From *Algorithm Engineering and Experimentation,*
Goodrich & McGeoch, Eds., Springer LNCS 1619, 1999, pp 246-265

A Self Organizing Bin Packing Heuristic

Janos Csirik[1], David S. Johnson[2], Claire Kenyon[3], Peter W. Shor[4], and Richard R. Weber[5]

[1] Department of Computer Sciences
University of Szeged, Szeged, Hungary
csirik@inf.u-szeged.hu
Supported in part by a Fullbright Fellowship, by DIMACS, and by AT&T Labs
[2] AT&T Labs – Research, Room C239,
Florham Park, NJ 07932-0971, USA
dsj@research.att.com
[3] Laboratoire de Recherche en Informatique
Bâtiment 490, Université Paris-Sud, 91405 Orsay Cedex, France
Claire.Kenyon@lri.fr
Supported in part by AT&T Labs
[4] AT&T Labs – Research, Room C237,
Florham Park, NJ 07932-0971, USA
shor@research.att.com
[5] Statistical Laboratory, Cambridge University
16 Mill Lane, Cambridge CB2 1SB, England
rrw1@cam.ac.uk
Supported in part by AT&T Labs

Abstract. This paper reports on experiments with a new on-line heuristic for one-dimensional bin packing whose average-case behavior is surprisingly robust. We restrict attention to the class of "discrete" distributions, i.e., ones in which the set of possible item sizes is finite (as is commonly the case in practical applications), and in which all sizes and probabilities are rational. It is known from [7] that for any such distribution the optimal expected waste grows either as $\Theta(n)$, $\Theta(\sqrt{n})$, or $O(1)$, Our new *Sum of Squares* algorithm (SS) appears to have roughly the same expected behavior in all three cases. This claim is experimentally evaluated using a newly-discovered, linear-programming-based algorithm that determines the optimal expected waste rate for any given discrete distribution in pseudopolynomial time (the best one can hope for given that the basic problem is NP-hard). Although SS appears to be essentially optimal when the expected optimal waste rate is sublinear, it is less impressive when the expected optimal waste rate is linear. The expected ratio of the number of bins used by SS to the optimal number appears to go to 1 asymptotically in the first case, whereas there are distributions for which it can be as high as 1.5 in the second. However, by modifying the algorithm slightly, using a single parameter that is tunable to the distribution in question (either by advanced knowledge or by on-line learning), we appear to be able to make the ratio go to 1 in all cases.

1 Introduction

In the classical one-dimensional bin packing problem, one is given a list $L = \{a_1, ..., a_n\}$ of items, with a size $s(a_i) \in [0,1]$ for each item in the list. One desires to pack the items into a minimum number of unit-capacity bins, i.e, a partition of the items into a minimum number of subsets such that the sum of the sizes of the items in each subset is one or less. This problem is NP-hard, so much research has concentrated on designing and analyzing polynomial-time approximation algorithms for it, i.e., algorithms that construct packings that use relatively few bins, although not necessarily the smallest possible number. Of special interest have been *on-line* algorithms, i.e., ones that must permanently assign each item in turn to a bin without knowing anything about the sizes or numbers of additional items, a requirement in many applications.

In this paper we concentrate on the average-case behavior of such algorithms. The key metrics with which we are concerned can be defined using the following notation. For a given algorithm A and list L, let $A(L)$ be the number of bins used when A packs L, let $s(L) = \sum_{a \in L} s(a)$, and let $OPT(L) \geq s(L)$ be the optimal number of bins. For a given probability distribution F on item sizes, let $L_n(F)$ be a random n-item list with item sizes chosen independently according to distribution F. Then the *asymptotic expected performance ratio* for A on F is

$$ER_A^\infty(F) \equiv \limsup_{n \to \infty} \left(E \left[\frac{A(L_n(F))}{OPT(L_n(F))} \right] \right)$$

and the *expected waste rate* for A on D is

$$EW_A^n(F) \equiv E\left[A(L_n(F)) - s(L_n(F))\right]$$

Note that because of the low variance of $s(L_n(F))$ for any fixed F, $EW_A^n(F) = o(n)$ implies $ER_A^\infty(F) = 1$ (although not necessarily vice versa). When the context is clear, we will often omit the "(F)" in the above notation.

To date, the most broadly effective practical on-line bin packing algorithm has been *Best Fit* (BF), in which each item is placed in the fullest bin that currently has room for it. Best Fit has been studied under a significant range of distributions. The classical results concern the continuous uniform distributions $U[0,b]$, where item sizes are uniformly distributed over the real interval $[0,b]$. For $b = 1$ we have $EW_A^n = \Theta(n^{1/2}(\log n)^{3/4})$ [13, 10], and for $b < 1$ experiments reported in [1, 3] suggest that $ER_A^n > 1$, with a maximum value of approximately 1.014, attained for $b \sim 0.79$.

More recently, the behavior of BF has been studied in [3, 5, 8] for the discrete uniform distributions $U\{j,k\}$, $1 \leq j < k$, in which the allowed item sizes are $1/k, 2/k, ..., j/k$, all equally likely. For $k \geq 3$ and $j = k-1$, BF's behavior for $U\{j,k\}$ approximately mimics that for $U[0,1]$, and we have $EW_A^n = \Theta(n^{1/2}(\log k)^{3/4})$ [4]. Moreover, for $j = k-2$ or $j < \sqrt{2k+2.25} - 1.5$, much better performance occurs and we have $EW_A^n = O(1)$ [3, 8]. However, there appears to exist a constant c such that $ER_A^n > 1$ for $c\sqrt{k} < j \leq k-3$ and k sufficiently large, with the behavior for $U\{j,k\}$ roughly mimicking that for $U[0, j/k]$.

If running time is no object, algorithms with significantly better expected behavior are possible. Rhee and Talagrand [11] have shown that for any fixed distribution F, there is an algorithm X_F such that $ER^{\infty}_{X_F}(F) = 1$ and such that if $EW^n_{OPT}(F) = o(n)$, then $EW^n_{X_F}(F) = O(n^{1/2}(\log n)^{3/4})$. Moreover, if one is willing to repeatedly solve instances of an NP-hard partitioning problem as part of the algorithm, this level of asymptotic performance can be attained *without* knowing the distribution F in advance, simply by obtaining better and better estimates of it as one goes along, i.e., by learning F on-line [12].

If one restricts attention to discrete distributions, i.e., ones in which the item sizes are all members of a fixed finite set of rational numbers and the corresponding probabilities are all rational numbers as well, even stronger results are possible. For discrete distributions F, the only possible values of $EW^n_{OPT}(F)$ are $\Theta(n)$, $\sqrt{n}$, and $O(1)$, as shown in [7], and for any fixed discrete distribution F there is a linear time on-line algorithm Y_F that has $EW^n_{Y_F}(F) = O(EW^n_{OPT}(F))$. As was the case with the algorithms X_F, the performance of the algorithms Y_F can also be obtained by a single distribution-free algorithm that learns the distribution as it goes along and repeatedly solves NP-hard problems.

Neither of these generic approaches seems practical, and even the distribution-specific algorithms X_F and Y_F are far too complicated to use. They require the (possibly repeated) construction of detailed and distribution-dependent models of multi-bin packings, into the slots of which the incoming items must be separately matched. In this paper we shall present a new and quite simple algorithm *Sum of Squares* (SS) that we conjecture approximately attains the same level of performance as the Y_F for any discrete distribution F, without knowing or attempting to learn F. (We say "approximately" because in some cases where $EW^n_{OPT} = O(1)$, the new algorithm can be shown to yield $EW^n_{SS} = \Omega(\log n)$.) Moreover, although SS like the Y_F's can have $ER^{\infty}_{SS}(F) > 1$ when $EW^n_{OPT} = \Theta(n)$, for each such distribution F, there is a simple-to-construct and practical variant SS_F that we conjecture does yield $ER^{\infty}_{SS_F}(F) = 1$.

