The Annals of Applied Probability
2010, Vol. 20, No. 2, 722–752
DOI: 10.1214/09-AAP635

THE LIMITING MOVE-TO-FRONT SEARCH-COST IN LAW OF LARGE NUMBERS ASYMPTOTIC REGIMES

BY JAVIERA BARRERA[1,3] AND JOAQUÍN FONTBONA[2,3]

Universidad Adolfo Ibáñez and Universidad de Chile

We explicitly compute the limiting transient distribution of the search-cost in the move-to-front Markov chain when the number of objects tends to infinity, for general families of deterministic or random request rates. Our techniques are based on a "law of large numbers for random partitions," a scaling limit that allows us to exactly compute limiting expectation of empirical functionals of the request probabilities of objects. In particular, we show that the limiting search-cost can be split at an explicit deterministic threshold into one random variable in equilibrium, and a second one related to the initial ordering of the list. Our results ensure the stability of the limiting search-cost under general perturbations of the request probabilities. We provide the description of the limiting transient behavior in several examples where only the stationary regime is known, and discuss the range of validity of our scaling limit.

1. Introduction. We consider the search-cost process in the move-to-front (MtF) Markov chain. A finite set of objects labeled $1, \dots, n$ is dynamically maintained as a serial list, and objects are requested at random instants with a given probability $p_i^{(n)}$, $i = 1, \dots, n$. Instantaneously after request, an object is moved to the front of the list, while the relative order of the other objects is left unchanged. The search-cost at a given instant is defined as the position in the list of the next requested object.

The exact and limiting behaviors of the move-to-front rule have received much attention in the computer science and discrete probability literature since the 1960s (see Fill [7] and Jelenković [10] for historical references). For a fixed and finite number of objects, the search-cost distribution has been studied by Fill [6, 7], Fill and Holst [8] and Flajolet, Gardy and Thimonier [9] for different deterministic request probability vectors $\mathbf{p}^{(n)} = (p_1^{(n)}, \dots, p_n^{(n)})$. To our knowledge, the limiting search-cost distribution as the number of objects goes to infinity is known only

Received November 2006; revised August 2009.

[1]Supported by Fondecyt Project 1060485.

[2]Supported by Fondecyt Project 1070743.

[3]Supported by BASAL-Conicyt CMM-U de Chile and Millennium Nucleus Information and Randomness ICM P04-069.

AMS 2000 subject classifications. Primary 60B10, 68W40; secondary 68P10.

Key words and phrases. Move-to-front rule, search-cost, law of large numbers, propagation of chaos.

for the stationary regime of the MtF Markov chain. This problem was first studied by Fill [6] for several types of deterministic request probabilities $\mathbf{p}^{(n)}$, and later by Barrera, Huillet and Paroissin [2, 3] and Barrera and Paroissin [1] for random request probabilities $\mathbf{p}^{(n)}$ defined by normalized samples of positive i.i.d. random variables. A different approach was adopted by Jelenković [10], who considered various scaling limits for the stationary search-cost when the optimal static distribution of objects in the list is specified.

In this article we explicitly compute the limiting law of the transient search-cost as the number n of objects tends to infinity, for a large class of deterministic or random request probabilities. This class includes several previously considered cases. To be more precise, let $w_1^{(n)}, \dots, w_n^{(n)} \geq 0$ be deterministic or random real numbers and consider the probability vector $\mathbf{p}^{(n)} = (p_i^{(n)})_{i=1}^n$ defined by

$$p_i^{(n)} = \frac{w_i^{(n)}}{\sum_{j=1}^n w_j^{(n)}}. \tag{1}$$

We call $p_i^{(n)}$ the "popularity" of object i. Let us denote by $\mathcal{P}(\mathbb{R}_+)$ the space of Borel probability measures in $\mathbb{R}_+$. We consider request probabilities $\mathbf{p}^{(n)}$ of this type that exhibit a weak law of large numbers behavior. That is, we assume that as n goes to infinity, the empirical measure

$$\hat{\nu}^{(n)} := \frac{1}{n} \sum_{i=1}^n \delta_{w_i^{(n)}} \in \mathcal{P}(\mathbb{R}_+), \qquad n \in \mathbb{N},$$

converges in distribution [as a random element of the Polish space $\mathcal{P}(\mathbb{R}_+)$] to a deterministic limit $P \in \mathcal{P}(\mathbb{R}_+)$. Moreover, we assume that $\int x \hat{\nu}^{(n)}(dx)$ converges to $\mu := \int x P(dx) \in (0, \infty)$ in distribution. (These conditions hold if, for instance, $w_i^{(n)} = w_i$, $i = 1, 2, \dots$, are i.i.d. random variables with finite mean μ.)

Denote by $\nu^{(n)}$ the empirical measure of the scaled vector $(n\mu)\mathbf{p}^{(n)}$. As we shall see, the computation of the search-cost distribution will involve continuous functionals of $\nu^{(n)}$. On the other hand, we will show that the sequence $\nu^{(n)}$ shares the same law of large numbers behavior and limit of $\hat{\nu}^{(n)}$. This fact is what we call "law of large numbers for random partitions of the interval." With these elements, we will be able to explicitly compute for all $t > 0$ the limiting distribution of the (suitably normalized) transient search-cost, in terms of the limiting probability measure P.

We will show the existence of an explicit deterministic threshold, depending only on P and t, such that the limiting transient and stationary search-costs have the same distribution in the event they fall below it. The limiting transient search-cost restricted to that event will be called "equilibrium part" of the transient search-cost, and its distribution (as well as the limiting stationary one) will depend only on the law P. Alternatively, we will call "out-of-equilibrium part" of the transient search-cost its restriction to the complementary event. The asymptotic behavior of the out-of-equilibrium part will depend on P, but also on the relative order of popularities in the list at $t = 0$. Concretely, three situations will be studied: (1) no information at all about the request probabilities is available at the beginning; (2) objects in the list are known to be initially arranged in decreasing order of popularity; and (3) objects are known to be initially arranged in increasing order of popularity. In all three cases, we find an explicit expression for the limiting law of the out-of-equilibrium part in terms of P.

As a consequence, we obtain an upper bound for the total variation distance between the transient and the stationary limiting search-cost distributions. The distance from equilibrium at time t turns out to behave like $O(\int xe^{-tx} P(dx))$ in each of the three situations we consider. The techniques we introduce also show that the limiting transient search-cost distribution is stable under perturbations of the request probabilities preserving both the limiting law P and the initial relative order in the list.

The rest of this article is organized as follows. In Section 2 we define the MtF Markov chain in continuous time and the associated search-cost process following the lines of Fill and Holst [8]. We also recall some nonasymptotic results about its law. In Section 3 we state our main results, namely Theorems 3.2 and 3.3, which provide the limiting expressions for the Laplace transforms of the two components of the suitably normalized transient search-cost. We deduce from them the limiting search-cost law in terms of P and its Laplace transform. We also present examples, we discuss connections with the Persistent-Caching-Algorithm introduced by Jelenković and Radovanović [11] and we compare our type of law of large numbers asymptotic with the fluid limit considered by Jelenković [10]. Furthermore, we prove stochastic order relations between the search-costs in the three situations (regarding the initial ordering) that are considered. In Section 4, we prove our law of large numbers for random partitions, we discuss its connection with the propagation of chaos property arising in the probabilistic study of mean field models, and we use those ideas to prove Theorems 3.2 and 3.3. In the last section, we discuss the scope of application and the limitations of our techniques.

Let us establish some notation. In the sequel, $\to^d$ means convergence in distribution, and the notation $\Longrightarrow$ stands for weak convergence of probability measures. We denote by δ_x the Dirac mass at some point x. The convention $\frac{0}{0} := 0$ is adopted throughout.

2. Preliminaries and notation.

Consider a list of n objects labeled $\{1, \dots, n\}$ and a permutation π of $\{1, \dots, n\}$. Assume that at time $t = 0$, object i is at position $\pi(i)$ of the list. Then, objects are requested at random instants $t > 0$ after which the list is instantaneously modified, by placing the requested object on its top. This is the MtF rule. It is customary to assume that different objects are requested at random instants given by independent standard Poisson processes in the line.

Let w_i denote the intensity at which object i is requested. The total number of requests up to time t defines a Poisson process, say $\tilde{N}_t$, of rate

$$w = \sum_{j=1}^{n} w_j.$$

By the strong Markov property, the probability that object i is requested at a given arrival time of $\tilde{N}_t$ is

$$p_i := \frac{w_i}{w}.$$

We call this quantity the "popularity" of object i.

We shall in the sequel work with the time-changed process

$$N_t := \tilde{N}_{t/w}$$

and its requests instants. This is the time scale considered, for instance, in [4, 8], and in our case it will simplify the asymptotic analysis by keeping a constant (unitary) total rate of requests. The request rate w_i of object i becomes p_i in the new time scale, but its popularity remains unchanged.

REMARK 2.1. Nevertheless, our statements will rely on hypotheses made on the parameters (w_i) and will be interpreted also in the original time scale (see Remark 3.4 below).

We denote by

$$S^{(n,i)}(t)$$

the position in the list at time t of object i, and by $I_k \in \{1, \dots, n\}$ the kth requested object (in chronological order). Thus, the label of the first object requested in the time interval $[t, +\infty)$ is $I_{N_{t-}+1}$. We are interested in the search-cost of that object. That is, in the random variable defined by

$$S^{(n)}(t) := \sum_{i=1}^{n} S^{(n,i)}(t) \mathbf{1}_{\{I_{N_{t-}+1}=i\}}.$$

Notice that although the processes $S^{(n,i)}(t)$ are left continuous, $S^{(n)}(t)$ is right continuous since the list is modified instantaneously after each request.

We will further need the following notation:

- R_t is the subset of $\{1, \dots, n\}$ consisting of objects that have been requested at least once in the time interval $[0, t[$.
- We decompose the search-cost $S^{(n)}(t)$ into two random variables:

$$S^{(n)}(t) = S^{(n)}_{\mathbf{e}}(t) + S^{(n)}_{\mathbf{o}}(t),$$

where

$$S_{\mathbf{e}}^{(n)}(t) := S^{(n)}(t)\mathbf{1}_{\{I_{N_{t^-}+1} \in R_t\}}$$

and

$$S_{\mathbf{o}}^{(n)}(t) := S^{(n)}(t)\mathbf{1}_{\{I_{N_{t^-}+1} \notin R_t\}}.$$

Thus, $S_{\mathbf{e}}^{(n)}(t)$ is the search-cost of the requested object if it has been requested at least once in $[0, t[$, and it is 0 otherwise. $S_{\mathbf{o}}^{(n)}(t)$ is defined conversely. The subscripts **e** and **o** respectively stand for "equilibrium" and "out of equilibrium." This decomposition and notation are inspired in Fill's work [7], where a coupling was introduced which simultaneously updates the list in stationary regime and an arbitrary second list. In that coupling, each object had the same search-cost in the two lists after its first request.

REMARK 2.2. Notice that an object has been requested before time t if and only if it stands at one of the first $|R_t|$ positions in the list. Therefore, we have

$$\{S_{\mathbf{e}}^{(n)}(t) > 0\} = \{S^{(n)}(t) \le |R_t|\}.$$

The next result will be used in the sequel.

PROPOSITION 2.1. *Let $\pi(i)$ be the position in the list of item i at time 0. For given real parameters $q_1, \dots, q_n \in [0, 1]$ let $(B_1(q_1)\dots, B_n(q_n))$ denote a vector of n independent Bernoulli random variables of such parameters.*

(a) *For all $k, i \in \{1, \dots, n\}$,*

$$\mathbb{P}\{S^{(n,i)}(t) = k, i \in R_t\} = \int_0^t p_i e^{-p_i u}\mathbb{P}\{J_{\mathbf{e}}^{(n)}(u) = k\}\, du,$$

where $J_{\mathbf{e}}^{(n)}(u) =^d \sum_{j=1, j\neq i}^n B_j(1 - e^{-p_j u})$.

