Average case analysis for the Probabilistic Bin Packing Problem

M. Bellalouna, S. Souissi, and B. Ycart

ABSTRACT: *In the Probabilistic Bin Packing Problem (PBPP), some items are randomly deleted after having been placed into bins. The problem is to rearrange the remaining items, using the a priori solution. The initial arrangement being done with the Next Fit Decreasing heuristic, we consider two procedures. In the first one, the NF algorithm is applied to the new list. In the second one, successive groups of bins are optimally rearranged. In both cases, we prove a law of large numbers and a central limit theorem for the number of occupied bins as the initial number of items tends to infinity.*

1 Introduction

Bin Packing is a classical NP-hard problem of optimization [10]: given items of sizes $(x_1, \ldots, x_n)$, all smaller than 1, one must pack them into bins of size 1, so as to minimize the total number of non-empty bins. Many approximation heuristics have been proposed and studied: see Coffman et al. [5] for a survey. We shall focus here on the Next Fit Decreasing (NFD) heuristic [1]. Firstly the items are ranked in decreasing order (this can be done in $O(n \log(n))$ time). Then they are put into bins according to the the Next Fit algorithm: one bin is open at a time; when a new item has to be placed, either it fits in the open bin or it does not, in which case the current box is closed and a new one is opened (it takes $O(n)$ time to place the items once sorted). In the average case analysis, the item sizes are random variables, and so is the number of non-empty bins. Its distribution has been thoroughly studied, in particular by Csirik et al. [7, 6], Hofri and Kamhi [13] and Rhee [17, 18] (see also section 5.2 of [11] or section 10.3 of [12]).

The idea of so called Probabilistic Combinatorial Optimisation Problems comes from Jaillet [14, 15] who introduced it for the Traveling Salesman Problem (see also [3, 4]). The Probabilistic Bin Packing Problem (PBPP) was first studied in [2]. The idea is the following. Assume that a list of n items has been given, and an *a priori* solution (exact or approximate) has been found for the BPP. Suppose now that some items randomly disappear from the list. Can the knowledge of the a priori solution for the full list be used to construct a solution for the reduced one? Can this be done efficiently without reopening simultaneously too many bins of the a priori solution?

The aim of this paper is to propose an average case analysis of the PBPP, when the a priori solution is obtained through the NFD heuristic. We shall deal with two sources of randomness. Firstly, the initial sizes of the items are independent and identically distributed random variables (i.i.d.r.v.'s) $(X_1, \ldots, X_n)$. Secondly, once they have been sorted in decreasing order and placed into bins by the NF algorithm, a random binary decision is taken: for each $i = 1, \ldots, n$ the item number i remains or disappears. To formalize this, we consider a n-tuple $(U_1, \ldots, U_n)$ of i.i.d.r.v.'s, uniformly distributed on $[0, 1]$. The two random vectors $(X_i)_{i=1,\ldots,n}$ and $(U_i)_{i=1,\ldots,n}$ are independent. The probability for an item to stay in the list may depend on its size: we denote by $\rho(x)$ the probability for an item of size x to stay in the list. It will be convenient to view disappearing items as objects whose size has become null. Thus the new list of sizes is $(Y_1, \ldots, Y_n)$, where for $i = 1, \ldots, n$:

$$Y_i = X_i \, \mathbb{I}_{\{U_i \leq \rho(X_i)\}} \ ,$$

denoting by $\mathbb{I}_A$ the indicator of an event A.

If the NFD heuristic has been used for the a priori solution, an obvious procedure immediately comes to mind. Since the initial items were ranked in decreasing order, so are the remaining ones, and it is fast and natural to apply again from scratch the NF algorithm to the list of remaining items. The average case analysis of this procedure is proposed in section 2. The total number of bins will be proved to satisfy a law of large numbers and a central limit theorem, and an explicit expression for the asymptotic mean and variance will be given (theorem 2.3).

However, it is not in the spirit of Probabilistic Combinatorial Problems not to use the a priori solution once the items have been randomly deleted. Therefore, we shall study another heuristic. Suppose the a priori NFD solution has been computed, its bins being numbered by order of opening. Once the items have been randomly deleted, vacancy is left in some of the bins. The Group Rearrangement (GR) procedure depends on a fixed integer m which is the number of bins to be opened simultaneously. Here is the algorithm.