For simplicity in what follows, we shall assume that all discrete distributions have been scaled up by an appropriate multiplier B to obtain an equivalent distribution where all item sizes are integers (and for which the bin capacity is B). For example, the scaled $U\{j,k\}$ distributions have item sizes $1, 2, ..., j$ and bin capacity k. This scaling leaves the values of ER^{∞}_A unchanged and only affects the constant of proportionality for EW^n_A. In Section 2, we describe SS and its original motivation, and present experimental results comparing it with BF for the distributions $U\{j,k\}$, $1 \le j < k = 100$. It was these results that first suggested to us SS's surprising effectiveness.

For the $U\{j,k\}$ distributions, the needed comparison values of ER^{∞}_{OPT} and EW^n_{OPT} are already known from theoretical results in [2,3]. For more general classes of discrete distributions, determining these values can be NP-hard. However, as we show in Section 3, the determination can be made by solving a small number of linear programs (LP's) with $O(B^2)$ variables and $O(B)$ constraints, a process that is feasible for B as large as 1000. We use this LP-based approach in Section 4, where we study a generalization of the $U\{j,k\}$ to what we call the *interval distributions* $U\{h..j,k\}$, $1 \le h \le j < k$, in which the bin capacity is k and the item sizes, all equally likely, are the integers s, $h \le s \le j$. Using linear programming, we first determine the values of ER^{∞}_{OPT} and EW^{n}_{OPT} for all such distributions with $k = 19$ or $k = 100$. Then, based on simulations with 10^5, 10^6 and 10^7 items, we estimate the corresponding values for SS. For $k = 19$ we do this for all relevant values of h and j; for $k = 100$ we do this for a challenging subset of the relevant values. In all cases tested our data is consistent with the hypothesis that $EW^{n}_{SS} = O(\max\{\log n, EW^{n}_{OPT}\})$, as claimed. The need for the $\log n$ option is illustrated by tests of the interval distribution $U\{2..3,9\}$, and we describe the conditions under which EW^{n}_{SS} can be proved to grow at least at this rate even though $EW^{n}_{OPT} = O(1)$.

The apparent success of SS far outstrips our original motivation for proposing it. In Section 5 we suggest an intuitive explanation for its behavior that views the operation of SS as a self-organizing process. This explanation is illustrated using detailed measurements of the algorithm's internal parameters under various distributions.

As observed above, SS can have $ER^{\infty}_{SS}(F) > 1$ when $EW^{n}_{SS}(F) = \Theta(n)$. In Section 6, after first presenting a sequence of distributions for which the limiting value of ER^{∞}_{SS} is 1.5, we report on various modification of SS aimed at reducing the value of ER^{∞}_{SS} when $EW^{n}_{OPT} = \Theta(n)$. Most importantly, we show how we can use the results of the LP computation that we performed to determine the value of $EW^{n}_{OPT}(F)$ to devise simple variants SS_F that appear to have $ER^{\infty}_{SS_F} = 1$. This approach can in turn be incorporated into a single polynomial-time "learning" algorithm that we conjecture has an asymptotic expected performance ratio of 1 for all discrete distributions F. Experimental results for the distributions considered in Section 4 are presented that appear to support this conjecture.

We conclude in Section 7 with a preview of the journal version of this paper, which will contain additional experimental and theoretical results, most importantly a proof by Jim Orlin of one our main conjectures.

2 The Sum of Squares Algorithm and $U\{j,k\}$

The sum of squares algorithm works as follows. Assume that our instance has been scaled so that it consists of integer-size items with an integral bin capacity B. Define $N(g)$ to be the number of bins in the current packing with gap g, $1 \le g < B$, where a bin has *gap* g if the items contained in it have total size $B - g$. Initially $N(g) = 0$, $1 \le g < B$. To pack the next item a_i, we place it in a bin (either a currently empty one or a partially full bin with gap at least $s(a_i)$) that will yield the minimum updated value of $\sum_{1 \le g < B} N(g)^2$. If there is a tie, we break it in favor of a candidate bin with the largest current total contents.

A naive motivation for this algorithm (and indeed, the one that led us to propose it in the first place) starts with a fact about Best Fit. This algorithm performs surprisingly well on symmetric discrete distributions, i.e., distributions in which for all sizes s, items of size s and $B - s$ occur with equal probability (e.g., see [3,4]). The reason for this is that typically when the next item to be packed will fit in *some* partially-filled bin, there already exists such a bin whose gap precisely equals the size of the new item. Hence most bins end up being perfectly packed, i.e., having gap 0, and there is very little total waste. How might one extend this behavior to non-symmetric distributions? One idea would be to use items that don't fit perfectly in some current gap to help build and maintain an inventory of gaps so that future items *will* be likely to find a perfect fit. In the absence of other information about the distribution, an initial goal for such an inventory would be to aim for equal numbers of bins for each possible gap. The sum of squares criterion would seem to do this, since given a set of variables whose sum is fixed, their sum of squares is minimized when the values are as close to equal as possible.

This argument does not apply precisely to bin packing, since it is the total number of items in the packing that is fixed, not the total number of bins, but the simplicity of the sum of squares criterion argues in its favor. An item of size s must either (1) start a new bin, in which case the sum of squares increases by $2N(B-s)+1$, (2) perfectly fill an old bin, in which case $N(s) > 0$ and the sum of squares decreases by $2N(s)-1$, or (3) go into a bin with gap g, $s < g < B$, in which case the best choice is a g which maximizes $N(g) - N(g-s)$, and the sum of squares decreases by $2(N(g) - N(g-s)) - 2$. Thus no squares need actually be computed. However, unless we find a way to improve on exhaustive search for (3), we will have a worst-case time of $\Theta(\min(n, B))$ per item as opposed to $\Theta(\log(\min(n, B)))$ for Best Fit. The question is whether this extra running time might provide us better average performance, as hoped.

Let us first consider $U\{j,k\}$ distributions. The behavior of EW^n_{OPT} for these distributions has been characterized in [2,3]: $EW^n_{OPT} = O(1)$ for $1 \le j \le k-2$ and $EW^n_{OPT} = \Theta(\sqrt{n})$ for $j = k-1$. For each of the $U\{j, 100\}$ distributions, $1 \le j \le 99$, and each $n \in \{10^5, 10^6, 10^7, 10^8\}$ we computed the average of $SS(L) - s(L)$ and $BF(L) - s(L)$ over a set of random n-item instances (100, 32, 10, and 3 instances respectively) to obtain estimates of EW^n_{SS} and EW^n_{BF}. Instances were generated using the "shift register" random number generator described in [9, pages 171–172]. Previous experiments have shown that for bin packing simulations, this choice is unlikely to introduce significant biases.

These were the first instances we tested after proposing SS, and the results are even better than we had hoped. Whereas Best Fit's performance is as suggested in Section 1, with EW^n_{BF} apparently growing as $\Theta(n)$ for each j, $25 \le j \le 97$, EW^n_{SS} appeared to be $O(1)$ for all j, $1 \le j \le 98$. This is the same range for which $EW^n_{OPT} = O(1)$, and the results for $j = 99$ were consistent with $EW^n_{SS} = O(\sqrt{n})$, again the same value as for EW^n_{OPT}. Figure 1 depicts the average waste for SS as a function of j on a log scale, with the averages for each value of n connected in a curve. The log scale is necessary since although $EW^n_{SS}(U\{j, 100\})$ does not appear to grow with n for $j \le 98$, it does grow substantially with j. Note that the curves for each of the four values of n all more or less coincide, except possibly for j very close to 100. The solid curve is for

$n = 10^8$, and its higher variability is due to the fact that we tested only three instances of this size.

Table 1 shows the specific averages obtained for $j \in \{24, 25, 60, 97, 98, 99\}$. The first two values of j were chosen as these represent the critical region for Best Fit, where EW_{BF}^n makes a transition from $O(1)$ to $\Theta(n)$. The results for $j = 60$ are typical (except in precise values) of the broad range of j between 25 and 96. The results for $97, 98, 99$ display a critical region for both algorithms, as EW_{SS}^n goes from $O(1)$ to $\Theta(\sqrt{n})$ and EW_{BF}^n goes from $\Theta(n)$ to $O(1)$ to $\Theta(\sqrt{n})$. Our experiments for these last three values of j were extended to include instances with $n = 10^9$, as the rate of convergence is much slower when j is close to k. Although the variance is still sufficiently large for $j = 98$ that we would need substantially more samples if we wanted to get good estimates of the constant to which the expected waste rates are converging, the fact that the ratios are bounded is strongly suggested by the data.

