Probabilistic Analysis of the Complexity of A^*

Nam Huyn, Rina Dechter and Judea Pearl

Cognitive Systems Laboratory, School of Engineering and Applied Science, University of California, Los Angeles, CA, U.S.A.

Recommended by Daniel G. Bobrow

ABSTRACT

This paper analyzes the number of nodes expanded by A^* as a function of the accuracy of its heuristic estimates by treating the errors $h^* - h$ as random variables whose distributions may vary over the nodes in the graph. Our model consists of an m-ary tree with unit branch costs and a unique goal state situated at a distance N from the root.

Two results are established:

(1) *for any error distribution, if A_1^* is stochastically more informed than A_2^*, then A_1^* is stochastically more efficient than A_2^*, and*

(2) *if the probability that the relative error be bounded away from zero is greater than $1/m$, then the average complexity of A^* is exponential with N, whereas if the probability of zero error is greater than $1 - 1/m$, the average complexity is $O(N)$.*

1. Introduction

Most AI programs employ some sort of 'heuristics', i.e., assertions inferred from simplified models. By its very nature, the utility of the heuristic used depends on the proximity between its underlying model and the reality of the problem at hand. However, very little is known about the relationship between the proximity of the model and the performance of the algorithms it serves to guide.

The A^* search algorithm is a model for studying this relationship quantitatively. Aside from providing a universal schema for casting search problems, A^* is perhaps the only heuristic search technique about which a theoretical body of knowledge has begun to crystalize [1]–[8]. One of the attractive features of A^* is the fact that both its mission (i.e., distance minimization) and its heuristics can be expressed quantitatively and so, instead of dealing with a

This work was supported in part by the National Science Foundation Grants MCS 78-07468 and MCS 78-18924.

Artificial Intelligence 15 (1980), 241–254

vague proximity relation between a model and reality, one can invoke formal measures of accuracy for distance-estimation procedures.

The first analysis on the effect of errors (i.e., inaccuracies) on the performance of A^* was conducted by Pohl [9] and has been pursued since by Munyer [10], Vanderbrug [11], Pohl [7], and Gaschnig [8]. The basic motivation for these studies has been the following enigma. When A^* employs a perfectly informed heuristic ($h = h^*$) it is propelled directly toward the goal without ever getting sidetracked, spending only N computational steps where N is the distance of the goal state. At the other extreme, when no heuristic at all is available ($h = 0$), the search becomes exhaustive, yielding an exponentially growing complexity. Between these two extremes there lies an unknown relationship between the accuracy of the heuristic estimates and the complexity of the search which they help control. Only sketchy information is known today about the character of this relationship, and it is entirely confined to worst-case-analyses.

It often happens that the designer of a problem-solving system has the option of controlling the quality of the heuristics employed, usually by resorting to more sophisticated models or by permitting lengthier computations. The question then arises, "Is it worth it?". Having a predictive model for the accuracy-complexity dependency would help the designer decide whether the added computation invested in improving accuracy would pay itself in reduced search complexity. Some results along this vein were obtained by Pohl [7] and Gaschnig [8]. For instance, if the relative error remains constant, then the search complexity is exponential. When the absolute error is constant the search complexity is linear.

These results, however, were derived for a worst case model, assuming that a clever adversary would distribute the errors in such a way that A^* would exhibit its poorest performance. A probabilistic extension of these analyses is warranted for two main reasons. First, worst case results are often suspected of being too conservative and so, one wonders whether the average performance of A^* would be significantly higher. Second, it is often hard to guarantee precise bounds on the magnitude of errors produced by a given heuristic, whereas probabilistic characterization of these magnitudes may be more natural.

This paper presents a probabilistic analysis of A^*'s performance, treating the heuristic estimates as random variables whose distributions may vary over the nodes of the graph. Section 2 defines the models for the search space and the heuristic distributions. Section 3 establishes a general formula for the mean complexity of A^* for a given characterization of error distributions. In Section 4, we formalize and prove the notion that if one heuristic is 'generally' more accurate than another, then the first would 'generally' give rise to a more efficient search. Section 5 is devoted to the analysis of the mean complexity of A^* when the relative errors have a positive probability of being bounded away from zero.

