



DALARNA
UNIVERSITY

**Working papers in transport, tourism, information technology and microdata
analysis**

GRASP and statistical bounds for heuristic solutions to combinatorial problems



Kenneth Carling

Mengjie Han

Editor: Hasan Fleyeh

Nr: 2016:06

Working papers in transport, tourism, information technology and microdata analysis
ISSN: 1650-5581

© Authors

Abstract: The quality of a heuristic solution to a NP-hard combinatorial problem is hard to assess. A few studies have advocated and tested statistical bounds as a method for assessment. These studies indicate that statistical bounds are superior to the more widely known and used deterministic bounds. However, the previous studies have been limited to a few metaheuristics and combinatorial problems and, hence, the general performance of statistical bounds in combinatorial optimization remains an open question. This work complements the existing literature on statistical bounds by testing them on the metaheuristic Greedy Randomized Adaptive Search Procedures (GRASP) and four combinatorial problems. Our findings confirm previous results that statistical bounds are reliable for the p -median problem, while we note that they also seem reliable for the set covering problem. For the quadratic assignment problem, the statistical bounds has previously been found reliable when obtained from the Genetic algorithm whereas in this work they found less reliable. Finally, we provide statistical bounds to four 2-path network design problem instances for which the optimum is currently unknown.

1 Introduction

Solutions to combinatorial problems have been extensively studied in the field of discrete optimization. The application to combinatorial problems are found on classical problems like traveling salesman problem, location problem, network design and scheduling problems. For a simple problem, the complete algorithm guarantees to find the optimal solution. However, for less simple problems, complete algorithms will fail to find the optimal solution in the polynomial time, in which case the problem is known as NP-hard and therefore computationally intractable.

Many algorithms for NP-hard problems have been developed in the last 30 years. Metaheuristics is a high level strategy for exploring search spaces by using different methods (Blum and Roli, 2003). Examples are GRASP, Tabu Search, Iterated Local Search, Variable Neighborhood Search, Genetic search and Simulated Annealing. In contrast to the complete algorithms, metaheuristics try to locally search for good solutions, thereby significantly reducing the combinations to evaluate. They usually start from some initial solution and iteratively try to replace the current solution by a

* Kenneth Carling is a professor in Statistics and Mengjie Han is a PhD in Microdata Analysis at the School of Technology and Business Studies, Dalarna University, SE-791 88 Falun, Sweden. *Corresponding author:* Mengjie Han, e-mail: mea@du.se, phone: +46-76-5835398.

better solution in an appropriately defined neighborhood of the current solution (Blum and Roli, 2003). In doing so, the guarantee of finding the optimal solutions to the problem has been sacrificed.

The prevailing practice is to run the metaheuristic for a pre-specified number of iterations or until improvement in the solution becomes infrequent. However for a specific problem instance, such practice does not readily lend itself to the assessment of the quality of the solution (Carling and Meng, 2016). One approach to assess the quality is to seek the deterministic bounds (Beasley, 1993; Lemaréchal and Oustry, 1999). This approach is however applicable to a few heuristics and sensitive to parameter settings. For an unsuitable parameter setting, the gap of the bounds is usually large and thus uninformative.

Another approach is to use the statistical property of the solution. Specifically, the statistical approach implies point estimation to the optimum with an associated uncertainty interval at some confidence level. Golden and Alt (1979) did pioneering work on statistical bounds followed by others in 1980s. Derigs (1985) conducted computer experiments with regard to the confidence limits for travelling salesmen and quadratic assignment problem instances by applying k -interchange heuristics. Although Derigs' (1985) conclusions were generally favorable for the statistical bounds, his work received little attention during three decades in the practice of combinatorial optimization. Recently, Giddings *et al* (2014) reviewed four statistical optimum estimation techniques which provides a comprehensive framework of optimum estimation. From their concluding discussion, it is clear that for statistical optimum estimation techniques to be well understood and practically useful in combinatorial optimization a great number of varying problem instance specific computer experiments are warranted. In addition to Derigs' (1985) computer experiments, Carling and Meng (2015, 2016) studied the statistical confidence for the p -median problems using several metaheuristics, while Meng (2015) tested statistical bounds using the Genetic algorithm on QAP. These studies suggest that statistical bounds do provide a good assessment of the quality in the heuristic solution with a greater precision than deterministic bounds.

