71/2007

Raport Badawczy
Research Report

RB/22/2007

On some average case analysis results of the set packing problem

K. Szkatuła

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Pracowni zgłaszający pracę:
prof. dr hab. inż. Krzysztof Kiwiel

Warszawa 2007

On some average case analysis results of the set packing problem

Krzysztof SZKATUŁA
Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warszawa, Poland
E-mail: Krzysztof.Szkatula@ibspan.waw.pl

December 4, 2007

Abstract

The paper deals with the well known set packing problem. It is assumed that some of the problem coefficients are realizations of mutually independent random variables. Average case (i.e. asymptotical probabilistic) properties of selected problem characteristics are investigated for the variety of possible instances of the problem. The important results of the paper are:

- Behavior of the optimal solution values of the set packing problem is presented in the special asymptotic case, where mutual asymptotical relation between m (number of elements of the packed set) and n (number of sets provided) is playing essential role.
- For the considered in the paper random model of the problem there is no feasible solution, but the trivial cases, with probability approaching 1, in the asymptotic case. However probability of reaching feasible solution is reasonably high (i.e. $\geqslant 2/e, 2/e \approx 0.736$); moreover it may be set arbitrary close to 1 (e.g. 0.999), but quality of approximation of the behavior of the optimal solution values may be very unsatisfactory then.

1 Introduction

Let us consider a m element set M and Φ a collection of n subsets M_i, $i = 1, \ldots, n$, of the set M, $\Phi = \{M_1, M_2, \ldots, M_n\}$. *Set packing problem* consists in finding set of disjoint subsets Ψ in $\Phi, \Psi \subseteq \Phi$, where, $M_i, M_k \in \Psi$ if and only if $M_i \cap M_k = \emptyset$, for every i, k, $i \neq k$, $i, k \in \{1, \ldots, n\}$. Set packing problem may be formulated as the binary multiconstraint knapsack problem, see Nemhauser and Wolsey [6]:

$$\begin{aligned} & z_{OPT}(n) = \max \sum_{i=1}^{n} c_i \cdot x_i \\ & \text{subject to} \quad \sum_{i=1}^{n} a_{ji} \cdot x_i \leqslant 1 \\ & \text{where} \quad j = 1, \ldots, m, \quad x_i = 0 \text{ or } 1 \end{aligned} \tag{1}$$

It is assumed that:

$$c_i > 0,\ a_{ji} = 0 \text{ or } 1,\ i = 1, \dots, n,\ j = 1, \dots, m.$$

In fact a_{ji}, $i = 1, \dots, n$, $j = 1, \dots, m$ are defining Φ, set of subsets of M, namely M_i, $i = 1, \dots, n$ in the following way

$$a_{ji} = \begin{cases} 1 & \text{if } j \in M_i \\ 0 & \text{if } j \notin M_i \end{cases},$$

where c_i is the certain value expressing the preference assigned to M_i. Let us observe that definition of the sets M_i, $i = 1, \dots, n$, does not require them to be disjoint. Namely if there exists $j \in \{1, \dots, m\}$, $k \neq l$, $k, l \in \{1, \dots, n\}$, such that $a_{jk} = a_{jl} = 1$, then $j \in M$ belongs to both M_k and M_l, i.e. $M_k \cap M_l \neq \emptyset$. Choice of x_i, fulfilling the constraints imposed in (1) is defining the packing of the set M into disjoint subsets M_i, $M_i \in \Psi$,where $M_i \cap M_k = \emptyset$ $i \neq k$, $i, k \in \{1, \dots, n\}$,for every M_i, $M_k \in \Psi$. Namely in (1)

$$\forall\ k,\ k, \in \{1, \dots, n\},\ M_k \in \Psi,\ \text{if and only if } \exists\ j \in M_k : a_{jk} \cdot x_k = 1.$$

Each of the constraints $\sum_{i=1}^{n} a_{ji} \cdot x_i \leqslant 1$, $j = 1, \dots, m$ is guaranteeing that each of the items j of the set M is assigned to maximum one of the subsets M_i, $M_i \in \Psi$. Optimisation criteria in (1) is securing the choice of best possible packing according to preferences expressed by c_i, $i = 1, \dots, n$. If $c_i = c$, $i = 1, \dots, n$, c - constant (e.g. $c = 1$), then optimisation problem seeks for the maximum amount of subsets M_i to pack set M, known as *Maximum Set Packing Problem.*

Set packing problem (1) is well known to be $\mathcal{NP}$ hard combinatorial optimisation problem, see Garey and Johnson [2]. Moreover Set Packing Problem is one of the 21 first Karp's $\mathcal{NP}$ complete problems, see [3]. There are also two closely related combinatorial problems, namely *set covering problem* and *set partitioning problem* (also known as exact covering),where in both of them one is looking for the subsets M_{kj}, $j = 1, ..., r$, of the collection Φ of n subsets of M_i, $i = 1, \dots, n$, where demand $\bigcup_{j=1}^{r} M_{kj} = M$ holds, moreover in the set partitioning problem there is additional demand, namely that all M_{kj} are pairwise disjoint, i.e. $M_{k_j} \cap M_{k_l} = \emptyset$, for every k_j, k_l, $k_j \neq k_l$, $j, l \in \{1, \dots, r\}$. Both problems may be also formulated as special cases of the binary multiconstraint knapsack problem, see Nemhauser and Wolsey [6]

Although set packing problem may be formulated as the binary multiconstraint knapsack problem, it is rather special case of it, see Martello and Toth [4]. Its peculiarity consists in 2 facts:

- All the constraints left hand sides coefficients are equal either to 1 or to 0:

$$a_{ji} = 0 \text{ or } 1,\ i = 1, \dots, n,\ j = 1, \dots, m.$$

- All of the constraints right hand sides coefficients are equal to 1.

