Linear Waste of Best Fit Bin Packing on Skewed Distributions

Claire Kenyon*
Université Paris-Sud
kenyon@lri.fr

Michael Mitzenmacher†
Harvard University
michaelm@eecs.harvard.edu

Abstract

We prove that Best Fit bin packing has linear waste on the discrete distribution $U\{j,k\}$ (where items are drawn uniformly from the set $\{1/k, 2/3, \cdots, j/k\}$) for sufficiently large k when $j = \alpha k$ and $0.66 \leq \alpha < 2/3$. Our results extend to continuous skewed distributions, where items are drawn uniformly on $[0,a]$, for $0.66 \leq a < 2/3$. This implies that the expected asymptotic performance ratio of Best Fit is strictly greater than 1 for these distributions.

1 Introduction

1.1 Background and results

In the bin packing problem, one is given a sequence $L_n = a_1, \ldots, a_n \in (0,1]$ of items and asked to pack them into bins of unit capacity so as to minimize the number of bins used. This problem is well known to be NP-hard, and a vast literature has developed around the design and analysis of efficient approximation algorithms for it. The most widely studied among these is the Best Fit algorithm, in which the items are packed on-line, with each successive item going into a partially filled bin with the smallest residual capacity large enough to accommodate it; if no such bin exists, a new bin is started.

Best Fit was first analyzed in the worst case in 1974 in [7], where it was proved that the number of bins used is always within a factor 1.7 of optimal, so that the asymptotic performance ratio of Best Fit is 1.7. For the uniform distribution on $[0,1]$, the expected asymptotic performance ratio of Best Fit is 1, and more precisely the expected waste of Best Fit is $\Theta(n^{1/2}\log^{3/4} n)$ [14, 10]. The waste is the total unused space, i.e. the difference between the number of bins used and the sum of the sizes of all the items.

To better understand Best Fit, researchers then turned their attention to the skewed distributions $U(0,a)$ where the item sizes are independent random variables uniform over the interval $[0,a]$ for some a strictly less than one.[1] For these distributions, the optimal packing is perfect in the sense that $\lim_{n\to\infty} E(OPT(L_n)/(a_1+a_2+\cdots+a_n)) = 1$. Therefore the expected asymptotic performance ratio of Best Fit is strictly greater than 1 if and only if the waste grows linearly in the number of items. Based on experimental evidence, it was conjectured that for all skewed distributions $U(0,a)$ the growth of the waste was linear [5, 1].

The discrete distributions $U\{j,k\}$ were introduced in [4] in the hope of gaining insight into the continuous case. Under distribution $U\{j,k\}$, items are drawn independently and uniformly from the set $\{1/k, 2/k, \ldots, j/k\}$. The distributions $U\{j,k\}$ approximate the continuous distribution $U(0,a)$ if one sets $j = ak$ and lets k go to infinity. Note that $U\{j,k\}$ can equivalently be thought of as the bins having capacity k and the item sizes being uniformly distributed on the integers $\{1,\ldots,j\}$; we generally use this formulation. Thanks to extensive experimental work [4, 6], several extreme cases have been analyzed under $U\{j,k\}$: when $j = k-1$, the expected waste is $\Theta(n^{1/2}\log k)$ [4]; when $j = k-2$, the expected waste is $O(1)$ [9] (a result which can also be extended to First Fit [2]); and the expected waste is also $O(1)$ when $j \leq \sqrt{2k+2.25} - 1.5$ [4]. The only case where the expected waste of Best Fit was proven to be linear was for the two distributions $U\{8,11\}$ and $U\{9,12\}$ [6]. Unfortunately, none of these results gave any information about the continuous distributions $U(0,a)$.

In this paper, we first study the discrete distributions $U\{j,k\}$ and prove that Best Fit has linear waste when k is large enough and $0.66k < j < (2/3)k$. We then proceed to prove our main result: Best Fit has linear waste for the continuous distributions $U(0,a)$ with $0.66 \leq a < 2/3$. This work therefore provides the first proof of linear waste for Best Fit under skewed continuous distributions.

*Part of this work was done while the author was visiting AT&T.

†Supported in part by an Alfred P. Sloan Research Fellowship and NSF CAREER Grant CCR-9983832. Part of this work was done while the author was visiting AT&T.

[1]Next Fit was analyzed under $U(0,a)$ by Karmakar in 1982 [8] using completely different techniques.

1.2 Proof techniques

In the discrete case, as in most previous work, we view the algorithm as a multi-dimensional Markov chain [6]. The states of the chain are non-negative integer vectors $s = (s_1, \ldots, s_{k-1})$, where s_i represents the current number of open bins of residual capacity i. It is a simple matter to write down the new vector s' that results from the arrival of any item $i \in \{1, \ldots, j\}$ when in state s. This defines an infinite Markov chain on $\mathbf{Z}_+^{k-1}$. The expected waste of Best Fit is directly related to the asymptotic behavior of this chain, which we analyze in detail.

