Improved Algorithms and Analysis for Secretary Problems and Generalizations*

Miklos Ajtai† Nimrod Megiddo‡ Orli Waarts§

July 19, 1995

In the classical secretary problem, n objects from an ordered set arrive in random order, and one has to accept k of them so that the final decision about each object is made only on the basis of its rank relative to the ones already seen. Variants of the problem depend on the goal: either maximize the probability of accepting the best k objects, or minimize the expectation of the sum of the ranks (or powers of ranks) of the accepted objects. The problem and its generalizations are at the core of tasks with a large data set, in which it may be impractical to backtrack and select previous choices.

Optimal algorithms for the special case of $k = 1$ are well known. Partial solutions for the first variant with general k are also known. In contrast, an explicit solution for the second variant with general k has not been known; even the question of whether or not the expected sum of powers of the ranks of selected items tends to infinity with n has been unresolved. We answer the above open questions by obtaining explicit algorithms. For each $z \geq 1$, the resulting expected sum of the zth powers of the ranks of the selected objects is at most[1] $k^{z+1}/(z+1) + C(z) \cdot k^{z+0.5} \log k$, whereas the best possible value at all is $k^{z+1}/(z+1) + O(k^z)$. Our methods are very intuitive and apply to some generalizations. We also derive a lower bound on the trade-off between the probability of selecting the best object and the expected rank of the selected object.

*A preliminary version of this work will be presented at the 36th Symposium on Foundations of Computer Science, November 1995.

†IBM Research Division, Almaden Research Center. Research supported in part by ONR Contract N-00014-94-C-0007. E-Mail: ajtai@almaden.ibm.com

‡IBM Research Division, Almaden Research Center, and School of Mathematical Sciences, Tel Aviv University, Research supported in part by ONR Contract N-00014-94-C-0007. E-Mail: megiddo@almaden.ibm.com

§Computer Science Division, U.C. Berkeley. Work supported in part by ONR Contract N-00014-94-C-0007 and NSF postdoctoral fellowship. Part of this work was done while this author was at IBM Almaden Research Center. E-Mail: waarts@cs.berkeley.edu

[1] $\log k \equiv \max\{1, \log_2 k\}$.

1. Introduction

In the classical *secretary* problem, n items or options are presented one by one in random order (*i.e.*, all $n!$ possible orders being equally likely). If we could observe them all, we could rank them totally with no ties, from best (rank 1) to worst (rank n). However, when the ith object appears, we can observe only its rank relative to the previous $i-1$ objects; the relative rank is equal to one plus the number of the predecessors of i which are preferred to i. We must accept or reject each object, irrevocably, on the basis of its rank relative to the objects already seen, and we are required to select k objects. The problem has two main variants. In the first, the goal is to maximize the probability of obtaining the best k objects. In the second, the goal is to minimize the expectation of the sum of the ranks of the selected objects or, more generally, for a given positive integer z, minimize the expectation of the sum of the zth powers of the ranks.

Solutions to the classical problem apply also in variety of more general situations. Examples include (i) the case where objects are drawn from some probability distribution; the interesting feature of this variant is that the decisions of the algorithms may be based not only on the relative rank of the item but also on an absolute "grade" that the item receives, (ii) the number of objects is not known in advance, (iii) objects arrive at random times, (iv) some limited backtracking is allowed: objects that were rejected may be recalled, (v) the acceptance algorithm has limited memory, and also combinations of these situations. In addition to providing intuition and upper and lower bounds for the above important generalizations of the problem, solutions to the classical problem also provide in many cases very good approximations, or even exact solutions (see [5, 14, 15] for survey and also [9]). Our methods can also be directly extended to apply for these generalizations.

The obvious application to choosing a best applicant for a job gives the problem its common name, although the problem (and our results) has a number of other applications in computer science. For any problem with a very large data set, it may be impractical to backtrack and select previous choices. For example, in the context of data mining, selecting records with best fit to requirements, or retrieving images from digital libraries. In such applications limited backtracking may be possible, and in fact this is one of the generalizations mentioned above. Another important application is when one needs to choose an appropriate sample from a population for the purpose of some study. In other applications the items may be jobs for scheduling, opportunities for investment, objects for fellowships, etc.

1.1 Background and Intuition

The problem has been extensively studied in the probability and statistics literature (see [5, 14, 15] for surveys and also [11]).

The case of $k = 1$

Let us first review the case of $k = 1$, *i.e.,* only one object has to be selected. Since the observer cannot go back and choose a previously presented object which, in retrospect, turns out to be the best, it clearly has to balance the risk of stopping too soon and accepting an apparently desirable object when an even better one might still arrive, against the risk of waiting for too long and then find that the best item had been rejected earlier.

It is easy to see that the optimal probability of selecting the best item does *not* tend to zero as n tends to infinity; consider the following stopping rule: reject the first half of the objects and then select the first relatively best one (if any). This rule chooses the best object whenever the latter is among the second half of the objects while the second best object is among the first half. Hence, for every n, this rule succeeds with probability greater than $1/4$. Indeed, it has been established ([8, 6, 3]) (see below) that there exists an optimal rule that has the following form: reject the first $r-1$ objects and then select the first relatively best one or, if none has been chosen through the end, accept the last object. When n tends to infinity, the optimal value of r tends to n/e, and the probability of selecting the best is approximately $1/e$. (Lindley showed the above using backward induction [8]. Later, Gilbert and Mosteller provided a slightly more accurate bound for r [6]. Dynkin established the result as an application of the theory of Markov stopping times [3].)

It is not as easy to see that the optimal expected rank of the selected object tends to a finite limit as n tends to infinity. Observe that the above algorithm (for maximizing the probability of selecting the best object) yields an expected rank of $n/(2e)$ for the selected item; the argument is as follows. With probability $1/e$, the best item is among the first n/e items, and in this case the algorithm selects the last item. The conditional expectation of the rank of the last object in this case is approximately $n/2$. Thus, the expected rank for the selected object in this algorithm tends to infinity with n. Indeed, in this paper we show that, surprisingly, the two goals are in fact in conflict (see Section 1.2).

It can be proven by backward induction that there exists an optimal policy for minimizing the expected rank of selected item that has the following form: accept an object if and only if its rank relative to the previously seen objects exceeds a certain threshold (depending on the number of objects seen so far). Note that while the optimal algorithm for maximizing the probability of selecting the best has to remember only the best object seen so far, the threshold algorithm has to remember all the previous objects. (See [12] for solutions where the observer is allowed to remember only one of the previously presented items.) This fact suggests that minimizing the expected rank is harder. Thus, not surprisingly, finding an approximate solution for the dynamic programming recurrence for this problem seems significantly harder than in the case of the first variant of the problem, *i.e.,* when the goal is to maximize the probability of selecting the best. Chow, Moriguti, Robbins, and Samuels, [2] showed that the optimal expected rank of the selected object is approximately 3.8695. The question of whether higher powers of the rank of the selected object tend to finite limits as n tends to infinity was resolved in [12]. It has also been shown that if the order of arrivals is determined by an adversary, then no algorithm can yield an expected rank better than $n/2$ [13].

The case of a general k

There has been much interest in the case where more than one object has to be selected. It is not hard to see that for every fixed k, the maximum probability of selecting the best k objects does not tend to zero as n tends to infinity. The proof is as follows. Partition the sequence of n objects into k disjoint intervals, each containing n/k consecutive items. Apply the algorithm for maximizing the probability of selecting the best object to each set independently. The resulting algorithm, selects the best item in each interval with probability e^{-k}. The probability that the best k objects belong to distinct intervals tends to $k!/k^k$ as n tends to infinity. For this first variant of the problem, the case of $k = 2$ was considered in [10]; Vanderbei [17], and independently Glasser, Holzager, and Barron [7], considered the problem for general k. They showed that there is an optimal policy with the following threshold form: accept an object with a given relative rank if and only the number of observations exceeds a critical number that depends on the number of items selected so far; in addition, an object which is worse than any of the already rejected objects need not be considered. Notice that this means that not all previously seen items have to be remembered, but only those that were already selected and the best among all those that were already rejected. This property is analogous to what happened in the $k = 1$ case, where the goal was to maximize the probability of selecting the best item. Both papers derive recursive relations using backward induction. General solutions to their recurrences are not known, but the authors give explicit solutions (*i.e.,* critical values and probability) for the case of $n = 2k$ [7, 17] and $n = 2k+1$ [7]. Vanderbei [17] also presents certain asymptotic results as n tends to infinity and k is fixed and also as both k and n tend to infinity so that $(2k - n)/\sqrt{n}$ remains finite.

In analogy to the case of $k = 1$, bounding the optimal expected sum of ranks of k selected items appears to be considerably harder than minimizing the probability of selecting the best k items. Also, here it is not obvious to see whether or not this sum tends to a finite limit when n tends to infinity. Backward induction gives recurrences that seem even harder to solve than those derived for the case of maximizing the probability of selecting the best k. Such equations were presented by Henke [9], but he was unable to approximate their general solutions.

Thus, the question of whether the expected sum of ranks of selected items tends to infinity with n has been open. There has not been any explicit solution for obtaining a bounded expected sum. Thus the second, possibly more realistic, variant of the secretary problem has remained open.

1.2 Our Results

In this paper we present a family of explicit algorithms for the secretary problem such that for each positive integer z, the family includes an algorithm for accepting items, where for all values of n and k, the resulting expected sum of the zth powers of the ranks of the accepted items is at most

$$\frac{k^{z+1}}{z+1} + C(z) \cdot k^{z+0.5} \log k \ ,$$

where $C(z)$ is a constant.[2]

Clearly, the sum of ranks of the zth powers of the best k objects is $k^{z+1}/(z+1) + O(k^z)$. Thus, the sum achieved by our algorithms is not only bounded by a value independent of n, but also differs from the best possible sum only by a relatively small amount. For every fixed k, this expected sum is bounded by a constant. Thus we resolve the above open questions regarding the expected sum of ranks and, in general, zth powers of ranks, of the selected objects.

Our approach is very different from the dynamic programming approach taken in most of the papers mentioned above. It has been more successful in obtaining explicit solutions to this classical problem, and can more easily be used to obtain explicit solutions for numerous generalizations.

[2] $\log k \equiv \max\{1, \log_2 k\}$.

We remark that our approach does not partition the items into k groups and select one item in each. Such a method is suboptimal since with high probability, a constant fraction of the best k items appear in groups where they are not the only ones from the best k. Therefore, this method rejects a constant fraction of the best k with high probability, and so the expected value of the sum of the ranks obtained by such an algorithm is greater by at least a constant factor than the optimal.

Since the expected sums achieved by our algorithms depend only on k and z and, in addition, the probability of our algorithms to select an object does not decrease with its rank, it will follow that the probabilities of our algorithms to actually select the best k objects depend only on k and z, and hence for fixed k and z, do not tend to zero when n tends to infinity. In particular, this means that for $k = z = 1$, our algorithms will select the best possible object with probability bounded away from zero.

In contrast, for any algorithm for the problem, if the order of arrival of items is the worst possible (*i.e.,* generated by an oblivious adversary), then the algorithm yields an expected sum of at least $kn^z 2^{-(z+1)}$ for the zth powers of the ranks of selected items. Our lower bound holds also for randomized algorithms.

Finally, in Section 1.1 we observed that an optimal algorithm for maximizing the probability of selecting the best object results in an unbounded expected rank of the selected object. As a second part of this work we show that this fact is not a coincidence: the two goals are in fact in conflict. No algorithm can simultaneously optimize the expected rank and the probability of selecting the best. We derive a lower bound on the trade-off between the probability of accepting the best object and the expected rank of the accepted item.

2. The Algorithms

In this section we describe a family of algorithms for the secretary problem, such that for each positive integer z, the family includes an algorithm for accepting objects, where the resulting expected sum of the zth powers of the ranks of accepted objects is

$$\frac{k^{z+1}}{z+1} + O(k^{z+0.5} \log k) \ .$$

In addition, it will follow that the algorithm accepts the best k objects with positive probability that depends only on k and z. Let z be the positive integer that we are given. Denote $p = 64 + \log^2 k$.

For the convenience of exposition, we assume without loss of generality that n is a power of 2. We partition the sequence $[1, \ldots, n]$ (corresponding to the objects in the order of arrival) into $m = \log n + 1$ consecutive intervals I_i $(i = 1 \ldots, m)$, so that

$$I_i = \begin{cases} \left[1 + n \sum_{j=1}^{i-1} 2^{-j}, n \sum_{j=1}^{i} 2^{-j}\right] & \text{if } 1 \le i \le m-1 \\ \{n\} & \text{if } i = m \end{cases}$$

In other words, the first $m-1$ intervals are $[1, \frac{n}{2}]$, $[\frac{n}{2}+1, \frac{3n}{4}], \ldots$, each containing a half of the remaining elements. The mth interval contains the last element. Note that $|I_i| = \lceil n/2^i \rceil$ $(i = 1, \ldots, m-1)$.

Let us refer to the first

$$m' = \max\{0, \lfloor \log(k/p) \rfloor\}$$

intervals as the *opening* ones, and let the rest be the *closing* ones. Note that since $p \geq 64$, the last five intervals are closing. For an opening I_i, the expected number of those of the top k objects in I_i is

$$|I_i| \cdot \frac{k}{n} = k/2^i \quad (i = 1, \ldots, m').$$

(The latter is not necessarily an integer.) Furthermore, for any $d \leq \sum_{j=1}^{m'} |I_i|$ (*i.e.*, d is in one of the opening intervals), the expected number of those of the top k objects among the first d to arrive is $d \cdot \frac{k}{n}$.

Let

$$p_i = \begin{cases} k2^{-i} & \text{if } i \leq m' \\ k2^{-m'} & \text{if } i = m' + 1 \\ 0 & \text{if } m' + 1 < i \leq m \end{cases}$$

Observe that $p_{m'+1} = k - \sum_{j=1}^{m'} p_j$.

We will refer to p_i as the *minimum number of acceptances required for* I_i $(i = 1, \ldots, m)$. Observe that for $i \leq m'$, $p_i \geq k \cdot 2^{-\log k/p} = p$. On the other hand, $p_{m'+1} = k2^{-m'} \leq k2^{-\log k/p+1} = 2p$.

Intuitively, during each interval the algorithm attempts to accept the expected number of top k objects that arrive during this interval, and in addition to make up for the number of objects that should have been accepted prior to the beginning of this interval but have not. Note that since $p_i = 0$ for $i > m' + 1$, then during such intervals the algorithm only attempts to make up for the number of objects that should have been accepted beforehand and have not.

Let us explain this slightly more formally. During each execution of the algorithm, at the beginning of each interval, the algorithm computes a *threshold* for acceptance, with the goal that by the time the processing of the last object of this interval is completed, the number of accepted objects will be at least the minimum number of acceptances required prior to this time. In particular, recall that for $i = 1, \ldots, m$, p_i denotes the minimum number of acceptances required for I_i. Given a "prefix" of an execution prior to the beginning of I_i $(i = 1, \ldots, m+1)$, let Q_j $(j = 0, \ldots, i-1)$, be the number of items accepted in I_j. Let $D_{i-1} = \max\left\{0, \sum_{j=1}^{i-1} p_j - \sum_{j=1}^{i-1} Q_j\right\}$. Roughly speaking, D_{i-1} is the difference between the minimum number of acceptances required prior to the beginning of I_i and the number of items that were actually accepted during the given prefix. Note that $D_0 = 0$.

Given a prefix of an execution prior to the beginning of I_i, let

$$A_i = \begin{cases} D_{i-1} + p_i & \text{if } \sum_{j=1}^{i-1} Q_j < k \\ 0 & \text{otherwise} \end{cases}$$

We refer to A_i computed at the beginning of I_i as the *acceptance threshold for* I_i in this execution. Loosely stated, given a prefix of execution of the algorithm prior to the beginning of I_i, A_i is the number of objects the algorithm has to accept during I_i in order to meet the minimum number required by the end of I_i. The algorithm will aim at accepting at least A_i objects during I_i. To ensure that it accepts that many, it attempts to accept a little more. In particular, during each opening interval I_i, the algorithm attempts to accept an expected number of $A_i + 6(z+1)\sqrt{A_i} \log k$.

As we will see, this ensures that the algorithm accepts at least A_i objects during this interval with probability of at least $k^{-5(z+1)}$. During each closing interval I_i, the algorithm attempts to accept an expected number of $32(z+1)A_i$. This ensures that the algorithm accepts at least A_i objects during this interval with probability of at least $2^{-5(z+1)(a_i+1)}$.

