1-y} \int_0^{(1-x)/(1-y)} dm \mathbb{E}[t^{-\beta^*} N_{m(t-\tau_1)}(y)] \right) \right\} \\
 &\leq C' + 2 \sup_{t>0} \left\{ \int_x^1 \frac{dy}{y} \int_0^1 dm \mathbb{E}[t^{-\beta^*} N_t(y)] + \int_0^{1/2} \frac{dy}{1-y} \int_0^1 dm \mathbb{E}[t^{-\beta^*} N_t(y)] \right\} \\
 &\quad + 2 \sup_{t>0} \int_{1/2}^x \frac{dy}{1-y} \int_0^{(1-x)/(1-y)} dmm^{\beta^*} \mathbb{E}[(mt)^{-\beta^*} N_{mt}(y)] \\
 &\leq C' + 8 \sup_{t>0} \int_0^1 dy \mathbb{E}[t^{-\beta^*} N_t(y)] + 2 \int_{1/2}^x \frac{dy}{1-y} \int_0^{(1-x)/(1-y)} dmm^{\beta^*} S(y) \\
 &\leq C_2 + \frac{2}{\beta^* + 1} (1-x)^{\beta^*+1} \int_{1/2}^x dy \frac{1}{(1-y)^{\beta^*+2}} S(y). \tag{13}
 \end{aligned}$$

Let us show that this implies that, for every  $x \in (0, 1)$ ,  $S(x) \leq 100C_2$ . Arguing by contradiction, suppose that there exists an  $a \in (\frac{1}{2}, 1)$  such that  $S(a) > 100C_2$ . Let  $S = \sup_{x \in [1/2, a]} S(x)$ . By (11),  $S$  is finite; there exists a  $b \in [\frac{1}{2}, a]$  such that  $S(b) \geq 0.9S$ . In particular,  $S(b) \geq 0.9 \sup_{x \in [1/2, b]} S(x)$  and  $S(b) > 90C_2$ . Applying (13) at  $b$ , we obtain

$$\begin{aligned}
 S(b) &\leq 90^{-1} S(b) + \frac{2}{\beta^* + 1} (1-b)^{\beta^*+1} \int_{1/2}^b dy \frac{1}{(1-y)^{\beta^*+2}} 0.9^{-1} S(b) \\
 &\leq 90^{-1} S(b) + \frac{2 \cdot 0.9^{-1}}{(\beta^* + 1)^2} S(b),
 \end{aligned}$$

leading to a contradiction since  $(\beta^* + 1)^2 > 2 \cdot 0.9^{-1} / (1 - 90^{-1})$ . Finally,  $S(x) \leq 100C_2$  for every  $x \in (0, 1)$ .

#### 4. The convergence at fixed $x \in (0, 1)$

We prove in this section that, when  $x \in [0, 1]$  is fixed,  $t^{-\beta^*} \mathbb{E}[N_t(x)]$  admits a finite limit as  $t \rightarrow \infty$ . The results of the preceding section do not directly apply since the place  $X_0(x)$  of  $x$  in the rectangle  $R_0(x)$  highly depends on  $x$ . Recall the notation  $\mathfrak{z}_k$  for the word composed of  $k$  0s,  $0 \cdots 0 \in \mathcal{A}$ . The guiding idea is that the splittings tend to make  $X_{\mathfrak{z}_k}(x)$  uniform and independent of  $M_{\mathfrak{z}_k}(x)$ .

##### 4.1. A key Markov chain

Fix  $x \in (0, 1)$ . To simplify notation, for every  $k \geq 1$ , we write  $X_k$  for  $X_{\mathfrak{z}_k}(x)$  and  $M_k$  for  $M_{\mathfrak{z}_k}(x)$ . We will focus on the process  $(X_k, M_k)_{k \geq 0}$ , which is obviously a homogeneous Markov chain starting from  $(x, 1)$  whose transition probability is given by (4) or (5). Let  $k \geq 1$ . We denote by  $\mathcal{F}_k$  the filtration generated by  $(X_i, M_i)_{1 \leq i \leq k}$ . It is easy to see that the transition probability depends only on  $X_k$ , that is,

$$\mathbb{E}[(X_{k+i}, M_{k+i})_{i \geq 1} \mid \mathcal{F}_k] = \mathbb{E}[(X_{k+i}, M_{k+i})_{i \geq 1} \mid X_k].$$