2. Model Definition

Following Pohl [7] and Gaschnig [8], our search space is modeled by a uniform m-ary tree T, with a unique start state S and a unique goal state G, situated at a distance N from S. The search tree is depicted in Fig. 1 where (without loss of generality) the solution path $(S, n^s_{N-1}, \ldots, n^s_i, \ldots, n^s_1, G)$ is represented on the extreme right. The trees $T_1, \ldots, T_i, \ldots, T_N$ are subtrees of T, one level removed from the solution path. Thus each 'off-course' subtree T_i is rooted at a direct successor of n^s_i which is off the solution path.

A^* searches for the goal state G using the following procedure. An evaluation function $f(\cdot)$ assigns a real value to each node n in T representing a heuristic estimate of the length of the optimal solution path constrained to contain n. All branches are assumed to be of unit distance, and f is of the form:

$$f(n) = g(n) + h(n)$$

where:

$g(n)$ is the length of the shortest path from S to n found by A^*.

$h(n)$ is a heuristic estimate of the minimal distance from n to G.

$h(n)$ is assumed to be *admissible*, i.e., $0 \leq h(n) \leq h^*(n)$ for all nodes in T, where $h^*(n)$ is the actual minimal distance from n to G.

ALGORITHM A^* (for trees)

Step 1: Mark S as 'OPEN' and compute $f(S)$.

Step 2: Choose an OPEN node n whose f value is minimal (resolving ties arbitrarily).

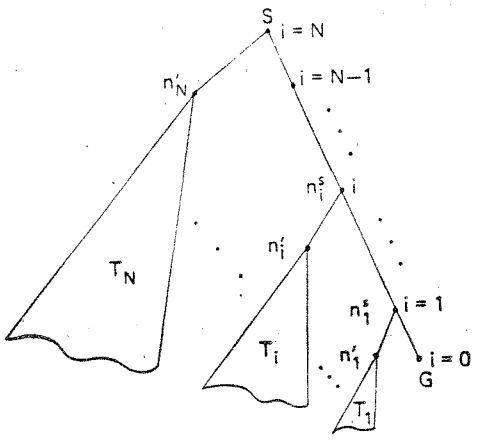

FIG. 1. Uniform tree model ($m = 2$).

Step 3: If n is the goal node G, then terminate.

Step 4: Mark n as 'CLOSED' and compute $f(n_i')$ for each son n_i' of n. Mark each such node as OPEN. Go to Step 2.

The execution of Step 4 is called an 'expansion' of node n, and the standard measure of complexity of A^* is the number of nodes, Z, expanded before termination.

Our probabilistic analysis of A^* assumes that each $h(n)$ can be treated as a random variable ranging over $[0, h^*(n)]$ and characterized by a distribution function $F_{h(n)}(x) = P[h(n) \leq x]$. We further assume that for any two nodes n_1 and n_2, $h(n_1)$ and $h(n_2)$ are conditionally independent, given their corresponding distances to the goal.

3. A General Formula for the Mean Complexity of A^*

Given a characterization of the distribution function $F_{h(n)}(x)$, our task is to compute $E(Z)$, the expected number of nodes expanded by A^*, as a function of $h^*(S) = N$, the length of the solution path.

Since $h(n)$ is admissible, a necessary condition for the expansion of a node n is $n \in$ OPEN and $f(n) \leq f^*(S)$, and a sufficient condition for its expansion is $n \in$ OPEN and $f(n) < f^*(S)$ (see [12]). Consequently we can write:

$$P(n \text{ expanded}) \leq P(n\text{'s parent expanded and } f(n) \leq N), \tag{1}$$

$$P(n \text{ expanded}) \geq P(n\text{'s parent expanded and } f(n) < N). \tag{2}$$

The gap between the bounds in (1) and (2) is created by the eventuality that some node n_i' on the solution path together with a node n off the solution path could both be OPEN and $f(n_i') = f(n) = N$. In such a case the expansion of n would be determined by the tie-breaking rule and not by the character of n. However, if $h(n)$ is a continuous random variable the likelihood of such eventuality is essentially zero, and the bounds in (1) and (2) become equalities. Moreover, since we are primarily interested in establishing bounds to $E(Z)$ we may, in case $h(n)$ is a discrete random variable, use (1) for deriving upper bounds and (2) for deriving lower bounds. In the remaining portion of this paper we will treat (2) as an equality with the understanding that, under special circumstances, (1) should be consulted.

