

Simple Random Sampling Estimation of the Number of Local Optima

Khulood Alyahya^(✉) and Jonathan E. Rowe

School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK
{kya020, J.E.Rowe}@cs.bham.ac.uk
<http://www.cs.bham.ac.uk>

Abstract. We evaluate the performance of estimating the number of local optima by estimating their proportion in the search space using simple random sampling (SRS). The performance of this method is compared against that of the jackknife method. The methods are used to estimate the number of optima in two landscapes of random instances of some combinatorial optimisation problems. SRS provides a cheap, unbiased and accurate estimate when the proportion is not exceedingly small. We discuss choices of confidence interval in the case of extremely small proportion. In such cases, the method more likely provides an upper bound to the number of optima and can be combined with other methods to obtain a better lower bound. We suggest that SRS should be the first choice for estimating the number of optima when no prior information is available about the landscape under study.

1 Introduction

Local search algorithms are widely used to find solutions to many optimisation problems either on their own or as a part of other metaheuristics. The neighbourhood operator they employ defines a structure over the search space; the properties of that structure can strongly influence their performance. One of these properties is the number of local optima, which combined with the additional knowledge of other properties such as the quality of the optima and the correlation between the basin size and fitness can give an indication of the structure difficulty. Nonetheless, knowing only the number of local optima can still provide some guidance in informing the choice of the neighbourhood operator. The knowledge of the number of local optima can also be used to study its growth behaviour, as the dimensionality increases, or across different values of problem parameters (e.g. phase transition control parameter). However, the number of local optima in a given instance is not known in advance and counting them is infeasible in most cases, apart from very small problem sizes. Therefore, the need for obtaining a statistical estimate of the number of local optima arises. Having an estimate of the total number of optima can also be helpful in commenting on the quality of the found local optima or the confidence that the global has been seen [17]. In the last two decades, a number of approaches have

been proposed for estimating the number of local optima in combinatorial optimisation problems (COPs) [4, 6, 8, 9, 15–17]. Most of these methods start from a random sample of different configurations and apply local search to them until a local optimum is reached. Some of the methods are non-parametric estimators such as jackknife and bootstrap [6], while others assume some parametric distribution of the basin sizes (e.g. gamma distributions) [8, 9]. However, each of these methods has its particular limitations and none of them provide a good estimate in all scenarios (e.g. when the basin sizes are different or when the number of optima is small). For example, the jackknife method [6] requires the sample size to increase as the number of optima increases, which is impractical since the number of optima grows exponentially or sub-exponentially with the problem size in most problems [13, 15]. One drawback of the bootstrap method is its computational demands to carry out the re-samplings [6]. The approach proposed by [8] models the basin sizes using gamma distribution and requires an estimate of the parameter value of the distribution, which may not be practical. Another possible limitation of all the methods that apply local search to an initial random sample is the time needed to converge to a local optimum. In many cases, this time is linear or superlinear in problem size [15, 21], but it can be exponential in other cases [5]. A review and an evaluation for several of these methods and others from the statistical literature can be found in [12].

The problem of estimating the number of local optima in COPs can be considered as the classical problem of estimating a population proportion in statistics. However, the use of this method to estimate the number of local optima is seldom found in the literature. It has been used to estimate number of optima in the multidimensional assignment problem [11], and in the quadratic assignment problem [19, 20]. [4] mentioned the attractiveness of the simplicity and the unbiased estimate provided by this method, but they argued against it as the required sample size can be very large when the proportion is exceedingly small. They also criticised that in such a case, the method is more likely to provide an upper bound estimate rather than a lower bound one. [12] recommends using it only when all or most of the sampled optima have been seen once, after applying local search to an initial sample of points. This method is problem-independent and we argue that it is the best for estimating the number of local optima in terms of simplicity, accuracy and computational requirement when their proportion is large. As mentioned before, the required sample size for an accurate estimate increases as the proportion decreases, which makes obtaining an accurate estimate very expensive. However, an upper bound on the number of optima in such cases can still be obtained with reasonable sample sizes, giving some useful information about the studied landscapes. In the rest of this paper, we refer to estimating the number of local optima by estimating their proportion as simple random sampling (SRS). To provide a baseline, we compare the performance of SRS with the performance of the jackknife method. In Sect. 2, we introduce some preliminaries. In Sect. 3, we describe SRS and jackknife, and discuss different choices of confidence intervals for SRS. In Sect. 4 we describe the experimental settings and discuss the results.