**Proposition 3.** Fix  $x \in (0, 1)$ . There exists a coupling of the chain  $(X_k, M_k)_{k \geq 0}$  with a random time  $T \in \mathbb{Z}_+$  such that, for any  $k \geq 0$ , conditionally on  $\{T \leq k\}$ , the random variable  $X_k$  is uniformly distributed over  $[0, 1]$ , independent of  $(M_i)_{1 \leq i \leq k}$  and of  $T$ . Furthermore, we have

$$E[1.15^T] < +\infty.$$

*Proof.* For any  $k \geq 1$ , we consider the event

$$E_k = \{M_k < X_{k-1} \wedge (1 - X_{k-1})\}.$$

Using the explicit densities (4) and (5), we see that, conditionally on  $\mathcal{F}_{k-1}$  and on the event  $E_k$  of probability  $-(X_{k-1} \wedge (1 - X_{k-1})) \ln(X_{k-1}(1 - X_{k-1}))$ , the conditional distribution of  $X_k$  is

$$\frac{1}{-\ln(X_{k-1}(1 - X_{k-1}))} \left( \frac{1}{1 - y} \mathbf{1}_{\{y \in (0, X_{k-1})\}} + \frac{1}{y} \mathbf{1}_{\{y \in (X_{k-1}, 1)\}} \right) dy.$$

In particular, conditionally on  $E_k$  and  $\mathcal{F}_{k-1}$ , the variable  $X_k$  is independent of  $M_k$  and has a density bounded below by  $-1/\ln(X_{k-1}(1 - X_{k-1}))$ . Thus, we can construct simultaneously with  $(X_k, M_k)_{k \geq 0}$  a sequence of random variables  $(B_k)_{k \geq 0} \in \{0, 1\}^{\mathbb{Z}_+}$  as follows. Suppose that we have constructed  $(X_i, M_i, B_i)_{0 \leq i \leq k-1}$ . Then, independently of  $\mathcal{F}_{k-1}$ , toss a Bernoulli variable of parameter  $-(X_{k-1} \wedge (1 - X_{k-1})) \ln(X_{k-1}(1 - X_{k-1}))$ . If 0 comes out, we consider that we are on the event  $E_k^c$ , then put  $B_k = 0$  and sample  $(X_k, M_k)$  with the conditional distribution on  $E_k^c$  and  $\mathcal{F}_{k-1}$ . If 1 comes out, we consider that we are on the event  $E_k$  and we proceed to the following.

1. First sample  $M_k$  from its distribution conditionally on  $E_k$  and  $\mathcal{F}_{k-1}$
2. Then, independently of  $M_k$ , toss a Bernoulli variable  $B_k$  of parameter  $-1/\ln(X_{k-1}(1 - X_{k-1}))$ . If  $B_k = 1$ , sample  $X_k$  uniformly from  $[0, 1]$  and independently of  $(M_1, \dots, M_k)$ . Otherwise, sample  $X_k$  with density

$$\frac{1}{-\ln(X_{k-1}(1 - X_{k-1})) - 1} \left( \left( \frac{1}{1 - y} - 1 \right) \mathbf{1}_{\{y \in (0, X_{k-1})\}} + \left( \frac{1}{y} - 1 \right) \mathbf{1}_{\{y \in (X_{k-1}, 1)\}} \right) dy,$$

independently of  $(M_1, \dots, M_k)$ .

The device provides us with a Markov chain  $(X_k, M_k, B_k)_{k \geq 0}$  such that the first two coordinates have the law of the process introduced before Proposition 3. We then let

$$T = \inf\{k \geq 0, B_k = 1\}.$$

By the definition of  $T$ , the random variable  $X_T$  is sampled uniformly over  $[0, 1]$  and independently of  $(M_1, \dots, M_T)$ . We deduce that the process  $(X_{T+i}, M_{T+i})_{i \geq 1}$  has the same distribution as the process  $(X_{3k}, M_{3k}(U), M_{3k}(U))_{k \geq 1}$  defined in Proposition 1; hence, an easy adaptation of Proposition 1 shows that, for every positive integer  $i$ ,  $X_{T+i}$  is uniformly distributed over  $[0, 1]$  independent of  $(M_1, \dots, M_{T+i})$  and of  $T$ . This proves the first part of Proposition 3.

