On Confidence Intervals for the Number of Local Optima

Anton V. Eremeev[1] and Colin R Reeves[2]

[1] Omsk Branch of Sobolev Institute of Mathematics,
13 Pevtsov str. 644099, Omsk, Russia.
eremeev@iitam.omsk.net.ru

[2] Coventry University, School of Mathematical & Information Sciences,
Priory Street. Coventry CV1 5FB, UK
C.Reeves@coventry.ac.uk

Abstract. The number of local optima is an important indicator of optimization problem difficulty for local search algorithms. Here we will discuss some methods of finding the confidence intervals for this parameter in problems where the large cardinality of the search space does not allow exhaustive investigation of solutions. First results are reported that were obtained by using these methods for NK landscapes, and for the low autocorrelation binary sequence and vertex cover problems.

1 Introduction

Many heuristic search methods for combinatorial optimization problems (COPs) are based on local search. Typically, search is focused on a *neighbourhood* of the current point in the search space, and various strategies can be used for deciding on whether or not to move to one of these neighbours, and if so, to which. In practice, such methods quickly converge to a local optimum, and the search must either begin again by starting from a new point, or use some 'metaheuristic' (e.g., tabu search) to guide the search into new areas. While such local search methods can be arbitrarily bad in the worst case [1], experimental evidence shows that they tend to perform rather well across a large range of COPs. However, experience suggests that the type and size of the neighbourhood can make a big difference to practical performance of the local search methods. Some recent theoretical results also suggest that the local optima are tightly connected with the attractors of the simple genetic algorithm [2].

The neighbourhood structure can be described by a graph, or by its related adjacency matrix, which enables us to give a precise meaning to the idea of a *landscape* (which results from the interaction of this structure with a particular problem instance), and to its characteristics, such as local (and global) optima, and to related properties of an optimum such as its basin of attraction. Formally, we have an optimization problem: given a function (the objective or goal function) of a set of decision variables, denoted by the vector x,

$$g : \mathcal{X} \mapsto I\!R,$$

find

$$\arg\max_{x \in \mathcal{X}} g.$$

We define a neighbourhood function

$$N : \mathcal{X} \mapsto 2^{\mathcal{X}},$$

whence the induced landscape $\mathcal{L}$ is the triple $\mathcal{L} = (\mathcal{X}, g, d)$ where d denotes a distance measure $d : \mathcal{X} \times \mathcal{X} \to I\!R^{+} \cup \{\infty\}$ for which is required that

$$\left.\begin{array}{l} d(x,y) \geq 0 \\ d(x,y) = 0 \Leftrightarrow x = y \\ d(x,u) \leq d(x,y) + d(y,u) \end{array}\right\} \quad \forall\, x, y, u \in \mathcal{X}.$$

Usually, the distance measure d is defined by the neighbourhood structure:

$$y \in N(x) \Leftrightarrow d(x,y) = 1.$$

A *locally optimal* vector $x^o \in \mathcal{X}$ is then one such that

$$g(x^o) \geq g(y) \quad \forall\, y \in N(x^o).$$

We shall denote the set of such optima as $\mathcal{X}^o$. (For a fuller discussion on landscapes, see [3] or [4]; for some illustrations of the way landscapes are formed, and how they depend on neighbourhood structure, see [5].)

Many things make an instance difficult, but one of them is the *number* of optima induced in the landscape $\nu = |\mathcal{X}^o|$. Besides this an estimate of ν might be useful in defining stopping criteria for local search based metaheuristics, and in experimental testing of the 'big valley' conjecture [6] (for such an estimate may show how representative is the sample of the local optima found).

A number of approaches have been proposed for estimation of ν: one may apply an assumption that the distribution of basin sizes is isotropic [7], or fit certain type of parametric distribution of the basin sizes (exponential, gamma, lognormal etc.) [8, 9]. Nonparametric estimates, such as the bootstrap or the jackknife, can also be employed [9, 10]. Assuming a particular type of distribution of basin sizes one can obtain the maximal likelihood estimate for ν, or a confidence interval for it. It is interesting that some of these approaches are actually based on the same methods as used by ecologists to estimate the number of animals in a population (see e.g. [11]).

It would be interesting to see how ν depends on the problem size: theoretical results show that many optimization problems have an exponential number of local optima in the worst-case (see [1] and references therein), but experimental results have usually been limited to low-dimensional problems where complete enumeration of solutions is possible. At present, information on experimental evaluations of the parameter ν is scarce even for widely used benchmarks.