2 Preliminaries

Search Space: The search space X is the finite set of all the candidate solutions. The fitness functions of all the studied problems in this paper are pseudo-Boolean functions, hence the search space size is $|X| = 2^n$.

Neighbourhood: A neighbourhood is a mapping $N : X \rightarrow P(X)$, that associates each solution with a set of candidate solutions, called neighbours, which can be reached by applying the neighbourhood operator once. The set of neighbours of x is called $N(x)$, and $x \notin N(x)$. We consider two different neighbourhood operators: the Hamming 1 operator ($H1$) and the 1+2 Hamming operator ($H1+2$). The neighbourhood of the $H1$ operator is the set of points that are reached by a 1-bit flip mutation of the current solution x , hence the neighbourhood size is $|N(x)| = n$. The neighbourhood of the $H1+2$ operator includes the Hamming one neighbours in addition to the Hamming two neighbours of the current solution x , which can be reached by a 2-bits flip mutation. The neighbourhood size for this operator is $|N(x)| = (n^2 + n)/2$.

Fitness Landscape: The fitness landscape of a combinatorial optimisation problem is a triple (X, N, f) , where f is the objective function $f : X \rightarrow R$, X is the search space and N is the neighbourhood operator function [18].

Local Optima: We define a local minimum $x^* \in X$ as $f(y) > f(x^*)$ for all $y \in N(x^*)$. A local maximum is defined analogously. We use the term local optimum to denote either a local maximum or a local minimum. We refer to the actual number of optima in a given landscape as v .

Local Search: The local search strategy we use is the best improving move, stopping when a local optimum is reached.

Basin of Attraction: The basin of attraction $B(x^*)$ for an optimum $x^* \in X$ is the set of points that leads to it after applying local search to them, $B(x^*) = \{x \in X \mid \text{local-search}(x) = x^*\}$.

3 Estimation Methods

3.1 Simple Random Sampling

Suppose that a random sample of size s is taken from the search space, and that Y optima has been observed in the sample ($0 \leq Y \leq s$), p is the unknown proportion of the optima in the search space. Since the sample size is fixed, and the sampled configurations are independent and have a constant probability of being an optimum given by p , then Y has a Binomial distribution, $B(s, p)$, with s trials and p success probability. The unbiased point estimate of the population proportion is given by $\hat{p} = Y/s$ and the estimated number of local optima can then be directly calculated by multiplying \hat{p} by the search space size $S = |X|$. There are several methods for computing confidence interval estimates for p ; the most referred ones are based on the approximation of the binomial distribution by the normal distribution [14]. A rule of thumb, that is frequently mentioned, is that the binomial distribution is suitable for approximation by the normal

distribution as long as $sp \geq 5$ and $s(1-p) \geq 5$ [2,22]. The most widely used confidence interval for p is the standard Wald confidence interval (CI_s) [2,14,22]:

$$CI_s = \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{s}} \quad (1)$$

where $z_{\alpha/2}$ is the z-score for $(1-\alpha)100\%$ confidence level and $z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{s}}$ is the error margin e . The error margin can be corrected for a finite population of size S to be equal to $e = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{s}} \sqrt{\frac{S-s}{S-1}}$, where the value $\sqrt{\frac{S-s}{S-1}}$ is the finite population correction (fpc) factor [22]. The value of fpc is approximately one when S is large compared to s , and is obviously equal to zero when $s = S$. The sample size for a desired confidence level and a desired margin of error can be determined for an infinite population by:

$$s_0 = \frac{z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{e^2} \quad (2)$$