(b) *For all $k, i \in \{1, \dots, n\}$,*

$$\mathbb{P}\{S^{(n,i)}(t) = k, i \notin R_t\} = \mathbb{P}\{J_{\mathbf{o}}^{(n)}(t) = k\}e^{-p_i t},$$

where $J_{\mathbf{o}}^{(n)}(t) =^d \sum_{j=1, j\neq i}^n B_j(1 - e^{-p_j t}\mathbf{1}_{\pi(i)<\pi(j)})$ and $\mathbf{1}_{\pi(i)<\pi(j)} = 1$ if object i precedes j in the initial permutation or 0 otherwise.

(c) *For all $k \in \{1, \dots, n\}$,*

$$\mathbb{P}\{S_{\mathbf{e}}^{(n)}(t) = k\} = \sum_{i=1}^n \int_0^t p_i^2 e^{-p_i u}\mathbb{P}\{J_{\mathbf{e}}^{(n)}(u) = k\}\, du.$$

(d) *For all $k \in \{1, \dots, n\}$,*

$$\mathbb{P}\{S_{\mathbf{o}}^{(n)}(t) = k\} = \sum_{i=1}^n p_i\mathbb{P}\{J_{\mathbf{o}}^{(n)}(t) = k\}e^{-p_i t}.$$

PROOF. The proof of relations (a) and (b) can be deduced from Proposition 2.1 in [8]. The basic ideas are to condition in the last instant $u \in]0, t]$ where object i has been requested and to consider the Poisson point process in reversed time starting from t (see Theorem 2.3.1.3. and Corollary 2.3.1.6 in [4] for a complete proof).

Relation (c) [resp. (d)] follows easily from (a) [resp. (b)], thanks to independence of the events $\{S^{(n,i)}(t) = k, i \in R_t\}$ [resp. $\{S^{(n,i)}(t) = k, i \notin R_t\}$] and $\{I_{N_{t^-}+1} = i\}$. □

3. Main statements, examples and consequences. For each n we next consider a random or deterministic vector of nonnegative real numbers

$$\mathbf{w}^{(n)} = \big(w_1^{(n)}, \dots, w_n^{(n)}\big).$$

Then, *conditionally* on $\mathbf{w}^{(n)}$, we define the MtF Markov chain and its search-cost $S^{(n)}(t)$ in the same way as was done in the previous section for deterministic request rates. Recall that the process $S^{(n)}(t)$ refers to the time-scale at which requests arrive at rate 1.

REMARK 3.1. By Proposition 2.1, the law of $S^{(n)}(t)$ conditional on $\mathbf{w}^{(n)}$ depends on that vector only through the popularities

$$\mathbf{p}^{(n)} = \big(p_1^{(n)}, \dots, p_n^{(n)}\big)$$

defined as in (1).

Let us recall the result obtained in [3] for request probability vectors given by normalized samples of positive i.i.d. random variables.

THEOREM 3.1. *Let $(w_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of nonnegative random variables with finite mean μ and Laplace transform $\phi(t)$ and, for each $n \in \mathbb{N}$, take $\mathbf{w}^{(n)} = (w_i)_{i=1}^n$.*

Let $S^{(n)}(\infty)$ be a random variable defined, conditionally on $\mathbf{w}^{(n)}$, as the search-cost associated with the MtF Markov chain in stationary regime. Then, when $n \to \infty$, we have the convergence

$$\frac{S^{(n)}(\infty)}{n} \to^d S^{(\infty)},$$

where $S^{(\infty)}$ is a random variable in $[0, 1]$ with density given by

$$f_{S^{(\infty)}}(x) = -\frac{1}{\mu}\frac{\phi''(\phi^{-1}(1-x))}{\phi'(\phi^{-1}(1-x))}\mathbf{1}_{[0,1-\mathbf{p}_0]},$$

and $\mathbf{p}_0 = \mathbb{P}(w_i = 0)$.

The proof of Theorem 3.1 relied on Laplace integral techniques. Our goal now is to describe the behavior of a suitable normalization of the random variable $S^{(n)}(t)$ when n goes to ∞. Furthermore, we will do this under an assumption naturally generalizing that of [3].

DEFINITION 3.1 (Condition LLN-P). We say that a sequence of (random or deterministic) vectors $\mathbf{w}^{(n)} = (w_1^{(n)}, \dots, w_n^{(n)})_{n\in\mathbb{N}}$ satisfies a law of large numbers with limiting law P (LLN-P for short), if there exist a probability measure $P \in \mathcal{P}(\mathbb{R}_+)$ with finite first moment $\mu \neq 0$ and positive random variables Z_n, such that the empirical measures

$$\hat{\nu}^{(n)} := \frac{1}{n}\sum_{i=1}^{n} \delta_{Z_n w_i^{(n)}}$$

converge in law to P, and their empirical means

$$\frac{1}{n}\sum_{i=1}^{n} Z_n w_i^{(n)}$$

converge in law to μ.

REMARK 3.2.

(i) Condition LLN-P may hold true with random variables Z_n's that are not identically equal to 1 and fail to hold if one takes $Z_n \equiv 1$ [see, e.g., (c) below].

(ii) LLN-P is equivalent to say that the sequence $(\hat{\nu}^{(n)})$ converges in distribution to the deterministic value P, when seen as random variables in the Polish space $\mathcal{P}_1(\mathbb{R}_+)$ of Borel probability measures with finite first moment, endowed with the Wasserstein distance W_1 (see Theorem 4.1).

(iii) In the time scale originally introduced in Section 2, the constant μ may be thought of as the (asymptotic in n) average request rate per object.

Provided that the empirical means converge in law, LLN-P holds in several situations. The following are some examples:

(a) $w_i^{(n)} = w_i$ for all $n \in \mathbb{N}$, with $(w_i)_{i\in\mathbb{N}}$ an ergodic process with invariant measure P, and $Z_n = 1$.

(b) $(w_1^{(n)}, \dots, w_n^{(n)})$, $n \in \mathbb{N}$, is an exchangeable and P-chaotic vector and $Z_n = 1$. Recall that a random vector $v = (v_i)_{i=1}^n$ in $\mathbb{R}^n$ is said to be *exchangeable* if the law of $(v_{\sigma(i)})_{i=1}^n$ is the same for any n-permutation σ. The notion of "P-chaotic vector" is recalled in Section 4.

(c) $w_i^{(n)} = i^\alpha$ for all $n \in \mathbb{N}$, $\alpha \in \mathbb{R}$ and $Z_n = n^{-\alpha}$. Indeed, for any $\varphi : \mathbb{R} \to \mathbb{R}$ continuous and bounded, one has $\frac{1}{n}\sum_{i=1}^n \varphi(\frac{i^\alpha}{n^\alpha}) \to \int_0^1 \varphi(x^\alpha)\, dx$, using the continuity of $x \mapsto x^\alpha$ in $(0, 1]$ for any α. By the obvious change of variable we get

$$P(dx) = \begin{cases} \dfrac{1}{\alpha} x^{1/\alpha - 1} \mathbf{1}_{[0,1]}(x)\,dx, & \text{if } \alpha > 0, \\ \delta_1(dx), & \text{if } \alpha = 0, \\ \dfrac{1}{|\alpha|} x^{1/\alpha - 1} \mathbf{1}_{[1,\infty)}(x)\,dx, & \text{if } \alpha < 0. \end{cases} \tag{2}$$

Thus, one can check that LLN-P holds for these $w_i^{(n)}$ if and only if $\alpha > -1$.

(d) Let q be a continuous probability density with compact support in $[0, c]$, and for each $n \in \mathbb{N}$ define

$$w_i^{(n)} := Q(ci/n) - Q(c(i-1)/n), \qquad i = 1, \dots, n,$$

where $Q(x) = \int_0^x q(y)\,dy$ is the primitive of q. Then, setting $Z_n := n/c$, for certain $x_i^{(n)} \in ((i-1)/n, i/n]$, $i = 1, \dots, n$, we have

$$\frac{1}{n} \sum_{i=1}^n \varphi(Z_n w_i^{(n)}) = \frac{1}{n} \sum_{i=1}^n \varphi(q(cx_i^{(n)})).$$

Hence, for any continuous and bounded function $\varphi : \mathbb{R} \to \mathbb{R}$, we get

$$\frac{1}{n} \sum_{i=1}^n \varphi(Z_n w_i^{(n)}) \to \frac{1}{c} \int_0^c \varphi(q(x))\,dx,$$

and so LLN-P holds in this case with $P(dx) = (\frac{1}{c} \mathbf{1}_{x \in [0,c]}\,dx) \circ q^{-1}$, the push-forward of the normalized Lebesgue's measure by q. (The convergence of empirical means is in this case trivial.)

Case (c) above was considered by Fill [6]. Case (d) is a particular instance of the "light tail" condition studied in Jelenković [10]. See the examples below for more details.

Before stating our main results, notice that from Proposition 2.1 the law of $S_{\mathbf{e}}^{(n)}(t)$ does not depend on the initial permutation π of the list, whereas that of $S_{\mathbf{o}}^{(n)}(t)$ does. This is simply due to the fact that the cost of searching an already requested object does not depend any more on its initial position. In turn, the value of the initial permutation π remains present in the law of $S_{\mathbf{o}}^{(n)}(t)$. By this reason, we need to study separately the two components of the transient search-cost. Let us state our main theoretical result on the equilibrium part.

THEOREM 3.2. *For $\lambda \geq 0$ and $t \geq 0$, define*

$$\begin{aligned} A_n(t, \lambda) &:= \mathbb{E}\big(\exp\{-\lambda S_{\mathbf{e}}^{(n)}(t)\} \mathbf{1}_{\{S_{\mathbf{e}}^{(n)}(t) > 0\}}\big) \\ &= \mathbb{E}\big(\exp\{-\lambda S^{(n)}(t)\} \mathbf{1}_{\{S^{(n)}(t) \leq |R_t|\}}\big). \end{aligned}$$

Then, if LLN-P holds, we have

$$\lim_{n\to\infty} A_n(n\mu t, \lambda/n) = \frac{1}{\mu}\int_0^t \int_{\mathbb{R}^+} x^2 e^{-xu} P(dx) \exp\{-\lambda(1-\phi(u))\}\, du,$$

where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is the Laplace transform of P.

REMARK 3.3. The Laplace transform of the limiting stationary search-cost obtained in [3] corresponds to the limit of the latter expression when $t \to \infty$.

By the above exposed reasons, some asymptotic assumptions on the initial ordering π of the list will be needed in order to observe a coherent limiting behavior of the out-of-equilibrium part of the transient search-cost $S_{\mathbf{o}}^{(n)}(t)$. Notice that any relevant property of π can be restated in terms of the vector of popularities $\mathbf{p}^{(n)}$, and one can therefore assume without loss of generality that π is equal to the identity permutation Id. We shall explicitly analyze three particular assumptions on $\mathbf{w}^{(n)}$ or (equivalently) on $\mathbf{p}^{(n)}$:

LLN-P-ex: LLN-P holds, $\pi = Id$ and $\mathbf{w}^{(n)}$ is exchangeable for each $n \in \mathbb{N}$.

LLN-P^-: LLN-P holds, $\pi = Id$ and $\mathbf{w}^{(n)}$ is decreasing a.s. for each $n \in \mathbb{N}$.