The first novel ingredient in this paper is Lemma 1, a simple but crucial observation which we formulated after examining detailed simulations: if the maximum item size $\max$ is less than $2/3$ times the bin capacity c, then Best Fit has at most one open bin with remaining capacity in the range $[c/3, c - \max)$. Hence we focus on values of j with $j < 2k/3$.

At a high level, our approach is surprisingly simple. Our goal is to show that s_1, the number of bins of residual capacity 1, grows linearly in n. For most configurations, the next incoming item will on average tend to increase s_1 or at least not decrease it. The only exceptions are configurations with open bins of remaining capacity $1, 2, \ldots, m$, and no bins of larger remaining capacity (up to $j+1$). Thanks to Lemma 1, this implies $m \leq k/3$. Intuitively, such configurations are then extremely short-lived, and inserting a few more items then typically modifies them into configurations in which s_1 is biased towards increasing. Thus the undesirable effects of these configurations should be amortized by running the Markov chain forward in time for a few steps.

In practice, running the Markov chain for C steps, there are j^C possible sequences to analyze, which would be computationally infeasible as j gets large. To get around this problem, our detailed analysis partitions the configurations into a constant number of groups. We then use stochastic domination, i.e. take the worst case configuration within each group. This worst case configuration is determined by dynamic programming: we successively find the worst configuration within each group given that there is one more item to be inserted, and from that calculate the worst case given that there are two more items to be inserted, and so on. This use of dynamic programming is commonly used in the analysis of Markov decision processes (see, e.g., [3, 12]) and has been used for contention resolution protocols [11]; as far as we know, this is the first time it has been applied to stochastic bin-packing. Although the derived dynamic program only has to deal with a constant number of cases, it is too large to be processed manually and so we ended up writing a computer program for it. The actual table filled in by the program has tens of thousands of entries.

The continuous model follows the same general argument. The result for discrete distributions cannot be applied directly to continuous distributions by simply letting k go to infinity, because when bin sizes are scaled to 1, our discrete result shows that s_1, the number of bins with remaining capacity $1/k$, grows linearly, but as $1/k$ goes to 0 the contribution to the waste is sublinear. Hence in the continuous case, instead of studying s_1, we focus on bins with remaining capacity in the range $(0, \epsilon]$ for some small constant ϵ, and suitably adapt the proof of the discrete case.

2 Analysis of discrete skewed distributions

2.1 The Markov chain

In this section we study the discrete distribution $U\{j, k\}$ where the bin capacity is k and items are picked uniformly at random from the set $\{1, 2, \ldots, j\}$. We focus specifically on the case where $j = \alpha k$ for $33/50 < \alpha < 2/3$, and we assume that j and k are sufficiently large so that our arguments hold throughout.

Let us first recall the associated Markov chain setting. We shall denote the state of the system at time t by $s(t) = (s_1(t), \ldots, s_{k-1}(t))$, where $s_i(t)$ is the number of open bins at time t with residual capacity exactly i. Initially, the state of the system is $s(0) = (0, \ldots, 0)$, reflecting the fact that there are no open bins. Let ℓ be the size of the next item inserted. Let i be the smallest index such that $i \geq \ell$ and $s_i(t) > 0$, if such exists: in this case, the algorithm inserts item ℓ into a bin with capacity i, so we have $s_i(t+1) = s_i(t) - 1$ and, if $i > \ell$, $s_{i-\ell}(t+1) = s_{i-\ell}(t) + 1$; all other components of $s(t)$ are unchanged. If no such i exists, then the algorithm inserts item ℓ into an empty bin, so we have $s_{k-\ell}(t+1) = s_{k-\ell}(t) + 1$ and all other components of s are unchanged. This completes the description of the Markov chain.

2.2 The difficult configurations

Our attack for proving instability is straightforward: we show that s_1, the number of almost full bins, is biased upward and hence tends to increase. Let X_t denote the number of ways to increase s_1 and $Y_t \in \{0, 1\}$ denote the number of ways to decrease s_1. The value s_1 increases exactly when an item of size x is inserted and we have $s_x = 0, s_{x+1} \neq 0$, and so X_t is exactly the number of such pairs (s_x, s_{x+1}) with $x \leq j$. At every time step, if $s_1(t) = 0$ we have $Y_t = 0$, and if $s_1(t) \neq 0$ then $Y(t) = 1$: namely, s_1 can decrease only when an item of size 1 arrives. The only situations where s_1 is biased downward are if s_1 has one way to decrease and no way to increase, i.e. if for some m, $s_1, s_2, \ldots, s_m \neq 0$ and $s_{m+1} = \ldots = s_{k-j+1} = 0$. We call these configurations where s_1 is biased downwards difficult configurations, as handling them is the challenge of the problem.

The lemma below will enable us to conclude that m must be less than $k/3$ in any difficult configuration.

2.3 The open range lemma

In this subsection, we demonstrate that one cannot have more than one bin with remaining capacity within a rather large range. We call the resulting lemma the *open range lemma*. The following fact is a classical basic property of Best Fit.