In this paper we suggest several methods of statistical evaluation of this parameter in problems where the large cardinality of the search space does not allow exhaustive investigation of solutions. The focus is on the theoretical basis and experimental testing of the methods on some well-known benchmarks.

The paper is organized as follows. Sect. 2 is devoted to the derivation of the computational formulae for the confidence intervals. After some general observations we analyse three methods of collecting the statistical information on an instance: the first repetition time, the maximum first repetition time and the Schnabel census procedure. The computational results obtained using these methods are reported in Sect. 3 for NK landscapes, the low autocorrelation binary sequence problem and the vertex cover problem. Finally the results obtained are discussed in Sect. 4.

2 Confidence Intervals for the Number of Optima

First let us consider the problem of numerical estimation of a confidence interval for the unknown distribution parameter ν, when the distribution function is easily computable for any value of an observable random variable T and any value of ν. Let the distribution function of T be $F_\mu(t) = P\{T \le t | \nu = \mu\}$ and assume that the observed value $\tau \in Z$ is sampled from this distribution.

In the case of real-valued ν, the $(1 - \alpha_1 - \alpha_2) \cdot 100\%$ confidence interval $[\nu_1(\tau), \nu_2(\tau)]$ for any given α_1, α_2 is found by solving the equations

$$1 - F_{\nu_1}(\tau) = \alpha_1, \quad F_{\nu_2}(\tau) = \alpha_2 \tag{1}$$

for ν_1 and ν_2 (see e.g.[12]). For our purposes to estimate the number of optima we need to modify this estimator for the situation where ν is integer.

Proposition 1. *A conservative* $(1 - \alpha_1 - \alpha_2) \cdot 100\%$ *confidence interval* $[\nu_1(\tau), \nu_2(\tau)]$ *for integer parameter* ν *is given by*

$$\nu_1(\tau) = \min\{\mu : 1 - F_\mu(\tau - 1) \ge \alpha_1\}, \quad \nu_2(\tau) = \max\{\mu : F_\mu(\tau) \ge \alpha_2\},$$

and then $P\{\nu < \nu_1(T)\} \le \alpha_1$, $P\{\nu_2(T) < \nu\} \le \alpha_2$.

Proof. Suppose that for some value ν the proposition is false. Then either $\alpha_1 < P\{\nu < \nu_1(\tau)\}$ or $\alpha_2 < P\{\nu_2(\tau) < \nu\}$ or both. In the first case we have:

$$\alpha_1 < P\{\nu < \min\{\mu : 1 - F_\mu(T - 1) \ge \alpha_1\}\} \le P\{1 - F_\nu(T - 1) < \alpha_1\} =$$

$$P\{T - 1 \ge \min\{\Theta : 1 - F_\nu(\Theta) < \alpha_1\}\} = P\{T \ge \Theta_{\min}(\nu, \alpha_1)\},$$

where $\Theta_{\min}(\nu, \alpha_1) = \min\{\Theta' : 1 - F_\nu(\Theta' - 1) < \alpha_1\}$. But $P\{T \ge \Theta_{\min}(\nu, \alpha_1)\} = 1 - F\{\Theta_{\min}(\nu, \alpha_1) - 1\} < \alpha_1$ by definition of $\Theta_{\min}$, which is a contradiction. In the second case

$$\alpha_2 < P\{\max\{\mu : F_\mu(T) \ge \alpha_2\} < \nu\} \le P\{F_\nu(T) < \alpha_2\} = P\{T < \theta_{\min}(\nu, \alpha_2)\},$$

where $\theta_{\min}(\nu, \alpha_2) = \min\{\theta : F_\nu(\theta) \ge \alpha_2\}$. But $P\{T < \theta_{\min}(\nu, \alpha_2)\} < \alpha_2$. □

Note that if $F_\nu(t)$ is monotone in ν, it is easy to compute the confidence interval suggested in Prop.1 using (e.g.) the bisection procedure. In what follows we will see that this monotonicity property holds for the estimators of ν we use.

2.1 The First Repetition Time

We consider the case where there are multiple independently initiated restarts of the local search procedure. Let $p_1, ..., p_\nu$ denote the probabilities of finding the corresponding local optima, $p = (p_1, ..., p_\nu)$ and let T denote the first repetition time, i.e. the index number of the restart which yields the first re-occurrence of a previously encountered optimum. In what follows the distribution function of T corresponding to the distribution vector p will be denoted by $F_p(t)$ and by the isotropic distribution vector we will mean $\overline{p} = (1/\nu, ..., 1/\nu)$.