If no prior information about p or no initial estimate of \hat{p} is available, then \hat{p} can conservatively be set to 0.5 where the expression $\hat{p}(1-\hat{p})$ is maximised. This will ensure that the sample size is at its maximum for the desired e . However, the proportion of optima is typically much smaller than that, thus it might be more wise to set p to a smaller value and set e to a much smaller value. The sample size can be corrected for a finite population by the following formula:

$$s_1 = \frac{s_0 S}{s_0 + (S-1)} \quad (3)$$

From Eq. (2) we can see that the sample size does not depend on the population size but only on the desired confidence level, the desired margin of error, and the estimate of p . The behaviour of Wald interval is poor when p is close to 0 or 1, and when $Y = 0$ or $Y = s$, the length of the Wald interval is zero [1,2,14]. The exact Clopper-Pearson interval (exact in the sense of using the binomial distribution rather than the approximation by the normal distribution) is an alternative method to consider in such cases. However, and because of the inherent conservativeness of exact methods, other approximate methods are more useful [1]. The *Agresti-Coull* confidence interval (CI_{AC}) is recommended for correcting the Wald interval. It recentres the Wald interval by adding the value $z_{\alpha/2}^2/2$ to Y so it becomes $\tilde{Y} = Y + z_{\alpha/2}^2/2$ and adding the value $z_{\alpha/2}^2$ to s to become $\tilde{s} = s + z_{\alpha/2}^2$. When the z-score for the 95% confidence level ($z_{0.05/2}^2 = 1.96$) is approximated to 2, the *Agresti-Coull* interval is equivalent to adding two successes and two failures to the sample [1,2]. The corrected point estimate is $\tilde{p} = \tilde{Y}/\tilde{s}$ and the confidence interval is given by:

$$CI_{AC} = \tilde{p} \pm z_{\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{s}}} \quad (4)$$

Using *Agresti-Coull* confidence interval, the SRS estimation of the number of local optima is given by:

$$\hat{v}^{SRS} = \tilde{p}S \quad (5)$$

3.2 Jackknife

Jackknife is a non-parametric method based on the idea of re-sampling to reduce the bias of the estimate. The use of jackknife to estimate the number of local optima was first proposed by [6]. We selected the jackknife method as a comparison baseline for two reasons: jackknife has an attractive simple and fast closed-form computation, and it is recommend to be used when the size of the sample is adequate with respect to v [6,12]. Starting from s different randomly sampled configurations and after applying local search to each one of them, the jackknife estimate of the number of local optima is given by:

$$\hat{v}^{JK} = \beta + \frac{s-1}{s}\beta_1 \quad (6)$$

where β_1 is the number of optima that have been seen once and $\beta = \sum_{i=1}^r \beta_i$ is the number of distinct optima seen. Note that this is a special case of the jackknife estimator where one point is left out of the original sample s at a time. A generalised estimator that considers leaving out $1, \dots, 5$ points at a time can be found in [3]. As pointed out by [17], the choice of the most suitable number of points to leave out in order to achieve a better estimate is problem-dependent.

4 Experiments

We obtain statistical estimates of the number of optima in randomly generated instances of the number partitioning problem and the 0–1 knapsack problem. The aim of the experiments is twofold: compare the estimates of SRS with that of jackknife, and examine the effect of the sample size on the accuracy of the SRS estimate. We compare the performance of the two methods using two sample sizes to allow for a fair comparison, since SRS uses at most $s(|N(x)| + 1)$ number of fitness evaluations compared to $s(|N(x)| + 1) + t|N(x)|$ fitness evaluations used by jackknife, where t is the total number of steps taken when descending(ascending) from each initial configuration. We describe the settings of the two sample sizes in more details in the results subsection.

4.1 Combinatorial Optimization Problems

Number Partitioning Problem (NPP). Given a set $W = \{w_1, \dots, w_n\}$ of m -bit positive integers (weights) drawn at random from the set $\{1, 2, \dots, M\}$ with $M = 2^m$, the goal is to partition W into two disjoint subsets S, S' such that the discrepancy between them $|\sum_{w_i \in S} w_i - \sum_{w_i \in S'} w_i|$ is minimised. The instances we study have weights drawn from a uniform distribution and $m = n$.