For the second part, we need to evaluate the tail of the random time  $T$ . We introduce the following variation. Let  $(\hat{X}_k)_{k \geq 0}$  be a Markov chain with space state  $(0, 1) \cup \{\partial\}$ , where  $\partial$  is a cemetery point. Informally, this chain is the chain  $(X_k)$  until we reach the time  $T$ , then it is killed and sent to the cemetery point. Thanks to the calculation presented at the beginning of the proof, it should be clear that, given  $X_{k-1}$  and conditionally on  $\{T \geq k - 1\}$ , the probability

of the event  $\{T = k\}$  is  $X_{k-1} \wedge (1 - X_{k-1})$ . Thus, the transition probability for the chain  $(\hat{X}_k)$  is defined as follows: for every  $x \in (0, 1)$ ,

$$p(x, dy) = x \wedge (1 - x)\delta_\partial + \left( \frac{1 - x}{(1 - y)^2} \mathbf{1}_{\{y \in (0, x)\}} + \frac{x}{y^2} \mathbf{1}_{\{y \in (x, 1)\}} - x \wedge (1 - x) \right) dy$$

and  $p(\partial, dy) = \delta_\partial$ . By the construction of this chain, the stopping time  $\hat{T} = \inf\{k \geq 1 : \hat{X}_k = \partial\}$  has the same distribution as  $T$ . In order to estimate  $\hat{T}$ , we define the potential function  $V : (0, 1) \cup \{\partial\} \rightarrow [1, \infty)$  as

$$V(x) = \mathbf{1}_{\{x=\partial\}} + \frac{10}{\sqrt{x}} \mathbf{1}_{\{x \in (0, 1/2)\}} + \frac{10}{\sqrt{1-x}} \mathbf{1}_{\{x \in [1/2, 1)\}}.$$

Then we can show that, for every  $x \in (0, 1) \cup \{\partial\}$ ,

$$\int p(x, dy)V(y) \leq 0.85V(x) + \mathbf{1}_{\{\partial\}}(x),$$

so that [12, Theorem 15.2.5] may be applied: there exists a  $\varepsilon > 0$  such that, for all  $x \in (0, 1)$ ,

$$\mathbb{E} \left[ \sum_{k=0}^{\hat{T}-1} V(\hat{X}_k) 1.15^k \right] \leq \varepsilon^{-1} 1.15^{-1} V(x),$$

from which we deduce that

$$\mathbb{E}[1.15^{\hat{T}}] < \infty$$

(note that the last quantity is not uniformly bounded for  $x \in (0, 1)$ ). This completes the proof.

In the remaining part of this section,  $x$  is fixed in  $(0, 1)$ . Returning to (3) and writing  $\bar{M}_k = M_1 M_2 \cdots M_k$  for the Lebesgue measure of  $R_{3k}(x)$ , we have

$$t^{-\beta^*} \mathbb{E}[N_t(x)] = t^{-\beta^*} (g_k(t) + 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T > k\}}] + 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T \leq k\}}]). \tag{14}$$

We will treat the last two terms of (14) separately.

#### 4.2. Study of $t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T > k\}}]$

We will see that  $t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T > k\}}]$  is arbitrarily small uniformly in  $t$  provided that the integer  $k$  is chosen large enough. Observe that

$$\begin{aligned} & t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T > k\}}] \\ & \leq t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t}(X_k) \mathbf{1}_{\{T > k\}}] \\ & = 2^k \mathbb{E}[\bar{M}_k^{\beta^*} (\bar{M}_k t)^{-\beta^*} \tilde{N}_{\bar{M}_k t}(X_k) \mathbf{1}_{\{T > k\}}] \\ & = 2^k \mathbb{E}[\bar{M}_k^{\beta^*} \mathbf{1}_{\{T > k\}} \mathbb{E}[(\bar{M}_k t)^{-\beta^*} \tilde{N}_{\bar{M}_k t}(X_k) \mid \sigma(\bar{M}_k, X_k, T)]]. \end{aligned}$$

Letting  $\phi$  be the map  $(s, u) \mapsto \mathbb{E}[s^{-\beta^*} N_s(u)]$ , we have

$$t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T > k\}}] \leq 2^k \mathbb{E}[\bar{M}_k^{\beta^*} \mathbf{1}_{\{T > k\}} \phi(\bar{M}_k t, X_k)].$$