Alg	n	Samples	$j = 24$	25	60	97	98	99
SS	10^5	100	223	223	884	23,350	28,510	34,286
	10^6	32	233	249	894	48,896	70,453	105,277
	10^7	10	212	217	797	64,997	150,291	343,958
	10^8	3	267	213	779	82,378	321,068	1,232,118
	10^9	3				68,719	184,328	3,512,397
BF	10^5	100	78	167	16,088	22,669	24,736	25,532
	10^6	32	76	831	154,460	59,015	77,831	88,258
	10^7	10	102	7,737	1,536,747	213,447	185,870	277,278
	10^8	3	67	75,546	15,340,879	1,800,011	254,235	1,081,251
	10^9	3				17,607,786	187,061	2,757,530

Table 1. Measured waste rates for SS and BF under distributions $U\{j, 100\}$.

As suggested by Figure 1 and Table 1, for fixed n the average waste for SS increases monotonically and fairly smoothly with j, but follows a more adventuresome path for BF. More details on the behavior of BF are reported in [3]. For now it is interesting to note on behalf of Best Fit that although the average waste for BF is enormously larger that that for SS when $25 \le j \le 97$ and EW_{BF}^n appears to grow linearly, the situation is different when EW_{BF}^n is sublinear, as it is for $1 \le j \le 24$ and for $j \in \{98, 99\}$. In these cases its value for fixed n is typically significantly lower than that for EW_{SS}^n, even though the latter has the same growth rate to within a constant factor.

Similar positive results for SS were obtained for $U\{j, k\}$ distributions with other values of k, leading us to conjecture that $EW_{SS}^n = O(EW_{OPT}^n)$ for all such distributions. Could something like this conjecture extend to even wider ranges of discrete distributions? A fundamental stumbling block to investigating this question lies in the fact that, in general, determining $EW_{OPT}^n(F)$ given F is an NP-hard problem. The $U\{j, k\}$ distributions are to date the most complicated special cases for which the answers have been obtained analytically. Fortunately, there is a way around this obstacle, at least for moderate values of B.

3 How to Determine EW_{OPT}^{n}

In order to test the conjecture made in the previous section, we need a way of determining $EW_{OPT}^{n}(F)$, given a discrete distribution F. It turns out that the slightly simpler question of whether $EW_{OPT}^{n}(F) = o(n)$ can be formulated as a surprisingly simple linear program related to standard network flow models. Suppose our discrete distribution consists of item sizes s_i, $1 \le i \le J$, with the probability that s_i occurs being p_i, and let B be the bin size. Our program will have $J(B+1)$ variables $v(i,g)$, $1 \le i \le J$ and $0 \le g \le B$, where $v(i,g)$ represents the rate at which items of size s_i go into bins with gap g. The constraints are:

$$\begin{aligned}
v(i,g) &= 0, && s_i > g \\
\sum_{g=1}^{B} v(i,g) &= p_i, && 1 \le i \le J \\
\sum_{i=1}^{J} v(i,g) &\le \sum_{j=1}^{J} v(j, g+s_j), && 1 \le g \le B-1
\end{aligned}$$

where the value of $v(j, g+s_j)$ when $g + s_j > B$ is taken to be 0 by definition for all j. The first set of constraints says that no item can go into a gap that is smaller than it. The second set says that all items must be packed. The third says that bins with a given gap are created at least as fast as they disappear. The goal is to minimize

$$\sum_{g=1}^{B-1} \left(g \cdot \left(\sum_{j=1}^{J} v(j, g+s_j) - \sum_{i=1}^{J} v(i,g) \right) \right)$$

that is, the rate at which waste space is created.

Let $c(F)$ be the optimal solution value for the above LP, and let $s(F) = \sum_{i=1}^{J} s_i p_i$ be the average item size under F. Then it can be shown based on results in [7] that $ER_{OPT}^{\infty} = c(F)/s(F)$ and if $c(F) = 0$, then EW_{OPT}^{n} is either $\Theta(\sqrt{n})$ or $O(1)$.

Moreover, in the latter case, the determination of which growth rate applies can be made by solving J additional LP's, one for each item size: In the LP for item size s_i, we add an additional variable $x \ge 0$, replace the constraint $\sum_{g=1}^{B} v(i,g) = p_i$ by $\sum_{g=1}^{B} v(i,g) = p_i + x$, add a constraint setting the original objective function to 0, and attempt to maximize x. If the optimal value for x is 0 in any of these LP's, then $EW_{OPT}^{n} = \Theta(\sqrt{n})$, otherwise it is $O(1)$, again by results in [7].

Using the software packages `AMPL` and `CPLEX`, we have created an easy-to-use system for generating, solving, and analyzing the solutions of these LP's, given B and a listing of the s_i's and p_i's, or given the parameters h, j, k of an interval distribution. In the next section we describe our results for such distributions.

4 Experiments with General Interval Distributions

In Section 2 we raised the question of whether our conjecture that $EW_{SS}^n = O(EW_{OPT}^n)$ holds for all distributions $U\{j,k\}$ might extend to broader classes of distributions. A natural class to consider is that of the interval distributions $U\{h..j,k\}$, as defined in Section 1.

Before summarizing our experiments with such distributions, however, we must first make admit that they caused us to make a slight modification to our conjecture. It turns out that there are interval distributions with $EW_{OPT}^n = O(1)$ for which EW_{SS}^n appears unavoidably to be $\Omega(\log(n))$. Consider for example $U\{2..3,9\}$, a simple distribution with $EW_{OPT}^n = O(1)$. Note that for $M >> \limsup_{n\to\infty} EW_{OPT}^n$, a sequence of M items of size 2 is likely to create $\Theta(M)$ bins with gap 1 under SS. But gaps of size 1 can never be filled, because there are no items of size 1. Sequences of M consecutive items of size 2 will be rare for large M, but instances with 2^M items can be expected to contain at least one such sequence. This implies that the expected waste will be $\Omega(\log n)$, although the constant of proportionality may be quite small. As an empirical verification, consider Table 2, which summarizes results for runs of SS for instances based on $U\{2..3,9\}$ with n ranging from 10^4 to 10^{10}. Note that the average waste does appear to grow roughly as $\Theta(\log n)$.

n	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
# Samples	10000	3162	1000	316	100	32	10
Average Waste	7.6	8.6	10.1	10.8	12.1	12.6	14.5
95% Conf. Int.	±0.1	±0.1	±0.2	±0.4	±0.8	±1.0	±1.9

Table 2. Measured average waste for SS under distributions $U\{2..3;9\}$.

We shall thus revise our conjecture about SS and split it into two parts:

Conjecture 1 $EW_{OPT}^n(F) = O(\sqrt{n})$ implies $EW_{SS}^n(F) = O(\sqrt{n})$.
Conjecture 2 $EW_{OPT}^n(F) = O(1)$ implies $EW_{SS}^n(F) = O(\log(n))$.

To test these conjectures, we investigated interval distributions $U\{h..j,k\}$ for two specific values of k, namely $k = 19$ and $k = 100$. For $k = 19$, we tested all pairs $h \le j < k$ with $h \le 9$ using the techniques of the previous section to determine ER_{OPT}^∞ and EW_{OPT}^n and then testing SS and BF on collections of randomly generated instances for the given distribution with $n \in \{10^5, 10^6, 10^7\}$. Pairs h, j with $h \ge 10$ were omitted since for these distributions BF, SS, and OPT all simply place one item per bin and unavoidably have a $\Theta(n)$ expected waste growth. The results are summarized in Table 3.