When the weights are drawn from a uniform distribution, the theoretical average proportion of the local optima in the $H1$ landscape is given by the following formula that was obtained using statistical mechanics analysis [7]:

$$\langle p \rangle^{\text{NPP}} = \sqrt{\frac{24}{\pi}} n^{-3/2} \quad (7)$$

0-1 Knapsack Problem (0-1KP) is defined as follows: given a knapsack of capacity C and a set of n items each with associated weight w_i and profit p_i , the aim is to find a subset of items that maximises $f(x) = \sum_{i=1}^n x_i p_i$, subject to $\sum_{i=1}^n x_i w_i \leq C$, where $x \in \{0, 1\}^n$, $C = \lambda \sum_{i=1}^n w_i$, and $0 \leq \lambda \leq 1$. Infeasible solutions that violate the given constraint are penalised by subtracting this value from the fitness function: $\text{Pen}(x) = \rho (\sum_{i=1}^n x_i w_i - C) + \sum_{i=1}^n p_i$, where $\rho = \max_{i=1, \dots, n} \{p_i\} / \min_{i=1, \dots, n} \{w_i\}$. The weights of the instances studied in this paper are drawn from a discretised normal distribution $\mathcal{N}(2^{n-1}, \frac{2^n}{10})$.

4.2 Results

The mean estimates of v in the two landscape of the 0-1KP is shown as n grows in Fig. 1 (note that some data points lie on top of each other). The estimates were obtained by the jackknife and SRS, and were averaged over 10 samples for each sample size. The sample sizes are set as follows: first we obtained the sample size s for each n from Eqs. (2) and (3) by setting $e = 0.005$, $\hat{p} = 0.3$ and $z_{\alpha/2} = 2.576$. Note that the sample size, only changes slightly as n increases, starting from $s = 45,701$ when $n = 18$, until it reaches $s = 55,351$ when $n = 100$. After obtaining s , we then set the small sample size of SRS to s and the small sample size of jackknife to $s - t + t/(|N(x)| + 1)$ (i.e. we subtract the fitness evaluations used when ascending from the sample budget). We set the large sample size of jackknife to s and the large sample size of SRS to $s + t - t/(|N(x)| + 1)$, where t is the total number of steps taken by jackknife with the large sample. The samples are drawn without replacement for $n \leq 24$. The figure shows that SRS using both small and large sample sizes accurately estimates the real proportions in both landscapes, apart from $n = 100$ in the $H1+2$ landscape. The discrepancy between estimates of the large and small samples in this case, in addition to the larger standard deviations, indicate that the proportion is small and that the sample size, in particular the small one is probably inadequate. As for the jackknife, both sample sizes quickly become inadequate as the number of optima

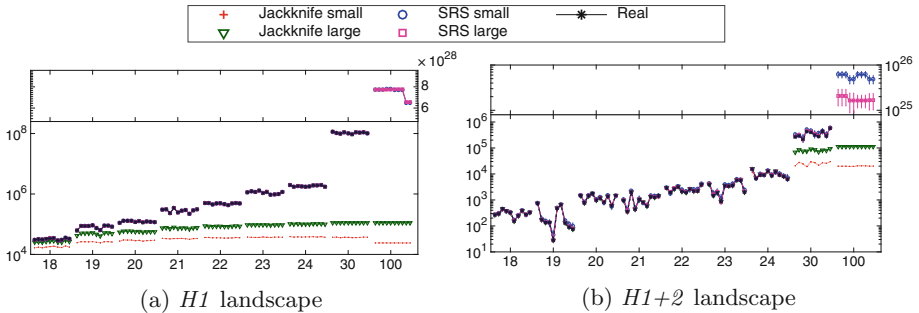


Fig. 1. SRS and Jackknife estimates of the optima number (in log scale) as the problem size grows. Each data point represents the average estimate of 10 samples from a single instance of 0-1KP. The error bars show the standard deviations.