The process by which algorithm A^* expands nodes in a search tree is analogous to a death–birth process (or more precisely a *branching process*) where, for any node n off the solution path, the condition for reproduction is $f(n) < f^*(S)$ and the death condition (prior to reproduction) is $f(n) \geq f^*(S)$.

The branching process model was originally used to study the problem of families' survival [13]. The process can be described as follows. At the top level, we have the 0th generation consisting of one member. This member gives

birth to a random number of sons of the first generation (or equivalently, the father gives birth to a fixed number of infants, a random number of which will survive to reproduce). Then, each surviving member of the first generation gives birth to a random number of sons of the second generation, etc. The number of sons of any member may also be 0 (i.e., all the sons may die), thus raising the possibility that the whole family be extinct after a certain number of generations.

The generation process is tree-structured (cf. Diagram 1).

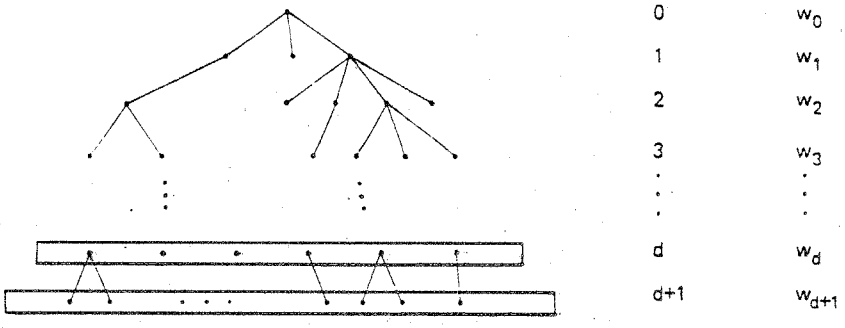

DIAGRAM 1.

Analytically, the kth member of generation d $(1 \leq k \leq w_d)$ gives birth to a random number $x_{k,d}$ of fertile sons which are members of generation $d+1$. Therefore, the size w_{d+1} of the $(d+1)$th generation is the sum of a random number (w_d) of random variables:

$$\begin{cases} w_{d+1} = x_{1,d} + \cdots + x_{k,d} + \cdots + x_{w_d,d} & \text{if } w_d \geq 1, \\ w_{d+1} = 0 & \text{if } w_d = 0. \end{cases}$$

In accordance with our previous assumption that $F_{h(n)}(x)$ depends only on $h^*(n)$, we may assume that w_d and $x_{k,d}$ are mutually independent and that the variables $\{x_{k,d}\}$, $k = 1, 2, \ldots, w_d$, are identically distributed, i.e., the fertility of each member of generation d may depend on d but is independent of both the size of the generation and its serial number. These conditions imply [14] that the expectation of w_{d+1} is given by the product:

$$E(w_{d+1}) = E(w_d) \cdot E(x_{k,d}).$$

Denoting by q_d the probability of infant survival for any member of generation d and assuming that exactly m infants are born to each parent, $x_{k,d}$ is a binomial

random variable with mean $E(x_{k,d}) = mq_{d+1}$. Consequently, $E(w_{d+1})$ becomes:

$$E(w_{d+1}) = E(w_d)mq_d$$

and, by induction over $d = 0, 1, 2, \ldots$ (with $E(w_0) = q_0$) we finally obtain:

$$E(w_d) = m^d q_d q_{d-1} \cdots q_0. \tag{3}$$

The expected number of nodes expanded by A^* could be calculated by applying this analysis to each of the off-course subtrees in Fig. 1. The expected number of nodes expanded at depth d of T_i is given by the product $m^d q_{i,d} \cdot q_{i,d-1} \cdots q_{i,0}$, where $q_{i,k}$ represents the probability that a node n_i at depth k of T_i is expanded once it enters OPEN (infant survival), i.e.:

$$q_{i,k} = P(f(n_{ik}) < N). \tag{4}$$

Summing this product over all levels d and over all off-course subtrees and adding the N nodes expanded along the solution path, we obtain a formula for the expected number of nodes expanded:

$$E(Z) = N + (m-1)\sum_{i=1}^{N}\sum_{d=0}^{\infty} m^d \prod_{k=0}^{d} q_{i,k} \tag{5}$$

where $q_{i,k}$ is given in (4). In Section 5 we will utilize eq. (5) to show that a fixed distribution of relative errors gives rise to an exponential complexity. However, prior to this analysis we wish to examine in what sense one heuristic can be said to be '*superior*' to another.

4. Comparative Merits of Random Heuristics

The problem of deciding whether one heuristic is better than another arises often. Clearly, if one heuristic consistently provides a more accurate estimate of h^*, it ought to be preferred. This is indeed the essence of a theorem by Gelperin [6] who showed that if for each node of the search graph $h_1(n) < h_2(n)$ and both are admissible, then every node expanded by A_2^* is also expanded by A_1^*, i.e., $h_2(n)$ is to be preferred. However, we seldom possess sufficient a priori knowledge to guarantee that the inequality $h_1(n) < h_2(n)$ holds for every node in the graph. Even when the improved accuracy of h_2 is a product of invoking more sophisticated computation procedures than h_1, the improvement is seldom guaranteed to take place at every node of the problem space. Generally, when $h(n)$ is made more accurate for some nodes, it may become less accurate for others. It is natural to then ask whether a statement of preference can be made in the case where the inequality $h_1 < h_2$ is only known to be a reasonably

probable but occasionally violated event. The formalization and affirmation of such a statement will be carried out in the rest of this section.

DEFINITION. Given two random variables X_1 and X_2, we say that X_2 is *stochastically greater* than X_1 (denoted by $X_1 \trianglelefteq X_2$) iff:

$$P(X_2 > x) \geq P(X_1 > x) \qquad \forall x \in R$$

or equivalently,

$$F_{X_1}(x) \geq F_{X_2}(x) \qquad \forall x \in R.$$

DEFINITION. Let A_1^* and A_2^* employ the heuristic functions h_1 and h_2, respectively. A_2^* is said to be *stochastically more informed* than A_1^* iff $h_1(n) \trianglelefteq h_2(n)$, $\forall n \in T$. Similarly, A_2^* is said to be *stochastically more efficient* than A_1^* iff $Z_2 \trianglelefteq Z_1$, where Z_1 and Z_2 are the number of nodes expanded by A_1^* and A_2^*, respectively.

An inspection of eqs. (5) and (4) reveals that if for every node n $h_2(n)$ is stochastically greater than $h_1(n)$, then h_2 would induce a lower $E(Z)$. This is so because every $q_{i,k}$ in the case of h_2 will be smaller than the corresponding factor for h_1. However, a much stronger statement can be made; $h_1(n) \trianglelefteq h_2(n)$ implies not merely preference in the mean, but also *stochastic* preference as is demonstrated in the following paragraphs.

THEOREM 1. *For any error distribution, if A_2^* is stochastically more informed than A_1^*, then A_2^* is stochastically more efficient than A_1^*.*

PROOF. Given a set of mutually independent random variables, the order $\trianglelefteq$ is preserved under addition within the set. In other words if X_1, Y_1, X_2, Y_2 are mutually independent, then $X_1 \trianglelefteq X_2$ and $Y_1 \trianglelefteq Y_2$ imply $X_1 + Y_1 \trianglelefteq X_2 + Y_2$.

