Limiting search cost distribution for the move-to-front rule with random request probabilities

Javiera Barrera, Thierry Huillet, Christian Paroissin

HAL Id: hal-00005427
https://hal.science/hal-00005427

Submitted on 17 Jun 2005

Limiting search cost distribution for the move-to-front rule with random request probabilities

Javiera Barrera[a][*]   Thierry Huillet[b]   Christian Paroissin[c]

(a) MAP5 - UMR 8145 CNRS, Université Paris 5-René Descartes, 45 rue des Saints-Pères, 75270 PARIS cedex 06, FRANCE.

(b) LPTM - UMR 8089 CNRS, Université de Cergy-Pontoise, 2 avenue Adolphe-Chauvin, 95302 CERGY-PONTOISE, FRANCE.

(c) LMA - UMR 5142 CNRS, Université de Pau et des Pays de l'Adour, Avenue de l'Université, BP 1155, 64013 PAU cedex, FRANCE.

Abstract

Consider a list of n files whose popularities are random. These files are updated according to the move-to-front rule and we consider the induced Markov chain at equilibrium. We give the exact limiting distribution of the search-cost per item as n tends to infinity. Some examples are supplied.

ccsd-00005427, version 1 - 17 Jun 2005

[*]Corresponding author

[1]E-mail Addresses: `jbarrera@dim.uchile.cl` (J. Barrera), `Thierry.Huillet@ptm.u-cergy.fr` (T. Huillet) and `cparoiss@univ-pau.fr` (C. Paroissin).

1 Introduction and model

Consider a list of n files which is updated as follows: at each unit of discrete time, a file is requested independently of the previous requests and is moved to the front of the list. This heuristic is called the move-to-front rule and was first introduced by [14] and [10] to sort files. Such strategy is used when the request probabilities are unknown, otherwise we would list the files in order to have decreasing request probabilities. The move-to-front rule induces a Markov chain over the permutations of n elements which has a unique stationary distribution, (see [3] and reference to the work of Hendricks, Dies and Letac therein). This distribution turns out to be the size-biased permutation of the request probabilities.

Here, we consider that these request probabilities are themselves random, as in a Bayesian analysis. Let $\omega = (\omega_i)_{i \in I\!\!N^*}$ be a sequence of iid positive random variables. The Laplace transform of a weight will be denoted by ϕ and its expectation by μ. For any $i \in I\!\!N^*$, ω_i represents the weight of the file i. We can construct request probabilities $\mathbf{p} = (p_1, .., p_n)$ as follows:

$$\forall i \in \{1, ..., n\} , \qquad p_i = \frac{\omega_i}{W_n} \quad \text{where} \quad W_n = \sum_{i=1}^{n} \omega_i .$$

Such random vector $\mathbf{p}$ is called a random discrete distribution [8].

Let us denote by S_n the search cost of an item (i.e. the position in the list of the requested item) when the underlying Markov chain is in steady state (the first position will be 0). For this model, [2] obtained exact and asymptotic formulae for the Laplace transform of S_n (some results were also extended to the case of independent random weights). In particular, they found the limit of the expectation and the variance of S_n. Moreover, in the case of i.i.d. gamma weights, [1], obtained the exact and asymptotic distribution of S_n, using an exact representation of the size-biased permutation arising from Dirichlet partitions. Note that [5] found the limiting distribution of S_n when weights are deterministic but non-identical, in some cases (uniform, Zipf's law, generalized Zipf's law, power law and geometric).

In section 2, we shall give a general formula for the density of the limiting search cost distribution S, provided that the expected weight is finite. Then we derive the moment function and the cumulative distribution function of S. We also discuss the relationship between the move-to-front rule and the least-recently-used strategy. In section 3 we study some examples for which computations can be done explicitly: both continuous and discrete distributions are considered.