Thanks to (10),  $\phi \leq C$ , so that the quantity in the last display is at most  $C2^k \mathbb{E}[\bar{M}_k^{\beta^*} \mathbf{1}_{\{T>k\}}]$ . Hölder's inequality yields, for every  $p > 1$ ,

$$C2^k \mathbb{E}[\bar{M}_k^{\beta^*} \mathbf{1}_{\{T>k\}}] \leq C2^k \mathbb{E}[\bar{M}_k^{\beta^* p}]^{1/p} \mathbb{E}[\mathbf{1}_{\{T>k\}}]^{1-1/p}.$$

The last term is easily treated: by Markov's inequality we have  $\mathbb{E}[\mathbf{1}_{\{T>k\}}] \leq 1.15^{-k} \mathbb{E}[1.15^T]$ . Concerning  $\mathbb{E}[\bar{M}_k^{\beta^* p}]$  we have

$$\begin{aligned} \mathbb{E}[\bar{M}_k^{\beta^* p}] &\leq \mathbb{E}[M_{\beta_2}(x)^{\beta^* p} \cdots M_{\beta_k}(x)^{\beta^* p}] \\ &= \int_0^1 f^{(x)}(y) \, dy \mathbb{E}[M_{\beta_1}(y)^{\beta^* p} \cdots M_{\beta_{k-1}}(y)^{\beta^* p}], \end{aligned}$$

where  $f^{(x)}$  is the density of  $X_1$  under  $\mathbb{P}$ . It is easy to see from (4) that  $f^{(x)}$  is bounded from above by  $(x \wedge (1-x))^{-1}$ . Hence,

$$\mathbb{E}[\bar{M}_k^{\beta^* p}] \leq \frac{1}{x \wedge (1-x)} \int_0^1 dy \mathbb{E}[\bar{M}_{k-1}(y)^{\beta^* p}].$$

Recall from Proposition 1 that when  $x = U$  is uniform over  $[0, 1]$  and independent of  $(Q(t))_{t \geq 0}$ , then  $M_{\beta_1}(U), \dots, M_{\beta_k}(U)$  are independent and distributed according to  $\mathbf{1}_{\{m \in [0, 1]\}} 2(1-m) \, dm$ . In particular,

$$\mathbb{E}[M_0(U)^{\beta^* p}] = \frac{2}{(\beta^* p + 1)(\beta^* p + 2)},$$

and, thus,

$$\int_0^1 dy \mathbb{E}[\bar{M}_{k-1}(y)^{\beta^* p}] = \left( \frac{2}{(\beta^* p + 1)(\beta^* p + 2)} \right)^{k-1}.$$

Gathering all these estimates, we obtain

$$\begin{aligned} &t^{-\beta^*} 2^k \mathbb{E}[N_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T>k\}}] \\ &\leq C2^k \left( \frac{1}{x \wedge (1-x)} \right)^{1/p} \left( \frac{2}{(\beta^* p + 1)(\beta^* p + 2)} \right)^{(k-1)/p} \mathbb{E}[1.15^T]^{1-1/p} 1.15^{-k(1-1/p)} \\ &= K_{p,x} \left( 2 \left\{ \frac{2}{(\beta^* p + 1)(\beta^* p + 2)} \right\}^{1/p} 1.15^{1/p-1} \right)^k, \end{aligned}$$

where  $K_{p,x}$  is a constant that depends only on  $p$  and  $x$  but not on  $k$ . Now, we can easily prove that, for  $p > 1$  sufficiently close to 1, the term in the curly brackets in the last display becomes strictly less than 1. Consequently, letting  $\varepsilon > 0$  be fixed, there exists a sufficiently large integer  $k$  such that, for every  $t > 0$ ,

$$t^{-\beta^*} 2^k \mathbb{E}[N_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T>k\}}] \leq \varepsilon. \tag{15}$$

