ECONOMETRIC INSTITUTE

PROBABILISTIC ANALYSIS OF ALGORITHMS

A.H.G. RINNOOY KAN

REPORT 8538/A

ERASMUS UNIVERSITY ROTTERDAM - P.O. BOX 1738 - 3000 DR ROTTERDAM - THE NETHERLANDS

PROBABILISTIC ANALYSIS OF ALGORITHMS

by

A.H.G. Rinnooy Kan

Abstract

An introductory and selective review is presented of results obtained through a probabilistic analysis of combinatorial algorithms. The emphasis is on asymptotic characteristics of optimal solution values, and on the relative and absolute error analysis for simple heuristics.

1. INTRODUCTION

Suppose that two thieves meet on a regular basis to divide the proceeds of their joint effort. Each stolen object has a specific dollar value and has to be assigned to one of the two. For obvious reasons, they are interested in a quick and fair partitioning scheme.

In spite of its apparent simplicity, the above combinatorial problem is not easy to solve if we insist on an optimal solution, i.e., one in which the difference between the values assigned to each thief is as small as possible. As is the case with most practical problems, this problem too is known to belong to the class of NP-complete problems. This implies that any optimization method for its solution could be expected to perform very poorly on some occasions: more formally, its worst case running time is likely to grow exponentially with problem size.

Hence, in choosing a solution scheme, the thieves will be forced into an unpleasant trade-off between two features of algorithmic quality: the computational effort (the smaller running time, the better) on one hand and the computational result (the smaller deviation from optimality, the better) on the other hand. Complexity theory indicates that we cannot insist on a simultaneous absolute guarantee for both, i.e., on a fast (polynomially bounded) running time as well as a zero deviation from the optimal solution value.

One possible way out of this dilemma is to change the perspective on the analysis by no longer demanding an absolute guarantee. For practical purposes an algorithm that, with respect to both effort and result, does well in the

majority of cases or even on average might be perfectly acceptable.

Probability theory provides the natural setting for such an analysis of algorithms. This analysis starts from a specification of what an average problem instance would look like, in terms of a probability distribution over the class of all instances. The running time and the solution value of a particular algorithm are then considered as random variables, whose behaviour can be studied and evaluated. This approach can therefore be viewed as the analytical counterpart to the familiar experimental analysis in which an algorithm is tried out on a set of supposedly representative test problems and evaluated statistically. Here we obtain the rigor of mathematical analysis at the expense of a certain naivete, in that only relatively straightforward solution methods can be analyzed probabilistically in full detail.

Although the probabilistic analysis of algorithms has only recently become an active research area, it has already generated an impressive number of publications. A concise survey of this area would require the prior introduction of many techniques from probability theory and could hardly do justice to the diversity of ideas and approaches that one finds in the literature. Fortunately, a recent annotated bibliography [Karp et al. 1984] provides an up to date survey of the available articles and publications. In view of the existence of this source of detailed information, no attempt at completeness will be made in what follows below. Rather, the nature of the analysis and of the results will be illustrated by some typical examples.

In Section 2, we consider the problem of the two thieves in more detail. It is, of course, none other than the well known PARTITION problem in which one seeks to minimize the size of the largest share. In Section 3, we review some representative results that are known for problems defined in the Euclidean plane. In Section 4, we examine the fertile area of optimization problems defined on graphs and networks. In Section 5, we indicate some possibilities for future work in this very lively research area.

We conclude this introduction by a short digression on modes of stochastic convergence, clearly an essential concept if we want to analyze the notion of a random variable such as the error of a heuristic going to 0 with increasing problem size.

Almost sure convergence of a sequence of random variables y_n to a constant c by definition means that $\Pr\{\lim_{n\to\infty} y_n = c\} = 1$; it is a strong form of stochastic convergence and implies the weaker convergence in probability,

which stands for $\lim_{n\to\infty} Pr\{|\underline{y}_n - c| > \varepsilon\} = 0$ for all $\varepsilon > 0$. The reverse implication holds under the additional assumption that, for all $\varepsilon > 0$,

$$\sum_{j=1}^{\infty} Pr\{|\underline{y}_n - c| > \varepsilon\} < \infty . \qquad (1)$$

Similarly, convergence of $\underline{y}_n$ to c <u>in expectation</u>, i.e.

$$\lim_{n\to\infty} |E(\underline{y}_n) - c| = 0 \qquad (2)$$

also implies convergence in probability, with the reverse implication holding under additional boundedness assumptions on $\underline{y}_n$.

2. THE PARTITION PROBLEM

Perhaps the simplest possible way to solve the PARTITION problem of the two thieves is to allow each thief to choose a particular item in turn until they have all been assigned. If the j-th item has value a_j ($j = 1, \ldots, n$), then this amounts to ordering the items according to decreasing a_j values $a^{(n)} \geq a^{(n-1)} \geq \ldots \geq a^{(1)}$; one thief receives $a^{(n)} + a^{(n-2)} + \ldots$, the other $a^{(n-1)} + a^{(n-3)} + \ldots$

This is clearly a fast heuristic method that may, however, produce a very inequitable result: in the worst case, the first thief may receive up to 50 percent more than the optimal partition would grant him. (Take $a_1 = 2$, $a_{2i} = a_{2i+1} = 2^{-1}$ ($i \geq 1$.) How about its average case behaviour? To answer that question, we specify a probability distribution over all problem instances by assuming that robberies are so frequent and haphazard that the a_j can be thought of as independent draws from a uniform distribution on, say, [0,1]. (Actually, many of the results mentioned below hold under much more general assumptions, though independence is always required.)

Under the uniform assumption, the optimal solution value $\underline{Z}_n^{OPT}$ of the partition problem (i.e., the smallest possible size of the larger share) turns out to be <u>almost surely</u> (a.s.) <u>asymptotic to</u> the lower bound $(\Sigma_{j=1}^{n} \underline{a}_j)/2$:

$$\frac{\underline{Z}_n^{OPT}}{(\Sigma_{j=1}^{n} \underline{a}_j)/2} \to 1 \qquad \text{(a.s.)} \qquad (3)$$

This result provides a first example of <u>asymptotic probabilistic value analysis</u>: for n large enough, the optimal solution value Z_n^{OPT} can be guessed

with increasing (relative) accuracy. What about the heuristic solution value Z_n^H, i.e. the size of the larger share under the heuristic scheme proposed above? We know that

$$\underline{Z}_n^H = \frac{1}{2} \sum_{j=1}^{n} \underline{a}_j + \frac{1}{2}[(\underline{a}^{(1)} - \underline{a}^{(2)}) + (\underline{a}^{(3)} - \underline{a}^{(4)}) + \ldots], \qquad (4)$$

and it is not hard to show that the difference between $\underline{Z}_n^H$ and $(\sum_{j=1}^{n} \underline{a}_j)/2$ a.s. converges to 1/4, so that also

$$\frac{\underline{Z}_n^H}{(\sum_{j=1}^{n} \underline{a}_j)/2} \to 1 \qquad \text{(a.s.)} \qquad (5)$$

and hence

$$\frac{\underline{Z}_n^H}{\underline{Z}_n^{OPT}} \to 1 \qquad \text{(a.s.)} \qquad (6)$$

This implies that the heuristic is <u>asymptotically optimal</u>: its <u>relative error</u> (i.e., its percentage deviation from the optimum) $(\underline{Z}_n^H - \underline{Z}_n^{OPT})/\underline{Z}_n^{OPT}$ a.s. goes to 0. Hence, a probabilistic analysis leads to a much more optimistic conclusion than a worst case oriented one.