The next proposition shows that the isotropic distribution is extremal in some sense.

Proposition 2. *$F_p(t)$ is minimal only at $p = \overline{p}$ for any $t \geq 2$.*

Proof. Let S be the set of all t-element sequences of the elements from $\{1, ..., \nu\}$ without repetitions, and $\sigma(i)$ is the i-th element of sequence σ. Then

$$F_p(t) = 1 - P\{T > t\} = 1 - \sum_{\sigma \in S} \prod_{i=1}^{t} p_{\sigma(i)}.$$

It is easy to see that $F_p(t)$ is a continuous function on a compact set $\Lambda = \{(p_1, p_2, ..., p_\nu) \geq 0 : \sum_{i=1}^{\nu} p_i = 1\}$ when t is fixed. Thus there exists a probability distribution yielding the minimum of $F_p(t)$ on Λ. We claim that except for $\overline{p}$ no other vector is minimal. Suppose the contrary: in the minimal vector p we have $p_i \neq p_j$ for some i, j. Without loss of generality we can assume that $i = 1, j = 2, p_2 = \gamma p_1$ and $\gamma < 1$. Let us consider a new distribution:

$$p'_1 = p'_2 = \frac{p_1 + p_2}{2}, p'_3 = p_3, ..., p'_\nu = p_\nu,$$

and its distribution function $F_{p'}(t)$. We shall now show that $F_{p'}(t) < F_p(t)$, which leads to a contradiction with the supposed optimality of p. Denote the subset of S consisting of the sequences containing element j by $S(j), j = 1, 2$. Then $F_p(t) = 1 - a(p) - b(p) - c(p) - d(p)$ where

$$a(p) = \sum_{\sigma \in S(1) \cap S(2)} \prod_{i=1}^{t} p_{\sigma(i)}, \qquad b(p) = \sum_{\sigma \in S(1) \setminus S(2)} \prod_{i=1}^{t} p_{\sigma(i)},$$

$$c(p) = \sum_{\sigma \in S(2) \setminus S(1)} \prod_{i=1}^{t} p_{\sigma(i)}, \qquad d(p) = \sum_{\sigma \in S \setminus (S(1) \cup S(2))} \prod_{i=1}^{t} p_{\sigma(i)}.$$

Analogously $F_{p'}(t) = 1 - a(p') - b(p') - c(p') - d(p')$ and $d(p') = d(p)$. Note that if the element 2 in all sequences of $S(2) \setminus S(1)$ was substituted by 1, this would yield the set of sequences $S(1) \setminus S(2)$. Thus, on the one hand

$$\begin{aligned} b(p) + c(p) &= \sum_{\sigma \in S(1) \setminus S(2)} \prod_{i=1}^{t} p_{\sigma(i)} + \gamma \left(\sum_{\sigma \in S(1) \setminus S(2)} \prod_{i=1}^{t} p_{\sigma(i)} \right) \\ &= (1 + \gamma) p_1 \cdot \sum_{\sigma \in S(1) \setminus S(2)} \prod_{i : \sigma(i) \neq 1} p_{\sigma(i)}. \end{aligned}$$

On the other hand $p'_1 = p'_2 = (p_1 + p_2)/2$, so that

$$b(p') + c(p') = (p_1 + p_2) \sum_{\sigma \in S(1) \setminus S(2)} \prod_{i:\sigma(i) \neq 1} p_{\sigma(i)},$$

and hence $b(p) + c(p) = b(p') + c(p')$. Now

$$a(p) = \gamma p_1^2 \cdot \sum_{\sigma \in S(1) \cap S(2)} \prod_{i:\sigma(i) \neq 1,2}^{t} p_{\sigma(i)},$$

$$a(p') = \left(\frac{p_1(1+\gamma)}{2}\right)^2 \sum_{\sigma \in S(1) \cap S(2)} \prod_{i:\sigma(i) \neq 1,2}^{t} p_{\sigma(i)}.$$

But $(1+\gamma)^2 > 4\gamma$ since $\gamma < 1$, so $a(p') > a(p)$ and $F_{p'}(t) < F_p(t)$. $\square$

This proposition is important on its own right but it also yields the monotonicity property of $F_{\overline{p}}(t)$ mentioned above:

Corollary 1. *$F_{\overline{p}}(t)$ is a strictly monotonically decreasing function on ν.*

Proof.
Apply the above result to the case $p = (1/(\nu-1), 1/(\nu-1), \ldots, 1/(\nu-1), 0)$. $\square$