Denote by $Z_1(n)$ and $Z_2(n)$ the number of nodes expanded by A_1^*, and A_2^*, respectively, within the subtree rooted at n. We want to prove that $Z_2(S) \trianglelefteq Z_1(S)$. Equivalently, since the number of nodes expanded in the off-course trees T_i (see Fig. 1) is independent, using the order preserving property of addition for independent random variables, we only need to show that $Z_2(n_i') \trianglelefteq Z_1(n_i')$ for one arbitrary T_i. Let n be any node at level d of that tree; it is sufficient to prove that $Z_2(n) \trianglelefteq Z_1(n)$ for all such n. We plan to prove this by a bottom-up induction on the level d in the tree.

(1) For d sufficiently large the statement is trivially true since $Z_2(n) = Z_1(n) = 0$ for all n at level d.

(2) Assume $Z_2(n) \trianglelefteq Z_1(n)$ for all n at level d; we wish to show that $Z_2(n) \trianglelefteq Z_1(n)$ for all nodes n at level $d-1$. Let nodes $n_1 \cdots n_k \cdots n_m$ be the direct descendents of a node n at level $d-1$ (cf. Diagram 2). For all integers $x \geq 0$

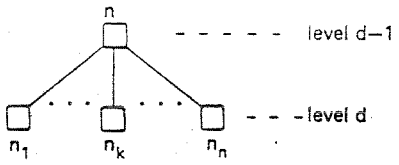

DIAGRAM 2.

we have:

$$P(Z(n)\leq x)=\begin{cases}1 & \text{if } A^* \text{ does not expand } n,\\ P\left(\sum_k Z(n_k)\leq x-1\right) & \text{if } A^* \text{ expands } n\end{cases}$$

or:

$$\begin{aligned}F_{Z(n)}(x) &= P(Z(n)\leq x)\\ &= [1-P(A^* \text{ expands } n)]\\ &\quad + P(A^* \text{ expands } n)P\left(\sum_k Z(n_k)\leq x-1 \,\middle|\, n \text{ expanded}\right).\end{aligned} \tag{6}$$

Since $h_1(n)\circledS h_2(n)\forall n$, and the path from X to n is unique, it is clear that (see (4)):

$$P(A_1^* \text{ expands } n)\geq P(A_2^* \text{ expands } n). \tag{7}$$

By induction hypothesis:

$$Z_2(n_k)\circledS Z_1(n_k) \quad \text{for all } k$$

but, since $Z(n_1)\cdots Z(n_k)\cdots Z(n_m)$ are conditionally independent:

$$\sum_k Z_2(n_k)\circledS \sum_k Z_1(n_k)$$

or:

$$P\left(\sum_k Z_1(n_k)\leq x-1\right)\leq P\left(\sum_k Z_2(n_k)\leq x-1\right). \tag{8}$$

Thus, using (7) and (8) in (6), we wish to show that $F_{Z_1(n)}(x)\leq F_{Z_2(n)}(x)$ or that:

$$\left.\begin{matrix}u_1\geq u_2\\ v_1\leq v_2\end{matrix}\right\}\Rightarrow 1-u_1+u_1v_1\leq 1-u_2+u_2v_2 \tag{9}$$

where:

$$u_i = P(A_i^* \text{ expands } n) \qquad i = \{1, 2\},$$

$$v_i = P\left(\sum_k Z_i(n_k) \leq x - 1\right) \qquad i = \{1, 2\}.$$

The implication in (9) can be reduced to:

$$\left.\begin{matrix} u_1 \geq u_2 \\ v_1 \leq v_2 \end{matrix}\right\} \Rightarrow (u_1 - u_2)(1 - v_1) + (v_2 - v_1)u_2 \geq 0$$

which is easily validated using the fact that v_1 stands for a probability and so, $v_1 \leq 1$.

We conclude, therefore, that $F_{Z_1(n)}(x) \leq F_{Z_2(n)}(x)$ for all nodes and all levels of a given off-course subtree of T. Consequently, $Z_1(S) \circledS Z_2(S)$ which proves Theorem 1.