1. Open the bins of the a priori solution by groups of m, one group at a time: first bins with numbers 1 to m, then $m + 1$ to $2m$, and so on...

2. For each group of m bins, rearrange the remaining items in an optimal way.

3. Eliminate those bins that have been emptied.

The average case analysis of the GR procedure is treated in section 3. Again, a law of large numbers and a central limit theorem for the total number of non-empty bins will be proved (theorem 3.1).

Of course, the GR procedure is neither faster, nor better on average than the NFD heuristic: both run in linear time, and the asymptotic mean number of bins is larger for the former than for the latter. However, numerical evidence shows that the difference is small. We are not able at this point to propose a similar study for the optimal a priori solution. But we consider our NFD results as a reason to believe that local rearrangements inside small sized groups of bins, such as in the GR procedure, may bring a fast and relatively good solution to the PBPP, when starting from an a priori solution, be it optimal or not.

2 PBPP by the NFD heuristic

In this section, we study the asymptotics of the total number of bins filled by the NFD heuristic, for items of random sizes, once some of them have been randomly deleted.

Two independent sequences of i.i.d.r.v.'s are given: $(X_i)_{i \geq 1}$ and $(U_i)_{i \geq 1}$. The X_i's are the sizes of the original items, and the U_i's are the random variables that decide of their deletion. The probability distribution function of the X_i's is denoted by F and the U_i's have uniform distribution on $[0, 1]$. A measurable function ρ, from $[0, 1]$ into itself is given. If x is an item size, $\rho(x)$ is its probability to remain in the new list. As already pointed out, it is convenient to consider deleted objects as items of size 0. Thus the new list of item sizes after random deletions becomes $(Y_i)_{i\geq 1}$, where for all $i \geq 1$:

$$Y_i = X_i \, \mathbb{I}_{\{U_i \leq \rho(X_i)\}} \; .$$

Notice that the Y_i's are still i.i.d.r.v.'s. Denote by A_n the number of bins used by the NFD algorithm to arrange the n items of sizes $Y_1, \ldots, Y_n$. The asymptotic study of A_n requires very little adaptation of the classical proof for uniformly distributed sizes, developped by Csirik et al. [6] (see Hofri [12] section 10.3.1, p.543 ff.). We shall review below the main arguments. We are aware of the more precise approach of Rhee [17], who gives a much better bounding for A_n than that of lemma 2.2. The reason why we chose Csirik et al.'s truncation technique is that it can also be used for the GR procedure, to be treated in section 3.

The first observation is that the number of bins depends more on their *types* rather than on their actual sizes.

Definition 2.1 *For $k \geq 1$, an item is said to be of type k if its size x is such that*

$$\frac{1}{k+1} < x \leq \frac{1}{k} \; .$$

Thus a bin can accomodate exactly k objects of type k. To account for deletions, we shall agree that an item of size 0 has type 0. With our probabilistic hypotheses, the item types are i.i.d.r.v.'s with values in $\mathbb{N}$. We shall denote by $p = (p_k)_{k\in\mathbb{N}}$ their distribution. For $k \geq 1$, the probability for an item to be of type k is

$$p_k = \int_{\frac{1}{k+1}}^{\frac{1}{k}} \rho(x)\, dF(x) \; ,$$

whereas its probability to be of type 0 (deletion) is

$$p_0 = 1 - \sum_{k=1}^{\infty} p_k = 1 - \int_0^{\infty} \rho(x)\, dF(x) \; .$$

As a particular case, if the original item sizes are uniformly distributed on $[0,1]$ and the function ρ is constant, one gets $p_0 = 1 - \rho$ and for $k \geq 1$:

$$p_k = \frac{\rho}{k(k+1)} \; .$$

The results that follow only depend on the distribution p.

Since the items are examined in decreasing order of size, all items of type 1 are treated first, and placed alone in as many bins. Then come type 2 items. The first of them possibly fits in the same bin as the last type 1 item, the others are placed 2 by 2 into new bins, and so on... It is intuitively clear that, apart from a few "frontier" bins that may contain items of different types, most bins will host a fixed number of items of the same type. Lemma 2.2 below gives bounds on the number of used bins, in terms of two functions of the item types.