Fact 1 *Any two open bins with remaining capacities g and g' must have $g + g' < k$.*

Lemma 1 [Open range lemma] *If the maximum item size j is strictly less than $2k/3$, then $s_{\frac{k}{3}} + \ldots + s_{k-j-1} \leq 1$.*

Proof: Note that initially $s_{k/3} + \ldots + s_{k-j-1} = 0$. Hence we need only show that when $s_{k/3} + \ldots + s_{k-j-1} = 1$ it cannot increase.

Consider any time t when $s_{k/3} + \ldots + s_{k-j-1} = 1$ and let $i \in \{k/3, \ldots, k-j-1\}$ be such that $s_i = 1$. Let i' be such that $k/3 \leq i' \leq k-j-1$. How can $s_{i'}$ increase? Note that $s_{i'}$ cannot increase by having an item of size $k - i'$ placed into an empty bin, since $k - i'$ is greater than j, the largest item size. Thus a bin with remaining capacity i' can only be introduced by adding some item of size x to a bin which already has a remaining capacity of g, with $g - x = i'$.

Assume then that at time t there is one bin with remaining capacity i and one with remaining capacity g. From Fact 1, we have $i + g < k$, so that

$$k - g > i \geq k/3. \tag{1}$$

By definition the remaining capacity g must be larger than i. Also, by the definition of Best Fit, item x would have been placed in the bin with remaining capacity i if it had fit there, rather than in the bin with remaining capacity g. So it must be that x does not fit in remaining capacity i:

$$x > i \geq k/3. \tag{2}$$

Now, by assumption

$$i' \geq \frac{k}{3}. \tag{3}$$

Summing inequalities (1), (2), and (3) we obtain $k - g + x + i' > k$, and hence $i' > g - x$, a contradiction.

■

We reiterate that the open range lemma simplifies the analysis, since it ensures that there is some well-defined range of values i where most of the values s_i must be 0, and hence that any difficult configuration must have $m < k/3$.

2.4 A stopping time framework

Recall that the difficult configurations are the ones such that for some $m < k/3$, $s_1, s_2, \ldots, s_m \neq 0$ and $s_{m+1} = \ldots = s_{k-j+1} = 0$. In these cases, we consider the evolution of the system over τ steps for some random *stopping time* τ, and will show that the expected number of ways for s_1 to increase is greater than 1 over these τ steps. For convenience, imagine the process starting at time 0, with time t corresponding to the moment after the tth item has been inserted. The stopping time will correspond to one of the following events:

1. Time step C has been reached, for some fixed constant C.
2. The coordinate s_1 increases or decreases.
3. The coordinate s_m becomes 0.
4. The coordinate s_{m+1} becomes positive.
5. For some $a > m$, the coordinate s_a becomes equal to 2.
6. For some $a > m + 1$, the coordinates s_a and s_{a-1} become positive.

The idea of this stopping time is as follows: we run the chain for at most C steps, or until some other (unlikely) event occurs. If s_1 is more likely to increase than decrease over the interval corresponding to the stopping time, then we can "collapse" the steps of the process until the stopping time into a single long superstep. Then s_1 is either unbiased or biased upwards over any normal step or any superstep of the chain, and this is sufficient to prove instability.

We include certain events in the stopping time that would affect our analysis. In particular, in our analysis, we consider that the number of ways for s_1 to increase is exactly $s_{m+1} + s_{m+2} + \ldots + s_{j+1}$; thus we assume that whenever some s_a for $k/2 > a > m$ becomes non-zero then it becomes 1, and that it corresponds to a "useful gap" for us (i.e. an entering item of size $a - 1$ will cause s_1 to increase). This fails to be true if both s_a and s_{a-1} become positive, or if s_a becomes 2. Rather than complicate our analysis further, we essentially remove these "edge effects" from analysis by introducing these events into the stopping time. As these events occur with very small (constant) probability (for large enough values of k compared to C) they will not affect our argument, as we explain, and so we will generally dismiss them in subsequent analysis.

We wish to show that the probability that s_1 increases over the interval of τ steps is greater than the probability that s_1 decreases. We now introduce another simplification. Let X_t be a random variable representing the number of ways for s_1 to increase at time t for $t \in [0, C)$. Then, up to lower order terms, the difference from the probability that s_1 increases to the probability that s_1 decreases is at least

$$\frac{\left(E\left[\sum_{t=0}^{C-1} X_t/C \mid \text{state at time } 0\right] - 1\right)}{j}.$$

Intuitively, the above formula shows that we can just count the number of ways s_1 can increase at each step and subtract the number of ways s_1 can decrease at each step over C steps in order to compute difference in the probability that s_1 increases rather than decreases over an interval that ends at a stopping time. Of course this calculation is not exact, since for example the stopping time might be reached before C steps. However, because the probability that an event of type $2, 3, 4, 5$, or 6 occurs over C steps is $O(1/j)$ (where the constant factor depends on C), it can be checked that the above expectation differs from the proper difference by an $O(1/j^2)$ lower order term. Hence, for our purposes, the above expression gives the proper bound.