LLN-P^+: LLN-P holds, $\pi = Id$ and $\mathbf{w}^{(n)}$ is increasing a.s. for each $n \in \mathbb{N}$.

Clearly, the assumption $\pi = Id$ is superfluous under LLN-P-ex, but we shall adopt it for notational convenience. The asymptotic behavior of the out-of-equilibrium part of the transient search-cost is stated in the following:

THEOREM 3.3. *For $\lambda \geq 0$ and $t \geq 0$, define*

$$\begin{aligned} B_n(t,\lambda) &:= \mathbb{E}\big(\exp\{-\lambda S_{\mathbf{o}}^{(n)}(t)\}\mathbf{1}_{\{S_{\mathbf{o}}^{(n)}(t)>0\}}\big) \\ &= \mathbb{E}\big(\exp\{-\lambda S^{(n)}(t)\}\mathbf{1}_{\{S^{(n)}(t)>|R_t|\}}\big) \end{aligned}$$

and define $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ as before. Then, $L(\mu, t, \lambda) := \lim_{n\to\infty} B_n(n\mu t, \lambda/n)$ exists in the following cases:

(i) *if LLN-P-ex holds, and then*

$$\begin{aligned} L(\mu,t,\lambda) &= \frac{|\phi'(t)|}{\mu}\left(\frac{e^{-\lambda(1-\phi(t))} - e^{-\lambda}}{\lambda\phi(t)}\right) \\ &= \frac{|\phi'(t)|}{\mu}\int_0^1 e^{-\lambda\phi(t)x}\, dx \exp\{-\lambda(1-\phi(t))\}; \end{aligned}$$

(ii) *if LLN-P^- holds, and then*

$$L(\mu,t,\lambda) = \frac{1}{\mu}\int_0^\infty x e^{-xt} \exp\left\{-\lambda\int_{x^+}^\infty e^{-yt} P(dy)\right\} P(dx) \exp\{-\lambda(1-\phi(t))\};$$

(iii) *if LLN-P^+ holds, and then*

$$L(\mu,t,\lambda)=\frac{1}{\mu}\int_0^\infty xe^{-xt}\exp\left\{-\lambda\int_0^x e^{-yt}P(dy)\right\}P(dx)\exp\{-\lambda(1-\phi(t))\}.$$

The proofs of Theorems 3.2 and 3.3 are deferred to the next section. They will rely on what we call a law of large numbers for random partitions of the interval. Let us now deduce the law of the limiting transient search-cost under the previous sets of hypotheses.

COROLLARY 3.1. *If LLN-P-ex holds, for each $t>0$ we have*

$$\frac{S^{(n)}(n\mu t)}{n}\to^d S(t),$$

where $S(t)$ satisfies the relation in distribution

$$S(t)=^{(d)} S^{(\infty)}\mathbf{1}_{\{S^{(\infty)}\le 1-\phi(t)\}}+U\mathbf{1}_{\{S^{(\infty)}>1-\phi(t)\}},\tag{3}$$

with $S^{(\infty)}$ defined in Theorem 3.1 and U a uniform random variable in $[1-\phi(t),1]$ *independent of $S^{(\infty)}$. Moreover, when $n\to\infty$ we have*

$$\mathbb{P}\big(S_{\mathbf{o}}^{(n)}(n\mu t)>0\big)\longrightarrow\mathbb{P}\big(S^{(\infty)}>1-\phi(t)\big)=\frac{|\phi'(t)|}{\mu}.$$

Finally, the random variable $S(t)$ has density

$$f_{S(t)}(x)=f_{S^{(\infty)}}(x)\mathbf{1}_{[0,1-\phi(t)]}+\frac{|\phi'(t)|}{\mu\phi(t)}\mathbf{1}_{[1-\phi(t),1]}$$

and, with $\|\cdot\|_{\mathrm{TV}}$ denoting the total variation distance, we have

$$\begin{aligned}\|\operatorname{law}(S(t))-\operatorname{law}(S^{(\infty)})\|_{\mathrm{TV}}&=\int_{1-\phi(t)}^{1-\mathbf{p}_0}\left|f_{S^{(\infty)}}(x)+\frac{|\phi'(t)|}{\mu\phi(t)}\right|dx+\frac{|\phi'(t)|\mathbf{p}_0}{\mu\phi(t)}\\&\le 2\frac{|\phi'(t)|}{\mu}.\end{aligned}$$

PROOF. The Laplace transform of $S^{(\infty)}$ is

$$\mathbb{E}\big(\exp\{-\lambda S^{(\infty)}\}\big)=\frac{1}{\mu}\int_0^\infty\phi''(u)\exp\{-\lambda(1-\phi(u))\}\,du$$

(see [3] or Remark 3.3). Now, from Theorem 3.2 we have

$$\lim_{n\to\infty}A_n(n\mu t,\lambda/n)=\frac{1}{\mu}\int_0^t\phi''(u)\exp\{-\lambda(1-\phi(u))\}\,du.$$

Taking $\lambda = 0$ and using Remark 2.2, we obtain

$$\begin{aligned}\lim_{n\to\infty} P\big(S^{(n)}(n\mu t) \le |R_{n\mu t}|\big) &= \lim_{n\to\infty} A_n(n\mu t, 0)\\ &= 1 - \frac{(-\phi'(t))}{\mu}\\ &= \mathbb{P}\big(S^{(\infty)} \le 1 - \phi(t)\big).\end{aligned}$$

On the other hand, since the Laplace transform of $\frac{S^{(n)}(n\mu t)}{n}$ conditional on the event $S^{(n)}(n\mu t) \le |R_{n\mu t}|$ is given by

$$\mathbb{E}\bigg(\exp\bigg\{-\lambda \frac{S^{(n)}(n\mu t)}{n}\bigg\}\bigg| S^{(n)}(n\mu t) \le |R_{n\mu t}|\bigg) = \frac{A_n(n\mu t, \lambda/n)}{A_n(n\mu t, 0)},$$

we obtain

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{E}\bigg(\exp\bigg\{-\lambda \frac{S^{(n)}(n\mu t)}{n}\bigg\}\bigg| S^{(n)}(n\mu t) \le |R_{n\mu t}|\bigg)\\ &\quad= \frac{1}{\mu + \phi'(t)} \int_0^t \phi''(u) \exp\{-\lambda(1 - \phi(u))\}\, du\\ &\quad= \mathbb{E}\big(\exp\{-\lambda S^{(\infty)}\} | S^{(\infty)} \le 1 - \phi(t)\big).\end{aligned}$$

Concerning the limiting behavior of B_n, we get in a similar way that

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\big(S^{(n)}(n\mu t) > |R_{n\mu t}|\big) &= -\frac{\phi'(t)}{\mu}\\ &= \mathbb{P}\big(S^{(\infty)} > 1 - \phi(t)\big),\end{aligned}$$

and

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\bigg(\exp\bigg\{-\lambda \frac{S^{(n)}(n\mu t)}{n}\bigg\}\bigg| S^{(n)}(n\mu t) > |R_{n\mu t}|\bigg) &= \lim_{n\to\infty} \frac{B_n(n\mu t, \lambda/n)}{B_n(n\mu t, 0)}\\ &= \bigg(\frac{e^{-\lambda(1-\phi(t))} - e^{-\lambda}}{\lambda \phi(t)}\bigg)\\ &= \mathbb{E}(\exp\{-\lambda U\}).\end{aligned}$$

Combining the previous limits yields

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{E}\bigg(\exp\bigg\{-\lambda \frac{S^{(n)}(n\mu t)}{n}\bigg\}\bigg)\\ &\quad= \lim_{n\to\infty} A_n(n\mu t, \lambda/n) + B_n(n\mu t, \lambda/n)\\ &\quad= \mathbb{E}\big(\exp\{-\lambda S^{(\infty)}\} \mathbf{1}_{\{S^{(\infty)} \le 1 - \phi(t)\}}\big)\\ &\qquad+ \mathbb{E}\big(\exp\{-\lambda U\}\big) P\big(S^{(\infty)} > 1 - \phi(t)\big)\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\big(\exp\{-\lambda S^{(\infty)}\}\mathbf{1}_{\{S^{(\infty)}\le 1-\phi(t)\}}\big)\\
&\quad + \mathbb{E}\big(\exp\{-\lambda U\}\mathbf{1}_{\{S^{(\infty)}>1-\phi(t)\}}\big)\\
&= \mathbb{E}\big(\exp -\lambda\{S^{(\infty)}\mathbf{1}_{\{S^{(\infty)}\le 1-\phi(t)\}} + U\mathbf{1}_{\{S^{(\infty)}>1-\phi(t)\}}\}\big).
\end{aligned}
$$

From the latter we obtain the density of $S(t)$, and then the total variation distance to equilibrium. The last asserted inequality follows from the well-known fact that $\|\operatorname{law}(X)-\operatorname{law}(Y)\|_{\mathrm{TV}} \le 2\mathbb{P}\{X\ne Y\}$ for any coupling of random variables (X, Y). □

COROLLARY 3.2. *Define two functions g_t and $\widetilde{g}_t$ by*

$$
g_t(y) = \int_0^y e^{-zt} P(dz) \quad \text{and} \quad \widetilde{g}_t(y) = g_t^{-1}(1-y) \qquad \text{if LLN-}P^- \text{ holds},
$$

or by

$$
g_t(y) = \int_{y^+}^\infty e^{-zt} P(dz) \quad \text{and} \quad \widetilde{g}_t(y) = (1-g_t)^{-1}(y) \qquad \text{if LLN-}P^+ \text{ holds}.
$$

(Here, g^{-1} stands for the generalized inverse of a nondecreasing right continuous function $g:\mathbb{R}_+\to\mathbb{R}_+$.) Then, under LLN-$P^-$ or LLN-P^+, for each $t>0$ we have

$$
\frac{S^{(n)}(n\mu t)}{n} \to^d S(t),
$$

where $S(t)$ has the density

$$
f_{S(t)}(x) = \mathbf{1}_{[0,1-\phi(t)]}(x) f_{S^{(\infty)}}(x) + \mathbf{1}_{[1-\phi(t),1]}(x)\frac{1}{\mu}\widetilde{g}_t(x). \tag{4}
$$

Moreover, we have

$$
\mathbb{P}\big(S_{\mathbf{o}}^{(n)}(n\mu t) > 0\big) \longrightarrow \mathbb{P}\big(S(t) > 1-\phi(t)\big) = \mathbb{P}\big(S^{(\infty)} > 1-\phi(t)\big) = \frac{|\phi'(t)|}{\mu},
$$

when $n\to\infty$, and for all $t\ge 0$, $\|\operatorname{law}(S(t)) - \operatorname{law}(S^{(\infty)})\|_{\mathrm{TV}} \le 2\frac{|\phi'(t)|}{\mu}$.

PROOF. If LLN-P^- holds, the result follows by using Theorem 3.3 and making the change of variable $z = 1 - g_t(x)$ to obtain

$$
\begin{aligned}
L(\mu, t, \lambda) &= \frac{1}{\mu}\int_0^\infty x e^{-xt} \exp\left\{-\lambda \int_{x^+}^\infty e^{-yt} P(dy)\right\} P(dx) \exp\{-\lambda\big(1-\phi(t)\big)\}\\
&= \frac{1}{\mu}\int_{1-\phi(t)}^1 \exp\{-\lambda z\} g_t^{-1}(1-z)\, dz.
\end{aligned}
$$

The remaining case is similar. □

REMARK 3.4. If $\tilde{S}^{(n)}(t)$ denotes the search cost of the MtF process in the original time-scale (see Section 2), it is clear that the previous results are equivalently stated replacing $S^{(n)}(n\mu t)$ by

$$\tilde{S}^{(n)}(n\mu t/w^{(n)}),$$

where $w^{(n)} = \sum_{j=1}^{n} w_j^{(n)}$.