Lemma 2.2 *Let $r \geq 1$ be an integer. Define the two functions ϕ_1 and ϕ_2, from $[0,1]$ into itself by:*

$$\phi_1(x) \quad = \quad \begin{cases} \frac{1}{k} & \text{if} \quad \frac{1}{k+1} < x \leq \frac{1}{k}\ ,\ k=1,\ldots,r-1\ , \\ 0 & \text{if} \quad 0 \leq x \leq \frac{1}{r} \end{cases}$$

$$\phi_2(x) \quad = \quad \begin{cases} \frac{1}{k} & \text{if} \quad \frac{1}{k+1} < x \leq \frac{1}{k}\ ,\ k=1,\ldots,r-1\ , \\ \frac{1}{r} & \text{if} \quad 0 < x \leq \frac{1}{r}\ , \\ 0 & \text{if} \quad x = 0\ . \end{cases}$$

Let $y_1,\ldots,y_n$ be n (possibly null) item sizes. Let a_n be the total number of bins required to arrange those items using the NFD heuristic. Then:

$$\sum_{i=1}^{n} \phi_1(y_i) - (r-1) \leq a_n \leq \sum_{i=1}^{n} \phi_2(y_i) + r\ .$$

Proof: For $k \geq 1$, let n_k be the number of type k items. The number of bins they will occupy is at least $\lfloor \frac{n_k}{k} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part. Hence the lower bound, neglecting items of size $\leq r$. For the upper bound, all items of type $k \leq r-1$ can be accomodated in at most $\lfloor \frac{n_k}{k} \rfloor + 1$ bins, and all items of type $k \geq r$ in at most $\lfloor \frac{n_k}{r} \rfloor + 1$ bins. □

When the input sizes are random, lemma 2.2 provides bounds on A_n in terms of two sums of random variables:

$$\sum_{i=1}^{n} \phi_1(Y_i) - (r-1) \leq A_n \leq \sum_{i=1}^{n} \phi_2(Y_i) + r\ . \tag{1}$$

In (1), both $S_{1,n} = \sum \phi_1(Y_i)$ and $S_{2,n} = \sum \phi_2(Y_i)$ are sums of bounded i.i.d.r.v.'s. Their asymptotic behavior (exponential tail inequalities, law of large numbers, central limit theorem) is described by basic results of probability theory (see [8, 9] or [16] as general references). These can be passed to A_n, through a careful choice of the free parameter r.

For $r \geq 1$ denote by q_r the tail probability for the distribution of types:

$$q_r = \sum_{k=r}^{\infty} p_k = 1 - \sum_{k=0}^{r-1} p_k\ .$$

We will assume that q_r decreases at least as fast as some negative power of r: there exist two positive constants c and α such that for all $r \geq 1$,

$$q_r \leq c r^{-\alpha}\ . \tag{2}$$

This is actually an assumption on both the behavior of the size distribution F close to 0, and the function ρ.

Theorem 2.3 *Under the previous hypotheses, denote by μ and σ^2 the following asymptotic mean and variance:*

$$\mu = \sum_{k=1}^{\infty} \frac{p_k}{k}\ . \tag{3}$$

$$\sigma^2 = \sum_{k=1}^{\infty} \frac{p_k}{k^2} - \mu^2 \ . \tag{4}$$

Then the following results hold for A_n*.*

1. *Exponential tail inequality: for all* $x > 0$ *and* $n \geq 1$*,*

$$\mathbb{P}[\,|A_n - n\mu| \geq x\sqrt{n}|\,] \leq \exp\left(-2(x - n^{-1/2} - 2\sqrt{c}\,n^{-\frac{\alpha}{2(2+\alpha)}})^2\right) \ . \tag{5}$$

2. *Law of large numbers:*

$$\lim_{n\to\infty} \frac{A_n}{n} = \mu \quad a.s. \tag{6}$$

3. *Central limit theorem:*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sqrt{n\sigma^2}}(A_n - n\mu) \leq x\right] = \Phi(x) \ , \tag{7}$$

where Φ *denotes the standard Gaussian probability distribution function.*

For the particular case of item sizes uniformly distributed on $[0,1]$ and a constant value of ρ, the asymptotic mean and variance can be expressed in terms of Riemann's Zeta function:

$$\zeta(u) = \sum_{i=1}^{\infty} \frac{1}{n^u} \ .$$

One gets:

$$\mu = \rho\,(\zeta(2) - 1) \simeq 0.645\,\rho \ ,$$

and:

$$\sigma^2 = \rho\,(\zeta(3) - \zeta(2) + 1) - \rho^2\,(\zeta(2) - 1)^2 \simeq 0.557\,\rho - 0.416\,\rho^2 \ .$$