From our definition of the stopping time, which excludes problematic events, it is easy to see that X_t, the number of ways for s_1 to increase, at each time step in this range is equal to $s_{m+1} + \ldots + s_{j+1}$, and thus we are reduced to analyzing this simple expression. We therefore show that there exists an integer constants $C > 0$ and a real constant $\epsilon > 0$ (independent of k) such that for sufficiently large k,

$$E\left[\frac{\sum_{t=0}^{C-1} s_{m+1}(t)+\ldots+s_{j+1}(t)}{C} \mid \text{state at time } 0\right] > 1+\epsilon. \quad (4)$$

Note that we require that the right hand side be greater than $1 + \epsilon$ so as to absorb the error term introduced by our approximations due to the stopping time. Hence, we show that up to amortizing slightly over time, the bin-packing procedure is more likely to increase s_1 than decrease it when it begins in a difficult configuration. We then show that this implies that s_1 must have a slight bias toward increasing over all steps; from this we conclude that s_1 and hence the waste grow linearly in the number of items n.

2.5 The analysis starting from a difficult configuration

We assume that $j = \alpha k$ for $\alpha \in (33/50, 2/3)$. We begin in a state where $s_1, s_2, \ldots, s_m \neq 0$ and $s_{m+1} = \ldots = s_{j+1} = 0$. In this state s_1 has no possibility to increase. Note that proving (4) will imply that for sufficiently large k the value s_1 diverges to infinity. Our proof is based on stochastic domination.

2.5.1 Using stochastic domination: a nonrigorous example

We first present an example of a simplified and nonrigorous analysis, which demonstrates how we attack the underlying Markov chain. For our simplified analysis, we will ignore the effect of non-empty bins with remaining capacity at least $k/2$. That is, we assume the number of such bins is 0 throughout. As we describe subsequently, such mostly empty bins complicate the analysis. Take the example where $k-j \geq m \geq k/4$. At every step, there are $j = k\alpha$ possibilities for the item arriving. Out of these possibilities, X_t has at least $j - k/2 = k(2\alpha - 1)/2$ ways of increasing, corresponding to insertions of items $k/2, k/2+1, \ldots, j$ (this is because of our assumption that there are never bins more than half empty). On the other hand, since s_m stays positive X_t has at most $k/2 - m$ ways of decreasing if it is non-zero and no way of decreasing if it is equal to 0. It is worth also noting that X_t has at most mX_t ways to decrease in general, since for bin than contributes to X_t, there are only m possible items that could enter and reduce its residual capacity to something less than m. (For this range of m, however, the bound of $k/2-m$ is better.) We now use stochastic domination. Following standard definitions (see, e.g. [13]) we say that X stochastically dominates Y and write $X \geq Y$ if $\Pr(X > u) \geq \Pr(Y > u)$ for all real values u. Intuitively, X is more likely to take on larger values than Y. It is simple to show (say via induction) that $X_t \geq Z_t$, where

$$\begin{aligned} Z_0 &= 0 \\ Z_{t+1} &= \begin{cases} Z_t + 1 & \text{w.p. } \frac{k(2\alpha-1)}{2k\alpha}, \\ Z_t - 1 & \text{w.p. } \frac{(k/2-m)}{k\alpha} \text{ if } Z_t \neq 0, 0 \text{ otherwise,} \\ Z_t & \text{with all remaining probability.} \end{cases} \end{aligned}$$

It is easily checked that the probability of Z_t increasing is smallest and the probability of Z_t decreasing is largest for the smallest value of α, $\alpha = 33/50$, and the smallest value of m, $m = k/4$. That is, $X_t \geq Z_t$, where

$$\begin{aligned} Z_0 &= 0 \\ Z_{t+1} &= \begin{cases} Z_t + 1 & \text{w.p. } 8/33, \\ Z_t - 1 & \text{w.p. } 25/66 \text{ if } Z_t \neq 0, 0 \text{ otherwise,} \\ Z_t & \text{with all remaining probability.} \end{cases} \end{aligned}$$

We check by explicit calculation that $E((Z_0 + Z_1 + \ldots + Z_{C-1})/C) > 1$ for a small constant C. (For $C = 30$, $E((Z_0 + Z_1 + \ldots + Z_{C-1})/C) \approx 1.007$; by calculating the stationary distribution, we see that as C gets large, $E((Z_0 + Z_1 + \ldots + Z_{C-1})/C)$ approaches $\frac{16}{9}$.) Hence, since X stochastically dominates Z, we have for a small constant C that

$$E((X_0+\ldots+X_{C-1})/C) \geq E((Z_0+\ldots+Z_{C-1})/C) > 1.$$

We can continue by breaking down the process into a finite number of similar cases, covering the entire range of m. (For example, we could take the cases $0 \leq m \leq k/8$ and $k/8 \leq m \leq k/4$.) In each case, we can bound the probability of X_t increasing and decreasing, so as to find a dominating simple one-dimensional random walk. It is therefore easy to check each specific case, simply by determining the distribution of Z_t over a reasonably small number of steps.