In this paper, our aim is to complement the existing knowledge about the performance of statistical bounds by conducting novel computer experiments of one metaheuristic algorithm, GRASP, on four combinatorial problems. The work is inspired by a recent study by Ribeiro *et al* (2013) who developed a probabilistic stopping rule for the GRASP heuristics. They studied the power of the statistical distribution and gave the likelihood of a better solution for a certain, additional number of iterations. This is informative if a pre-set threshold, β , is set since the expected number of better solutions in the next run can be estimated. However, they found the approach to fail when $\beta \leq 10^{-4}$ for QAP and set covering problem instances. Since the gap between the statistical bounds can be used as a stopping criterion, this work is also complementing the work on stopping rules for GRASP of Ribeiro *et al* (2013).

The remaining of the paper is organized as follows. Section 2 gives a short description of the methodology; Section 3 introduces GRASP and the problem instances; Section 4 gives the results; and the final section provides concluding remarks.

2 Methodology

The statistical optimum estimation techniques (SOETs) are extensively reviewed by Giddings *et al* (2014). For future (computer) studies of SOET they proposed an evaluation framework with associated notation. A problem instance I is a particular combinatorial optimization problem with fixed constraint and objective function coefficients (see also Meng, 2015). A problem class \mathcal{J} is a group of problem instances where $I \in \mathcal{J}$. Similarly, \mathcal{H} is the collection of all possible heuristics. For a particular heuristic algorithm $H \in \mathcal{H}$, the number of replicates, n , is usually arbitrarily determined by the experience. SOETs is a set consisting of the complete combination of problems and algorithms: $\mathcal{J} \times \mathcal{H}^n$. Thus, the task of summarizing all possible $\mathcal{J} \times \mathcal{H}^n$ is not straightforward, because same heuristics could perform completely differently on different problems. For this reason, it may be necessary to give customized statistical bound for each element in $\mathcal{J} \times \mathcal{H}^n$.

They also noted that studies on statistical bounds for combinatorial problems are rare. This far, Derigs (1985) found that statistical bounds are more useful than deterministic bounds for harder problems and that the k -interchange heuristics he studied gave more competitive statistical bounds for QAP than TSP. Derigs (1985) also pointed out that the Weibull hypothesis derived from extreme-value theory is a proper approach. Carling and Meng (2016) examined the Weibull (and the Gumbel) point estimator of the optimum as well as the first and second order Jackknifing (JK) point estimator on 40 p -median problem instances using Simulated Annealing as metaheuristic. Moreover, they proposed a SR statistic (hereafter referred to as SR) to determine if statistical bounds could be deemed reliable (and cautioned for the use of statistical bounds unless SR is smaller than 4). They also found the second order JK estimator preferable for both point and interval estimation. Further, Carling and Meng (2015) found the statistical bounds obtained from Simulated Annealing and Vertex Substitution to be reliable and much more efficient than deterministic bounds derived from Lagrangian relaxation for the same 40 p -median problem instances as above. Furthermore, they claimed that statistical bounds may be computed based on a small number of replicates ($n = 10$). Meng (2015) showed that the 4th order JK-estimator is an attractive SOET for difficult QAP instances when using the Genetic algorithm. Meng (2015) also compared statistical bounds to several methods to derive deterministic bounds and concluded that in general statistical bounds were more precise.

We follow the notation in Carling and Meng (2016), but generalize them from the p -median to all problem instances originally studied by Ribeiro *et al*, 2013. They are p -

median problem (PMP), quadratic assignment problem (QAP), set k-covering problem (SCP) and 2-path network design problem (2-PNDP). We introduce them in details in Section 3. The following notation applies throughout the paper:

z_p : a feasible solution of a specific instance (I);

A : the set of all feasible solutions to I , $A = \{z_1, z_2, \dots, z_N\}$;

$g(z_p)$: the value of the objective function at z_p ;

θ : the minimum of $g(z_p)$, $\theta = \min_A g(z_p)$;

$\hat{\theta}$: an estimator of θ ;

n : the number of replicates of the metaheuristic with unique random starting values;

\tilde{x}_i : the heuristic solution of the i th replicate, $i = 1, 2, \dots, n$;

$\tilde{x}_{(i)}$: the i th order statistic of n heuristic solutions.

To align with the work of Carling and Meng (2016, 2015), Meng (2015) and Ribeiro *et al* (2013), we will examine the Weibull point estimator, $\hat{\theta}_W$, and the second order JK point estimator, $\hat{\theta}_{JK}^{(2)}$. For the Weibull approach, $\hat{\theta}_W = \tilde{x}_{(1)}$ is the statistical upper bound. The statistical lower bound is $\hat{\theta}_W - \hat{c}$ with the confidence level of $(1 - e^{-n})$, where \hat{c} is the estimated shape parameter of Weibull distribution (Wilson *et al*, 2004). One good choice is $\hat{c} = \tilde{x}_{[0.63(n+1)]} - (\tilde{x}_{(1)}\tilde{x}_{(n)} - \tilde{x}_{(2)}^2)/(\tilde{x}_{(1)} + \tilde{x}_{(n)} - 2\tilde{x}_{(2)})$ as suggested by Derigs (1985).

