Asymptotic Approximation of the Move-To-Front Search Cost Distribution and Least-Recently-Used Caching Fault Probabilities

Predrag R. Jelenković

Bell Laboratories
Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974

ABSTRACT

Consider a finite list of items $n = 1, 2, \ldots, N$, that are requested according to an i.i.d. process. Each time an item is requested it is moved to the front of the list. The associated search cost C^N for accessing an item is equal to its position before being moved. If the request distribution converges to a proper distribution as $N \to \infty$, then the stationary search cost C^N converges in distribution to a limiting search cost C.

We show that, when the (limiting) request distribution has a heavy tail (e.g., generalized Zipf's law) $\mathbb{P}[R = n] \sim c/n^\alpha$ as $n \to \infty$, $\alpha > 1$, then the limiting stationary search cost distribution $\mathbb{P}[C > n]$, or, equivalently, the Least-Recently-Used (LRU) caching fault probability, satisfies

$$\lim_{n\to\infty} \frac{\mathbb{P}[C > n]}{\mathbb{P}[R > n]} = \left(1 - \frac{1}{\alpha}\right)\left[\Gamma\left(1 - \frac{1}{\alpha}\right)\right]^\alpha \nearrow \quad e^\gamma \text{ as } \quad \alpha \to \infty,$$

where Γ is the Gamma function and γ $(= 0.5772\ldots)$ is Euler's constant.

When the request distribution has a light tail $\mathbb{P}[R = n] \sim ce^{-\lambda n^\beta}$ as $n \to \infty$ $(c, \lambda, \beta > 0)$, then

$$\lim_{n\to\infty} \frac{\mathbb{P}[C_f > n]}{\mathbb{P}[R > n]} = e^\gamma,$$

independently of c, λ, β, where C_f is a fluid approximation of C.

We experimentally demonstrate that the derived asymptotic formulas yield accurate results for lists of finite sizes. This should be contrasted with the exponential computational complexity of Burville and Kingman's (1973) exact expression for finite lists. The results also imply that the fault probability of LRU caching is asymptotically at most a factor e^γ (≈ 1.78) greater than for the optimal static arrangement.

Keywords: Self-organizing lists, move-to-front searching, least-recently-used caching (paging), heavy-tailed distributions, light-tailed distributions.

`Submitted for publication to Annals of Applied Probability.`

May 25, 1997; Revised: May 11, 1998.

Asymptotic Approximation of the Move-To-Front Search Cost Distribution and Least-Recently-Used Caching Fault Probabilities

Predrag R. Jelenković

Bell Laboratories
Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974

1. Introduction

One of the most commonly encountered problems in modern distributed network environment is efficient information retrieval (e.g., the Internet Web searching). As a solution to this problem an entire spectrum of different heuristic dynamically organizing data structures have been proposed. Among the proposed algorithms, the most basic ones are the *Move-To-Front* (MTF) self-organizing searching algorithm and the corresponding *Least-Recently-Used* (LRU) caching scheme. The main objective of this paper is to obtain an analytic asymptotic characterization of the MTF search cost distribution function or, equivalently, the LRU caching fault probabilities.

As sketched in the abstract, the MTF algorithm can be informally described as follows. Assume that there is a finite linear list of items (say $1, 2, 3, \dots, N$, sequentially ordered from first to last) and a sequence of requests for the items of the list. Each time a requested item is found at the nth position in the list it is brought to the first position and items in positions $1, \dots, n-1$ are moved one position down. One performance measure is the search cost function which is defined to be the position of the requested item. The caching scheme that corresponds to MTF is the LRU algorithm. For this scheme, it is assumed that n items are kept in fast memory (cache) and that the remaining $N - n$ items are stored in slow memory. Each time a request for an item is made, fast memory is searched first. If the item is not found there it is brought from slow memory and replaced with the least recently used item in the cache. The performance quantity of interest for this algorithm is the LRU fault probability; i.e., the probability that the requested item is not in the cache. It can be shown that computing the LRU fault probability is the same as computing the MTF search cost distribution (the details on this connection will be discussed in Section 2.1).

The performance analysis of self-organizing data structures (e.g., lists, trees) has a long history. Basic references can be found in Knuth (1973) [22]. In the analysis of self-organizing lists there have been two approaches: combinatorial and probabilistic analysis. For the combinatorial (amortized, competitive) analysis the reader is referred to Bentley and McGeoch (1985) [2] and Sleator and Tarjan (1985) [29]. Recent results and references for this approach can be found in Borodin et al. (1995) [7] and Irani et al. (1996) [18]. In this paper we will concentrate on probabilistic analysis.

Early work on the probabilistic analysis of the MTF rule dates back to McCabe (1965) [23]. He computed the expected value and variance of the cost function for finite lists. In Hendricks (1972) [17], a Markov chain on the state space of all permutation on the elements of the list was analyzed and the stationary distribution was derived. Rivest (1976) [28] shows that in stationarity the transposition rule (search algorithm in which the requested item is moved only one position closer to the front) is more efficient (in a certain sense) than the MTF rule. Bitner (1979) [5] investigated the transient behavior of the expected cost function. The n-step transition probabilities for the underlying Markov chain were derived in Nelson (1977) [24] (see also [27, 11]). Spectral analysis was conducted in Phatarfod (1991) [26] (see also Fill (1996) [11]). A combinatorial expression for the distribution function of the search cost was first derived by Burville and Kingman (1973) [8]. An integral form of the Laplace transform of the search cost distribution function was computed in Flajolet et al. (1992) [14]. The authors derive this result using combinatorial techniques and formal languages. The same result was rederived in Fill (1996) [12] using a Poisson embedding technique. A comprehensive list of references on the probabilistic analysis of the LRU and MTF algorithms can be found in Bodell (1997) [6] (for the early work see also Coffman et al. (1973) [21]).

The principal contribution of this paper is the complete asymptotic characterization, both for the light- and heavy-tailed case (as defined in the abstract), of the MTF search cost distribution and the LRU cache fault probabilities. One of the main mathematical techniques used for the heavy-tailed case is Hardy-Littlewood-Karamata Tauberian theorem for an asymptotic inversion of the Laplace transform. In addition, we develop a novel fluid limit approach for analyzing self-organizing data structures. This approach, combined with a direct Laplace inversion, yields a characterization of the search cost distribution in the case of light tails. The practical implication of these results is that, for a large class of distributions considered in this paper, the LRU caching scheme is asymptotically only $e^{\gamma} \approx 1.78$ times worse than the optimal static arrangement.

The remainder of this paper is organized as follows. Section 2 formally defines the problem and gives a short technical note on existing results. A summary of the main results from the literature on the stationary distribution of the search cost distribution is given in Subsection 2.1. Optimality of MTF and LRU algorithms is discussed in Subsection 2.2. The main asymptotic results, Theorems 3 and 6, are presented in Section 3 and 4, respectively. Extensive simulation experiments that verify the accuracy of the asymptotic approximation formula are presented in Section 5. The paper concludes in Section 6. To simplify the reading process, the majority of the technical proofs are given in Section 7.

2. Problem Definition and Historical Notes

Consider a finite set of items $L = \{1, \dots, N\}$, and a sequence of i.i.d. requests $\{R_t^N, t = 0, 1, \dots\}$ distributed as $q_r = \mathbb{P}[R_t^N = r], 1 \le r \le N$. The dynamics of the MTF algorithm is described as follows. If at time t the requested item r $(R_t^N = r)$ is at the nth position of the list, then it is brought to the first position and items in positions $1, \dots, n-1$ are moved one position down. The process of list updates can be modeled as a Markov chain $\{\sigma_t^N, t \geq 0\}$ on the state space of all list permutations. A search cost process $\{C_t^N, t \geq 0\}$ is defined such that C_t^N represents the position in the list of the item requested at time t. The notation R^N, C^N, σ^N, will be used to denote the random variables that have the same distribution as the stationary distribution of R_t^N, C_t^N, σ_t^N, respectively. Our main objective is to derive a simple asymptotic characterization of the search cost distribution $\mathbb{P}[C^N > n]$.

McCabe (1965) [23] derived the following formula for the expected value of the search cost distribution

$$\mathbb{E}C^N = 1 + 2\sum_{r<k} \frac{q_r q_k}{q_r + q_k}. \tag{2.1}$$

Bitner (1979) [5] obtained the transient expected cost

$$\mathbb{E}C_t^N = 1 + 2\sum_{r<k} \frac{q_r q_k}{q_r + q_k} + \sum_{r<k} \frac{(q_r - q_k)^2(1 - q_r - q_k)^t}{2(q_r + q_k)}. \tag{2.2}$$

The stationary distribution for σ_t^N was first computed by Hendricks (1972) [17]

$$\mathbb{P}[\sigma^N = x] = q_{x(1)} \frac{q_{x(2)}}{1 - q_{x(1)}} \frac{q_{x(3)}}{1 - q_{x(1)} - q_{x(2)}} \cdots \frac{q_{x(N)}}{1 - q_{x(1)} - q_{x(2)} - \cdots - q_{x(N-1)}}, \tag{2.3}$$

where x is a particular permutation (state) of the list, and $x(i)$ is the item at the ith position in the permutation. Spectral analysis of the Markov chain $\{\sigma_t^N, t \geq 0\}$ was performed in Phatarfod (1991) [26]. Phatarfod showed that the eigenvalues of the transition matrix of this Markov chain are the $2^N - N$ distinct numbers $\sum_{i \in S} q_i$, $S \subset \{1, \ldots, N\}$. In the same paper Phatarfod derived the multiplicities of eigenvalues (see also Fill (1996) [11]).

2.1 Stationary Distribution of the Search Cost and the Fault Probability of the Least Recently Used Caching

This subsection presents the results from the literature on the stationary distribution of the MTF search cost and the fault probability for the LRU caching. A formal connection between these two quantities is well known (e.g., see [14, 12]). Arguments in support of this connection are also supplied within this subsection.

Burville and Kingman (1973) [8] derived the following combinatorial formula

$$\mathbb{P}[C^N = n] = \sum_{r=1}^{N} \sum_{a=0}^{n-1} (-1)^{n-1-a} \binom{N-1-a}{n-1-a} \sum_{A:|A|=a, r \notin A} \frac{q_r^2}{(1 - Q_A)}, \tag{2.4}$$

where $Q_A \stackrel{\text{def}}{=} \sum_{r \in A} q_r$. The connection between LRU caching and MTF searching can be demonstrated as follows. Denote by $D(k, N)$ the fault probability in the LRU caching scheme with the cache size k. Then we claim

$$D(k, N) = \mathbb{P}[C^N > k].$$

Here is a simple argument that justifies this claim. We can imagine that k elements in the cache are arranged in increasing order of their last access times. Each time there is a request for an item that is not in the cache, the item is brought to the cache and the last element of the cache is moved to the slow memory. The claim is that the fault probability $D(k, N)$ stays the same if the remaining $N - k$ items in the slow memory are arranged in any specific order. In particular, they can be arranged in increasing order of their last access times. It is clear that the obtained algorithm is the same as the MTF algorithm and that $D(k, N) = \mathbb{P}[C^N > k]$. For those who are still not convinced by the preceding argument one can obtain the expression for $D(k, N)$ directly from equation (2.4) as follows. Compute $\mathbb{P}[C^N \le k] = \sum_{n=1}^{k} \mathbb{P}[C^N = n]$, by using (2.4), interchange the sums with respect to n and a, and use the identity (which can be proved easily by induction on k)

$$\sum_{n=a+1}^{k} (-1)^{n-1-a} \binom{N-1-a}{n-1-a} = (-1)^{k-1-a} \binom{N-2-a}{k-1-a}.$$

This derivation leads to Corollary 5.2 from [14], which, for convenience, is stated here.

$$1 - D(k, N) = \mathbb{P}[C^N \le k] = \sum_{r=1}^{N} \sum_{a=0}^{k-1} (-1)^{k-1-a} \binom{N-2-a}{N-1-k} \sum_{A:|A|=a, r \notin A} \frac{q_r^2}{(1-Q_A)}. \tag{2.5}$$

Unfortunately, except for relatively small N, n, and a, the preceding formulas (2.4) and (2.5) are not suitable for numerical evaluation. This is due to a combinatorial explosion; as pointed out in [14] the evaluation of $\mathbb{P}[C^N > k]$ takes about $N^k/k!$ operations, which for example for $N = 1000$ and $k = 20$ computes to roughly 10^{40} operations. This is clearly infeasible. Furthermore, when $k = bN, 0 < b \le 1$, application of Stirling's formula shows that $N^{bN}/(bN)!$ grows exponentially in N, i.e., (2.5) (or equivalently (2.4)) has exponential complexity.