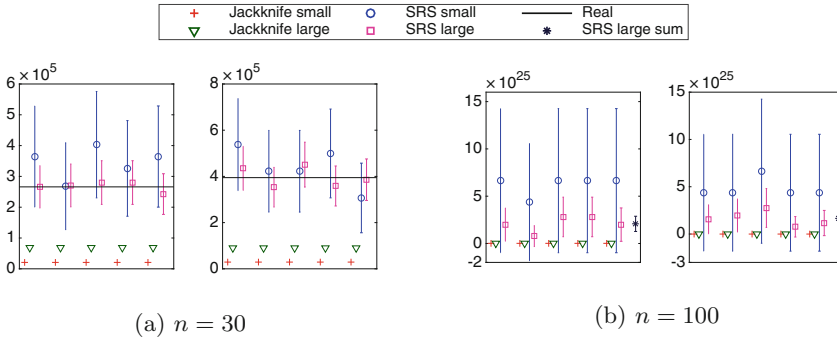


Fig. 2. Each figure shows the estimates of the number of optima in a single instance of 0-1KP, and each data point shows the estimate of a single sample. The error bars around SRS estimates are the 95 % CI_{AC} .

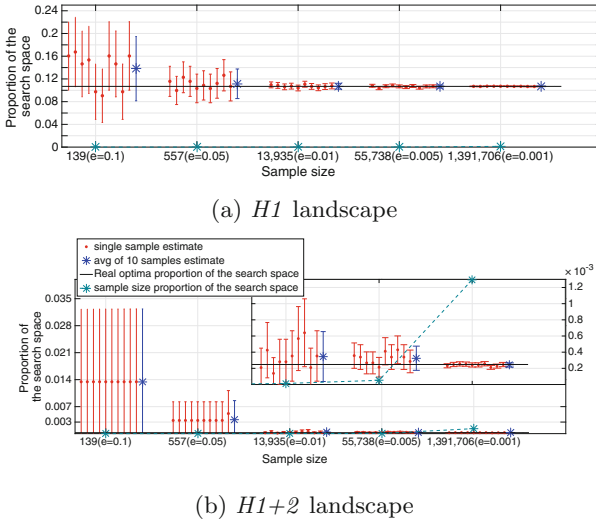


Fig. 3. SRS estimates of the optima proportion versus s . The sample sizes are obtained from Eqs. (2) and (3) by setting $\hat{p} = 0.3$ and $z_{\alpha/2} = 2.576$ (corresponding to 99 % confidence level). The results are for a single instance of 0-1KP of size $n = 30$. The error bars are the 95 % CI_{AC} .

seen once quickly grows with n until all the optima that have been seen were only seen once. Thus, the method fails to provide accurate estimates and grossly underestimates v . This is more noticeable in the $H1$ landscape where v is large. The CI_{AC} of SRS estimates are very narrow in $H1$ landscape across all n , but they get wider as n increases in the $H1+2$ landscape. In Fig. 2, we look closely at the results of four instances of size $n = 30, 100$ from Fig. 1. The figure shows the confidence interval around 5 estimates of each method with each sample size. The width of the CI_{AC} decreased with the large sample size as expected.

Table 1. NPP sample sizes

n	24	30	100	1000
s $e = \langle p \rangle^{\text{NPP}}$	276	388	2,395	75,915
$e = \frac{\langle p \rangle^{\text{NPP}}}{5}$	6,889	9,697	59,855	1,897,856
$e = \frac{\langle p \rangle^{\text{NPP}}}{10}$	27,520	38,785	239,420	7,591,421