2.2 The Maximal First Repetition Time

If the average repetition time turns out to be short compared to the number of restarts, the sequence of solutions obtained in a multiple restart procedure may be split into a series of subsequences, where each subsequence ends with its first re-occurrence. Let s denote the number of such subsequences and let the random variable T_j denote the length of j-th subsequence (including the re-occurring element). In this case the maximal first repetition time $T^{(s)} = \max_j T_j$ appears to be a more 'stable' estimate for ν compared to the simple first repetition time procedure discussed above. Denote $F_p^{(s)}(t) = P\{T^{(s)} \leq t\}$, then

$$F_p^{(s)}(t) = P\{T_j \leq t \ \ \forall j = 1, ..., s\} = \left[1 - \sum_{\sigma \in S} \prod_{i=1}^{t} p_{\sigma(i)}\right]^s,$$

where S is the set of sequences defined the same way as in Prop.2. Obviously, for a fixed s the isotropic distribution $\overline{p}$ will yield the minimum of $F_p^{(s)}(t)$ again, and it is easy to see that $F_{\overline{p}}^{(s)}(t) = (1 - \binom{\nu}{t} t!/\nu^t)^s$.

Owing to the monotonicity of $F_{\overline{p}}(t)$ (and $F_{\overline{p}}^{(s)}(t)$) in ν, in both cases the confidence interval for the unknown parameter ν can be estimated using Prop.1. Note that if α_2 is set to 0, in these confidence intervals ν_1 becomes a lower bound for ν with a confidence level $100(1-\alpha_1)\%$ and in view of Prop.2 this lower bound is valid for any probability distribution p on the set of local optima. In what follows we will use $L(t, s, \alpha)$ for the $100(1-\alpha)\%$-confidence lower bound constructed by Prop.1 with $\alpha_1 = \alpha, \alpha_2 = 0$ given the observed value $T^{(s)} = t$ and $F_\nu(t) = (1 - \binom{\nu}{t} t!/\nu^t)^s$.

If the computational resources are limited to r restarts, which all produce distinct outcomes, a valid $100(1-\alpha)\%$-confidence lower bound for ν could be obtained assuming that the next (hypothetical) restart will yield one of the previously visited optima. It is interesting to see how fast this lower bound $L(r+1, 1, \alpha)$ grows with the number of restarts. Note that computing $L(r+1, 1, \alpha)$ we seek for the minimal ν such that $\binom{\nu}{r} r!/\nu^r \geq \alpha$. But

$$\binom{\nu}{r}\frac{r!}{\nu^r} = \frac{\nu!}{(\nu-r)!\nu^r} = \exp\left\{\sum_{j=1}^{r} \ln\left(1 - \frac{r-j}{\nu}\right)\right\} \leq$$

$$\exp\left\{-\sum_{j=1}^{r} \frac{r-j}{\nu}\right\} = \exp\left\{-\frac{r(r+1)}{2\nu}\right\} \leq \exp\left\{-\frac{r^2}{2\nu}\right\},$$

so $L(r+1, 1, \alpha) \geq 0.5r^2/\ln(\alpha^{-1})$, and e.g. for $\alpha = 0.05$ we have $L(r+1, 1, 0.05) \approx 0.167r^2$. For large values of t this simple estimate is a good approximation to $L(r+1, 1, \alpha)$.

2.3 Schnabel Census Procedure

A common problem in ecology is to provide an estimate of the size of a population of animals by using the results of repeated samples. One such procedure is known as the *Schnabel census* (see e.g. [11]): we take repeated samples of size n (at suitable intervals of time) and count the number of *distinct* animals seen. The usual assumption is that the probability of catching any particular animal is the same. (Note that this is not the same as the subsequence idea discussed in the previous section: although it would be simple to map the subsequence approach onto the Schnabel census, the probability distribution is different.) We can readily apply this idea to the estimation of ν by repeated restarts of a local search procedure where we take $n = 1$.

Let r be the number of restarts, and K be the random variable for the number of distinct local optima found, while $F(k, \nu, r) = P\{K \leq k\}$ is the distribution function for K in case of the isotropic distribution $\overline{p}$.

For Schnabel census an *approximate* approach to computing the confidence interval for ν via a Normal approximation of the underlying distribution is well-studied (see e.g. [7, 11]). The following proposition provides a basis for an *exact* computation of the confidence intervals for ν from the observed value of K in a way suggested by Prop.1.