The second order JK point estimator is $\hat{\theta}_{JK}^{(2)} = 3\tilde{x}_{(1)} - 3\tilde{x}_{(2)} + 3\tilde{x}_{(3)}$, whereas the upper statistical bound is $\tilde{x}_{(1)}$. The lower statistical bound is obtained by $\hat{\theta}_{JK}^{(2)} - 3\sigma^*(\hat{\theta}_{JK}^{(2)})$ where $\sigma^*(\hat{\theta}_{JK}^{(2)})$ is the standard deviation of $\hat{\theta}_{JK}^{(2)}$ obtained from bootstrapping the n heuristic solutions. Carling and Meng (2016) proposed using 1,000 bootstrap samples such that the confidence level is nominally 99.9%.

3 GRASP and the problem instance

We give a template for the GRASP and the test problem instances. GRASP is an iterative randomized sampling metaheuristic and is first introduced by Feo and Resende (1995). Three test problems: PMP, QAP and SCP are used to evaluate the statistical bounds since the optimal solutions are available in the public libraries (Beasley, 1990 and Burkard *et al*, 1997). The fourth problem 2-PNDP and its four instances is a random data set provided to us by Ribeiro *et al* (2013). The optimal solutions to these four problem instances are unknown, but Ribeiro *et al* (2013) have provided us with their best heuristic solutions for comparison. We will give our suggested benchmarks of the optimal solutions in Section 4.

3.1 Description of GRASP

<p>GRASP</p> <ol style="list-style-type: none"> 1 Instance Input; 2 While (stopping criteria not reached){ 3 Construction stage(RCL) → Solution 1; 4 Local search stage(Solution 1) → Solution 2; 5 Update Solution 2 → Better Solution; 6 Check stopping criteria}; 7 Return Best Solution; <p>End</p>

Figure 1 Pseudo-code of GRASP

GRASP is an iterative algorithm with random starting values. The pseudo-code is shown in Figure 1. Line 1 inputs a specific problem instance by identifying the feasible solutions z_p and the object function $g(z_p)$. Line 2-6 run the construction stage and the local search stage when the stopping criteria is not reached. The number of iterations is pre-set, but can be completely different for each problem. Line 7 returns the best solutions.

The construction and local search stages are problem-dependent and should be customized for each problem (Ribeiro and Rosseti, 2007). In the construction stage, the probabilistic component is characterized by randomly selecting one of the best candidates in the list, *restricted candidate list* (RCL), but not necessary the top candidate (Feo and Resende, 1995). RCL is formed by a number of the elements which add smallest incremental cost to the current solution. RCL is updated each time one element is selected until a feasible solution is formed. The construction solution is random in each iteration and is regarded as the input of the local search stage.

The local search stage tries to improve the solution by searching in neighbor solutions. The definition of neighbor is flexible and also problem-dependent. The local optimum solution is found when the local search stopping criteria is reached.

Like other metaheuristics, the value of applying GRASP on combinatorial problems is to avoid being trapped in local minima. This will be at the cost of less efficient local search. The parallel starting procedure is able to search the solutions in a wide range, though new iterations will not learn from previous iterations. In order to reduce the computing cost, the parameter settings are problem dependent and to be discussed in Section 4.

3.2 Problems and instances

Four combinatorial problems are to be described and studied. We follow Ribeiro *et al* (2013) and examine the same four instances for each of the four problems.

3.2.1 p -median problem

The discrete p -median model was first introduced by Hakimi (1964) and it is NP-hard (Kariv and Hakimi, 1979). The goal with the model is to find p supply nodes which minimize the summed distances between each demand node and their nearest supply node. This problem can be formulated to minimize

$$f_1 = \sum_{i=1}^q \sum_{j=1}^q w_i d_{ij} x_{ij}$$

subject to $\sum_{j=1}^q x_{ij} = 1$ and $\sum_{j=1}^q x_{jj} = p$, where f is the value of objective function. q is the number of demand locations and also stipulated to be the number of candidate nodes. w_i is the weight of each demand location. d_{ij} is the distance from demand location i to the center j . x_{ij} is a binary variable: taking on 1 if location i is allocated to center j . Test instances are shown in Table 1 (Beasley, 1990)¹. The number of candidate nodes are between 200 and 800 and the number of supply nodes are between 67 and 200.