### 4.3. Conclusion

Observe that we have, for every  $t > 0$ ,

$$\begin{aligned} &t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T \leq k\}}] \\ &= 2^k \mathbb{E}[\mathbf{1}_{\{T \leq k\}} \mathbb{E}[t^{-\beta^*} \tilde{N}_{\bar{M}_k t - F_k}(X_k) \mid \sigma(\bar{M}_k, F_k, T)]] \\ &= 2^k \mathbb{E}[\mathbf{1}_{\{T \leq k\}} (\bar{M}_k - t^{-1} F_k)_+^{\beta^*} \mathbb{E}[(\bar{M}_k t - F_k)_+^{-\beta^*} \tilde{N}_{\bar{M}_k t - F_k}(X_k) \mid \sigma(\bar{M}_k, F_k, T)]], \end{aligned}$$

where  $y_+$  denotes  $y \vee 0$ . By Proposition 3, on the event  $\{T \leq k\}$ , the random variable  $X_k$

is uniformly distributed over  $[0, 1]$  and independent of  $M_1, \dots, M_k$  and, thus, of  $\bar{M}_k$ . It is also independent of  $F_k$  and  $T$ . Hence, letting  $\theta$  be the map  $s \mapsto \mathbb{E}[s_+^{-\beta^*} N_s(U)]$ , where  $U$  is a random variable uniformly distributed on  $(0, 1)$  independent of  $N$ , we have

$$t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T \leq k\}}] = 2^k \mathbb{E}[\mathbf{1}_{\{T \leq k\}} (\bar{M}_k - t^{-1} F_k)_+^{\beta^*} \theta(\bar{M}_k t - F_k)].$$

Applying Proposition 2,  $\theta(\bar{M}_k t - F_k)$  a.s. tends to a finite limit as  $t \rightarrow \infty$ . Hence, by dominated convergence,  $t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\bar{M}_k t - F_k}(X_k) \mathbf{1}_{\{T \leq k\}}]$  has a finite limit as  $t \rightarrow \infty$ . We deduce, from this fact, (14), and (15), that

$$\limsup_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[N_t(x)] - \liminf_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[N_t(x)] \leq \varepsilon.$$

Since this inequality holds for every  $\varepsilon > 0$ ,  $t^{-\beta^*} \mathbb{E}[N_t(x)]$  has a finite limit as  $t \rightarrow \infty$ , which we denote by

$$n_\infty(x) = \lim_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[N_t(x)].$$

### 5. Identifying the limit

In this section we show that  $x \mapsto n_\infty(x)$  is proportional to  $x \mapsto (x(1-x))^{\beta^*/2}$  using a fixed-point argument for the integral equation (see also [4, Section 4.1] for a similar application). The normalizing constant will come from the  $\mathbb{L}^1$ -norm of  $x \mapsto (x(1-x))^{\beta^*/2}$  and the constant of Proposition 2.

Combining (2) with the densities computed in (4), we deduce that

$$\begin{aligned} t^{-\beta^*} \mathbb{E}[N_t(x)] &= t^{-\beta^*} \mathbb{P}(t \geq \tau_1) \\ &+ 2 \left( \int_x^1 \frac{dy}{y} \int_0^{x/y} dm m^{\beta^*} \mathbb{E}[(mt)^{-\beta^*} N_{m(t-\tau_1)}(y)] \right. \\ &\quad \left. + \int_0^x \frac{dy}{1-y} \int_0^{(1-x)/(1-y)} dm m^{\beta^*} \mathbb{E}[(mt)^{-\beta^*} N_{m(t-\tau_1)}(y)] \right). \end{aligned}$$

Thanks to Lemma 2, we obtain, by dominated convergence,

$$n_\infty(x) = \frac{2}{\beta^* + 1} \left( x^{\beta^*+1} \int_x^1 dy \frac{1}{y^{\beta^*+2}} n_\infty(y) + (1-x)^{\beta^*+1} \int_0^x dy \frac{1}{(1-y)^{\beta^*+2}} n_\infty(y) \right).$$

In other words, if we define

$$g_x(y) = \frac{2}{\beta^* + 1} \left( x^{\beta^*+1} \frac{1}{y^{\beta^*+2}} \mathbf{1}_{\{x < y < 1\}} + (1-x)^{\beta^*+1} \frac{1}{(1-y)^{\beta^*+2}} \mathbf{1}_{\{0 < y < x\}} \right),$$

we have

$$n_\infty(x) = \int_0^1 dy g_x(y) n_\infty(y).$$

Let  $G$  be the operator that maps a function  $f \in \mathbb{L}^1[0, 1]$  to the function