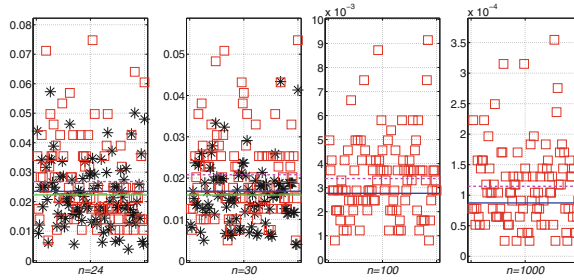
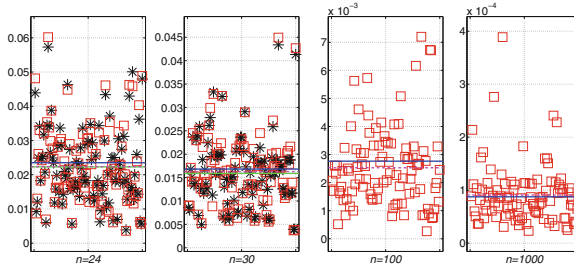
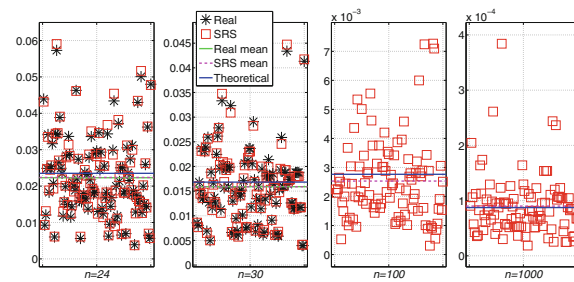
(a) $e = \langle p \rangle^{\text{NPP}}$ (b) $e = \frac{\langle p \rangle^{\text{NPP}}}{5}$ (c) $e = \frac{\langle p \rangle^{\text{NPP}}}{10}$

Fig. 4. Optima proportion in the $H1$ landscape of NPP for different vales of n . SRS estimates are shown when the sample size is obtained with 3 different desired error margins e (shown in Table 1). The results are for 100 random instances for each n . Obtaining the real proportion was only computationally feasible for $n = 24, 30$. The theoretical mean proportions are obtained from Eq. (7).

The SRS large sample size for $n = 30$ is around 2×10^5 and around 3×10^5 for $n = 100$. Obtaining the real number of optima was infeasible for $n = 100$ (note that methods that exploit some knowledge of f can obtain v of larger n than that feasible by exhaustive search of X [10]), therefore we show the estimate of SRS with a larger sample size by setting Y to the sum of the number of optima found in all the large samples and s to the sum of the large sample sizes. The outcome \hat{v}^{SRS} of both instances are around 10^{-5} . The very wide CI_{AC} with negative lower bounds around the small sample size estimates of SRS in $n = 100$ indicate that the proportion is much smaller than what SRS can precisely estimate with this sample size. In such a case, the \hat{v}^{SRS} more likely provides an upper bound to v . However, we suggest combining the results of the two methods in such cases by using the result of the jackknife method for a better lower bound than just zero.

Figures 3 and 4 show how the accuracy of SRS estimates increases as the desired error margin e decreases. Decreasing e consequently increases s . The figures also show how SRS is able to accurately estimate the fraction of v with relatively small s . As we mentioned before, the required s does not directly depend on n , but since the fraction of v usually declines as n grows [7], the required s will increase with n as shown in Table 1. The values of s in Table 1 are obtained from Eqs. (2) and (3) by setting $\hat{p} = \langle p \rangle^{NPP}$ (obtained from Eq. (7)), $z_{\alpha/2} = 2.576$ and e as shown in the table. In both problems and in both landscapes, most of the basin sizes are small and only very few ones are large.

5 Conclusions

Simple random sampling with the CI_{AC} provides a simple way to obtain an unbiased statistical estimate of the number of local optima. The accuracy of the obtained estimate depends on the sample size s , which can be determined for a desired margin of error e . A negative lower bound of the CI_{AC} usually indicates that the proportion is smaller than the desired e . In such a case, s can be increased considering that it only costs at most $|N(x)| + 1$ fitness evaluations per configuration. This is practical as long as the proportion is not exceedingly small. Alternatively, the estimate of SRS can be used as an upper bound as it is more likely to provide an overestimate in such cases. It can be combined with the estimate of another method that applies local search to an initial sample for a lower bound other than zero (since these methods usually tend to provide an underestimate [12]). We recommend that SRS should be the first method to use for estimating the number of optima, especially when no prior information is available about the problem being studied.