We make the distinction between opening and closing intervals in order to restrict the expected rank of the accepted objects. If I_i is closing, then A_i may be much smaller than $\sqrt{A_i}\log k$. Let

$$B_i = \begin{cases} A_i + 6(z+1)\sqrt{A_i}\log k & \text{if } I_i \text{ is opening} \\ 32(z+1)(A_i) & \text{if } I_i \text{ is closing.} \end{cases}$$

In order to accept an expected number of B_i objects during interval I_i, the algorithm will accept the dth item if it is one of the approximately $B_i \cdot 2^i d/n$ top ones among the first d. Since the order of arrival of the items is random, the rank of the dth object relative to the first d ones is distributed uniformly in the set $\{1, \ldots, d\}$. Therefore, the dth object will be accepted with probability of $B_i 2^i/n$, and hence, since $|I_i| = \lceil n/2^i \rceil$, the expected number of objects accepted during I_i is indeed B_i.

If at some point during the execution of the algorithm, the number of slots that still have to be filled equals the number of items that have not been processed yet, all the remaining items will be accepted regardless of rank. Analogously, if by the time the dth item arrives all slots have already been filled, this item will not be accepted.

Finally, the algorithm does not accept any of the first $\lceil n/(8\sqrt{k}) \rceil$ items except in executions during which the number of slots becomes equal to the number of items before $\lceil n/(8\sqrt{k}) \rceil$ items have been processed. Roughly speaking, this modification will allow to bound the expected rank of the dth item in terms of its rank relative to the first d items.

The above leads to our algorithm, which we call *Select*.

Algorithm Select: *The algorithm processes the items, one at a time, in their order of arrival. At the beginning of each interval I_i, the algorithm computes A_i as described above. When the dth item ($d \in I_i$) arrives, the algorithm proceeds as follows.*

(i) *If all slots have already been filled then the object is rejected.*

(ii) *Otherwise, if $d > \lceil n/(8\sqrt{k}) \rceil$, then*

  (a) *If $i \le m'$, the dth item is accepted if it is one of the top $\lfloor (A_i + 6(z+1)\sqrt{A_i}\log k)2^i d/n \rfloor$ items among the first d.*

  (b) *If $i > m'$, the algorithm accepts the dth item if it is one of the top $\lfloor 32(z+1)(A_i)2^i d/n \rfloor$ items among the first d.*

(iii) *Otherwise, if the number of slots that still have to be filled equals the number of items left (i.e., $n-d-1$), the dth item is accepted.*

We refer to acceptances under (3) , *i.e.*, when the number of slots that still have to be filled equals the number of items that remained to be seen, as *mandatory*, and to all other acceptances as *elective*. For example, if the dth item arrives during I_1, and the latter is opening, then the item is accepted electively if and only if it is one of the approximately

$$\begin{aligned}\lfloor (A_1 + 6(z+1)\sqrt{A_1}\log k)\cdot(2d/n)\rfloor &= \lfloor (k/2 + 6(z+1)\sqrt{k/2}\log k)\cdot(2d/n)\rfloor \\ &= \lfloor (k + 12(z+1)\sqrt{k/2}\log k)\cdot(d/n)\rfloor\end{aligned}$$

top objects among the first d. In general, if the dth object arrives during an opening I_i, then the object is accepted electively if and only if it is one of the approximately

$$\lfloor (2^i A_i + 6(z+1)\cdot 2^i\sqrt{A_i}\log k)\cdot(d/n)\rfloor$$

top objects among the first d.

3. Analysis of Algorithm Select

Very loosely stated, the proof proceeds as follows. In Section 3.1 we show that for $i = 1, \ldots, m+1$ ($m = \log n + 1$), with high probability, $D_{i-1} = 0$. Observe that this implies that for $i = 1, \ldots, m$, with high probability, A_i is approximately p_i, *i.e.*,

$$A_i \approx \begin{cases} 2^{-i}k & \text{if } i \le m' \\ 2^{-m'}k \le 2p & \text{if } i = m'+1 \\ 0 & \text{if } i > m'+1\ . \end{cases}$$

In Section 3.2 we show that if the dth object arrives during an opening I_i, then the conditional expectation of the zth power of its rank, given that it is accepted electively, is not greater than $2^{iz}\frac{1}{z+1}A_i^z + c_4(z)2^{iz}A_i^{z-0.5}\log k$, for some constant $c_4(z)$ (depending on z); if I_i is closing, this conditional expectation is not greater than $c_6(z)2^{iz}A_i^z$ for some constant $c_6(z)$. In Section 3.3 these results of Sections 3.1 and 3.2 are combined and it is established that if the dth object arrives during an opening I_i, then its conditional expected zth power of rank, given that it is accepted electively, is at most

$$\frac{k^z}{z+1} + c(z)2^{i/2}k^{z-0.5}\log k$$

for some constant $c(z)$. If I_i is closing, that conditional expected zth power of rank is at most $c'(z)k^z$, for some constant $c'(z)$, if $i = m'+1$, and is approximately 0 otherwise. From this it will follow that the expected sum of the zth powers of ranks of the electively accepted objects is $\frac{1}{z+1}k^{z+1} + O(k^{z+0.5}\log k)$. In addition we use the result of Section 3.1 to show that the expected sum of the zth powers of ranks of mandatorily accepted objects is $O(k^{z+0.5}\log k)$. Thus the expected sum of the zth powers of ranks of the accepted objects is $\frac{1}{z+1}k^{z+1} + O(k^{z+0.5}\log k)$.

In addition, from the fact that the expected sum of the zth powers of ranks of the accepted objects is bounded by a value that depends only on k and z, it will also follow that the algorithm accepts the top k objects with probability that depends only on k and z.

3.1 Bounding the A_is

In this section we show that for $i = 1, \ldots, m$, with high probability, A_i is very close to p_i. More precisely, we say that a prefix of execution prior to the end of the ith interval is *smooth*, if for each $j = 1 \ldots, i$, the value computed for A_i in this prefix is $\leq |I_j|$. We distinguish between smooth and nonsmooth executions.

In Section 3.1.1 we show that for an opening interval I_i, in executions whose prefix prior to the end of the $i-1$th interval is smooth, the probability that $A_i > 2^j p_i$ decreases exponentially with j (Part 1 of Lemma 3.3). For a closing I_i, in executions whose prefix prior to the end of the $i-1$th interval is smooth, the probability that $A_i > 2^j p_{m'+1}$ decreases exponentially both with j and with i (Part 2 of Lemma 3.3). Part 1 and Part 2 of Lemma 3.3 will follow, respectively, from Lemmas 3.1 and 3.2 that show that in executions whose prefix prior to the end of the ith interval is smooth, in I_i the algorithm accepts A_i objects with high probability (where A_i is computed for the prefix of the execution). Intuitively, the restriction to smooth executions is necessary since at most $|I_i|$ objects can be selected in I_i. Lemma 3.3 implies that for each $i = 1, \ldots, m$, in executions whose prefix prior to the end of the ith interval is smooth, with high probability, by the end of I_i the number of objects that were already accepted is not smaller than the minimum number of acceptances required prior to this point. The latter holds even if I_i started at a disadvantage in the sense that the minimum number of acceptances required prior to I_i was greater than the number of objects that were actually accepted by that point.

Clearly, Lemma 3.3 implies that in smooth executions, with high probability, A_i is very close to p_i. To complete the proof that A_i is close to p_i, Section 3.1.2 shows that nonsmooth executions are rare. In particular, Section 3.1.2 uses Lemma 3.3 to show that in executions whose prefix prior to the end of the $(i-1)$st interval is smooth, the probability that $A_i > |I_i|$ is less than $c(z)n^{-2.5(z+1)}$ for some constant $c(z)$ (Lemmas 3.4 and 3.5). The case of $k \geq n/2$ is excluded (Lemma 3.5) and thus handled separately later (Section 3.3).

3.1.1 Smooth Prefixes

Denote by E_i the prefix of an execution E prior to the end of I_i. Note that E_m is E. We say that E_i is smooth, if for $j = 1, \ldots, i$, A_j computed in E_i is $\leq |I_j|$. Denote by M_{E_i} the event in which E_i is smooth.

Lemma 3.1 *For every $i \leq m'$ and for any value a_i of A_i,*

$$\mathrm{Prob}\left\{D_i > 0 \mid \{A_i = a_i\} \cap \mathrm{M}_{E_i}\right\} < k^{-5(z+1)} .$$

Proof: Note that $D_i > 0$ only if the number of objects accepted in I_i is less than a_i.

Overview Loosely stated, the algorithm accepts the dth object electively if it is one of the top $\lfloor (A_i + 6(z+1)\sqrt{A_i}\log k)\frac{2^i d}{n} \rfloor$ objects among the first d. Since the objects arrive in a random order, the rank of the dth object within the set of first d is distributed uniformly and hence it will be accepted electively with probability not less than $\lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n} \rfloor / d$. Moreover, the rank of the dth object within the set of the first d is independent of the arrival order of the first

$d-1$, and hence is independent of whether or not any previous object in this interval, say the d_1th one, is one of the top $\lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)2^i d_1/n \rfloor$ objects among the first d_1. The rest of the proof follows from computing the expected number of accepted candidates and Chernoff inequality.

We proceed with the actual proof. Suppose $i \leq m'$, and let a_i be the acceptance threshold computed for I_i in a given execution. Recall that if the dth object arrives during I_i while there are still empty slots, $d > \lceil n/(8\sqrt{k}) \rceil$, and $i \leq m'$, then the algorithm accepts the object electively if it is one of the top $\lfloor (A_i + 6(z+1)\sqrt{A_i}\log k)\frac{2^i d}{n} \rfloor$ objects among the first d. (If either $d \leq \lceil n/(8\sqrt{k}) \rceil$ or there are no empty slots when the dth object arrives, it may not be accepted electively.) Since the objects arrive in a random order, the rank of the dth object within the set of first d is distributed uniformly and hence it will be accepted electively with probability not less than $\min\{1, \lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n} \rfloor / d\}$. Moreover, the rank of the dth object within the set of the first d is independent of the arrival order of the first $d-1$. Hence this rank is independent of whether or not any previous object in this interval, say the d_1th one, is one of the top $\lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)2^i d_1/n \rfloor$ objects among the first d_1.

Without loss of generality we may assume that

$$\min\{1, \lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n} \rfloor / d\} = \lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n} \rfloor / d < 1 \ .$$

For, if for some $d \in I_i$, $\lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n} \rfloor / d \geq 1$, then $a_i + 6(z+1)\sqrt{a_i}\log k \geq \frac{n}{2^i}$, and hence, for each $d \in I_i$, $\lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n} \rfloor \geq 1$. In this case each object in I_i is accepted with probability 1 unless all slots have already been filled. If all slots are filled, then $D_i = 0$, and we are done. Otherwise, $Q_i = |I_i|$. It follows from the definition that $D_i \leq a_i - Q_i$, and hence $D_i \leq a_i - |I_i|$. Since by the lemma assumption, $a_i \leq |I_i|$, it follows that D_i is non positive.

The rest of the proof follows directly from Chernoff's inequality. Formally, suppose the gth object is the first in I_i, *i.e.*, $g = 1 + n\sum_{j=1}^{i-1} 2^{-j}$. Define $X_1, \ldots, X_{|I_i|}$ to be independent random $(0,1)$-variables such that

$$\mathrm{Prob}\{X_t = 1\} \ = \ \frac{\lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i(t+g-1)}{n} \rfloor}{t+g-1}$$

for $t+g-1 > \lceil n/(8\sqrt{k}) \rceil$. It follows from the reasoning above that if the dth object is in an opening I_i, then the probability that the dth object is accepted electively is not less than $\mathrm{Prob}\{X_{d-g+1} = 1\}$. The independence of the order of arrival of the first $d-1$ objects also implies that

$$\mathrm{Prob}\{D_i > 0\} \leq \mathrm{Prob}\{Q_i < a_i\} \ \leq \ \mathrm{Prob}\left\{\sum_{t=1}^{|I_i|} X_t < a_i\right\} \ .$$

Thus, to complete the proof, we will show that $\mathrm{Prob}\{\sum_{t=1}^{|I_i|} X_t < a_i\} < k^{-5(z+1)}$. To this end, we first establish:

Claim: $\sum_{t=1}^{|I_i|} \mathrm{Prob}\{X_t = 1\} \geq a_i + 5(z+1)\sqrt{a_i}\log k$.

Proof: We distinguish two cases.

Case I: $i > 1$.

$$\begin{aligned}
\sum_{t=1}^{|I_i|} \mathrm{Prob}\{X_t = 1\} &= \sum_{t=1}^{|I_i|} \frac{\lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i(t+g-1)}{n}\rfloor}{t+g-1} \\
&\geq \sum_{t=1}^{|I_i|} \frac{(a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i(t+g-1)}{n} - 1}{t+g-1} \\
&= \sum_{t=1}^{n2^{-i}} (a_i + 6(z+1)\sqrt{a_i}\log k)2^i/n - \sum_{t=1}^{n2^{-i}} \frac{1}{t+g-1} \\
&\geq a_i + 6(z+1)\sqrt{a_i}\log k - n2^{-i}\cdot\frac{2}{n} \\
&= a_i + 6(z+1)\sqrt{a_i}\log k - 2^{-i+1} \\
&\geq a_i + 5(z+1)\sqrt{a_i}\log k.
\end{aligned}$$

The first inequality follows since for $i > 1$, we have $g > n/2$. Hence, $g > \lceil n/(8\sqrt{k})\rceil$, and the same holds for $t + g - 1$ for each $t \in I_i$. Thus, by definition,

$$\mathrm{Prob}\{X_t = 1\} = \frac{\lfloor (a_i + 6(z+1)\sqrt{a_i}\log k)\frac{2^i(t+g-1)}{n}\rfloor}{t+g-1}.$$

The third inequality follows since (i) $i < m$ because, as noted in Section 2., I_m is closing, and (ii) $|I_i| = n/2^i$ $(i = 1, \ldots, m-1)$. The fourth inequality follows from the fact that for $i > 1$, we have $g > n/2$. The last inequality follows because, as noted in Section 2., since $i \leq m'$, $p_i \geq k2^{-\log k/p} = p$, and hence also $k, a_i \geq p$. Since $p \geq 64$, $\sqrt{a_i}\log k \geq \sqrt{p}\log p \geq 1 \geq 2^{-i+1}$.

Case II: $i = 1$.

$$\begin{aligned}
&\sum_{t=1}^{|I_1|} \mathrm{Prob}\{X_t = 1\} \\
= \quad & \sum_{t=\lceil n/(8\sqrt{k})\rceil+1}^{|I_1|} \frac{\lfloor (a_1 + 6(z+1)\sqrt{a_1}\log k)\frac{2t}{n}\rfloor}{t} \\
\geq \quad & \sum_{t=\lceil n/(8\sqrt{k})\rceil+1}^{|I_1|} \frac{(a_1 + 6(z+1)\sqrt{a_1}\log k)\frac{2t}{n} - 1}{t} \\
= \quad & \sum_{t=\lceil n/(8\sqrt{k})\rceil+1}^{n/2} (a_1 + 6(z+1)\sqrt{a_1}\log k)2/n - \sum_{t=n/8\sqrt{k}}^{n/2} 1/t \\
\geq \quad & (a_1 + 6(z+1)\sqrt{a_1}\log k)\left(1 - \frac{2}{8\sqrt{k}} - \frac{2}{n}\right) - 4\sqrt{k} \\
\geq \quad & a_1 + 6(z+1)\sqrt{a_1}\log k - \left((a_1 + 6(z+1)\sqrt{a_1}\log k)\cdot\frac{1}{2\sqrt{k}} + 4\sqrt{k}\right) \\
= \quad & a_1 + 6(z+1)\sqrt{a_1}\log k - \left((k/2 + 6(z+1)\sqrt{k/2}\log k)\cdot\frac{1}{2\sqrt{k}} + 4\sqrt{k}\right)
\end{aligned}$$

$$
\begin{aligned}
&= a_1 + 6(z+1)\sqrt{a_1}\log k - \sqrt{k}\left(\frac{1}{4} + \frac{6(z+1)\log k}{\sqrt{8}\sqrt{k}} + 4\right) \\
&\geq a_1 + 6(z+1)\sqrt{a_1}\log k - \sqrt{k}\left(\tfrac{1}{4} + \tfrac{21}{16}(z+1) + 4\right) \\
&\geq a_1 + 6(z+1)\sqrt{a_1}\log k - (z+1)\sqrt{\frac{k}{2}}\log k \\
&= a_1 + 5(z+1)\sqrt{a_1}\log k.
\end{aligned}
$$