$$G(f)(x) = \int_0^1 dy g_x(y) f(y).$$

In particular,  $n_\infty$  is a fixed point of  $G$ . It is easy to check that  $x \in (0, 1) \mapsto g_x(\cdot) \in \mathbb{L}^1[0, 1]$  is continuous for the  $\mathbb{L}^1$ -norm. Furthermore, Lemma 2 ensures that  $|n_\infty(x)| \leq C$  for every  $x \in (0, 1)$ . As a consequence,  $x \mapsto n_\infty(x)$  is continuous over  $(0, 1)$ . An easy computation shows that, for every  $y \in (0, 1)$ ,  $\int_0^1 dx g_x(y) = 1$ . Let  $p$  be another fixed point of  $G$  having the same integral as  $n_\infty$ . Then

$$\begin{aligned} \int_0^1 dx |n_\infty(x) - p(x)| &= \int_0^1 dx \left| \int_0^1 dy g_x(y) (n_\infty - p)(y) \right| \\ &\leq \int_0^1 dx \int_0^1 dy g_x(y) |n_\infty(y) - p(y)| \\ &= \int_0^1 dy |n_\infty(y) - p(y)|, \end{aligned}$$

which shows that the inequality is in fact an equality. Hence,  $n_\infty - p$  has an almost everywhere (a.e.) constant sign. As we know that the integral of  $n_\infty - p$  is 0, we deduce that  $n_\infty = p$  a.e. Straightforward calculations prove that  $p_0: x \mapsto (x(1-x))^{\beta^*/2}$  is also a fixed point of  $G$  in the  $\mathbb{L}^1$ -norm, so that

$$n_\infty(x) = \|n_\infty\|_1 \|p_0\|_1^{-1} (x(1-x))^{\beta^*/2} \quad \text{a.e.}$$

Since  $n_\infty$  and  $p_0$  are continuous, we can remove the a.e. statement (observe that  $n_\infty(0) = n_\infty(1) = 0$  by Theorem 2). Plainly,

$$\|p_0\|_1 = \frac{\Gamma^2(\beta^*/2 + 1)}{\Gamma(\beta^* + 2)}.$$

On the other hand, (10) and the dominated convergence theorem ensure that

$$\|n_\infty\|_1 = \lim_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[N_t(U)],$$

which was computed in Proposition 2:

$$\|n_\infty\|_1 = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)}.$$

*Proof of Theorem 1.* To sum up, we have, for every  $x \in [0, 1]$ ,

$$t^{-\beta^*} \mathbb{E}[N_t(x)] \rightarrow \frac{\Gamma(2\beta^* + 2)\Gamma(\beta^* + 2)}{2\Gamma^3(\beta^* + 1)\Gamma^2(\beta^*/2 + 1)} (x(1-x))^{\beta^*/2} \quad \text{as } t \rightarrow \infty.$$

Applying Lemma 1, Theorem 1 is proved.

## 6. Extensions and comments

### 6.1. Various convergences

In this paper we proved only a convergence in mean of  $t^{-\beta^*} N_t(x)$ . We may wonder whether this quantity also converges in distribution, in probability, or even almost surely. A more interesting question is the following: does the process  $((t^{-\beta^*} N_t(x))_{x \in [0, 1]}, t > 0)$  converge in distribution in the Skorokhod sense to a random function  $(\mathcal{C}(x))_{x \in [0, 1]}$  as  $t \rightarrow \infty$ ? Observe that if it does then there exists a random point  $U$  uniformly distributed over  $(0, 1)$  such that  $\mathcal{C}(U) = 0$ , where  $U$  corresponds to the point  $x_1$  of the first atom of  $\Pi$  ( $N_t(x_1)$  is indeed of order  $t^{\sqrt{2}-1}$  by Theorem 2).

**Conjecture 1.** We have the functional limit law  $(t^{-\beta^*} N_t(x))_{x \in [0,1]} \rightarrow (\mathcal{C}(x))_{x \in [0,1]}$  as  $t \rightarrow \infty$  in  $(\mathbb{D}([0, 1]), \|\cdot\|_\infty)$ , where  $\mathcal{C}$  satisfies the distributional fixed-point equation