References

1. Agresti, A., Coull, B.A.: Approximate is better than “exact” for interval estimation of binomial proportions. *Am. Statistician* **52**(2), 119–126 (1998)
2. Brown, L.D., Cai, T.T., DasGupta, A.: Interval estimation for a binomial proportion. *Stat. Sci.* **16**(2), 101–117 (2001)

3. Burnham, K.P., Overton, W.S.: Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**(3), 625–633 (1978)
4. Caruana, R., Mullin, M.: Estimating the number of local minima in big, nasty search spaces. In: *Proceedings of IJCAI-1999 Workshop on Statistical Machine Learning for Large-Scale Optimization* (1999)
5. Englert, M., Röglin, H., Vöcking, B.: Worst case and probabilistic analysis of the 2-Opt algorithm for the TSP. *Algorithmica* **68**(1), 190–264 (2013)
6. Eremeev, A.V., Reeves, C.R.: Non-parametric estimation of properties of combinatorial landscapes. In: Cagnoni, S., Gottlieb, J., Hart, E., Middendorf, M., Raidl, G.R. (eds.) *EvoIASP 2002, EvoWorkshops 2002, EvoSTIM 2002, EvoCOP 2002, and EvoPlan 2002*. LNCS, vol. 2279, pp. 31–40. Springer, Heidelberg (2002)
7. Ferreira, F.F., Fontanari, J.F.: Probabilistic analysis of the number partitioning problem. *J. Phys. A: Math. Gen.* **31**(15), 3417 (1998)
8. Garnier, J., Kallel, L.: How to detect all maxima of a function. In: Kallel, L., Naudts, B., Rogers, A. (eds.) *Theoretical Aspects of Evolutionary Computing*, pp. 343–370. Springer, Heidelberg (2001)
9. Garnier, J., Kallel, L.: Efficiency of local search with multiple local optima. *SIAM J. Discret. Math.* **15**(1), 122–141 (2002)
10. Goldman, B.W., Punch, W.F.: Hyperplane elimination for quickly enumerating local optima. In: Chicano, F., et al. (eds.) *EvoCOP 2016*. LNCS, vol. 9595, pp. 154–169. Springer, Heidelberg (2016)
11. Grundel, D.A., Krokhmal, P.A., Oliveira, C.A.S., Pardalos, P.M.: On the number of local minima for the multidimensional assignment problem. *J. Comb. Optim.* **13**(1), 1–18 (2007)
12. Hernando, L., Mendiburu, A., Lozano, J.A.: An evaluation of methods for estimating the number of local optima in combinatorial optimization problems. *Evol. Comput.* **21**(4), 625–658 (2013)
13. Mathias, K.E., Whitley, L.D.: Transforming the search space with gray coding. In: *IEEE WCCI*, pp. 513–518, vol. 1 (1994)
14. Pires, A.M., Amado, C.: Interval estimators for a binomial proportion: comparison of twenty methods. *REVSTAT-Stat. J.* **6**(2), 165–197 (2008)
15. Prügel-Bennett, A., Tayarani-N, M.-H.: Maximum satisfiability: anatomy of the fitness landscape for a hard combinatorial optimization problem. *IEEE Trans. Evol. Comput.* **16**(3), 319–338 (2012)
16. Reeves, C.R.: Direct statistical estimation of GA landscape properties. *Found. Genet. Algorithms* **6**, 91–107 (2001)
17. Reeves, C.R., Eremeev, A.V.: Statistical analysis of local search landscapes. *J. Oper. Res. Soc.* **55**(7), 687–693 (2004)
18. Stadler, P.F., Stephens, C.R.: Landscapes and effective fitness. *Comments Theor. Biol.* **8**(4–5), 389–431 (2002)
19. Tayarani-N, M.-H., Prügel-Bennett, A.: On the landscape of combinatorial optimization problems. *IEEE Trans. Evol. Comput.* **18**(3), 420–434 (2014)
20. Tayarani-N, M.-H., Prügel-Bennett, A.: Quadratic assignment problem: a landscape analysis. *Evol. Intell.* **8**(4), 165–184 (2015)
21. Tovey, C.A.: Hill climbing with multiple local optima. *SIAM J. Algebraic Discrete Methods* **6**(3), 384–393 (1985)
22. Triola, M.F.: *Elementary Statistics*, 12th edn. Pearson, Upper Saddle River (2012)