What about the <u>absolute error</u> $\underline{Z}_n^H - \underline{Z}_n^{OPT}$ or the <u>absolute difference</u> between the two shares? Neither of these two quantities goes to 0 for the above heuristic, so there is room for improvement. A slightly more sophisticated scheme would be to allow each thief in turn to select items until the value of his share exceeds the value of the current share of his colleague. Even from a worst case point of view, this is a much more reasonable approach: the larger share can never exceed its smallest possible size by more than 16 2/3 percent [Graham 1969]. (For a worst case example, take e.g., $a_1 = a_2 = 3$, $a_3 = a_4 = a_5 = 2$.) In a probabilistic sense, the difference is even more impressive. Of course, the relative error again goes to 0, but the absolute difference between the shares $\underline{d}_n^H$ (and hence the absolute error) also satisfies

$$\underline{d}_n^H \to 0 \qquad \text{(a.s.)} \qquad (7)$$

[Frenk and Rinnooy Kan 1983]. To prove this result, one observes that

$$\underline{d}_n^H \leq \max\{\underline{d}_{n-1}^H - \underline{a}^{(1)}, \underline{a}^{(1)}\} \qquad (8)$$

which, after repeated application, yields that

$$\begin{aligned} \underline{d}_n^H &\leq \max_{1 \leq k \leq n} \{\underline{a}^{(k)} - \textstyle\sum_{j=1}^{k-1} \underline{a}^{(j)}\} \\ &\leq \underline{a}^{([\delta n])} + \max\{\underline{a}^{(n)} - \textstyle\sum_{j=1}^{[\delta n]} \underline{a}^{(j)}, 0\} \end{aligned} \tag{9}$$

for all $\delta > 0$. The first term converges a.s. to δ and can therefore be made arbitrarily small; for any fixed δ, the second term converges a.s. to 0 since $\sum_{j=1}^{[\delta n]} a^{(j)} = O(n)$ and $a^{(n)}$ is $o(n)$ for every distribution with finite first moment.

The two results presented so far demonstrate the importance of the theory of order statistics for the analysis of heuristics that involve the sorting of numbers; priority rules generally fall into this class.

Can we do still better? One weakness of a result such as (7) is its asymptotic nature, i.e. its validity only for sufficiently large values of n. At the very least, one would like to know the rate at which $\underline{d}_n^H$ converges to 0. It can be shown [Frenk and Rinnooy Kan 1984] that

$$\limsup_{n \to \infty} \frac{\underline{d}_n^H}{\log \log n/n} < \infty \qquad \text{(a.s.)}, \tag{10}$$

and a simple argument shows that the rate of convergence for this heuristic has to be at least $1/n$ in expectation [Karp, 1984]. This rate is good but not quite good enough: indeed, it is also known that $\underline{d}_n^{OPT}$, the smallest possible absolute difference, satisfies

$$\limsup_{n \to \infty} \frac{\underline{d}_n^{OPT}}{n 2^{-n}} < \infty \qquad \text{(a.s.)} \tag{11}$$

[Karmarkar et al. 1984]. Hence, the exponential effort that may be required for the computation of the optimal partition is at least rewarded by an exponential decrease to 0 of the difference between the two shares. Can this also be achieved a.s. in polynomial time?

The answer to this question is unknown, but the previous heuristic can again be improved upon by essentially assigning two items at a time and compensating for their difference. More precisely, in the first iteration $a^{(n)}$ would be assigned to one thief, and $a^{(n-1)}$ to the other. The two terms would then be replaced by a single item of value $a^{(n)} - a^{(n-1)}$ and the process would be

repeated on the new set of n-1 items until only one item remains; its value represents the difference between the two shares. A simple backtracking procedure establishes the partition in terms of the original items.

In the worst case, this heuristic method is not better than the previous one. The probabilistic analysis of its performance is quite difficult: as on so many other occasions, each step in the algorithm conditions the probability distribution encountered in the succeeding steps in a complicated fashion. Since the difference of two independent uniformly distributed random variables follows a triangular distribution, there is no distributional invariance throughout the steps of this method and yet such invariance is an essential prerequisite for a successful analysis. One way to overcome this kind of obstacle is to change the algorithm so that (with high probability) the value produced will not be affected but its modified behaviour can be analyzed rigorously. In this particular case, the new versions of the algorithm works roughly as follows. In iteration m, it deals with numbers in an interval $[0, \alpha_m]$, which is first partitioned into subintervals of size α_{m+1}. The original differencing method is then applied to random pairs of numbers taken from the set S_i found in the i-th subinterval ($i = 1, \ldots, \alpha_m/\alpha_{m+1}$), until a set $S_i' \subset [0, \alpha_{m+1}]$ is obtained and at most one of the numbers originally in S_i remains. The original differencing method is then used on the set of remaining numbers to reduce them to a single number in $[0, \alpha_{m+1}]$.

Now, before the method can be applied recursively to $[0, \alpha_{m+1}]$, the issue of distributional invariance has to be addressed. To do so, we assume (inductively) that $S = \cup S_i$ was divided into two subsets, a subset G of 'good' points that can be assumed to come from a uniform distribution over $[0, \alpha_m]$ and a subset B of 'bad' points. During the application of the differencing method to the sets S_i the numbers entering S_i' are labeled good only if they are obtained as the difference of two good numbers in S_i. As a result, we know that the good numbers in $S' = \cup S_i'$ follow a distribution related in the triangular one. In a final step, a subset of these numbers is relabeled 'bad' so that the remaining good ones are again uniformly distributed over $[0, \alpha_{m+1}]$. All that remains to be shown is that for appropriate choices of α_m, enough good numbers remain for the method to reach the number of iterations that is required for a good result.

Through this approach, it could be established in [Karwarkar & Karp 1982] that

$$\limsup_{n\to\infty} \frac{\underline{d}_n^H}{n^{-\log n}} < \infty \text{ (a.s.)} \qquad (12)$$

Thus, in O(n log n) time this method guarantees a rate of convergence that is superpolynomial, yet subexponential. It is tempting to conjecture that this is best possible for a polynomial time heuristic!

We have dealt with this simple example in some detail, since it exhibits many of the ingredients typically encountered in a probabilistic analysis:

- a combinatorial problem that may be difficult to solve to optimality (the PARTITION problem is NP-complete);
- a probability distribution over all problem instances, that generates the problem data as realizations of independent and identically distributed (i.i.d.) random variables (the $\underline{a}_j$ are independent and uniform over [0,1]);
- a probabilistic value analysis that yields a description of the asymptotic optimal solution value as a simple function of the problem data ($\underline{Z}_n^{OPT}/(n/4) \to 1$ (a.s.));
- simple, fast heuristics whose relative error or absolute error may decrease to 0 in some stochastic sense or may be otherwise well behaved;
- a rate of convergence analysis that allows further differentiation among the heuristics.

The state of the art for a particular problem class can conveniently be monitored by means of the above concepts. Consider, for example, the MULTIKNAPSACK problem $\max\{\Sigma_j c_j x_j \mid \Sigma_j a_{ij} x_j \leq b_i \ (i = 1, \ldots, m), x_j \in \{0,1\} \ (j = 1, \ldots, n)\}$, which is a generalization of PARTITION (take $m = 1$, $c_j = a_{1j}$ $(j = 1, \ldots, n)$, $b_1 = (\sum_{j=1}^{n} a_{1j})/2$).