The first inequality follows since for $i = 1$ we have $g = 1$, and hence by definition,

$$\text{Prob}\{X_t = 1\} = \frac{\lfloor (a_1 + 6(z+1)\sqrt{a_1}\log k)\frac{2t}{n} \rfloor}{t}$$

for every $t > \lceil n/(8\sqrt{k}) \rceil$. The third inequality follows since, (i) as noted above, I_m is closing, and since in our case I_1 is opening, we have $1 < m$, and (ii) $vertI_i| = n2^{-i}$ $(i = 1, \ldots, m-1)$. The fifth inequality follows since $\frac{1}{4\sqrt{k}} \geq \frac{2}{n}$, because (i) as noted above, if $i \leq m'$ then $p_i \geq p$, and since $p_1 = \frac{k}{2}$, we have $k \geq 2p$, and (ii) $p \geq 64 + \log^2 k$, so $8\sqrt{k} \leq \sqrt{p}\sqrt{k} \leq k \leq n$. The sixth inequality follows since $a_1 = p_1 = \frac{k}{2}$. The eighth inequality follows since $k \geq 2p \geq 128$. The ninth inequality follows because $\log k \geq 7$ since, as noted above, $k \geq 128$. The last inequality follows since $a_1 = k/2$. ∎

An inequality related to Chernoff's states:

Let $X_1, \ldots, X_n$ be independent random $(0,1)$-variables with $\text{Prob}\{X_i = 1\} = p_i$, $0 < p_i < 1$. *Let $X = \sum_{i=1}^n X_i$ and $\mu = \sum_{i=1}^n p_i$. Then, for $\delta \in [0,1]$,*

$$\text{Prob}\{X < (1-\delta)\mu\} < \exp(-\tfrac{1}{2}\mu\delta^2).$$

Using the claim, we apply Chernoff inequality to our X_t's to get

$$
\begin{aligned}
\text{Prob}\left\{\sum_{t=1}^{|I_i|} X_t < a_i\right\} &= \text{Prob}\left\{\sum_{t=1}^{|I_i|} X_t < (a_i + 5(z+1)\sqrt{a_i}\log k)\left(1 - \frac{5(z+1)\sqrt{a_i}\log k}{a_i + 5(z+1)\sqrt{a_i}\log k}\right)\right\} \\
&< \exp\left(-\frac{25(z+1)^2 a_i \log^2 k}{2(a_i + 5(z+1)\sqrt{a_i}\log k)}\right) \\
&\leq \exp\left(-\frac{25 a_i (z+1)^2 \log^2 k}{4 a_i (z+1)\log k}\right) \\
&< \exp\left(-5(z+1)\log k\right) < k^{-5(z+1)}\ .
\end{aligned}
$$

The third inequality follows since as noted above, $a_i, k \geq p \geq 64$, and hence (i) $a_i \geq 7\sqrt{a_i}$, and (ii) $\log k \geq 1$. ∎

Lemma 3.2 *If $n \geq 16$, then for every $i > m'$,*

$$\text{Prob}\{D_i > 0 \mid \{A_i = a_i\} \cap \mathrm{M}_{E_i}\} < 2^{-5(z+1)(a_i+1)}\ .$$

Proof: Suppose $i > m'$, and let a_i be the acceptance threshold computed for I_i.

First, observe that $A_i > 0$ ($i = m'+1, \ldots, m$) implies $A_i \geq 1$. For, by definition

$$A_i = \begin{cases} D_{i-1} + p_i & \text{if } \sum_{j=1}^{i-1} Q_j < k \\ 0 & \text{otherwise} \end{cases}$$

Hence, for $i = m'+1$, if $k - \sum_{j=1}^{m'} Q_j \leq 0$, then $A_{m'+1} = \cdots = A_m = 0$, and the observation follows. Assume $k - \sum_{j=1}^{m'} Q_j > 0$. Then

$$A_{m'+1} = D_{m'} + p_{m'+1} = \max\{0, \sum_{j=1}^{m'} p_j - \sum_{j=1}^{m'} Q_j\} + p_{m'+1} = \max\{p_{m'+1}, k - \sum_{j=1}^{m'} Q_j\} \geq k - \sum_{j=1}^{m'} Q_j \ .$$

Since k and Q_j are integers, and by our assumption $k - \sum_{j=1}^{m'} Q_j > 0$, this means $A_{m'+1} \geq 1$. For $i > m'+1$, if $k - \sum_{j=1}^{i-1} Q_j \leq 0$, then $A_i = 0$ by definition. Otherwise, $A_{m'+1} > 0$, and hence as reasoned above is $\geq k - \sum_{j=1}^{m'} Q_j$. Thus

$$A_i = D_{i-1} = \max\{0, A_{m'+1} - \sum_{j=m'+1}^{i-1} Q_j\} \geq k - \sum_{j=1}^{m'} Q_j - \sum_{j=m'+1}^{i-1} Q_j = k - \sum_{j=1}^{i-1} Q_j \ .$$

Thus, since k and Q_j are integers, and $k - \sum_{j=1}^{i-1} Q_j > 0$ by assumption, then $A_j \geq 1$.

If $a_i = 0$, the lemma follows since $D_i \leq a_i$. Thus, assume that $a_i \geq 1$. The proof is analogous to that of Lemma 3.1.

Recall that for $d > \lceil n/(8\sqrt{k}) \rceil$, if the dth object arrives during a closing I_i while there are still empty slots, then the object is accepted electively if it is one of the top $\lfloor 32(z+1)A_i 2^i d/n \rfloor$ objects among the first d. (If either $d \leq \lceil n/(8\sqrt{k}) \rceil$, or there are no empty slots, this object is not accepted electively.) Since the rank of the dth object in the set of the first d is uniformly distributed, it will be accepted electively with probability not less than $\min\{1, \lfloor 32(z+1)a_i 2^i d/n \rfloor / d\}$.

As in the proof of Lemma 3.1, we may assume that $\min\{1, \lfloor 32(z+1)a_i 2^i d/n \rfloor / d\} = \lfloor 32(z+1)a_i 2^i d/n \rfloor / d < 1$. We apply again Chernoff's inequality. Suppose the gth object is the first to arrive during I_i, *i.e.*, $g = 1 + n\sum_{j=1}^{i-1} 2^{-j}$. Let $X_1, \ldots, X_{|I_i|}$ be independent random $(0,1)$-variables such that

$$\mathrm{Prob}\{X_t = 1\} = \lfloor 32(z+1)a_i 2^i (t+g-1)/n \rfloor / (t+g-1)$$

for $t + g - 1 > \lceil n/(8\sqrt{k}) \rceil$. It follows that if the dth object arrives during I_i and $i > m'$, then the probability that it is accepted electively is not less than $\mathrm{Prob}\{X_{t-g+1} = 1\}$. It also follows that $\mathrm{Prob}\{D_i > 0\} \leq \mathrm{Prob}\{Q_i < a_i \mid A_i = a_i\} \leq \mathrm{Prob}\{\sum_{t=1}^{|I_i|} X_t < a_i\}$. Thus, to complete the proof, we will show that $\mathrm{Prob}\{\sum_{t=1}^{|I_i|} X_t < a_i\} < 2^{-5(z+1)(a_i+1)}$.

To show this we first prove:

Claim: $\sum_{t=1}^{|I_i|} \mathrm{Prob}\{X_t = 1\} \geq 16(z+1)a_i$.

Proof: Again, we distinguish two cases.

Case I: $i > 1$.

$$\begin{aligned}
\sum_{t=1}^{|I_i|} \text{Prob}\{X_t = 1\} &= \sum_{t=1}^{|I_i|} \frac{\lfloor 32(z+1)a_i \frac{2^i(t+g-1)}{n} \rfloor}{t+g-1} \\
&\geq \sum_{t=1}^{|I_i|} \frac{32(z+1)a_i \frac{2^i(t+g-1)}{n} - 1}{t+g-1} \\
&= \sum_{t=1}^{\lceil n2^{-i} \rceil} 32(z+1)a_i \frac{2^i}{n} - \sum_{t=1}^{\lceil n2^{-i} \rceil} \frac{1}{t+g-1} \\
&\geq 32(z+1)a_i - \lceil n2^{-i} \rceil \frac{1}{n/2} \\
&\geq 32(z+1)a_i - 2^{-i+2} \\
&\geq 16(z+1)a_i.
\end{aligned}$$

The first inequality follows since for $i > 1$, we have $g > n/2$; thus $g > \lceil n/(8\sqrt{k}) \rceil$, and hence so is $t + g - 1$ for each t in I_i; thus by definition, $\text{Prob}\{X_t = 1\} = \lfloor 32(z+1)a_i 2^i(t+g-1)/n \rfloor/(t+g-1)$. The third inequality follows since $|I_i| = \lceil n/2^i \rceil$. The fourth inequality follows again from the fact that for $i > 1$, we have $g > n/2$. The last inequality follows since, as noted in the beginning of the proof, we may assume $a_i \geq 1$. Thus, $8(z+1)a_i \geq 2 \geq 2^{-i+2}$.

Case II: $i = 1$.

$$\begin{aligned}
\sum_{t=1}^{|I_1|} \text{Prob}\{X_t = 1\} &= \sum_{t=\lceil n/(8\sqrt{k}) \rceil + 1}^{|I_1|} \frac{\lfloor 32(z+1)a_1(2t/n) \rfloor}{t} \\
&\geq \sum_{t=\lceil n/(8\sqrt{k}) \rceil + 1}^{|I_1|} \frac{32(z+1)a_1(2t/n) - 1}{t} \\
&= \sum_{t=\lceil n/(8\sqrt{k}) \rceil + 1}^{\lceil n/2 \rceil} 32(z+1)a_1 \cdot \frac{2}{n} - \sum_{t=\lceil n/(8\sqrt{k}) \rceil + 1}^{\lceil n/2 \rceil} \frac{1}{t} \\
&\geq 32(z+1)a_1 \left(1 - \frac{2}{8\sqrt{k}} - \frac{2}{n}\right) - 4\sqrt{k} \\
&= 32(z+1)a_1 - \left(\frac{64(z+1)a_1}{8\sqrt{k}} + \frac{64(z+1)a_1}{n} + 4\sqrt{k}\right) \\
&= 32(z+1)k - k \cdot \left(\frac{8(z+1)}{\sqrt{k}} + \frac{64(z+1)}{n} + \frac{4}{\sqrt{k}}\right) \\
&\geq 32(z+1)k - k \cdot 16(z+1) \\
&\geq 16(z+1)a_1 \ .
\end{aligned}$$

The first inequality follows since for $i = 1$ we have $g = 1$, and hence by definition,

$$\text{Prob}\{X_t = 1\} = \lfloor 32(z+1)a_1 2^1 t/n \rfloor / t \quad \text{for} \quad t > \lceil n/(8\sqrt{k}) \rceil \ .$$

The third inequality follows since $\lceil n/2 \rceil$ objects arrive during I_1. The sixth inequality follows because, since $1 = i > m'$ in this case, we get $m' = 0$; thus by definition, $a_1 = p_1 = k2^{-m'} = k$. The seventh inequality follows since $n \geq 16$ and $z \geq 1$. ∎

Using the claim, we apply Chernoff's inequality to get

$$\begin{aligned}
\text{Prob}\left\{\sum_{t=1}^{|I_i|} X_t < a_i\right\} &= \text{Prob}\left\{\sum_{t=1}^{|I_i|} X_t < 16(z+1)a_i \cdot \left(1 - \frac{16(z+1)-1}{16(z+1)}\right)\right\} \\
&< \exp\left(-\tfrac{1}{2} \cdot 16(z+1)a_i \cdot \left(\frac{16(z+1)-1}{16(z+1)}\right)^2\right) \\
&< \exp\left(-\tfrac{1}{2} \cdot 16(z+1)a_i \cdot (9/10)\right) \le 2^{-5(z+1)(a_i+1)} .
\end{aligned}$$

The third inequality follows from $z \ge 1$. The last inequality follows since, as noted in the beginning of the proof, we may assume that $a_i \ge 1$. ■

Lemma 3.3

(i) *For* $i \le m'$,

$$\text{Prob}\{A_i > k2^{-i}(2^j - 1) \mid \mathrm{M}_{E_{i-1}}\} \le k^{-5(z+1)j} .$$

(ii) *If* $n \ge 16$, *then for* $i > m'$, $j \ge 0$,

$$\text{Prob}\{A_i > k2^{-m'}(2^j - 1) \mid \mathrm{M}_{E_{i-1}}\} \le k^{-5(z+1)j}(2^{-5(z+1)})^{i-m'-1} .$$

Proof: For the proof of (1), suppose $i \le m'$. Without loss of generality we may assume that $j \ge 1$, because for $j \le 0$, we have $k^{-5(z+1)j} \ge 1$, and (1) follows.

Recall that the minimum number of acceptances required for an opening interval I_i is $p_i = k2^{-i}$. Thus if $A_i > k2^{-i}$, then $D_{i-1} > 0$. Moreover,

$$\frac{k}{2^i}(2^j - 1) = \frac{k}{2^i}(1 + 2 + \cdots + 2^{j-1}) = p_i + p_{i-1} + \cdots + p_{i-j+1}.$$

By induction, if $A_i > k2^{-i}(2^j - 1)$, then $D_{i-1}, D_{i-2}, \ldots, D_{i-j}$ are positive. Thus, it is enough to bound

$$\text{Prob}\left\{(D_{i-1} > 0) \cap (D_{i-2} > 0) \cap \cdots \cap (D_{i-j} > 0) \mid \mathrm{M}_{E_{i-1}}\right\} .$$

Note that the above events $D_a > 0$ are mutually dependent, and are conditioned on $\mathrm{M}_{E_{i-1}}$. However, both the dependency and the conditioning are working in our favour. Thus, Lemma 3.1 implies that each of the underlying events $\{D_q > 0\}$ $(q = 1, \ldots, i-1)$, occurs with probability less than $k^{-5(z+1)}$. Clearly, each of the events $\{D_q > 0\}$ $(q \le 0)$ occurs with probability 0 and hence less than $k^{-5(z+1)}$. Thus,

$$\text{Prob}\{A_i > k2^{-i}(2^j - 1)\} \le (k^{-5(z+1)})^j = k^{-5(z+1)j} .$$

For the proof of (2), suppose $i > m'$. Recall that

$$p_{m'+2} = \cdots = p_m = 0$$

and

$$p_{m'+1} = k2^{-m'} .$$

Thus, if $A_i > k2^{-m'}(2^j - 1)$, then we must have

$$D_{m'} > k2^{-m'}(2^j - 2)$$

and

$$D_{m'+1}, \ldots, D_{i-1} > 0 \ .$$

Lemma 3.2 implies that for each q ($q = m'+1, \ldots, m$), the underlying event $\{D_q > 0\}$ occurs with probability less than $2^{-5(z+1)}$. Again the dependency and the conditioning on $\mathrm{M}_{E_{i-1}}$ are working in our favour. Thus, if $j = 0$, then

$$\mathrm{Prob}\{A_i > k2^{-m'}(2^j - 1) \mid \mathrm{M}_{E_{i-1}}\} \leq (2^{-5(z+1)})^{i-m'-1} = k^{-5(z+1)j}(2^{-5(z+1)})^{i-m'-1}.$$

To complete the proof, assume $j \geq 1$. Then $D_{m'} > k2^{-m'}(2^j - 2) \geq 0$. Lemma 3.2 implies that the underlying event $\{D_{m'} > 0\}$ occurs with probability less than $k^{-5(z+1)}$. Moreover, since $D_{m'} \leq A_{m'}$, it follows that $D_{m'} > k2^{-m'}(2^j - 2)$ implies $A_{m'} > k2^{-m'}(2^j - 2) \geq k2^{-m'}(2^{j-1} - 1)$. The first part of the lemma implies thus that for $j \geq 1$, the underlying event $\{A_{m'} > k2^{-m'}(2^{j-1} - 1)\}$ occurs with probability at most $k^{-5(z+1)(j-1)}$. Hence

$$\mathrm{Prob}\{A_i > k2^{-m'}(2^j - 1) \mid \mathrm{M}_{E_{i-1}}\} \leq k^{-5(z+1)(j-1)} \cdot k^{-5(z+1)} \cdot (2^{-5(z+1)})^{i-m'-1} = k^{-5(z+1)j}(2^{-5(z+1)})^{i-m'-1}.$$

■

3.1.2 Nonsmooth Executions

Lemma 3.4 *If $i \leq m'$, then*

$$\mathrm{Prob}\{\neg\mathrm{M}_{E_i} \cap \mathrm{M}_{E_{i-1}}\} \leq 2^{5(z+1)} n^{-2.5(z+1)} \ .$$

Proof:

$$\begin{aligned}
\mathrm{Prob}\{\neg\mathrm{M}_{E_i} \cap \mathrm{M}_{E_{i-1}}\} &= \mathrm{Prob}\{A_i > |I_i| \cap \mathrm{M}_{E_{i-1}}\} \\
&= \mathrm{Prob}\{A_i > |I_i| \mid \mathrm{M}_{E_{i-1}}\} \cdot \mathrm{Prob}\{\mathrm{M}_{E_{i-1}}\} \\
&\leq \mathrm{Prob}\{A_i > |I_i| \mid \mathrm{M}_{E_{i-1}}\} \\
&\leq \mathrm{Prob}\{A_i > n2^{-i} \mid \mathrm{M}_{E_{i-1}}\} \ .
\end{aligned}$$

The first inequality follows from the definition of $\neg\mathrm{M}_{E_i} \cap \mathrm{M}_{E_{i-1}}$. The last inequality follows since by definition, $|I_i| = \lceil n2^{-i} \rceil$ ($i = 1, \ldots, m$).