To alleviate this problem, in Flajolet et al. (1992) [14] a compact integral representation of the Laplace transform of the search cost distribution function was derived. In the same paper Cauchy contour integration was proposed for efficient inversion of the search cost distribution function. Fill (1996) [12] rederived the same result using the Poisson embedding technique. This result reads as

$$\mathbb{E}e^{-sC^N} = e^{-s} \int_{t=0}^{\infty} e^{-t} \left[\sum_{r=1}^{N} \frac{q_r^2}{1 + e^{-s}(e^{q_r t} - 1)} \right] \left[\prod_{r=1}^{N} (1 + e^{-s}(e^{t q_r} - 1)) \right] dt,$$

for any $s > 0$.

In Fill and Holst (1996) [13] the limiting search cost C as $N \to \infty$, was investigated. In order to state this result, choose a probability distribution sequence $\mathbb{P}[R = r] = q_r, 1 \le r < \infty, \sum_{r=1}^{\infty} q_r = 1$. Next, construct a sequence of MTF algorithms with finite number of elements N, whose request probabilities are given as $\mathbb{P}[R_t^N = r] = q_r/q_N^+, 1 \le r \le N$, where $q_N^+ = \sum_{r=1}^{N} q_r$. Then, Fill obtained the following result (Proposition 4.4, [12]; in the same paper, he also considered the case when $q_N^+ \to \infty$ as $N \to \infty$, and showed that appropriately scaled C^N converges to a proper limit).

Theorem 1 *The sequence of search costs C^N converges in distribution to C, as $N \to \infty$, and the Laplace transform of C is given as*

$$\mathbb{E}e^{-sC} = e^{-s} \int_{t=0}^{\infty} \sum_{i=1}^{\infty} q_i^2 e^{-q_i t} \left[\prod_{r: r \neq i} (1 - (1 - e^{-s})(1 - e^{-q_r t})) \right] dt. \tag{2.6}$$

This result is the basis of our further investigation.

2.2 Optimality of MTF and LRU Algorithms

Before starting with the analysis let us give a few known results about the optimality of the MTF and LRU algorithms. Note that without loss of generality we can assume that the request probabilities q_r form a monotonically non-increasing sequence (if this is not the case we can always relabel the items in such a way that the new sequence q_r is non-increasing). Now, it is clear that the optimal algorithm is the one that keeps the items static and ordered in decreasing order of their request probabilities. The search cost of this optimal algorithm is $\mathbb{E}R^N$. Next, by using $q_r/(q_r + q_k) < 1$ in (2.1) we easily derive an upper bound (see pp. 399, [22])

$$\mathbb{E}R^N \leq \mathbb{E}C^N \leq 2\mathbb{E}R^N.$$

Using Hilbert's inequalities in Chung et al. (1988) [9] the upper bound was improved from 2 to $\pi/2$.

Similarly, for the caching system with a cache size n the optimal algorithm keeps n most frequent items in the cache all of the time. Now, $\mathbb{P}[C^N > n]/\mathbb{P}[R^N > n]$ gives the ratio between the LRU fault probability and the fault probability for the optimal static arrangement. A bound on this quantity was obtained by Franaszek and Wagner (1974) (Theorem 3, [15]).

Theorem 2

$$\frac{\mathbb{P}[C^N > n]}{\mathbb{P}[R^N > n]} \leq 1 + n \frac{\mathbb{P}[R^N \leq n]}{1 + (n-1)\mathbb{P}[R^N > n]}$$

for arbitrary $q_1, \ldots, q_N$.

At this point, and for the rest of the paper, we introduce the following customary notation. For any two real functions $a(t)$ and $b(t)$ and fixed $t_0 \in \mathbb{R} \cup \{\infty\}$ we will use $a(t) \sim b(t)$ as $t \to t_0$ to denote $\lim_{t \to t_0} a(t)/b(t) = 1$. Similarly, we say that $a(t) \gtrsim b(t)$ as $t \to t_0$, if $\liminf_{t \to t_0} a(t)/g(t) \geq 1$; $a(t) \lesssim b(t)$ has a complementary definition.

For the case when R^N, C^N converge to proper distributions R, C, respectively, as $N \to \infty$, Theorem 2 implies the following. If $\mathbb{E}R < \infty$, then

$$\frac{\mathbb{P}[C > n]}{\mathbb{P}[R > n]} \lesssim n \quad \text{as} \quad n \to \infty. \tag{2.7}$$

If $\mathbb{P}[R = n] \sim c/n^\alpha, 1 < \alpha < 2$, then

$$\frac{\mathbb{P}[C > n]}{\mathbb{P}[R > n]} \lesssim \frac{n^{2-\alpha}}{c} \quad \text{as} \quad n \to \infty. \tag{2.8}$$

Under the additional tail conditions on $\mathbb{P}[R = n]$, our main results, Theorems 3 and 6, will show that the bounds in (2.7) and (2.8) can be replaced by a constant $e^\gamma \approx 1.78$, where γ is Euler's constant.

3. Heavy Tails

This section presents a straightforward asymptotic characterization of the search cost distribution function $\mathbb{P}[C > n]$ for the case when the request distribution has a heavy (polynomial) tail. The main result is stated in Theorem 3. The primary technique that is used is Karamata's Tauberian/Abelian theory for the asymptotic inversion of the Laplace transforms.

In order to be able to obtain the asymptotic inversion of (2.6) we need to derive a simplified (asymptotic) representation of the infinite sum and infinite products that appear in (2.6). To this end we define the following density function

$$f(t) \stackrel{\text{def}}{=} \sum_{r=1}^{\infty} q_r^2 e^{-q_r t}, \tag{3.1}$$

whose asymptotic behavior is described in the subsequent lemma.

Lemma 1 *Assume that $q_r \sim c/r^\alpha$ as $r \to \infty$, with $\alpha > 1$ and $c > 0$. Then*

$$f(t) \sim \frac{c^{\frac{1}{\alpha}}}{\alpha} \Gamma\left(2 - \frac{1}{\alpha}\right) t^{-2+\frac{1}{\alpha}} \quad \text{as} \quad t \to \infty,$$

where Γ is the Gamma function.

Proof: Given in Section 7.1. ◇

Our next object of investigation is

$$g(t) \stackrel{\text{def}}{=} \sum_{r=1}^{\infty} (1 - e^{-q_r t}). \tag{3.2}$$

$g(t)$ also has a straightforward asymptotic characterization.

Lemma 2 *Assume that $q_r \sim c/r^\alpha$ as $r \to \infty$, with $\alpha > 1$ and $c > 0$. Then*

$$g(t) \sim \Gamma\left(1 - \frac{1}{\alpha}\right) c^{\frac{1}{\alpha}} \; t^{\frac{1}{\alpha}} \quad \text{as} \quad t \to \infty.$$

Proof: Given in Section 7.1. ◇

Actually, a stronger version of Lemma 2 holds. Let A be any set of indexes $A \subset \mathbb{N}$ with cardinality $|A|$. Denote with $g_A(t) = \sum_{\{r: r \notin A\}} (1 - e^{-q_r t})$.

Corollary 1 *Assume that $q_r \sim c/r^\alpha$ as $r \to \infty$, with $\alpha > 1$ and $c > 0$. Then for any fixed finite ℓ,*

$$g_A(t) \sim \Gamma\left(1 - \frac{1}{\alpha}\right) c^{\frac{1}{\alpha}} \ t^{\frac{1}{\alpha}} \quad \text{as} \quad t \to \infty,$$

uniformly in all $A \subset \mathbb{N}$ such that $|A| \le \ell$.

Proof: The proof follows immediately from Lemma 2 and

$$g(t) - \ell \le g_A(t) \le g(t).$$

$\diamond$

The preceding technical results led up to the following main result of this section.

Theorem 3 *Assume that $q_r \sim c/r^\alpha$ as $r \to \infty$, with exponent $\alpha > 1$ and $c > 0$. Then*

$$\mathbb{P}[C > n] \sim \left(1 - \frac{1}{\alpha}\right)\left[\Gamma\left(1 - \frac{1}{\alpha}\right)\right]^\alpha \mathbb{P}[R > n] \quad \text{as} \quad n \to \infty. \tag{3.3}$$

Furthermore, if we denote the constant of proportionality in the preceding equation as $K(\alpha)$, then $K(\alpha)$ is monotonically increasing with

$$\lim_{\alpha \to \infty} K(\alpha) = e^\gamma \approx 1.78107, \quad \lim_{\alpha \downarrow 1} K(\alpha) = 1, \tag{3.4}$$

where γ is the Euler constant, i.e., $\gamma = \lim_{n \to \infty} \sum_{i=1}^n i^{-1} - \log n = 0.5772156649\ldots$.

Remark: This theorem implies that

$$\mathbb{P}[C > n] \lesssim e^\gamma \mathbb{P}[R > n] \quad \text{as} \quad n \to \infty. \tag{3.5}$$

The plot of $K(\alpha)$ is given in Figure 1.
Proof: A rigorous proof of this theorem is given in Section 7.2. $\diamond$
Heuristic sketch of the proof: First, for all sufficiently small s, and all sufficiently large t,

$$\prod_{r:r \ne i} (1 - (1 - e^{-s})(1 - e^{-q_r t})) \approx e^{-s \sum_{r=1}^\infty (1 - e^{-q_r t})}. \tag{3.6}$$

Now, by utilizing the above approximation and Lemmas 1 and 2, one can obtain the following informal approximation for sufficiently small s

$$\begin{aligned} &\int_{t_0}^\infty \sum_{i=1}^\infty q_i^2 e^{-q_i t} \left[\prod_{r:r \ne i} (1 - (1 - e^{-s})(1 - e^{-q_r t}))\right] dt \\ &\approx \int_{t_0}^\infty \frac{c^{\frac{1}{\alpha}}}{\alpha} \Gamma\left(2 - \frac{1}{\alpha}\right) t^{-2 + \frac{1}{\alpha}} e^{-s\Gamma\left(1 - \frac{1}{\alpha}\right) c^{\frac{1}{\alpha}} \ t^{\frac{1}{\alpha}}} dt. \end{aligned} \tag{3.7}$$

By changing the variable of integration to $u = \Gamma\left(1 - \frac{1}{\alpha}\right) c^{\frac{1}{\alpha}} \ t^{\frac{1}{\alpha}}$ in the integral above, we

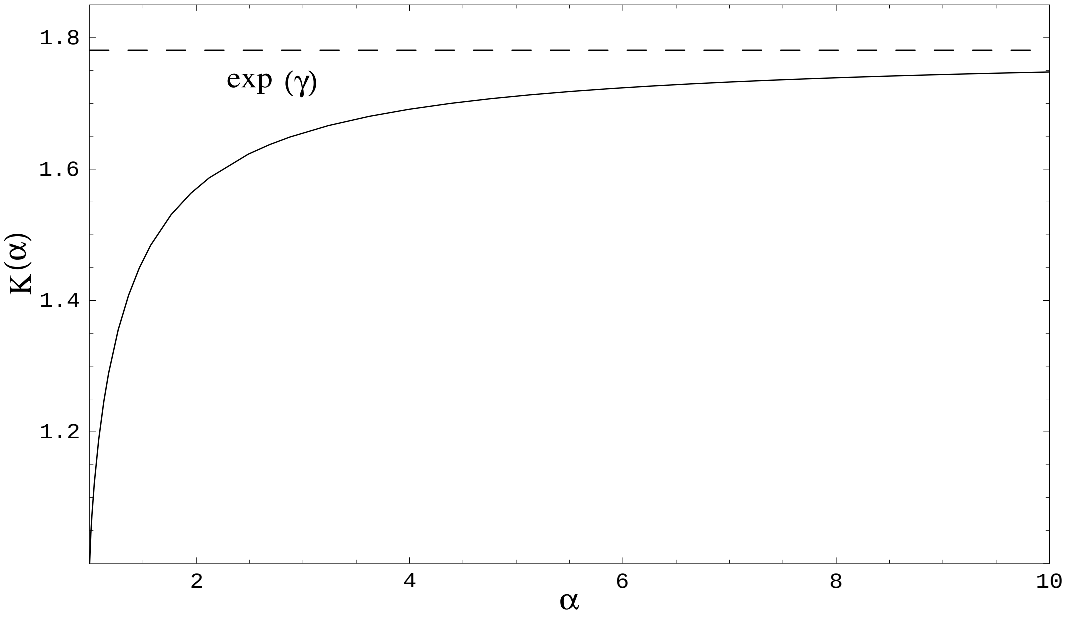

Figure 1: Function $K(\alpha)$.

conclude that the integral in (3.7) is approximately equal to

$$\left(1-\frac{1}{\alpha}\right)\left[\Gamma\left(1-\frac{1}{\alpha}\right)\right]^{\alpha}\int_{u_0}^{\infty}\frac{c}{u^{\alpha}}e^{-su}du,$$

where $u_0=\Gamma\left(1-\frac{1}{\alpha}\right)c^{\frac{1}{\alpha}}\ t_0^{\frac{1}{\alpha}}$. From this, the observation that $\int_{u_0}^{\infty}\frac{c}{u^{\alpha}}e^{-su}du\approx\mathbb{E}\left(e^{-sR}1[R>u_0]\right)$, and the uniqueness of the inverse of the Laplace transform, we can roughly conclude that the tail of C should be proportional to the tail of R with a constant of proportionality given in the equation above. Unfortunately, to make these arguments rigorous much lengthier analysis is required. In particular, one has to investigate the asymptotic behavior of the derivatives of $\mathbb{E}e^{-sC}$. A complete proof that utilizes Karamata's Tauberian/Abelian theorem is provided in Section 7.2. $\diamond$

4. Fluid Limit and Light Tails

In this section in Theorem 4 we develop a fluid limit approximation of the search cost function. For the heavy-tailed case the validity of this approximation is demonstrated in Theorem 5. The main result is presented in Theorem 6. This theorem shows that for the light-tailed case the ratio between the tail of the search cost distribution and the request distribution is asymptotically *invariant* with respect to the shape of the request distribution function.