Proof: For $j = 1, 2$, we shall denote by μ_j the expectation of $\phi_j(Y_i)$, and by σ_j^2 its variance:

$$\mu_1 = \sum_{k=1}^{r-1} \frac{p_k}{k} \quad , \quad \mu_2 = \sum_{k=1}^{r-1} \frac{p_k}{k} + \frac{q_r}{r} \ ,$$

$$\sigma_1^2 = \sum_{k=1}^{r-1} \frac{p_k}{k^2} - \mu_1^2 \quad , \quad \sigma_2^2 = \sum_{k=1}^{r-1} \frac{p_k}{k^2} + \frac{q_r}{r^2} - \mu_2^2 \ .$$

Obviously as r tends to infinity, μ_1 and μ_2 tend to μ, whereas σ_1^2 and σ_2^2 tend to σ^2. From (1), the following bounds on the PDF of A_n are easily deduced.

$$\begin{aligned} \mathbb{P}[S_{2,n} - n\mu_2 \leq x + r + n\tfrac{q_r}{r}] &\leq \mathbb{P}[A_n - n\mu \leq x] \\ &\leq \mathbb{P}[S_{1,n} - n\mu_1 \leq x - r - n\tfrac{q_r}{r}] \ . \end{aligned} \tag{8}$$

As sums of bounded i.i.d.r.v.'s, both $S_{1,n}$ and $S_{2,n}$ satisfy classical exponential tail inequalities, such as Hoeffding's [16]:

$$\mathbb{P}[|S_{j,n} - n\mu_j| \geq z\sqrt{n}] \leq 2e^{-2z^2} \ .$$

Both also satisfy the central limit theorem:

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sqrt{n}}(S_{j,n} - n\mu_j) \leq x\right] = \Phi(\frac{x}{\sigma_j}) \ .$$

To derive from (8) the corresponding results for A_n, one needs to let the parameter r tend to infinity, as a function of n. Recall from (2) that $q_r \leq c r^{-\alpha}$. For all $n \geq 1$, we set $r = r(n)$ as follows:

$$r(n) = \lfloor \sqrt{c}\, n^{\frac{1}{2+\alpha}} \rfloor + 1 \ . \tag{9}$$

One readily checks that:

$$n\frac{q_{r(n)}}{r(n)} \leq \sqrt{c}\, n^{\frac{1}{2+\alpha}} \ .$$

From there, (5) easily follows from Hoeffding's inequality applied to $S_{1,n}$ and $S_{2,n}$. To derive the strong law of large numbers from an exponential inequality such as (5) is an easy application of the Borel-Cantelli lemma. For the central limit theorem (7), one just has to divide variables by $\sqrt{n}$ in (8), and let n tend to infinity: with our choice of $r(n)$, both $\frac{r(n)}{\sqrt{n}}$ and $\frac{\sqrt{n}\, q_{r(n)}}{r(n)}$ tend to 0. □

3 The Group Rearrangement procedure

We now turn to the GR procedure, in which bins are taken by groups of m, each group being optimally rearranged after deletions. As we saw before, in the a priori solution most bins contain items of the same type. Therefore, most groups of bins will be homogeneous in the sense that each bin of the group contains exactly k items of type k; let us call "k-group" such a group of m bins. We need to restrict slightly our assumption on the function ρ: we assume now that it is constant for objects of the same type, and denote by ρ_k the probability for an item of type k to remain in the list.