In the general formulation of the binary multiconstraint knapsack problem it is only required that all of the knapsack problem coefficients, i.e. goal function, constraints left and right hand sides, are non-negative or, in order to avoid unclear interpretations, strictly positive. The latter especially applies to goal function and constraints right hand sides coefficients.

2 Definitions

The following definitions are necessary for the further presentation:

Definition 1 *We denote $V_n \approx Y_n$, where $n \to \infty$, if*

$$Y_n \cdot (1 - o(1)) \leqslant V_n \leqslant Y_n \cdot (1 + o(1))$$

when V_n, Y_n are sequences of numbers, or

$$\lim_{n\to\infty} P\{Y_n \cdot (1 - o(1)) \leqslant V_n \leqslant Y_n \cdot (1 + o(1))\} = 1$$

when V_n is a sequence of random variables and Y_n is a sequence of numbers or random variables, where $\lim_{n\to\infty} o(1) = 0$ as it is usually presumed.

Definition 2 *We denote $V_n \preceq Y_n (V_n \succeq W_n)$ if*

$$V_n \leqslant (1 + o(1)) \cdot Y_n \ (V_n \geqslant (1 - o(1)) \cdot W_n)$$

when V_n, Y_n (W_n) are sequences of numbers, or

$$\lim_{n\to\infty} P\{V_n \leqslant (1 + o(1)) \cdot Y_n\} = 1 \ (\lim_{n\to\infty} P\{V_n \geqslant (1 - o(1)) \cdot W_n\} = 1)$$

when V_n is a sequence of random variables and Y_n (W_n) is a sequence of numbers or random variables, where $\lim_{n\to\infty} o(1) = 0$.

Definition 3 *We denote $V_n \approxeq Y_n$ if there exist constants $c'' \geqslant c' > 0$ such that*

$$c' \cdot Y_n \preceq V_n \preceq c'' \cdot Y_n$$

where Y_n, V_n are sequences of numbers or random variables.

The following random model of (1) will be considered in the paper:

- m, n, $0 < n \leqslant m!$, are arbitrary positive integers and moreover $n \to \infty$.
- c_i, a_{ji}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, are realizations of mutually independent random variables and moreover c_i, are uniformly distributed over $(0, 1]$ and $P\{a_{ji} = 1\} = p$, where $0 < p \leq 1$.

Let us observe that asymptotical relations $0 < n \leqslant m!$ and $n \to \infty$ requires that also $m \to \infty$. As the matter of fact mutual asymtotical relation of the values of m and n may vary between 2 extreme cases $n/m \approx 0$ or $n \approx m!$ as $n \to \infty$

Under the assumptions made about c_i, a_{ji}, and taking into account (1) the following always hold

$$0 \leqslant z_{OPT}(n) \leqslant \sum_{i=1}^{n} c_i \leqslant n, \tag{2}$$

Moreover, from the strong law of large numbers it follows that

$$\sum_{i=1}^{n} c_i \approx E(c_1) \cdot n = n/2, \ \sum_{i=1}^{n} a_{ji} \approx p \cdot n. \tag{3}$$

Therefore, it is justified to enhance formulas (2) and (3) in the following way:

$$0 \leqslant z_{OPT}(n) \preceq n/2, \ \sum_{i=1}^{n} a_{ji} \preceq 1, \ \text{if } p < \frac{1}{n} \text{ or } \sum_{i=1}^{n} a_{ji} \succeq 1 \text{ when } p > \frac{1}{n}. \quad (4)$$

Formula (4) shows that random model of set packing problem (1) is complete in the sense that nearly all possible instances of the problem are considered.

The growth of $z_{OPT}(n)$ - value of the optimal solution of the problem (1) may be influenced by the problem coefficients, namely:

$$n, \ m, \ c_i, \ a_{ji}, \ \text{where } i = 1, \dots, n, \ j = 1, \dots, m.$$

We have assumed that c_i, a_{ji} are realizations of the random variables and therefore their impact on the $z_{OPT}(n)$ growth is in this case indirect. Moreover, we have also assumed that m, n are arbitrary positive integers and $n \to \infty$.

The main aim of the present paper is to perform probabilistic analysis of the considered class of random set packing problems in the asymptotical case, i.e. when $n \to \infty$. Probabilistic analysis has 2 strategic goals, namely:

- To examine existence of the feasible solutions.
- To investigate asymptotic behaviour of $z_{OPT}(n)$.