We distinguish two cases.

Case I: $k \leq \sqrt{n}$.

$$\begin{aligned}
\mathrm{Prob}\{A_i > n2^{-i} \mid \mathrm{M}_{E_{i-1}}\} &= \mathrm{Prob}\{A_i > (k2^{-i})\frac{n}{k} \mid \mathrm{M}_{E_{i-1}}\} \\
&\leq \mathrm{Prob}\{A_i > (k2^{-i})\sqrt{n} \mid \mathrm{M}_{E_{i-1}}\} \\
&\leq \mathrm{Prob}\{A_i > (k2^{-i})(2^j - 1) \mid \mathrm{M}_{E_{i-1}}\} \ ,
\end{aligned}$$

where $j = \lceil \frac{1}{2} \log n \rceil - 1$.

From Part (1) of Lemma 3.3 it follows that

$$\begin{aligned}
\mathrm{Prob}\{A_i > (k2^{-i})(2^j - 1) \mid \mathrm{M}_{E_{i-1}}\} &\leq k^{-5(z+1)j} \\
&\leq 2^{-5(z+1)(\lceil \frac{1}{2}\log n \rceil - 1)} \\
&\leq 2^{-5(z+1)(\frac{1}{2}\log n - 1)} \\
&\leq 2^{5(z+1)} \cdot n^{-2.5(z+1)} .
\end{aligned}$$

The second inequality follows from $k \geq 2$ because (i) since $i \leq m'$, we have $p_i \geq p$ and hence also $k \geq p$, and (ii) by definition, $p \geq 64$.

Case II: $k \geq \sqrt{n}$.

$$\mathrm{Prob}\{A_i > n2^{-i} \mid \mathrm{M}_{E_{i-1}}\} \leq \mathrm{Prob}\{A_i > k2^{-i} \mid \mathrm{M}_{E_{i-1}}\} \leq (\sqrt{n})^{-5(z+1)} \leq n^{-2.5(z+1)} .$$

The first inequality follows from Part (1) of Lemma 3.3. ■

Lemma 3.5 *If $n \geq 16$, $k \leq \frac{1}{2}n$, and $i > m'$, then*

$$\mathrm{Prob}\{\neg \mathrm{M}_{E_i} \cap \mathrm{M}_{E_{i-1}}\} \leq 2^{10(z+1)} n^{-2.5(z+1)} .$$

Proof:

$$\begin{aligned}
\mathrm{Prob}\{\neg \mathrm{M}_{E_i} \cap \mathrm{M}_{E_{i-1}}\} &= \mathrm{Prob}\{A_i > |I_i| \cap \mathrm{M}_{E_{i-1}}\} \\
&= \mathrm{Prob}\{A_i > |I_i| \mid \mathrm{M}_{E_{i-1}}\} \cdot \mathrm{Prob}\{\mathrm{M}_{E_{i-1}}\} \\
&\leq \mathrm{Prob}\{A_i > |I_i| \mid \mathrm{M}_{E_{i-1}}\} \\
&\leq \mathrm{Prob}\{A_i > n2^{-i} \mid \mathrm{M}_{E_{i-1}}\}.
\end{aligned}$$

The first inequality follows from the definition of $\neg \mathrm{M}_{E_i} \cap \mathrm{M}_{E_{i-1}}$. The last inequality follows from the definition $|I_i| = \lceil n2^{-i} \rceil$ $(i = 1, \ldots, m)$.

We distinguish two cases.

Case I: $k \leq \sqrt{n}$.

$$\begin{aligned}
\mathrm{Prob}\{A_i > n2^{-i} \mid \mathrm{M}_{E_{i-1}}\} &= \mathrm{Prob}\{A_i > (k2^{-m'}) \cdot \frac{n}{k} 2^{-i+m'} \mid \mathrm{M}_{E_{i-1}}\} \\
&\leq \mathrm{Prob}\{A_i > (k2^{-m'}) \cdot \sqrt{n} 2^{-i+m'} \mid \mathrm{M}_{E_{i-1}}\} \\
&\leq \mathrm{Prob}\{A_i > k2^{-m'} \cdot (2^j - 1) \mid \mathrm{M}_{E_{i-1}}\} ,
\end{aligned}$$

where $j = \max\{0, \lceil \frac{1}{2} \log n \rceil - i + m' - 1\}$.

We distinguish again two cases.

Case I.a. $\lceil \frac{1}{2} \log n \rceil - i + m' - 1 \leq 0$. In this case, $i \geq \lceil \frac{1}{2} \log n \rceil + m' - 1$. From Part (2) of Lemma 3.3 it follows that

$$\begin{aligned}
\mathrm{Prob}\{A_i > k2^{-m'} \cdot (2^j - 1) \mid \mathrm{M}_{E_{i-1}}\} &\leq k^{-5(z+1)j} 2^{-5(z+1)(i-m'-1)} \\
&\leq 2^{-5(z+1)(i-m'-1)} \\
&\leq 2^{-5(z+1)(\lceil \frac{1}{2}\log n \rceil - 2)} \\
&\leq 2^{10(z+1)} n^{-2.5(z+1)} .
\end{aligned}$$

Case I.b: $\lceil \frac{1}{2}\log n\rceil - i + m' - 1 \geq 0$. From Part (2) of Lemma 3.3 it follows that

$$\text{Prob}\{A_i > k2^{-m'} \cdot (2^j - 1) \mid \text{M}_{E_{i-1}}\} \leq k^{-5(z+1)j} 2^{-5(z+1)(i-m'-1)}.$$

We distiguish two cases.

Case I.b.1: $k \leq 2$.

By definition

$$A_i \leq p_i + D_{i-1} = p_i + \max\left\{0, \sum_{j=1}^{i-1} p_j - \sum_{j=1}^{i-1} Q_j\right\} \leq \sum_{j=1}^{i} p_j \leq k \leq 2\ .$$

On the other hand, by our assumption $A_i \geq n2^{-i}$. Thus, $n2^{-i} \leq 2$, and hence $i \geq \log n - 1$. In addition, $m' = 0$ since by definition, $m' = \max\{0, \lfloor \log(k/p)\rfloor\}$ and $p \geq 64$. Thus,

$$k^{-5(z+1)j} 2^{-5(z+1)(i-m'-1)} \leq 2^{-5(z+1)(i-m'-1)} \leq 2^{-5(z+1)(\log n - 2)} \leq 2^{10(z+1)} n^{-5(z+1)}\ .$$

Case I.b.2: $k \geq 2$.

$$\begin{aligned} k^{-5(z+1)j} 2^{-5(z+1)(i-m'-1)} &\leq 2^{-5(z+1)(\lceil \frac{1}{2}\log n\rceil - i + m' - 1)} \cdot 2^{-5(z+1)(i-m'-1)} \\ &\leq 2^{-5(z+1)(\frac{1}{2}\log n - 2)} \\ &\leq 2^{10(z+1)} \cdot n^{-2.5(z+1)}. \end{aligned}$$

Case II: $k \geq \sqrt{n}$. We distinguish again two cases.

Case II.a: $i = m' + 1$.

$$\begin{aligned} \text{Prob}\{A_i > n2^{-i} \mid \text{M}_{E_{i-1}}\} &= \text{Prob}\{A_i > n2^{-m'-1} \mid \text{M}_{E_{m'}}\} \\ &\leq \text{Prob}\{A_i > k2^{-m'} \mid \text{M}_{E_{m'}}\} \\ &\leq k^{-5(z+1)} \\ &\leq n^{-2.5(z+1)}\ . \end{aligned}$$

The second inequality follows from the lemma assumption that $k \leq \frac{1}{2}n$. The third inequality follows from Part (2) of Lemma 3.3.

Case II.b: $i > m' + 1$.

$$\begin{aligned} \text{Prob}\{A_i > n2^{-i} \mid \text{M}_{E_{i-1}}\} &\leq \text{Prob}\{A_i > 0 \mid \text{M}_{E_{i-1}}\} \\ &\leq \text{Prob}\{D_{m'+1} > 0 \mid \text{M}_{E_{m'+1}}\} \\ &\leq 2^{-5(z+1)(a_{m'+1}+1)} \\ &\leq 2^{-5(z+1)p} \\ &\leq 2^{-5(z+1)\log^2 k} \\ &\leq k^{-5(z+1)} \\ &\leq (\sqrt{n})^{-5(z+1)} \\ &\leq n^{-2.5(z+1)}\ . \end{aligned}$$

The third inequality follows from Lemma 3.2. The fourth inequality follows since by definition of $A_{m'+1}$, if it is ≥ 0, then it is $\geq k2^{-m'} \geq p$. The fifth inequality follows from $p \geq \log^2 k$. ∎

The case of $k \geq n/2$ is excluded (Lemma 3.5) and thus handled separately later (Section 3.3).

3.2 Expected zth powers of Ranks

Let us denote by R_d the random variable of the rank of the dth object. We define the *arrival rank* of the dth object as its rank within the set of the first d objects, *i.e.,* one plus the number of better objects seen so far. Denote by S_d the random variable of the arrival rank. Denote by NA_d the event in which the dth object is accepted electively. In this section we show that there exist constants $c_4(z), c_5(z)$ and $c_6(z)$ such that if the dth object arrives during an opening interval I_i, then

$$\mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\}) \le 2^{iz} \cdot \frac{a_i^z}{z+1} + c_4(z) 2^{iz} a_i^{z-\frac{1}{2}} \log k + c_5(z) \sqrt{\frac{n}{d}} 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k$$

(Lemma 3.9); and if I_i is closing, then

$$\mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\}) \le c_6(z) 2^{iz} a_i^z$$

(Lemma 3.10). To prove the above, we first prove a technical lemma (Lemma 3.6) showing that for fixed d and s, if $r \ge \frac{n}{d}s + \frac{n}{d}\sqrt{s}$, then $\mathrm{Prob}\{R_d = r \mid S_d = s\}$ decreases exponentially with r. This lemma will be used to prove Lemma 3.7, that states roughly that there exists a constant $c_2(z)$ such for every s

$$\mathcal{E}(R_d^z \mid S_d = s) \le \left(\frac{n}{d}\right)^z s^z + c_2(z) \left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}} \log k \ .$$

Lemma 3.9 will follow by combining the result of Lemma 3.7 with the fact that given that the object is accepted electively during an opening interval I_i and $A_i = a_i$, then S_d is distributed uniformly in the set $\{1, 2, \ldots, \lfloor (a_i + 6(z+1)\sqrt{a_i} \log k) 2^i d/n \rfloor\}$. Lemma 3.10 will follow analogously by combining the result of Lemma 3.7 with the fact that given that the object is accepted electively during a closing interval I_i and $A_i = a_i$, then S_d is distributed uniformly in the set $\{1, 2, \ldots, \lfloor 32(z+1) a_i 2^i d/n \rfloor\}$.

Lemma 3.6 *For all s and $j \ge \frac{n}{d}\sqrt{s}$,*

$$\mathrm{Prob}\left\{R_d = \frac{n}{d}s + j \mid S_d = s\right\} \le \exp\left(-\frac{jd}{8n\sqrt{s}}\right) .$$

Proof: Clearly,

$$\mathrm{Prob}\left\{R_d = \frac{n}{d}s + j \mid S_d = s\right\} = \frac{\binom{\frac{n}{d}s+j-1}{s-1}\binom{n-\frac{n}{d}s-j}{d-s}}{\binom{n-1}{d-1}} \ .$$

Define

$$\alpha = \frac{\binom{\frac{n}{d}s-1}{s-1}\binom{n-\frac{n}{d}s}{d-s}}{\binom{n-1}{d-1}} \ .$$

Then

$$\begin{aligned} & \mathrm{Prob}\left\{R_d = \frac{n}{d}s + j \mid S_d = s\right\} \\ = \quad & \frac{(\frac{n}{d}s + j - 1)!}{(\frac{n}{d}s + j - s)!} \cdot \frac{1}{(s-1)!} \cdot \frac{(n - \frac{n}{d}s - j)!}{(n - \frac{n}{d}s - j - d + s)!} \cdot \frac{1}{(d-s)!} \cdot \frac{1}{\binom{n-1}{d-1}} \end{aligned}$$

$$
\begin{aligned}
&= \alpha \frac{(\frac{n}{d}s+j-1)\cdots(\frac{n}{d}s)}{(\frac{n}{d}s+j-s)\cdots(\frac{n}{d}s+1-s)} \cdot \frac{(n-\frac{n}{d}s-j-d+s+1)\cdots(n-\frac{n}{d}s-d+s)}{(n-\frac{n}{d}s-j+1)\cdots(n-\frac{n}{d}s)} \\
&\le \alpha \left(\frac{\frac{n}{d}s}{\frac{n}{d}s+1-s}\right)^{\frac{j}{2}} \cdot \left(\frac{\frac{n}{d}s+\frac{j}{2}}{\frac{n}{d}s+\frac{j}{2}-s}\right)^{\frac{j}{2}} \cdot \left(\frac{n-\frac{n}{d}s-d+s}{n-\frac{n}{d}s}\right)^{\frac{j}{2}} \cdot \left(\frac{n-\frac{n}{d}s-\frac{j}{2}-d+s}{n-\frac{n}{d}s-\frac{j}{2}}\right)^{\frac{j}{2}} \\
&\le \alpha \left(\frac{1}{1-\frac{d}{n}}\right)^{\frac{j}{2}} \cdot \left(\frac{1}{1-\frac{d}{n}\frac{s}{s+\frac{d}{n}\frac{j}{2}}}\right)^{\frac{j}{2}} \cdot \left(1-\frac{d}{n}\cdot\frac{1-\frac{s}{d}}{1-\frac{s}{d}}\right)^{\frac{j}{2}} \cdot \left(1-\frac{d}{n}\cdot\frac{1-\frac{s}{d}}{1-\frac{s}{d}-\frac{j}{2n}}\right)^{\frac{j}{2}} .
\end{aligned}
$$

Let us denote

$$
\rho(n,d,s,j) = -\ln\left(1-\frac{d}{n}\cdot\frac{s}{s+\frac{d}{n}\frac{j}{2}}\right) + \ln\left(1-\frac{d}{n}\cdot\frac{1-\frac{s}{d}}{1-\frac{s}{d}-\frac{j}{2n}}\right) ,
$$

so we can write

$$
\text{Prob}\left\{R_d = \frac{n}{d}s+j \mid S_d = s\right\} \le \alpha e^{\rho(n,d,s,j)\frac{j}{2}} .
$$

Observe that $\alpha \le 1$. Thus, to complete the proof, we show that $\rho(n,d,s,j) \le -\frac{d}{4n\sqrt{s}}$.

However,

$$
\rho(n,d,s,j) = \sum_{t=1}^{\infty}\left(\left(\frac{s}{s+\frac{d}{n}\frac{j}{2}}\right)^t - \left(\frac{1-\frac{s}{d}}{1-\frac{s}{d}-\frac{j}{2n}}\right)^t\right)\left(\frac{d}{n}\right)^t \cdot \frac{1}{t} .
$$

Thus, to complete the proof, it suffices to show that

$$
\text{(i)} \qquad \frac{s}{s+\frac{d}{n}\frac{j}{2}} - \frac{1-\frac{s}{d}}{1-\frac{s}{d}-\frac{j}{2n}} \le -\frac{1}{4\sqrt{s}} ,
$$

$$
\text{and (ii) for each } t>1, \ \left(\frac{s}{s+\frac{d}{n}\frac{j}{2}}\right)^t \le \left(\frac{1-\frac{s}{d}}{1-\frac{s}{d}-\frac{j}{2n}}\right)^t .
$$

For the proof of (i),

$$
\frac{s}{s+\frac{d}{n}\frac{j}{2}} - \frac{1-\frac{s}{d}}{1-\frac{s}{d}-\frac{j}{2n}} = \frac{s-\frac{s^2}{d}-\frac{js}{2n}-s-\frac{dj}{2n}+\frac{s^2}{d}+\frac{sj}{2n}}{s-\frac{s^2}{d}-\frac{sj}{2n}+\frac{dj}{2n}-\frac{sj}{2n}-\frac{dj^2}{4n^2}} \le -\frac{\frac{dj}{2n}}{s+\frac{dj}{2n}} \le -\frac{\frac{\sqrt{s}}{2}}{s+\frac{\sqrt{s}}{2}} < -\frac{1}{4\sqrt{s}}
$$

The second inequality follows from the fact that the denominator is a product of two positive factors (since each of these factors originated from the factorials terms in the beginning of the proof). The third inequality follows since by the statement of the lemma, $j \ge \frac{n}{d}\sqrt{s}$.