Let us assume that the $\underline{c}_j$ and $\underline{a}_{ij}$ are i.i.d. uniform on [0,1] and that $b_i = n\beta_i$ is constant. As above, we are interested in the optimal solution value as a function of $\beta = (\beta_1, \ldots, \beta_m)$ and in heuristics whose error vanishes asymptotically with high probability.

The analysis of this problem in [Meanti et al. 1984] is of interest in that exploits the close relationship (in probability) between certain difficult nonconvex combinatorial optimization problems such as MULTIKNAPSACK and their convex LP relaxations obtained by replacing the constraints $x_j \in \{0,1\}$ by $0 \leq x_j \leq 1$. It is easy to see that the absolute difference between the

solution values of these two problems is bounded by m, so that the relative error that we make by focusing on the LP relaxation goes to 0. But the LP relaxation (or, rather, its dual) is much easier to analyze: its value is given by

$$\min_{\lambda} \underline{L}_n(\lambda) \tag{13}$$

with

$$\underline{L}_n(\lambda) = \Sigma_{i=1}^{m} \lambda_i b_i + \max\{\Sigma_j (\underline{c}_j - \Sigma_{i=1}^{m} \lambda_i \underline{a}_{ij}) x_j \mid 0 \leq x_j \leq 1\} \tag{14}$$

where the maximization problem in (14) is solved by setting

$$\underline{x}_j(\lambda) = \begin{cases} 1 & \text{if } \underline{c}_j - \Sigma_{i=1}^{m} \lambda_i \underline{a}_{ij} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Results from convex analysis can then be used to establish that the optimal solution value Z_n^{OPT} satisfies

$$\frac{\underline{Z}_n^{OPT}}{n} \to \min_{\lambda} L(\lambda) \quad \text{(a.s.)}, \tag{16}$$

where

$$L(\lambda) = \Sigma_{i=1}^{m} \lambda_i \beta_i + \frac{1}{n} E\underline{c}^T \underline{x}(\lambda) - \frac{1}{n} \Sigma_{i=1}^{m} \lambda_i E\underline{a}_i^T \underline{x}(\lambda) \tag{17}$$

is a convex, twice differentiable function with a unique minimum that can actually be computed in closed form in some simple cases (e.g., for m=1).

Not surprisingly, successful heuristics for this problem also have a strong LP flavor (cf. [Frieze & Clarke 1981]). A natural one to consider is the generalized greedy heuristic in which x_j's are set equal to 1 in order of non-increasing ratio's $c_j / \Sigma_{i=1}^{m} \lambda_i a_{ij}$. If the λ_i are chosen to be equal to the values minimizing the right hand side of (16), then the relative error of this greedy method goes to 0 a.s. A heuristic whose absolute error vanishes asymptotically is not known, however, and further analysis of the model reveals puzzling differences between the minimization and maximization version of MULTIKNAPSACK that still have to be resolved. None the less, the probabilistic analysis of the model yields surprisingly high returns.

What are the strong and the weak points of the approach? On one hand, the algorithmic insights nicely complement the more traditional worst case analysis, with an emphasis that is not so much on an exact guarantee that a certain running time or a certain error will not be exceeded, as on an explanation why the algorithm may perform so much better in practice than the worst case analysis would seem to suggest. On the other hand, the results are obtained under probabilistic assumptions that can always be questioned, and are usually only valid for 'sufficiently large' problem sizes, with little indication of how large these might have to be. Rate of convergence results somewhat compensate for this latter deficiency.

Apart from its contribution to error analysis, a probabilistic value analysis may, however, yield additional benifits. An estimate of the optimal solution value can, for instance, be used in a branch and bound procedure to replace weak upper or lower bounds, at the possible expense of sacrificing optimality but at the gain of a significant improvement in running time [Derigs 1984]. Or it may be used in a two stage stochastic programming problem, where the first stage decision determines the value of some parameters of the second stage problem. A probabilistic value analysis may then reveal how the second stage solution value depends on these parameters so that they can be given optimal values in the first stage [Stougie 1985].

As a fist step towards a classification of results in probabilistic analysis, let us note that the PARTITION problem and its generalization, the MULTIKNAPSACK problem, both have a problem input that consists of numbers. The probabilistic analysis of algorithms for these problems usually assumes that these are independently generated from a fixed distribution. To close this section, we briefly review some typical results that were obtained for other problems in this category.

The MULTIMACHINE SCHEDULING problem is an appropriate starting point, since it can be viewed as yet another generalization of PARTITION: it is the problem to distribute jobs with processing times $a_1, \ldots, a_n$ among m identical machines so as to minimize the maximal sum of processing times assigned to any machine (the makespan); the PARTITION problem corresponds to the case that $m=2$.

The second heuristic proposed for PARTITION can be viewed as a special case of the LPT (Largest Processing Time) heuristic for MULTIMACHINE SCHEDULING, in which jobs are assigned to the earliest available machine in order of nonincreasing a_j. If the a_j are i.i.d. uniform on [0,1], the proof technique based on order statistics can again be applied to show that the absolute error of the LPT heuristic goes to 0 a.s.; the optimal solution value is asymptotic to $\frac{1}{2}$ n/m. The order in which jobs are assigned to machines turns out to be really essential: an arbitrary list scheduling heuristic will have a relative error that goes to 0 a.s. but an absolute error that does not. The Karp-Karmarkar heuristic can be extended to the case of arbitrary m to improve the rate of convergence to optimality from 1/n for the LPT rule to $n^{-\log n}$. We refer to [Coffman et al. 1983] for additional references.

The famous BIN PACKING problem is in a sense dual to MULTIMACHINE SCHEDULING: here, the makespan is fixed (say, equal to 1) and the objective now is to find the minimum number of bins (machines) into which the items (jobs) of size a_1, ..., a_n can be packed. The probabilistic analysis of this problem is again usually carried out under the assumption that the jobs are i.i.d. uniform on [0,1]. It has yielded many beautiful results. To give one example, consider the heuristic that inspects the items in order of decreasing a_j and matches each item a_k with the largest unassigned item a_ℓ satisfying $a_k + a_\ell \leq 1$. To analyze the performance of this heuristic, we consider the set of all a_j on the interval [0,1] and replace each item a_k larger than 1/2 by $1 - a_k$, marking it with a '+'; the items smaller than 1/2 get a '-' sign. The heuristic now amounts to pairing each '+' with the largest '-' to its left; the number of poorly filled bins (i.e., bins with only one item) is related to the excess of '+'s over '-'s. But the sequence of '+'s and '-'s can be viewed as generated by flips of a coin, and results from the theory of random walks can be invoked to show that the expected total number of bins used is $n/2 + O(\sqrt{n})$; since the expected optimal number of bins is known to be $n/2 + \Omega(\sqrt{n})$, the relative error of this heuristic goes to 0 in expectation [Frederickson 1980; Knodel 1981].

This analysis reveals an interesting connection between matching problems on the interval [0, 1/2] and bin packing methods. In a similar vein, a connection can be established between on line bin packing methods (i.e., methods in which items arrive in arbitrary order and have to be irrevocably assigned to a bin right away; the previous method is off line) and matching problems on a

square, where one of the dimensions corresponds to item size and the other dimension represents time, i.e., it indicates the order in which the items arrive for packing. In this case, feasibility dictates that each point can only be matched to a right upward neighbour; this matching problem was studied in a different context in [Karp et al. 1985], and the clever extension and refinement of these results to on-line bin packing can be found in [Shor 1984]. Further references can be found in the useful survey [Coffman et al. 1983] that was already quoted above.