Consider a k-group. The number of remaining items in all its m bins, has binomial distribution with parameters mk and ρ_k. Let us denote by $\pi_{k,m} = (\pi_{k,m}(i))_{i=0,\ldots,m}$ the probability distribution of the number of remaining bins, once the k-group has been rearranged:

$$\pi_{k,m}(i) = \begin{cases} (1-\rho_k)^{mk} & \text{for } i = 0 \ , \\ \displaystyle\sum_{l=(i-1)k+1}^{ik} \binom{mk}{l} \rho_k^l (1-\rho_k)^{mk-l} & \text{for } i = 1, \ldots, m \ . \end{cases}$$

The expectation and the variance of $\pi_{k,m}$ will be denoted by $e_{k,m}$ and $v_{k,m}$ respectively.

$$e_{k,m} = \sum_{i=0}^{m} i\, \pi_{k,m}(i) \quad \text{and} \quad v_{k,m} = \sum_{i=0}^{m} i^2\, \pi_{k,m}(i) - e_{k,m}^2 \ .$$

Let p_k^* be the *initial* proportion of type k items: for $k \geq 1$,

$$p_k^* = F(\frac{1}{k}) - F(\frac{1}{k+1}) \; .$$

We shall make the same assumption on the tail of p^* as we did for p: there exist two positive constants c and α such that for all $r \geq 1$,

$$q_r^* = \sum_{k=r}^{\infty} p_k^* \leq c r^{-\alpha} \; . \tag{10}$$

Let $B_{n,m}$ be the number of remaining bins after the GR procedure. The asymptotics of $B_{n,m}$ is described in the following result.

Theorem 3.1 *Under the previous hypotheses, denote by μ_m and σ_m^2 the following asymptotic mean and variance.*

$$\mu_m = \sum_{k=1}^{\infty} p_k^* \, \frac{e_{k,m}}{km} \; , \tag{11}$$

$$\sigma_m^2 = \sum_{k=1}^{\infty} p_k^* \left(\frac{v_{k,m}}{km} + \frac{e_{k,m}^2}{k^2 m^2} \right) \; - \; \mu_m^2 \; . \tag{12}$$

Then the following results hold for $B_{n,m}$.

1. *Law of large numbers:*

$$\lim_{n \to \infty} \frac{B_{n,m}}{n} = \mu_m \quad a.s. \tag{13}$$

2. *Central limit theorem:*

$$\lim_{n \to \infty} \mathbb{P} \left[\frac{1}{\sqrt{n \sigma_m^2}} (B_{n,m} - n\mu_m) \leq x \right] = \Phi(x) \; . \tag{14}$$

Essentially, $B_{n,m}$ behaves asymptotically as a sum of n i.i.d.r.v.'s, each having expectation μ_m and variance σ_m^2; they can be viewed as the individual contributions of the original n items to the final packing. Indeed, for n large, a typical item belongs to a k-group with probability p_k^*. The contribution of km such items (one k-group) to $B_{n,m}$ has expectation $e_{k,m}$ and variance $v_{k,m}$; hence each contribution should have expectation $\frac{e_{k,m}}{km}$, and variance $\frac{v_{k,m}}{km}$. Thus the expected squared contribution of an object of type k should be $\frac{v_{k,m}}{km} + \frac{e_{k,m}^2}{k^2 m^2}$. So σ_m^2 can be seen as the variance for the contribution of a typical item to $B_{n,m}$.

Clearly, as m increases, the space wasted by the GR procedure compared to the NFD heuristic diminishes. In particular, the asymptotic expectation μ_m defined by (11) tends to μ. One may wonder what is the difference for small values of m. To get a partial answer, we computed numerically $\mu_m - \mu$, for $m = 2 \ldots, 5$, in the particular case where the item sizes are uniformly distributed on $[0, 1]$ and the probability ρ is a constant. Figure 1 shows a plot of $\mu_m - \mu$ as a function of ρ.

It turns out that the difference between the global algorithm (NFD) and the local one (GR) is relatively small, even for $m = 2$. In order to understand why, let us fix m and ρ_k, and look at the asymptotic behavior of $e_{k,m}$ as k increases. The law of large numbers implies that $e_{k,m}$ converges to i for all values of ρ_k in the interval $]\frac{i-1}{m}, \frac{i}{m}]$, i ranging from 1 to m. In other terms, as k increases, $e_{k,m}$ approaches $\lfloor m\rho_k \rfloor + 1$. So μ_m is actually close to the following sum :

$$\mu_m \simeq \sum_{k=1}^{\infty} p_k^* \frac{\lfloor m\rho_k \rfloor + 1}{km} ,$$

to be compared with

$$\mu = \sum_{k=1}^{\infty} p_k^* \frac{\rho_k}{k} .$$

This also accounts for the modes in $\mu_m - \mu$, plotted as a function of ρ (figure 1).