$$\begin{aligned}
 (\mathcal{C}(x))_{x \in [0,1]} \stackrel{D}{=} & \left( \mathbf{1}_{\{x < U_0\}} \left\{ (U_0 U_1)^{\beta^*} \mathcal{C}^{(00)} \left( \frac{x}{U_0} \right) + (U_0(1 - U_1))^{\beta^*} \mathcal{C}^{(01)} \left( \frac{x}{U_0} \right) \right\} \right. \\
 & + \mathbf{1}_{\{x > U_0\}} \left\{ ((1 - U_0)U_1)^{\beta^*} \mathcal{C}^{(10)} \left( \frac{x - U_0}{1 - U_0} \right) \right. \\
 & \left. \left. + ((1 - U_0)(1 - U_1))^{\beta^*} \mathcal{C}^{(11)} \left( \frac{x - U_0}{1 - U_0} \right) \right\} \right)_{x \in [0,1]},
 \end{aligned}$$

where  $U_0, U_1, \mathcal{C}^{(00)}, \mathcal{C}^{(01)}, \mathcal{C}^{(10)}$ , and  $\mathcal{C}^{(11)}$  are independent,  $U_0$  and  $U_1$  are uniformly distributed on  $[0, 1]$ , and  $\mathcal{C}^{(00)}, \mathcal{C}^{(01)}, \mathcal{C}^{(10)}$ , and  $\mathcal{C}^{(11)}$  all have the same distribution as  $\mathcal{C}$ .

### 6.2. Multidimensional case

The strategy adopted in Subsection 3.1 may be generalized to higher dimensions. As for the convergence in mean of the number of hyper-rectangles crossed by a fixed affine subspace having a direction generated by some vectors of the canonical basis, our approach may also be followed.

### 6.3. Quadtree as a model of random geometry

On top of its numerous applications in theoretical computer science, the model of a random quadtree may be considered as a model of random geometry. More precisely, we can view, for  $t \geq 0$ , the set of rectangles  $Q(t)$  as a random graph, assigning length 1 to each edge of the rectangles. We denote this graph by  $\tilde{Q}(t)$ . A natural question would be to understand the metric behavior of  $\tilde{Q}(t)$  as  $t \rightarrow \infty$ . The study of the graph distance  $L_t$  in  $\tilde{Q}(t)$  between the upper-left and upper-right corners would be a first step in understanding the global geometry of  $\tilde{Q}_t$ . Observe that Theorem 2 already shows that  $L_t$  is less than the order  $t^{\sqrt{2}-1}$ .

### Acknowledgements

We would like to express our gratitude to Philippe Flajolet who introduced us to the problem of partial match query. We are indebted to Nicolas Broutin and to Ralph Neininger for fruitful discussions. We also deeply thank Jean Bertoin for his careful reading of the first versions of this work.

### References

- [1] BERTOIN, J. (2006). *Random Fragmentations and Coagulation Processes* (Camb. Stud. Adv. Math. **102**). Cambridge University Press.
- [2] BERTOIN, J. AND GNEDIN, A. V. (2004). Asymptotic laws for nonconservative self-similar fragmentations. *Electron. J. Prob.* **9**, 575–593.
- [3] CHERN, H.-H. AND HWANG, H.-K. (2003). Partial match queries in random quadtrees. *SIAM J. Comput.* **32**, 904–915.
- [4] CURIEN, N. AND LE GALL, J.-F. (2011). Random recursive triangulations of the disk via fragmentation theory. To appear in *Ann. Prob.*
- [5] FINKEL, R. A. AND BENTLEY, J. L. (1974). Quad trees a data structure for retrieval on composite keys. *Acta Informatica* **4**, 1–9.
- [6] FLAJOLET, P. AND SEDGEWICK, R. (2009). *Analytic Combinatorics*. Cambridge University Press.
- [7] FLAJOLET, P., GONNET, G., PUECH, C. AND ROBSON, J. M. (1993). Analytic variations on quadtrees. *Algorithmica* **10**, 473–500.
- [8] FLAJOLET, P., LABELLE, G., LAFOREST, L. AND SALVY, B. (1995). Hypergeometrics and the cost structure of quadtrees. *Random Structures Algorithms* **7**, 117–144.



- [9] ISERLES, A. AND LIU, Y. (1997). Integro-differential equations and generalized hypergeometric functions. *J. Math. Anal. Appl.* **208**, 404–424.
- [10] LIU, Q. (1997). Sur une équation fonctionnelle et ses applications: une extension du théorème de Kesten-Stigum concernant des processus de branchement. *Adv. Appl. Prob.* **29**, 353–373.
- [11] LIU, Q. (2001). Asymptotic properties and absolute continuity of laws stable by random weighted mean. *Stoch. Process. Appl.* **95**, 83–107.
- [12] MEYN, S. AND TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press.