Consider a density function q on $[0,\infty)$, and the corresponding distribution function $Q(t)=\int_0^t q(u)du$. Assume that the request probabilities are given as $q_r=Q(r)-Q(r-1), r\geq 1$.

Now, construct a sequence of MTF algorithms with request probabilities

$$q_r^k = Q(r/k) - Q((r-1)/k) \qquad k, r \geq 1.$$

For each of the constructed MTF schemes let $C(k)$ be the stationary search cost random variable. Informally, the kth MTF scheme is constructed from the original one by dividing each item r into k items with request probabilities $q_{r+(i/k)}, 1 \leq i \leq k$. In order to compare the derived schemes with the original one we will scale the search cost of the kth scheme as $C(k)/k$. Now, we show that $C(k)/k$ converges in distribution to a proper (fluid) limit as $k \to \infty$. Assume that q is monotonically decreasing and continuous. (Note that this assumption is not restrictive, i.e., for any given monotonically decreasing sequence of request probabilities we can always choose a continuous function q such that $q_r = Q(r) - Q(r-1), r \geq 1$.)

Theorem 4 *The sequence $C(k)/k$ converges in distribution to the fluid limit C_f (finite with probability 1), as $k \to \infty$, whose Laplace transform is given as*

$$\mathbb{E}e^{-sC_f} = \int_0^\infty \left(\int_0^\infty q^2(u) e^{-q(u)t} du \right) e^{-s\int_0^\infty (1-e^{-q(u)t})du} dt \qquad s > 0. \tag{4.1}$$

Proof: Given in Section 7.3. $\diamond$

We term C_f the fluid limit approximation of C (recall that C_f is obtained by dividing each item into infinitely many smaller items, i.e., the items become divisible like fluid). The accuracy of this fluid approximation is demonstrated in the following theorem. The theorem shows that in the heavy-tailed context the fluid limit search cost distribution behaves asymptotically the same as the original search cost distribution.

Theorem 5 *If $q(u) \sim c/u^\alpha, \alpha > 1, c > 0$, then*

$$\mathbb{P}[C > n] \sim \mathbb{P}[C_f > n] \text{ as } n \to \infty.$$

In order to investigate the asymptotic behavior of the distribution function of C_f we define (in the same spirit as in (3.1) and (3.2)) the following functions:

$$f_f(t) \stackrel{\text{def}}{=} \int_0^\infty q^2(u) e^{-q(u)t} du \qquad g_f(t) \stackrel{\text{def}}{=} \int_0^\infty (1 - e^{-q(u)t}) du, \tag{4.2}$$

where the subscript f refers to the fluid limit.
Proof of Theorem 5: Similarly to Lemmas 1 and 2, one can show that

$$f_f(t) \sim \frac{c^{\frac{1}{\alpha}}}{\alpha} \Gamma\left(2 - \frac{1}{\alpha}\right) t^{-2+\frac{1}{\alpha}} \quad \text{as} \quad t \to \infty,$$

and

$$g_f(t) \sim \Gamma\left(1 - \frac{1}{\alpha}\right) c^{\frac{1}{\alpha}} \; t^{\frac{1}{\alpha}} \quad \text{as} \quad t \to \infty.$$

By using exactly the same procedure as in the proof of Theorem 3 one can complete the proof of this theorem. In order to avoid duplications we omit this derivation. $\diamond$

Theorem 6 **(e^γ law)** *If $q(u) \sim ce^{-\lambda u^\beta}$ as $u \to \infty$, for some positive constants c, λ, β, then*

$$\mathbb{P}[C_f > n] \sim e^\gamma \mathbb{P}[R > n] \text{ as } n \to \infty,$$

for **any** *(fixed) choice of the parameters c, λ, β.*

Remarks: **(i)** Note that $q(u) \sim ce^{-\lambda u^\beta}$ is a large class of distributions, containing most of the well known light-tailed distributions, e.g., Weibull, exponential, and Normal. **(ii)** Observe that polynomial ($\sim c/u^\alpha$) and Weibull distributions ($\sim ce^{-\lambda u^\beta}, 0 < \beta < 1$) belong to the class of "subexponential" distributions. In the queueing context these distributions demonstrate the same asymptotic behavior (see Jelenković et al. [20]). **(iii)** Based on the discussion in Section 2.2 we conclude that, under the conditions of this theorem and Theorem 3, the LRU fault probability is only a factor $e^\gamma \approx 1.78$ larger than for the optimal static setup.

Lemma 3 *If $q(u) \sim ce^{-\lambda u^\beta}$ as $u \to \infty$, where $(c, \lambda, \beta) > 0$, then the first derivative of g_f behaves asymptotically as*

$$g_f'(t) \sim \frac{(\log ct)^{\beta^{-1}-1}}{t\beta\lambda^{1/\beta}} \quad \text{as} \quad t \to \infty.$$

Proof: Given in Section 7.4. ◇

Lemma 4 *If $q(u) \sim ce^{-\lambda u^\beta}$ as $u \to \infty$, for any $(c, \lambda, \beta) > 0$, then f_f behaves asymptotically as*

$$f_f(t) \sim \frac{(\log ct)^{\beta^{-1}-1}}{t^2\beta\lambda^{1/\beta}} \quad \text{as} \quad t \to \infty.$$

Proof: Given in Section 7.4. ◇

Corollary 2 *If $q(u) \sim ce^{-\lambda u^\beta}$ as $u \to \infty$, for any $(c, \lambda, \beta) > 0$, then*

$$\frac{f_f(t)}{g_f'(t)} \sim \frac{1}{t} \quad \text{as} \quad t \to \infty.$$

Proof: Follows directly from Lemmas 3 and 4. ◇

Lemma 5 *For any $d > -1$, and $t > 0$,*

$$\int_0^t \frac{1 - e^{-x}}{x} (\log(t/x))^d dx - \frac{(\log(t))^{d+1}}{d+1} \sim \gamma(\log(t))^d \text{ as } t \to \infty.$$

Proof: Given in Section 7.4. ◇

Lemma 6 *If $q(u) \sim ce^{-\lambda u^\beta}$ as $u \to \infty$, $(c, \lambda, \beta) > 0$, then the inverse of $g_f(t)$ behaves asymptotically as*

$$g_f^{-1}(v) \sim e^{-\gamma} c^{-1} e^{\lambda v^\beta} \quad \text{as} \quad v \to \infty.$$

Proof: For any $\delta > 0$ we can choose u_0 such that for all $u > u_0$, $c(1-\delta)e^{-\lambda u^\beta} \le q(u) \le c(1+\delta)e^{-\lambda u^\beta}$; let $c_\delta = (1+\delta)c$. Then

$$g_f(t) \le \int_0^\infty (1 - e^{-c_\delta t e^{-\lambda u^\beta}})du + u_0.$$

Now, by changing the variable of integration to $x = c_\delta t e^{-\lambda u^\beta}$ in the integral above, we compute

$$g_f(t) \le \frac{1}{\beta \lambda^{1/\beta}} \int_0^{c_\delta t} \frac{1-e^{-x}}{x} (\log(c_\delta t/x))^{1/\beta - 1} dx + u_0. \tag{4.3}$$

Here, by choosing $d = 1/\beta - 1$ and applying Lemma 5 in (4.3) we obtain

$$g_f(t) \le \lambda^{-1/\beta} (\log c_\delta t)^{1/\beta} + \frac{\gamma}{\beta\lambda^{1/\beta}} (\log c_\delta t)^{1/\beta - 1}(1+o(1)) \quad \text{as} \quad t \to \infty. \tag{4.4}$$

Consequently, by using the above expression we compute

$$\begin{aligned} g_f(e^{-\gamma} c_\delta^{-1} e^{\lambda u^\beta}) &\le \lambda^{-1/\beta}(\lambda u^\beta - \gamma)^{1/\beta} + \frac{\gamma}{\beta\lambda^{1/\beta}}(\lambda u^\beta - \gamma)^{1/\beta - 1}(1+o(1)) \quad \text{as} \quad u \to \infty \\ &= u\left(1 - \frac{\gamma}{\beta\lambda u^\beta}(1+o(1))\right) + \frac{\gamma}{\beta \lambda u^{\beta-1}}(1+o(1)) \quad \text{as} \quad u \to \infty \\ &= u\left(1 + o\left(\frac{1}{u^\beta}\right)\right) \quad \text{as} \quad u \to \infty. \end{aligned} \tag{4.5}$$

If we introduce a new variable $v \equiv v(u) = u(1 + o(\frac{1}{u^\beta}))$ as $u \to \infty$, then

$$\begin{aligned} v^\beta \left(\frac{v}{u} - 1\right) &= v^\beta o\left(\frac{1}{u^\beta}\right) = o(1) \quad \text{as} \quad u \to \infty \\ &\Rightarrow\; u = v\left(1 + o\left(\frac{1}{v^\beta}\right)\right) \quad \text{as} \quad v \to \infty. \end{aligned}$$

Finally, by replacing the preceding expression in (4.5) it directly follows that

$$\begin{aligned} g_f^{-1}(v) &\ge e^{-\gamma} c_\delta^{-1} e^{\lambda(v(1+o(v^{-\beta})))^\beta} \quad \text{as} \quad v \to \infty \\ &= e^{-\gamma} c_\delta^{-1} e^{\lambda(v^\beta + o(1))} \quad \text{as} \quad v \to \infty, \end{aligned}$$

which implies that

$$\liminf_{v\to\infty} g_f^{-1}(v) e^{\gamma} c e^{-\lambda v^\beta} \ge (1+\delta)^{-1}.$$

Finally, by passing $\delta \to 0$ we obtain the lower bound of the proof of the lemma. To prove the upper bound we use

$$g_f(t) \ge \int_0^\infty (1 - e^{-c_\delta t e^{-\lambda u^\beta}})du - u_0,$$

where $c_\delta = (1-\delta)c$; by repeating exactly the same arguments as for the lower bound we obtain

$$\limsup_{u\to\infty} g_f^{-1}(v) e^\gamma c e^{-\lambda v^\beta} \le 1.$$

This finishes the proof of the lemma. $\diamond$

Finally, we are ready to supply the proof of the theorem.

Proof of Theorem 6: By changing the variable of integration in Theorem 4 to $v = g_f(t)$ we obtain

$$\mathbb{E}e^{-sC_f} = \int_0^\infty \frac{f_f(g_f^{-1}(v))}{g_f'(g_f^{-1}(v))} e^{-sv} dv.$$

Thus, by the uniqueness of the Laplace transform inverse, we conclude that the density of C_f is equal to

$$q_f(v) = \frac{f_f(g_f^{-1}(v))}{g_f'(g_f^{-1}(v))}.$$

Now by using Corollary 2 we derive

$$q_f(v) \sim \frac{1}{g_f^{-1}(v)} \quad \text{as} \quad v \to \infty,$$

which by application of Lemma 6 yields

$$q_f(v) \sim e^\gamma c e^{-\lambda v^\beta} \quad \text{as} \quad v \to \infty.$$

This concludes the proof of the theorem. $\diamond$

5. Simulation Experiments

In this section we illustrate our main results (Theorems 3 and 6) with several simulation examples. Note that the asymptotic results were obtained first by passing the list size N to infinity in Theorem 1, then by investigating the tail of the limiting search cost distribution as n goes to infinity (or, equivalently, the LRU cache fault probability as the cache size n grows). Thus, it can be expected that the asymptotic expressions from Theorems 3 and 6 will give a reasonable approximation for $\mathbb{P}[C^N > n]$ when both n and N are large and N is significantly larger than n. However, it is surprising how accurately these approximations work for relatively *small* values of N and *almost all* values of $n \le N$.

The experiments were conducted on a modern multiprocessor Silicon Graphics computer. We have used C++ programming language with a standard 48-bit pseudo random number generator. The initial position of the items in the list was chosen uniformly at random. In each experiment, before we have conducted the measurements, we have allowed a certain amount of time τ for the system to reach its steady state. In general, we have adopted a heuristic for choosing τ such that $|\mathbb{E}C_\tau^N - \mathbb{E}C^N|/\mathbb{E}C^N < 1\%$ (recall that $\mathbb{E}C_\tau^N$ is given by equation (2.2)). Typically, τ was smaller than 10^6 time units; the only exception was the first experiment where the convergence to stationarity was very slow and we had to choose $\tau = 10^8$ to achieve $|\mathbb{E}C_\tau^N - \mathbb{E}C^N|/\mathbb{E}C^N < 5\%$. Then, after waiting τ units of time, in every experiment we have measured the search cost density between $10^8 - 10^{10}$ time units. The measured data is presented in the remainder of this section.