A final number problem that deserves to be mentioned is LINEAR PROGRAMMING, not because it is a hard combinatorial problem (the Khachian method solves it in polynomial time) but because probabilistic analysis played such a vital role in understanding the excellent average performance of the simplex method. The history of the analysis illustrates the importance of an appropriate probabilistic model: ultimately, the concept of a random polytope as being generated by m fixed hyperplanes in $\mathbb{R}^n$ and m coin flips to determine the direction of the corresponding inequalities reduced the computation of the average number of simplex pivots to a combinatorial counting argument. Within this model, various simplex variants admit of a quadratic upper bound on the expected number of iterations (including those in Phase I), which takes us very close to the behaviour observed in practice [Haimovich 1983; Adler et al. 1983; Adler & Megiddo 1983; Todd 1983].

In the next section, we turn to problems with a geometric flavor, whose probabilistic analysis involves random sets of points in the Euclidean plane.

3. EUCLIDEAN PROBLEMS

In this section we are concerned with problems whose input includes a set of n points in the Euclidean plane. The most famous problem of this type is surely the TRAVELLING SALESMAN problem, which is to find the shortest tour connecting the n points. It has a venerable history, of which the probabilistic analysis forms one of the most recent chapters (see [Lawler et al. 1975]).

To carry out such an analysis, one usually assumes the points to be uniformly distributed over a fixed region, e.g., the unit (1 × 1) square. Under such an assumption, it is not difficult to arrive at an intuitive probabilistic value

analysis for the TRAVELLING SALESMAN problem. For n large, the length of an optimal tour through a 2 × 2 square with 4n points will be approximately 4 times as large as the optimal tour in a unit square with n points. Scaling back the 2 × 2 square to a unit one, we conclude that the optimal tour length Z_n^{OPT} is likely to grow proportionally to $\sqrt{n}$. Indeed, a heuristic from [Few 1955] shows that its value is bounded deterministically from above by $\sqrt{(2n)}$, and it is not hard to show that there exists a positive constant c such that it is a.s. bounded by $c\sqrt{n}$ from below.

The actual convergence argument is much more difficult and was first provided in [Beardwood et al. 1959]: as expected,

$$\frac{Z_n^{OPT}}{\sqrt{n}} \to \beta \quad \text{(a.s.)} \tag{18}$$

where β is a constant that has been estimated empirically to be about 0.765.

The proof of this result involves a technique useful in a broader context: rather than viewing the problem on a fixed set of n points, uniformly distributed on a square, they are assumed to be generated by a Poisson process of intensity 1. The advantage is that point sets in disjoint regions are now fully independent; the disadvantage that an expression for, say, the expected routelength F(t) in a square $[0,t]\times[0,t]$ has to be converted back into a result for the expected length EZ_n^{OPT} of a tour through n points, using the relation

$$F(t) = \Sigma_{n=1}^{\infty} tEZ_n^{OPT} e^{-t^2} \frac{(t^2)^n}{n!} . \tag{19}$$

Certain Tauberian theorems allow one to do precisely that.

In the specific case of the TRAVELING SALESMAN problem, F(t) is computed through the use of a heuristic that embodies the intuitive insight mentioned above. The square $[0,t]\times[0,t]$ is divided into m^2 equal size subsquares; a tour in each subsquare has expected length F(t/m). These tours can be linked to form a feasible solution to the original problem by adding segments of length O(tm), and hence

$$F(t) \leq m^2 F(t/m) + O(tm) \tag{20}$$

It is not hard to see that this imples that $F(t)/t^2$ converges to a constant β and to conclude from (19) that $EZ_n^{OPT} = \beta\sqrt{n}$. The argument is completed by proving the variance of Z_n^{OPT} to be O(1) through the Efron-Stein inequality; Chebyshev's inequality and the Borel-Cantelli lemma then yield (18) [Karp & Steele 1985].

Given the result of the probabilistic value analysis, it now becomes attractive to search for heuristics whose absolute error is $o(\sqrt{n})$; their relative error then goes to 0 almost surely. As in the case of other Euclidean problems, partitioning heuristics do precisely that. Generally, in these heuristics the square is appropriately partitioned into subregions (e.g. rectangles), subproblems defined by the points in each subregion are analyzed separately, and a feasible solution to the problem as a whole is composed out of the separate results.

For the TRAVELLING SALESMAN problem, one partitioning approach is to execute an alternating sequence of horizontal and vertical cuts through the point with current median vertical and horizontal coordinate respectively, until the resulting rectangles contain no more than $\sqrt{(\log n)}$ points. Each of these subproblems is solved to optimality by some enumerative technique (say, dynamic programming, which takes $O(n^{\varepsilon})$ time per rectangle, and hence $O(n^{1+\varepsilon})$ time overall, for all $\varepsilon > 0$). The resulting tours define a connected graph with even degree at each point; the Euler walk that visits each edge of this graph can be converted into a tour of no greater length by eliminating multiple visits to one point. The difference between the length of this tour and the optimal one can be shown to be of the same order as the total perimeter of the rectangles generated, which is easily seen to be $o(\sqrt{n})$ in this case. Thus, the relative error of the heuristic goes to 0 a.s. [Karp 1977, Karp and Steele 1985].

Not much is known about the rate of convergence to optimality of this heuristic, nor is any heuristic known whose absolute error goes to 0 asymptotically. The partitioning approach, however, has been generalized to a variety of other Euclidean problems. Consider, for example, the CAPACITATED ROUTING problem, where the points have to be equally distributed among salesmen, each of whom can visit no more than q customers before returning to

a common depot. If we assume that the i-th customer is at distance r_i from the depot ($\underline{r}_i$ i.i.d., with finite first moment $\underline{E}r$), then it is not hard to prove that the optimal solution value satisfies

$$Z_n^{OPT} \geq \max\{T_n^{OPT}, 2\frac{\Sigma_{i=1}^{r} r_i}{q}\} \qquad (21)$$

where T_n^{OPT} is the length of a single travelling salesman tour through the n customers. A tour partitioning heuristic, in which T_n is optimally divided into [n/q] consecutive segments containing q customers each, can be shown to yield a value no more than the sum of the two terms on the right hand side of (21), so that - in view of (18) - $\underline{Z}_n^{OPT}$ is a.s. asymptotic to $2n\underline{E}r/q$. As a byproduct of this result, one obtains that the tour partitioning heuristic (of the '<u>route first, cluster second</u>') type has relative error going to 0 a.s.. A similar result holds for certain region partitioning heuristics ('<u>cluster first, route second</u>'): as indicated above, their absolute error is dominated by the total perimeter of the subregions generated, which in this case is $o(\sqrt{(n/q)})$ and hence vanishes asymptotically relative to $\underline{Z}_n^{OPT}$.

All these results presuppose that q is a constant that does not grow with n; if it does, than the results hold as long as $q = o(\sqrt{n})$. Above this threshold, $\underline{Z}_n^{OPT}$ behaves as in (18), since at that point the total cost of moving among the groups of customers, T_n, starts to dominate the total cost $2\Sigma_{i=1}^{n} r_i/q$ of reaching these groups from the depot (cf. (21)).

The ubiquitous presence of partitioning techniques in probabilistic Euclidean analysis points in the direction of a common generalization, which was indeed provided in [Steele 1981]. There, (18) is generalized to problems of arbitrary dimension, provided that the objective function is <u>subadditive</u> on sets of points (and that a few other technical conditions are satisfied). The TRAVELLING SALESMAN problem is one example, the MATCHING problem (find the segments linking each point to a neighbour, of minimum total length) is another one: see [Papadimitriou 1978] for a probabilistic value analysis and a heuristic that together establish the optimal solution value to be a.s. asymptotic to $\varepsilon\sqrt{n}$, with $\varepsilon \in [0.25, 0.40106]$.