In the construction stage of GRASP, one of the nodes in RCL is selected and added into the solution until all p nodes are included. In the local search stage, nodes in the neighbors are examined. Substitution occurs when a low-cost neighbor is found (Resende and Werneck 2004).

Table 1 Test instances of the p -median problem

instances	q	number of edges	p
pmed10	200	800	67
pmed15	300	1800	100
pmed25	500	5000	167
pmed30	800	7200	200

3.2.2 Quadratic assignment problem

QAP was first defined by Koopmans and Beckmann (1957) and it is NP-hard (Sahni and Gonzalez, 1976). Consider two square matrixes $\mathbf{A} = (a_{ij})_{N_Q}$ and $\mathbf{B} = (b_{ij})_{N_Q}$. A set Y of all permutations of $\{1, 2, \dots, N_Q\}$ is formed by $N_Q!$ elements and $y \in Y$. The objective function of QAP is to minimize

¹ The file is organized in the OR-Library and the test instances are available for downloading at <http://people.brunel.ac.uk/~mastjib/jeb/orlib/files/>.

Table 2 Test instances of the quadratic assignment problem

instances	N_Q
tai30a	30
tai35a	35
tai40a	40
tai50a	50

$$f_2 = \sum_{i=1}^{N_Q} \sum_{j=1}^{N_Q} a_{ij} b_{y(i)y(j)}$$

Four problem instances are given in Table 2 including the instance name and size (Burkard *et al*, 1997)². When $N_Q > 15$, the QAP becomes very difficult to solve. The computational experiments with GRASP is similar to the PMP, though only interchanges are taken for the local search.

3.2.3 Set k-covering problem

SCP is a NP-complete problem of covering the rows of an m_s -row, n_s -column zero-one matrix (S_{ij}) by a subset of columns at minimum cost (Beasley, 1987 and Aho *et al* 1974). Suppose $x_j = 1$ if column j is in the solution with cost c_j , otherwise $x_j = 0$. The object is to minimize

$$f_3 = \sum_{j=1}^{n_s} c_j x_j$$

where $\sum_{i=1}^{m_s} s_{ij} x_j \geq k$ guarantees that each row is covered at least k times. The test instances are described in Table 3 (Beasley, 1990)³. Implementation for the GRASP is illustrated in Pessôa *et al* (2013).

Table 3 Test instances of the set k -covering problem

instances	m_s	n_s	k
scp42	200	1000	2
scp47	200	1000	2
scp55	200	2000	2
scpa2	300	3000	2

² The test instances can be download from QAPLIB at <http://anjos.mgi.polymtl.ca/qaplib/inst.html>.

³ See footnote 1.

3.2.4 2-path network design problem

Table 4 Test instances of the 2-path network design problem

instances	$ V $	$ E $	K
2pndp50	50	1225	500
2pndp70	70	2415	700
2pndp90	90	4005	900
2pndp200	200	19900	2000

The fourth problem does not have a known optimum to the four instances, which means that the statistical bounds cannot be checked with the optimum. For these problem instances we will provide statistical bounds as a complement to the few studies that have tried to provide a solution to the problem instances of 2-PNDP (Dahl and Johannessen, 2004; Ribeiro *et al*, 2013). Given a nonnegative edge supply graph (E, V) and a subset of $|V|$ pairs of nodes (demand graph), the 2-PNDP is to find the minimum weighted edges in the supply graph such that every pairs in the demand graph are connected by only one or two edges, i.e. minimizing

$$f_4 = \sum_{e \in E}^n c_e x_e$$

where c_e is the weight of the selected edge and $x_e \in \{0,1\}$ indicates if the edge is selected from E . Four test instances are shown in Table 4 (Ribeiro *et al*, 2013)⁴.

Riberio and Rosseti (2002, 2007) described the application of GRASP on 2-PNDP with path-relinking. Path-relinking re-searches best individual path without considering the cost increment in the current solution. After all node pairs in the demand graph find their new paths, the unnecessary paths are dropped. Path-relinking may be viewed as a constrained local search strategy applied to the current solution and is implemented each time before the neighborhood local search is conducted.

In the construction stage, a random order of demand pairs are constructed. Starting from the first pair, edges with the lowest weights are added to the solution set for each pair in the demand graph. The construction stage stops once all demand pairs are added. Path-relinking is done before the local search. In the local search stage, a random order of demand pairs is generated. Each demand pair is picked from the beginning to the end and better paths are searched for each pair. If better path with lower cost is found, the current solution is updated. When all the demand pairs are evaluated, the first local search is temporarily stopped. The value of the objective function is compared to the value before path relinking and local search. If the former

⁴ The data set is available by contacting the authors at mea@du.se.

value is lower, then it is required to re-start path relinking and more local search in a new circle. We iteratively run the path-relinking and local search until value of the objective function does not change.