The "expected" waste rates for OPT in Table 3 are theorems, as determined using the LP's of Section 3, whereas those for SS and BF are for the most part conjectures with which our data is consistent. (We do have proofs for some of the

j	Alg	$h=1$	2	3	4	5	6	7	8	9
18	OPT	$\sqrt{n}$	n	n	n	n	n	n	n	n
	SS	$\sqrt{n}$	n	n	n	n	n	n	n	n
	BF	$\sqrt{n}$	n	n	n	n	n	n	n	n
17	OPT	1	$\sqrt{n}$	n	n	n	n	n	n	n
	SS	1	$\sqrt{n}$	n	n	n	n	n	n	n
	BF	1	$\sqrt{n}$	n	n	n	n	n	n	n
16	OPT	1	n	$\sqrt{n}$	n	n	n	n	n	n
	SS	1	n	$\sqrt{n}$	n	n	n	n	n	n
	BF	n	n	$\sqrt{n}$	n	n	n	n	n	n
15	OPT	1	1	n	$\sqrt{n}$	n	n	n	n	n
	SS	1	$\log n$	n	$\sqrt{n}$	n	n	n	n	n
	BF	n	n	n	$\sqrt{n}$	n	n	n	n	n
14	OPT	1	1	$\sqrt{n}$	n	$\sqrt{n}$	n	n	n	n
	SS	1	$\log n$	$\sqrt{n}$	n	$\sqrt{n}$	n	n	n	n
	BF	n	n	n	n	$\sqrt{n}$	n	n	n	n
13	OPT	1	1	1	n	n	$\sqrt{n}$	n	n	n
	SS	1	$\log n$	$\log n$	n	n	$\sqrt{n}$	n	n	n
	BF	n	n	n	n	n	$\sqrt{n}$	n	n	n
12	OPT	1	1	1	n	n	n	$\sqrt{n}$	n	n
	SS	1	$\log n$	$\log n$	n	n	n	$\sqrt{n}$	n	n
	BF	n	n	n	n	n	n	$\sqrt{n}$	n	n
11	OPT	1	1	1	1	n	n	n	$\sqrt{n}$	n
	SS	1	$\log n$	$\log n$	$\log n$	n	n	n	$\sqrt{n}$	n
	BF	n	n	n	n	n	n	n	$\sqrt{n}$	n
10	OPT	1	1	1	1	n	n	n	n	$\sqrt{n}$
	SS	1	$\log n$	$\log n$	$\log n$	n	n	n	n	$\sqrt{n}$
	BF	1	n	n	n	n	n	n	n	$\sqrt{n}$
9	OPT	1	1	1	n	n	n	n	n	n
	SS	1	$\log n$	$\log n$	n	n	n	n	n	n
	BF	1	n	n	n	n	n	n	n	n
8	OPT	1	1	1	1	n	n	n	n	
	SS	1	$\log n$	$\log n$	$\log n$	n	n	n	n	
	BF	1	n	n	n	n	n	n	n	
7	OPT	1	1	1	1	n	n	n		
	SS	1	$\log n$	$\log n$	$\log n$	n	n	n		
	BF	1	n	n	n	n	n	n		
6	OPT	1	1	1	n	n	n			
	SS	1	$\log n$	$\log n$	n	n	n			
	BF	1	n	n	n	n	n			
5	OPT	1	1	1	n	n				
	SS	1	$\log n$	$\log n$	n	n				
	BF	1	n	n	n	n				
4	OPT	1	1	1	n					
	SS	1	$\log n$	$\log n$	n					
	BF	1	n	n	n					
3	OPT	1	1	n						
	SS	1	$\log n$	n						
	BF	1	n	n						
2	OPT	1	n							
	SS	1	n							
	BF	1	n							

Table 3. Orders of magnitude of the measured waste rates under distributions $U\{h..j, 19\}$.

$h = 1$ entries for BF, in particular those for $j \in \{2, 3, 4, 17, 18\}$ [3,8].) Overall the data is consistent with Conjectures 1 and 2. The values of n tested were not sufficiently large for our measurements to make a convincing case for the $\log n$ growth rates reported for SS in the table; in many cases one might just as well have conjectured $EW_{SS}^n = O(1)$. However, in each of these cases the same sort of argument as was used above for $U\{2..3, 9\}$ applies.

Let us now consider the distributions $U\{h..j, 100\}$. Figure 2 graphically represents the values for EW_{OPT}^n for such distributions, where an entry of "–" represents $O(1)$, an entry of "+" represents $\Theta(\sqrt{n})$, and an entry of "·" represents $\Theta(n)$. Note that this picture appears to be a refinement of the structure apparent in Table 2. Moreover, if one ignores the distinction between –'s and +'s, it is a fairly accurate discretization of the results for the continuous uniform distributions $U[a, b]$, $0 \le a \le b \le 1$, depicted in Figure 5.2 of [6], which partitions the unit square into regions depending on whether $ER_{OPT}^{\infty}(U[a, b])$ is equal to or greater than 1.

There are far too many $U\{h..j, 100\}$ distributions for us to test SS and BF on them all. We therefore settled for testing isolated examples plus what looks like a challenging slice through Figure 2 – the distributions with $h = 18$. This is a particularly interesting value for h because of the large number of transitions that EW_{OPT}^n makes as j goes from h up to $k - 1$. In all cases, our experimental results were consistent with Conjectures 1 and 2. There is not room here to present the results in detail, but a high-level view of the performance of SS, BF and OPT is presented in Figure 3. This figure graphs the average values $BF(L)/s(L)$ and $SS(L)/s(L)$ over three instances with $n = 10^8$ for each distribution $U\{18..j, 100\}$, $18 \le j \le 99$, and compares these to the value of $\lim_{n\to\infty} E[OPT(L_n)/s(L_n)]$, as determined by the LP's of Section 3. The figure gives a good indication of those j that produce linear waste for each algorithm, i.e., those yielding average values of $A(L)/s(L)$ greater than 1. (For this statistic, the variance between instances is insignificant, so that three instances suffice to give good estimates.) Note that SS has linear waste in precisely those cases where the optimal packing does, although typically the average value is significantly greater. BF has linear waste for all values of j except 82 (the one value that yields a symmetric distribution).

5 *SS* as a Self-Organizing Algorithm

Why does SS do so well? Clearly the idea that SS is simply making sure bins are available into which new items will fit exactly does not suffice as an explanation for performance as good as that we have observed. For example, under $U\{25, 100\}$ there are no items available that will fit exactly into gaps of size exceeding 25, even though the algorithm will tend to produce bins with those gaps if none exist. A possibly better explanation is the following. Because of the sum of squares criterion, the continuing creation of bins with a given gap will be inhibited unless there is some way for bins with that gap to continually disappear. One way for a bin to disappear is for it to have its gap exactly filled;

it then no longer contributes to any of the $N(g)$'s. Another way for a bin to disappear is for it to have its gap reduced to one that already disappears for another reason, for instance if the next *two* items it receives will together fill its gap exactly, or the next three, etc. Thus the algorithm will be driven to favor the creation of precisely those gaps that can (eventually) lead to perfectly packed bins, and the sum of squares criterion is possibly providing a subtle feedback mechanism to maintain the production of the various gaps at the appropriate rates. In other words, it can be thought of as organizing itself for a maximum rate of production of perfectly packed bins.

A likely way to see this organization in action would be to look at the average gap counts $N(g)$ for $1 \le g \le B-1$ as a function of n. Figures 4 through 7 display such *gap count profiles* for $n \in \{100, 200, 800, 50K, 100K, 200K, 400K, 800K\}$ for four different interval distributions. Here, averages are taken over 1,000 separate instances. In each figure, the solid line represents the profile for $n = 800K$. The variety in patterns confirms that SS is indeed organizing itself differently for different distributions, although only the profiles for $U\{18..27, 100\}$ (Figure 4) seem to correlate naturally with the above explanation. Indeed, for $U\{1..19, 100\} = U\{19, 100\}$, all the average counts are 1 or less, so any inhibiting effects of the counts must be fairly subtle. Further study is clearly needed.

6 Improving on the Performance of SS when $EW^n_{OPT} = \Theta(n)$

Extrapolating from the results reported in Sections 2 and 4, one might propose that Conjectures 1 and 2 of Section 4 both hold for *all* discrete distributions. However, although this would imply that $ER^\infty_{SS} = 1$ whenever $EW^n_{OPT} = o(n)$, it still allows the possibility that $ER^\infty_{SS} > 1$ when $EW^n_{OPT} = \Theta(n)$. We have already remarked that this appears to be the case for several of the $U\{18..j, 100\}$ distributions, as illustrated by Figure 3. A simple distribution for which ER^∞_{SS} provably exceeds 1 is $F = U\{2..2, 5\}$, i.e., the distribution in which all items have size 2 and $B = 5$. Here an optimal packing places two items in each bin for a total of $\lceil n/2 \rceil$ bins, whereas SS will create a bin with a single item in it for every two bins that contain two, yielding $ER^\infty_{SS} = 1.2$.