Proposition 3. *$F(k, \nu, r)$ is a non-increasing function on ν for any $k \in \mathbf{Z}$.*

Proof. By definition, $F(k, \nu, r) = \sum_{i=1}^{k} p(i, \nu, r)$, where

$$p(i, \nu, r) = P\{K = i\} = \frac{\nu! S(r, i)}{(\nu - i)!\nu^r}$$

(see e.g. [13]). $S(r,i)$ stands here for the Stirling number of the second kind [14], i.e. the number of all possible partitions of r-element set into i nonempty subsets. Let us consider the difference $\Delta_\nu(k) = F(k,\nu+1,r) - F(k,\nu,r)$. Obviously $\Delta_\nu(0) = \Delta_\nu(r) = 0$. We shall prove that $\Delta_\nu(k) \le 0$ for $k = 0,...,r$ by showing that $\Delta_\nu(k)$ is a non-increasing function on k when $0 \le k \le k^*$, and non-decreasing when $k^* \le k \le r$, where $k^* = \nu(1-\nu^{r-1})/(\nu+1)^{r-1}$. We will carry the proof by induction in two steps: a) for $k \le k^*$ and b) for $k^* \le k$ (note that these ranges overlap at $k = k^*$).

a) The basis of induction is with $k = 0$. Now

$$\Delta_\nu(k) = \Delta_\nu(k-1) + p(k,\nu+1,r) - p(k,\nu,r) < 0,$$

since $\Delta_\nu(k-1) \le 0$ by induction and

$$p(k,\nu+1,r) - p(k,\nu,r) = \frac{k-1}{\nu}\left(\frac{\nu+1}{\nu}\right)^{r-1} \frac{(\nu+1)!S(r,k)}{(\nu+1-k)!(\nu+1)^r} < 0.$$

b) For $k \ge k^*$ we use the basis $k = r$. Here

$$\Delta_\nu(k) = \Delta_\nu(k+1) - p(k+1,\nu+1,r) + p(k+1,\nu,r) \le 0,$$

since $\Delta_\nu(k+1) \le 0$ by induction and

$$p(k+1,\nu+1,r) - p(k+1,\nu,r) = \frac{k}{\nu}\left(\frac{\nu+1}{\nu}\right)^{r-1} \frac{(\nu+1)!S(r,k+1)}{(\nu-k)!(\nu+1)^r} \ge 0.$$

□

For the Schnabel census we also conjecture a statement similar to Prop.2. This claim is supported by experiments, but its proof is not known to us.

3 Computational Experiments

The estimators presented above have been tested in experiments with three sets of optimization problems of different kinds. Among them there are some NK landscapes (see e.g. [15]), the low autocorrelation binary sequence (LABS) problems [16] and some vertex cover problems (VCP) converted from the maximum clique instances suggested in [17]. In all cases we used a 'bit-flip' neighbourhood, i.e. $N(x) = \{y : d_H(x,y) = 1\}$ where $d_H(x,y)$ is Hamming distance. For each instance of NK landscapes and the LABS problems, $r = 1000$ restarts of local search were made, while the VCP instances demanded more of trials. Here we set $r = 10000$ where the computational load was not too great. The data from some of the experiments discussed below have been already used in [10]. For some low-dimensional problems it was possible to perform a complete enumeration of solutions—in these cases the actual ν-values are reported as well.

3.1 *NK* Landscapes

The experiments with NK landscapes of increasing ruggedness were performed for K ranging from 2 to 12 while keeping $N = 15$. The graph in Fig. 1 shows the total number of local optima and the two-sided confidence intervals for the Schnabel census with the isotropic assumption $p = \overline{p}$. These are displayed around the maximum likelihood estimate 'Sch.ML' [7, 10]. This graph also contains the first repetition time lower bounds $L(t_1, 1, \alpha)$, $L(\max_{j=1,\ldots,s} t_j, s, \alpha)$ computed on the basis of the same data from $r = 1000$ restarts (where s is the number of subsequences observed in r restarts). In all cases the confidence level is 95%.