2 Limiting search cost distribution

The early analysis of the heuristic move-to-front focused on the expected search cost, see [10], [8] and [7], for instance. Later, researchers paid much attention to the (transient and stationary) distribution of the search cost ([6]). Some of them investigated the limiting behavior as the number n of items tends to infinity (see [5]). In a more recent article, [2] obtained an integral representation of the Laplace transform of S_n in the Bayesian model described in the introduction. Their main theorem is the following:

Theorem 2.1 *For a sequence ω of iid positive random variables,*

$$\forall s \geqslant 0, \qquad \phi_{S_n}(s) = n \int_0^\infty \int_t^\infty \phi''(r) \left[\phi(r) + e^{-s} \left(\phi(r-t) - \phi(r)\right)\right]^{n-1} dr\, dt .$$

In the same article the integral representation for the two first moments of S_n were derived. Moreover, they obtained a point-wise asymptotic equivalent for the Laplace transform of S_n and the limit of the first two moments of S_n/n when the number n of items tends to infinity. From theorem 2.1, we can obtain the following closed-form expression for the density function of the limiting distribution of Sn/n:

Theorem 2.2 *For a sequence ω of iid positive random weights with finite expectation μ,*

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{d} S ,$$

where S is a continuous random variable with the following density function f_S:

$$f_S(x) = -\frac{1}{\mu} \frac{\phi'' \left(\phi^{-1}(1-x)\right)}{\phi' \left(\phi^{-1}(1-x)\right)} \, \mathbb{1}_{[0,1-p_0]}(x) , \tag{1}$$

where $p_0 = \mathbb{P}(\omega_i = 0)$ and ϕ^{-1} is the inverse function of ϕ.

Remark 2.1 *The quantity p_0 can be interpreted as follows: p_0 is the probability that an item is never requested. At stationarity, one expects that any such item will be at the bottom of the list: np_0 is the mean number of unrequested items. So it is not surprising that the support of S is not the entire unit interval. Note that if the distribution of the weight is continuous, then $p_0 = 0$.*

Proof We have to prove that S_n/n converges in distribution, as n tends to infinity, to a certain random variable that will be denote by S. First, observe that:

$$\forall s \geqslant 0, \quad \phi_{S_n/n}(s) = \phi_{S_n}\left(\tfrac{s}{n}\right) .$$

So we are now interested in the limit of $\phi_{S_n}(s/n)$.

For any reals a and b such that $0 \leqslant a \leqslant b \leqslant \infty$, let:

$$I_n(a,b) = \int_a^b \phi''(r) \left[\phi(r) + e^{-s/n}(\phi(r-t) - \phi(r))\right]^{n-1} dr .$$

If $b = \infty$, then we will omit this parameter, i.e. $I_n(a) = I_n(a,\infty)$. Using these notations, theorem 2.1 gives:

$$\phi_{S_n}\left(\tfrac{s}{n}\right) = n \int_0^\infty I_n(t)\, dt . \tag{2}$$

We now decompose $I_n(t)$ into two parts: $I_n(t) = I_n(t, t+\varepsilon) + I_n(t+\varepsilon)$. We will prove that $nI_n(t+\varepsilon, \infty)$ tends to 0 when n tends to infinity:

$$\begin{aligned}
nI_n(t+\varepsilon, \infty) &= n \int_{t+\varepsilon}^\infty \phi''(r) \left[e^{-s/n}(\phi(r-t) + (1 - e^{-s/n})\phi(r))\right]^{n-1} dr , \\
&\leqslant n \int_{t+\varepsilon}^\infty \phi''(r)\phi(r-t)^{n-1} dr , \\
&\leqslant -n\phi(\varepsilon)^{n-1}\phi'(t+\varepsilon) ,
\end{aligned}$$

since ϕ is decreasing. Then $\lim_{n\to\infty} nI_n(t+\varepsilon, \infty) = 0$, for all $\varepsilon > 0$.