We close this section by discussing an interesting class of problems that cannot quite be handled by Steele's techniques, i.e. LOCATION problems, in

which k depots have to be located so as to minimize the average distance between each point and its closest depot (the k-MEDIAN problem) or the maximum of these distances (the k-CENTER problem). The probabilistic value analysis for both problems lead to surprisingly similar results, provided that $k = O(n/\log n)$ (the case $k = \alpha n$ is partially open; see [Hochbaum & Steele 1981]). Both optimal solution values are asymptotically proportional to $1/\sqrt{k}$, albeit for different constants of proportionality.

The analysis leading to these results is of special interest, since it relies heavily on the similarity between the original (discrete) problem for large n and the continuous problem in which customer demand is not concentrated in a finite set of points but spread uniformly and continuously over the entire region. A simple partitioning heuristic in which a depot is located in each of k nonempty subsquares of size $1/\sqrt{k}$ by $1/\sqrt{k}$ each already provides an $O(1/\sqrt{k})$ upper bound on both optimal solution values. An asymptotically optimal heuristic, however, is only obtained by partitioning the region in regular hexagons (the honeycomb heuristic), with the constants of proportionality being determined by the optimal solution value of the continuous problem with k=1 over one such hexagon. This heuristic actually solves the continuous problem to optimality, and a detailed error analysis shows that, for n sufficiently large, its relative error in the discrete case becomes vanishingly small [Haimovich 1984, Zemel 1984].

4. GRAPHS AND NETWORKS

We now turn to the rich area of combinatorial optimization problems defined on graphs and networks. One of the reasons for the wide variety of probabilistic results for this class of problems is the existence of a substantial theory dealing with random graphs. There are two popular definitions of this concept: $G_{n,p}$ is defined to be the (undirected) graph for n vertices for which each of the $\frac{1}{2}n(n-1)$ edges occurs independently with equal probability p; G_n^N is defined by assuming that each of the $\binom{n(n-1)/2}{N}$ undirected graphs on n vertices occurs with equal probability (see [Bollobas 1985] for a survey of the theory). Especially for structural graph optimization problems, in which we are interested in graph properties that depend only on the node-edge incidence structure, random graphs provide a natural probability distribution over the set of all problem intances of size n.

Continuing in the spirit of the previous two sections, we again refer to [Karp et al. 1985] for a list of recent references in this area and review only a few typical probabilistic analyses of heuristics for NP-complete problems, whose expected performance compares favorably with the limits set by worst case analysis. In doing so, we (reluctantly) exclude many beautiful results on problems of CONNECTIVITY and MATCHING that can be solved in worst case polynomial time.

A typical example of a difficult structural problem is the CLIQUE problem of finding a complete subgraph of G that has maximal size $\omega(G)$. To carry out a probabilistic value analysis of $\omega(G_{n,p})$ for fixed p, we observe that the expected number of cliques of size k in such a graph is equal to $\binom{n}{k}p^{k(k-1)/2}$. We could expect the maximal clique of size k to occur when the expected number is approximately equal to 1, i.e., when (from Stirling's approximation of k!)

$$\frac{1}{\sqrt{2\pi k}}\left(\frac{nep^{(k-1)/2}}{k}\right)^k \approx 1. \tag{14}$$

The left hand side of (14) decreases very rapidly as k increases and passes through the value 1 where

$$\frac{nep^{(k-1)/2}}{k} = 1 \tag{15}$$

i.e. when

$$k = 2\log_{1/p}n + 2\log_{1/p}e - 2\log_{1/p}k + 1 \tag{16}$$

so that, approximately, $k \approx k(n,p)$ with

$$k(n,p) = 2\log_{1/p}n - 2\log_{1/p}\log_{1/p}n + 2\log_{1/p}(e/2) + 1. \tag{17}$$

This estimate turns out to be very sharp indeed. In [Matula 1976], it is proved that, for all $\varepsilon > 0$,

$$\lim_{n\to\infty}\Pr\{[k(n,p) - \varepsilon] \leq Z_n^{OPT} = \omega(G_{n,p}) \leq [k(n,p) + \varepsilon]\} = 1. \tag{18}$$

so that, for large enough n, the size of the largest clique can be predicted to be one of two consecutive integers with high probability of success.

This precise probabilistic value analysis again encourages the search for a fast heuristic whose absolute error compares favorably to $2 \log_{1/p} n$. Consider, for instance, the sequential greedy algorithm, which considers the vertices of G in arbitrary order and adds a vertex to the current clique if it is adjacent to all its members. For an analysis of the performance of this method, one observes that the expected number of trials to increase the clique size from j to j+1 is $1/p^j$, so that we might guess the ultimate clique size Z^H to satisfy

$$\sum_{j=0}^{Z^H-1} 1/p^j = \frac{1 - 1/p^{Z^H}}{1 - 1/p} = n \qquad (19)$$

i.e.,

$$Z^H \approx \log_{1/p} n. \qquad (20)$$

A more precise analysis shows that, indeed, this greedy approach a.s. yields a clique of size $(1/2 - \varepsilon)\underline{Z}_n^{OPT}$ [Grimmett and McDiarmid 1975]. Thus, the relative error does not go to 0, but is almost surely close to 50 percent. (There is by the way, no known polynomial time heuristic with any constant worst case bound on the relative error.)

The above result has immediate implications for the problem to find the INDEPENDENT SET in G of maximal size; it coincides with the maximal size clique in the complement of G. Again, the sequential greedy approach, which picks up each successive vertex that is not adjacent to any member of the current independent set, produces an independent set whose size is a.s. close to 50 percent of the optimal value. The COLORING problem, which is to partition the vertices of G into the smallest possible number $\chi(G)$ of independent sets, is much harder to analyze: the asymptotic optimal solution value $\underline{Z}_n^{OPT} = \chi(G_{n,p})$ is known for $p = 1/2$ [Korzanov 1980], though that (Russian) announcement has not been verified. The heuristic method, which greedily finds an independent set as above, deletes it and repeats on the remaining graph does poorly [McDiarmid 1979] but good enough to get within a factor of $(2 + \varepsilon)$ a.s. [Grimmett and McDiarmid 1975].

The other class of structural graph problems for which probabilistic analysis has been successful is the HAMILTONIAN CIRCUIT problem of searching for a

simple cycle containing all vertices. The emphasis here is more on conditions under which such a cycle exists (e.g., a.s. in G_n^N when $N = \frac{1}{2} n \log n + \frac{1}{2} n \log\log n + cn$ [Komlos and Szemeredi 1983]) and less on the development of fast heuristics that, with high probability, would be successful in finding such a cycle if it existed. However, the heuristic principle of extension and rotation has been applied to this class of problems with considerable success [Posa 1976, Angluin & Valiant 1979]. The general idea is as follows. Given a path of vertices $(v_0, \ldots, v_k)$, one of the neigbours of v_k, say w, sampled randomly and the edge $\{v_k, w\}$ is added to the path. If w is not in the path, it is adjoined and the method is applied to w. If $w = v_\ell$ $(0 \leq \ell \leq k-1)$, then the edge $\{v_\ell, v_{\ell+1}\}$ is removed from the path and the method is applied to $v_{\ell+1}$. If N exceeds the threshold by a sufficient amount (e.g., $N = cn \log n$, for large enough c) this method will be successful with high probability.