5. Exponential Bounds to the Average Complexity

Assuming that the relative errors $Y(n) = [h^*(n) - h(n)]/h^*(n)$ are independent random variables $(1 \geq Y(n) \geq 0)$ with arbitrary distribution functions $F_{Y(n)}(y) = P(Y(n) \leq y)$, we wish to examine the conditions under which the set of distributions $\{F_{Y(n)}(y)\}$ lead to an exponential growth of $E(Z)$ (eq. (5)). Our plan is to begin with the case where $Y_n(y)$ are identically distributed over all nodes with a well-behaved distribution function $F_Y(y)$, determine under what conditions $F_Y(y)$ would give rise to an exponential $E(Z)$, and then identify those sets $\{F_{Y(n)}(y)\}$ which could be uniformly bounded from above by $F_Y(y)$. Our choice for $F_Y(y)$ is the piecewise linear functions:

$$F_Y(y) = L(\beta, \varepsilon; y) = \begin{cases} 1 & y > \varepsilon. \\ \beta + \dfrac{(1-\beta)}{\varepsilon} y & 0 \leq y \leq \varepsilon. \\ 0 & y < 0 \end{cases}$$

where $0 \leq \beta \leq 1$ and $0 < \varepsilon \leq 1$. This distribution, with its associated density function $f_Y(y)$, is shown in Fig. 2. It characterizes a mixed random variable

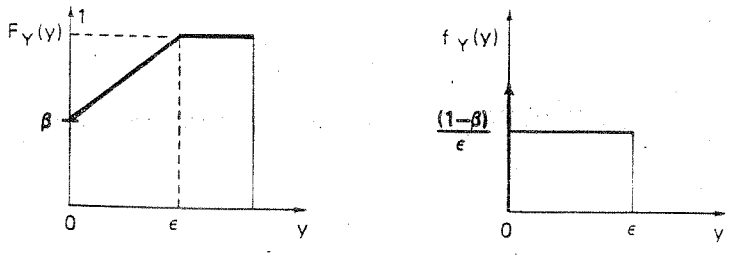

FIG. 2.

whose continuous part is uniformly distributed between 0 and ε, having a probability β of being identically equal to zero. Aside from permitting a simplification of (5), such functions can be used to bound from above any reasonable distribution function.

For a node $n_{i,k}$ at depth k of T_i we have:

$$\begin{aligned} h^*(n_{i,k}) &= i+1+k, \\ g(n_{i,k}) &= N-i+k+1 \end{aligned} \tag{11}$$

and so (4) becomes:

$$q_{ik} = P(h(n) < i-k-1) = 1 - F_{Y(n)}\left(\frac{2(k+1)}{i+k+1}\right). \tag{12}$$

Using the linear distribution of (10), q_{ik} simplifies to:

$$q_{ik} = \begin{cases} (1-\beta)\left(1-\dfrac{2(k+1)}{\varepsilon(i+k+1)}\right) & k \leq \dfrac{\varepsilon i}{2-\varepsilon} - 1, \\ 0 & k > \dfrac{\varepsilon i}{2-\varepsilon} - 1 \end{cases} \tag{13}$$

and (5) becomes:

$$\begin{aligned} E(Z) = N + (m-1) \sum_{i=1}^{N} \sum_{d=0}^{\lfloor \varepsilon i/(2-\varepsilon)-1 \rfloor} (1-\beta)^{d+1} m^d \\ \times \prod_{k=0}^{d} \left[1 - \frac{2(k+1)}{\varepsilon(i+k+1)}\right]. \end{aligned} \tag{14}$$

The right-hand side of (14) would decrease by taking only one term from the double summation, e.g., $i = N$ and $d = d_0$, where the choice of d_0 will be made at a later stage. Thus, (14) becomes:

$$E(Z) \geq N + (m-1) m^{d_0} (1-\beta)^{d_0+1} \exp \sum_{k=0}^{d_0} \log \left[1 - \frac{2(k+1)}{(N+k+1)\varepsilon}\right]. \tag{15}$$