Clearly, (ii) follows from (i). ∎

Lemma 3.7 *There exist constants $c_2(z)$, $c_3(z)$ and $c_{18}(z)$ such that for all $d \ge \frac{n}{k}$ and s,*

$$
\mathcal{E}(R_d^z \mid S_d = s) \le \left(\frac{n}{d}\right)^z \left(1+\frac{c_3(z)}{k}\frac{d}{n}\right) s^z + c_2(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c_{18}(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k .
$$

Proof: Define $j_0 = 32z\frac{n}{d}\sqrt{s}\ln k$.

$$
\begin{aligned}
& \mathcal{E}(R_d^z \mid S_d = s) \\
= \;& \sum_{r=1}^{\infty} r^z \cdot \mathrm{Prob}\{R_d = r \mid S_d = s\} \\
\leq \;& \left(\frac{n}{d}s + j_0\right)^z \cdot \mathrm{Prob}\left\{R_d < \frac{n}{d}s + j_0 \mid S_d = s\right\} + \sum_{r=\frac{n}{d}s+j_0}^{\infty} r^z \cdot \mathrm{Prob}\{R_d = r \mid S_d = s) \\
= \;& \left(\frac{n}{d}s + j_0\right)^z \cdot \mathrm{Prob}\left\{R_d < \frac{n}{d}s + j_0 \mid S_d = s\right\} \\
& + \sum_{j=j_0}^{\infty} \left(\frac{n}{d}s + j\right)^z \cdot \mathrm{Prob}\left\{R_d = \frac{n}{d}s + j \mid S_d = s\right\} \\
\leq \;& \left((\frac{n}{d})^z s^z + c(z)(\frac{n}{d})^z s^{z-\frac{1}{2}}\log k + c'(z)(\frac{n}{d})^z s^{z/2}\log^z k\right) \cdot \mathrm{Prob}\left\{R_d < \frac{n}{d}s + j_0 \mid S_d = s\right\} \\
& + \sum_{j=j_0}^{\infty} \left((\frac{n}{d})^z s^z + c''(z)(\frac{n}{d})^{z-1}s^{z-1}j + j^z\right) \cdot \mathrm{Prob}\left\{R_d = \frac{n}{d}s + j \mid S_d = s\right\} \\
= \;& \left(\frac{n}{d}\right)^z s^z + \left(c(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c'(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k\right) \cdot \mathrm{Prob}\left\{R_d < \frac{n}{d}s + j_0 \mid S_d = s\right\} \\
& + \sum_{j=j_0}^{\infty} \left(c''(z)(\frac{n}{d})^{z-1}s^{z-1}j + j^z\right) \cdot \mathrm{Prob}\left\{R_d = \frac{n}{d}s + j \mid S_d = s\right\} \\
\leq \;& \left(\frac{n}{d}\right)^z s^z + c(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c'(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k \\
& + c''(z)\left(\frac{n}{d}\right)^{z-1} s^{z-1} \cdot \sum_{j=j_0}^{\infty} j \cdot \mathrm{Prob}\left\{R_d = \frac{n}{d}s + j \mid S_d = s\right\} \\
& + \sum_{j=j_0}^{\infty} j^z \cdot \mathrm{Prob}\left\{R_d = \frac{n}{d}s + j \mid S_d = s\right\} \\
\leq \;& \left(\frac{n}{d}\right)^z s^z + c(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c'(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k \\
& + c''(z)\left(\frac{n}{d}\right)^{z-1} s^{z-1} \cdot \sum_{j=j_0}^{\infty} j \cdot \exp\left(\frac{-jd}{8n\sqrt{s}}\right) \\
& + \sum_{j=j_0}^{\infty} j^z \cdot \exp\left(\frac{-jd}{8n\sqrt{s}}\right) \\
\leq \;& \left(\frac{n}{d}\right)^z s^z + c(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c'(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k \\
& + c''(z)\left(\frac{n}{d}\right)^{z-1} s^{z-1} \cdot \frac{8n\sqrt{s}}{d} \cdot \sum_{t=\frac{j_0 d}{8n\sqrt{s}}}^{\infty} \frac{8n\sqrt{s}}{d} t e^{-t} \\
& + \frac{8n\sqrt{s}}{d} \cdot \sum_{t=\frac{j_0 d}{8n\sqrt{s}}}^{\infty} \left(\frac{8n\sqrt{s}}{d}\right)^z t^z e^{-t}
\end{aligned}
$$

$$
\begin{aligned}
\leq\ & \left(\frac{n}{d}\right)^z s^z + c(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c'(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k \\
& + c^{iv}(z)\left(\frac{n}{d}\right)^{z-1} s^{z-1}\cdot\left(\frac{8n\sqrt{s}}{d}\right)^2\cdot \ln k \exp\left(\frac{-j_0 d}{8n\sqrt{s}}\right) \\
& + c^{v}(z)\left(\frac{8n\sqrt{s}}{d}\right)^{z+1}\cdot \ln^z k \exp\left(\frac{-j_0 d}{8n\sqrt{s}}\right) \\
\leq\ & \left(\frac{n}{d}\right)^z s^z + c(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c'(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k \\
& + c^{iv}(z)\left(\frac{n}{d}\right)^{z-1} s^{z-1}\cdot\left(\frac{8n\sqrt{s}}{d}\right)^2\cdot \ln k \frac{1}{k^4} + c^{v}(z)\left(\frac{8n\sqrt{s}}{d}\right)^{z+1}\cdot \ln^z k\frac{1}{k^4} \\
\leq\ & \left(\frac{n}{d}\right)^z s^z + c(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c'(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k \\
& + \frac{c^{iv}(z)}{k}\left(\frac{n}{d}\right)^{z-1} s^z + \frac{c^{vi}(z)}{k}\left(\frac{n}{d}\right)^{z-1} s^{\frac{z+1}{2}} \\
\leq\ & \left(\frac{n}{d}\right)^z\left(1+\frac{c_3(z)}{k}\frac{d}{n}\right)s^z + c_2(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k + c_{18}(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k\ ,
\end{aligned}
$$

where $c_2(z), c_3(z), c_{18}(z), c(z), c'(z), c''(c), c^{iv}(z), c^v(z)$ and $c^{vi}(z)$ are constants, and we are done. The seventh inequality follows from Lemma 3.6. The eighth inequality follows because by definition of j_0, $\frac{j_0 d}{8n\sqrt{s}} \geq 4z$, and $i^z e^{-i}$ decreases montonically with i for $i \geq z$. The inequality before the last follows since (i) $\frac{n}{d} \leq k$ because $d \geq \frac{n}{k}$ by the lemma assumption, and (ii) $\ln^z(k) \leq c(z)k$. The last inequality follows because $s^{(z+1)/2} \leq s^z$ for $z \geq 1$. ■

Lemma 3.8 *For every $x \geq 1$, there exists a constant $c_{20}(x)$, such that for all intervals I_i and for all values a_i of A_i, if the dth object arrives during I_i, and $d \geq n/\sqrt{k}$, then*

$$
\left(\frac{n}{d}\right)^{x/2} 2^{ix/2} a_i^{x/2}\log^x k \leq c_{20}(x)\sqrt{\frac{n}{d}}2^{i(x-\frac{1}{2})}a_i^{x-\frac{1}{2}}\log k\ .
$$

Proof: If $a_i = 0$, the lemma follows. Thus assume $a_i > 0$. It suffices to prove

$$
\left(\frac{n}{d}\right)^{(x-1)/2}\log^{x-1} k \leq c_{20}(x)2^{i(x-1)/2}a_i^{(x-1)/2}\ ,
$$

where $c_{20}(x)$ is a constant. We distinguish two cases.

Case I: $i = 1$.

$$
\left(\frac{n}{d}\right)^{(x-1)/2}\log^{x-1} k \leq k^{(x-1)/4}\log^{x-1} k \leq ck^{(x-1)/2} \leq c_{20}(x)a_1^{(x-1)/2} \leq c_{20}(x)2^{i(x-1)/2}a_1^{(x-1)/2}\ ,
$$

where $c, c_{20}(x)$ are constants. The first inequality follows since $d \geq n/\sqrt{k}$ by the lemma assumption. The second inequality follows since $\log k \leq ck^{1/4}$ for some constant c. The last inequality follows since by definition $a_1 \geq k/2$. (In particular, if I_1 is opening then $a_1 = k/2$, and otherwise we get that $m' = 0$; thus by definition, $a_1 = p_1 = k2^{-m'} = k$.)

Case II: $i > 1$. We distinguish two cases.

Case II.a: I_i is opening (*i.e.*, $i \leq m'$).

$$\left(\frac{n}{d}\right)^{(x-1)/2} \log^{x-1} k \leq 2^{(x-1)/2} \log^{x-1} k \leq c_{20}(x) a_i^{(x-1)/2} \leq c_{20}(x) 2^{i(x-1)/2} a_i^{(x-1)/2} ,$$

where $c_{20}(x)$ is a constant. The first inequality follows since for $i > 1$, $d > n/2$. The second inequality follows because (i) since I_i is an opening interval, we have (as mentioned in Section 2.) that $p_i \geq p$, and hence also $a_i \geq p$; and (ii) $p > \log^2 k$ by definition.

Case II.b: I_i is closing (*i.e.*, $i > m'$).

$$\begin{aligned}
\left(\frac{n}{d}\right)^{(x-1)/2} \log^{x-1} k &\leq 2^{(x-1)/2} \log^{x-1} k \\
&\leq 2^{(x-1)/2} (ck/p)^{(x-1)/2} \\
&\leq c_{20}(x) \left(2^{\log(k/p)}\right)^{(x-1)/2} \\
&\leq c_{20}(x) 2^{i(x-1)/2} \\
&\leq c_{20}(x) 2^{i(x-1)/2} a_i^{(x-1)/2},
\end{aligned}$$

where c and $c_{20}(x)$ are constants. The first inequality follows since for $i > 1$, $d > n/2$. The second inequality follows since $p = \log^2 k + 64$, and hence $\frac{k}{p} \geq c \log^2 k$ for some constant c. The fourth inequality follows ¿From $i > m'$. The last inequality follows since $a_1 \geq 1$, because (i) as observed in the beginning of the proof of Lemma 3.2, for closing I_i, $A_i > 0$ implies that $A_i \geq 1$, and (ii) by our assumption in the beginning of the proof, $a_i \neq 0$. ∎

Lemma 3.9 *There exist constants $c_4(z)$ and $c_5(z)$ such that for all opening intervals I_i (i.e., $i \leq m'$), for every value a_i of A_i, if the dth object arrives during I_i and $d \geq \frac{n}{k}$, then*

$$\mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\}) \leq 2^{iz} \frac{1}{z+1} a_i^z + c_4(z) 2^{iz} a_i^{z-\frac{1}{2}} \log k + c_5(z) \sqrt{\frac{n}{d}} 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k \ .$$

Proof: Recall that if the dth object arrives during an opening interval I_i, it is accepted electively only if it is one of the top $\lfloor (a_i + 6(z+1)\sqrt{a_i} \log k) 2^i d/n \rfloor$ among the first d. Obviously, S_d is distributed uniformly in $\{1, \ldots, d\}$, so given $\mathrm{NA}_d \cap \{A_i = a_i\}$, S_d takes on any of the values in the set $\{1, \ldots, \lfloor (a_i + 6(z+1)\sqrt{a_i} \log k) 2^i d/n \rfloor\}$ with equal probability of $(\lfloor (a_i + 6(z+1)\sqrt{a_i} \log k) 2^i d/n \rfloor)^{-1}$. Thus,

$$\begin{aligned}
& \mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\}) \\
= & \sum_{s=1}^{\lfloor (a_i + 6(z+1)\sqrt{a_i} \log k) 2^i d/n \rfloor} \mathrm{Prob}\,\{S_d = s \mid \mathrm{NA}_d \cap \{A_i = a_i\}\} \cdot \mathcal{E}(R_d^z \mid S_d = s) \\
\leq & \frac{1}{\lfloor (a_i + 6(z+1)\sqrt{a_i} \log k) \frac{2^i d}{n} \rfloor} \cdot \sum_{s=1}^{\lfloor (a_i + 6(z+1)\sqrt{a_i} \log k) \frac{2^i d}{n} \rfloor} \left(\frac{n}{d}\right)^z \left(1 + \frac{c_3(z)}{k} \frac{d}{n}\right) s^z
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad + c_2(z)\left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}}\log k \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad + c_{18}(z)\left(\frac{n}{d}\right)^z s^{z/2}\log^z k \\
\le\ & \left(\frac{n}{d}\right)^z \cdot \left(1+\frac{c_3(z)}{k}\frac{d}{n}\right)\cdot\frac{1}{z+1}\left((a_i+6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n}\right)^z \\
& + c^{iv}(z)\left(\frac{n}{d}\right)^z \cdot \left(1+\frac{c_3(z)}{k}\frac{d}{n}\right)\cdot\left((a_i+6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n}\right)^{z-1} \\
& + c_2(z)\left(\frac{n}{d}\right)^z \left((a_i+6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n}\right)^{z-\frac{1}{2}}\log k \\
& + c_{18}(z)\left(\frac{n}{d}\right)^z \left((a_i+6(z+1)\sqrt{a_i}\log k)\frac{2^i d}{n}\right)^{z/2}\log^z k \\
\le\ & \left(\frac{n}{d}\right)^z \cdot \left(1+\frac{c_3(z)}{k}\frac{d}{n}\right)\cdot\frac{1}{z+1}a_i^z\left(\frac{2^i d}{n}\right)^z \\
& + c(z)\left(\frac{n}{d}\right)^z \cdot a_i^{z-\frac{1}{2}}\log k\left(\frac{2^i d}{n}\right)^z \\
& + c'(z)\left(\frac{n}{d}\right)^z \cdot a_i^{z/2}\log^z k\left(\frac{2^i d}{n}\right)^z) \\
& + c''(z)\left(\frac{n}{d}\right)^z \left(a_i\frac{2^i d}{n}\right)^{z-\frac{1}{2}}\log k \\
& + c'''(z)\left(\frac{n}{d}\right)^z \left(a_i\frac{2^i d}{n}\right)^{z/2}\log^z k \\
\le\ & \left(1+\frac{c_3(z)}{k}\frac{d}{n}\right)\cdot 2^{iz}\frac{1}{z+1}a_i^z + c(z)2^{iz}a_i^{z-\frac{1}{2}}\log k + c'(z)2^{iz}a_i^{z/2}\log^z k \\
& + c''(z)\sqrt{\frac{n}{d}}2^{i(z-\frac{1}{2})}a_i^{z-\frac{1}{2}}\log k + c'''(z)\left(\frac{n}{d}\right)^{z/2}2^{iz/2}a_i^{z/2}\log^z k\ ,
\end{aligned}
$$

where $c(z), c'(z), c''(z), c'''(z)$ and $c^{iv}(z)$ are constants. The second inequality follows from Lemma 3.7. The fourth inequality follows from $\sqrt{a_i}\log k \le a_i$ because, (i) since I_i is an opening interval, we have, as mentioned in Section 2., that $p_i \ge p$ and hence also $a_i \ge p$ and $k \ge p$, and (2) $p \ge 64+\log^2 k$ by definition.