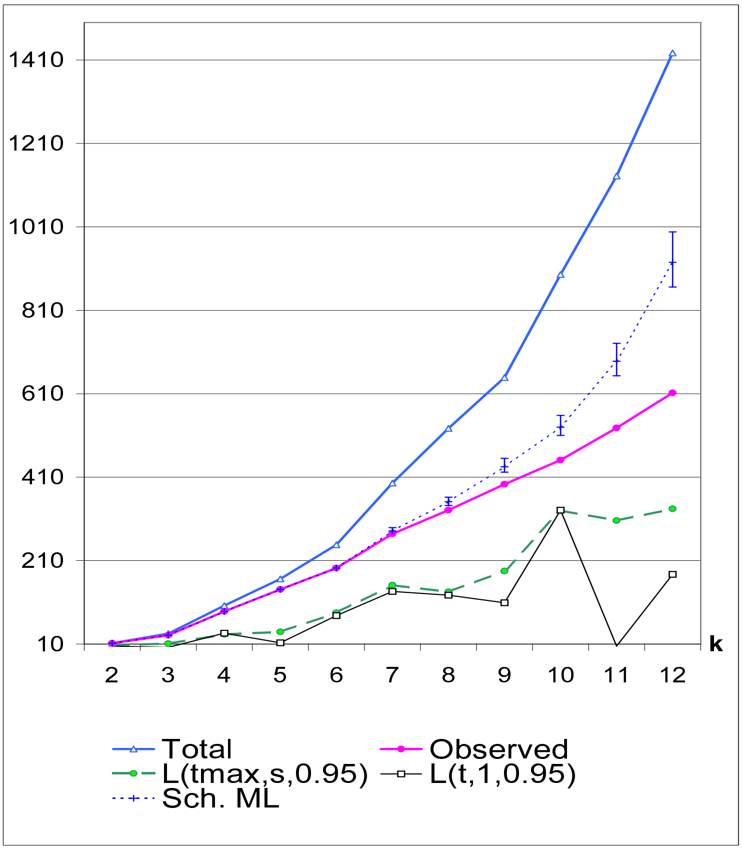

Fig. 1. Number of local optima in NK landscapes

The curve of true values of ν in this figure diverges significantly from the Schnabel census estimate, which clearly shows that the isotropic assumption is invalid for these NK landscapes. The lower bounds based on the first repetition time computed from our data go below the curve of number of observed optima. We shall see that in situations with much greater values of ν, lower bounds based on first repetition time may be more informative.

3.2 Low Autocorrelation Binary Sequences

The LABS problem consists in finding a vector $y \in \{-1, +1\}^N$ that minimizes $E = \sum_{j=1}^{N-1} c_j^2$, where c_j is the autocorrelation of the sequence $(y_1, ..., y_N)$ at lag j, i.e. $c_j = \sum_{i=1}^{N-j} y_i y_{i+j}$.

The graph in Fig. 2 shows the results for LABS problems analogously to Fig. 1, but here the sequence length N ranges from 7 to 29. For instances with $N > 24$ the complete enumeration of the solution space was computationally too expensive, so to illustrate the further behaviour of ν we also provide the non-parametric bootstrap estimate proposed in [10].

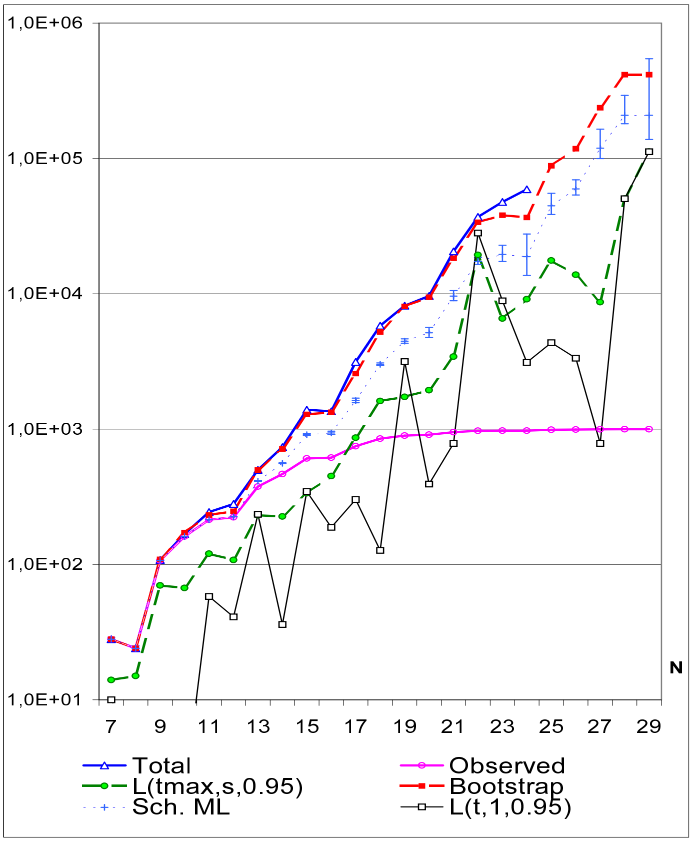

Fig. 2. Number of local optima in LABS problems

We see that, as in the case of NK landscapes, the lower bound $L(t_1, 1, \alpha)$ is less stable than $L(\max_{j=1,\ldots,s} t_j, s, \alpha)$, although quite naturally for large values of ν these estimates coincide.