4 Results

In this section we present the main findings in estimating the optimal solutions. We have run $n = 100$ replicates for each problem instance for the same reason as Carling and Meng (2015, 2016). Considering that the neighbors for local search are highly problem-dependent and the computing efforts for the number of the local searching stage increases as the problem complexity increases, the number of local search, \mathcal{L} , varies by problem instance. For example, we take $\mathcal{L} = 20,000$ for *pmed10* and $\mathcal{L} = 150,000$ for *pmed30* while \mathcal{L} varies between 1 million and 1.5 million for the most difficult problem being QAP.⁵ There is no simple way to compare our setting of \mathcal{L} on the instances to previous work on statistical bounds, but Carling and Meng (2015) provided information on their choice of iterations for Simulated Annealing and Vertex Substitution in terms of computing time. Our setting in the implementation of GRASP on the instances implies that the metaheuristic runs for equally long time as in Carling and Meng’s (2015) experiments for PMP and similarly for the SCP. The setting for QAP implies that the computational time is much larger.

As pointed out in Section 3, the local search stage for 2-PNDP differs from the other three problem. It varies in each iteration and is influenced by the solutions in the construction stage. Thus, there is no specific \mathcal{L} for 2-PNDP.

One *experiment* is defined as running GRASP $n = 100$ times for each problem instance. After each experiment, we also evaluate the statistic SR:

$$SR = \frac{1000\sigma(\tilde{x}_i)}{\hat{\theta}_{JK}^{(2)}}$$

SR measures the similarity amongst the solutions potentially allowing one to judge how far out in the tail and close to the optimum solutions they are. Hence, a large SR indicates the \tilde{x}_i :s are not near to the optimum. As noted above, $SR < 4$ is a sufficient condition for statistical bounds to be reliably applied for simulated annealing and vertex substitution heuristics on PMP (Carling and Meng, 2015, 2016). For GRASP it proved difficult to lower SR to a desirable level of about 4 within a reasonable computing time, and for most instances SR was found in the range of 5 to 10.

⁵ For SCP, the local searches varied between 7,000 to 10,000 with the most searches for the most difficult instance.

Table 5 Point estimators, relative bias (%) and SR

instances	θ	$bias(\hat{\theta}_{JK}^{(2)})$	$bias(\hat{\theta}_W)$	SR
pmed10	1255	-1.59	0.80	7.10
pmed15	1729	6.36	6.36	5.70
pmed25	1828	10.30	12.58	6.90
pmed30	1989	3.02	4.02	2.24
tai30a	909073	25.72	22.34	9.39
tai35a	1211001	23.98	25.15	8.17
tai40a	1569685	25.59	28.99	7.97
tai50a	2469398	30.78	33.85	1.69
scp42	1205	25.73	26.56	9.94
scp47	1115	11.66	21.52	8.95
scp55	550	16.36	20.00	6.38
scpa2	560	41.07	39.29	3.98

Next we evaluate both the point estimator and the coverage rate, i.e. whether the statistical bounds contain the optimum or not. Table 5 shows the point estimators $\hat{\theta}_{JK}^{(2)}$ and $\hat{\theta}_W$. We show the relative bias in millesimal, i.e. $bias(\hat{\theta}) = 1000 \times \frac{\hat{\theta} - \theta}{\theta}$.

In Table 5, θ is the known optimal value. Column 3-4 is the relative difference between the estimator and the optimal value. And the last column is SR. The results coincide with those in Ribeiro *et al* (2013) in terms of the distance to the optimal solutions. PMP is the easiest problem and gives smallest bias among three problems. QAP and SCP are harder problems. And $\hat{\theta}_{JK}^{(2)}$ for PMP always provides lower point estimation than $\hat{\theta}_W$. For QAP $\hat{\theta}_{JK}^{(2)}$ outperforms in three most complex instances: *tai30a*, *tai40a* and *tai50a* and also so on the three simpler instances of SCP. It is unsurprising that the harder instances of the same problem had lower SR considering that substantially more local searches were imposed on them.