2.5.2 The important parameters describing a configuration

The above analysis captures the fundamental flavor of our argument: simplify the high-level behavior of the Markov process according to the number of ways that s_1 can increase, and consider this behavior over a small number of steps. Unfortunately, the analysis above fails to take into consideration the behavior of the process when "mostly empty" bins arise. That is, it ignores possibilities when $s_i > 0$ for $i \geq k/2$.

To see that these cases have an important impact, consider the case where $j = 0.34k$, and $m = 0.32k$, and $s_{0.64k} = 1$. In this case it is impossible for any of $s_{m+1}, s_{m+2}, \ldots, s_{k/2}$ to increase on the next step. Thus it is important to keep track of the presence of bins of remaining capacity greater than $k/2$. The following fact, easily derived from Fact 1, proves fundamental:

Fact 2 *There can be only one open bin with remaining capacity at least $k/2$.*

Let us call a non-empty bin with remaining capacity at least $k/2$ a *light* bin. Fact 2 says there is at most one light bin. Hence we can consider two kinds of states: those with a light bin, and those without. In fact, we refine our analysis further, to three kinds of states, by dividing states with a light bin into two subtypes. Let us call a light bin *helpful* if its remaining capacity is at most $j+1$. A helpful light bin can immediately lead to an increase in s_1, if an appropriately sized item arrives. Similarly we call a light bin *unhelpful* if its remaining capacity is greater than $j+1$.

We may now represent the state of our bin-packing process by a triple (X_t, A_t, B_t), where X_t is again the number of ways for s_1 to increase at time t, A_t is a 0/1 random variable representing whether or not there is a light helpful bin, and B_t is a 0/1 random variable representing whether or not there is a light unhelpful bin. Note that when $A_t = 1$, we must have that $X(t) \geq 1$, since by definition a helpful light bin provides one way for s_1 to increase.

2.5.3 The dynamic program

We then use a computer program to assist in computing successive lower bounds for the value of the expression $E_c = E(\sum_{t=0}^{c-1} X_t/c \mid \text{ state at time } 0)$ until we find a sufficiently large C so that $E_C > 1+\epsilon$. The interesting point of our calculation is that in ambiguous situations we use a worst-case analysis, similar in spirit to allowing an oblivious adversary some limited power in deciding the flow of the process. For example, suppose $A_t = 1$, and an item of size in the range $[0.5k, j]$ arrives. Such an item could be placed in the light bin; alternatively, such an item may prove too large for the light bin, and instead cause X_t to increase. The effect of the item depends on the exact residual capacity of the light bin and the value of the entering item; however, we have failed to include the capacity of the light bin in our state.

In such ambiguous situations we can compute E_{t+1} from E_t by taking the worst-case possibility for E_t. This simplifies our case analysis, in that we need not try to distinguish further subcases when $A_t = 1$; however, it complicates it, in that we must consider several various possibilities at each step. This makes any direct analysis very complicated, requiring us to use computer analysis.

Finally, note that we must apply this analysis over the space of all (m, α) pairs, but we may wish to break up this space into several cases, as we did in Section 2.5.1. We suggest the approach we have taken. Suppose one focuses on a specific value of α (such as $\alpha = 33/50$), and proves $E_C > 1+\epsilon$ by splitting up the possible values of m over a small number of ranges. We claim that then there are small constants $\delta, \epsilon' > 0$ such that $E_C > 1+\epsilon'$ for $\alpha' \in [\alpha - \delta, \alpha + \delta]$. This is because with very high probability, the Markov chain will behave the same over C steps regardless of whether $j = \alpha k$ or $j = \alpha' k$; the small probability that the two behave differently is absorbed in the ϵ, ϵ'. Hence it suffices to try a sufficiently dense subset of α values in the range $[33/50, 2/3)$, and by the "continuity" implied by the above argument, we may conclude $E_C > 1+\epsilon$ for a suitable ϵ everywhere in the interval.

The list of cases is substantial enough that we do not include it in this extended abstract. (A sample case is included in the appendix.) Luckily we have found that calculating up to E_{100} is sufficient. (Indeed, the worst case appears to be the lower end of the interval, $33/50$; we cannot expand the range beyond $2/3$ simply because the open range lemma ceases to apply.) Hence our calculations, when aided by a machine, are actually relatively straightforward.

We note that we have chosen the barrier $33/50$ for convenience, and we have not tried to determine the exact range for which the argument holds. It appears that additional work detailing the cases would be required, however, to extend the lower bound of the range below $0.65 = 13/20$.

2.6 Wrapping up the proof

Note that showing $E\left[\sum_{t=0}^{C-1} \frac{X_t}{C} \mid \text{state at time } 0\right] > 1+\epsilon$ immediately proves that the waste for $U\{j,k\}$ diverges. We now prove that, given that we have shown that $E\left[\sum_{t=0}^{C-1} \frac{X_t}{C} \mid \text{state at time } 0\right] > 1+\epsilon$, then for sufficiently large k, the waste for $U\{j,k\}$ grows linearly.