3 Lagrange and dual estimations

When the knapsack problem, with one or many constraints, is considered then Lagrange function and the corresponding dual problems, see Averbakh [1], Meanti, Rinnooy Kan, Stougie and Vercellis [5], Szkatuła [7] and [8] are very useful tools to perform various kind of analyses of the original problem. In the case of set packing problem Lagrange function of the problem (1) may be formulated as follows:

$$\begin{aligned} L_n(x) &= \sum_{i=1}^{n} c_i \cdot x_i + \sum_{j=1}^{m} \lambda_j \cdot \left(1 - \sum_{i=1}^{n} a_{ji} \cdot x_i\right) = \\ &= \sum_{j=1}^{m} \lambda_j + \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{m} \lambda_j \cdot a_{ji}\right) \cdot x_i \end{aligned}$$

where $x = [x_1, \dots, x_n]$ and $\Lambda = [\lambda_1, \dots, \lambda_m]$ - vector of Lagrange multipliers. Moreover, let for every Λ, $\lambda_j \geq 0$, $j = 1, \dots, m$:

$$\phi_n(\Lambda) = \max_{x \in \{0,1\}^n} L_n(x, \Lambda) = \max_{x \in \{0,1\}^n} \left\{ \sum_{j=1}^{m} \lambda_j + \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{m} \lambda_j a_{ji}\right) x_i \right\}.$$

Taking the following notation:

$$x_i(\Lambda) = \begin{cases} 1 & \text{if } c_i - \sum\limits_{j=1}^{m} \lambda_j \cdot a_{ji} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

$$c_i(\Lambda) = \begin{cases} c_i & \text{if } c_i - \sum\limits_{j=1}^{m} \lambda_j \cdot a_{ji} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$a_{ji}(\Lambda) = \begin{cases} a_{ji} & \text{if } c_i - \sum\limits_{j=1}^{m} \lambda_j \cdot a_{ji} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

we have for every Λ, $\lambda_j \geq 0$, $j = 1, \ldots, m$:

$$\begin{aligned} \phi_n(\Lambda) &= \sum_{j=1}^{m} \lambda_j + \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{m} \lambda_j \cdot a_{ji} \right) \cdot x_i(\Lambda) = \\ &= \sum_{j=1}^{m} \lambda_j + \sum_{i=1}^{n} \left(c_i(\Lambda) - \sum_{j=1}^{m} \lambda_j \cdot a_{ji}(\Lambda) \right) \end{aligned}$$

Obviously for $i = 1, \ldots, n,\ j = 1, \ldots, m,$

$$c_i(\Lambda) = c_i \cdot x_i(\Lambda), \quad a_{ji}(\Lambda) = a_{ji} \cdot x_i(\Lambda).$$

Dual problem to set packing problem (1) maybe formulated as follows:

$$\Phi_n^* = \min_{\Lambda \geq 0} \phi_n(\Lambda). \tag{6}$$

For every $\Lambda \geq 0$ the following holds:

$$z_{OPT}(n) \leq \Phi_n^* \leq \phi_n(\Lambda) = z_n(\Lambda) + \sum_{j=1}^{m} \lambda_j (1 - s_j(\Lambda)). \tag{7}$$

Let us denote:

$$\begin{aligned} z_n(\Lambda) &= \sum_{i=1}^{n} c_i \cdot x_i(\Lambda) = \sum_{i=1}^{n} c_i(\Lambda), s_j(\Lambda) = \sum_{i=1}^{n} a_{ji} \cdot x_i(\Lambda) = \sum_{i=1}^{n} a_{ji}(\Lambda), \\ S_{nm}(\Lambda) &= \sum_{j=1}^{m} \lambda_j \cdot s_j(\Lambda),\ \tilde{\Lambda}(m) = \sum_{j=1}^{m} \lambda_j. \end{aligned}$$

By definition of $c_i(\Lambda)$ and $a_{ji}(\Lambda)$, see also (5), we have:

$$c_i(\Lambda) \geq \sum_{j=1}^{m} \lambda_j \cdot a_{ji}(\Lambda),\ i = 1, \ldots, n,$$

and therefore

$$z_n(\Lambda) \geq S_{nm}(\Lambda). \tag{8}$$

For certain Λ, $x_i(\Lambda)$ given by (5) may provide feasible solution of (1), i.e.:

$$s_j(\Lambda) \leq 1 \quad \text{for every} \quad j = 1, \dots, m. \tag{9}$$

Then:

$$z_n(\Lambda) \leq z_{OPT}(n) \leq \Phi_n^* \leq \phi_n(\Lambda) = z_n(\Lambda) + \tilde{\Lambda}(m) - S_{nm}(\Lambda). \tag{10}$$

If (9) holds, then the below inequality also holds:

$$\tilde{\Lambda}(m) - S_{nm}(\Lambda) \geq 0.$$

From (8) we get:

$$\frac{\phi_n(\Lambda)}{z_n(\Lambda)} = \frac{z_n(\Lambda)}{z_n(\Lambda)} + \frac{\tilde{\Lambda}(m) - S_{nm}(\Lambda)}{z_n(\Lambda)} \leq 1 + \frac{\tilde{\Lambda}(m) - S_{nm}(\Lambda)}{S_{nm}(\Lambda)}.$$

Therefore if (9) holds, then the following inequality also holds:

$$1 \leq \frac{z_{OPT}(n)}{z_n(\Lambda)} \leq \frac{\Phi_n^*}{z_n(\Lambda)} \leq \frac{\phi_n(\Lambda)}{z_n(\Lambda)} \leq \frac{\tilde{\Lambda}(m)}{S_{nm}(\Lambda)}. \tag{11}$$

Formula (11) shows, that if there exits such a set of Lagrange multipliers $\Lambda(n)$ which is fulfilling the formula (9) and if the formula below holds:

$$\lim_{n\to\infty} \frac{\tilde{\Lambda}(m)}{S_{nm}(\Lambda(n))} = 1 \tag{12}$$

then $x_i(\Lambda(n))$, $i = 1, \dots, n$, given by (5), is the asymtotically sub-optimal solution of the set packing problem (1). Moreover the value of $z_n(\Lambda(n))$ is an asymptotical approximation of the optimal solution value of the set packing problem i.e. $z_{OPT}(n)$.