Now we will estimate $I_n(t, t + \varepsilon)$. Let $h_n(r, t) = \phi(r) + e^{-s/n}(\phi(r - t) - \phi(r))$. For a fixed value of t, the function $h_n(\cdot, t)$ behaves as ϕ. In particular $\frac{\partial h_n}{\partial r}$ is an increasing function for $r \in [t, t + \varepsilon]$. Then we obtain the following bounds:

$$\frac{\partial h_n}{\partial r}(t, t) \leqslant \frac{\partial h_n}{\partial r}(r, t) \leqslant \frac{\partial h_n}{\partial r}(t + \varepsilon, t),$$

and:

$$\phi''(t + \varepsilon) \leqslant \phi''(r) \leqslant \phi''(t).$$

Hence, we can bound $I_n(t, t + \varepsilon)$ by:

$$
\begin{aligned}
I_n(t, t + \varepsilon) &= \int_t^{t+\varepsilon} \phi''(r)\, (h_n(r, t))^{n-1}\, \frac{\partial h_n}{\partial r}(r, t)\frac{\partial h_n}{\partial r}(r, t)^{-1}\, dr \\
&\leqslant \phi''(t)\frac{\partial h_n}{\partial r}(t, t)^{-1} \int_t^{t+\varepsilon} (h_n(r, t))^{n-1}\, \frac{\partial h_n}{\partial r}(r, t)\, dr \\
&\leqslant \phi''(t)\frac{\partial h_n}{\partial r}(t, t)^{-1}\frac{1}{n}\left[(h_n(t + \varepsilon, t))^n - (h_n(t, t))^n \right].
\end{aligned}
$$

Proceeding similarly, we can find a lower bound:

$$I_n(t, t + \varepsilon) \geqslant \phi''(t + \varepsilon)\frac{\partial h_n}{\partial r}(t + \varepsilon, t)^{-1}\frac{1}{n}\left[(h_n(t + \varepsilon, t))^n - (h_n(t, t))^n \right].$$

Then, for any $\varepsilon > 0$, one can prove the following limits hold:

$$
\begin{aligned}
\lim_{n\to\infty} (h_n(t + \varepsilon, t))^n &= 0, \\
\lim_{n\to\infty} (h_n(t, t))^n &= \exp\left[-s(1 - \phi(t)) \right], \\
\lim_{n\to\infty} \frac{\partial h_n}{\partial r}(t, t) &= \phi'(\varepsilon), \\
\lim_{n\to\infty} \frac{\partial h_n}{\partial r}(t + \varepsilon, t) &= \phi'(0).
\end{aligned}
$$

Replacing these limits in the equations above, we have computed upper and lower bounds of $I_n(t, t + \varepsilon)$. In other words, if the limit of $nI_n(t, t + \varepsilon)$ exists, then it is bounded by:

$$-\frac{\phi''(t + \varepsilon)}{\phi'(0)} \exp\left(-(1 - \phi(t))s \right) \leqslant \lim_{n\to\infty} nI_n(t, t + \varepsilon) \leqslant -\frac{\phi''(t)}{\phi'(\varepsilon)} \exp\left(-(1 - \phi(t))s \right).$$

This is true for any $\varepsilon > 0$; then letting ε tends to 0, we have:

$$\lim_{n\to\infty} nI_n(t) = \frac{\phi''(t)}{\mu} \exp\left(-(1 - \phi(t))s \right).$$

Replacing this limit in equation (2) we obtain

$$\lim_{n\to\infty} \phi_{S_n/n}(s) = \frac{1}{\mu} \int_0^\infty \phi''(t)e^{-(1 - \phi(t))s}\, dt, \tag{3}$$

which will be denoted by $\phi_S(s)$. Although this limit a priori is not necessarily the Laplace transform of a random variable, according to the Continuity theorem (page 431 Ch. XIII in [4]), one has to check that $\lim_{s\to 0} \phi_S(s) = 1$, which can be proved by using the dominated convergence theorem.