3.1. *Examples and applications*. In what follows, we give the limiting distribution of the transient search-cost for examples of random or deterministic request probabilities. The first ones are examples where explicit computations can be easily done.

(1) Let $w_i \sim \text{Bernoulli}(p)$, then

$$f_{S(t)}(x) = \frac{1}{p}\mathbf{1}_{[0,p(1-e^{-t}))}(x) + \frac{e^{-t}}{1-p+pe^{-t}}\mathbf{1}_{[p(1-e^{-t}),1]}(x).$$

(2) Let $w_i \sim \text{Gamma}(1,\alpha)$, then

$$f_{S(t)}(x) = \left(1+\frac{1}{\alpha}\right)(1-x)^{1/\alpha}\mathbf{1}_{[0,u(t))}(x) + (1+t)^{-1}\mathbf{1}_{[u(t),1]}(x),$$

with $u(t) = 1-(1+t)^{-\alpha}$.

(3) If $w_i \sim \text{Geometric}(p)$, then

$$f_{S(t)}(x) = \frac{2(1-x)-p}{1-p}\mathbf{1}_{[0,u(t))}(x) + \frac{pe^{-t}}{1-(1-p)e^{-t}}\mathbf{1}_{[u(t),1]}(x),$$

where $u(t) = \frac{(1-p)(1-e^{-t})}{p+(1-p)(1-e^{-t})}$.

(4) If $w_i = 1$ or equivalently, $w_i \sim \delta_1$, we get $f_{S(t)}(x) = 1$ (using any of LLN-P-ex, LLN-P^+ or LLN-P^-). That is, the limiting search cost is uniform for all $t \geq 0$.

The stationary distributions associated with examples (5)(i) and (6)(i) below were first studied in Fill [6], whereas the stationary regimes of examples (5)(ii) and (6)(ii) were considered by Barrera, Huillet and Paroissin in [3]. The description of the stationary behavior of example (5)(i) is also included in Theorem 2 of Jelenković and Radovanović [11], in the more general context of the Persistent-Access-Caching (PAC) algorithm introduced therein (see the more detailed discussion below on the PAC algorithm and the Last-Recently-Used rule).

(5) Let $\alpha \in (-1,0)$ and define

$$P_\alpha(dx) = -\frac{1}{\alpha}x^{1/\alpha-1}\mathbf{1}_{[1,\infty)}(x)\,dx \qquad \text{(Pareto law)},$$

$$\phi(s) = -\frac{1}{\alpha}\int_1^\infty e^{-xs}x^{1/\alpha-1}\,dx,$$

$$g_t(y) = -\frac{1}{\alpha}\int_1^y e^{-xt}x^{1/\alpha-1}\,dx.$$

(i) If $w_i = i^\alpha$, we have using (2) that

$$f_{S(t)}(x) = -(\alpha+1)\frac{\phi''(\phi^{-1}(1-x))}{\phi'(\phi^{-1}(1-x))}\mathbf{1}_{[0,1-\phi(t))}(x) + (\alpha+1)g_t^{-1}(1-x)\mathbf{1}_{[1-\phi(t),1]}(x).$$

(ii) If w_i are i.i.d. with law P_α, then

$$f_{S(t)}(x) = -(\alpha+1)\frac{\phi''(\phi^{-1}(1-x))}{\phi'(\phi^{-1}(1-x))}\mathbf{1}_{[0,1-\phi(t))}(x) + (\alpha+1)\frac{|\phi'(t)|}{\phi(t)}\mathbf{1}_{[1-\phi(t),1]}(x).$$

(6) Let $\alpha > 0$ and set now

$$\phi(s) = \frac{1}{\alpha}\int_0^1 e^{-xs}x^{1/\alpha-1}\,dx,$$
$$g_t(y) = \frac{1}{\alpha}\int_y^1 e^{-xt}x^{1/\alpha-1}\,dx.$$

(i) If $w_i = i^\alpha$, we have by (2) that

$$f_{S(t)}(x) = -(\alpha+1)\frac{\phi''(\phi^{-1}(1-x))}{\phi'(\phi^{-1}(1-x))}\mathbf{1}_{[0,1-\phi(t))}(x) + (\alpha+1)(1-g_t)^{-1}(x)\mathbf{1}_{[1-\phi(t),1]}(x).$$

(ii) If w_i are i.i.d. with law Beta$(1, 1/\alpha)$, then

$$f_{S(t)}(x) = -(\alpha+1)\frac{\phi''(\phi^{-1}(1-x))}{\phi'(\phi^{-1}(1-x))}\mathbf{1}_{[0,1-\phi(t))}(x) + (1+\alpha)\frac{|\phi'(t)|}{\phi(t)}\mathbf{1}_{[1-\phi(t),1]}(x).$$

It was remarked in [3] that example (5)(i) shares the same stationary distribution as example (5)(ii), and example (6)(i) the same as that of (6)(ii). We observe here that the equilibrium parts of the transient search-costs of examples (5)(i) and (5)(ii) coincide as well, as happens also with examples (6)(i) and (6)(ii). In turn, their out of equilibrium transient search-cost are different. In Section 5 we discuss and explain these facts in the light of the new techniques that will be shortly introduced.

To motivate our last example, we recall that Jelenković [10] considered a continuum (thus infinite) list of objects representing an "efficient static" or popularity decreasing arrangement of objects. More precisely, a probability measure Q on $\mathbb{R}_+$ with decreasing density q is used therein to specify the probability $q(x)\,dx$ that an object lying at position $x \in \mathbb{R}_+$ is requested. The stationary search-cost was studied by approximating the continuum list by discrete albeit countably infinite lists $(Q^n)_{n\in\mathbb{N}}$, (a "fluid limit"). That is, for each n, object at position i/n, $i \in \mathbb{N}$, is requested with probability $Q^n(i) = Q((i+1)/n) - Q(i/n)$. In this case the transient search cost $S(t)$ can be obtained by our approach when the continuum list has finite length.

(7) Let $Q(dx) = q(x)\,dx$ be supported in $[0, c]$ and $w_i^{(n)}, i = 1, \dots, n$, be defined as in (d) above. If q is a continuous decreasing probability density, then LLN-P^- holds and we get

$$f_{S(t)}(x) = -c\frac{\phi''(\phi^{-1}(1-x))}{\phi'(\phi^{-1}(1-x))}\mathbf{1}_{[0,1-\phi(t))}(x) + cg_t^{-1}(1-x)\mathbf{1}_{[1-\phi(t),1]}(x),$$

where

$$\phi(s) = \frac{1}{c}\int_0^c e^{-q(x)s}\,dx \quad \text{and}$$
$$g_t(y) = \frac{1}{c}\int_0^y e^{-q(x)t}\,dx.$$

Observe that in this case $\mu = 1/c$ since $P(dx) = (\frac{1}{c}\mathbf{1}_{x\in[0,c]}\,dx) \circ q^{-1}$, so that the Laplace transform computed in Theorem 3.2 reads

$$\int_0^t \left(\int_0^c q^2(u)e^{-q(u)s}\,du\right)\exp\left(-\lambda\left(1 - \frac{1}{c}\int_0^c e^{-q(u)s}\,du\right)\right)ds.$$

The limit of this expression when $t \to \infty$ is exactly formula 4.1 in [10] evaluated in $s = \lambda/c$.

We notice that the fluid limit approximation of Jelenković [10] also corresponds to a law of large numbers asymptotic, in the sense that the (infinite) empirical measures $\frac{1}{n}\sum_{i\in\mathbb{N}}\delta_{i/n}$ approach dx in $\mathbb{R}_+$ as the space scale $1/n$ goes to 0. However, in our case the search-cost is defined in terms of the relative position in a finite list, whereas in [10] it is understood as the absolute position in a possibly infinite list. (The reader familiar with particle systems will recognize a similar difference between the hydrodynamic and mean field limit formalisms; it is from the latter that we have borrowed the law of large numbers terminology; see next section.) Although we can "simulate" the fluid limit for compactly supported measures Q [as example (7) shows] the general case is not tractable with our techniques (see the discussion in Section 5).

To finish the discussion on related works, we remark that our main results also describe the transient behavior of particular instances of the Least-Recently-Used (LRU) caching rule, which dynamically selects a collection of frequently accessed documents and stores them in a low cost access place. Indeed, the probability that at time $n\mu t$ the requested document is not found among the δn selected ones (and a fault occurs) corresponds to the probability $\mathbb{P}(S^{(n)}(n\mu t) > \delta n)$ that the search-cost in the MtF scheme is bigger than δn (for a more detailed discussion of this relation we refer to [10]). Consequently, from Corollary 3.1 we can, for instance, compute the transient asymptotic fault probability under assumption LLN-P-ex:

$$\mathbb{P}(S(t) > \delta) = \begin{cases} \dfrac{|\phi'(\eta_\delta)|}{\mu}, & \text{if } \eta_\delta < t, \\ \dfrac{1-\delta}{\mu} \frac{\phi'(t)}{\phi(t)}, & \text{if } \eta_\delta \geq t, \end{cases} \tag{5}$$

with $\eta_\delta = \phi^{-1}(1-\delta)$. Under LLN-$P^+$ or LLN-P^-, thanks to Corollary 3.2, the same value is obtained for the case $\eta_\delta < t$, and an integral expression in terms of $\tilde{g}_t$ (which can be written explicitly) in the case $\eta_\delta \geq t$.

We remark that the PAC algorithm introduced in [11] generalizes the LRU rule by updating the list in a similar way, but only if the requested item at time t has already been requested $k-1$ times in the time interval $((t-\beta) \vee 0, t)$. Thus, by taking $t = \infty$ in formula (5) (see Remark 3.3) we obtain a generalization of the stationary result of Theorem 2 of [11] in the case $k = 1, \beta > 0$ [the latter corresponding to the particular ϕ given by the Pareto law of example (5)(i)]. Moreover, for the case $k = 1$ and $w_i = i^\alpha$, $\alpha \in (-1, 0)$ of Theorem 2 of [11], we obtain the transient asymptotic fault probability. This is given by

$$\begin{aligned} &\mathbb{P}(S(t) > \delta) \\ &\quad = \begin{cases} -\dfrac{\alpha+1}{\alpha} \eta_\delta^{-(1+1/\alpha)} \Gamma(1+1/\alpha, \eta_\delta), & \text{if } \eta_\delta < t, \\ -\dfrac{\alpha+1}{\alpha} t^{-(1+1/\alpha)} [\Gamma(1+1/\alpha, t) - \Gamma(1+1/\alpha, t\varepsilon_{\delta,t})], & \text{if } \eta_\delta \geq t, \end{cases} \end{aligned}$$

where $\Gamma(z, y) := \int_y^\infty x^{z-1} e^{-x}\, dx$ is the incomplete Gamma function, and $\varepsilon_{\delta,t} := g_t^{-1}(1-\delta)$ with g_t as in example (5)(i).

3.2. *Stochastic order relations.* In the remainder of this section we shall establish some stochastic order relations between the three situations LLN-P-ex, LLN-P^+ and LLN-P^-. Recall that given two real valued random variable X and Y, we say that X *is stochastically smaller than* Y, if for all $z \in \mathbb{R}$, one has $\mathbb{P}(X \leq z) \geq \mathbb{P}(Y \leq z)$. This is written $X \preceq Y$.