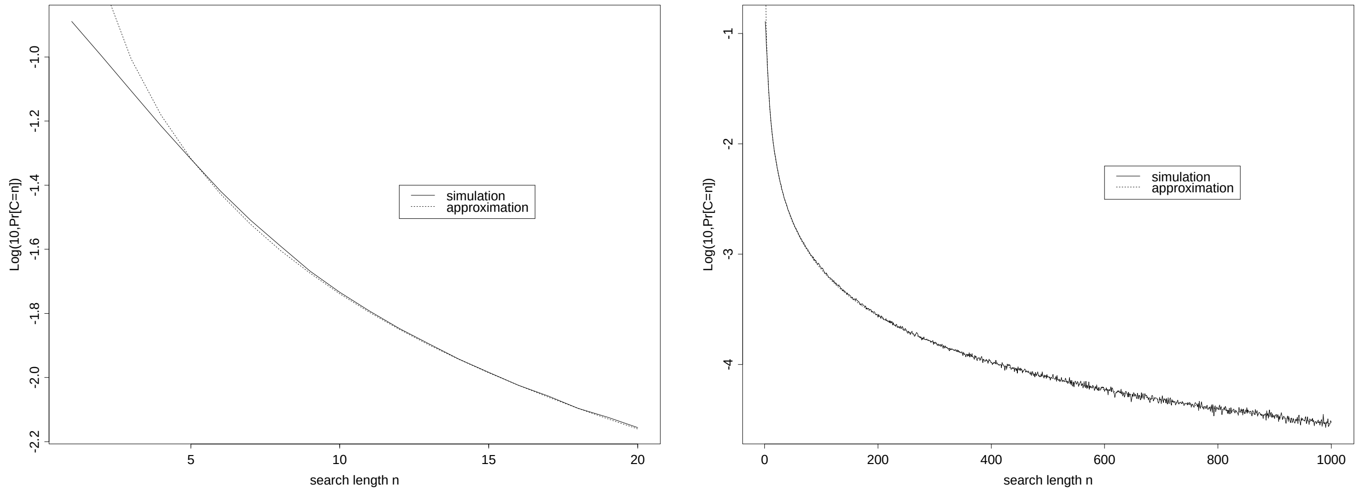

Figure 2: Illustration for Example 1.

Example 1 (heavy tails) In this example we will illustrate the heavy-tailed case from Theorem 3, i.e., the case when the request distribution obeys a generalized Zipf's law, $\mathbb{P}[R^N = n] = c_N/n^\alpha, 1 \leq n \leq N$. In this case we will use the approximation $\mathbb{P}[C^N = n] \approx \frac{K(\alpha)c_N}{n^\alpha}$.

In the first experiment we considered $\alpha = 1.4, N = 10^6$. The search cost was measured for 10^8 time units. The expected value of the search cost is bigger than $\mathbb{E}C^N \geq \mathbb{E}R^N \approx 2100$, from which the expected number of item look ups is bigger than 10^{11}. Hence, it took more than *3 days!* on a modern high speed computer to complete this simulation. The simulation results are displayed with a solid line in Figure 2. The left-hand side of the figure represents a zoomed in view for small values of C^N. The right-hand side plot in the same figure represents a zoomed out view of the same experiment. On the other hand, it is needless to say that it takes negligible computer time to evaluate the normalization constant for the Zipf's law distribution $c_N = 0.322004$, and the asymptotic proportionality constant $K(\alpha) = 1.42362$, which together yield the approximation of the search cost density $K(\alpha)c_N/n^\alpha = 0.4584139/n^\alpha$. The plot of the approximation is represented on the same figure with a dashed line. From the figure we can see that the approximation converges very fast to the actual distribution, i.e., it already becomes accurate for $\mathbb{P}[C^N = 5]$. Hence, the approximation is *almost identical*, except for $n < 5$, to the simulation results.
Similar experiments are repeated for $\alpha = 3, N = 10^3$, and $\alpha = 4, N = 100$; the measurements were 10^9 and 10^{10} time units long, respectively. The results are displayed on the left-hand and right-hand side of Figure 3, respectively. The corresponding approximations are presented with dashed lines on the same figure. The accuracy of the approximation is apparent.

Example 2 (light tails) In this example we illustrate the light-tailed search distribution behavior that was asymptotically characterized by Theorem 6. As suggested by the asymptotic result for all the experiments we use the same asymptotic approximation $\mathbb{P}[C^N = n] \approx$

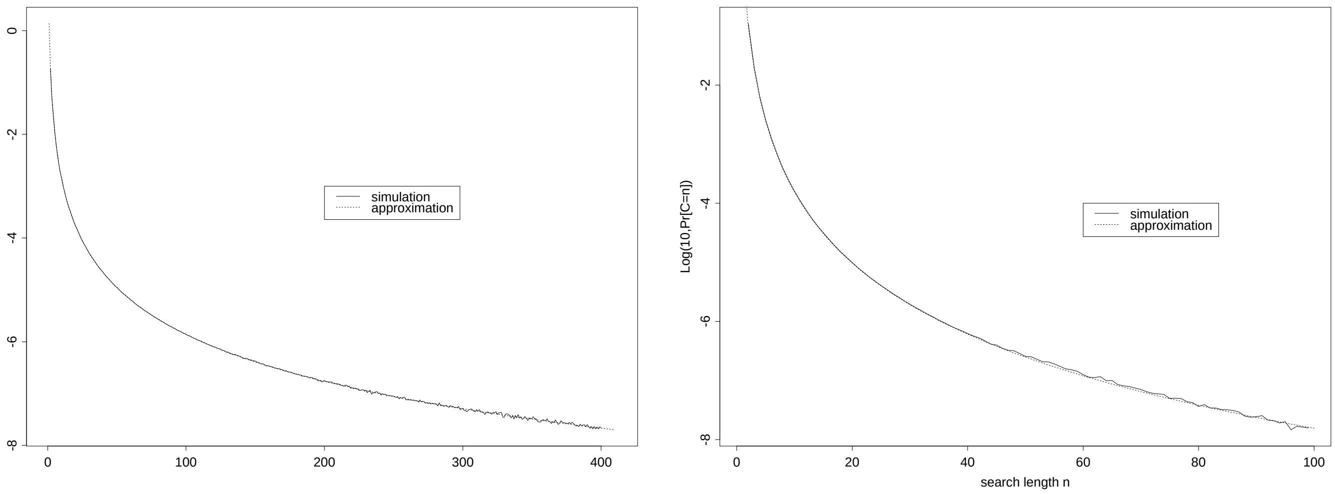

Figure 3: Illustration for Example 1.

$e^{\gamma}\mathbb{P}[R^N = n]$. Figure 4 contains the geometric (exponential) case for four different values of the geometric distribution parameter $p = 0.75, 0.8, 0.85, 0.9$; the observation interval was 10^9 time units for all the experiments. Again, the agreement between the approximation and the simulation results is evident.

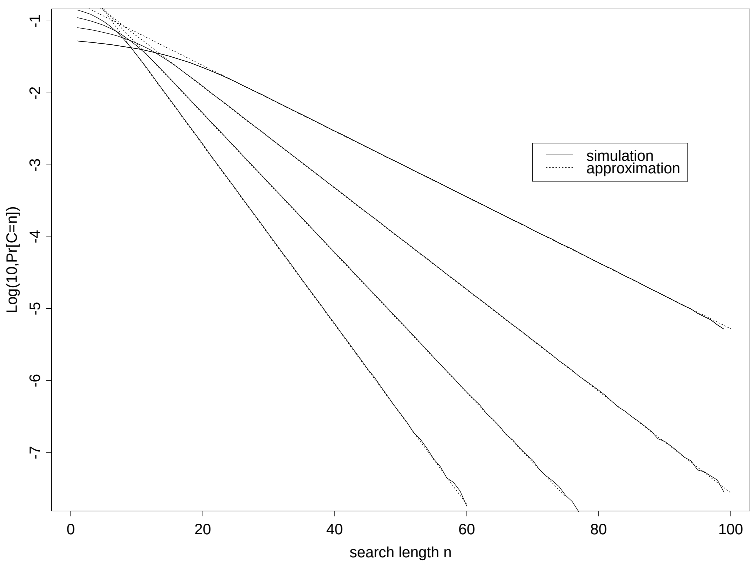

Figure 4: Illustration for Example 2.

Similar precision was observed for the distribution that has a Weibull tail $P[R^N = n] = c_N e^{-\sqrt{n}}$, $N = 1000$; see the left hand side of Figure 5. The experiment for a Normal-like tail $P[R^N = n] = c_N e^{-0.005n^2}$, $N = 100$ is presented on the right hand-side of the same figure. The measurements were conducted for 10^9 and 10^{10} time units, respectively.

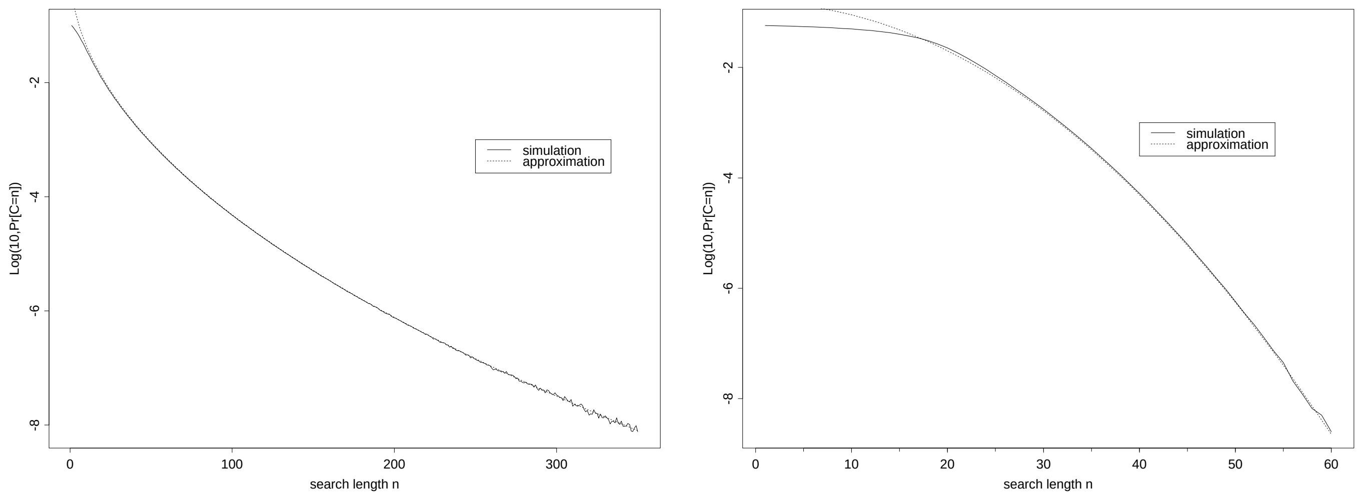

Figure 5: Illustration for Example 2.

6. Conclusion

In this paper we obtained a complete asymptotic characterization of the MTF search cost distribution function or, equivalently, the LRU caching fault probability, for both heavy and light tails. In both cases the tail of the MTF search cost distribution is asymptotically directly proportional to the tail of the request distribution with an explicitly computable constant of proportionality.

In the heavy-tailed (polynomial) case, the constant is a function of the polynomial exponent. As the tail becomes lighter, the constant increases to $e^{\gamma} \approx 1.78$, where γ is Euler's constant. In the light-tailed case the constant is *invariant* with respect to the request distribution shape and is always equal to e^{γ}.

We expect that the new asymptotic techniques developed in this paper will be useful for the analysis of more complex data structures.

Acknowledgments

I wish to thank my colleague Alan Weiss for reading and commenting on the first draft of this paper. I am also grateful to the anonymous reviewer for his/her detailed proofreading.