4 Probabilistic analysis

In the present section of the paper some probabilistic properties of the set packing problem (1) will be investigated. Let us observe that due to the assumptions made the following holds, for $i = 1, \dots, n$, $j = 1, \dots, m$:

$$P\{a_{ji} = 1\} = p,\ P\{a_{ji} = 0\} = 1 - p,\ P\{a_{ji}(\Lambda) = 1\} = 1 - P\{a_{ji}(\Lambda) = 0\},$$

$$P(c_i < x) = \begin{cases} 0 & \text{when } x \leqslant 0 \\ x & \text{when } 0 < x \leqslant 1 \\ 1 & \text{when } x \geqslant 1 \end{cases}. \tag{13}$$

Moreover for the random variable $\sum_{k=1,k\neq j}^{m} a_{ji}$, due to the binomial distribution, the following holds for every r - integer, $0 \leqslant r \leqslant m-1$:

$$P\left\{\sum_{k=1,k\neq j}^{m} a_{ki} = r\right\} = \binom{m-1}{r} \cdot p^r \cdot (1-p)^{m-r-1}. \tag{14}$$

Let us also assume that

$$\Lambda = \{\lambda, \cdots, \lambda\}, \text{ i.e. } \lambda_j = \lambda,\ \lambda \geqslant 0,\ j = 1, \cdots, m.$$

Lemma 1 *If a_{ji} are realizations of mutually independent random variables where $P\{a_{ji} = 1\} = p$, $0 < p \leq 1$, then*

$$P\{a_{ji}(\Lambda) = 1\} = p - p\sum_{r=0}^{m-1} \binom{m-1}{r} \cdot p^r \cdot (1-p)^{m-r-1} \min\{1, \lambda(r+1)\}.$$

If, moreover, $\lambda \leqslant 1/m$ then:

$$P\{a_{ji}(\Lambda) = 1\} = p \cdot (1 - \lambda \cdot (m \cdot p + 1 - p)).$$

Proof. From (5), (13) and (14) and taking into account that random variable $\sum_{k=1,k\neq j}^{m} a_{ji}$ may take any integer value r from the range $[0, m-1]$ with the probability given in (14) it follows that:

$$\begin{aligned} P\{a_{ji}(\Lambda) = 0\} &= P\left\{a_{ji} = 0 \cup a_{ji} = 1 \cap c_i < \lambda \cdot \left(\sum_{k=1,k\neq j}^{m} a_{ji} + 1\right)\right\} = \\ &= 1 - p + p \cdot P\left\{c_i < \lambda \cdot \left(\sum_{k=1,k\neq j}^{m} a_{ji} + 1\right)\right\} = \\ &= 1 - p + p\sum_{r=0}^{m-1} \binom{m-1}{r} \cdot p^r \cdot (1-p)^{m-r-1} \min\{1, \lambda(r+1)\}. \end{aligned}$$

Due to the (13) the first formula of the Lemma is proven. Because

$$\binom{m-1}{r} = \frac{(m-1)!}{r! \cdot (m-1-r)!},$$

then when $\lambda \leqslant 1/m$ the following holds

$$P\{a_{ji}(\Lambda) = 0\} = 1 - p + \lambda \sum_{r=0}^{m-1} \frac{(m-1)! \cdot (r+1)}{r! \cdot (m-1-r)!} \cdot p^{r+1} \cdot (1-p)^{m-r-1} \tag{15}$$

Let us observe that for every integers l, m, $l, > 1$, $m \geqslant 2$, and $0 \leqslant p \leqslant 1$ the following hold

$$\begin{aligned} \sum_{k=0}^{l} \binom{l}{k} \cdot p^k \cdot (1-p)^{l-k} &= (p + 1 - p)^l = 1 \\ r + 1 &= m - (m - 1 - r). \end{aligned}$$

Using the above mentioned formulas (15) may be rewritten as:

$$\begin{aligned} P\{a_{ji}(\Lambda)=0\} &= 1-p+\lambda\cdot p\left(\sum_{r=0}^{m-1}\frac{(m-1)!\cdot m}{r!\cdot(m-1-r)!}\cdot p^r\cdot(1-p)^{m-1-r}-\right.\\ &\left.-\sum_{r=0}^{m-1}\frac{(m-1)!\cdot(m-1-r)}{r!\cdot(m-1-r)!}\cdot p^r\cdot(1-p)^{m-1-r}\right)=\\ &= 1-p+\lambda\cdot p\left(m\sum_{r=0}^{m-1}\binom{m-1}{r}\cdot p^r\cdot(1-p)^{m-1-r}-\right.\\ &\left.-p\cdot(m-1)\cdot(1-p)\sum_{r=0}^{m-2}\binom{m-2}{r}\cdot p^r\cdot(1-p)^{m-2-r}\right)=\\ &= 1-p+\lambda\cdot p\cdot(m-(m-1)\cdot(1-p))=\\ &= 1-p+\lambda\cdot p\cdot(m\cdot p+1-p). \end{aligned}$$