Thus, to complete the proof, it suffices to show that there exists a constant c such that

$$(\mathrm{i}) \qquad a_i^z/k \le a_i^{z-\frac{1}{2}}\log k\ ,$$

$$(\mathrm{ii}) \qquad a_i^{z/2}\log^z k \le a_i^{z-\frac{1}{2}}\log k\ ,$$

$$\text{and (iii)}\ \left(\frac{n}{d}\right)^{z/2}2^{iz/2}a_i^{z/2}\log^z k \le c\sqrt{\frac{n}{d}}2^{i(z-\frac{1}{2})}a_i^{z-\frac{1}{2}}\log k\ .$$

For the proof of (i), since $a_i \leq k$ we have that $a_i^z/k \leq a_i^{z-1}$. Since, as noted above, $a_i \geq p$ and $k \geq p$, we have that a_i and $k \geq \log^2 k + 64$, and thus $a_i^{z-1} \leq a_i^{z-\frac{1}{2}} \log k$, and (i) follows. (ii) follows from $z \geq 1$ and $a_i \geq \log^2 k$. (iii) follows from Lemma 3.8. ■

Lemma 3.10 *There exists a constant $c_6(z)$, such that for all closing intervals I_i (i.e., $i > m'$), for all values a_i of A_i, if the dth object arrives during I_i, and $d \geq \frac{n}{\sqrt{k}}$, then*

$$\mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\}) \leq c_6(z) 2^{iz} a_i^z \ .$$

Proof: The proof is analogous to that of Lemma 3.9. Recall that if the dth object arrives during a closing I_i, then it is accepted electively if it is one of the top $\lfloor 32(z+1)a_i 2^i d/n \rfloor$ among the first d. Given $\mathrm{NA}_d \cap \{A_i = a_i\}$, R_d is uniformly distributed in the set $\{1, \ldots, \lfloor 32(z+1)a_i 2^i d/n \rfloor\}$. Thus,

$$\begin{aligned}
& \mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\}) \\
= & \sum_{s=1}^{\lfloor 32(z+1)a_i 2^i d/n \rfloor} \mathrm{Prob}\{S_d = s \mid \mathrm{NA}_d \cap \{A_i = a_i\}\} \cdot \mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{S_d = s\}) \\
\leq & \frac{1}{\lfloor 32(z+1)a_i 2^i d/n \rfloor} \cdot \sum_{s=1}^{\lfloor 32(z+1)a_i 2^i d/n \rfloor} \left(\frac{n}{d}\right)^z \left(1 + \frac{c_3(z)}{k}\frac{d}{n}\right) s^z + c_2(z) \left(\frac{n}{d}\right)^z s^{z-\frac{1}{2}} \log k \\
& \qquad + c_{18}(z) \left(\frac{n}{d}\right)^z s^{z/2} \log^z k \\
\leq & \left(\frac{n}{d}\right)^z \left(1 + \frac{c_3(z)}{k}\frac{d}{n}\right) \left(32(z+1)a_i \frac{2^i d}{n}\right)^z + c_2(z) \left(\frac{n}{d}\right)^z \left(32(z+1)a_i \frac{2^i d}{n}\right)^{z-\frac{1}{2}} \log k \\
& + c_{18}(z) \left(\frac{n}{d}\right)^z \left(32(z+1)a_i \frac{2^i d}{n}\right)^{z/2} \log^z k \\
\leq & c(z) 2^{iz} a_i^z + c'(z) \sqrt{\frac{n}{d}} 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k + c''(z) \left(\frac{n}{d}\right)^{z/2} 2^{iz/2} a_i^{z/2} \log^z k \\
\leq & c_6(z) 2^{iz} a_i^z \ ,
\end{aligned}$$

where $c(z), c'(z), c''(z)$ and $c_6(z)$ are constants. The second inequality follows from Lemma 3.7. The last inequality follows from Lemma 3.8 that implies that there is a constant c''' such that $(\frac{n}{d})^{z/2} 2^{iz/2} a_i^{z/2} \log^z k \leq c''' \sqrt{\frac{n}{d}} 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k$ and that $\sqrt{\frac{n}{d}} 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k \leq c''' 2^{iz} a_i^z$. ■

3.3 Expected Sum of Ranks

In this section we show that the expected sum of the zth powers of ranks of the k accepted objects is

$$\frac{1}{z+1} k^{z+1} + O(k^{z+0.5} \log k)$$

(Theorem 3.1). This will follow by adding up the expected sum of the zth powers of ranks of electively accepted objects (Lemmas 3.15), and the expected sum of the zth powers of ranks of mandatorily accepted objects (Lemma 3.17).

In Section 3.3.1 we bound the expected sum of the zth powers of ranks of electively accepted objects. In particular, denote by SUMZ_i the sum of the zth powers of ranks of objects that are accepted electively during I_i. We first use Lemmas 3.9 and 3.10 of Section 3.2 to show that there exist constants $c_7(z)$ and $c_8(z)$ such that if I_i is opening, then

$$\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a_i) \leq 2^{iz}\frac{1}{z+1}a_i^{z+1} + c_7(z)2^{iz}a_i^{z+\frac{1}{2}}\log k$$

(Lemma 3.11); and if I_i is closing, then

$$\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a_i) \leq c_8(z)2^{iz}a_i^{z+1}$$

(Lemma 3.12). Lemma 3.11 is then combined with Part 1 of Lemma 3.3 and with Lemma 3.4, to show that there exist a constant $c_9(z)$ such that if I_i is opening, then

$$\mathcal{E}(\mathrm{SUMZ}_i) \leq 2^{-i}\frac{k^{z+1}}{z+1} + c_9(z)2^{-i/2}k^{z+\frac{1}{2}}\log k$$

(Lemma 3.15). Lemma 3.12 is combined with Part 2 of Lemma 3.3 and with Lemma 3.5, to show that there exists a constant $c_{10}(z)$ such that if I_i is closing, then

$$\mathcal{E}(\mathrm{SUMZ}_i) \leq c_{10}(z)2^{-i}k^{z+1}$$

(Lemma 3.14). The expected sum of the zth powers of ranks of electively accepted objects is obtained by summing up these results over all intervals (Lemma 3.15).

Section 3.3.2 bounds the expected sum of the zth powers of ranks of mandatorily accepted objects. It first shows that if in execution E some object $d \in I_i$ is accepted mandatorily, then the prefix of E prior to the end of I_{i+1}, is not smooth (Lemma 3.16). Lemmas 3.4 and 3.5 of Section 3.1.2, imply that, for each I_i, the probability that a prefix of execution E prior to the end of I_i is not smooth, is at most $c(z)n^{-2.5(z+1)}\log n$, where $c(z)$ is a constant. This bound applies thus also for the probability that objects will be mandatorily accepted in I_i. Lemma 3.17 combines this bound with the facts that the rank of an object never exceeds n, and the number of accepted objects is at most $k \leq n$, to show that the expected sum of the zth powers of ranks of mandatorily accepted objects is $O(k^{z+0.5}\log k)$. The case of $k \geq \frac{1}{2}n$ is handled without the use of Lemma 3.5, since this lemma excludes it.

In addition, the fact that the expected sum of the zth powers of ranks of accepted objects is bounded by a value that does not depend on n will imply that the algorithm accepts the top k objects with positive probability that does not depend on n (Corollary 3.1).

3.3.1 Elective Acceptances

Lemma 3.11 *There exists a constant $c_7(z)$ such that for all opening intervals I_i and for all values a_i of A_i,*

$$\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a_i) \leq 2^{iz}\cdot\frac{a_i^{z+1}}{z+1} + c_7(z)2^{iz}a_i^{z+\frac{1}{2}}\log k\ .$$

Proof: Suppose I_i is opening. Then,

$$\begin{aligned}
& \mathcal{E}(\mathrm{SUMZ}_i | A_i = a_i) \\
= \ & \sum_{d \in I_i} \mathrm{Prob}\{\mathrm{NA}_d \mid A_i = a_i\} \cdot \mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\}) \\
\leq \ & (a_i + 6(z+1)\sqrt{a_i}\log k) \cdot \frac{2^i}{n} \sum_{\substack{d \in I_i \\ d > \lceil n/(8\sqrt{k}) \rceil}} 2^{iz} \cdot \frac{a_i^z}{z+1} + \left(c_4(z)2^{iz} + c_5(z)2^{i(z-\frac{1}{2})}\sqrt{\frac{n}{d}}\right) a_i^{z-\frac{1}{2}} \log k \\
\leq \ & (a_i + 6(z+1)\sqrt{a_i}\log k) \cdot \left(2^{iz} \cdot \frac{a_i^z}{z+1} + c_4(z)2^{iz} a_i^{z-\frac{1}{2}} \log k\right) \\
& +(a_i + 6(z+1)\sqrt{a_i}\log k) \cdot \frac{2^i}{n} \cdot c_5(z) 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k \cdot \sum_{\substack{d \in I_i \\ d > \lceil n/(8\sqrt{k}) \rceil}} \sqrt{\frac{n}{d}} \\
= \ & 2^{iz} \cdot \frac{a_i^{z+1}}{z+1} + \left(c_4(z) + \frac{6(z+1)}{z+1}\right) 2^{iz} a_i^{z+\frac{1}{2}} \log k + 6(z+1)c_4(z)2^{iz} a_i^z \log^2 k \\
& +(a_i + 6(z+1)\sqrt{a_i}\log k) \cdot \frac{2^i}{n} \cdot c_5(z) 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k \cdot \sum_{\substack{d \in I_i \\ d > \lceil n/(8\sqrt{k}) \rceil}} \sqrt{\frac{n}{d}} \\
\leq \ & 2^{iz} \cdot \frac{a_i^{z+1}}{z+1} + (c_4(z) + 6 + c_4(z)) 2^{iz} a_i^{z+\frac{1}{2}} \log k \\
& +(a_i + 6(z+1)\sqrt{a_i}\log k) \cdot \frac{2^i}{n} \cdot c_5(z) 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k \cdot \sum_{\substack{d \in I_i \\ d > \lceil n/(8\sqrt{k}) \rceil}} \sqrt{\frac{n}{d}} \ .
\end{aligned}$$

The second inequality follows since, as explained above, if the dth object arrives during an opening I_i and $d > \lceil n/(8\sqrt{k}) \rceil$, then

$$\mathrm{Prob}(\mathrm{NA}_d \mid A_i = a_i) \leq (a_i + 6(z+1)\sqrt{a_i}\log k)2^i/n \ .$$

The value of $\mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\})$ is given by Lemma 3.9. If $d \leq \lceil n/(8\sqrt{k}) \rceil$, the dth object is not accepted electively. The third inequality follows since $i < m$ (as observed in Section 2., I_m is always closing), and hence $|I_i| = n/2^i$. The last inequality follows from $\sqrt{a_i} \geq \log k$. The latter is true because since I_i is opening, we have, as mentioned in Section 2., $p_i \geq p$ and hence also $a_i \geq p$. However, by definition $p \geq \log^2 k$.

To complete the proof, it suffices to show that there is a constant $c(z)$ such that

$$(a_i + 6(z+1)\sqrt{a_i}\log k) \cdot \frac{2^i}{n} \cdot 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k \cdot \sum_{\substack{d \in I_i \\ d > \lceil n/(8\sqrt{k}) \rceil}} \sqrt{\frac{n}{d}} \leq c(z) 2^{iz} a_i^{z+\frac{1}{2}} \log k \ .$$

For $i > 1$, $\frac{n}{d} \leq 2$, and hence

$$\begin{aligned}
& (a_i + 6(z+1)\sqrt{a_i}\log k) \cdot \frac{2^i}{n} \cdot 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k \cdot \sum_{\substack{d \in I_i \\ d > \lceil n/(8\sqrt{k}) \rceil}} \sqrt{\frac{n}{d}} \\
\leq \ & (a_i + 6(z+1)\sqrt{a_i}\log k) \cdot \frac{2}{n} \cdot 2^{i(z-\frac{1}{2})} a_i^{z-\frac{1}{2}} \log k \cdot \sqrt{2}\frac{n}{2} \\
\leq \ & c(z) 2^{i(z-\frac{1}{2})} a_i^{z+\frac{1}{2}} \log k \ .
\end{aligned}$$

The last inequality follows since $\sqrt{a_i} \geq \log k$, by the same reasoning as above.

For $i = 1$,

$$\begin{aligned}
\sum_{\substack{d \in I_1 \\ d > \lceil n/(8\sqrt{k}) \rceil}} \sqrt{\frac{n}{d}} &\leq \sqrt{\frac{n}{\frac{n}{\sqrt{k}}}} + \sqrt{\frac{n}{\frac{n}{\sqrt{k}}+1}} + \sqrt{\frac{n}{\frac{n}{\sqrt{k}}+2}} + \cdots + \sqrt{\frac{n}{n}} \\
&\leq \frac{n}{\sqrt{k}} \cdot \left(\sqrt{\frac{n}{\frac{n}{\sqrt{k}}}} + \sqrt{\frac{n}{2\frac{n}{\sqrt{k}}}} + \sqrt{\frac{n}{3\frac{n}{\sqrt{k}}}} + \cdots + \sqrt{\frac{n}{\sqrt{k}\frac{n}{\sqrt{k}}}} \right) \\
&\leq \frac{n}{\sqrt{k}} \cdot \sqrt{\sqrt{k}} \cdot \sum_{i=1}^{\sqrt{k}} \sqrt{\frac{1}{i}} \leq cn \ .
\end{aligned}$$

■

Lemma 3.12 *There exists a constant $c_8(z)$ such that for all closing intervals I_i, for all acceptance thresholds a_i computed for I_i,*

$$\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a_i) \leq c_8(z) 2^{iz} a_i^{z+1} \ .$$

Proof: The proof is analogous to that of Lemma 3.11. Suppose I_i is closing. Then,

$$\begin{aligned}
\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a_i) &= \sum_{d \in I_i} \mathrm{Prob}\{\mathrm{NA}_d \mid A_i = a_i\} \cdot \mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\}) \\
&\leq 32(z+1) a_i \frac{2^i}{n} \cdot \sum_{\substack{d \in I_i \\ d > \lceil n/(8\sqrt{k}) \rceil}} c_6(z) 2^{iz} a_i^z \\
&\leq 32(z+1) a_i \frac{2^i}{n} \cdot \lceil \frac{n}{2^i} \rceil \cdot c_6(z) 2^{iz} a_i^z \leq c_8(z) 2^{iz} a_i^{z+1},
\end{aligned}$$

where $c_8(z)$ is a constant. The second inequality follows since, as explained above, if the dth object arrives during a closing interval I_i, and $d > \lceil n/(8\sqrt{k}) \rceil$, then

$$\mathrm{Prob}(\mathrm{NA}_d \mid A_i = a_i) \leq 32(z+1) a_i 2^i / n \ ,$$

and $\mathcal{E}(R_d^z \mid \mathrm{NA}_d \cap \{A_i = a_i\})$ is given by Lemma 3.10; if $d \leq \lceil n/(8\sqrt{k}) \rceil$, the dth object is not accepted electively. The third inequality follows since $|I_i| \leq \lceil n/2^i \rceil$. ■

Lemma 3.11 is combined with Part 1 of Lemma 3.3 and with Lemma 3.4 to show:

Lemma 3.13 *There exists a constant $c_9(z)$ such that for all opening intervals I_i,*

$$\mathcal{E}(\mathrm{SUMZ}_i) \leq 2^{-i} \frac{k^{z+1}}{z+1} + c_9(z) 2^{-i/2} k^{z+\frac{1}{2}} \log k \ .$$

Proof: Suppose I_i is opening. Then,

$$\begin{aligned}
\mathcal{E}(\mathrm{SUMZ}_i) &= \sum_{a=1}^{\infty} \mathrm{Prob}\{A_i = a\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \\
&= \sum_{a=1}^{\infty} \left(\{\mathrm{Prob}\{A_i = a \cap \neg \mathrm{M}_{E_{i-1}}\} + \mathrm{Prob}\{A_i = a \cap \mathrm{M}_{E_{i-1}}\}\right) \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \\
&\leq \mathrm{Prob}\{\neg \mathrm{M}_{E_{i-1}}\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k) + \sum_{a=1}^{\infty} \mathrm{Prob}\{A_i = a \cap \mathrm{M}_{E_{i-1}}\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \ .
\end{aligned}$$

The last inequality follows since $A_i \leq k$.

To complete the proof, we show that there exist constants $c(z)$ and $c'(z)$ such that:

$$(1) \quad \mathrm{Prob}\{\neg \mathrm{M}_{E_{i-1}}\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k) \leq c(z) 2^{-i/2} k^{z+\frac{1}{2}} \log k \ ;$$

$$(2) \quad \sum_{a=1}^{\infty} \mathrm{Prob}\{A_i = a \mid \mathrm{M}_{E_{i-1}}\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \leq 2^{-i} \frac{k^{z+1}}{z+1} + c'(z) 2^{-i/2} k^{z+\frac{1}{2}} \log k \ .$$

For the proof of (1),

$$\begin{aligned}
\mathrm{Prob}\{\neg \mathrm{M}_{E_{i-1}}\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k) &\leq 2^{5(z+1)} n^{-2.5(z+1)} \log n \cdot \left(2^{iz} \frac{1}{z+1} k^{z+1} + c_7(z) 2^{iz} k^{z+\frac{1}{2}} \log k\right) \\
&\leq c(z) n^{-1.5(z+1)} \log n 2^{iz} \\
&\leq c(z) 2^{-i/2} k^{z+\frac{1}{2}} \log k \ ,
\end{aligned}$$

where $c(z)$ is a constant. For the first inequality, Lemma 3.4 is used to bound $\mathrm{Prob}\{\neg \mathrm{M}_{E_{i-1}}\}$, and Lemma 3.11 is used to bound $\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k)$. The second inequality follows ¿From $k \leq n$. The last inequality follows because (i) $2^{i(z+\frac{1}{2})} \leq 2n^{z+\frac{1}{2}}$ since $i \leq \log n + 1$, and (ii) $\log k \geq 1$, because since I_i is opening, we have $p_i \geq p \geq 64$ and hence also $k \geq 64$.