Theorem 1 *The number of s_1 bins for Best Fit bin packing under the discrete distribution $U\{j,k\}$, $33k/50 < j < 2k/3$, grows linearly in n for sufficiently large k.*

Proof: Let Y_t be the indicator function of the event $s_1(t) \neq 0$, and define $Z(t) = X_t - Y_t$. The distribution of $s_1(t)$ given the state at time $t-1$ is:

$$s_1(t) = \begin{cases} s_1(t-1)+1 & \text{w.p. } X_{t-1}/j \\ s_1(t-1)-1 & \text{w.p. } Y_{t-1}/j. \end{cases}$$

Thus $E(s_1(t) \mid \text{state at } t-1) = s_1(t-1) + Z(t-1)$, and summing gives $E(s_1(T)) = E(\sum_0^{T-1} Z(t))$.

We always have $Z(t) \geq -1$. The "bad" configurations are exactly the ones for which $Z(t) = -1$. Consider running the chain for n steps, and divide time into *supersteps*. A superstep is simply a normal step of the chain, except in the case where we reach a state where $Z(t) = -1$. In this case all the steps from this point until the stopping time are combined into a superstep; in fact, we call this a *long superstep*. Every short superstep has $Z(t) \geq 0$ and every long superstep has $E(\sum_{t\in \text{ superstep}} Z(t)) > 0$.

To show linear waste, we note the following: there will be linear waste over n steps as long as there are (on average) $\geq cn$ short supersteps which have $Z(t) > 0$, or as long as there are (on average) $\geq cn$ long supersteps. Note that this is equivalent to saying that there are at least $c'n$ steps with $Z(t) > 0$, as each such step is either an appropriate short superstep or falls in a long superstep (of constant size).

Hence we must show that $Z(t) > 0$ for a constant fraction of the steps, on average. This may appear obvious; however, a priori it is possible that all of the s_i except those determined in the open range lemma are greater than 0 for almost all steps, in which case we might have $Z(t) = 0$ for almost every step. We can show that this is not the case, simply by showing that the area adjacent to the open range guaranteed by the open range lemma must also be open a constant fraction of the time. Specifically, let $\gamma = 2/3 - \alpha$. We show that all of $s_{k-j}, s_{k-j+1}, s_{k-j+\gamma j/2}$ are simultaneously 0 for an expected constant fraction of the time steps. From this it is easy to conclude that an expected constant fraction of the steps are short supersteps with $Z(t) > 0$, since from a state where $s_{k-j}, s_{k-j+1}, s_{k-j+\gamma j/2}$ are all simultaneously 0 we achieve a state where exactly two non-neighboring s_i in the range are non-zero in two time steps with constant probability. In fact the probability is approximately $\gamma^2/4$; we simply need the immediate insertion of two items into empty bins that yield bins with remaining capacity in this range.

To show that all of $s_{k-j}, s_{k-j+1}, s_{k-j+\gamma j/2}$ are simultaneously 0 for an expected constant fraction of the time steps, we show that $S = s_{k-j} + s_{k-j+1} \ldots + s_{k-j+\gamma j/2}$ is stochastically dominated by a random walk that is biased downward. This follows since when the range $s_{k/3+\gamma/4}, \ldots, s_{k-j-1}$ are all 0, there are at most $\gamma j/2$ entering item sizes that increase S when $S \neq 0$, corresponding to when items are placed into empty bins. No other items can increase S, by the same logic as the open range lemma. Of course, at least $3\gamma j/4$ possible item sizes decrease S; namely, any item size in the range $[k/3+\gamma/4, k-j-1]$. (Note there may be at most one non-zero entry in the range $s_{k/3+\gamma/4}, \ldots, s_{k-j-1}$, by Lemma 1, in which case we do not know that $3\gamma j/4$ possible entering item sizes decrease S. However, it is easy to show that this non-zero entry disappears after an expected constant number of steps, and no such non-zero entry can return as long as $s_{k-j} + s_{k-j+1} \ldots + s_{k-j+\gamma j/2} > 0$; hence it has a trivial effect on the argument.)

Hence we either have that some constant fraction of the supersteps are going to be short supersteps such that $Z(t) > 0$, or a constant fraction of the supersteps are long supersteps, and so $E(s_1(T)) = \Omega(T)$. Note that the constant factor implied by the Ω notation is in fact independent of k.

■

3 Analysis of continuous skewed distributions

Our above analysis can easily be extended to show that linear waste occurs when bins have size 1 and the items are uniform over the real interval $(0, a]$, for a in the range $33/50 \leq a < 2/3$. We believe this is the first non-trivial probabilistic analysis of Best-Fit for the continuous case when the interval is other than $(0, 1]$.