For any bin packing algorithm A, let us define

$$\max ER^\infty_A \equiv \sup\{ER^\infty_A(F) : F \text{ is a discrete distribution}\}$$

Generalizing the above example to the sequence of distributions $U\{2..2, 2m+1\}$, $m \to \infty$, we have $\max ER^\infty_{SS} \ge 1.5$.

It is thus natural to search for variants on SS that retain its good behavior when $EW^n_{OPT} = o(n)$, while yielding smaller values of $\max ER^\infty_A$. One idea is to add an additional "bin closing" rule to SS. By *closing* a bin we mean declaring it off limits for further items and removing it from the $N(g)$ counts. In SS, the closing rule is simply to close a bin with gap 0, i.e., one that is completely full, as soon as it is created. When $EW^n_{OPT} = \Theta(n)$, even the optimal packing ends up with $\Theta(n)$ incompletely-packed bins, so it might make sense for our algorithm to close some such incompletely-packed bins as well.

We have investigated several variants of SS based on *ad hoc* closing rules that seem to outperform the basic algorithm. However, by far the best results we have obtained via this approach use closing rules that are tailored to the distribution at hand. It appears that for each discrete distribution F there is a variant $SS1_F$, based on a distribution-dependent closing rule, such that $ER^{\infty}_{SS1_F}(F) = 1$.

The closing rule in question is derived from the optimal solution to the LP for F presented in Section 3, Let $v^*(i,g)$ be the value of the variable $v(i,g)$ in this solution, $1 \le i \le j$ and $0 \le g \le B$. For $0 \le g < B$, define

$$r_g \equiv \sum_{j=1}^{J} v(j, g+s_j) - \sum_{i=1}^{J} v(i,g)$$

Note that the r_g are non-negative due to the constraints of the LP, and they can be interpreted as the rate at which bins with final gap g are produced in an optimal packing. Our closing rule is the following: When a bin with gap g is created, check to see if the current number of closed bins with gap g is less than nr_g, where n is the number of items in the current packing. If so, close the bin.

Figure 8 depicts the measured average values of $SS1_F(L)/s(L)$ for the distributions $U\{18..j, 100\}$ and $n \in \{10^5, 10^6, 10^7, 10^8\}$, with the standard numbers of instances for each n. The average values of $A(L)/s(L)$ for each n are linked by a dashed line. As in Figure 3, the solid line connects the values of $\lim_{n\to\infty} E[OPT(L_n)/s(L_n)]$. Note that for each value of j, the average values of $SS1_F(L)/s(L)$ do seem to be converging to the expected ratio for OPT.

Interestingly, if we assume that Conjecture 1 holds for all discrete distributions, there is for each F a (less efficient) variant $SS2_F$ on $SS1_F$ that *provably* will have $ER^{\infty}_{SS2_F}(F) = 1$. Suppose F is as above, and let $R = \sum_{g=1}^{B-1} g \cdot r_g$, the rate at which a unit quantity of waste is produced in an optimal packing. The key idea behind $SS2_F$ is to imagine what would happen if (additional) unit-size items arrived at the same rate. It is not difficult to see that, after normalizing so that the sum of the probabilities again equals 1 (i.e., dividing by $1+R$), one would have a new probability distribution F' for which the solution value $c(F')$ for the LP of Section 3 equals 0. Hence $EW^n_{OPT}(F') = O(\sqrt{n})$. Thus, by Conjecture 1, SS will also have $O(\sqrt{n})$ expected waste for F'. However, note that we can simulate the SS packing of an instance generated according to F' by simply taking a list generated according to F and randomly introducing additional *imaginary* items of size 1 at rate R. The space taken up by these items represents wasted space in the actual packing, but we already knew such waste was unavoidable. The total expected waste is thus $EW^n_{OPT} + O(\sqrt{n})$, and so $ER^{\infty}_{SS2_F}(F) = 1$, as claimed.

Typically $SS2_F$ can be significantly slower than $SS1_F$, since it actually has to pack the imaginary items, and even though standard priority queue data structures can be used to reduce the time for packing them, there may well be a large number of them, especially when B is large. Moreover, the additional running time (and the theoretical guarantee, assuming Conjecture 1) does not seem to be justified by the results. We tested $SS2_F$ on the same test bed of $U\{18..j, 100\}$ instances as we did for $SS1_F$, and the resulting chart of average values of $A(L)/s(L)$ looked essentially identical to Figure 8.

What the *SS2* approach does have going for it is simplicity, and its potential as the basis for a learning algorithm in which one starts with no knowledge of F, but instead learns about it as one goes along. This is the type of approach proposed by [12] in a purely theoretical context. Here however it is quite practical, especially in the context of *SS2*, which only has one distribution-specific parameter to adjust, the rate $r(F)$ at which imaginary 1's are introduced. In our implementation, which we shall call SS^+, we pause to estimate the distribution at ever-increasing intervals, the ith stoppage occurring immediately after $10B * (2^i - 1)$ real items have arrived. During the pause, an estimate F_i of the distribution is derived based on the counts of items of each size received to date and the LP of Section 3 is solved for F_i. The resulting rate $r(F_i)$ is then used until the next stoppage. (Note that by starting with $r = 0$ and waiting for $10B$ items to have arrived before our first adjustment, we guarantee that the algorithm runs in polynomial time.) In tests, SS^+ worked essentially as well as the distribution-specific algorithms $SS1_F$ and $SS2_F$ on the $U\{18..j, 100\}$ distributions, again yielding essentially the same chart as Figure 8.

In practice, SS^+ is much slower than the algorithms that are given F in advance, because of the need to repeatedly construct and solve linear programs. It also requires access to an LP solver. Fortunately, the fact that *SS2* has only one distribution-specific parameter suggests an alternative approach: instead of learning F, why not simply learn $r(F)$? If Conjecture 2 is true for all discrete distributions, we should get good feedback on our estimates of r: If our current r is too small, then waste should clearly grow linearly, whereas if our current r is too big, Conjecture 2 in conjunction with results of [7] says that waste should grow only as $O(\log(n))$. If one waits for enough items to arrive and be packed, one should be able to distinguish between these two cases, and with some complicated algorithm engineering we have managed to implement this approach with some degree of success. The resulting algorithm SSA is too complicated to present in detail here, but Figure 9 summarizes the results for it on the same test bed covered for $SS1_F$ in Figure 8. Although the results are not quite as good as those for SS^+, they do again appear to be converging to the optimal expected waste.

7 Directions of Ongoing Research

Although the experiments reported in this paper have all been for selected interval distributions $U\{h..j, k\}$, we have in fact tested SS on a variety of distributions proposed to us, including various Zipf distributions and randomly-generated discrete distributions, as we shall describe more fully in the journal version of this paper. So far, we have been unable to find counterexamples to either of our two conjectures. In the case of Conjecture 1, this is not surprising, since Jim Orlin, after hearing a talk covering the results and open problems of this paper, has proved Conjecture 1. The journal version of this paper will add Jim as a co-author and include this result. It will also include a detailed proof that the LP's of Section 3 have the properties we claimed for them, as well as more detailed coverage of some of the other theoretical issues we have raised, such as bounds on $\max ER^{\infty}_{SS}$ and an investigation of the worst-case behavior of SS for arbitrary lists.

Another interesting question is how sacrosanct is the exponent of 2 in the definition of SS? What if we tried instead to minimize the sum of cubes, or the sum of 1.5 powers? Preliminary experiments suggest that these variants also satisfy Conjectures 1 and 2, although we should point out that Orlin's proof applies only when the exponent is precisely 2.