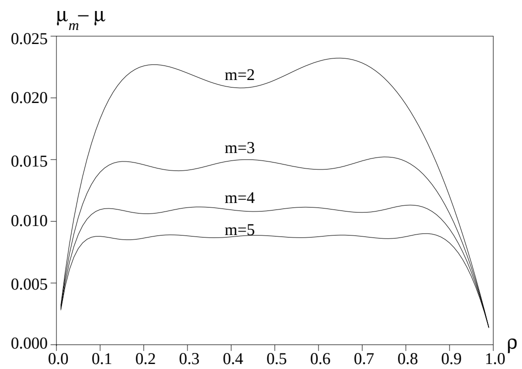

Figure 1: Asymptotic difference between the NFD heuristic and the GR procedure, for $m = 2, \ldots, 5$.

Proof: Controlling $B_{n,m}$ by sums of independent random variables is not as straightforward as for A_n, though a similar truncation technique will be used. We shall first describe the lower bound, then the upper bound.

To get a lower bound, we restrict ourselves to k-groups, with $k < r$: $B_{n,m}$ is certainly larger than the number of remaining bins after rearrangement of those k-groups, and suppression of all other groups. Denote by N_k the number of items of type k in the original list: the distribution of N_r is binomial with parameters n and p_k^*. If there are N_k items of type k, then the number of k-groups is certainly larger than

$$G_k = \lfloor \frac{1}{m}(\lfloor \frac{N_k}{k} \rfloor - 1) \rfloor - 1 . \qquad (15)$$

For all $k \geq 1$, consider independent sequences of i.i.d.r.v.'s $(Z_{m,k}^{(l)})_{l\geq 1}$, all independent from the X_i's, where $Z^{(l)}_{m,k}$ has distribution $\pi_{m,k}$. Define $S_{1,n}$ as the sum

$$S_{1,n} = \sum_{k=1}^{r-1} \sum_{l=1}^{G_k} Z^{(l)}_{m,k} \ .$$

The previous reasoning shows that $S_{1,n}$ is smaller than $B_{n,m}$ in the stochastic ordering sense: for all x,

$$\mathbb{P}[S_{1,n} \leq x] \geq \mathbb{P}[B_{n,m} \leq x] \ .$$

As in the proof of theorem 2.3, we need to control the difference between $n\mu_m$ and $\mathbb{E}[S_{1,n}]$. By Wald's theorem, one has

$$\mathbb{E}[S_{1,n}] = \sum_{k=1}^{r-1} \mathbb{E}[G_k]\, e_{k,m} \ .$$

By definition of G_k, one has

$$\mathbb{E}[G_k] \geq \frac{1}{m}\left(\frac{\mathbb{E}[N_k]}{k} - 2\right) - 2 = \frac{np^*_k}{km} - \frac{2}{m} - 2 \ .$$

Remarking that $e_{k,m} \leq m$ for all k, one gets

$$\begin{aligned} n\mu_m - \mathbb{E}[S_{1,n}] &= \sum_{k=1}^{r-1} e_{k,m}\left(\frac{np^*_k}{km} - \mathbb{E}[G_k]\right) + \sum_{k=r}^{\infty} e_{k,m}\frac{np^*_k}{km} \\ &\leq \sum_{k=1}^{r-1} e_{k,m}\left(\frac{2}{m} + 2\right) + \sum_{k=r}^{\infty} e_{k,m}\frac{np^*_k}{rm} \\ &\leq (r-1)(2+2m) + \frac{nq^*_r}{r} \ . \end{aligned}$$

Under (10), the same choice of $r(n)$ as in the proof of theorem 2.3 ensures that this difference is small compared to $\sqrt{n}$:

$$r(n) = \lfloor \sqrt{c}\, n^{\frac{1}{2+\alpha}} \rfloor + 1 \ . \qquad (16)$$

We also need to check that $Var[S_{1,m}] - n\sigma^2_m = o(n)$. One has:

$$Var[S_{1,n}] = \sum_{k=1}^{r-1} \mathbb{E}[G_k]\, v_{k,m} + Var[G_k]\, e^2_{k,m} + \sum_{k\neq h=1}^{r-1} Cov[G_k, G_h]\, e_{k,m}e_{h,m} \ .$$