For the proof of (2),

$$\begin{aligned}
&\sum_{a=1}^{\infty} \mathrm{Prob}\{A_i = a \mid \mathrm{M}_{E_{i-1}}\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \\
\leq\ & \mathrm{Prob}\{A_i = 0\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = 0) \\
&+ \sum_{j=1}^{\infty} \mathrm{Prob}\left\{k 2^{-i}(2^{j-1} - 1) < A_i \leq k 2^{-i}(2^j - 1) \mid \mathrm{M}_{E_{i-1}}\right\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k 2^{-i}(2^j - 1)) \\
\leq\ & \sum_{j=1}^{\infty} k^{-5(z+1)(j-1)} \cdot \left(2^{iz} \frac{1}{z+1} (k^{z+1} 2^{-i(z+1)} (2^j - 1)^{z+1}) + c_7(z) 2^{iz} k^{z+\frac{1}{2}} 2^{-i(z+\frac{1}{2})} 2^{j(z+\frac{1}{2})} \log k\right) \\
\leq\ & 2^{-i} \frac{1}{z+1} k^{z+1} \cdot \left(1 + 2^{z+1} \sum_{j=1}^{\infty} k^{-5(z+1)j} 2^{j(z+1)}\right) + c_7(z) 2^{-i/2} k^{z+\frac{1}{2}} \log k \cdot \left(1 + 2^{z+1} \sum_{j=1}^{\infty} k^{-5(z+1)j} 2^{j(z+1)}\right) \\
\leq\ & 2^{-i} \frac{1}{z+1} k^{z+1} \cdot \left(1 + 2^{z+1} \sum_{j=1}^{\infty} ((k/2)^{z+1})^{-j}\right) + c_7(z) 2^{-i/2} k^{z+\frac{1}{2}} \log k \cdot \left(1 + 2^{z+1} \sum_{j=1}^{\infty} ((k/2)^{z+1})^{-j}\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \quad 2^{-i}\frac{k^{z+1}}{z+1} + c''(z)2^{-i}k^{z} + c'''(z)2^{-i/2}k^{z+\frac{1}{2}}\log k \\
&\leq \quad 2^{-i}\frac{k^{z+1}}{z+1} + c'(z)2^{-i/2}k^{z+\frac{1}{2}}\log k\ ,
\end{aligned}
$$

where $c'(z), c''(z)$ and $c'''(z)$ are constants. For the second inequality, Lemma 3.3 is used to bound $\mathrm{Prob}\left\{k2^{-i}(2^{j-1}-1) < A_i \leq k2^{-i}(2^{j}-1) \mid \mathrm{M}_{E_{i-1}}\right\}$, and Lemma 3.11 is used to bound $\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k2^{-i}(2^j-1))$. The last two inequalities follow because, since I_i is opening, we have $p_i \geq p$ and hence also $k \geq p$. Thus $k \geq 64$ and $\log k \geq 1$. ■

Analogously,

Lemma 3.14 *If $n \geq 16$, then there exists a constant $c_{10}(z)$ such that for any closing interval I_i,*

$$\mathcal{E}(\mathrm{SUM}_i) \leq c_{10}(z)2^{-i}k^{z+1}\ .$$

Proof: The proof is analogous to that of Lemma 3.13. Suppose I_i is closing. Then,

$$
\begin{aligned}
\mathcal{E}(\mathrm{SUMZ}_i) \quad &= \quad \sum_{a=1}^{\infty}\mathrm{Prob}\{A_i = a\}\cdot\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \\
&= \quad \sum_{a=1}^{\infty}\left(\mathrm{Prob}\{A_i = a \cap \neg\mathrm{M}_{E_{i-1}}\} + \mathrm{Prob}\{A_i = a \cap \mathrm{M}_{E_{i-1}}\}\right)\cdot\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \\
&\leq \quad \mathrm{Prob}\{\neg\mathrm{M}_{E_{i-1}}\}\cdot\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k) + \sum_{a=1}^{\infty}\mathrm{Prob}\{A_i = a \mid \mathrm{M}_{E_{i-1}}\}\ .
\end{aligned}
$$

The last inequality follows since $A_i \leq k$.

To complete the proof, we prove that there exist constants $c(z)$ $c'(z)$ such that:

$$(1)\quad \mathrm{Prob}\{\neg\mathrm{M}_{E_{i-1}}\}\cdot\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k) \leq c(z)2^{-i}k^{z+1}\ ;$$

$$(2)\quad \sum_{a=1}^{\infty}\mathrm{Prob}\{A_i = a \mid \mathrm{M}_{E_{i-1}}\}\cdot\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \leq c'(z)2^{-i}k^{z+1}\ .$$

For the proof of (1), we distinguish between two cases.

Case I: $k \leq \frac{1}{2}n$.

$$\mathrm{Prob}\{\neg\mathrm{M}_{E_{i-1}}\}\cdot\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k) \leq 2^{10(z+1)}n^{-2.5(z+1)}\log n\cdot c_8(z)2^{iz}k^{z+1} \leq c(z)2^{-i}k^{z+1}\ ,$$

where $c(z)$ is a constant. The first inequality follows from Lemmas 3.5 and 3.12. The last inequality follows because (i) $k \leq n$, and (ii) $2^{i(z+1)} \leq 2^{z+1}n^(z+1)$ since $i \leq \log n + 1$.

Case II: $k \geq \frac{1}{2}n$.

$$
\begin{aligned}
\mathrm{Prob}\{\neg\mathrm{M}_{E_{i-1}}\}\cdot\mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k) \quad &\leq \quad \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k) \leq \lceil n2^{-i}\rceil\cdot n^{z} \leq 2(2k)2^{-i}\cdot(2k)^{z} \\
&\leq \quad c(z)2^{-i}k^{z+1}
\end{aligned}
$$

where $c(z)$ is a constant. For the second inequality observe that since $|I_i| = \lceil n2^{-i} \rceil$, and the maximum rank of any object is n, we have that the sum of the zth powers of the ranks of objects accepted during I_i is bounded above by $\lceil n2^{-i} \rceil n^z$. The third inequality follows since by our assumption $k \geq \frac{1}{2}n$.

For the proof of (2),

$$
\begin{aligned}
& \sum_{a=1}^{\infty} \operatorname{Prob}\{A_i = a \mid \mathrm{M}_{E_{i-1}}\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = a) \\
\leq \; & \operatorname{Prob}\{A_i = 0\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = 0) \\
& + \sum_{j=1}^{\log k+1} \operatorname{Prob}\left\{k2^{-m'}(2^{j-1}-1) < A_i \leq k2^{-m'}(2^j - 1) \mid \mathrm{M}_{E_{i-1}}\right\} \cdot \mathcal{E}(\mathrm{SUMZ}_i \mid A_i = k2^{-m'}(2^j-1)) \\
\leq \; & \sum_{j=1}^{\infty} k^{-5(z+1)(j-1)} (2^{-5(z+1)})^{i-m'-1} \cdot c_8(z) 2^{iz} \cdot k^{z+1} 2^{-m'(z+1)} (2^j - 1)^{z+1} \\
\leq \; & c'(z) 2^{-(4z+5)i} 2^{(4z+4)m'} k^{z+1} \\
= \; & c'(z) 2^{-(4z+4)(i-m')} 2^{-i} k^{z+1} \\
\leq \; & c'(z) \cdot 2^{-i} k^{z+1} \ ,
\end{aligned}
$$

where $c'(z)$ is a constant. The second inequality follows from Lemmas 3.3 and 3.12. The third inequality follows since if $k \leq 4$ then the sum is clearly constant, and if $k \geq 4$ then $k^{-5j}(2^j-1) \leq 2^{-j}$, and the sum converges. The last inequality follows since $i > m'$ by definition of closing intervals. ∎

The following lemma completes the proof of the upper bound on the sum of the ranks of the electively accepted objects. It sums up the expected sum of ranks of electively accepted objects over all intervals.

Lemma 3.15

$$\sum_{i=1}^{m} \mathcal{E}(\mathrm{SUMZ}_i) \leq \frac{k^{z+1}}{z+1} + O(k^{z+0.5} \log k) \ .$$

Proof: If $n \leq 16$, the assertion is immediate. For $n > 16$,

$$
\begin{aligned}
\sum_{i=1}^{m} \mathcal{E}(\mathrm{SUMZ}_i) & \leq \sum_{i=1}^{m'} \left(2^{-i} \frac{k^{z+1}}{z+1} + c_9(z) 2^{-i/2} k^{z+\frac{1}{2}} \log k \right) + \sum_{i=m'+1}^{m} c_{10}(z) 2^{-i} k^{z+1} \\
& \leq \frac{k^{z+1}}{z+1} (1 - 2^{-\log(k/p)}) + c_9(z) 2 k^{z+\frac{1}{2}} \log k + c_{10}(z) 2^{-\log(k/p)} k^{z+1} \\
& \leq \frac{k^{z+1}}{z+1} (1 - 2^{-\log(k/p)}) + c_9(z) 2 k^{z+\frac{1}{2}} \log k + c_{10}(z) k^z (\log^2 k + 64) \\
& \leq \frac{k^{z+1}}{z+1} + c_{11}(z) k^{z+0.5} (\log k + 1) \\
& = \frac{k^{z+1}}{z+1} + O(k^{z+0.5} \log k) \ ,
\end{aligned}
$$

where $c_{11}(z)$ is a constant. The first inequality follows from Lemmas 3.13 and 3.14. The third inequality follows since by definition $p = \log^2 k + 64$. ∎

3.3.2 Mandatory Acceptances

This section bounds the expected sum of mandatorily accepted objects. We first observe:

Lemma 3.16 *If the dth object is mandatorily accepted in execution E during I_i, then $\neg \mathrm{M}_{E_{i+1}}$.*

Proof: First we claim that $i < m - 1$. For, as noted in Section 2., I_{m-1}, I_m are closing. Thus, if there is still an empty slot in E by the time the dth object arrives during I_j ($j = m-1, m$), then $a_j > 0$. As observed at the beginning of the proof of lemma 3.2, in this case $a_j \geq 1$, since for closing I_j, $A_j > 0$ implies $A_j > 1$. Hence, this object will be electively accepted if it is among the top $\lfloor 32(z+1)a_j 2^j d/n \rfloor$ objects seen so far, and hence, it will be electively accepted if it is among the top

$$\lfloor 32(z+1)a_j 2^j d/n \rfloor \geq \lfloor 32(z+1)2^{\log n}(n-1)/n \rfloor \geq \lfloor 32(z+1)n(n-1)/n \rfloor \geq n \ ,$$

objects seen so far. Thus, it will be electively accepted and hence not mandatorily accepted.

Assume the dth object is mandatorily accepted in E during I_i. By definition of mandatorily accepted, this implies that the number of open slots just before the dth object's arrival equals to the total number of objects remaining to be seen, *i.e.*, $n - d + 1$. Since, as shown above, $i < m-1$, it follows that I_{i+1} exists. Thus, at the beginning of I_{i+1}, the number of open slots equals to the total number of objects remaining to be seen. Moreover, since $i < m - 1$, the number of objects that remain to be seen just before the beginning of I_{i+1} is exactly $2|I_{i+1}|$. Thus,

$$\sum_{j=1}^{i} Q_j = k - 2|I_{i+1}| \ .$$

Therefore,

$$D_i = \sum_{j=1}^{i} p_j - \sum_{j=1}^{i} Q_j = \sum_{j=1}^{i} p_j - (k - 2|I_{i+1}|) \ .$$

But

$$a_{i+1} = D_i + p_{i+1} = \sum_{j=1}^{i} p_j - (k - 2|I_{i+1}|) + p_{i+1} = 2|I_{i+1}| - \sum_{j=i+2}^{m} p_j \ .$$

To complete the proof, it suffices to show that $2|I_{i+1}| - \sum_{j=i+2}^{m} p_j > |I_{i+1}|$. We distinguish beween two cases.

Case I: $i \geq m'$. In this case, $\sum_{j=i+2}^{m} p_j = 0$, and hence

$$2|I_{i+1}| - \sum_{j=i+2}^{m} p_j = 2|I_{i+1}| > |I_{i+1}| \ .$$

Case II: $i < m'$. In this case, it follows directly from the definition of p_j that $\sum_{j=i+2}^{m} p_j = p_{i+1} = k2^{-i-1}$. If $k = n$, then clearly all objects are electively accepted and none is mandatorily accepted. Thus, assume $k < n$. Then $k2^{-i-1} < n2^{-i-1} = |I_{i+1}|$. Thus,

$$2|I_{i+1}| - \sum_{j=i+2}^{m} p_j > 2|I_{i+1}| - |I_{i+1}| = |I_{i+1} \ .$$

■

Denote by SUMDZ_i the sum of the zth powers of ranks of objects that are accepted mandatorily during I_i.

Lemma 3.17 *There exist constants $c_{21}(z)$ and $c_{22}(z)$ such that*

$$\sum_{i=1}^{m} \mathcal{E}(\mathrm{SUMDZ}_i) \leq c_{21}(z) k^{z+\frac{1}{2}} \log k + c_{22}(z) \ .$$

Proof: Again, if $n \leq 16$, the assertion is immediate. For $n > 16$, we argue as follows. The number of accepted objects in I_i at most $|I_i|$, and the rank of any object is of course not greater than n. Thus,

$$\begin{aligned}
\sum_{i=1}^{m} \mathcal{E}(\mathrm{SUMDZ}_i) &= \sum_{i=1}^{m-1} \mathcal{E}(\mathrm{SUMDZ}_i) \\
&\leq \sum_{i=1}^{m-1} \mathrm{Prob}\{\neg \mathrm{M}_{E_{i+1}}\} \cdot |I_i| \cdot n^z \\
&\leq \sum_{i=1}^{m'-1} \mathrm{Prob}\{\neg \mathrm{M}_{E_{i+1}}\} \cdot n2^{-i} \cdot n^z + \sum_{i=m'-1}^{m-1} \mathrm{Prob}\{\neg \mathrm{M}_{E_{i+1}}\} \cdot n2^{-i} \cdot n^z \ .
\end{aligned}$$

The first inequality follows since as shown in the beginning of the proof of Lemma 3.16, no object is accepted mandatorily in I_m. The second inequality follows from Lemma 3.16. The third inequality follows since for $i < m$, $|I_i| = n2^{-i}$.

To complete the proof, we show that there exist constants $c(z)$ and $c'(z)$ and $c''(z)$ such that:

$$(1) \quad \sum_{i=1}^{m'-1} \mathrm{Prob}\{\neg \mathrm{M}_{E_{i+1}}\} \cdot n2^{-i} \cdot n^z \leq c(z) \ ;$$

$$(2) \quad \sum_{i=m'-1}^{m-1} \mathrm{Prob}\{\neg \mathrm{M}_{E_{i+1}}\} \cdot n2^{-i} \cdot n^z \leq c'(z) k^{z+\frac{1}{2}} \log k + c''(z) \ .$$

For the proof of (1),

$$\begin{aligned}
\sum_{i=1}^{m'-1} \mathrm{Prob}\{\neg \mathrm{M}_{E_{i+1}}\} \cdot n2^{-i} \cdot n^z &\leq \sum_{i=1}^{m'-1} c(z) n^{-2.5(z+1)} \log n \cdot n2^{-i} \cdot n^z \\
&\leq \sum_{i=1}^{m'-1} c(z) n^{-1.5(z+1)} \log n \cdot 2^{-i} \leq c(z) \ ,
\end{aligned}$$

where $c(z)$ is a constant. The first inequality follows from Lemma 3.4.

For the proof of (2), we distinguish between two cases:

Case I: $k \geq \frac{1}{2}n$.

$$\begin{aligned}
\sum_{i=m'-1}^{m-1} \text{Prob}\{\neg \text{M}_{E_{i+1}}\} \cdot n2^{-i} \cdot n^z &\leq \sum_{i=m'-1}^{m-1} n2^{-i} \cdot n^z \\
&\leq (m-1) \cdot n2^{-m'+1} \cdot n^z \\
&= \log n \cdot n2^{-m'+1} \cdot n^z \\
&\leq \log(2k) \cdot (2k)2^{-m'+1} \cdot (2k)^z \\
&\leq c'''(z)k^{z+1}2^{-\log(k/p)} \log k \\
&= c'''(z)k^{z+1}(p/k) \log k \\
&= c'''(z)k^z(\log^2 k + 64) \log k \\
&\leq c'(z)k^{z+\frac{1}{2}} \log k + c''(z) \ ,
\end{aligned}$$

where $c'''(z), c'(z)$, and $c''(z)$ are constants. The fourth inequality follows from the assumption $k \geq \frac{1}{2}n$. The fifth inequality follows since by definition $m' = \lfloor \log(k/p) \rfloor$. The inequality before the last follows since $p = \log^2 k + 64$ by definition.