Finally above formulas can be summarized as:

$$P\{a_{ji}(\Lambda)=0\}=1-p+\lambda\cdot p\cdot(m\cdot p+1-p). \qquad (16)$$

Due to the formulas (13) and (16) we have

$$\begin{aligned} P\{a_{ji}(\Lambda)=1\} &= 1-P\{a_{ji}(\Lambda)=0\}=\\ &= p-\lambda\cdot p\cdot(m\cdot p+1-p)=p\cdot(1-\lambda\cdot(m\cdot p+1-p)). \end{aligned}$$

■

As the direct consequence of the above formulas we have

$$E(a_{ji}(\Lambda))=1\cdot P\{a_{ji}(\Lambda)=1\}+0\cdot P\{a_{ji}(\Lambda)=0\}=P\{a_{ji}(\Lambda)=1\}. \qquad (17)$$

Now instead of Λ we will consider $\Lambda(n)$. It does mean that for every value of integer n, we may consider different vector $\Lambda(n)=\{\lambda(n),\cdots,\lambda(n)\}$, $\lambda(n)\geqslant 0$. For every j, $j=1,\cdots,m$, we have:

$$\begin{aligned} E(s_j(\Lambda(n))) &= \sum_{i=1}^{n}E(a_{ji}(\Lambda(n)))=n\cdot P\{a_{ji}(\Lambda(n))=1\}= \qquad (18)\\ &= n\cdot p(1-\lambda(n)\cdot(m\cdot p+1-p)). \end{aligned}$$

Lemma 2 *For every α, $\alpha>0$ there exists m' n', m', n' $>,1$ such that for every $m\geqslant m'$ and $n\geqslant n'$, the following choice of $\lambda(n)$:*

$$\lambda(n)=\frac{1-\alpha/(n\cdot p)}{m\cdot p+1-p} \text{ is solving the equations } E(s_j(\Lambda(n)))=\alpha.$$

Corollary 1 *If $E(s_j(\Lambda(n)))=\alpha$, then $P\{a_{ji}(\Lambda(n))=1\}=\alpha/n$.*

Proof. Proof of Lemma and Corollary follows immediately from formulas (17) and (18) and following fact that for all $m\geqslant m'$ and $n\geqslant n'$:

$$\lambda(n)\leqslant\frac{1}{m}.$$

■

Solution of the set packing problem (1) given by formula (5) is feasible if and only if the formula (9) holds.

Theorem 1 *For every α, $\alpha > 0$ there exists m', n', m', $n' > 1$, such that for $\Lambda(n)$, providing $E(s_j(\Lambda(n))) = \alpha$, the following hold*

$$P\{s_j(\Lambda(n)) \leqslant 1\} = \left(1 - \frac{\alpha}{n}\right)^{n-1} \cdot (1 + \alpha - \frac{\alpha}{n})$$

Moreover for every fixed value of $\alpha, \alpha > 0$, we have

$$\lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\} = \frac{1+\alpha}{e^{\alpha}}$$

Proof. As it was already mentioned solution of problem (1) given by formula (5) is feasible if and only if formula (9) holds i.e. $s_j(\Lambda(n)) = 0$ or $s_j(\Lambda(n)) = 1$. For every $\Lambda(n)$, random variable $s_j(\Lambda(n)) = \sum_{i=1}^{n} a_{ji}(\Lambda(n))$ may take any integer value r from the range $[0, n]$ with the probability given by the following formula:

$$P\left\{\sum_{i=1}^{n} a_{ji}(\Lambda(n)) = r\right\} = \binom{n}{r} \cdot \tilde{p}^{r} \cdot (1-\tilde{p})^{n-r}, \text{ where } \tilde{p} = P\{a_{ji}(\Lambda(n)) = 1\}.$$

From the above formula and Corollary 1 it follows that

$$\begin{aligned} P\{s_j(\Lambda(n)) \leqslant 1\} &= P\left\{\sum_{i=1}^{n} a_{ji}(\Lambda(n)) = 0 \cup \sum_{i=1}^{n} a_{ji}(\Lambda(n)) = 1\right\} = \\ &= \left(1-\frac{\alpha}{n}\right)^{n} + \alpha\left(1-\frac{\alpha}{n}\right)^{n-1} = \left(1-\frac{\alpha}{n}\right)^{n-1} \cdot (1+\alpha-\frac{\alpha}{n}) \end{aligned} \qquad (19)$$

The proof is finished by observing that $\lim_{n\to\infty} (1 - \frac{\alpha}{n})^{n-1} = e^{-\alpha}$ and $\lim_{n\to\infty} \frac{\alpha}{n} = 0$ ■

Corollary 2 $P\{s_j(\Lambda(n)) \leqslant 1\} = 1$ *if and only if $n = 1$. When $\alpha \to 0$ as $n \to \infty$ then*

$$\lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\} = 1.$$

However if α, $\alpha > 0$, is a constant then:

$$\lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\} < 1 \qquad (20)$$

Proof. Formula (20) follows immediately from the Theorem 1. ■

The above Theorem 1 and Corollary 2 to it have interesting interpretation, which may be observed on few examples presented below:

Example 1

$$\begin{aligned} \textit{When } \alpha = 0.01 \textit{ then } \lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\} &= 0.999 \\ \textit{When } \alpha = 0.1 \textit{ then } \lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\} &= 0.995 \\ \textit{When } \alpha = 0.5 \textit{ then } \lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\} &= 0.9098 \\ \textit{When } \alpha = 1 \textit{ then } \lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\} &= \frac{2}{e} \approx 0.736 \end{aligned}$$

Interpretation of the above examples is following. The closer the value of α is to 1, i.e. set packing problem (1) right-hand-side values the better approximation of the optimal solution values may be provided, however with less satisfactory value of the $\lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\}$. However, for any value α, $0 < \alpha \leqslant 1$, $\lim_{n\to\infty} P\{s_j(\Lambda(n)) \leqslant 1\} \succeq 2/e$, where $2/e \approx 0.736$. Due approximations of the optimal solution values are provided in the next section.

5 Behavior of the optimal solution values

In order to analyse the behaviour of the optimal solution value of the set packing problem (1) one may need to exploit the probabilistic properties of the random variables $c_i(\Lambda(n))$, $i = 1, \cdots, n$. The construction of the random variables $c_i(\Lambda(n))$ is defined by formulas (5) and (13) respectively. Distribution functions of the random variables $c_i(\Lambda(n))$, $i = 1, \cdots, n$ are given by the following formulas, where $0 < x \leq 1$:

$$\begin{aligned} P\{c_i(\Lambda(n)) < x\} &= P\{c_i < x \cup c_i \geq x \cap c_i \leq \Lambda(n) \cdot \sum_{j=1}^{m} a_{ji}\} = \quad (21) \\ &= x + P\{x \leq c_i \leq \Lambda(n) \cdot \sum_{j=1}^{m} a_{ji}\}. \end{aligned}$$

Let us observe that $P\{x \leq c_i \leq \Lambda(n) \cdot \sum_{i=1}^{n} a_{ji}\}$ is by definition equal to zero if $c_i < x$ or $c_i > \Lambda(n) \cdot \sum_{i=1}^{n} a_{ji}$. Therefore (21) may be rewritten as

$$\begin{aligned} P\{c_i(\Lambda(n)) < x\} &= x + \sum_{r=1}^{m} P\{x \leq c_i \leq \Lambda(n) \cdot r \cap \sum_{j=1}^{m} a_{ji} = r\} = \quad (22) \\ &= x + \sum_{r=1}^{m} (r\Lambda(n) - x)_+ P\{\sum_{j=1}^{m} a_{ji} = r\}. \quad (23) \end{aligned}$$

The above formula may enable us to calculate the mean value of the random variables $c_i(\Lambda(n))$, $i = 1, \cdots, n$. Namely:

$$\begin{aligned} E(c_i(\Lambda(n))) &= \int_0^1 x \cdot d(P\{c_i(\Lambda(n)) < x\}) = \quad (24) \\ &= \frac{1}{2} + \int_0^{\Lambda(n)\cdot m} x \cdot \left(\sum_{r=1}^{m} (r\Lambda(n) - x)'_+ \cdot P\{\sum_{j=1}^{m} a_{ji} = r\} \right) = \\ &= \frac{1}{2} + \sum_{k=1}^{m} \int_{\Lambda(n)\cdot(k-1)}^{\Lambda(n)\cdot k} x \left(\sum_{r=k}^{m} (r\Lambda(n) - x)'_+ \cdot P\{\sum_{j=1}^{m} a_{ji} = r\} \right) dx = \\ &= \frac{1}{2} - \sum_{k=1}^{m} \int_{\Lambda(n)\cdot(k-1)}^{\Lambda(n)\cdot k} x \cdot P\{\sum_{j=1}^{m} a_{ji} = r\} dx \end{aligned}$$

Let us observe that, similarly to the formula (14), the random variable $\sum_{k=1}^{m} a_{ji}$, due to its binomial distribution, has the following distribution function for every

r - integer, $0 \leqslant r \leqslant m$:

$$P\left\{\sum_{k=1}^{m} a_{ki} = r\right\} = \binom{m}{r} \cdot p^r \cdot (1-p)^{m-r} \text{ and moreover } \left(\sum_{k=1}^{r}(2k-1)\right) = r^2.$$

Therefore the formula (24) could be further simplified as follows:

$$\begin{aligned} E(c_i(\Lambda(n))) &= \frac{1}{2} - \sum_{k=1}^{m}\left(\int\limits_{\Lambda(n)\cdot(k-1)}^{\Lambda(n)\cdot k} x dx\right) \cdot \left(\sum_{r=k}^{m}\binom{m}{r} \cdot p^r \cdot (1-p)^{m-r}\right) = \\ &= \frac{1}{2} - \frac{(\Lambda(n))^2}{2}\sum_{k=1}^{m}(2k-1) \cdot \left(\sum_{r=k}^{m}\binom{m}{r} \cdot p^r \cdot (1-p)^{m-r}\right) = \\ &= \frac{1}{2} - \frac{(\Lambda(n))^2}{2}\sum_{r=1}^{m}\left(\sum_{k=1}^{r}(2k-1)\right) \cdot \left(\binom{m}{r} \cdot p^r \cdot (1-p)^{m-r}\right) = \\ &= \frac{1}{2} - \frac{(\Lambda(n))^2}{2}\sum_{r=1}^{m} r^2 \cdot \left(\binom{m}{r} \cdot p^r \cdot (1-p)^{m-r}\right). \end{aligned}$$