7. Proofs

7.1 Proofs of Lemma 1 and 2

Proof of Lemma 1: Let us first prove the case $q_r = c/r^\alpha$. Observe that for $t > 0$, the function $(c^2/r^{2\alpha})\, e^{-ct/r^\alpha}$ is increasing in r for $r < (ct/2)^{1/\alpha}$, it is decreasing for $r > (ct/2)^{1/\alpha}$, and it has its maximum $4e^{-2}/t^2$ for $r = (ct/2)^{1/\alpha}$. Using these observations we can obtain the following set of equations. Let $l(t) \stackrel{\text{def}}{=} \lfloor (ct/2)^{1/\alpha} \rfloor$, where $\lfloor x \rfloor$ represents the integer part of x. Then

$$\begin{aligned}
f(t) &\le \sum_{r=1}^{l(t)-1} \left(c^2/r^{2\alpha}\right) e^{-ct/r^\alpha} + \sum_{r=l(t)+1}^{\infty} \left(c^2/r^{2\alpha}\right) e^{-ct/r^\alpha} + 4e^{-2}/t^2 \\
&\le \sum_{r=1}^{l(t)-1} \int_r^{r+1} \left(c^2/u^{2\alpha}\right) e^{-ct/u^\alpha} du + \sum_{r=l(t)+1}^{\infty} \int_{r-1}^{r} \left(c^2/u^{2\alpha}\right) e^{-ct/u^\alpha} du + 4e^{-2}/t^2 \\
&\le \int_1^\infty \left(c^2/u^{2\alpha}\right) e^{-ct/u^\alpha} du + 4e^{-2}/t^2 \\
&\le \int_0^\infty \left(c^2/u^{2\alpha}\right) e^{-ct/u^\alpha} du + 4e^{-2}/t^2.
\end{aligned} \tag{7.1}$$

By changing variables to $v = ct/u^\alpha$ in the integral in (7.1), we compute

$$\int_0^\infty \left(c^2/u^{2\alpha}\right) e^{-ct/u^\alpha} du = \frac{c^{\frac{1}{\alpha}}}{\alpha} \Gamma\left(2 - \frac{1}{\alpha}\right)\ t^{-2+\frac{1}{\alpha}}. \tag{7.2}$$

Finally, by substituting (7.2) in (7.1), and using $t^{-2}/t^{\alpha^{-1}-2} = o(1)$ as $t \to \infty$, we obtain

$$\limsup_{t\to\infty} f(t) t^{2-\frac{1}{\alpha}} \le \frac{c^{\frac{1}{\alpha}}}{\alpha} \Gamma\left(2 - \frac{1}{\alpha}\right) \tag{7.3}$$

Similarly, one can derive the lower bound

$$\begin{aligned}
f(t) &\ge \sum_{r=1}^{l(t)} \left(c^2/r^{2\alpha}\right) e^{-ct/r^\alpha} + \sum_{r=l(t)+1}^{\infty} \left(c^2/r^{2\alpha}\right) e^{-ct/r^\alpha} \\
&\ge \int_0^\infty \left(c^2/u^{2\alpha}\right) e^{-ct/u^\alpha} du - 4e^{-2}/t^2
\end{aligned} \tag{7.4}$$

Thus, by replacing (7.2) in (7.4), and by taking the limit infimum with respect to t we conclude

$$\liminf_{t\to\infty} f(t) t^{2-\frac{1}{\alpha}} \ge \frac{c^{\frac{1}{\alpha}}}{\alpha} \Gamma\left(2 - \frac{1}{\alpha}\right). \tag{7.5}$$

Now, the combination of (7.4) and (7.3) yields the proof for the case $q_r = c/r^\alpha$.

For the general case ($q_r \sim c/r^\alpha$), for any $c > \epsilon > 0$ we can choose $r_0 > 0$ such that, for all $r \geq r_0$, $-\epsilon < q_r r^\alpha - c < \epsilon$. Using this we obtain

$$\begin{aligned} f(t) &\leq \frac{r_0 4e^{-2}}{t^2} + \sum_{r=r_0+1}^{\infty} \frac{(c+\epsilon)^2}{r^{2\alpha}} e^{-(c-\epsilon)t/r^\alpha} \\ &\leq \frac{r_0 4e^{-2}}{t^2} + \frac{(c+\epsilon)^2}{(c-\epsilon)^2} \sum_{r=1}^{\infty} \frac{(c-\epsilon)^2}{r^{2\alpha}} e^{-(c-\epsilon)t/r^\alpha} \end{aligned} \tag{7.6}$$

Consequently, by applying what we already have proved in (7.3-7.6) we arrive at

$$\limsup_{t\to\infty} f(t) t^{2-\frac{1}{\alpha}} \leq \frac{(c+\epsilon)^2}{(c-\epsilon)^2} \frac{(c-\epsilon)^{\frac{1}{\alpha}}}{\alpha} \Gamma\left(2 - \frac{1}{\alpha}\right). \tag{7.7}$$

Finally, by passing $\epsilon \to 0$, we prove the extension of (7.3). Similarly, starting with

$$f(t) \geq -\frac{r_0 4e^{-2}}{t^2} + \sum_{r=1}^{\infty} \frac{(c-\epsilon)^2}{r^{2\alpha}} e^{-(c+\epsilon)t/r^\alpha}$$

we derive the analog of (7.5). This completes the proof of the lemma. $\diamond$

Proof of Lemma 2: Similarly to the proof of Lemma 1, let us first consider the case $q_r = c/r^\alpha$. Note that for $t > 0$, $1 - e^{-ct/r^\alpha}$ is monotonically decreasing in r. Then,

$$\begin{aligned} g(t) &= \sum_{r=1}^{\infty} \int_r^{r+1} (1 - e^{-ct/r^\alpha}) du \\ &\geq \sum_{r=1}^{\infty} \int_r^{r+1} (1 - e^{-ct/u^\alpha}) du \\ &= \int_1^{\infty} (1 - e^{-ct/u^\alpha}) du \\ &= \frac{1}{\alpha} c^{\frac{1}{\alpha}} t^{\frac{1}{\alpha}} \int_0^{ct} (1 - e^{-v}) v^{-\frac{1}{\alpha}-1} dv \\ &= 1 - e^{-ct} + c^{\frac{1}{\alpha}} t^{\frac{1}{\alpha}} \int_0^{ct} v^{-\frac{1}{\alpha}} e^{-v} dv \end{aligned} \tag{7.8}$$

where in the last equality we have used integration by parts with $U = 1 - e^{-v}, dU = e^{-v}dv, V = -\alpha v^{-\frac{1}{\alpha}}, dV = v^{-\frac{1}{\alpha}-1}dv$. Since $\int_0^{ct} v^{-\frac{1}{\alpha}} e^{-v} dv \to \Gamma\left(1 - \frac{1}{\alpha}\right)$ as $t \to \infty$, from (7.8) it follows that

$$\liminf_{t\to\infty} \frac{g(t)}{t^{\frac{1}{\alpha}}} \geq \Gamma\left(1 - \frac{1}{\alpha}\right) c^{\frac{1}{\alpha}}. \tag{7.9}$$

Similarly,

$$
\begin{aligned}
g(t)-(1-e^{-ct}) &= \sum_{r=2}^{\infty}\int_{r-1}^{r}(1-e^{-ct/r^{\alpha}})du \\
&\leq \sum_{r=2}^{\infty}\int_{r-1}^{r}(1-e^{-ct/u^{\alpha}})du \\
&= \int_{1}^{\infty}(1-e^{-ct/u^{\alpha}})du. \qquad (7.10)
\end{aligned}
$$

Finally, by replacing (7.8) in (7.10) we prove the upper bound, i.e.,

$$
\limsup_{t\to\infty}\frac{g(t)}{t^{\frac{1}{\alpha}}} \leq \Gamma\left(1-\frac{1}{\alpha}\right)c^{\frac{1}{\alpha}}, \qquad (7.11)
$$

which together with (7.9) finishes the proof for the case $q_r = c/r^{\alpha}$.

To prove the general case ($q_r \sim c/r^{\alpha}$), for any $0<\epsilon<c$, we choose $r_0>0$, such that for all $r\geq r_0$, $-\epsilon < q_r r^{\alpha} - c < \epsilon$. Since, $1-e^{-ct/r^{\alpha}}$ is monotonically increasing in c and bounded above by 1, it follows that

$$
-r_0+\sum_{r=1}^{\infty}(1-e^{-(c-\epsilon)t/r^{\alpha}}) \leq g(t) \leq r_0 + \sum_{r=1}^{\infty}(1-e^{-(c+\epsilon)t/r^{\alpha}}),
$$

which, by applying (7.8), (7.9), (7.10), (7.11), and by passing $\epsilon\to 0$, implies the conclusion of the lemma. $\diamond$

7.2 Proof of Theorem 3

As we have already mentioned the proof of this result is based on Karamata's Tauberian/Abelian theorem for distribution functions of regular variation. This theorem relates the tail behavior of a distribution function to the asymptotic behavior of its Laplace transform at the origin. For convenience we state the following result (Theorem 7) which is a weaker version of the theorem due to Bingham and Doney [3] ([4], pp. 333).

Let F be a distribution function on $[0,\infty)$, and let $\tilde{F}$ be its Laplace-Stieltjes transform. Denote by $\mu_m = \mathbb{E}X^m = \int_{[0,\infty)} x^m dF(x), m = 0,1,\ldots$. When $\mu_m<\infty$, $\tilde{F}(s)$ may be expanded in a Taylor series as far as s^m term:

$$
\tilde{F}(s) = \sum_{k=0}^{m}\mu_k(-s)^k/k! + o(s^m) \quad \text{as} \quad s\downarrow 0.
$$

To compare the tail behavior of F with the behavior of $\tilde{F}$ at the origin, one needs to eliminate the Taylor polynomial $\sum_{k=0}^{m}\mu_k(-s)^k/k!$. This may be done by subtraction or repeated differentiation. The following theorem has a wide application in probability, e.g., some recent applications to queueing can be found in [19].

Theorem 7 *Let $m\in\mathbb{N}_0$, and $\alpha = m+\beta$.*

(i) If $0 < \beta < 1$, then the following two asymptotic relations are equivalent:

$$(-1)^{m+1}\tilde{F}^{(m+1)}(s) \sim \alpha\Gamma[1-\beta]\frac{c}{s^{1-\beta}} \text{ as } s \downarrow 0, \tag{7.12}$$

$$1 - F(x) \sim \frac{c}{x^{\alpha}} \text{ as } x \to \infty, \tag{7.13}$$

where Γ stands for the gamma function and $\tilde{F}^{(m+1)}$ denotes the $(m+1)$st derivative of $\tilde{F}$.

(ii) If $\beta = 1$, then (7.13) is equivalent to

$$(-1)^{m+1}\tilde{F}^{(m+1)}(s) \sim -\alpha c \log s \text{ as } s \downarrow 0. \tag{7.14}$$

Proof: (i) is just a special case of Theorem 8.1.6, pp. 333, [4].

(ii) For $\alpha = m+1$ the implication (7.13) $\Rightarrow$ (7.14) can be proved easily by direct evaluation of $(-1)^{m+1}\tilde{F}^{(m+1)}(s)$; we skip the details.

For the reverse implication (7.14) $\Rightarrow$ (7.13), we have that, by Theorem 8.1.6, pp. 333, [4], (7.14) implies

$$\int_0^x t^{m+1} dF(t) \sim (m+1)c \log x \quad \text{as} \quad x \to \infty.$$

By changing the variables $u = t^{m+1}$, $y = x^{m+1}$, in the integral above, and by defining $F_1(y) \stackrel{\text{def}}{=} F(y^{1/(m+1)})$ we get that

$$\int_0^y u dF_1(u) \sim c \log y \quad \text{as} \quad y \to \infty. \tag{7.15}$$

Now, by the remark after Corollary 8.1.7, pp. 335, [4], (7.15) is equivalent to

$$1 - F_1(y) \sim \frac{c}{y} \quad \text{as} \quad y \to \infty,$$

which by $F_1(x^{m+1}) \equiv F(x)$ implies (7.13). This finishes the proof. $\diamond$

For any set of indices $A = \{i_1, \ldots, i_k\}$ let us define

$$\Pi_A(s,t) \stackrel{\text{def}}{=} \prod_{r: r \notin A} (1 - (1 - e^{-s})(1 - e^{-q_r t}));$$

when A is the empty set we denote $\Pi_A(s,t)$ simply as $\Pi(s,t)$. Let $|A|$ denote the cardinality of A.

Lemma 7 *For any $\epsilon > 0$ and any set of indices $A, |A| \le \ell < \infty$, there exist $s_0 > 0, t_0 < \infty$ such that for all $0 < s < s_0$, and $t > t_0$,*

$$e^{-s(1+\epsilon)c_1 t^{\frac{1}{\alpha}}} \le \Pi_A(s,t) \le e^{-s(1-\epsilon)c_1 t^{\frac{1}{\alpha}}},$$

where $c_1 = \Gamma(1 - (1/\alpha))c^{1/\alpha}$.