References

1. J. L. Bentley, D. S. Johnson, F. T. Leighton, and C. C. McGeoch. An experimental study of bin packing. In *Proceedings of the 21st Annual Allerton Conference on Communication, Control, and Computing*, pages 51–60, Urbana, 1983. University of Illinois.
2. E. G. Coffman, Jr., C. Courcoubetis, M. R. Garey, D. S. Johnson, P. W. Shor, R. R. Weber, and M. Yannakakis. Bin packing with discrete item sizes, part I: Perfect packing theorems and the average case behavior of optimal packings. *SIAM J. Disc. Math.* Submitted 1997.
3. E. G. Coffman, Jr., C. A. Courcoubetis, M. R. Garey, D. S. Johnson, L. A. McGeogh, 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*, pages 230–240. ACM Press, 1991.
4. E. G. Coffman, Jr., D. S. Johnson, P. W. Shor, and R. R. Weber. Bin packing with discrete item sizes, part IV: Average-case behavior of best fit. In preparation.
5. E. G. Coffman, Jr., D. S. Johnson, P. W. Shor, and R. R. Weber. Markov chains, computer proofs, and best fit bin packing. In *Proceedings of the 25th ACM Symposium on the Theory of Computing*, pages 412–421, New York, 1993. ACM Press.
6. E. G. Coffman, Jr. and G. S. Lueker. *Probabilistic Analysis of Packing and Partitioning Algorithms.* Wiley, New York, 1991.
7. C. Courcoubetis and R. R. Weber. Stability of on-line bin packing with random arrivals and long-run average constraints. *Prob. Eng. Inf. Sci.*, 4:447–460, 1990.
8. C. Kenyon, Y. Rabani, and A. Sinclair. Biased random walks, Lyapunov functions, and stochastic analysis of best fit bin packing. *J. Algorithms*, 27:218–235, 1998.
9. D. E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms.* 2nd Edition, Addison-Wesley, Reading, MA, 1981.
10. T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica*, 9:161–187, 1989.
11. W. T. Rhee and M. Talagrand. On line bin packing with items of random size. *Math. Oper. Res.*, 18:438–445, 1993.
12. W. T. Rhee and M. Talagrand. On line bin packing with items of random sizes – II. *SIAM J. Comput.*, 22:1251–1256, 1993.
13. P. W. Shor. The average-case analysis of some on-line algorithms for bin packing. *Combinatorica*, 6(2):179–200, 1986.

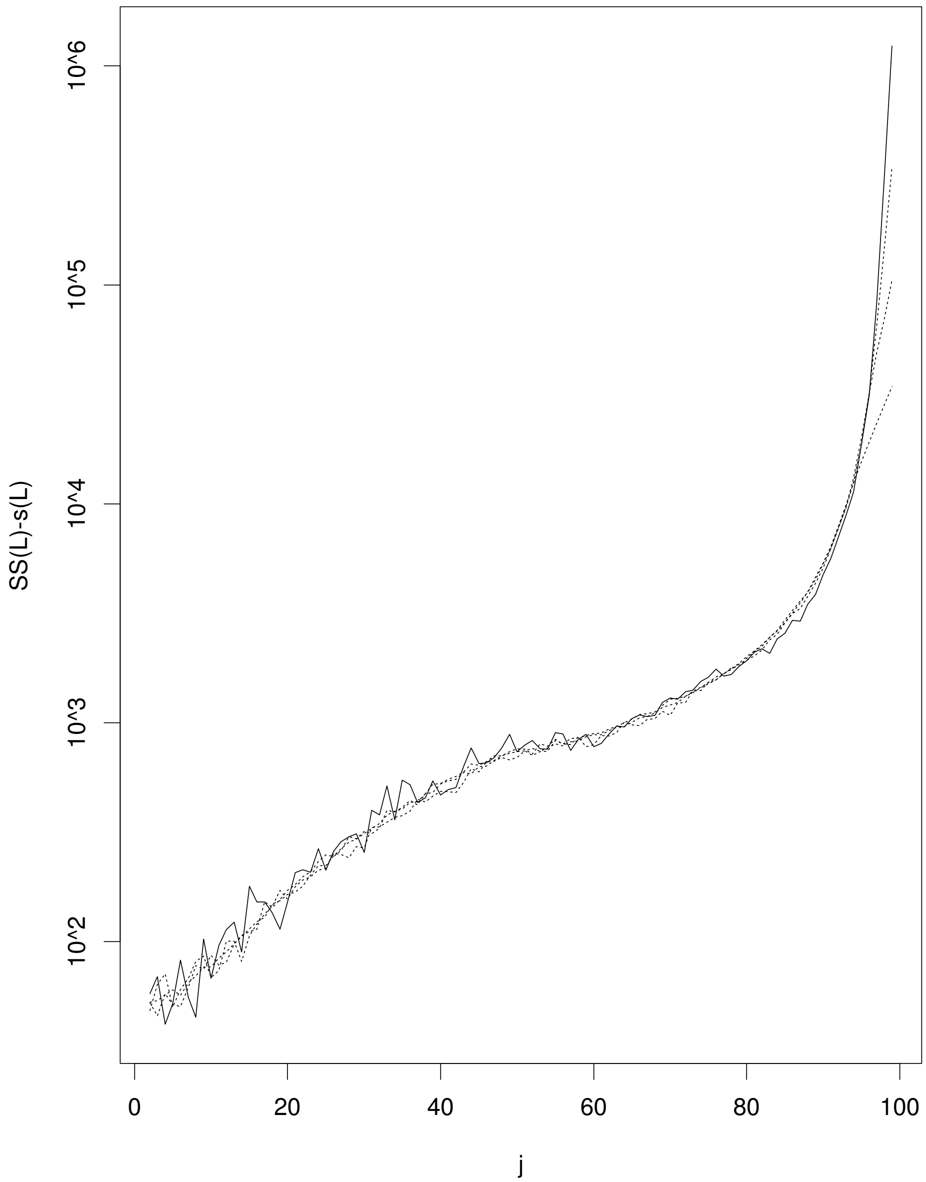

Fig. 1. Average values of $SS(L) - s(L)$ for distributions $U\{j, 100\}$, $1 \le j \le 98$ and $n \in \{10^5, 10^6, 10^7, 10^8\}$.

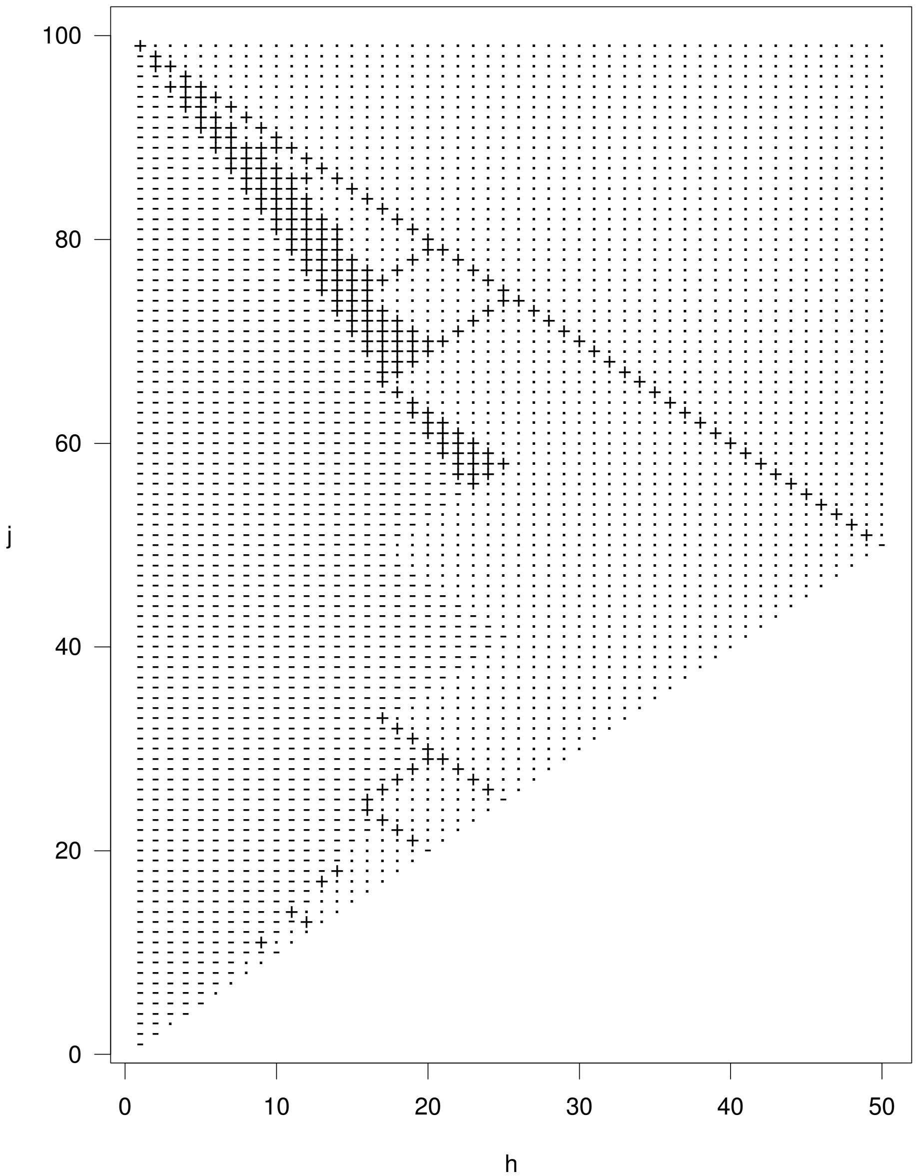

Fig. 2. $EW^n_{OPT}(U\{h..j,100\})$: "$-$" means $O(1)$, "+" means $\Theta(\sqrt{n})$, and "$\cdot$" means $\Theta(n)$.