Notice now that the three assumptions can be seen as a priori information of different type about the initial positions of objects in the list. More precisely, LLN-P-ex can be read as having no a priori knowledge at all, whereas LLN-P^- can be interpreted as the relative order of popularities being known, and objects being placed at time 0 in decreasing order (intuitively, this is an efficient statical ordering). Accordingly, assumption LLN-P^+ can be interpreted as the least efficient order at time 0, if the relative order of popularities is known. In this direction, Fill and Holst proved in Corollary 4.2 of [8] that for a given finite request probability vector, the transient search-cost is stochastically larger than that of the same vector rearranged in decreasing order, and smaller than when it is arranged in increasing order. We shall prove that similar stochastic order relations hold in the large numbers limit, by using the explicit expressions for $f_{S(t)}$ we have already found.

COROLLARY 3.3. *Let $S^{ex}(t)$, $S^{+}(t)$ and $S^{-}(t)$ denote the limiting transient search-cost $S(t)$ respectively under the assumptions, LLN-P-ex, LLN-P^{+} and LLN-P^{-}. Then, we have*

$$S^{-}(t) \preceq S^{ex}(t) \preceq S^{+}(t).$$

PROOF. From Corollaries 3.1 and 3.2, we just need to prove that

$$\begin{aligned}\mathbb{P}\{1-\phi(t) \le S^{-}(t) \le x\} &\ge \mathbb{P}\{1-\phi(t) \le S^{ex}(t) \le x\}\\ &\ge \mathbb{P}\{1-\phi(t) \le S^{+}(t) \le x\}\end{aligned}$$

for all $x \in [1-\phi(t), 1]$. The first inequality is equivalent to

$$\int_{g_t^{-1}(1-x)}^{\infty} ze^{-zt} P(dz) \ge \frac{|\phi'(t)|}{\phi(t)}\big(x-1+\phi(t)\big)$$

for all $x \in [1-\phi(t), 1]$, where $g_t(y) = \int_0^y e^{-zt} P(dz)$. This will follow if we can prove that

$$\frac{\int_{y+}^{\infty} ze^{-zt} P(dz)}{|\phi'(t)|} \ge \frac{\int_{y+}^{\infty} e^{-zt} P(dz)}{\phi(t)}$$

for all $y \ge 0$ or, equivalently, that

$$\frac{\int_0^y ze^{-zt} P(dz)}{|\phi'(t)|} \le \frac{\int_0^y e^{-zt} P(dz)}{\phi(t)}. \tag{6}$$

Observe that both sides have the same points of discontinuity, as functions of y. Therefore, by suitably approximating P, we may assume that $P(dz)$ has a continuous density $f(z)$ which is strictly positive. Write $a(y) = \int_0^y ze^{-zt} f(z)\,dz$ and $b(y) = \int_0^y e^{-zt} f(z)\,dz$. We need to check that

$$h(y) := \frac{a(y)}{a(\infty)} - \frac{b(y)}{b(\infty)} \le 0.$$

Since h is differentiable and $h(0) = h(\infty) = 0$, it is enough to prove that h has a unique critical point y_0 and that $h(y_0) \le 0$. By the assumption on P, the condition $h'(y_0) = 0$ is satisfied if and only if $y_0 = \frac{a(\infty)}{b(\infty)}$. But then, $h(y_0) \le 0$ is the same as

$$\frac{\int_0^{a(\infty)/b(\infty)} ze^{-zt} f(z)\,dz}{a(\infty)} \le \frac{\int_0^{a(\infty)/b(\infty)} e^{-zt} f(z)\,dz}{b(\infty)},$$

which is trivially true. We conclude that $\mathbb{P}\{1-\phi(t) \le S^{-}(t) \le x\} \ge \mathbb{P}\{1-\phi(t) \le S^{ex}(t) \le x\}$ for all $x \ge 0$. The remaining inequality is easily seen to follow also from (6). □

4. Law of large numbers for random partitions of the interval and proofs of Theorems 3.2 and 3.3. The main ingredient in the proofs of Theorems 3.2 and 3.3 will be what we call a "law of large numbers for random partitions of the interval." To illustrate this idea, consider first $(w_i)_{i\in\mathbb{N}}$ i.i.d. random variables in $\mathbb{R}_+$ of law P with finite mean $\mu > 0$, and the probability vector $\mathbf{p}^{(n)} = (p_i^{(n)})$ defined by

$$p_i^{(n)} := \frac{w_i}{\sum_{j=1}^n w_j}, \qquad i = 1, \dots, n.$$

Then, by the strong law of large numbers, we have

$$\big(n\mu p_1^{(n)}, \dots, n\mu p_k^{(n)}\big) \longrightarrow (w_1, \dots, w_k)$$

almost surely when $n \to \infty$. In particular, any $k \le n$ fixed coordinates of the vector $n\mu\mathbf{p}^{(n)}$ become independent as n tends to infinity, and the limiting law of each of them converges to P. The following result due to H. Tanaka implies that the empirical measures

$$\nu^{(n)} := \frac{1}{n}\sum_{i=1}^n \delta_{n\mu p_i^{(n)}},$$

converge to P, as n goes to infinity.

PROPOSITION 4.1. *For each $n \in \mathbb{N}$, let $X^{(n)} = (X_1^{(n)}, \dots, X_n^{(n)})$ be an exchangeable random vector in $\mathbb{R}^n$ with law P_n. Then, the following assertions are equivalent*:

(i) *There exists a probability measure P in $\mathbb{R}$ such that for all $k \in \mathbb{N}$, when $n \to \infty$,*

$$\operatorname{law}\big(X_1^{(n)}, \dots, X_k^{(n)}\big) \Longrightarrow P^{\otimes k}.$$

(ii) *The random variables $\frac{1}{n}\sum_{i=1}^n \delta_{X_i^{(n)}}$ [taking values in the polish space $\mathcal{P}(\mathbb{R})$] converge in law as n goes to infinity to a deterministic limit equal to P.*

A sequence of probability measures P_n satisfying condition (i) of Proposition 4.1 is said to be P-chaotic, or to have the propagation of chaos property with limiting law P. This is a central property in the probabilistic study of mean field models. For further background on these topics and a proof of Proposition 4.1, we refer the reader to Sznitman's course [14].

We now prove that the same conclusion about $\nu^{(n)}$ can be obtained under a weaker assumption on the vectors $(w_i^{(n)})_{i=1}^n$, $n \in \mathbb{N}$. Namely, we have:

THEOREM 4.1 (L.L.N. for random partitions of the interval). *Assume that $(\mathbf{w}^{(n)})_{n\in\mathbb{N}}$ satisfy condition LLN-P and let $(\mathbf{p}^{(n)})_{n\in\mathbb{N}}$ be defined as in* (1). *Then,* *the empirical measure*

$$\nu^{(n)} := \frac{1}{n}\sum_{i=1}^{n}\delta_{n\mu p_i^{(n)}}$$

converges in law to the deterministic limit P.

This amounts to say that if LLN-P holds for $(w_i^{(n)})_{i=1}^n$ and some sequence (Z_n), then it also holds for $(p_i^{(n)})_{i=1}^n$ and the sequence $(Z_n') := (n\mu)$ (the convergence of the empirical means of $\nu^{(n)}$ being trivial).

PROOF OF THEOREM 4.1. The proof is simple by using the Wasserstein distance W_1 in the space $\mathcal{P}_1(\mathbb{R})$ of Borel probability measures on $\mathbb{R}$ with finite first moment. Recall that

$$W_1(m,m') = \inf_Q \int_{\mathbb{R}^2} |x-y| Q(dx,dy),$$

where the inf is taken over all Borel probability measures Q on $\mathbb{R}^2$ with first and second marginal laws in $\mathcal{P}_1(\mathbb{R})$ respectively equal to m and m' (i.e., couplings of m and m'). Then, W_1 is a distance inducing the weak topology, strengthened with the convergence of first-order moments (see, e.g., [16]). Let us define

$$Q^{(n)} := \frac{1}{n}\sum_{i=1}^{n}\delta_{(n\mu p_i^{(n)}, Z_n w_i^{(n)})}$$

which is a coupling of $\hat{\nu}^{(n)}$ and $\nu^{(n)}$. Then, on the event $\{\sum_{j=1}^n Z_n w_j^{(n)} > 0\}$ we have

$$\begin{aligned}\int_{\mathbb{R}^2} |x-y| Q^{(n)}(dx,dy) &= \frac{1}{n}\sum_{i=1}^{n}\left|n\mu p_i^{(n)} - Z_n w_i^{(n)}\right| \\ &= \frac{1}{n}\sum_{i=1}^{n} Z_n w_i^{(n)}\left|\frac{n\mu}{\sum_{j=1}^n Z_n w_j^{(n)}} - 1\right| \\ &= \left|\mu - \frac{\sum_{i=1}^n Z_n w_i^{(n)}}{n}\right|,\end{aligned}$$

from where

$$W_1(\nu^{(n)}, \hat{\nu}^{(n)}) \le \left|\frac{\sum_{i=1}^n Z_n w_i^{(n)}}{n} - \mu\right|.$$

We deduce that

$$W_1(\nu^{(n)}, P) \le \left|\frac{\sum_{i=1}^n Z_n w_i^{(n)}}{n} - \mu\right| + W_1(\hat{\nu}^{(n)}, P).$$

Now, from LLN-P we have $\frac{\sum_{i=1}^{n} Z_n w_i^{(n)}}{n} \to \mu$ and $\hat{\nu}^{(n)} \to P$, both in probability (the second with respect to W_1). On the other hand, we have $\mathbb{P}\{\sum_{j=1}^{n} Z_n w_j^{(n)} = 0\} \to 0$ as $n \to \infty$, since $\mu > 0$. We deduce that $\nu^{(n)}$ converges to P in probability with respect to W_1, and therefore also with respect to the usual weak topology. □

REMARK 4.1. Assumption LLN-P together with Theorem 4.1, imply that for any bounded continuous function $\Psi : \mathcal{P}(\mathbb{R}_+) \to \mathbb{R}$, one has

$$\mathbb{E}(\Psi(\nu^n)) \to \Psi(P) \qquad \text{when } n \to \infty.$$

We shall systematically rely on this fact in the proof of Theorems 3.2 and 3.3, in order to compute the limits of quantities of the form $\mathbb{E}(f_n(\mathbf{p}^{(n)}))$, for adequate functions $f_n : \mathbb{R}_+^n \to \mathbb{R}$, $n \in \mathbb{N}$. Namely, we will use the fact that

$$\mathbb{E}\big(f_n\big(\mathbf{p}^{(n)}\big)\big) \to \Psi(P)$$

whenever $f_n(\mathbf{p}^{(n)})$ is equal (or asymptotically close enough) to $\Psi(\nu^n)$, for some bounded continuous Ψ not depending on n.

We shall also need the following lemma on the size-biased picking of probability measures on $\mathbb{R}^+$. Recall that given $\mathbf{m} \in \mathcal{P}(\mathbb{R}_+)$ with $0 < \langle \mathbf{m}, x \rangle < \infty$, the *size-biased picking of* $\mathbf{m}$ is the law

$$\overline{\mathbf{m}}(dy) := \frac{y}{\langle \mathbf{m}, x \rangle} \mathbf{m}(dy);$$

it is obtained from $\mathbf{m}$ as in the *waiting time paradox* (see, e.g., Feller [5], Chapter VI).