The statistical bounds further state the precision of the estimators. To examine this empirically, we take a sample of size 10 with replacement from $n = 100$ replicates of each instance. This means that we approximate the true distribution of solutions obtained from infinitely many replicates on an instance, by the empirical distribution obtained from 100 replicates in order to reduce the time required for an experiment and allocate this time for conducting more experiments. We do this 1000 times and compute the lower bounds for the two estimators and compute the proportion of times that the lower bound and the upper bound contains θ . Further, the average interval is the mean value of 1000 lower bounds and upper bounds. In Table 6, coverage and average statistical bounds for each problem instance are given for both estimators.

Table 6 Statistical bounds

instances	Cov. $\hat{\theta}_{JK}^{(2)}$	Bounds $\hat{\theta}_{JK}^{(2)}$	Cov. $\hat{\theta}_W$	Bounds $\hat{\theta}_W$	SR
pmed10	1.00	[1231, 1262]	0.84	[1246, 1262]	7.10
pmed15	0.99	[1715, 1746]	0.40	[1729, 1746]	5.70
pmed25	0.95	[1808, 1864]	0.22	[1838, 1864]	6.90
pmed30	0.99	[1980, 2001]	0.28	[1992, 2001]	2.24
tai30a	0.69	[904K, 932K]	0.13	[918K, 932K]	9.39
tai35a	0.82	[1202K, 1234K]	0.13	[1218K, 1234K]	8.17
tai40a	0.45	[1569K, 1625K]	0.01	[1594K, 1625K]	7.97
tai50a	0.03	[2507K, 2567K]	0.02	[2536K, 2567K]	1.69
scp42	0.92	[1182, 1248]	0.26	[1212, 1248]	9.94
scp47	0.90	[1099, 1144]	0.27	[1122, 1144]	8.95
scp55	0.90	[543, 563]	0.21	[554, 563]	6.38
scpa2	0.45	[559, 588]	0.09	[576, 588]	3.98

As shown in Table 6, $\hat{\theta}_{JK}^{(2)}$ always produces higher coverage but wider intervals than $\hat{\theta}_W$. This is straightforward because one cannot expect for both good intervals and good coverage. Even though all coverage are greater than 0, QAP still performs worst in terms of coverage among those three problems. One reason is that GRASP tends to be trapped in the local optima and is difficult to escape. On the contrary, PMP and SCP are easier to cover the optimal value than QAP even though SCP is classified as harder problem. This can be seen from the high coverage for PMP and relatively high coverage for SCP. In this sense, 1000 samples and $SR < 10$ are enough to cover the known optimum.

Table 7 Point estimator and statistical bounds for 2-pndp

instances	θ_H	$\hat{\theta}_{JK}^{(2)}$	bounds $\hat{\theta}_{JK}^{(2)}$	$\hat{\theta}_W$	bounds $\hat{\theta}_W$	SR
2pndp50	316	322	[312, 323]	322	[320, 323]	3.66
2pndp70	463	466	[461, 467]	467	[465, 467]	2.71
2pndp90	646	635	[626, 638]	638	[634, 638]	3.91
2pndp200	1379	1381	[1374, 1384]	1384	[1380, 1384]	2.42