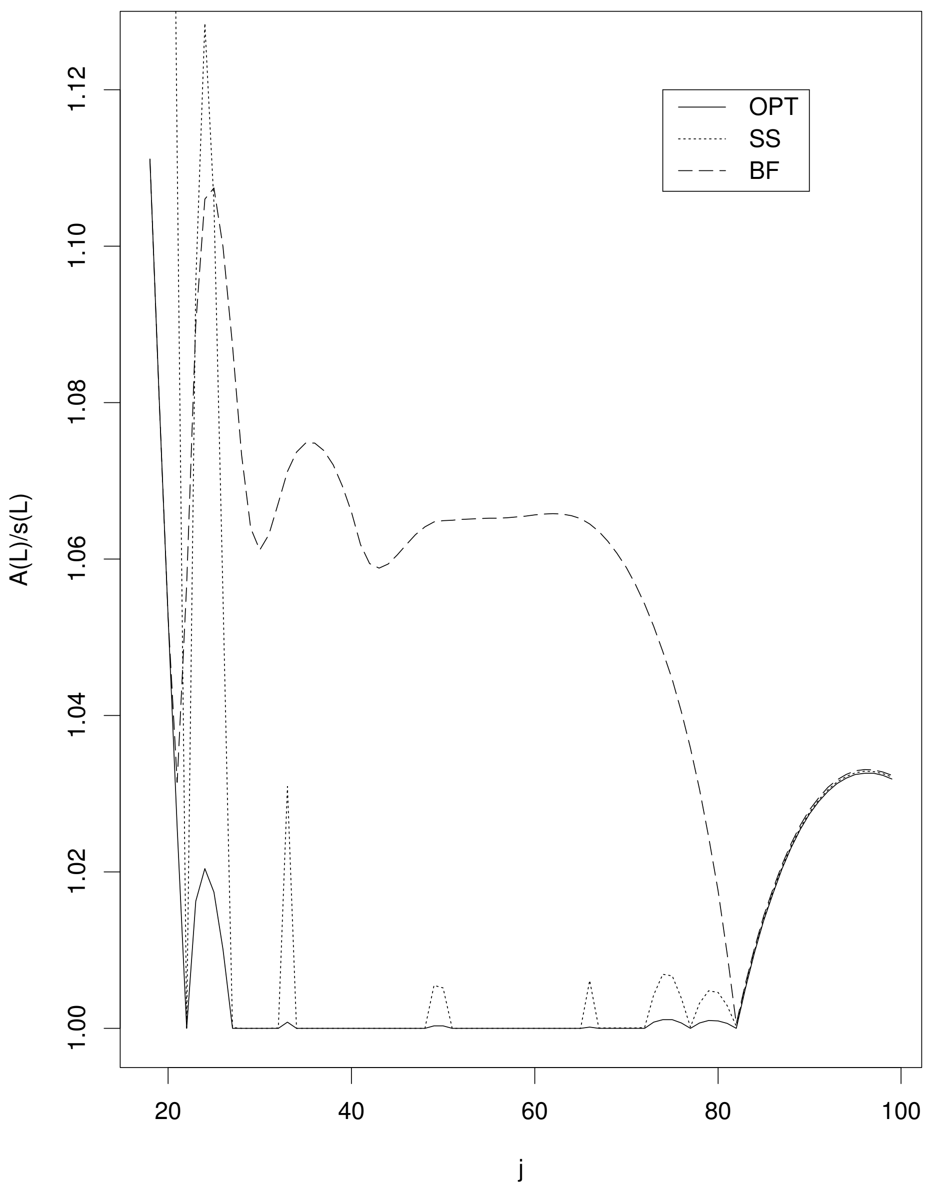

Fig. 3. Average values of $A(L)/s(L)$ for distributions $U\{18..j, 100\}$, $18 \leq j \leq 99$.

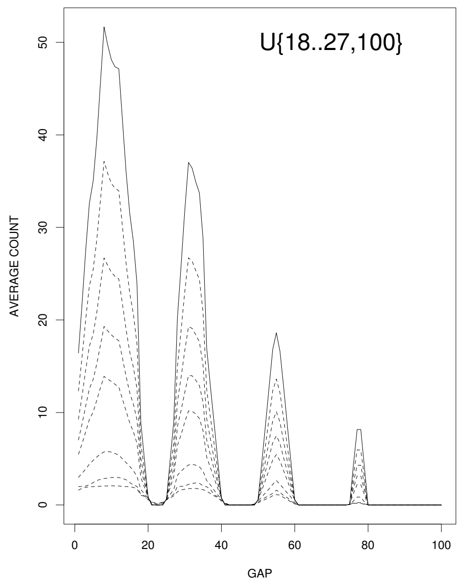

Fig. 4. Gap profiles for $U\{18..27, 100\}$.

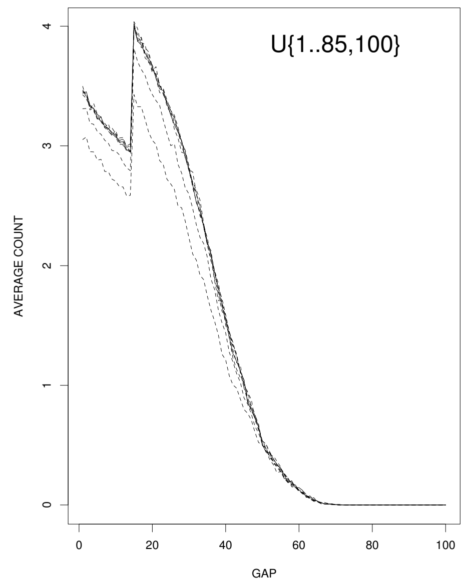

Fig. 6. Gap profiles for $U\{1..85, 100\}$.

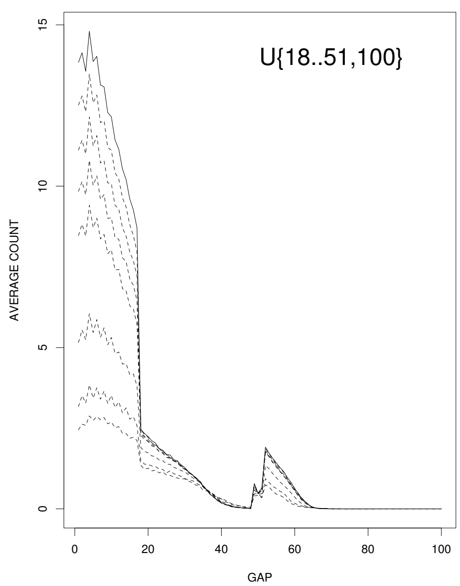

Fig. 5. Gap profiles for $U\{18..51, 100\}$.