Proof: First let us observe that for any $|A| \leq \ell$

$$\Pi(s,t) \leq \Pi_A(s,t) \leq e^{s\ell}\Pi(s,t). \tag{7.16}$$

From this we see that for all sufficiently small s (and fixed ℓ) $\Pi(s,t)$ uniformly approximates $\Pi_A(s,t)$. Therefore to finish the proof it is enough to prove that the lemma is satisfied for $\Pi(s,t)$. Now

$$\begin{aligned}\log \Pi(s,t) &= \sum_{r=1}^{\infty} \log(1-(1-e^{-s})(1-e^{q_r t})) \\ &\leq -(1-e^{-s})\sum_{r=1}^{\infty}(1-e^{q_r t}),\end{aligned} \tag{7.17}$$

where in the inequality above we use the inequality $\log(1+x) \leq x$. Here, by applying Lemma 2 in (7.17), and observing that $(1-e^{-s}) \sim s$, as $s \to 0$ we finish the proof of the upper bound, i.e.,

$$\Pi(s,t) \leq e^{-s(1-\epsilon)c_1 t^{\frac{1}{\alpha}}} \tag{7.18}$$

Similarly, to prove the lower bound we can use the inequality $x - x^2 \leq \log(1+x)$, $x \in [-0.683, 0]$. Therefore, by choosing s sufficiently small such that $1 - e^{-s} \leq 0.683$ we obtain

$$\begin{aligned}\log \Pi(s,t) &\geq -(1-e^{-s})\sum_{r=1}^{\infty}(1-e^{q_r t}) \\ &\qquad -(1-e^{-s})^2\sum_{r=1}^{\infty}(1-e^{q_r t})^2 \\ &\geq -(1-e^{-s})(2-e^{-s})\sum_{r=1}^{\infty}(1-e^{q_r t}).\end{aligned} \tag{7.19}$$

Clearly, for any $\epsilon > 0$ we can choose sufficiently small s such that $(1-e^{-s})(2-e^{-s}) \leq (1+\epsilon)s$. When this is replaced in (7.19) and by application of Lemma 2 we obtain the lower bound inequality of the lemma. This finishes the proof. $\diamond$

Lemma 8 *For any $\epsilon > 0$, for all $i \geq 1$, and for any fixed $m \geq 0$, there exist $s_0 > 0, t_0 < \infty$ such that for all $0 < s < s_0$, and $t > t_0$,*

$$(1-\epsilon)(c_1 t^{\frac{1}{\alpha}})^m e^{-s(1+\epsilon)c_1 t^{\frac{1}{\alpha}}} \leq (-1)^m \frac{\partial^m}{\partial s^m} e^{-s}\Pi_i(s,t) \leq (1+\epsilon)(c_1 t^{\frac{1}{\alpha}})^m e^{-s(1-\epsilon)c_1 t^{\frac{1}{\alpha}}},$$

where $c_1 = \Gamma(1-(1/\alpha))c^{1/\alpha}$ and $\Pi_i(s,t) \equiv \Pi_{\{i\}}(s,t)$.

Proof: Let us first investigate the form of the mth derivative of $e^{-s}\Pi_i(s,t)$. For convenience of notation for any set $A = \{i_1, \ldots, i_k\}$ we denote with $\Pi_{i_1,\ldots,i_k}(s,t) \equiv \Pi_A(s,t)$; also,

$\sum_{i_k/i_{k-1},\dots,i_1} \equiv \sum_{i_k:i_k\notin\{i_{k-1},\dots,i_1\}}$. For $m=1$ simple algebra gives

$$\frac{\partial}{\partial s}\left[e^{-s}\Pi_i(s,t)\right] = -e^{-s}\Pi_i - e^{-2s}\sum_{k_1/i}(1-e^{-q_{k_1}t})\Pi_{k_1 i}.$$

Similarly, for $m=2$

$$\frac{\partial^2}{\partial s^2}\left[e^{-s}\Pi_i(s,t)\right] = e^{-s}\Pi_i + 3e^{-2s}\sum_{k_1/i}(1-e^{-q_{k_1}t})\Pi_{k_1 i} + e^{-3s}\sum_{k_1/i}(1-e^{-q_{k_1}t})\sum_{k_2/k_1,i}(1-e^{-q_{k_2}t})\Pi_{k_1k_2i}.$$

Following this derivation, one can easily prove the following claim. We skip the details.

Claim 1 *For any $m\geq 1$ there exist a set of nonnegative integers $d_1^m,\dots,d_{m-1}^m$, such that*

$$\begin{aligned}\frac{\partial^m}{\partial s^m}\left[e^{-s}\Pi_i(s,t)\right] \;=\; (-1)^m \Bigg[& e^{-s}\Pi_i + d_1^m e^{-2s}\sum_{k_1/i}(1-e^{-q_{k_1}t})\Pi_{k_1 i} + \cdots \\ & + d_{m-1}^m e^{-ms}\sum_{k_1/i}(1-e^{-q_{k_1}t})\sum_{k_2/k_1,i}(1-e^{-q_{k_2}t})\cdots\sum_{k_{m-1}/k_1,\dots,k_{m-2},i}(1-e^{-q_{k_{m-1}}t})\Pi_{k_1\cdots k_{m-1}i} \\ & + e^{-(m+1)s}\sum_{k_1/i}(1-e^{-q_{k_1}t})\sum_{k_2/k_1,i}(1-e^{-q_{k_2}t})\cdots\sum_{k_m/k_1,\dots,k_{m-1},i}(1-e^{-q_{k_m}t})\Pi_{k_1\cdots k_m i}\Bigg].\end{aligned} \tag{7.20}$$

Finally, by applying Lemma 7 and Corollary 1 we finish the proof of this lemma. $\diamond$

Here, we are ready to **finish the proof of Theorem 3**: Observe that if $\mathbb{P}[R=n]\sim c/n^\alpha$, $\alpha>1$, then $\mathbb{P}[R>n]\sim(c/(\alpha-1))n^{-\alpha+1}$, as $n\to\infty$.

Let us first consider the case $\alpha=m+\beta$, $0<\beta<1$, $m\in\mathbb{N}$. Note that we can take the mth derivative of $\mathbb{E}e^{-sC}$ by interchanging the order of differentiation and integration/summation. Justification for this interchange follows by Dominated Convergence Theorem and the following bound.

Lemma 9 *For any (fixed) $s>0$, integer $\ell\geq 0$, there exist $h_0>0$, $\delta\equiv\delta(s)>0$, $\theta\equiv\theta(s)$, such that for all $A=\{i_1,\dots,i_k\}$, $|A|\leq\ell$, $0<|h|<h_0$,*

$$\left|\frac{\Pi_A(s,t)-\Pi_A(s+h,t)}{h}\right| \leq \theta t^{\frac{1}{\alpha}}e^{-\delta t^{\frac{1}{\alpha}}}. \tag{7.21}$$

Proof of this lemma is given at the end of this section.

Thus, by taking the nth derivative of $\mathbb{E}e^{-sC}$, and applying Lemma 1 and Lemma 8, we can choose, for any $\epsilon>0$, $t_0<\infty$, $s_0>0$ such that for all $0<s<s_0$, and $t>t_0$

$$\begin{aligned}(-1)^m\frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC} \;&\geq\; (1-\epsilon)\int_{t_0}^\infty \frac{c^{\frac{1}{\alpha}}}{\alpha}\Gamma\left(2-\frac{1}{\alpha}\right)t^{-2+\frac{1}{\alpha}}\left(c_1t^{\frac{1}{\alpha}}\right)^m e^{-s(1+\epsilon)c_1t^{\frac{1}{\alpha}}}dt \\ &=\; (1-\epsilon)c_1^n c_2\int_{t_0}^\infty t^{-2+\frac{m+1}{\alpha}}e^{-s(1+\epsilon)c_1t^{\frac{1}{\alpha}}}dt,\end{aligned} \tag{7.22}$$

where $c_1 = \Gamma(1-1/\alpha)c^{1/\alpha}$ and $c_2 = \frac{c^{\frac{1}{\alpha}}}{\alpha}\Gamma\left(2-\frac{1}{\alpha}\right)$; now, with the change of variables $u = s(1+\epsilon)c_1 t^{\frac{1}{\alpha}}$ we arrive at

$$\begin{aligned}(-1)^m \frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC} &\geq (1-\epsilon)c_2 c_1^m \alpha (s(1+\epsilon)c_1)^{\beta-1}\int_{s(1+\epsilon)c_1 t_0^{\frac{1}{\alpha}}}^{\infty} u^{-\beta}e^{-u}du \\ &\gtrsim (1-\epsilon)c_2 c_1^m \alpha (s(1+\epsilon)c_1)^{\beta-1}\Gamma(1-\beta) \text{ as } s\downarrow 0. \end{aligned} \tag{7.23}$$

Finally by taking the limit with respect to ϵ we get

$$\liminf_{s\downarrow 0}\left\{(-1)^m\left[\frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC}\right]s^{m+1-\alpha}\right\} \geq c_1^{\alpha-1}c_2\alpha\Gamma(m+1-\alpha). \tag{7.24}$$

For the upper bound we use

$$\begin{aligned}(-1)^m \frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC} &\leq \eta(t_0) + (1+\epsilon)c_1^m c_2 \int_{t_0}^{\infty} t^{-2+\frac{m+1}{\alpha}} e^{-s(1-\epsilon)c_1 t^{\frac{1}{\alpha}}}dt \\ &\lesssim (1+\epsilon)c_2 c_1^m \alpha\Gamma(m+1-\alpha)((1-\epsilon)c_1 s)^{\alpha-m-1} \text{ as } s\downarrow 0, \end{aligned} \tag{7.25}$$

where $\eta(t_0)$ is a sufficiently large constant. Now, by letting $\epsilon \to 0$ in (7.25) and combining it with (7.24) we obtain

$$\lim_{s\downarrow 0}\left\{(-1)^m\left[\frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC}\right]s^{m+1-\alpha}\right\} = c_1^{\alpha-1}c_2\alpha\Gamma(1-\beta). \tag{7.26}$$

In conclusion, by applying Karamata's Theorem 7 **(i)** we derive

$$\mathbb{P}[C>n] \sim \frac{\alpha c_1^{\alpha-1}c_2}{c}\frac{c}{\alpha-1}n^{-\alpha+1} \sim \frac{\alpha c_1^{\alpha-1}c_2}{c}\mathbb{P}[R>n] \quad \text{as} \quad n\to\infty,$$

which by replacing c_1 and c_2 yields the proof of the case $\alpha = m+\beta, 0<\beta<1$.

For integer $\alpha = m+1$ by combining the same reasoning as in (7.23) and (7.25) one can easily obtain

$$\begin{aligned}\frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC} &\sim (-1)^m c_1^m c_2(m+1)\int_{sc_1 t_0^{1/\alpha}}^{\infty} u^{-1}e^{-u}du \\ &\sim (-1)^m c_1^m c_2 (m+1)\log(1/s) \quad \text{as} \quad s\downarrow 0. \end{aligned} \tag{7.27}$$

Finally by applying Theorem 7 **(ii)** we obtain the proof for integer α and conclude the proof of the expression (3.3) of the theorem.

At this point, we are going to prove that $K(\alpha)$ is monotonically increasing in α for $\alpha > 1$, with its limitis at 1 and ∞ given by (3.4). Observe that

$$\begin{aligned}K(\alpha) &= \left(1-\frac{1}{\alpha}\right)\left[\Gamma\left(1-\frac{1}{\alpha}\right)\right]^{\alpha} \\ &= \left(1-\frac{1}{\alpha}\right)^{-\alpha+1}\left[\Gamma\left(2-\frac{1}{\alpha}\right)\right]^{\alpha}, \end{aligned} \tag{7.28}$$

where in the last equality we have used the identity $\Gamma(x+1) = x\Gamma(x)$. Next, the monotonicity of $K(\alpha)$ will follow if we prove that

$$\log K(\alpha) = -(\alpha - 1)\log\left(1 - \frac{1}{\alpha}\right) + \alpha \log \Gamma\left(2 - \frac{1}{\alpha}\right) \tag{7.29}$$

is monotonically increasing for $\alpha > 1$. By taking a derivative in (7.29) we arrive at

$$\frac{d}{d\alpha}\log K(\alpha) = \frac{-1}{\alpha} - \log\left(1 - \frac{1}{\alpha}\right) + \log \Gamma\left(2 - \frac{1}{\alpha}\right) + \frac{1}{\alpha}\psi^{(0)}\left(2 - \frac{1}{\alpha}\right), \tag{7.30}$$

where $\psi^{(k)}, k = 0, 1, \ldots,$ are Polygamma functions (see equation 6.4.1, pp. 260, [1]). Furthermore, since $\psi^{(0)}(1) = -\gamma$ (Euler's constant) and $\Gamma(1) = 1$, by passing $\alpha \to \infty$ in (7.30) we conclude

$$\frac{d}{d\alpha}\log K(\alpha) \to 0 \quad \text{as} \quad \alpha \to \infty. \tag{7.31}$$

Fortunately, the second derivative of $\log K(\alpha)$ takes the following relatively simple form

$$\frac{d^2}{d\alpha^2}\log K(\alpha) = -\frac{1}{(\alpha - 1)\alpha^2} + \frac{1}{\alpha^3}\psi^{(1)}\left(2 - \frac{1}{\alpha}\right). \tag{7.32}$$

Now, we intend to prove that (7.32) is negative for all $\alpha > 1$; for this we use the following claim.