A suitable change of variable $y = 1 - \phi(r)$ in equation (3) gives:

$$\phi_S(s) = -\frac{1}{\mu} \int_0^{1-p_0} \frac{\phi''\left(\phi^{-1}(1 - y) \right)}{\phi'\left(\phi^{-1}(1 - y) \right)} e^{-ys}\, dr,$$

where for the integral limits we used the property that $\phi(\infty) = p_0$ (see [4] remark in theorem 1(a) page 439 Ch. XIII). Therefore, we have that:

$$f_S(y) = -\frac{1}{\mu}\frac{\phi''\left(\phi^{-1}(1-y)\right)}{\phi'\left(\phi^{-1}(1-y)\right)}\, 1\!\!1_{[0,1-p_0]}(y)$$

is the probability density of S. $\hfill\square$

As a corollary to this theorem, we can compute the q-th moment and the cumulative distribution function (c.d.f.) of S:

Corollary 2.1 *For any $q \in \mathbb{R}$*

$$E[S^q] = \frac{1}{\mu}\int_0^\infty (1 - \phi(t))^q \phi''(t)\, dt\;,$$

and, for any $x \in [0, 1]$,

$$\mathbb{P}(S \leqslant x) = \left(\frac{1}{\mu}\int_0^{\phi^{-1}(1-x)} \phi''(t)\, dt\right) 1\!\!1_{[0,1-p_0]}(x) + 1\!\!1_{(1-p_0,1]}(x)\;.$$

One could be interested in the cumulative distribution function of S (or more precisely in the survival function), since the move-to-front rule is related to the least-recently-used strategy (see [7] for instance). Indeed, many operating systems or softwares use a memory (also called cache) that could be quickly addressed (think of a web browser, for instance). Hence, one needs to define a strategy to organize it. Let us consider that the cache is made of k files. The least-recently-used strategy is the following: at each unit of discrete time, a file is requested and is moved in front of the cache; if the file was not just previously in the cache, then the last file is deleted from the cache and all other files are shifted by one position to the right; if the file was just previously in the cache, then the file is moved exactly as in the move-to-front rule. So, the move-to-front rule can be viewed as a special case of the least-recently-used strategy for which the length of the cache is equal to the number of files ($k = n$). An important question arises: what is the probability that the requested file is not in the cache? The probability of this event is called the page default; we will denote it by π_k in the sequel. Because of the link between the move-to-front rule and the least-recently-used strategy (as underlined above), we clearly have that $\pi_k = \mathbb{P}(S_n \geqslant k)$. So, if we assume that the cache length is proportional to the number of files, say $k = \alpha n$ with $\alpha \in [0, 1]$ fixed, for a large collection of files, the following approximation holds:

$$\pi_{\alpha n} \simeq \frac{1}{\mu}\int_0^{\phi^{-1}(1-\alpha)} \phi''(t)\, dt$$

if $\alpha < p_0$ and $\pi_{\alpha n} \simeq 1$ otherwise.

3 Examples

In this section, we study some examples for which we are able to do explicitly all computations. We will consider both continuous and discrete distribution for the random weights.

Example 3.1 Suppose that the weights have the Dirac distribution at point mass 1 (in other words, weights are deterministic and are equally requested). Then $\phi(r) = e^{-r}$, the expectation $\mu = 1$ and $p_0 = 0$, we deduce that:

$$f_{S_1}(x) = 1\!\!1_{[0,1]}(x)\;.$$

5

Thus, S_1 has the uniform distribution over $[0,1]$: this result was already proved in (theorem 4.2, p. 198 of [5]). The k-th moment (with $k \in \mathbb{R}_+$) and the c.d.f. of S_1 is:

$$\mathbb{E}[S_1^k] = \frac{1}{k+1} \qquad \text{and} \qquad \forall x \in [0,1], \quad F_{S_1}(x) = \mathbb{P}(S_1 \leqslant x) = x .$$