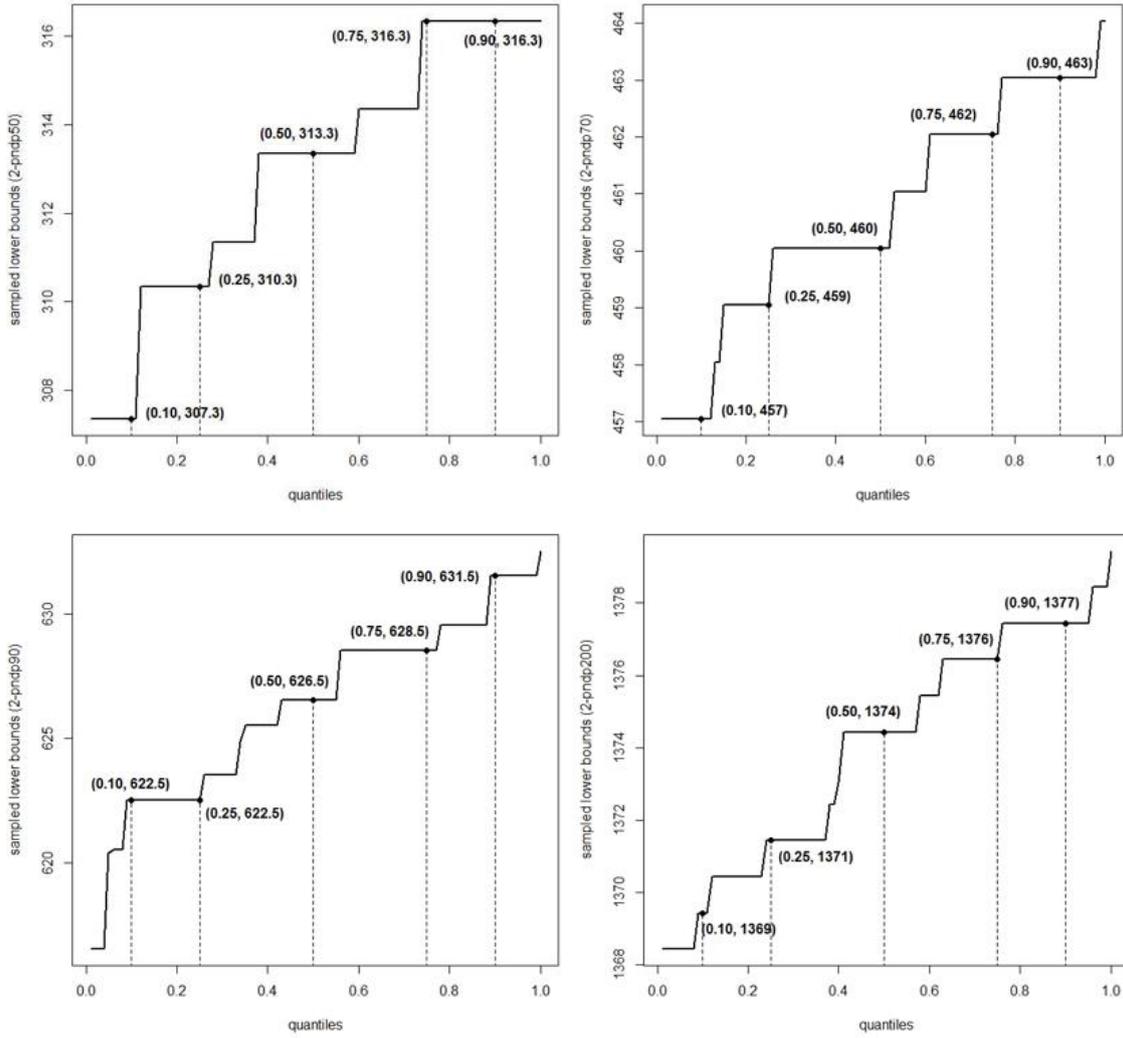


Figure 2 Samples lower bounds for $\hat{\theta}_{JK}^{(2)}$ to 2-pndp

Like PMP, 2-PNDP is also classified as simple problem in Ribeiro *et al* (2013) because a threshold $\beta = 10^{-5}$ quantile lower tail values are obtained. Thus, the true coverage should be similar to the PMP. Since optimal solutions are unknown and only best heuristic solutions θ_H are known (Ribeiro *et al*, 2013), we give our suggested statistical bounds for them. In Table 7, we show that the improved solution to 2-pndp90 is found and fairly tight intervals are constructed for each instance.

Figure 2 gives the distribution of the 1000 sampled lower bounds for $\hat{\theta}_{JK}^{(2)}$ to all four 2-pndp instances. The x-axis is the quantiles of the sorted values and y-axis is the value of the objective function. In the left bottom panel, for example, one can evaluate the solutions of 2-pndp90 by looking at the staircase graph. A solution with $f_4 = 628$ (if feasible) will fall in our suggested bounds and it will be covered by over 75% of 1000 sampled lower bounds.

6 Conclusions

In this paper, we studied two SOETs: truncation-point approach and extreme-value-theory when GRASP is applied on four combinatorial optimization problems. The optimal values are known for PMP, QAP and SCP. The point estimator and statistical bounds for them are constructed. We examined two estimators, $\hat{\theta}_{JK}^{(2)}$ and $\hat{\theta}_W$, for each problem instance. $\hat{\theta}_{JK}^{(2)}$ gives better estimation than $\hat{\theta}_W$ and the bounds of $\hat{\theta}_{JK}^{(2)}$ are reliable in terms of high coverage for PMP and SCP. However, less reliable bounds are found for QAP when GRASP heuristic is applied.

For the 2-PNDP, we give our suggested interval estimation by using the SR criteria since the optimal value is unknown and studies on it are infrequent. We also provide a benchmark for the future research.

Regarding the SR statistic that has been found important for judging the reliability of statistical bounds, we have had to content ourselves with SR in the range of 5 to 10 in this study of GRASP. This is a higher value possible indicating that the statistical bounds might be unreliable (Carling and Meng, 2016, 2015 advocated $SR < 4$ as a threshold for simulated annealing and vertex substitution for PMP). For problem instances with relatively low number of local searches, SR is difficult reduce to further without adding much more experiments. Hence, it remains an open question on how tight statistical bounds can be achieved by GRASP for the three combinatorial problems studied here with instances of known optimum.