Using the definition (15) of G_k, one easily gets:

$$\mathbb{E}[G_k] = \frac{np^*_k}{km} + O(1) \ , \ Var[G_k] = \frac{np^*_k(1-p^*_k)}{k^2m^2} + O(1) \ ,$$

$$Cov[G_k, G_h] = -\frac{np^*_kp^*_h}{khm^2} + O(1) \ .$$

From this one deduces:

$$
\begin{aligned}
Var[S_{1,n}] &= n\sum_{k=1}^{r-1} p_k^* \left(\frac{v_{k,m}}{km} + \frac{e_{k,m}^2}{k^2m^2}\right) - n\left(\sum_{k=1}^{r-1} p_k^* \frac{e_{k,m}}{km}\right)^2 + O(r^2) \\
&= n\,\sigma_m^2 + o(n)\ ,
\end{aligned}
$$

still using expression (16) for $r(n)$.

Let us now turn to the upper bound. The number of remaining bins $B_{n,m}$ certainly increases if one neglects to rearrange non homogeneous groups. It also increases if all items of size $\leq 1/r$ are replaced by items of size $1/r$ in the original list and none of them disappears. Let M_r denote the number of items of type $\geq r$. The upperbound $S_{2,n}$ is the following:

$$
S_{2,n} = S_{1,n} + (M_r/r + rm)\ .
$$

One has for all x:

$$
\mathbb{P}[B_{n,m} \leq x] \geq \mathbb{P}[S_{2,n} \leq x]\ .
$$

As before, one can check that $\mathbb{E}[S_{2,n}] = n\mu_m + o(\sqrt{n})$ and $Var[S_{2,n}] = n\sigma_m^2 + o(n)$.

To finish the proof along the same lines as that of theorem 2.3, we need to check that the law of large numbers and the central limit theorem hold for $S_{1,n}$ and $S_{2,n}$, which are sums of random numbers of r.v.'s. We shall do it for $S_{1,n}$; similar arguments hold for $S_{2,n}$. The law of large numbers is the easy part. By formula (15), G_k increases a.s. to infinity and

$$
\lim_{n\to\infty} \frac{G_k}{n} = \frac{p_k^*}{km} \quad \text{a.s.}
$$

Hence:

$$
\lim_{n\to\infty} \frac{1}{n}\sum_{l=1}^{G_k} Z_{k,m}^{(l)} = \lim_{n\to\infty} \frac{G_k}{n}\frac{1}{G_k}\sum_{l=1}^{G_k} Z_{k,m}^{(l)} = \frac{p_k^* e_{k,m}}{km} \quad \text{a.s.}
$$

Using again the expression (16) for $r(n)$, it follows that

$$
\lim_{n\to\infty} \frac{S_{1,n}}{n} = \mu_m \quad \text{a.s.}
$$

The central limit theorem is not as straightforward. Here are the main steps.

We first check that the vector $(G_k)_{1\leq k\leq r-1}$ is asymptotically normal. Consider the vector $(N_1,\ldots,N_{r-1},M_r)$. Its distribution is multinomial, with parameters n and $(p_1^*,\ldots,p_{r-1}^*,q_r^*)$. From there, and formula (15), it follows that the vector $(G_k)_{1\leq k\leq r-1}$ is asymptotically normal (G_k essentially behaves as $N_k/(km)$). More precisely, define for all $k \geq 1$

$$
\widetilde{p}_k = \frac{p_k^*}{km} \quad \text{and} \quad H_k = \frac{G_k - n\widetilde{p}_k}{\sqrt{n}}\ .
$$

The distribution of the random vector $(H_k)_{1\leq k\leq r-1}$ converges to the multidimensional Gaussian distribution with null expectation ond covariance matrix $C = (c_{k,h})_{1\leq k,h\leq r-1}$, given by:

$$
c_{k,k} = \frac{p_k^*(1-p_k^*)}{k^2m^2} \quad \text{and} \quad c_{k,h} = -\frac{p_k^* p_h^*}{khm^2}\ ,\ \text{for } k \neq h\ .
$$