Example 3.2 Suppose that the weights have the Gamma distribution with parameter $\alpha > 0$. In this example, the random vector $(p_1, \ldots, p_n)$ has the symmetric Dirichlet distribution $D_n(\alpha)$ (see [15] or [9]). In such a case, $p_0 = 0$, $\mu = \alpha$ and $\phi(r) = (1+r)^{-\alpha}$. Computations give:

$$f_{S_2}(x) = \left(1 + \frac{1}{\alpha}\right)(1-x)^{1/\alpha}\, \mathbb{1}_{[0,1]}(x) ,$$

which is the density function of the Beta distribution with parameters $(1, 1+1/\alpha)$. Note that this result has already been proved by [1] with a specific technique using properties of Dirichlet distribution (in this case we were able not only to find the limiting search cost distribution but also the transient search cost distribution for any finite n). The k-th moment (with $k \in \mathbb{R}_+$) of S_2 is:

$$\mathbb{E}[S_2^k] = \frac{\Gamma(k+1)\Gamma(2+\frac{1}{\alpha})}{\Gamma(2+k+\frac{1}{\alpha})} .$$

In particular, we have $\mathbb{E}[S_2] = \frac{\alpha}{2\alpha+1}$ and $\mathrm{Var}[S_2] = \frac{(\alpha+1)\alpha^2}{(3\alpha+1)(2\alpha+1)^2}$. One can also compute the c.d.f of S_2 and, for any $x \in [0,1]$, we get:

$$F_{S_2}(x) = \mathbb{P}(S_2 \leqslant x) = 1 - (1-x)^{1+1/\alpha} .$$

We can easily deduce that, for any $x \in [0,1]$, $\bar{F}_{S_2}(x) \leqslant \bar{F}_{S_1}(x)$, where $\bar{F}_{S_1}(\cdot) = 1 - F_{S_1}(\cdot)$. So we have $S_2 \preceq_{st} S_1$ (where $\preceq_{st}$ denotes the usual stochastic ordering; see [12] or [13], for instance).

Example 3.3 Suppose that the weights have the Geometric distribution on $\mathbb{N}$ with parameter $p \in (0,1)$. In such case, $p_0 = p$, $\mu = (1-p)/p$ and $\phi(r) = p/(1-(1-p)e^{-r})$. Elementary computations give:

$$f_{S_3}(x) = \frac{2(1-x) - p}{1-p}\, \mathbb{1}_{[0,1-p]}(x) .$$

The k-th moment (with $k \in \mathbb{R}_+$) of S_3 is:

$$\mathbb{E}[S_3^k] = \frac{(2+pk)(1-p)^k}{(k+1)(k+2)} .$$

In particular, we have $\mathbb{E}[S_3] = \frac{(2+p)(1-p)}{6}$ and $\mathrm{Var}[S_3] = \frac{(1-p)^2(2+2p-p^2)}{36}$. One can also compute the c.d.f of S_3 and, for any $x \in [0,1]$, get:

$$F_{S_3}(x) = \mathbb{P}(S_3 \leqslant x) = \frac{x(2-p-x)}{1-p}\, \mathbb{1}_{[0,1-p]}(x) + \mathbb{1}_{(1-p,1]}(x) .$$

Hence, from the above expression, one can check that $S_4 \preceq_{st} S_1$.

Example 3.4 Suppose that the weights have the Poisson distribution with parameter λ. In such case, $p_0 = e^{-\lambda}$, $\mu = \lambda$ and $\phi(r) = \exp(\lambda e^{-r} - 1)$. Simple computations give:

$$f_{S_4}(x) = \frac{\ln(1-x) + \lambda + 1}{\lambda}\, \mathbb{1}_{[0,1-e^{-\lambda}]}(x) .$$

Using formula 1.6.5.3 of [11] (page 244), one can compute the k-th moment (with $k \in \mathbb{N}$) of S_4:

$$\mathbb{E}[S_4^k] = \frac{1}{\lambda(k+1)}\left[\lambda + (1-e^{-\lambda})^{k+1} - \sum_{i=1}^{k+1}\frac{(1-e^{-\lambda})^i}{i}\right] .$$