Let us observe that the following formula holds for $0 < p \leq 1$ and $m = 1, 2, \ldots$

$$\sum_{r=1}^{m} r^2 \cdot \left(\binom{m}{r} \cdot p^r \cdot (1-p)^{m-r}\right) = m \cdot p \cdot (1 + p \cdot (m-1))$$

From Lemma 2 (where $E(s_j(\Lambda(n))) = \alpha$, and $\lambda(n) = \frac{1-\alpha/(n\cdot p)}{m\cdot p+1-p}$) and due to the formula (7) we will therefore receive

$$\begin{aligned} E(z_n(\Lambda)) &= \frac{n}{2}\left(1 - \left(\frac{1-\alpha/(n\cdot p)}{m\cdot p+1-p}\right)^2 \cdot m \cdot p \cdot (m \cdot p + 1 - p)\right) = \\ &= \frac{n}{2}\left(1 - \frac{m\cdot p \cdot (1-\frac{\alpha}{n\cdot p})^2}{m\cdot p+1-p}\right) = \frac{n}{2}\left(1 - \frac{(1-\frac{\alpha}{n\cdot p})^2}{1+(1-p)/(m\cdot p)}\right). \end{aligned}$$

If (9) holds then due to the formulas (10) and (11), where $\tilde{\Lambda}(m) = \sum_{j=1}^{m} \lambda_j(n) = m \cdot \lambda(n)$, $E(S_{nm}(\Lambda(n))) = \alpha \cdot m \cdot \lambda(n)$, one may receive much stronger results for $0 < \alpha \leqslant 1$, namely:

$$1 \leqslant E\left(\frac{z_{OPT}(n)}{z_n(\Lambda(n))}\right) \leqslant \frac{1}{\alpha}, \text{ where } E\left(\frac{\tilde{\Lambda}(m,n)}{S_{nm}(\Lambda(n))}\right) = \frac{1}{\alpha} \text{ and} \tag{25}$$

$$E(z_n(\Lambda(n))) = \frac{n}{2}\left(1 - \frac{(1-\alpha/(n\cdot p))^2}{1+(1-p)/(m\cdot p)}\right). \tag{26}$$

Formulas (25) and (26) may provide us with some estimations of the set packing problem (1) optimal solution values $z_{OPT}(n)$ growth, when $n \to \infty$.

Corresponding to Example 1 estimations of the $E\left(\frac{z_{OPT}(n)}{z_n(\Lambda(n))}\right)$ for the different values of α are provided in the Example below, where appropriate value of $E(z_n(\Lambda(n)))$ is given in the formula (26):

Example 2

$$\begin{aligned}
\textit{When } \alpha &= 0.01 \textit{ then } 1 \leqslant E\left(\frac{z_{OPT}(n)}{z_n(\Lambda(n))}\right) \leqslant 100 \textit{ with approx. probablity } 0.999 \\
\textit{When } \alpha &= 0.1 \textit{ then } 1 \leqslant E\left(\frac{z_{OPT}(n)}{z_n(\Lambda(n))}\right) \leqslant 10 \textit{ with approx. probablity } 0.995 \\
\textit{When } \alpha &= 0.5 \textit{ then } 1 \leqslant E\left(\frac{z_{OPT}(n)}{z_n(\Lambda(n))}\right) \leqslant 2 \textit{ with approx. probablity } 0.9098 \\
\textit{When } \alpha &= 1 \textit{ then } E\left(\frac{z_{OPT}(n)}{z_n(\Lambda(n))}\right) = 1 \textit{ with approx. probablity } \frac{2}{e} \approx 0.736.
\end{aligned}$$

Since $n \leqslant m!$ and moreover $n \to \infty$ then obviously also $m \to \infty$. According to formula (26) asymptotic growth of the $E(z_n(\Lambda(n)))$ may be influenced by both n and m. Let us consider the following mutual asymptotic dependence of the both parameters:

$$n = \beta \cdot m^{\gamma}, \text{ where } \beta \text{ is constants, } 0 < \gamma \leqslant m, \ \beta > 0. \tag{27}$$

If $0 < \gamma \leqslant m$ then condition $n \leqslant m!$ is always fulfilled asymptotically since, due to the Stirling's formula, for every constant $\beta > 0$ there exist constant $m' \geqslant 1$ such that for all $m \geqslant m'$ the inequality $n \leqslant m!$ holds. .