Using the inequality $\log(1-q) \geq -q/(1-q)$ we get:

$$\begin{aligned} \sum_{k=0}^{d_0} \log\left[1 - \frac{2(k+1)}{(N+k+1)\varepsilon}\right] &\geq -\sum_{k=0}^{d_0} \frac{2(k+1)}{N\varepsilon - (2-\varepsilon)(k+1)} \\ &\geq -\sum_{k=0}^{d_0} \frac{2(k+1)}{N\varepsilon - (2-\varepsilon)(d_0+1)} \\ &= -\frac{(d_0+2)(d_0+1)}{N\varepsilon - (2-\varepsilon)(d_0+1)}. \end{aligned} \tag{16}$$

Substituting in (15) and choosing:

$$d_0 = \lfloor \alpha N \rfloor, \qquad \alpha < \frac{\varepsilon}{2-\varepsilon} \tag{17}$$

we obtain:

$$E(Z) \geq (m-1)\exp\left\{N\left[\alpha \log[m(1-\beta)] - \frac{\alpha^2}{\varepsilon - (2-\varepsilon)\alpha} + O(1/N)\right]\right\}. \tag{18}$$

For $m(1-\beta) > 1$, the expression in the bracket is maximized by the choice:

$$\alpha = \frac{\varepsilon}{2-\varepsilon}\left(1 - \sqrt{\frac{1}{1+(2-\varepsilon)\log[m(1-\beta)]}}\right) \tag{19}$$

yielding:

$$\begin{aligned} E(Z) &\geq (m-1)\exp\left\{\frac{N\varepsilon}{(2-\varepsilon)^2}\left[\sqrt{1+(2-\varepsilon)\ln[m(1-\beta)]} - 1\right]^2 + O(1)\right\} \\ &= (m-1)[B(\varepsilon, \beta, m)]^{Ne[1+O(1/N)]} \end{aligned} \tag{20}$$

where

$$B(\varepsilon, \beta, m) = \exp\left[\frac{\sqrt{1+(2-\varepsilon)\log[m(1-\beta)]} - 1}{2-\varepsilon}\right]^2 > 1. \tag{21}$$

For example, $B(1, 0, 2) = 1.095$ and $B(1, 0, 10) = 1.95$.

If $m(1-\beta) \leq 1$, (14) gives:

$$E(Z) \leq N + (m-1)N \sum_{d=0}^{\lceil N(\varepsilon/2-\varepsilon)\rceil} (1-\beta)^d m^d \prod_{k=0}^{d}\left[1 - \frac{2(k-1)}{\varepsilon(N+k+1)}\right]. \tag{22}$$

But, since the product term is lower than unity, $E(N)$ can be upper bounded by:

$$E(Z) \leq N[1 + O(1/N)]. \tag{23}$$

Eqs. (20) and (23) support the following theorem.

THEOREM 2. *If the relative errors in a uniform m-ary tree are independent and identically distributed with a uniform distribution over* $[0, \varepsilon]$, *then the average complexity of* A^* *is given by*:

$$E(Z) = \begin{cases} O[\exp(cN)] & \text{if} \quad P(h = h^*) < 1 - 1/m, \\ O(N) & \text{if} \quad P(h = h^*) \geq 1 - 1/m \end{cases} \tag{24}$$

where c is a positive constant depending on $m(1-\beta)$ *and* ε.

Theorem 2 generalizes the results of Gaschnig [8] from the worst case to the average case analysis. Gaschnig has shown that if at all nodes off the solution path the relative errors stay at their maximal value above some fixed constant ε, then the complexity of A^* is exponential. Theorem 2 implies that the average complexity follows a similar behavior when the relative errors are uniformly distributed over the interval $[0, \varepsilon]$. Evidently, not much is gained by diffusing the errors smoothly over the interval. The only time the complexity of A^* reduces to a polynomial is when the likelihood that h literally coincides with h^* is sufficiently high at all nodes.

This result is not unique to the uniform distribution, but generalizes to distributions of arbitrary shape as long as a single distribution governs all the nodes in the graph.