For $k \geq 1$, consider the partial sum

$$\widetilde{S}_{k,m} = \sum_{l=1}^{G_k} Z_{k,m}^{(l)} .$$

Using the classical technique of characteristic functions, one can show that the distribution of the vector

$$\left(\frac{1}{\sqrt{n}} (\widetilde{S}_{k,m} - n\widetilde{p}_k e_{k,m}) \right)_{1 \leq k \leq r-1}$$

converges to the multidimensional Gaussian distribution with null expectation and covariance matrix

$$D_1 + D_2 C D_2 ,$$

where D_1 and D_2 are the following diagonal matrices.

$$D_1 = Diag((\widetilde{p}_k v_{k,m})_{1 \leq k \leq r-1}) \quad \text{and} \quad D_2 = Diag((e_{k,m})_{1 \leq k \leq r-1}) .$$

Summing coordinates, it follows that $S_{1,n}$ is asymptotically normal, for any fixed r. There remains to let $r = r(n)$ tend to infinity, using (16). The already given estimates on $\mathbb{E}[S_{1,n}]$ and $Var[S_{1,n}]$, yield that

$$\frac{1}{\sqrt{n\sigma_m^2}} (S_{1,n} - n\mu_m)$$

converges in distribution to the standard Gaussian distribution. □

References

[1] B.S. Baker and E.G. Coffman Jr. A tight asymptotic bound for next fit decreasing bin packing. *SIAM Journal on Algebraic and Discrete Methods*, 2:147–152, 1981.

[2] M. Bellalouna. *Problèmes d'optimisation combinatoire probabilistes*. PhD thesis, Ecole Nationale des Ponts et Chaussées, Paris, 1993.

[3] D. Bertsimas. *Probabilistic Combinatorial Optimisation Problems*. PhD thesis, Massachusetts Institute of Technology, Cambridge, Mass., 1988.

[4] D. Bertsimas, P. Jaillet, and A. Odoni. A priori optimization. *Operations Research*, 38:1019–1033, 1990.

[5] E.G. Coffman, Jr., M.R. Garey, and D.S. Johnson. Approximation algorithms for bin packing - a survey. In D. Hochbaum, editor, *Approximation algorithms for NP-Hard problems*, pages 46–93. PWS Publishing, Boston, 1996.

[6] J. Csirik, J.B.G. Frenk, A.M. Frieze, G. Galambos, and A.H.G. Rinnooy Kan. A probabilistic analysis of the Next-Fit decreasing bin packing heuristic. *Oper. Res. Lett.*, 5(5):233–236, 1986.

[7] J. Csirik and E. Mate. The probabilistic behaviour of the NFD bin packing algorithm. *Acta Cyber.*, 7:241–245, 1986.

[8] W. Feller. *An introduction to probability theory and its applications*, volume I. Wiley, London, 3rd edition, 1968.

[9] W. Feller. *An introduction to probability theory and its applications*, volume II. Wiley, London, 2nd edition, 1971.

[10] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-Completeness.* Freeman, San Francisco, 1979.

[11] M. Hofri. *Probabilistic analysis of algorithms.* Springer-Verlag, New York, 1987.

[12] M. Hofri. *Analysis of algorithms – Mathematical methods, computational tools.* Oxford University Press, Oxford, 1995.

[13] M. Hofri and S. Kamhi. A stochastic analysis of the NFD bin packing algorithm. *Journal of Algorithms*, 7:489–509, 1986.

[14] P. Jaillet. A priori solution of a traveling salesman problem in which a random subset of the customers are visited. *Operations Research*, 36:929–936, 1988.

[15] P. Jaillet. Analysis of probabilistic combinatorial optimization problems in Euclidean spaces. *Mathematics of Operation Research*, 18:51–71, 1993.

[16] V.V. Petrov. *Limit theorems of probability theory.* Oxford University Press, Oxford, 1995.

[17] W.T. Rhee. Probabilistic analysis of the next fit decreasing algorithm for bin-packing. *Oper. Res. Lett.*, 6(4):189–191, 1987.

[18] W.T. Rhee. Correction to: Probabilistic analysis of the next fit decreasing algorithm for bin-packing. *Oper. Res. Lett.*, 7:211, 1988.

Monia Bellalouna, Salma Souissi
CRISTAL pôle GRIFT, Ecole Nationale des Sciences de l'Informatique, Tunisia
{monia.bellalouna,salma.souissi}@ensi.rnu.tn
Bernard Ycart
MAP5 CNRS UMR 8145, Université Paris 5, France
ycart@math-info.univ-paris5.fr