Claim 2 *For any $\alpha \geq 1$,*

$$\psi^{(1)}\left(2 - \frac{1}{\alpha}\right) \leq \frac{4\alpha^2}{(2\alpha - 1)^2}. \tag{7.33}$$

Proof: Note that (7.33) is equivalent to

$$\psi^{(1)}(z) \leq \frac{4}{z^2}, \tag{7.34}$$

for all $1 \leq z < 2$. By using the integral representation given in equation 6.4.1, pp. 260, [1], of $\psi^{(1)}(z)$ we arrive at

$$\psi^{(1)}(z) = \int_0^\infty \frac{te^{-zt}}{1 - e^{-t}}dt, \tag{7.35}$$

which by the change of variable $t = u/z$ is equivalent to

$$\psi^{(1)}(z) = \frac{1}{z^2}\int_0^\infty \frac{ue^{-u}}{1 - e^{-u/z}}du. \tag{7.36}$$

For $1 \le z \le 2$ the integral in (7.36) is bounded by

$$\int_0^\infty \frac{ue^{-u}}{1-e^{-u/2}}du \le \frac{20}{7}\int_0^1 e^{-u}du + \frac{20}{7}\int_1^\infty ue^{-u}du = \frac{20}{7}(1+e^{-1}) < 4,$$

since $1-e^{-u/2}$ is monotonically increasing and $1-e^{-u/2} > 7u/20$, for $0 \le u \le 1$. This finishes the proof of Claim 2.

Next, by replacing (7.33) in (7.32) we arrive at

$$\begin{aligned}\frac{d^2}{d\alpha^2}\log K(\alpha) &\le \frac{1}{\alpha}\left(-\frac{1}{(\alpha-1)\alpha} + \frac{4}{(2\alpha-1)^2}\right)\\ &= \frac{-1}{\alpha^2(\alpha-1)(2\alpha-1)^2} < 0,\end{aligned}$$

for $\alpha > 1$. This implies that $d(\log K(\alpha))/d\alpha$ is monotonically decreasing (for $\alpha > 1$), which in combination with (7.31) yields

$$\frac{d}{d\alpha}\log K(\alpha) \ge 0 \quad \text{for} \quad \alpha > 1.$$

Thus, $\log K(\alpha)$ is monotonically increasing for $\alpha > 1$, and therefore the same holds for $K(\alpha)$.

The limits in (3.4) follow by straightforward application of equation 6.1.34, pp. 256, [1]. This concludes the proof of the theorem. $\diamond$

Proof of Lemma 9: Let $1 > h > 0$. Then, by applying elementary inequalities we obtain

$$\begin{aligned}0 < 1 - \frac{\Pi_A(s+h,t)}{\Pi_A(s,t)} &\le 1 - \frac{\Pi(s+h,t)}{\Pi(s,t)}\\ &= 1 - e^{-\log\frac{\Pi(s,t)}{\Pi(s+h,t)}}\\ &\le \log\frac{\Pi(s,t)}{\Pi(s+h,t)} &(7.37)\\ &= \sum_{r=1}^{\infty}\log\left(\frac{1-(1-e^{-s})(1-e^{-q_r t})}{1-(1-e^{-s-h})(1-e^{-q_r t})}\right)\\ &\le \sum_{r=1}^{\infty}\log\left(\frac{1-(1-e^{-s})(1-e^{-q_r t})}{1-(1-e^{-s})(1-e^{-q_r t})-he^{-s}(1-e^{-q_r t})}\right) &(7.38)\\ &= -\sum_{r=1}^{\infty}\log\left(1-\frac{he^{-s}(1-e^{-q_r t})}{1-(1-e^{-s})(1-e^{-q_r t})}\right)\\ &\le -\sum_{r=1}^{\infty}\log(1-h(1-e^{-q_r t})), &(7.39)\end{aligned}$$

where in (7.37), (7.38) we have used $1-e^{-x} \le x, x \ge 0$, and in (7.39) we applied $1-(1-e^{-s})(1-e^{-q_r t}) \ge e^{-s}$. By using the inequality $-\log(1-x) \le 2x$ for $0 \le x \le 0.79$, in (7.39) we derive

$$0 < 1 - \frac{\Pi_A(s+h,t)}{\Pi_A(s,t)} \leq 2h \sum_{r=1}^{\infty} (1 - e^{-q_r t}), \tag{7.40}$$

for all $0 < h < h_0 = 0.79$. Consequently, combining (7.39) with Lemma 2 yields

$$0 < 1 - \frac{\Pi_A(s+h,t)}{\Pi_A(s,t)} \leq h\theta_1 t^{\frac{1}{\alpha}}, \tag{7.41}$$

for a sufficiently large constant θ_1 and all $0 < h < h_0$. In addition, equations (7.16), (7.17) and Lemma 2 produce for any fixed $s > 0$

$$\Pi_A(s,t) \leq \theta_2 e^{-\delta t^{\frac{1}{\alpha}}}, \tag{7.42}$$

for some finite $\theta_2 \equiv \theta_2(s)$, $\delta \equiv \delta(s) > 0$, and all $A = \{i_1, \ldots, i_k\}, |A| \leq \ell$. Finally, equations (7.41) and (7.42) give the proof of the lemma for $0 < h \leq h_0 = 0.79$ ($\theta = \theta_1\theta_2$). The proof of the lemma when $h < 0$ is completely analogous, and therefore we leave it out. $\diamond$

7.3 Proof of Theorem 4

To prove this result it is enough to show that $\mathbb{E}e^{-sC(k)/k}$ converges to the expression in (4.1) and that $\mathbb{E}e^{-sC_f} \to 1$ as $s \downarrow 0$ (see Theorem 6.6.3, pp. 190, [10]). From equation (2.6) one obtains

$$\begin{aligned}
\mathbb{E}e^{-s\frac{C(k)}{k}} &= e^{-\frac{s}{k}} \int_{t=0}^{\infty} \sum_{i=1}^{\infty} (q_i^k)^2 e^{-q_i^k t} \left[\prod_{r:r\neq i} (1 - (1 - e^{-\frac{s}{k}})(1 - e^{-q_r^k t})) \right] dt \\
&= e^{-\frac{s}{k}} \int_{v=0}^{\infty} \sum_{i=1}^{\infty} k(q_i^k)^2 e^{-kq_i^k v} \left[\prod_{r:r\neq i} (1 - (1 - e^{-\frac{s}{k}})(1 - e^{-kq_r^k v})) \right] dv,
\end{aligned} \tag{7.43}$$

where the last equality follows by the change of variable $v = t/k$. First, we show that for each (fixed) $v \geq 0$

$$\lim_{k\to\infty} \sum_{i=1}^{\infty} k(q_i^k)^2 e^{-kq_i^k v} = \int_0^{\infty} q(u)^2 e^{-q(u)v} du. \tag{7.44}$$

In order to prove (7.44), observe that the monotonicity of q implies

$$\begin{aligned}
\sum_{i=1}^{\infty} k(q_i^k)^2 e^{-kq_i^k v} &\leq \sum_{i=1}^{\infty} \left[q((i-1)/k)^2/k \right] e^{-q(i/k)v} \\
&\leq \frac{q(0)}{k} + \int_0^{\infty} q(u)^2 e^{-q(u+2/k)v} du \\
&\to \int_0^{\infty} q(u)^2 e^{-q(u)v} du \quad \text{as} \quad k \to \infty,
\end{aligned}$$

where the last asymptotic relation follows by Dominated Convergence Theorem. Analogously, we prove the lower bound and finish the proof of (7.44). In a completely similar manner one can prove that for each (fixed) $v \geq 0$, $s > 0$, and uniformly in i

$$\prod_{r:r\neq i}(1-(1-e^{-\frac{s}{k}})(1-e^{-kq_r^k v}))dv \to e^{-s\int_0^\infty (1-e^{-q(u)v})du} \quad \text{as} \quad k\to\infty; \tag{7.45}$$

we omit the details.

Now, by applying Fatou's lemma (pp. 212, [25]) and equations (7.44) and (7.45) we derive

$$\liminf_{k\to\infty} \mathbb{E}e^{-s\frac{C(k)}{k}} \geq \int_0^\infty \left(\int_0^\infty q^2(u)e^{-q(u)t}du\right) e^{-s\int_0^\infty (1-e^{-q(u)t})du}dt.$$

For the upper bound, note that for all i the infinite products in (7.43) are upper bounded by 1, and therefore, for any $v_0 > 0$

$$\begin{aligned}\mathbb{E}e^{-s\frac{C(k)}{k}} &\leq e^{-\frac{s}{k}}\int_{v=0}^{v_0}\sum_{i=1}^\infty k(q_i^k)^2 e^{-kq_i^k v}\left[\prod_{r:r\neq i}(1-(1-e^{-\frac{s}{k}})(1-e^{-kq_r^k v}))\right]dv \\ &\quad + \int_{v_0}^\infty \sum_{i=1}^\infty k(q_i^k)^2 e^{-kq_i^k v};\end{aligned} \tag{7.46}$$

by applying Fubini's theorem in the second integral in (7.46) and using the monotonicity of q we derive

$$\begin{aligned}\int_{v_0}^\infty \sum_{i=1}^\infty k(q_i^k)^2 e^{-kq_i^k v}dv &= \sum_{i=1}^\infty q_i^k e^{-kq_i^k v_0} \\ &\leq \sum_{i=1}^\infty \frac{q((i-1)/k)}{k}e^{-q(i/k)v_0} \\ &\leq \frac{q(0)}{k} + \int_0^\infty q(u)e^{-q(u+2/k)v_0}du \\ &\to \int_0^\infty q(u)e^{-q(u)v_0}du \quad \text{as} \quad k\to\infty;\end{aligned} \tag{7.47}$$

again the last asymptotic relation follows by dominated convergence. By replacing (7.47) in (7.46) and by using the Dominated Convergence Theorem in the first integral in (7.46) we arrive at

$$\limsup_{k\to\infty} \mathbb{E}e^{-s\frac{C(k)}{k}} \leq \int_{v=0}^{v_0}\left(\int_0^\infty q^2(u)e^{-q(u)v}du\right)e^{-s\int_0^\infty (1-e^{-q(u)v})du}dv + \int_0^\infty q(u)e^{-q(u)v_0}du. \tag{7.48}$$

Finally, by passing $v_0 \to \infty$ in (7.48) and by using

$$\int_0^\infty q(u)e^{-q(u)v_0}du \to 0 \quad \text{as} \quad v_0\to\infty$$

we show that $\mathbb{E}e^{-s\frac{C(k)}{k}}$ converges to the desired limit in (4.1) as $k \to \infty$. To complete the proof we need to show that $\mathbb{E}e^{-sC_f} \to 1$ as $s \downarrow 0$. But, this follows by Monotone Convergence Theorem and

$$\int_0^\infty \left(\int_0^\infty q^2(u) e^{-q(u)t} du \right) dt = 1,$$

where the last equality is implied by Fubini's theorem. $\diamond$

7.4 Proofs of Lemma 3, 4, and 5

Proof of Lemma 3: By dominated convergence, from the definition of $g_f(t)$ it follows that

$$g_f'(t) = \int_0^\infty q(u) e^{-q(u)t} du. \tag{7.49}$$

First, let us assume that $q(u) = ce^{-\lambda u^\beta}$. By changing the variables to $x = te^{-\lambda u^\beta}$ in (7.49), we compute

$$g_f'(\frac{t}{c}) = \frac{c}{t\beta\lambda^{1/\beta}} \int_0^t e^{-x} (\log(t/x))^d dx, \tag{7.50}$$

where $d \stackrel{\text{def}}{=} (1/\beta) - 1$ (> -1). Thus, to complete the proof, it is enough to show that for $d > -1$

$$\int_0^t e^{-x} (\log(t/x))^d dx \sim (\log t)^d \quad \text{as} \quad t \to \infty. \tag{7.51}$$

To finish this let us decompose the integral above into three integrals:

$$\int_0^t e^{-x} (\log(t/x))^d dx = \int_0^{1/\log t} + \int_{1/\log t}^{\log t} + \int_{\log t}^t \stackrel{\text{def}}{=} I_1(t) + I_2(t) + I_3(t). \tag{7.52}$$

Let us first investigate the asymptotic behavior of $I_2(t)$. Assume first that $d \geq 0$. Then

$$I_2(t) \leq (\log(t \log t))^d \int_{1/\log t}^{\log t} e^{-x} dx \sim (\log t)^d \tag{7.53}$$

as $t \to \infty$; similarly for the lower bound

$$I_2(t) \geq \left(\log \left(\frac{t}{\log t} \right) \right)^d \int_{1/\log t}^{\log t} e^{-x} dx \sim (\log t)^d \tag{7.54}$$

as $t \to \infty$. For $0 > d > -1$ the inequalities in (7.53) and (7.54) will hold with the inequalities being reversed. Thus, we have proved that

$$I_2(t) \sim (\log t)^d \quad \text{as} \quad t \to \infty. \tag{7.55}$$

For $I_1(t)$ we have the following set of estimates:

$$\begin{aligned} I_1(t) &\le \int_0^{1/\log t} (\log(t/x))^d dx \\ &= t\int_{\log(t\log t)}^{\infty} u^d e^{-u} du \\ &\sim t(\log(t\log t))^d \frac{1}{t\log t} = o((\log t))^d \quad \text{as} \quad t\to\infty, \end{aligned} \tag{7.56}$$

where the last asymptotic relation follows from equation (6.5.32), pp. 263, [1]. Similarly, one can easily prove that

$$I_3(t) = o((\log t)^d) \quad \text{as} \quad t\to\infty. \tag{7.57}$$

Finally, by combining equations (7.55-7.57) we conclude the proof of the case $q(u) = ce^{-\lambda u^\beta}$.