3.3 Vertex Cover Problems

The computational results for the vertex cover problems are presented in Table 1. The VCP is a well-known NP-hard problem, formulated as follows. Let $G = (V, E)$ be a graph with a set of vertices V and the edge set E. A subset $C \subseteq V$ is called a vertex cover of G if every edge has at least one endpoint in C. The vertex cover problem is to find a vertex cover C^* of minimal cardinality.

Table 1. Estimates of the number of local optima in VCP

Name	$\lvert V\rvert$	k	$L(t_{\max}, s, \alpha)$	$L(t, 1, \alpha)$	Bootstrap	Sch. 2-sided
john8-4-4	70	4509	244	572	7145	5248-5398
brock200-2	200	972	5554	346	33761	12214-26212
mann a9	45	9406	41535	7824	$1.5{\cdot}10^5$	74808-87445
keller4	171	9805	$1.9{\cdot}10^5$	$1.4{\cdot}10^5$	$4.9{\cdot}10^5$	$2.2{\cdot}10^5$-$2.9{\cdot}10^5$
brock200-4	200	999	81200	81200	$8.3{\cdot}10^5$	$9.0{\cdot}10^4$-$2.0{\cdot}10^7$
john16-2-4	120	9939	$4.3{\cdot}10^5$	$1.6{\cdot}10^5$	$1.6{\cdot}10^6$	$6.4{\cdot}10^5$-$1.1{\cdot}10^6$
san200-0.72	200	9945	$8.5{\cdot}10^5$	$1.2{\cdot}10^6$	$1.8{\cdot}10^6$	$7.0{\cdot}10^5$-$1.2{\cdot}10^6$
san200-0.92	200	9995	$3.0{\cdot}10^6$	$4.1{\cdot}10^6$	$1.2{\cdot}10^7$	$4.3{\cdot}10^6$-$3.1{\cdot}10^7$
C125.9	125	9999	$1.5{\cdot}10^7$	$1.5{\cdot}10^7$	10^8	$9.0{\cdot}10^6$-$1.1{\cdot}10^9$
gn200p0.9-55	200	9999	$1.1{\cdot}10^6$	$1.1{\cdot}10^6$	10^8	$9.0{\cdot}10^6$-$1.1{\cdot}10^9$

In order to include this problem into the unconstrained optimization framework accepted throughout this paper we will assume that the solutions set is the set of all $|V|$-element binary strings where ones indicate the vertices belonging to the corresponding subset C. The infeasible solutions are penalized by summing their cardinality with the number of uncovered edges in E multiplied by $|V|$ (by uncovered we mean those edges with both endpoints outside the set C). In our study we take some widely used instances of maximum clique problem from the DIMACS benchmark set [17] and transform them into VCP instances using the standard reduction (see e.g. [18]).

It turned out that large-scale instances in this set are characterised by large values of ν, so in order to provide at least some repetitions in the visited local optima we limited our consideration to graphs with not more than 200 vertices. In the case $r = 1000$ even the smaller problems often yielded no repetitions at all, thus r was set equal to 10000 for all instances except for 'brock200-2' and 'brock200-4'. For the problems 'gn200p0.9-44', 'mann a27', 'san200-0.9-1' and 'san200-0.9-3' we still found $k = r = 10000$, so in this case we have $L(\max_{j=1,\ldots,s} t_j, s, 0.95) = L(t_1, s, 0.95) \approx 1.67 \cdot 10^6$. Analysis of these data indicates that the number of local optima in many VCP instances of similar dimensions depends heavily upon the particular structure inherent in the benchmark generator and its settings.

4 Conclusions and Further Work

Our aim in this work was to develop and test the applicability of statistical techniques for estimation of the number of local optima based on the assumption that their basins of attraction have equal size. The techniques developed do not employ any approximations, unlike their former versions in [7] and thus these tools should yield more precise estimates in general (although at expense of longer computations).

We have shown that the estimators based on the first repetition time are actually applicable for arbitrary distribution of basins of attraction size, when they are used for finding a lower bound on the number of local optima. A similar situation is conjectured for the Schnabel census procedure although we have no formal proof for it so far. A simple analysis indicates that the proposed techniques are able to yield an estimate that grows at most quadratically in the number of restarts. This is a serious limitation, and thus a subject for future research (for many problems in the worst case the number of local optima may grow exponentially in problem size [1]).