6

In particular, we have $\mathbb{E}[S_4] = \frac{1}{2} - \frac{1-e^{-2\lambda}}{4\lambda}$. One can also compute the c.d.f of S_4 and, for any $x \in [0, 1]$, we get:

$$F_{S_4}(x) = \mathbb{P}(S_4 \leqslant x) = \left(x - \frac{1}{\lambda}(1 - x)\ln(1 - x)\right)\mathbb{1}_{[0,1-e^{-\lambda}]}(x) + \mathbb{1}_{(1-e^{-\lambda},1]}(x) \ .$$

Thus, from the expression above, one can deduce that $S_4 \preceq_{st} S_1$.

From the study of these four examples, one can observe that both S_2, S_3 and S_4 are stochastically smaller than S_1. Hence, the following conjecture looks appealing:

Conjecture 3.1 *Let S be the limiting distribution of the search cost associated to a sequence ω of iid positive random variables. Then, $S \preceq_{st} S_1$ where S_1 is a random distribution having the uniform distribution on the unit interval.*

This conjecture is compatible with some remarks in [2], more precisely with proposition 3.1 therein. Indeed, if the conjecture is right, then as a consequence we have $\mathbb{E}[S] \leqslant \mathbb{E}[S_1] = \frac{1}{2}$. And this is precisely what is stated in proposition 3.1. This conjecture can be interpreted as follows: the case with Dirac weights corresponds to the worst case. Despite our conjecture seems to be true, its proof seems to be difficult.

References

[1] J. Barrera, T. Huillet, and C. Paroissin, *Size-biased permutation of dirichlet partitions and search-cost distribution*, Probab. Engrg. Inform. Sci. **19** (2005), 83–97.

[2] J. Barrera and C. Paroissin, *On the distribution of the stationary search cost for the move-to-front with random weights*, J. Appl. Prob. **41** (2004), no. 1, 250–262.

[3] P. Donnelly, *The heaps process, libraries, and size-biased permutations*, J. Appl. Probab. **28** (1991), no. 2, 321–335.

[4] W. Feller, *An introduction to probability theory and its aplications*, vol. II, John Wiley, London, 1971.

[5] J. Fill, *Limits and rates of convergence for the distribution of search cost under the move-to-front rule*, Theoret. Comput. Sci. **1-2** (1996), 185–206.

[6] J.A. Fill and L. Holst, *On the distribution of search cost for the move-to-front rule*, Random Structures Algorithms **8** (1996), no. 3, 179–186.

[7] P. Flajolet, D. Gardy, and L. Thimonier, *Birthday paradox, coupon collectors, caching algorithms and self-organizing search*, Discrete Appl. Math. **39** (1992), no. 3, 207–229.

[8] J.F.C. Kingman, *Random discrete distributions*, J. R. Statist. Soc. **B37** (1975), 1–22.

[9] S. Kotz, N. Balakrishnan, and N.L. Johnson, *Continuous multivariate distributions, volume 1: Models and applications*, John Wiley, London, 2000.

[10] J. McCabe, *On serial files with relocatable records*, Operat. Res. **13** (1965), 609–618.

[11] A.P. Prudnikov, Y.A. Brychkov, and O.I. Marichev, *Integrals and series*, vol. 1: elementary functions, Gordon and Breach Science Publishers, New-York, 1992.

[12] M. Shaked and J.G. Shanthikumar, *Stochastic orders and their applications*, Academic Press Inc., Boston, 1994.

[13] D. Stoyan, *Comparison methods for queues and other stochastic models*, Wiley, Chichester, 1983.

[14] M.L. Tsetlin, *Finite automata and models of simple forms of behavior*, Russian Math. Surveys **18** (1963), no. 4, 1–27.

[15] S.S. Wilks, *Mathematical statistics*, John Wiley, London, 1962.