Under the above assumption the following Lemma holds

Lemma 3 *If asymptotical dependence (27) holds then:*

$$E(z_n(\Lambda(n))) \approx \frac{2 \cdot \alpha + \beta \cdot (1-p) \cdot m^{\gamma-1}}{2 \cdot p} \textit{ when } n \to \infty \tag{28}$$

Proof. When (27) holds then (26) may be reformulated as follows:

$$E(z_n(\Lambda(n))) = \frac{2m \cdot \alpha \cdot \beta \cdot p + m^{\gamma} \cdot \beta^2 \cdot p \cdot (1-p) - \alpha^2 \cdot m^{-\gamma+1}}{2\beta \cdot p \cdot (m \cdot p + 1 - p)}$$

Taking into account previously made assumptions on α, β, γ and p proof of the formula (28) is straightforward. ■

Corollary 3 *Depending on the value of γ, $0 < \gamma \leqslant m$, the following cases of the asymtotical behaviour of $E(z_n(\Lambda(n)))$ may be distinguished:*

$$\lim_{m \to \infty} E(z_n(\Lambda(n))) = \begin{cases} \dfrac{\alpha}{p} & \textit{when } 0 < \gamma < 1 \\ \dfrac{2\alpha + \beta \cdot (1-p)}{2p} & \textit{when } \gamma = 1 \\ \infty & \textit{when } \gamma > 1 \end{cases} \tag{29}$$

Due to the formulas (11) and (25) $E(z_n(\Lambda(n)))$ is reasonable asymptotic approximation of the optimal solution of the set packing problem (1) i.e. $E\left(z_{OPT}(n)\right)$. The above Lemma and Corollary, especially formulas (28) and (29), provides interesting insight into asymptotical behavior of the value of $E(z_n(\Lambda(n)))$. Namely:

$$\text{When } n = o(m) \text{ then } \lim_{m\to\infty} E(z_n(\Lambda(n))) = \frac{\alpha}{p}.$$

It does mean that in this case values of β and γ are neglectable so is the mutual asymptotic dependence of both n and m.

$$\text{When } n \approxeq m \text{ then } E(z_n(\Lambda(n))) \approx \frac{2\alpha + \beta\cdot(1-p)}{2p}.$$

In this case level of proximity of n and m is substantial and is expressed by value β.

$$\text{When } m = o(n) \text{ then } E(z_n(\Lambda(n))) \approx \frac{\beta\cdot(1-p)}{2\cdot p}\cdot m^{\gamma-1}$$

In the latter case dependence on α is neglectable, β and p are defining constant multiplier.

In 2 first cases, where $\gamma \leqslant 1$, there is no asymptotical influence of the value of m (and therefore of n either) on the asymptotical value of $E(z_n(\Lambda(n)))$. However in the case when $\gamma > 1$, there is very strong dependence from both m and γ.

On the other hand parameters α, and p have substantial influence on the asymptotical behavior of $E(z_n(\Lambda(n)))$, when $\gamma \leqslant 1$. Namely the bigger is value of α, $\alpha > 0$, and/or smaller is value of p, $0 < p \leqslant 1$, the bigger is value of $E(z_n(\Lambda(n)))$. Consequence of the above statement is following

- The bigger is value of α the less probability of feasibility of the corresponding solution of the set packing problem (1) is, see Theorem 1.
- The smaller the value of p is the sparser the initial subsets M_i, $i = 1, \cdots, n$, of the original set M may be.

6 Concluding remarks

In the present paper some results describing probabilities properties of the set packing problem (1) are summarized.

In the paper distribution functions of the various random variables representing important problems characteristics are presented. Moreover some results concerning the feasibility of the received solutions and estimations of the set packing problem (1) optimal solution values $z_{OPT}(n)$ growth, when $n \to \infty$ are provided.

Examples 1 and 2 shows that the higher is accuracy of approximation of the optimal solution value the lower is probability of the feasibility of corresponding solution. For example when $\alpha = 0.5$ the quality of approximation is pretty tolerable, with relatively high probability of the feasibility of the solution. Moreover when $\alpha = 1$ the quality of approximation is very good with reasonable probability of the feasibility of the solution, approximately equal to 0.736. Lemma 3 shows possible asymptotical behavior of the optimal solution values when there is certain mutual asymptotic dependence of the parameters n and m.

Some of the important avenues for the future research is convergence of the approximate solutions to the optimal solution and possibility of investigating realistic approximations of their values. The considered model of the general *Set Packing Problem* seems to be most appropriate for the case of *Maximum Set Packing Problem*, which is especially suitable, due to optimisation criteria introduced.

References

[1] I. Averbakh. Probabilistic properties of the dual structure of the multidimensional knapsack problem and fast statistically effcient algorithms. *Mathematical Programming*, 65:311–330, 1994.

[2] M. Garey and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, San Francisco, 1979.

[3] R. Karp. Reducibility among combinatorial problems. In R. Miller and J. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, New York, 1972.

[4] S. Martello and P. Toth. *Knapsack Problems: Algorithms and Computer Implementations.* Wiley & Sons, 1990.

[5] M. Meanti, A. R. Kan, L. Stougie, and C. Vercellis. A probabilistic analysis of the multiknapsack value function. *Mathematical Programming*, 46:237–247, 1990.

[6] G. Nemhauser and L. Wolsey. *Integer and Combinatorial Optimization.* John Wiley & Sons Inc., New York, 1988.

[7] K. Szkatuła. On the growth of multi-constraint random knapsacks with various right-hand sides of the constraints. *European Journal of Operational Reserch*, 73:199–204, 1994.

[8] K. Szkatuła. The growth of multi-constraint random knapsacks with large right-hand sides of the constraints. *Operations Research Letters*, 21:25–30, 1997.