Acknowledgement

We are grateful to Xiangli Meng who provided his computer code to our disposal as a starting point for this study. We are also grateful to Celso Ribeiro, Isabel Rosseti, and Reinaldo Souza who provided us with code for generating the 2-NDPD instances and their best solutions to the instances, as well as answering questions on details in their implementation of GRASP for the combinatorial problems.

References

- Aho, A.V., Hopcroft, J.E. and Ullman, J.D., 1974. The Design and Analysis of Computer Algorithms. *Addison-Wesley Publishing Company*, Reading, Mass.
- Beasley, J.E., 1987. An algorithm for the set covering problem. *European Journal of Operational Research* 31, 85-93.
- Beasley, J.E., 1990. OR library: distributing test problems by electronic mail. *Journal of the operational research society* 41(11), 1069-1072.
- Beasley, J.E., 1993. Lagrangean heuristics for location problems. *European Journal of Operational Research* 65(3), 383-399.

- Blum, C. and Roli, A., 2003. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys* 35(3), 268-308.
- Burkard, R.E., Karisch, S.E. and Rendl, F., 1997. QAPLIB--A quadratic assignment problem library. *Journal of Global Optimization* 10(4), 391-403.
- Carling, K and Meng, X., 2015. Confidence in heuristic solutions? *Journal of Combinatorial Optimization* 63(2), 381-399.
- Carling, K and Meng, X., 2016. On statistical bounds of heuristic solutions to location problems. *Journal of Combinatorial Optimization* 31(4), 1518-1549.
- Dahl, G. and Johannessen, B., 2004. The 2-path network problem. *Networks* 43(3), 190-199.
- Derigs, U., 1985. Using confidence limits for the global optimum in combinatorial optimization. *Operations Research* 33(5), 1024-1049.
- Feo, T.A. and Resende, M.G.C., 1995. Greedy randomized adaptive search procedures. *Journal of Global Optimization* 6, 10-133.
- Giddings, A.P., Rardin, R.L. and Uzsoy, R., 2014. Statistical optimum estimation techniques for combinatorial optimization problems: a review and critique. *Journal of Heuristics* 20(3), 329-358.
- Golden, B.L. and Alt, F.B., 1979. Interval estimation of a global optimum for large combinatorial optimization. *Naval Research Logistics Quarterly* 26(1), 69-77.
- Hakimi, S.L., 1964. Optimum locations of switching centers and the absolute centers and medians of graph. *Operations Research* 12(3), 450-459.
- Kariv, O. and Hakimi, S.L., 1979. An algorithmic approach to network location problems. II: The p -medians. *SIAM Journal on Applied Mathematics* 37(3), 539-560.
- Koopmans, T.C. and Beckmann, M.J., 1957. Assignment problems and the location of economic activities. *Econometrica* 25, 53-76.
- Lemaréchal, C. and Oustry, F., 1999. Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization. Technical Report 3710, INRIA Rhône-Alpes.
- Meng, X., 2015. Statistical bounds of genetic solutions to quadratic assignment problems. Working paper in transport, tourism, information technology and microdata analysis, 2015:02.

Pessôa, L.S., Resende, M.G.C. and Riberio C.C., 2013. A hybrid Lagrangean heuristic with GRASP and path-relinking for set k -covering. *Computers & Operations Research* 40(12), 3132-3146.

Ribeiro, C.C. and Rosseti, I., 2002. A parallel GRASP heuristic for the 2-path network design problem. *Lecture Notes in Computer Science* 2400, 922-926.

Ribeiro, C.C. and Rosseti, I., 2007. Efficient parallel cooperative implementations of GRASP heuristics. *Parallel Computing* 33, 21-35.

Ribeiro, C.C., Rosseti, I. and Souza, R. C., 2013. Probabilistic stopping rules for GRASP heuristics and extensions. *Intl. Trans. In Op. Res.* 20, 301-323.

Resende, M.G.C. and Werneck, R.F., 2004. A hybrid heuristic for the p -median problem. *Journal of Heuristics* 10, 59-88.

Sahni, S. and Gonzalez, T., 1976. P-complete approximation problems. *Journal of the Association of Computing Machinery* 23, 555-565.

Wilson, D.S., King, R.E. and Wilson, J.R., 2004. Case study on statistically estimating minimum makespan for flow line scheduling problems. *Eur. J. Oper. Res.* 155, 439-454.