The result of Theorem 1 cannot be extended immediately as k grows to infinity to yield the continuous case, since if we scale back the bin sizes in the $U\{j,k\}$ model so that all bins have size 1, Theorem 1 only says that the number of bins with remaining capacity $1/k$ grows linearly. A more careful argument avoids this, giving us waste that grows independent of k. Rather than extend the discrete case, however, we simply outline the corresponding argument for the continuous case.

We first note that the open range lemma that simplifies our analysis has a continuous analog.

Lemma 2 *Let $1/2 < a < 2/3$. Then when the bin capacity is 1 and the item sizes are drawn from $(0, a]$, there can only be one bin with remaining capacity in the interval $(1/3, 1 - a)$.*

Moving from the discrete case to the continuous case requires some care. The primary difference is that for the continuous case, we consider the creation and deletion of bins with remaining capacity $[\delta/\gamma, \delta)$ for some suitably small constant δ and suitably large constant γ. We call the number of such bins $s_{\delta,\gamma}$. Note that the choices of δ and γ are dependent on a. Also note that we consider bins with remaining capacity in the range $[\delta/\gamma, \delta)$ rather than $[0, \delta)$, because if we allow the lower bound to go to 0, we cannot give strong enough statements about the waste represented by these bins.

We sketch the proof, which follows the same outline as before, in that we use our analysis of a chain (X_t, A_t, B_t). Whereas before X_t represented the number of ways to increase s_1, it should now represent the number of ways that $s_{\delta,\gamma}$ can be increased. That is, suppose that there is a bin with remaining capacity $x < a+\delta$, and no bin with remaining capacity in the range $[x-\delta, x)$. Then this bin represents a way that $s_{\delta,\gamma}$ can increase, as any incoming item in the range $[x - \delta, x - \delta/\gamma)$ increases $s_{\delta,\gamma}$. Assuming without loss of generality that all open bins have distinct remaining capacities, X_t is simply the number of distinct values x with this property at time t. The values A_t and B_t then have similar meanings as in the discrete case.

We must change our definition of the stopping time as well. For example, in the discrete case we did not wish to allow two bins with remaining capacities i and $i + 1$ to both count for X_t, and we introduced a stopping time to prevent such an event. We do the equivalent here; should we introduce over multiple steps of the process two bins with remaining capacities x and $x + \rho$ for $\rho < \delta$ that could count for X_t, the event signals the stopping time.

Given these modifications, we can show with the same argument as in the discrete case that even when $X_t = 0$, over the next C steps we have $E_C = E\left[\sum_{t=0}^{C-1} \frac{X(t)}{C} \mid \text{state at time } 0\right] > 1 + \epsilon'$ for some constants ϵ' and C. Indeed, we may use the same dynamic programming formulation as in the discrete case, where the evolution of (X_t, A_t, B_t) is analyzed independent of k; by making δ sufficiently small and γ sufficiently large, the effect on the analysis can be made arbitrarily small. As before, the constant ϵ' also handles any problems introduced by using stopping times.

We may conclude that the rate at which bins with remaining capacity in the range $[\delta/\gamma, \delta)$ are created is slightly larger (by a small constant factor) than the rate at which they disappear. Hence the rate at which $s_{\delta,\gamma}$ increases is slightly larger than the rate at which it decreases. Since the remaining capacity of a bin contributing to $s_{\delta,\gamma}$ is at least the constant δ/γ, this implies linear waste.

Theorem 2 *Best Fit bin packing with item sizes chosen uniformly from the range $[0, a]$ has linear waste, for $33/50 \leq a < 2/3$.*

4 Conclusions and open problems

We have shown linear waste for discrete distributions of the form $U\{j, k\}$, where $j = \alpha k$ for α in the range $[0.66, 2/3)$, and from this derived an argument for the continuous case $(0, a]$ for a in the above range. Our analysis depends on a careful breakdown of the underlying Markov chain. Although we feel that the complexity of our argument is in fact necessitated by the complex behaviors of the system, it may be possible to simplify our argument substantially.

One open problem is to determine the extent of the range of values for which our argument functions. It is clear that the range $[0.66, 2/3)$ can be expanded, in both directions. Our current argument appears to work down to values of α (or a, in the continuous case) of almost $0.65 = 13/20$ and could perhaps be improved further. Extending the range upwards would be more delicate since Lemma 1 would no longer apply.

It seems likely that our techniques might also apply to First Fit. It is conjectured that First Fit has linear waste on input distributions $[0, a]$ for $a < 1$ as well.

References

[1] Problems in Analysis of Algorithms Pages. At `http://pauillac.inria.fr/algo/AofA/Problems/index.html`. Problem 5, posted by Ed Coffman.

[2] S. Albers and M. Mitzenmacher. Average case analyses of First Fit and Random Fit bin packing. In *Proceedings of the 9th Annual ACM Symposium on Discrete Algorithms*, pp. 290–299, 1998.

[3] D.P. Bertsekas and J.N. Tsitsiklis. *Neuro-Dynamic Programming*, Athena Scientific, Massachusetts, 1996.