We now turn briefly to number problems on weighed graphs, i.e., graphs with weights on the edges, an area which mixes features addressed in Section 2 with the theory of random graphs. Here, most results refer to problems that admit of a worst case polynomially bounded algorithm. A typical example is provided by the LINEAR ASSIGNMENT problem of minimizing $\Sigma_i c_{i\pi(i)}$ over all permutations π. If the c_{ij} are i.i.d. with distribution F, then a probabilistic value analysis can be arrived at by viewing the problem as weighted matching problem on a directed bipartite graph, with weights $\underline{b}_{ij}$ on edge (i,j) and $\underline{d}_{ji}$ on edge (j,i) such that $\underline{c}_{ij} \stackrel{d}{=} \min\{\underline{b}_{ij}, \underline{d}_{ji}\}$. If we now remove all edges except the s outgoing ones of minimal weight at each vertex and disregard edge orientations in the resulting bipartite graph, a result by Walkup shows that a perfect matching will be present with probability $1 - O(1/n)$ if s=2 and $1 - O(2^{-n})$ if $s > 2$. Hence, Z_n^{OPT} is essentially determined by the small order statistics of F. In the case that the $\underline{c}_{ij}$ are uniform on [0,1], this yields that $E\underline{Z}_n^{OPT} \leq 3$ [Walkup 1979]; generally, $E\underline{Z}_n^{OPT}$ is asymptotic to $nF^{-1}(1/n)$ [Frenk & Rinnooy Kan 1985].

Many other problems in this category have also been succesfully analyzed. For instance, in [Frieze 1982], the MINIMUM SPANNING TREE problem is studied under the assumption that the graph is complete and the edge weights are i.i.d. uniform on [0,1]; the expected optimal solution value is equal

to $\sum_{k=1}^{\infty} 1/k^3 \approx 1.2202$. We also refer to [Perl 1977] for an analysis of the SHORTEST PATH problem; further references can be found in [Karp et al. 1985].

An NP-complete problem that belongs in this category is the ASYMMETRIC TRAVELLING SALESMAN problem, defined on a complete directed graph. We refer to [Steele 1985] for a probabilistic value analysis for a Euclidean variant of this problem n. The optimal solution value is asymptotic to $\zeta\sqrt{n}$ in expectation; see also [Karp 1979] for a heuristic that patches the subcycles appearing in the linear assignment relaxation together to achieve a relative error going to 0 in expectation. Perhaps the most peculiar result has been obtained for its generalization, the QUADRATIC ASSIGNMENT problem

$$\max\{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} c_{ij}d_{k\ell}x_{ik}x_{j\ell} \mid \sum_{i=1}^{n} x_{ij} = 1, (j = 1, \ldots, n);$$

$$\sum_{j=1}^{n} x_{ij} = 1 \ (i = 1, \ldots, n),\ x_{ij} \in \{0,1\}\}.$$

In [Burkhard and Fincke 1982; Frenk et al. 1982], it is shown that for this problem with $\underline{c}_{ij}$ and $\underline{d}_{k\ell}$ i.i.d., the ratio of the best and the worst possible solution value tends to 1 in probability. It shows an unexpected side benefit of probabilistic analysis, in that it clearly indicates how not to generate test problems for an empirical analysis of heuristic solution methods!

5. CONCLUDING REMARKS

In this final section, we first discuss a limitation of the preceding survey in that the probabilistic behaviour discussed there was only caused by factors extraneous to the algorithm itself. The algorithm itself could also contribute to the randomness of its outcome by containing random steps, i.e., steps whose result depends partially on a random mechanism.

A very early example of an algorithm that flips coins is the quicksort method for SORTING. In each step, the algorithm picks a number a_j from the set $\{a_1, \ldots, a_n\}$ to be sorted; it divides the set into the subset smaller than a_j and the subset larger than a_j, and repeats recursively on each subset of cardinality greater than two. The worst case and best case running time of the method are easily seen to be $O(n^2)$ and $O(n \log n)$ respectively; the average running time is $O(n \log n)$ under a variety of distributional assumptions but

also under the assumption that for a fixed input the splitting number a_j is chosen randomly from the set of candidates. Notice that the behaviour of the algorithm is now a random variable independent of any distributional assumption, avoiding the always controversial issue of what a random instance of a particular problem looks like.

The formal study of randomized algorithms is far from complete, and in particular the real power of randomization remains a mysterious issue; for instance, it is not clear to what extent (if any) the class of problems that can be solved in randomized polynomial time (i.e. fast with high reliability) strictly includes the class of problems that can be solved in worst case polynomial time. A recent annotated bibliography [Maffioli et al. 1985] provides a useful survey of the area. We restrict ourselves once again to a discussion of a few examples that highlight the variety of results produced so far.

Historically, PRIMALITY TESTING was the first successful application of randomization. In the algorithm in [Rabin 1980], a number is submitted to k tests and declared to be prime if it passes all of them, with the probability of it being composite after all equal to 2^{-k}. Such an algorithm is called a Monte Carlo method, in contrast to a Las Vegas method in which the algorithm never produces an incorrect answer, but may - with small probability - produce no answer at all. The method for GRAPH ISOMORPHISM in [Babai 1979] is of this nature. The two examples above are of special interest in that they concern two problems whose computational complexity (polynomially solvable or NP-complete) is still unknown.

Generally, randomization may produce a speed-up of a polynomial computation at the expense of a little uncertainty, as for instance for a fast BIPARTITE MATCHING algorithm in [Ibarra & Moran 1981] with a small probability of error, or an $O(|V|^2)$ expected time algorithm for CIRCUIT COVERING ([Itai & Rodeh 1978]) that with small probability will not terminate. For NP-hard problems, there are other potential benefits. Statistical inference has been suggested as a way to estimate the optimal solution value from a sample of (supposedly independent) local minima, i.e., feasible solutions whose value cannot be further improved by a local improvement method ([Dannenbring 1977]). And the fashionable simulated annealing approach can be viewed as a randomized version

of such a local improvement heuristic, in which neighbouring solutions that decrease the quality of the current feasible solution are also accepted, albeit with a small probability that is appropriately decreasing over time ([Kirkpatrick et al. 1983]).

It should be clear by now that the area of probabilistic analysis still harbors many interesting research challenges. The purpose of the preceding sections has, again, not been to provide an exhaustive review, but to provide some typical examples that convey the flavour of this area. They have ranged from the very complete insight we have into various solution methods for the PARTITION problem to the less satisfactory state of the art for the CLIQUE and the COLORING problem. Clearly, a lot of problems and a lot of algorithms await investigation. It is not hard to formulate open questions for probabilistic analysis; so far, however, it has turned out to be quite hard to come up with satisfactory answers for any but the simplest heuristics.

A particularly fascinating possibility is the development of a complexity theory that would lead to a class of problems for which solution to optimality in polynomial expected time is as unlikely as the equality of P and NP. A first step in that direction can be found in [Levin 1984], where a TILING problem is introduced, together with a probability distribution over its problem instances, such that any other problem with a (mildly restricted type of) probability distribution is reducible to the TILING problem.

To establish completeness for other problems in this class is a major challenge of considerable interest. After all, the reasonable average behaviour of enumerative methods (and the remarkable success of a nonenumerative method based on computations in an integer lattice [Lagarias & Odlyzko 1983]) to solve some NP-complete problems and the apparent impossibility to find such algorithms for other NP-complete problems, still defy theoretical explanation!