LEMMA 4.1. *Let* $(\mathbf{m}_n)$ *be a sequence of probability measures on* $\mathbb{R}_+$ *with finite means and weakly converging to a probability measure* $\mathbf{m} \neq \delta_0$. *Assume moreover that* $\langle \mathbf{m}, x \rangle < \infty$ *and that* $\langle \mathbf{m}_n, x \rangle \to \langle \mathbf{m}, x \rangle$ *when n goes to* ∞. *Then, we have*

$$\overline{\mathbf{m}}_n \Longrightarrow \overline{\mathbf{m}}.$$

PROOF. Since $\langle \overline{\mathbf{m}}_n, 1 \rangle \longrightarrow \langle \overline{\mathbf{m}}, 1 \rangle$ as n goes to ∞, it is enough to prove that $\langle \overline{\mathbf{m}}_n, f \rangle \longrightarrow \langle \overline{\mathbf{m}}, f \rangle$ for each continuous function f with compact support. Since for such f the function $xf(x)$ is continuous and bounded, this follows from the assumptions. □

PROOF OF THEOREM 3.2. In what follows, we drop for notational simplicity the superscript (n) of the popularity $p_i^{(n)} = p_i$.

From Proposition 2.1, it holds that

$$A_n(n\mu t, \lambda/n) = \mathbb{E}\left(\int_0^{n\mu t} e^{-u} \sum_{i=1}^{n} p_i^2 \left(\prod_{j=1, j\neq i}^{n} \big(1 + (e^{p_j u} - 1)e^{-\lambda/n}\big)\right) du\right).$$

Let $g_n, h_n : \mathbb{R}_+^2 \to \mathbb{R}$ be functions defined by

$$g_n(u,x) = \frac{x^2 e^{-xu}}{1-(1-e^{-xu})(1-e^{-\lambda/n})},$$
$$h_n(u,x) = n\log\bigl(1-(1-e^{-xu})(1-e^{-\lambda/n})\bigr).$$

Making the right change of variable we can write $A_n(n\mu t, \lambda/n)$ as

$$A_n(n\mu t,\lambda/n) = \frac{1}{\mu}\mathbb{E}\Biggl(\int_0^t \frac{1}{n}\sum_{i=1}^n g_n(u, n\mu p_i)\exp\Biggl(\frac{1}{n}\sum_{i=1}^n h_n(u,n\mu p_i)\Biggr)du\Biggr)$$
$$= \frac{1}{\mu}\mathbb{E}\biggl(\int_0^t \langle \nu^{(n)}, g_n(u,\cdot)\rangle \exp(\langle \nu^{(n)}, h_n(u,\cdot)\rangle)\,du\biggr).$$

Now define $g(u,x) = x^2 e^{-xu}$ and $h(u,x) = -(1-e^{-xu})\lambda$. Then, if

$$\widetilde{A}_n(n\mu t,\lambda/n) := \frac{1}{\mu}\mathbb{E}\biggl(\int_0^t \langle \nu^{(n)}, g(u,\cdot)\rangle \exp(\langle \nu^{(n)}, h(u,\cdot)\rangle)\,du\biggr),$$

we see that

$$|\widetilde{A}_n(n\mu t,\lambda/n) - A_n(n\mu t,\lambda/n)| \le I_1(t,\lambda) + I_2(t,\lambda),$$

with $I_1(t,\lambda)$ and $I_2(t,\lambda)$ defined by

$$I_1(t,\lambda) = \frac{1}{\mu}\mathbb{E}\biggl(\int_0^t \langle \nu^{(n)}, |g_n(u,\cdot) - g(u,\cdot)|\rangle \exp(\langle \nu^{(n)}, h_n(u,\cdot)\rangle)\,du\biggr),$$
$$I_2(t,\lambda) = \frac{1}{\mu}\mathbb{E}\biggl(\int_0^t \langle \nu^{(n)}, g(u,\cdot)\rangle \bigl|\exp(\langle \nu^{(n)}, h_n(u,\cdot) - h(u,\cdot)\rangle) - 1\bigr|\,du\biggr).$$

On the other hand, we have the following estimates for n large enough:

$$\begin{aligned}|g_n(u,x) - g(u,x)| &\le g(u,x)\frac{1-e^{-\lambda/n}}{1-(1-e^{-ux})(1-e^{-\lambda/n})} \\ &\le 2g(u,x)\lambda/n\end{aligned} \tag{7}$$

(we use the bound $\frac{1-e^{-\alpha}}{1-c(1-e^{-\alpha})} \le e^{\alpha} - 1$ for $c \in [0,1], \alpha \ge 0$), and

$$\begin{aligned}|h_n(u,x) - h(u,x)| &\le 2\lambda\biggl\{\biggl(\frac{\log(1-(1-e^{-xu})(1-e^{-\lambda/n}))}{(1-e^{-xu})(1-e^{-\lambda/n})} + 1\biggr) + \lambda/n\biggr\} \\ &\le \frac{8\lambda^2}{n}.\end{aligned} \tag{8}$$

In the last line, we have used the bound

$$\biggl|\frac{\log(1-c(1-e^{-\alpha}))}{c(1-e^{-\alpha})} + 1\biggr| \le 2(1-e^{-\alpha})$$

for all $c \in [0, 1]$ and $(1 - e^{-\alpha}) \le 1/2$.

Estimates (7) and (8) imply that for large enough n, we have

$$I_1(t, \lambda) \le \frac{2}{\mu}\mathbb{E}\left(\int_0^t \langle \nu^{(n)}, g(u, \cdot)\rangle\, du\right)\lambda/n$$

and

$$\begin{aligned} I_2(t, \lambda) &\le \frac{1}{\mu}\mathbb{E}\left(\int_0^t \langle \nu^{(n)}, g(u, \cdot)\rangle \langle \nu^{(n)}, |h_n(u, \cdot) - h(u, \cdot)|\rangle\, du\right) \\ &\le \frac{8}{\mu}\mathbb{E}\left(\int_0^t \langle \nu^{(n)}, g(u, \cdot)\rangle\, du\right)\frac{\lambda^2}{n}. \end{aligned}$$

Since $\int_0^t \langle \nu^{(n)}, g(u, \cdot)\rangle\, du = \int_{\mathbb{R}_+} x(1 - e^{-xt})\nu^{(n)}(dx) \le \mu$ by Fubini's theorem, we get from the previous estimates that

$$|\widetilde{A}_n(n\mu t, \lambda/n) - A_n(n\mu t, \lambda/n)| \le \frac{C}{n}$$

for all n large enough. Consequently, we just need to prove that

$$\lim_{n\to\infty} \widetilde{A}_n(t, s) = \frac{1}{\mu}\mathbb{E}\left(\int_0^t \langle P, g(u, \cdot)\rangle \exp\langle P, h(u, \cdot)\rangle\, du\right). \tag{9}$$

Let us set

$$\begin{aligned} \triangle(t, n) := \bigg|\mathbb{E}\left(\int_0^t \langle \nu^{(n)}, g(u, \cdot)\rangle \exp\left(\langle \nu^{(n)}, h(u, \cdot)\rangle\right) du\right) \\ - \int_0^t \langle P, g(u, \cdot)\rangle \exp\langle P, h(u, \cdot)\rangle\, du\bigg|. \end{aligned}$$

For each $\delta > 0$, since $h(u, x) \le 0$ we have the estimate

$$\begin{aligned} \triangle(t, n) \le \bigg|\mathbb{E}\left(\int_\delta^t \langle \nu^{(n)}, g(u, \cdot)\rangle \exp\langle \nu^{(n)}, h(u, \cdot)\rangle\, du\right) \\ - \int_\delta^t \langle P, g(u, \cdot)\rangle \exp\langle P, h(u, \cdot)\rangle\, du\bigg| \\ + \int_0^\delta \mathbb{E}\langle \nu^{(n)}, g(u, \cdot)\rangle\, du + \int_0^\delta \langle P, g(u, \cdot)\rangle\, du. \end{aligned}$$

Observe that for each $u > 0$ the functions $g(u, \cdot)$ and $h(u, \cdot)$ are continuous and bounded. Moreover, for each $\delta > 0$, the restriction of g to $[\delta, \infty]$ is uniformly bounded. Thus, by using dominated convergence, the mapping

$$\nu \mapsto F(\nu) := \int_\delta^t \langle \nu, g(u, \cdot)\rangle \exp\langle \nu, h(u, \cdot)\rangle\, du$$

is seen to be continuous and bounded on $\in \mathcal{P}(\mathbb{R}_+)$. Thanks to LLN-$P$ and Theorem 4.1, we deduce that

$$\mathbb{E}\big(F(\nu^{(n)})\big) \to F(P) \qquad \text{when } n \text{ goes to } \infty$$

and, consequently, we get that for any $\delta > 0$

$$\limsup_{n\to\infty} \triangle(t,n) \le \sup_{n\in\mathbb{N}} \int_0^\delta \mathbb{E}\langle \nu^{(n)}, g(u,\cdot)\rangle\, du + \int_0^\delta \langle P, g(u,\cdot)\rangle\, du. \tag{10}$$

In order to prove (9) it is therefore enough to establish that the two terms on the r.h.s. of inequality (10) go to 0 with δ. Notice that the second term is equal to

$$\begin{aligned}\int_{\mathbb{R}_+}\left(\int_0^\delta x^2 e^{-xu}\, du\right) P(dx) &= \int_{\mathbb{R}_+} x P(dx) - \int_{\mathbb{R}_+} x e^{-x\delta} P(dx)\\ &= \mu\big(\overline{\phi}(0) - \overline{\phi}(\delta)\big),\end{aligned}$$

where $\overline{\phi}(s) := \frac{1}{\mu}\int_{\mathbb{R}_+} x e^{-sx} P(dx)$ is the Laplace transform of the size-biased picking of P. Thus, that term goes to 0 with δ by continuity of $\bar{\phi}$.

To tackle the first term on the r.h.s in (10), we consider the intensity measures associated with the random measures $\nu^{(n)}$. That is, the (deterministic) probability measures defined for each $n \in \mathbb{N}$ by

$$\langle \mathbf{m}_n, f\rangle := \mathbb{E}\langle \nu^{(n)}, f\rangle.$$

Notice that $\mathbf{m}_n$ has mean μ for all $n \in \mathbb{N}$. On the other hand, if we denote by $\overline{\mathbf{m}}_n$ the size-biased picking of $\mathbf{m}_n$, we get through similar computations as before that

$$\int_0^\delta \mathbb{E}\langle \nu^{(n)}, g(u,\cdot)\rangle\, du = \mu\big(\overline{\phi}_n(0) - \overline{\phi}_n(\delta)\big),$$

with $\overline{\phi}_n(s) := \frac{1}{\mu}\int_{\mathbb{R}_+} x e^{-sx}\mathbf{m}_n(dx)$ the Laplace transform of $\overline{\mathbf{m}}_n$.

Consequently, what we need to prove is that

$$\lim_{\delta\to 0}\sup_{n\in\mathbb{N}} |\overline{\phi}_n(\delta) - \overline{\phi}_n(0)| = 0. \tag{11}$$

But from LLN-P and Theorem 4.1, for all $f \in C_b(\mathbb{R})$ we have that

$$\langle \mathbf{m}_n, f\rangle = \mathbb{E}\langle \nu^{(n)}, f\rangle \to \langle P, f\rangle$$

since the mapping $\nu \mapsto \langle \nu, f\rangle$ is continuous and bounded. In other words, the sequence $\mathbf{m}_n$ converges weakly to P. With Lemma 4.1 we deduce that the sequence $\overline{\mathbf{m}}_n$ is weakly convergent, and therefore, by standard properties of the Laplace transform, the family of functions $(\overline{\phi}_n)_{n\in\mathbb{N}}$ is equicontinuous. Clearly, this implies that (11) holds, and the proof is finished. □

In the remaining proof we shall use the following result.