[4] E.G. Coffman Jr., C. Courcoubetis, M.R. Garey, D.S. Johnson, L.A. McGeoch, P.W. Shor, R.R. Weber and M. Yannakakis. Fundamental discrepancies between average-case analyses under discrete and continuous distributions: A bin-packing case study. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, 1991, pp. 230–240.

[5] E.G. Coffman Jr., M.R. Garey and D.S. Johnson. Approximation algorithms for bin packing: A survey. In *Approximation Algorithms for NP-Hard Problems*, D.S. Hochbaum, ed., PWS Publishing Company, Boston, 1996.

[6] E.G. Coffman Jr., D.S. Johnson, P.W. Shor and R.R. Weber. Markov chains, computer proofs and average-case analysis of Best Fit bin packing. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, 1993, pp. 412–421.

[7] D.S. Johnson, A. Demers, J.D. Ullman, M.R. Garey and R.L. Graham. Worst case performance bounds for simple one-dimensional packing algorithms. *SIAM Journal on Computing* **3** (1974), pp. 299–325.

[8] N. Karmakar. Probabilistic Analysis of Some Bin-Packing Problems. IEEE FOCS 1982, pp. 107-111.

[9] C. Kenyon, Y. Rabani, and A. Sinclair. Biased Random Walks, Lyapunov Functions, and Stochastic Analysis of Best Fit Bin Packing. *Journal of Algorithms*, Vol. 27, No. 2, 1998, pp. 218-235.

[10] F.T. Leighton and P.W. Shor. Tight bounds for minimax grid matching with application to the average case analysis of algorithms. *Combinatorica* **9** (1989), pp. 161–187.

[11] J. Mosely and P. A. Humblet. A class of efficient contention resolution algorithms for multiple access channels. *IEEE Transactions on Communications* **33:2** (1985), pp. 145–151.

[12] M. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley and Sons, New York, 1994.

[13] M. Shaked and J. G. Shanthikumar. *Stochastic Orders and Their Applications.* Academic Press, Inc., San Deigo, 1994.

[14] P.W. Shor. The average case analysis of some on-line algorithms for bin packing. *Combinatorica* **6** (1986), pp. 179–200.

Appendix: Case analysis

The following presents part of our case analysis, for the case where $1/4 \leq m \leq 1/3$. Let the entering item have weight wk, where $0 \leq w \leq \alpha$. Recall that X_t is the number of ways to increase s_1; A_t is 1 if and only if there is a helpful light bin, and 0 otherwise; B_t is 1 if and only if there is an unhelpful light bin, and 0 otherwise.

- Case 1: $A_t = 0, B_t = 0$.
  - If $w \leq 1/4$, no change.
  - If $1/4 \leq w \leq 1/3$, the adversary decides between setting X_{t+1} to $X_t - 1$ (if possible) or setting B_{t+1} to 1.
  - If $1/3 \leq w \leq 1 - \alpha$:
    * If $X_t = 0$, set B_{t+1} to 1.
    * If $X_t > 0$, the adversary decides between setting X_{t+1} to $X_t - 1$ or setting B_{t+1} to 1.
  - If $1 - \alpha \leq w \leq 1/2$:
    * If $X_t = 0$, set A_{t+1} to 1.
    * If $X_t > 0$, the adversary decides between setting X_{t+1} to $X_t - 1$ or setting A_{t+1} to 1.
  - If $1/2 \leq w \leq \alpha$, set X_{t+1} to $X_t + 1$.
- Case 2: $A_t = 1$.
  - If $w \leq 1/4$, no change.
  - If $1/4 \leq w \leq 1/3$, the adversary decides between setting X_{t+1} to $X_t - 1$ (if possible) or setting A_{t+1} to 0 (either increasing X_t or not.)
  - If $1/3 \leq w \leq 1/2$:
    * If $X_t = 0$, set A_{t+1} to 0, and adversary decides whether to increase X_t or not.
    * If $X_t > 0$, the adversary decides between setting A_{t+1} to 0 (either increasing X_t, or not) and setting X_{t+1} to $X_t - 1$.
  - If $1/2 \leq w \leq \alpha$, the adversary decides between setting X_{t+1} to $X_t + 1$ or setting A_t to 0.
- Case 3: $B_t = 1$.
  - If $w \leq 1/4$, no change.
  - If $1/4 \leq w \leq 1/3$, the adversary decides between setting X_{t+1} to $X_t - 1$ (if possible) or setting B_{t+1} to 0 (either increasing X_t, or not)
  - If $1/3 \leq w \leq 1/2$
    * If $X_t = 0$, set B_{t+1} to 0, and adversary decides whether to increase X_t or not.
    * If $X_t > 0$, the adversary decides between setting B_{t+1} to 0 (either increasing X_t, or not) and setting X_{t+1} to $X_t - 1$.
  - If $1/2 \leq w \leq \alpha$, set B_{t+1} to 0.

We require similar analysis for the following ranges of m: $0 \leq m \leq 12/100$, $12/100 \leq m \leq 22/100$, and $22/100 \leq m \leq 25/100$.