Acknowledgements

This paper was written during a visit to the Department of Industrial Engineering and Operations Research and the School of Business Administration of the University of California at Berkeley, with partial support from a NATO

Senior Scientist Fellowship. I gratefully acknowledge various useful discussions with Dick Karp and Dorit Hochbaum.

References

Adler, I., R.M. Karp, R. Shamir (1983). A Simplex Variant Solving an m × d Linear Program in $O(\min(m^2,d^2))$ Expected Number of Pivot Steps. Report UCB CSD 83/158, Computer Science Division, University of California, Berkeley.

Adler, I., N. Megiddo (1983). A Simplex Algorithm Whose Average Number of Steps is Bounded Between Two Quadratic Functions of the Smaller Dimension. Technical Report, University of California, Berkeley.

Angluin, D., L.G. Valiant (1979). Fast Probabilistic Algorithms for Hamiltonian Circuits and Matchings. J. Comput. Systems Sci. 19, 155-193.

Babai, L. (1979). Monte Carlo Algorithms in Graph Isomorphism Testing. Rapport de Recherches DMS 79-10, Département de Mathématique et de Statistique, Université de Montréal.

Beardwood, J., J.H. Halton, J.M. Hammersley (1959). The Shortest Path through Many Points. Proc. Cambridge Philos. Soc. 55, 299-327.

Bollobás, B. (1985). Lectures on Random Graphs. (to appear)

Burkard, R.E., U. Fincke (1982). Probabilistic Asymptotic Properties of Quadratic Assignment Problems. Z. Oper. Res. 27, 73-81.

Coffman, E.G., Jr., M.R. Garey, D.S. Johnson (1983). Approximation Algorithms for Bin-Packing - An Updated Survey. Bell Laboratories, Murray Hill, NJ.

Dannenbring (1977). Procedures for estimating optimal solution values for large combinatorial problems. Management Sci. 23, 1273-1283.

Derigs, U. (1984). Private communication.

Few, L. (1955). The Sortest Path and the Shortest Road through n Points. Mathematika 2, 141-144.

Frederickson, G.N. (1980). Probabilistic Analysis for Simple One- and Two-Dimensional bin packing algorithms. Inform. Process. Lett. 11, 156-161.

Frenk, J.B.G., M. van Houweninge, A.H.G. Rinnooy Kan (1982). Asymptotic Properties of the Quadratic Assignment Problem. Math. of Oper. Res. 10, 100-116.

Frenk, J.B.G., A.H.G. Rinnooy Kan (1983). The Asymptotic Optimality of the LPT Heuristic. Math. Oper. Res. (to appear).

Frenk, J.B.G., A.H.G. Rinnooy Kan (1984), Order Statistics and the Linear Assignment Problem. Econometric Institute, Erasmus University, Rotterdam.

Frenk, J.B.G., A.H.G. Rinnooy Kan (1984). On the Rate of Convergence to Optimality of the LPT Rule. Discr. Appl. Math. (to appear).

Frieze, A.M., M.R.B. Clarke (1984). Approximation Algorithms for the m-dimensional 0-1 Knapsack Problem. Eur. J. Oper. Res. 15, 100-109.

Frieze, A.M. (1982). On the Value of a Random Minimum Spanning Tree Problem. Technical report, Department of Computer Science and Statistics, Queen Mary College, University of London.

Golden, B. (1985). Chapter 5. E.L. Lawler, J.K. Lenstra, D.B. Shmoys (eds.). The Traveling Salesman Problem. Wiley, Chichester.

Graham, R.L. (1969). Bounds on Multiprocessing Timing Anomalies. SIAM J. Appl. Math. 17, 263-269.

Grimett, G., C. McDiarmid (1975). On colouring random graphs. Math. Proc. Cambridge Philos. Soc. 77, 313-324.

Haimovich, M. (1983). The Simplex Method is Very Good! - On the Expected Number of Pivot Steps and Related Properties of Random Linear Programs. Columbia University, New York.

Haimovich, M. (1984). Ph.D. Thesis, M.I.T.

Haimovich, M., A.H.G. Rinnooy Kan (1984). Bounds and Heuristics for Capacitated Routing Problems. Math. Oper. Res. (to appear).

Hochbaum, D.S., J.M. Steele (1981). Steinhaus' Geometric Location Problem for Random Samples in the Plane. Adv. in Appl. Probab. 14, 56-67.

Ibarra, O.H., S. Moran (1981). Deterministic and Probabilistic Algorithms for Maximum Bipartite Matching via Fast Matrix Multiplications. Inform. Process. Lett. 13, 12-15.

Itai, A. & M. Rodeh (1978). Covering a Graph by Circuits. G. Ausiello, C. Böhm (eds.) Automata, Languages and Programming, Lecture Notes in Computer Science 62, Springer, Berlin, 289-299.

Karmarkar, N., R.M. Karp (1982). The Differencing Method of Set Partitioning. Math. Oper. Res. (to appear).

Karmarkar, N., R.M. Karp, G.S. Lueker, A. Odlyzko (1984). The Probability Distribution of the Optimal Value of a Partitioning Problem. Bell Laboratories, Murray Hill, NJ.

Karp, R.M. (1977). Probabilistic Analysis of Partitioning Algorithms for the Traveling Salesman Problem in the Plane. Math. Oper. Res. 2, 209-224.

Karp, R.M. (1979). A Patching Algorithm for the Nonsymmetric Traveling Salesman Problem. SIAM J. Comput. 8, 561-573.

Karp, R.M. (1984). Private communication.

Karp, R.M., J.K. Lenstra, C.J.H. McDiarmid, A.H.G. Rinnooy Kan (1985). Chapter 6. M. OhEigeartaigh, J.K. Lenstra, A.H.G. Rinnooy Kan (eds.). Combinatorial Optimization: Annotated Bibliographies, Wiley, Chichester.

Karp, R.M., J.M. Steele (1985). Chapter 6. E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, D.B. Shmoys (eds.). The Traveling Salesman Problem. Wiley, Chichester.

Kirkpatrick, S., C.D. Gelat Jr. and M.P. Vecchi (1983). Optimization by Simulated Annealing. Science 220, 671-680.

Knödel, W. (1981). A Bin Packing Algorithm with Complexity O(n log n) and Performance 1 in the Stochastic Limit. J. Gruska, M. Chytil (eds.). Mathematical Foundations of Computer Science 1981, Lecture Notes in Computer Science 118, Springer, Berlin, 369-378.

Komlős, J., E. Szemerédi (1983). Limit Distribution for the Existence of Hamiltonian Cycles in Random Graphs. Discrete Math. 43, 55-63.

Korsunov, A.D. (1980). The Chromatic Number of n-vertex Graphs. Metody Diskret. Analiz. 35, 14-44, 104 (in Russian).

Lagarias, J.C., A.M. Odlyzko (1983). Solving Low-Density Subset Sum Problems. Proc. 24th Annual IEEE Symp. Foundations of Computer Science, 1-10.

Lawler, E.L., J.K. Lenstra, A.H.G. Rinnooy Kan, D.B. Shmoys (eds.) (1985). The Traveling Salesman Problem, Wiley, Chichester.

Levin, L.A. (1984). Problems, Complete in 'Average' Instance. Proc. 16th Annual ACM Symp. Theory of Computing, 465.

Maffioli, F., M.G. Speranza, C. Vercellis (1985). Chapter 7. M. OhEigeartaigh, J.K. Lenstra, A.H.G. Rinnooy Kan (eds.) Combinatorial Optimization: Annotated Bibliographies, Wiley, Chichester.