LEMMA 4.2. *Let F_m denote the distribution function of $m \in \mathcal{P}(\mathbb{R})$, and*

$$F_m^{-1}(x) := \inf\{t \ge 0 : F_m(t) \ge x\}$$

be its generalized inverse. Assume that $m_k \in \mathcal{P}(\mathbb{R})$ converges weakly to m. Then, $F_{m_k}^{-1}(x)$ converges to $F_m^{-1}(x)$ for dx—almost every $x \in [0,1]$.

PROOF. By Lemma 21.2 in van der Waart [15], $F_{m_k}^{-1}(x)$ converges to $F_m^{-1}(x)$ for all x at which F_m^{-1} is continuous. Since F_m^{-1} is increasing, this fails to happen for at most countably many points $x \in [0, 1]$. The statement follows. □

PROOF OF THEOREM 3.3. Recall that we always take $\pi = Id$. From Proposition 2.1 we have

$$B_n(n\mu t, \lambda/n) = \mathbb{E}\left(\sum_{i=1}^n p_i e^{-n\mu t} \prod_{j=1, j\neq i}^n [\mathbf{1}_{i<j} + (e^{n\mu p_j t} - \mathbf{1}_{i<j})e^{-\lambda/n}]\right).$$

Since $\sum_j p_j = 1$, we can rewrite

$$\begin{aligned} &B_n(n\mu t, \lambda/n) \\ &\quad = \mathbb{E}\left(\sum_{i=1}^n p_i e^{-n\mu p_i t} \prod_{j=1, j\neq i}^n [1 - (1 - e^{-\lambda/n})(1 - \mathbf{1}_{i<j} e^{-n\mu p_j t})]\right). \end{aligned}$$

Let us define

$$\tilde{B}_n := \mathbb{E}\left(\frac{1}{n\mu} \sum_{i=1}^n n\mu p_i e^{-n\mu p_i t} \exp\left\{-\lambda/n \sum_{j=1, j\neq i}^n (1 - e^{-n\mu p_j t}\mathbf{1}_{i<j})\right\}\right).$$

It is elementary to check that $|B_n(n\mu t, \lambda/n) - \tilde{B}_n| \le \frac{C}{n}$, so we shall study the term $\tilde{B}_n$. We have that

$$\begin{aligned} \tilde{B}_n = \mathbb{E}\Bigg(&\exp\left\{-\lambda/n \sum_{j=1}^n 1 - e^{-n\mu p_j t}\right\} \\ &\times \frac{1}{n\mu} \sum_{i=1}^n n\mu p_i e^{-n\mu p_i t} \exp\left\{-\lambda/n \sum_{j=1}^i e^{-n\mu p_j t}\right\} e^{\lambda/n}\Bigg). \end{aligned}$$

Therefore, thanks to the bound $xe^{-xt} \le \frac{1}{t}$ we have

$$|e^{-\lambda/n}\tilde{B}_n - L(\mu, t, \lambda)| \le \frac{1}{\mu\lambda}\mathbb{E}|\Psi(\nu^n)| + |\mathbb{E}(\hat{\mathcal{L}}_n(\mu, t, \lambda)) - \hat{\mathcal{L}}(\mu, t, \lambda)|,$$

with

$$\Psi(m) := \exp\left\{-\lambda \int_{\mathbb{R}_+} 1 - e^{-xt} m(dx)\right\} - \exp\{-\lambda(1 - \phi(t))\},$$

$$\hat{\mathcal{L}}_n(\mu, t, \lambda) := \frac{1}{n\mu} \sum_{i=1}^n n\mu p_i e^{-n\mu p_i t} \exp\left\{-\lambda/n \sum_{j=1}^i e^{-n\mu p_j t}\right\}$$

and $\hat{\mathcal{L}}(\mu, t, \lambda)$ defined as follows:

$$\hat{\mathcal{L}}(\mu, t, \lambda) = \frac{-\phi'(t)}{\mu} \int_0^1 e^{-\lambda\phi(t)x}\,dx \qquad \text{if LLN-}P\text{-ex holds,}$$

$$\hat{\mathcal{L}}(\mu, t, \lambda) = \frac{1}{\mu} \int_0^\infty x e^{-xt} \exp\left\{-\lambda \int_x^\infty e^{-yt} P(dy)\right\} P(dx) \qquad \text{if LLN-}P^- \text{ holds}$$

or

$$\hat{\mathcal{L}}(\mu, t, \lambda) = \frac{1}{\mu} \int_0^\infty x e^{-xt} \exp\left\{-\lambda \int_0^x e^{-yt} P(dy)\right\} P(dx) \qquad \text{if LLN-}P^+ \text{ holds.}$$

Since Ψ is continuous and bounded in $\mathcal{P}(\mathbb{R}_+)$ and $\Psi(P) = 0$, we get by LLN-P and Theorem 4.1 that $\mathbb{E}|\Psi(\nu^n)| \to 0$ when $n \to \infty$. Thus, we just have to prove that

$$\mathbb{E}(\hat{\mathcal{L}}_n(\mu, t, \lambda)) \longrightarrow \hat{\mathcal{L}}(\mu, t, \lambda).$$

The exchangeable case. Notice that under LLN-P-ex,

$$\begin{aligned}
&\mathbb{E}(\hat{\mathcal{L}}_n(\mu, t, \lambda)) \\
&\quad = \mathbb{E}\left(\frac{1}{n\mu} \sum_{i=1}^n \frac{1}{n!} \sum_{\sigma \in \Pi} n\mu p_{\sigma(i)} e^{-n\mu p_{\sigma(i)} t} \right. \\
&\qquad\qquad \left. \times \exp\left\{-\lambda/n \sum_{j=1}^i e^{-n\mu p_{\sigma(j)} t}\right\}\right) \\
&\quad = \mathbb{E}\left(\frac{1}{n\mu} \sum_{i=1}^n \sum_{k=1}^n \frac{1}{n!} \sum_{\sigma \in \Pi, \sigma(i) = k} n\mu p_{\sigma(i)} e^{-n\mu p_{\sigma(i)} t} \right. \\
&\qquad\qquad \left. \times \exp\left\{-\lambda/n \sum_{j=1}^i e^{-n\mu p_{\sigma(j)} t}\right\}\right) \\
&\quad = \mathbb{E}\left(\frac{1}{n\mu} \sum_{k=1}^n n\mu p_k e^{-n\mu p_k t} \right. \\
&\qquad\qquad \left. \times \frac{1}{n} \sum_{i=1}^n \frac{1}{(n-1)!} \sum_{\sigma \in \Pi, \sigma(i) = k} \exp\left\{-\lambda/n \sum_{j=1}^i e^{-n\mu p_{\sigma(j)} t}\right\}\right).
\end{aligned}$$

Since by LLN-P and Theorem 4.1,

$$\mathbb{E}\left(\frac{1}{n\mu} \sum_{k=1}^n n\mu p_k e^{-n\mu p_k t}\right) \int_0^1 e^{-\lambda\phi(t)x}\,dx \longrightarrow \frac{-\phi'(t)}{\mu} \int_0^1 e^{-\lambda\phi(t)x}\,dx$$

when $n \to \infty$, it is enough to show that $\delta_n(\mu, t, \lambda)$ goes to 0 when $n \to \infty$, where

$$\begin{aligned}\delta_n(\mu, t, \lambda) &:= \mathbb{E}(\hat{\mathcal{L}}_n(\mu, t, \lambda)) - \mathbb{E}\left(\frac{1}{n\mu}\sum_{k=1}^{n} n\mu p_k e^{-n\mu p_k t}\right)\int_0^1 e^{-\lambda\phi(t)x}\,dx \\ &= \mathbb{E}(\hat{\mathcal{L}}_n(\mu, t, \lambda)) - \frac{1}{\mu}\mathbb{E}\left(\int_{\mathbb{R}_+} x e^{-xt}\nu^{(n)}(dx)\right)\int_0^1 e^{-\lambda\phi(t)x}\,dx.\end{aligned}$$

Let us write for $i = 1, \dots, n-1$, and a permutation σ of $\{1, \dots, n\}$,

$$\alpha_t^\sigma(i, n) := \sum_{j=1}^{i} e^{-n\mu p_{\sigma(j)} t} \quad \text{and} \quad \alpha_t(n, n) := \sum_{j=1}^{n} e^{-n\mu p_j t}.$$

Define furthermore

$$\begin{aligned}I_n^k = \frac{1}{n}\sum_{i=1}^{n}\Bigg(&\exp\left\{-\frac{\lambda}{n}\alpha_t(n, n)\frac{i}{n}\right\} \\ &\times \frac{1}{(n-1)!}\sum_{\sigma\in\Pi, \sigma(i)=k}\left[\exp\left\{-\lambda/n\left[\alpha_t^\sigma(i, n) - \frac{i}{n}\alpha_t(n, n)\right]\right\} - 1\right]\Bigg),\end{aligned}$$

$$II_n = \frac{1}{n}\sum_{i=1}^{n}\exp\left\{-\frac{\lambda}{n}\alpha_t(n, n)\frac{i}{n}\right\} - \exp\left\{-\lambda\phi(t)\frac{i}{n}\right\}$$

and

$$III_n = \frac{1}{n}\sum_{i=1}^{n}\exp\left\{-\lambda\phi(t)\frac{i}{n}\right\} - \int_0^1 e^{-\lambda\phi(t)x}\,dx.$$

Then, we have

$$\begin{aligned}|\delta_n(\mu, t, \lambda)| &\le \left|\mathbb{E}\left(\frac{1}{n\mu}\sum_{k=1}^{n} n\mu p_k e^{-n\mu p_k t}(I_n^k + II_n + III_n)\right)\right| \\ &\le \frac{1}{t\mu}\left[\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}|I_n^k| + \mathbb{E}|II_n| + |III_n|\right]\end{aligned}$$

thanks to the bound $xe^{-xt} \le \frac{1}{t}$. Term III_n clearly goes to 0 when $n \to \infty$. On the other hand, we have

$$\mathbb{E}|II_n| \le \frac{\lambda}{n}\sum_{i=1}^{n}\frac{i}{n}\mathbb{E}\left|\frac{1}{n}\alpha_t(n, n) - \phi(t)\right| \le \frac{\lambda}{2}\mathbb{E}\left|\int_{\mathbb{R}_+} e^{-xt}\nu^n(dx) - \phi(t)\right|.$$

The mapping $\nu \mapsto |\int_{\mathbb{R}_+} e^{-xt}\nu(dx) - \int_{\mathbb{R}_+} e^{-xt}P(dx)|$ being continuous and bounded on $\mathcal{P}(\mathbb{R}_+)$, the latter term goes to 0 by LLN-P and Theorem 4.1.