Case II: $k \leq \frac{1}{2}n$.

$$\begin{aligned}
\sum_{i=m'-1}^{m-1} \text{Prob}\{\neg \text{M}_{E_{i+1}}\} \cdot n2^{-i} \cdot n^z &\leq \sum_{i=m'-1}^{m-1} c'''(z)n^{-2.5(z+1)} \log n \cdot n2^{-i} \cdot n^z \\
&\leq (m-1) \cdot c'''(z)n^{-1.5(z+1)}2^{-m'+1} \log n \\
&= \log n \cdot c(z)n^{-1.5(z+1)}2^{-m'+1} \log n \\
&\leq c''(z) \ ,
\end{aligned}$$

where $c''(z)$ and $c'''(z)$ are constants. The first inequality follows from Lemmas 3.4 and 3.5. ∎

Lemmas 3.15 and 3.17 imply:

Theorem 3.1 *The expected sum of ranks of accepted objects is at most*

$$\frac{1}{z+1}k^{z+1} + O(k^{z+0.5} \log k) \ .$$

Corollary 3.1 *Algorithm Select accepts the best k objects with positive probability that depends only on k and z.*

Proof: Theorem 3.1 implies that the expected sum of the zth powers of ranks of accepted objects is bounded by a value that is independent of n. Thus, there is some value r that does not depend on n such that with probability $\geq 1/2$, all accepted objects are of rank $\leq r$. Clearly, the probablity of acceptance decreases monotonically with the object's rank. Therefore, the probability that the k accepted objects are the best k objects is at least $\frac{1}{2}/\binom{r}{k}$. ∎

4. Trade-Off between Small Expected Rank and Large Probability of Accepting the Best

Theorem 4.1 *Let p_0 be the maximum possible probability of selecting the best object. There is a $c > 0$ so that for all $\epsilon > 0$ and all sufficiently large n, if A is an algorithm that selects one of n objects, and the probability p_A that A selects the best one is greater than $p_0 - \epsilon$, then the expected rank of the selected object is at least c/ϵ.*

Proof: Suppose that contrary to our assertion there is an algorithm A that selects the best object with probability of at least $p_0 - \epsilon$ and yet the expected value of the rank of the selected object is less than c/ϵ. Starting from A, we construct another algorithm R so that R selects the best object with a probability $> p_0$.

Denote by OPT the following algorithm: Let n/e objects pass, and then accept the first object that is better than anyone seen so far. If no object was accepted by the time the last object arrives, accept the last object. For n sufficiently large, this algorithm accepts the best object with the highest possible probability, and hence with probability p_0 [8].[3]

We define R by modifying A. The definition will depend on parameters $c_1 > d > 0$. We will assume that d is a sufficiently large absolute constant and c_1 is sufficiently large with respect to d. R will accept an object if at least one of the following conditions is satisfied:

(i) A accepts the object after time n/d and by time $n - c_1\epsilon n$ and the object is better than anybody else seen earlier;

(ii) OPT accepts the object whereas A accepted earlier somebody who, at the time of acceptance, was known not to be the best one (that is there was a better one before);

(iii) OPT accepts the object and A has already accepted somebody by time n/d;

(iv) the object comes after time $n - c_1\epsilon n$, it is better than anybody else seen before and R has not yet accepted anybody based on the rules (1), (2), (3);

(v) the object is the nth object and R has not accepted yet any object.

Notation: Denote by BA, BR, and BOPT the events in which A, R and OPT, repectively, accept the best object. Denote by B1, B2, and B3 the events in which the best object appears in the intervals $[1, n/d]$, $(n/d, t_0 = n - c_1\epsilon]$, and $(t_0, n]$, respectively. Denote by IA1, IA2 and IA3 the events in which A makes a selection in the intervals $[1, n/d]$, $(n/d, t_0 = n - c_1\epsilon]$, and $(t_0, n]$, respectively.

We distinguish between two cases.

Case I: Prob{IA1} $\geq 3\epsilon/p_0$.

[3] In fact, $r = [(n - \frac{1}{2})e^{-1} + \frac{1}{2}]$ is a better approximation to r than ne^{-1} although the difference is never more than 1 [6]. We ignore this difference for the sake of simplicity.

Claim 4.1

$$\text{Prob}\{\text{BR} \mid \mathit{IA1}\} \geq P\{\text{BA} \mid \mathit{IA1}\} + p_0/2.$$

Proof: Suppose that A made a selection by time n/d. According to rule (3), in this case R will accept an object that arrives after time n/d if and only if OPT accepts this object. By choosing d sufficiently large, we have that objects are accepted by OPT only after time n/d. Thus, if A made a selection by time n/d, R will accept the object if and only if OPT accepts it. Thus,

$$\text{Prob}\{\text{BR} \mid \text{IA1}\} = \text{Prob}\{\text{BOPT} \mid \text{IA1}\} = \text{Prob}\{\text{BOPT}\} \geq p_0.$$

The second inequality follows since the probability that OPT accepts the best object is independent of the order of arrival of the first n/d objects, and hence independent of whether or not A makes a selection by time n/d. On the other hand,

$$\text{Prob}\{\text{BA} \mid \text{IA1}\} \leq \text{Prob}\{\text{B1}\} \leq 1/d.$$

Thus, by choosing d to be sufficiently large the claim follows. ■

Claim 4.2

$$\text{Prob}\{\text{BR} \mid \mathit{IA2}\} \geq \text{Prob}\{\text{BA} \mid \mathit{IA2}\}.$$

Proof: The claim follows immediately from the fact that if A picks the best object between n/d and t_0, then this object must be the best seen so far, and hence by rule (1), R picks the same object. ■

Claim 4.3

$$\text{Prob}\{\text{BR} \mid \mathit{IA3}\} \geq \text{Prob}\{\text{BA} \mid \mathit{IA3}\}.$$

Proof: If IA3 holds then neither A nor R have accepted anybody till time t_0. Let X be the event when A chooses no later than R. By the definition of R we have that if $X \cap \text{IA3}$ holds then either A accepts an object that already at the moment of acceptance is known not to be the best, or A and R accept the same object. Thus,

$$\text{Prob}\{\text{BR} \mid \text{IA3} \cap X\} \geq \text{Prob}\{\text{BA} \mid \text{IA3} \cap X\} \ .$$

To complete the proof, it suffices to show that

$$\text{Prob}\{\text{BR} \mid \text{IA3} \cap \neg X\} \geq \text{Prob}\{\text{BA} \mid \text{IA3} \cap \neg X\} \ .$$

Suppose that $\text{IA3} \cap \neg X$ holds and R accepts an object at some time $t > t_0$. By definition, A has not accepted anybody yet, and the object accepted by R at t is better than anyone else seen earlier. Thus, if a better object than the one accepted by R arrives after time t, this means that the best object arrives after time t. Since the objects arrive in a random order, the rank of each dth arriving object within the set of first d is distributed uniformly. Hence, the probability that the best object will arrive after time t is at most $(n-t)/n \leq c_1 \epsilon n$. Notice that this probability is independent of the ordering of the first t objects, and hence is independent of the fact that R has accepted the tth object. Therefore the probability that the object accepted by R is indeed the best object is at least $1 - c_1 \epsilon n$, while the probability that A accepts the best one later is smaller than $c_1 \epsilon n$. Thus, for any fixed choice of t and fixed order of the first t objects (with the property $\text{IA3} \cap \neg X$), the probability of BR is larger than BA, and hence $\text{Prob}\{\text{BR} \mid \text{IA3} \cap \neg X\} \geq \text{Prob}\{\text{BR} \mid \text{IA3} \cap \neg X\}$. ■

Now we can complete the proof of Case I:

$$
\begin{aligned}
& \text{Prob}\{\text{BR}\} \\
= \; & \text{Prob}\{\text{BR} \mid \text{IA1}\} \cdot \text{Prob}\{\text{IA1}\} \\
& + \text{Prob}\{\text{BR} \mid \text{IA2}\} \cdot \text{Prob}\{\text{IA2}\} \\
& + \text{Prob}\{\text{BR} \mid \text{IA3}\} \cdot \text{Prob}\{\text{IA3}\} \\
\geq \; & (\text{Prob}\{\text{BA} \mid \text{IA1}\} + p_0/2) \cdot \text{Prob}\{\text{IA1}\} \\
& + \text{Prob}\{\text{BA} \mid \text{IA2}\} \cdot \text{Prob}\{\text{IA2}\} \\
& + \text{Prob}\{\text{BA} \mid \text{IA3}\} \cdot \text{Prob}\{\text{IA3}\} \\
= \; & \text{Prob}\{\text{BA}\} + (p_0/2) \cdot \text{Prob}\{\text{IA1}\} \\
\geq \; & p_0 - \epsilon + (p_0/2) \cdot 3\epsilon/p_0 = p_0 + \epsilon/2 \; .
\end{aligned}
$$

The second inequality follows from Claims 4.1, 4.2 and 4.3. The fourth inequality follows from (i) $\text{Prob}\{\text{BA}\} \geq p_0 - \epsilon$ by the theorem assumption and (ii) $\text{Prob}\{\text{IA1}\} \geq 3\epsilon/p_0$ by Case I assumption.

Case II: $\text{Prob}\{\text{IA1}\} < 3\epsilon/p_0$.

Denote by BR1, BR2, and BR3 the events when R picks the best object and its selections are in the interval $[1, n/d], (n/d, t_0]$ and $(t_0, n]$, respectively. Denote by BA1, BA2, and BA3 the corresponding events for A.

Since by the assumption of this case $\text{Prob}\{\text{IA1}\} < 3\epsilon/p_0$, we have

$$(1) \qquad \text{Prob}\{\text{BA1}\} < 3\epsilon/p_0 \; .$$

If A picks the best object between n/d and t_0, then this object must be the best seen so far, and hence by rule (1), R picks the same object. Thus

$$(2) \qquad \text{Prob}\{\text{BR2}\} \geq \text{Prob}\{\text{BA2}\} \; .$$

By choosing d sufficiently large, we have that objects are accepted by OPT only after time n/d. Observe that in that case, if the second best comes by time n/d and the best comes after time t_0, then R accepts the best object. The probability that the second best object arrives by time n/d is $1/d$, and the conditional probability that the best object comes after time t_0, given that the second best comes by time n/d, is at least $c_1\epsilon$. It thus follows:

$$(3) \qquad \text{Prob}\{\text{BR3}\} \geq c_1\epsilon/d \; .$$

For bounding $\text{Prob}\{\text{BA3}\}$, we first use the assumption that the expected rank of the object selected by A is less than c/ϵ, to show:

Claim 4.4

$$\text{Prob}\{\mathit{IA3}\} \leq 1/(2d) \; .$$

Proof: Each of the $1/(10dc_1\epsilon)$ objects with a rank smaller than $1/(10dc_1\epsilon)$ arrives after time $t_0 = n - c_1\epsilon n$ with probability of at most $c_1\epsilon$. Therefore, with probability of at least $1 - 1/(10d)$, all objects that arrive after time t_0 are of rank larger than $1/(10dc_1\epsilon)$. Hence, if the probability of IA3 had been greater than $1/(2d)$, then the expected value of the rank would have been larger than c'/ϵ for some absolute constant $c' > 0$. Take the c of the theorem to be equal to c', and we get a contradiction to the assumption that the expected rank of the selected object is at most c/ϵ. ■

Let B3 denote the event in which the best object arrives in interval $(t_0, n]$. Then $\text{Prob}\{\text{BA3}\} \leq \text{Prob}\{\text{IA3}\} \cdot \text{Prob}\{\text{B3} \mid \text{IA3}\}$. But B3 is independent of the order of arrival of the first t_0 objects and hence independent on whether or not A has accepted an object by time t_0. Thus, Claim 4.4 implies that $\text{Prob}\{\text{IA3}\} \cdot \text{Prob}\{\text{B3} \mid \text{IA3}\} = \text{Prob}\{\text{IA3}\} \cdot \text{Prob}\{\text{B3}\} \leq \frac{1}{2d} \cdot c_1\epsilon$. Thus,

$$\text{Prob}\{\text{BA3}\} \leq c_1\epsilon/(2d) \ . \tag{4}$$

Equations (1) to (4) imply

$$\begin{aligned}
& \text{Prob}\{\text{BR}\} - \text{Prob}\{\text{BA}\} \\
= \ & \text{Prob}\{\text{BR1}\} - \text{Prob}\{\text{BA1}\} + \text{Prob}\{\text{BR2}\} - \text{Prob}\{\text{BA2}\} + \text{Prob}\{\text{BR3}\} - \text{Prob}\{\text{BA3}\} \\
\geq \ & -3\epsilon/p_0 + c_1\epsilon/d - c_1\epsilon/(2d) \\
= \ & c_1\epsilon/(2d) - 3\epsilon/p_0 \geq 2\epsilon \ .
\end{aligned}$$

(The last inequality follows from our assumption that c_1 is sufficiently large with respect to d.) Therefore

$$\text{Prob}\{\text{BR}\} \geq \text{Prob}\{\text{BA}\} + 2\epsilon \geq p_0 - \epsilon + 2\epsilon > p_0 \ .$$

■

5. Deterministic Arrivals

In this section we consider the case where the order of arrivals is not random but is determined by an adversary that knows the algorithm, *i.e.,* an oblivious adversary. We show that against such an adversary, no algorithm can obtain an expected sum of the zth powers of ranks of selected items that is less than $kn^z/2^{z+1}$. In particular, this expected sum tends to infinity with n. This lower bound holds also for randomized algorithms.

Given an algorithm A, we construct a sequence over which the expected sum of zth powers of ranks of objects selected by A is at least $kn^z/2^{z+1}$. Without loss of generality assume that n is even. Let p be the expected number of acceptances prior to the time the $(n/2)$th object is seen (inclusive), in case the ranks of the arriving objects are monotonically increasing. If $p \leq k/2$, then construct a sequence of objects such that the best $n/2$ objects are the first to arrive, and they arrive in order of increasing rank. Clearly, the expected number of objects accepted during the second half is at least $k/2$, and each such object is of rank larger than $n/2$. It thus follows that the average rank of accepted objects is at least $(k/2) \cdot (n/2)^z = kn^z/2^{z+1}$. The case of $p > k/2$ is analogous. The sequence is constructed so that the worst $n/2$ objects are the first to arrive, and they arrive in order of increasing rank. It again follows that the expected sum of z-powers of ranks of accepted objects is at least $kn^z/2^{z+1}$.

Acknowledgements

We are indebted to Eugen Dynkin, John Preater, Yossi Rinott, Mike Saks, Steve Samuels, and Robert Vanderbei for helpful references.

References

[1] M. Ajtai, N. Megiddo, and O. Waarts. Improved algorithms and analysis for secretary problems and generalizations. Manuscript, available from the authors.

[2] Y. S. Chow, H. Robbins, S. Moriguti, and S. M. Samuels. Optimal selection based on relative rank (the "secretary problem"). *Israel J. Math.* 2, 81–90, 1964.

[3] E. B. Dynkin. The optimum choice of the instant for stopping a Markov process. . *Soviet Math.* 4, 627–629, 1963.

[4] T. S. Ferguson. Who solved the secretary problem? *Statistical Science* 4, 282–296, 1989.

[5] P. R. Freeman. The secretary problem and its extensions: A review. *International Statistical Review* 51, 189–206, 1983.

[6] J. Gilbert and F. Mosteller. Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.* 61, 35–73, 1966.

[7] K. S. Glasser, R. Holzagar, and A. Barron. The d Choice secretary problem. *Commun. Statist. Sequential Analysis* 2(3), 177–179, 1983.

[8] D. V. Lindley. Dynamic programming and decision theory. *Appl. Statist.* 10, 39–52, 1961.

[9] M. Henke. Sequentialle Auswahlprobleme bei Unsicherheit. Meisenheim: Anton Hain Verlag, 1970.

[10] M. L. Nikolaev. A generalization of the best choice problem. *Theory Probability Appl.* 22, 187–190, 1977.

[11] J. Preater. On multiple choice secretary problems. *Mathematics of Operations Research* 19(3), 597–602, August 1994.

[12] H. Rubin and S. M. Samuels. The finite memory secretary problem. *Ann. Probab.* 5, 627–635, 1977.

[13] S. M. Samuels. Optimal counter strategies for the secretary problem. *IMS Bull.* 7, 93 (Abstract 78t-43).

[14] S. M. Samuels. Secretary problems. In *Handbook of sequential analysis.* B. K. Gosh and P. K. Sen, eds. Marcel Dekker, New York. 381–405 (chapter 16), 1991.

[15] S. M. Samuels. Secretary problems as a source of benchmark sounds. *Stochastic Inequalities IMS Lecture Notes - Monograph Series,* 22, 371–387, 1993.

[16] D. D. Sleator and R. E. Tarjan. Amortized efficiency of list updates and paging rules. *Communication of ACM* 28(2), 202–208, 1985.

[17] R. J. Vanderbei. The optimal choice of a subset of a population. *Mathematics of Operations Research* 5(4), 481–486, November 1980.