Matula, D.W. (1976). The Largest Clique Size in a Random Graph. Technical report CS7608, Department of Computer Science, Southern Methodist University, Dalles TX.

McDiarmid, C. (1979). Colouring Random Graphs Badly. R.J. Wilson (ed.) Graph Theory and Combinatorics, Pitman Research Notes in Mathematics 34, Pitman, London 76-86.

Meanti, M., A.H.G. Rinnooy Kan, L. Stougie, C. Vercellis (1984). A Probabilistic Analysis of the Multiknapsack Value Function. Econometric Institute, Erasmus University, Rotterdam.

Papadimitriou, C.H. (1978). The Complexity of the Capacitated Tree Problem. Networks 8, 217-230.

Perl, Y. (1977). Average Analysis of Simple Path Algorithms. Technical report

UIUCDCS-R-77-905, Department of Computer Science, University of Illinois at Urbana-Champaign.

Pósa, L. (1976). Hamiltonian Circuits in Random Graphs. Discrete Math. 14, 359-364.

Rabin, M.O. (1980). Probabilistic Algorithms for Testing Primality. J. Number Theory 12, 128-138.

Steele, J.M. (1981). Subadditive Euclidean Functionals and Nonlinear Growth in Geometric Probability. Ann. Probab. 9, 365-376.

Steele, J.M. (1985). Probabilistic Algorithm for the Directed Traveling Salesman Problem. Math. Oper. Res. (to appear).

Stougie, L. (1985), Ph.D. Thesis. Econometric Institute, Erasmus University, Rotterdam.

Todd, M.J. (1983). Polynomial Expected Behavior of a Pivoting Algorithm for Linear Complementarity and Linear Programming Problems. Technical report 595, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

Zemel, E. (1984). Probabilistic Analysis of Geometric Location Problems. Ann. Oper. Res. I.

LIST OF REPORTS 1985

8500 "Publications of the Econometric Institute Second Half 1984: List of Reprints 378-400, Abstracts of Reports".

8501/O **R.M. Karp, J.K. Lenstra, C.J.H. McDiarmid and A.H.G. Rinnooy Kan,** "Probabilistic analysis of combinatorial algorithms: an annotated bibliography", 26 pages.

8502/E **M.E. Homan,** "Verschillen in consumptie tussen één en tweekostwinnerhuishoudens: een eerste analyse", 13 pages.

8503/O **A.W.J. Kolen,** "The round-trip p-center and covering problem on a tree", 17 pages.

8504/O **H.C.P. Berbee, C.G.E. Boender, A.H.G. Rinnooy Kan, C.L. Scheffer, R.L. Smith and J. Telgen,** "Hit-and-run algorithms for the identification of nonredundant linear inequalities", 32 pages.

8505/S **H. Brozius and L. de Haan,** "On limiting laws for the convex hull of a sample", 10 pages.

8506/O **J.B.G. Frenk and A.H.G. Rinnooy Kan,** "On the rate of convergence, to optimality of the LPT rule - postscript", 5 pages.

8507/E **P. Kooiman, H.K. van Dijk and A.R. Thurik,** "Likelihood diagnostics and Bayesian analysis of a micro-economic disequilibrium model for retail services", 35 pages.

8508/O **C.G.E. Boender, A.H.G. Rinnooy Kan and J.R. de Wit,** "A Bayesian procedure for the (s,Q) inventory problem", 26 pages.

8509/E **B.M.S. van Praag, S. Dubnoff and N.L. van der Sar,** "From judgments to norms: measuring the social meaning of income, age and educaton", 35 pages.

8510/S **B. Bode, J. Koerts and A.R. Thurik,** "On shopkeepers' pricing behaviour", 19 pages.

8511/M **H. Bart, I. Gohberg and M.A. Kaashoek,** "Exponentially dichotomous operators and inverse Fourier transforms", 70 pages.

8512/M **J. Brinkhuis,** "Normal integral bases and embedding of fields", 13 pages.

8513/E **P.M.C. de Boer,** "On the relationship between Revankar's and Lu-Fletcher's production function", 5 pages.

8514/S **L. de Haan,** "Extremes in higher dimensions: the model and some statistics", 15 pages.

8515 Publications of the Econometric Intitute First Half 1985: List of Reprints 401-414, Abstracts of Reports.

8516/A **A.R. Thurik and A. Kleijweg,** "Cyclical effects in retail labour productivity", 20 pages.

8517/C **B.M.S. van Praag and J. van Weeren,** "The impact of past experiences and anticipated future on individual income judgements", 24 pages.

8518/A **A.R. Thurik and J. Koerts,** "Behaviour of retail entrepreneurs". 16 pages.

8519/B **M. Hazewinkel, J.F. Kaashoek and B. Leynse,** "Pattern formation for a one dimensional evolution equation based on Thom's River Basin Model". 24 pages.

8520/A **H.K. van Dijk, T. Kloek and C.G.E. Boender,** "Posterior moments computed by mixed integration", 25 pages.

8521/B **J. Brinkhuis,** "Testing concavity and quasi-concavity is easy", 12 pages.

8522/A **R. Harkema,** "Minimum sample size requirements for maximum likelihood estimation of some demand models", 25 pages.

8523/B **A.C.F. Vorst,** "The general linear group of discrete Hodge algebras", 10 pages.

8524/A **A.M.H. Gerards and A.W.J. Kolen,** "Polyhedral combinatorics in combinatorial optimization", 25 pages.

8525/A **B.J. Lageweg, J.K. Lenstra, A.H.G. Rinnooy Kan and L. Stougie,** "Stochastic integer programming by dynamic programming", 19 pages.

8526/A **J.B. Orlin,** "A dual version of Tardo's algorithm for linear programming", 9 pages.

8527/A **B.M.S. van Praag,** "Household cost functions and equivalence scales", 24 pages (report 8424/E revised)

8528/A **S. Schim van der Loeff,** "Limited information maximum likelihood estimation of a subsystem of nonlinear equations", .. pages.

8529/A **B.S. van der Laan and A.S. Louter,** "On the number and the amount of damage of a passenger car traffic accidents in the Netherlands", ..pages.

8530/A **B.S. van der Laan and A.S Louter,** "A statistical model for the costs of passenger car traffic accidents", .. pages.

8531/A **B.M.S. van Praag, J. de Leeuw and T. Kloek,** "The population-sample decomposition approach to multivariate estimation methods", 34 pages.

8532/A **M.E. Homan, B.M.S. van Praag and A.J.M. Hagenaars,** "Household cost functions and the value of home production in one-and two earner families", .. pages.

8533/B **R.J. Stroeker,** "An inequality for YFF's analogue of the Brocard angel of a plane triangle", 18 pages.

8534/A **J. Bouman,** "Testing nonnested linear hypothesis: A Bayesian approach based on imcomletely specified prior distributions". pages.

8535/A **P.M.C. de Boer, R. Harkema and B.J. van Heeswijk,** "Estimating foreign trade functions, a comment and a correction". pages.

8536/A **C.G.E. Boender and A.H.G. Rinnooy Kan,** "Bayesian stopping rules for multistart global optimization methods", 30 pages.

8537/A **A.H.G. Rinnooy Kan and G.T. Timmer,** "The multi level single linkage method for unconstrained and constrained global optimization", 14 pages.

8538/A **A.H.G. Rinnooy Kan,** "Probabilistic analysis of algorithms", 26 pages.

Until report 8515 this series was devided in E(conometrics), S(tatistics), M(athematics) and O(perations Research). From report 8516 on it will be

A. Economics, Econometrics and Operations Research

B. Mathematics

C. Miscellaneous