Now, by exchangeability $\mathbb{E}|I_n^k|$ does not depend on k, and moreover, setting $\alpha_t(i,n) := \sum_{j=1}^{i} e^{-n\mu p_j t}$, we have

$$\begin{aligned}
\mathbb{E}|I_n^k| &\le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|\exp\left\{-\lambda/n\left[\alpha_t(i,n) - \frac{i}{n}\alpha_t(n,n)\right]\right\} - 1\right| \\
&\le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|\lambda/n\left[\sum_{j=1}^{i} e^{-n\mu p_j t} - \frac{i}{n}\sum_{k=1}^{n} e^{-n\mu p_k t}\right]\right| \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|\lambda/n\sum_{j=1}^{i}\left(e^{-n\mu p_j t} - \int e^{-xt}\nu^{(n)}(dx)\right)\right|,
\end{aligned}$$

and so

$$\mathbb{E}|I_n^k| \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|\lambda/n\sum_{j=1}^{i}\left(e^{-n\mu p_j t} - \phi(t)\right)\right| + \frac{\lambda}{2}\mathbb{E}\left|\phi(t) - \int e^{-xt}\nu^{(n)}(dx)\right|.$$

Thus, we just have to check that $IV_n := \frac{\lambda}{n^2}\sum_{i=1}^{n}\mathbb{E}|\sum_{j=1}^{i}(e^{-n\mu p_j t} - \phi(t))|$ goes to 0. Indeed, we have

$$\begin{aligned}
IV_n &\le \frac{\lambda}{n^2}\sum_{i=1}^{n}\left[\mathbb{E}\left(\sum_{j=1}^{i}\left(e^{-n\mu p_j t} - \phi(t)\right)\right)^2\right]^{1/2} \\
&= \frac{\lambda}{n^2}\sum_{i=1}^{n}\left[\sum_{j=1}^{i}\mathbb{E}\left(e^{-n\mu p_j t} - \phi(t)\right)^2\right. \\
&\qquad\qquad \left. + \sum_{k=1}^{i}\sum_{l=1, l\neq k}^{i}\mathbb{E}\left(e^{-n\mu p_l t} - \phi(t)\right)\left(e^{-n\mu p_k t} - \phi(t)\right)\right]^{1/2} \\
&\le \frac{2\lambda}{\sqrt{n}} + \frac{\lambda}{n^2}\sum_{i=1}^{n}\left[i(i-1)\left|\mathbb{E}\left(e^{-n\mu p_1 t} - \phi(t)\right)\left(e^{-n\mu p_2 t} - \phi(t)\right)\right|\right]^{1/2}.
\end{aligned}$$

Therefore,

$$IV_n \le \frac{2\lambda}{\sqrt{n}} + \lambda\left|\mathbb{E}\left(e^{-n\mu p_1 t} - \phi(t)\right)\left(e^{-n\mu p_2 t} - \phi(t)\right)\right|^{1/2}.$$

By LLN-P and Proposition 4.1(i) with $k = 2$, we conclude that the latter term goes to 0. This finishes the proof in the exchangeable case.

The monotone cases. We consider the case when LLN-P^+ holds, the decreasing case being similar. Notice that if $F_n^{-1}(x) := \inf\{t \ge 0 : F_n(t) \ge x\}$ is the generalized inverse of $F_n(x) = \nu^{(n)}([0,x])$, we have that

$$\hat{\mathcal{L}}_n(\mu, t, \lambda) := \frac{1}{\mu}\int_0^1 F_n^{-1}(x) e^{-F_n^{-1}(x)t}\exp\left\{-\lambda\int_0^{i_n(x)} e^{-F_n^{-1}(y)t}\,dy\right\}dx,$$

where $i_n(x) = \frac{\lceil nx \rceil}{n}$ and $\lceil \cdot \rceil$ is the ceiling function. On the other hand, under the law dx the generalized inverse $F^{-1} : [0, 1] \to \mathbb{R}_+$ of F is a random variable of law P. We thus have

$$\hat{\mathcal{L}}(\mu, t, \lambda) := \frac{1}{\mu} \int_0^1 F^{-1}(x) e^{-F^{-1}(x)t} \exp\left\{-\lambda \int_0^x e^{-F^{-1}(y)t}\, dy\right\} dx.$$

Thanks to this and the bound $xe^{-xt} \le \frac{1}{t}$, we get that

$$\begin{aligned}
&|\mathbb{E}(\hat{\mathcal{L}}_n(\mu, t, \lambda)) - \hat{\mathcal{L}}(\mu, t, \lambda)| \\
&\quad \le \frac{\lambda}{t\mu} \mathbb{E}\left(\int_0^1 \int_0^x |e^{-F_n^{-1}(y)t} - e^{-F^{-1}(y)t}|\, dy\, dx\right) \\
&\qquad + \frac{\lambda}{t\mu} \int_0^1 \int_x^{i_n(x)} e^{-F^{-1}(y)t}\, dy\, dx \\
&\qquad + \frac{1}{\mu} \mathbb{E}\left(\int_0^1 |F_n^{-1}(x) e^{-F_n^{-1}(x)t} - F^{-1}(x) e^{-F^{-1}(x)t}|\, dx\right) \\
&\quad \le \frac{\lambda}{t\mu} \mathbb{E}\left(\int_0^1 |e^{-F_n^{-1}(y)t} - e^{-F^{-1}(y)t}|\, dy\right) + \frac{\lambda}{nt\mu} \\
&\qquad + \frac{1}{\mu} \mathbb{E}\left(\int_0^1 |F_n^{-1}(x) e^{-F_n^{-1}(x)t} - F^{-1}(x) e^{-F^{-1}(x)t}|\, dx\right).
\end{aligned}$$

Therefore, and thanks also to LLN-P and Theorem 4.1, it is enough to prove that the bounded functionals on $\mathcal{P}(\mathbb{R}_+)$

$$\nu \mapsto \int_0^1 |e^{-F_\nu^{-1}(y)t} - e^{-F^{-1}(y)t}|\, dy$$

and

$$\nu \mapsto \int_0^1 |F_\nu^{-1}(x) e^{-F_\nu^{-1}(x)t} - F^{-1}(x) e^{-F^{-1}(x)t}|\, dx$$

are continuous, since they both vanish at $\nu = P$. This follows by dominated convergence and Lemma 4.2. The proof of the theorem is finished. $\square$

5. Concluding remarks. The limiting stationary regime of the MtF search-cost as the number of objects tend to infinity has been considered by several authors. One of the motivations is to compare efficiency among different popularity distributions when equilibrium is reached. Nonetheless, the rate at which equilibria are reached should also account for efficiency considerations. This was one of the motivations of the present article.

We have developed a general framework for studying the limiting dynamical behavior of the MtF search-cost when requests rates are sampled from empirical probability measures that asymptotically approach a specified law P. In this law of large numbers asymptotic regime, popularities of objects are comparable, in the sense that their asymptotic average is finite and nonnull. By this reason, although the transient behavior depends on the initial ordering, is not considerably sensitive to it. This can be seen in the fact that a common convergence rate to equilibrium $O(\int xe^{-tx}P(dx))$ was obtained in the three representative situations considered.

Our techniques also ensure the asymptotic stability under perturbations of request rates that preserve P. Namely, the limiting expectations of functionals of $(p_i^{(n)})_{i=1}^n$, which are symmetric functions of $\nu^{(n)}$, depend only on P. This was the case of the equilibrium part of the transient search-cost. The out-of-equilibrium transient search-cost involved in turn nonsymmetric functionals of $(p_i^{(n)})_{i=1}^n$, which yielded different limits according to the different "enumerations" of objects. This is the explanation for the coincidences and discrepancies pointed out in examples (5) and (6) of Section 3. Nevertheless, under the assumption of exchangeability one still might replace nonsymmetric functionals by symmetrized versions of them (as in the proof of Theorem 3.3) and obtain "symmetric" limits.

It is in principle possible to use our techniques in the asymptotic analysis of other sorting algorithms, at least in those cases where the corresponding relevant variables depend on the empirical measures of the popularities or of the request rates. However it is not obvious to identify which functionals of the empirical measures are involved.

On the other hand, the law of large numbers asymptotic behavior we have described corresponds to a very particular scaling limit, in the sense explained before. Therefore, it a priori excludes deterministic cases of interest such as the Zipf laws $w_i = i^\alpha$ with $\alpha \le -1$ or scaling approximations of the Poisson–Dirichlet distribution (see, e.g., [13], Chapter 9, and Joyce and Tavaré [12] for the limits of symmetric linear functionals of these random partitions). Neither the fluid limit approximation of the search-cost studied by Jelenković [10] is covered in its whole generality by our approach. Indeed, if Q has unbounded support, one might try to approximate Q by compactly supported laws as in point (d) of Section 3. But if one chooses therein $c = c_n$ diverging with n, the empirical means vanish as n goes to infinity.

In these examples, the transient dynamics may be of particular interest, since they exhibit coexistence of microscopic and macroscopic popularities which is likely to affect the convergence to equilibrium. A similar question could be of interest in the context of the PAC algorithm introduced in [11]. The splitting of the transient search-cost we introduced here could be useful in those cases. A combination of ideas in [3] and results in [12] could help to extend part of our arguments to Poisson–Kingman type asymptotics, although additional difficulties arise. The computation of the search-cost law involved highly nonlinear functionals of the empirical measures, which cannot be deduced from the asymptotics results obtained in [12]. These and related questions are addressed in progressing works by the authors.

Acknowledgments. The authors thank the anonymous referee for drawing our attention to the papers [10] and [11], so as for several suggestions that allowed us to improve the presentation of this work.

REFERENCES

[1] BARRERA, J. and PAROISSIN, C. (2004). On the distribution of the search cost for the move-to-front rule with random weights. *J. Appl. Probab.* **41** 250–262. MR2036286

[2] BARRERA, J., HUILLET, T. and PAROISSIN, C. (2005). Size-biased permutation of Dirichlet partitions and search-cost distribution. *Probab. Engrg. Inform. Sci.* **19** 83–97. MR2104552

[3] BARRERA, J., HUILLET, T. and PAROISSIN, C. (2006). Limiting search cost distribution for the move-to-front rule with random request probabilities. *Oper. Res. Lett.* **34** 557–563. MR2238404

[4] BODELL, J. (1997). Cost of searching—probablistic analysis of the self organizing Move-to-Front and Move-to-Root sorting rules. Ph.D. thesis, Mathematics Dept., Royal Institute of Technology, Sweden.

[5] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **II**. Wiley, New York. MR0210154

[6] FILL, J. A. (1996). Limits and rates of convergence for the distribution of search cost under the move-to-front rule. *Theoret. Comput. Sci.* **164** 185–206. MR1411204

[7] FILL, J. A. (1996). An exact formula for the move-to-front rule for self-organizing lists. *J. Theoret. Probab.* **9** 113–160. MR1371073

[8] FILL, J. A. and HOLST, L. (1996). On the distribution of search cost for the move-to-front rule. *Random Structures Algorithms* **8** 179–186. MR1603279

[9] FLAJOLET, P., GARDY, D. and THIMONIER, L. (1992). Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Appl. Math.* **39** 207–229. MR1189469

[10] JELENKOVIĆ, P. R. (1999). Asymptotic approximation of the move-to-front search cost distribution and least-recently used caching fault probabilities. *Ann. Appl. Probab.* **9** 430–464. MR1687406

[11] JELENKOVIĆ, P. R. and RADOVANOVIĆ, A. (2008). The persistent-access-caching algorithm. *Random Structures Algorithms* **33** 219–251. MR2436847

[12] JOYCE, P. and TAVARÉ, S. (1992). A convergence theorem for symmetric functionals of random partitions. *J. Appl. Probab.* **29** 280–290. MR1165214

[13] KINGMAN, J. F. C. (1993). *Poisson Processes*. *Oxford Studies in Probability* **3**. Oxford Univ. Press, New York. MR1207584

[14] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX*—1989. *Lecture Notes in Math.* **1464** 165–251. Springer, Berlin. MR1108185

[15] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistcial and Probabilistic Mathematics* **3**. Cambridge Univ. Press. MR1652247

[16] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483

ESCUELA DE INGENIERÍA Y CIENCIAS
UNIVERSIDAD ADOLFO IBÁÑEZ
AVDA. DIAGONAL LAS TORRES 2640
EDIFICIO C
PEÑALOLEN
CHILE
E-MAIL: javiera.barrera@uai.cl

DEPARTAMENTO DE INGENIERÍA MATEMÁTICA AND CENTRO DE MODELAMIENTO MATEMÁTICO
DIM-CMM, UMI(2807) UCHILE-CNRS
UNIVERSIDAD DE CHILE
CASILLA 170-3, CORREO 3
SANTIAGO
CHILE
E-MAIL: fontbona@dim.uchile.cl