For the general case $q(u) \sim ce^{-\lambda u^\beta}$ as $u\to\infty$, for any $\epsilon > 0$ we can choose u_0, such that for all $u > u_0$, $(1-\epsilon)ce^{\lambda u^\beta} \le q(u) \le (1+\epsilon)ce^{\lambda u^\beta}$. Using this in conjunction with the inequality $xe^{-xt} \le (1/t)e^{-1}, x\ge 0$, and the case $q(u) = ce^{-\lambda u^\beta}$, we obtain

$$\begin{aligned} g_f'(t) &\le u_0 e^{-1}/t + \frac{1+\epsilon}{1-\epsilon}\int_0^\infty (1-\epsilon)ce^{-\lambda u^\beta} e^{-t(1-\epsilon)ce^{-\lambda u^\beta}} du \\ &\sim \frac{1+\epsilon}{1-\epsilon}\frac{(\log(1-\epsilon)ct)^{1/\beta-1}}{t\beta\lambda^{1/\beta}} \quad \text{as} \quad t\to\infty. \end{aligned}$$

Similarly, one can obtain the lower bound. Finally by passing $\epsilon\to 0$ we obtain the conclusion of the theorem. $\diamond$

Proof of Lemma 4: By assuming that $q(u) = ce^{-\lambda u^\beta}$, and by changing the variable of integration to $x = te^{-\lambda u^\beta}$ in (7.49), as in the proof of Lemma 3, we compute

$$f_f(\frac{t}{c}) = \frac{c^\lambda}{t^2\beta\lambda^{1/\beta}}\int_0^t xe^{-x}(\log(t/x))^d dx, \tag{7.58}$$

where $d \stackrel{\text{def}}{=} (1/\beta) - 1$ (> -1). Thus, to complete the proof it is enough to show that for $d > -1$

$$\int_0^t xe^{-x}(\log(t/x))^d dx \sim (\log t)^d \quad \text{as} \quad t\to\infty. \tag{7.59}$$

To finish this let us decompose the integral above into three integrals:

$$\int_0^t xe^{-x}(\log(t/x))^d dx = \int_0^{1/\log t} + \int_{1/\log t}^{\log t} + \int_{\log t}^t \stackrel{\text{def}}{=} I_1(t) + I_2(t) + I_3(t). \tag{7.60}$$

The arguments from here are exactly the same as in the proof of Lemma 3. We skip the details. $\diamond$

Proof of Lemma 5: It is easy to compute that

$$\begin{aligned}\int_0^t \frac{1-e^{-x}}{x}(\log(t/x))^d dx - \frac{(\log(t))^{d+1}}{d+1} &= \int_0^1 \frac{1-e^{-x}}{x}(\log(t/x))^d dx - \int_1^t \frac{e^{-x}}{x}(\log(t/x))^d dx \\ &\stackrel{\text{def}}{=} I_1(t) - I_2(t). \end{aligned} \tag{7.61}$$

By changing the variable of integration to $u = t/x$ in $I_1(t)$ we obtain

$$I_1(t) = \int_t^\infty \frac{1-e^{-t/u}}{u}(\log(u))^d du,$$

which can be decomposed in

$$\begin{aligned} I_1(t) &= \int_t^{t\log^2 t} \frac{1-e^{-t/u}}{u}(\log(u))^d du + \int_{t\log^2 t}^\infty \frac{1-e^{-t/u}}{u}(\log(u))^d du \\ &\stackrel{\text{def}}{=} I_{11}(t) + I_{12}(t). \end{aligned} \tag{7.62}$$

Consider first the case $d \geq 0$. It is easy to see that

$$\begin{aligned} I_{11}(t) &\leq (\log(t\log^2 t))^d \int_t^{t\log^2 t} \frac{1-e^{-t/u}}{u} du \\ &= (\log(t\log^2 t))^d \int_{1/(\log^2 t)}^1 \frac{1-e^{-x}}{x} dx \\ &\sim (\log(t))^d \int_0^1 \frac{1-e^{-x}}{x} dx \quad \text{as} \quad t \to \infty. \end{aligned} \tag{7.63}$$

Similarly, one gets the lower bound

$$\begin{aligned} I_{11}(t) &\geq (\log(t))^d \int_{1/(t\log^2 t)}^1 \frac{1-e^{-x}}{x} dx \\ &\sim (\log(t))^d \int_0^1 \frac{1-e^{-x}}{x} dx \quad \text{as} \quad t \to \infty. \end{aligned} \tag{7.64}$$

Also, by using the inequality $1 - e^{-x} \leq x$, for $x \geq 0$, we obtain

$$\begin{aligned} I_{12}(t) &\leq t \int_{t\log^2 t}^\infty \frac{1}{u^2}(\log(u))^d du \\ &= t \int_{\log(t\log^2 t)}^\infty x^d e^{-x} dx \\ &\sim t(\log(t\log^2 t))^d e^{-\log(t\log^2 t)} \quad \text{as} \quad t \to \infty \\ &= o((\log(t))^d) \quad \text{as} \quad t \to \infty; \end{aligned} \tag{7.65}$$

the asymptotics in (7.65) follows from [1], pp. 263, equation 6.5.32. Next, we investigate

$$\begin{aligned}
I_2(t) &= \int_1^{\log(t)} \frac{e^{-x}}{x}(\log(t/x))^d dx + \int_{\log t}^{t} \frac{e^{-x}}{x}(\log(t/x))^d dx \\
&\stackrel{\text{def}}{=} I_{21}(t) + I_{22}(t)
\end{aligned} \tag{7.66}$$

Here, the asymptotic behavior of $I_{21}(t)$ is determined by

$$I_{21}(t) \leq (\log(t))^d \int_1^{\log(t)} \frac{e^{-x}}{x} dx \sim (\log(t))^d \int_1^{\infty} \frac{e^{-x}}{x} dx, \tag{7.67}$$

and

$$I_{21}(t) \geq (\log(t/\log(t)))^d \int_1^{\log(t)} \frac{e^{-x}}{x} dx \sim (\log(t))^d \int_1^{\infty} \frac{e^{-x}}{x} dx. \tag{7.68}$$

The estimate for $I_{22}(t)$ is given by

$$I_{22}(t) \leq (\log(t/\log(t)))^d \frac{1}{t \log t} = o(\log(t)^d) \quad \text{as} \quad t \to \infty. \tag{7.69}$$

Finally, by combining equations (7.61-7.69) it follows

$$\begin{aligned}
I_1(t) - I_2(t) &\sim (\log(t))^d \left(- \int_1^{\infty} \frac{e^{-x}}{x} + \int_0^1 \frac{1 - e^{-x}}{x} dx \right) \quad \text{as} \quad t \to \infty \\
&= \gamma (\log(t))^d,
\end{aligned}$$

where the last equality follows from [16], pp. 946, equation 8.367 (12). This finishes the proof for the case $d \geq 0$.

For the case $-1 < d < 0$ the inequalities in (7.63) and (7.64) are going to be reversed, but the asymptotic behavior of I_{11} is still going to be the same. Since (7.65) still holds $I_1(t)$ will behave asymptotically the same as for the $d \geq 0$ case. Similarly, the inequalities in (7.67) and (7.68) are going to be reversed, but the asymptotic behavior of $I_{21}(t)$ is unchanged. The asymptotic upper bound for $I_{22}(t)$ is given by

$$\begin{aligned}
I_{22}(t) &\leq \frac{1}{t} \int_{\log(t)}^{t} \frac{1}{x} (\log(t/x))^d dx \\
&= \frac{1}{t} \int_0^{\log(t/\log t)} u^d du \\
&= \frac{1}{t} \frac{(\log t/\log t)^{d+1}}{d+1} = o((\log t)^d) \quad \text{as} \quad t \to \infty.
\end{aligned}$$

This implies that $I_2(t)$ will have the same asymptotics as in the $d \geq 0$ case, and therefore (7.70) holds. This completes the proof of the lemma. ◇

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions.* Dover Publications, Inc., New York, 1974.

[2] J. L. Bentley and C. C. McGeoch. Amortized analyses of self-organizing sequential search heuristics. *Communications of the ACM*, 28(4):404–411, 1985.

[3] N. H. Bingham and R. A. Doney. Asymptotic properties of super-critical branching processes I: the Galton-Watson process. *Advances in Applied Probability*, 6:711–731, 1974.

[4] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation.* Cambridge University Press, Cambridge, 1987.

[5] J. R. Bitner. Heuristics that monotonically organize data structures. *SIAM J. Comput.*, 8:82–110, 1979.

[6] J. Bodell. *Cost of Searching: Probabilistic Analysis of the Self-Organizing Move-to-Fornt and Move-to-Root Sorting Rules.* PhD thesis, Department of Mathematics, Royal Institute of Technology, Stockholm, April 1997.

[7] W. A. Borodin, S. Irani, P. Raghavan, and B. Schieber. Competitive paging with locality of reference. *Journal of Computer and System Science*, 50(2):244–258, 1995.

[8] P. J. Burville and J. F. C. Kingman. On a model for storage and search. *Journal of Applied Probability*, 10:697–701, 1973.

[9] F. R. K. Chung, D. J. Hajela, and P. D. Seymour. Self-organizing sequential search and hilbert's inequalities. *Journal of Computer and System Sciences*, 36(2):148–157, 1996.

[10] Kai Lai Chung. *A Course in Probability Theory.* Academic Press, Inc., 1974.

[11] J. A. Fill. An exact formula for the move-to-front rule for self-organizing lists. *Journal of Theoretical Probability*, 9(1):113–159, 1996.

[12] J. A. Fill. Limits and rate of convergence for the distribution of search cost under the move-to-front rule. *Theoretical Computer Science*, 164:185–206, 1996.

[13] J. A. Fill and L. Holst. On the distribution of search cost for the move-to-front rule. *Random structures and algorithms*, 8(3):179, 1996.

[14] P. Flajolet, D. Gardy, and L. Thimonier. Birthday paradox, coupon collector, caching algorithms and self-organizing search. *Discrete Applied Mathematics*, 39:207–229, 1992.

[15] P. A. Franaszek and T. J. Wagner. On some distribution-free aspects of paging algorithm performance. *Journal of Applied Probability*, 10:697–701, 1973.

[16] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals Series and Products.* Academic Press, New York and London, 1967.

[17] W. J. Hendricks. The stationary distribution of an interesting Markov chain. *Journal of Applied Probability*, 9:231–233, 1972.

[18] S. Irani, A. R. Karlin, and S. Philips. Strongly competitive algorithms for paging with locality of reference. *SIAM J. Comput.*, 25(3):477–497, June 1996.

[19] P. R. Jelenković. *The Effect of Multiple Time Scales and Subexponentiality on the Behavior of a Broadband Network Multiplexer.* PhD thesis, Graduate School of Arts and Sciences, Columbia University in the City of New York, Center for Telecommunications Research Technical Report 465-96-31, 1996.

[20] P. R. Jelenković and A. A. Lazar. Subexponential asymptotics of a markov-modulated random walk with queueing applications. *Journal of Applied Probability*, 35(2), June 1998, to appear.

[21] E. G. Coffman Jr. and P. J. Denning. *Operating Systems Theory.* Prentice-Hall, 1973.

[22] D. E. Knuth. *The Art of Computer Programming, Vol. 3: Sorting and Searching.* Addison-Wesley, 1973.

[23] J. McCabe. On serial files with relocatable records. *Operations Research*, 13:609–618, 1965.

[24] P. R. Nelson. Single-shelf library-type Markov chains with infinitely many books. *Journal of Applied Probability*, 14:298–308, 1977.

[25] Billingsley P. *Probability and Measure.* John Wiley & Sons, Inc., 1986.

[26] R. M. Phataford. On the matrix occuring in a linear search problem. *Journal of Applied Probability*, 28:336–346, 1991.

[27] R. M. Phataford. On the transition probabilities of the move-to-front scheme. *Journal of Applied Probability*, 31:570–574, 1994.

[28] R. Rivest. On self-organizing sequential search heuristics. *Communications of the ACM*, 19(2):63–67, 1976.

[29] D. D. Sleator and R. E. Tarjan. Amortized efficiency of list update and paging rules. *Communications of the ACM*, 28(2):202–208, 1985.