The computational experiments demonstrate the application of the new techniques to well-known benchmark problem instances of medium size. These results show that the lower bounds are always adequate, although the two-sided confidence intervals are often irrelevant if the isotropic assumption is invalid.

Future research might also deal with the questions of using the proposed techniques in stopping rules for local search heuristics with multiple restarts. In particular, the lower bounds could give the necessary termination conditions, and if all local optima have an equal chance to be found, sufficient conditions could be formulated as well. Also these techniques may be used for the classification of instances of an optimization problem in terms of their difficulty for local search algorithms when test problems are being generated. The open issues mentioned in this paper are also the subject of future research.

Acknowledgment

Part of this research was carried out when the first author was visiting the UK on a grant from the Royal Society. The authors thank Sergey Klokov for the helpful comments on this paper.

References

1. Tovey, C.A.: Local Improvement on Discrete Structures. In: Aarts, E., Lenstra, J. K. (eds.): Local Search in Combinatorial Optimization. John Wiley & Sons (1997) 57-90
2. Reeves, C.R.: The Crossover Landscape and the Hamming Landscape for Binary Search Spaces. To appear in Proc. of Foundations of Genetic Algorithms - VII. (2003)
3. Reeves C.R.: Fitness Landscapes and Evolutionary Algorithms. In: Fonlupt, C., Hao, J-K., Lutton, E., Ronald, E. and Schoenauer, M. (eds.): Artificial Evolution: 4th European Conference. Lecture Notes in Computer Science, Vol. 1829. Springer-Verlag, Berlin (2000) 3-20

4. Stadler, P.F.: Towards a Theory of Landscapes. In: Lopéz-Peña, R., Capovilla, R., García-Pelayo, R., Waelbroeck, H. and Zertuche, F. (eds.): Complex Systems and Binary Networks. Springer-Verlag, Berlin (1995) 77-163
5. Reeves, C.R.: Landscapes, Operators and Heuristic Search. Annals of Operations Research. **86** (1999) 473-490
6. Boese, K.D., Kahng, A.B., Muddu S.: A New Adaptive Multi-Start Technique for Combinatorial Global Optimizations. Operations Research Letters. **16** (1994) 101-113
7. Reeves, C.R.: Estimating the Number of Optima in a Landscape, Part I: Statistical Principles. Coventry University Technical Report SOR#01-03 (2001)
8. Garnier, J., Kallel, L.: How to Detect All Maxima of a Function? In: Proceedings of the Second EVONET Summer School on Theoretical Aspects of Evolutionary Computing (Anvers, 1999). Springer, Berlin (2001) 343-370
9. Reeves, C.R.: Estimating the Number of Optima in a Landscape, Part II: Experimental Investigations. Coventry University Technical Report SOR#01-04 (2001)
10. Eremeev, A.V., Reeves C.R.: Non-Parametric Estimation of Properties of Combinatorial Landscapes. In: Cagnoni, S., Gottlieb, J., Hart, E., Middendorf, M. and Raidl, G. (eds.): Applications of Evolutionary Computing: Proceedings of EvoWorkshops 2002. Lecture Notes in Computer Science, Vol. 2279. Springer-Verlag, Berlin Heidelberg (2002) 31-40
11. Seber G.A.F.: The Estimation of Animal Abundance. Charles Griffin, London (1982)
12. Mood, A. M., Graybill, F.A., Boes, D.C.: Introduction to the Theory of Statistics. 3rd edn. McGraw-Hill, New York (1973)
13. Johnson, N.L., Kotz, S.: Discrete Distributions. Wiley, New York (1969)
14. Liu, C.L.: Introduction to Combinatorial Mathematics. McGraw-Hill, New York (1968)
15. Kauffman, S.A.: Adaptation on Rugged Fitness Landscapes. In: Lectures in the Sciences of Complexity, Vol. I of SFI studies, Addison-Wesley (1989) 619-712
16. Golay, M.J. E.: Series for Low-Autocorrelation Binary Sequences. IEEE Trans. Inform. Theory. **23** (1977) 43-51
17. Johnson, D.S., Trick, M.A.: Introduction to the Second DIMACS Challenge: Cliques, Coloring, and Satisfiability. In: Johnson, D.S., Trick, M.A. (eds.): Cliques, Coloring, and Satisfiability. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 26. AMS (1996) 1-10
18. Garey, M.R., Johnson, D.S.: Computers and Intractability. A Guide to the Theory of NP-Completeness. W.H. Freeman and Company, San Francisco (1979)