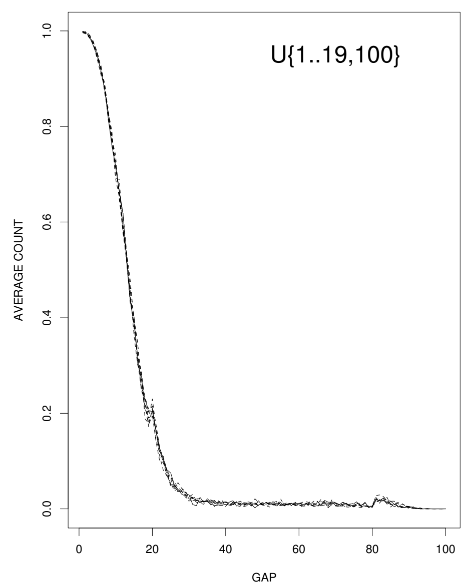

Fig. 7. Gap profiles for $U\{1..19, 100\}$.

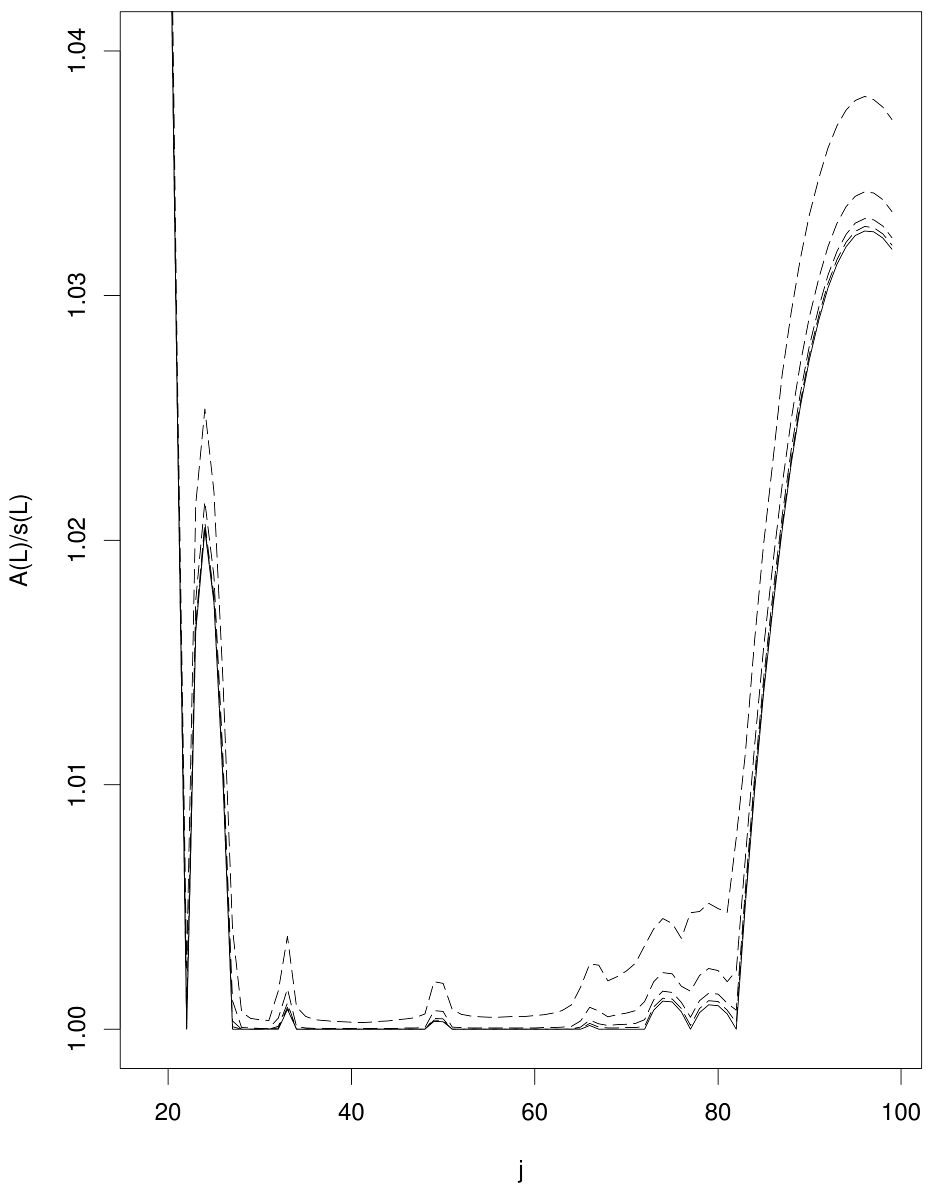

Fig. 8. Average values of $A(L)/s(L)$ for $SS1_F$ when $F = U\{18..j, 100\}$, $18 \le j \le 99$ and $n \in \{10^5, 10^6, 10^7, 10^8\}$

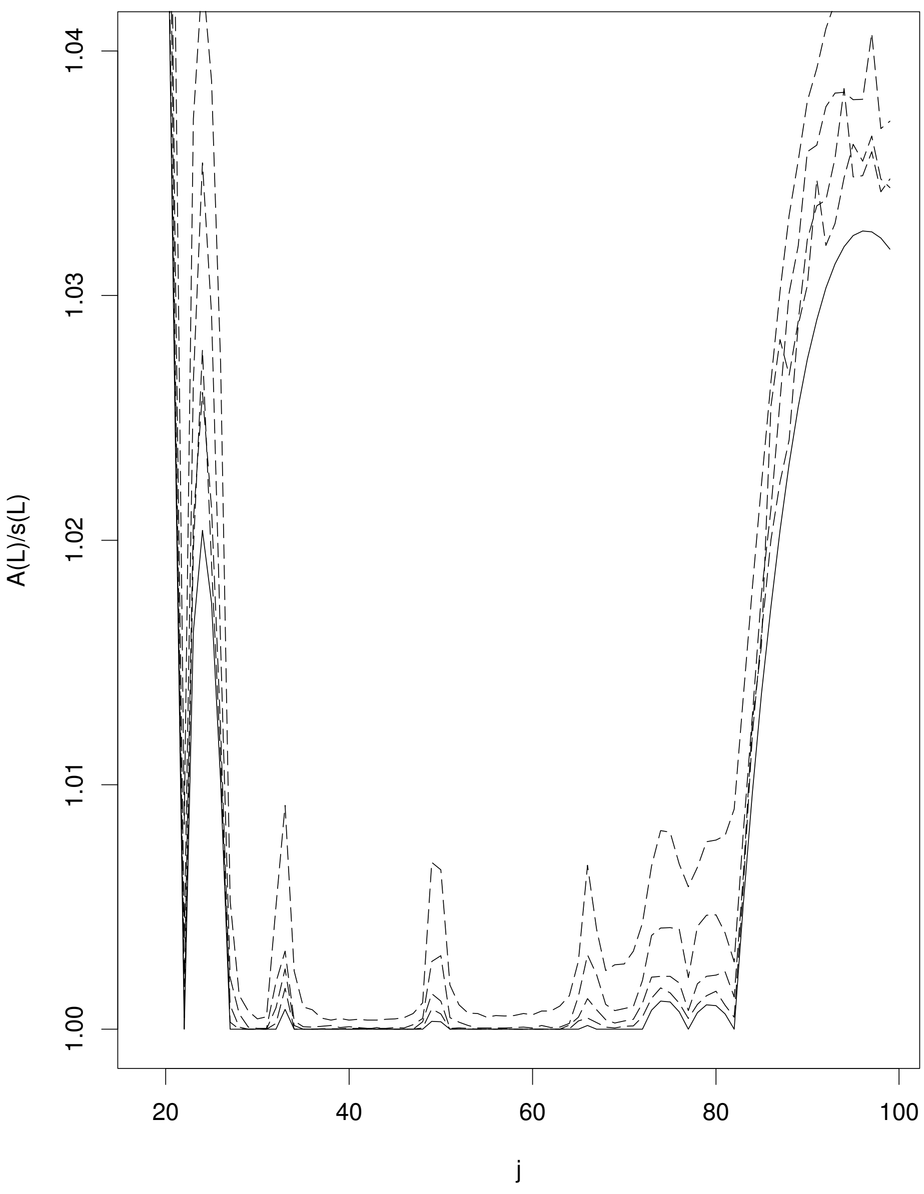

Fig. 9. Average values of $A(L)/s(L)$ for SSA when $F = U\{18..j, 100\}$, $18 \le j \le 99$ and $n \in \{10^5, 10^6, 10^7, 10^8\}$