COROLLARY 1. *If the relative errors in a uniform m-ary tree are independent and identically distributed random variables with an arbitrary distribution* $F_Y(y)$, then the average complexity of A^* is identical to that given by (24).

PROOF. If $P(h = h^*) < 1 - 1/m$, then $F_Y(0) = 1/m - \delta$ where δ is some positive quantity. Such a function can always be bounded from above by a linear function $L(1 - 1/m - \delta/2, \varepsilon; y)$ and, therefore, the complexity of A^* would be not lower than that of (24), i.e., exponential.

If $P(h = h^*) \geq 1 = 1/m$, then the product term in (5) would be strictly smaller than $(m)^{-d}$ for any $d > 0$ leading to a geometrical series with:

$$E(Z) = O(N).$$

These results can be further generalized to characterize the conditions under which errors with varying distributions would give rise to exponential complexity.

THEOREM 3. *If, for every node in the tree, the probability that the relative error exceeds some fixed positive quantity* ε *is greater than* $1/m$, *then the average complexity of* A^* *is exponential in* N.

PROOF. Consider the set of random variables $\{Y_n\}$ where Y_n stands for the relative error at node n, and n ranges over all nodes of the tree. The condition $P(Y_n > \varepsilon) < 1/m$ translates to $F_{Y_n}(\varepsilon) < 1 - 1/m$. But every distribution function $F_{Y_n}(y)$ satisfying the latter inequality can be bounded from above by a linear function $L(\beta, \varepsilon; y)$ with $\beta < 1 - 1/m$ which renders the mean complexity of A^* exponential. Thus, the entire set $\{Y_n\}$ would also give rise to an exponentially growing average complexity.

Theorem 3 implies that in order to avoid an exponential growth of $E(Z)$, the magnitude of the typical errors in the tree should increase slower than

linearly with the distance from the goal. For example, if the absolute errors $h^*(n) - h(n)$ are bounded by a fixed quantity δ, the condition of Theorem 3 would not be satisfied. For every $\varepsilon > 0$, we can find n sufficiently remote from the goal for which the error $h(n) - h^*(n)$ is below $\varepsilon h^*(n)$. Indeed, for this type of error both Pohl [7] and Gaschnig [8] have established a linear worst-case complexity, thus guaranteeing linear mean complexity.

6. Conclusions and Pending Problems

In light of the results reported in this paper, a natural question to ask is how accurate the estimates must be in order to guarantee a polynomial complexity. By a more elaborate manipulation of eq. (5) we have recently discovered [15] that the required accuracy is *logarithmic*, i.e., if the typical error increases faster than logarithmically in the distance to the goal then, regardless of the shape of the distribution functions, the mean complexity of A^* grows faster than N^k for any finite k. Moreover, if the typical error grows like $[\phi(h^*)]$, then $E(Z)$ would increase at the rate $\exp[c\phi(N)]$. Thus, highly precise heuristics must be devised if the search complexity is to be contained within reasonable growth rates. Most physical measurements are subject to a constant relative error, statistical inferences are usually characterized by $O(N^{1/2})$ error law (N is the number of random elements), and logarithmic precision is a rare commodity.

So far the average-case results obtained by a probabilistic analysis stand in striking similarity to those obtained by worst-case analysis. It is interesting to explore how widespread this similarity is in unstructured combinatorial search problems. One area where probabilistic analysis seems necessary is the treatment of non-admissible heuristics. Since most of the useful and the more accurate heuristics occasionally overestimate the distance to the goal, and since the magnitude of the overestimation error can be bounded only probabilistically, the relationship between accuracy and complexity in this case is of greater practical significance. This relationship is further complicated in the case of graphs where overestimation implies a finite likelihood of obtaining a suboptimal solution path. Probabilistic analysis may establish then the tradeoff between the mean search complexity and the mean increase in path length.

REFERENCES

1. Hart, P., Nilsson, N. and Raphael, B., A formal basis for the heuristic determination of minimum cost paths, *IEEE Trans. Systems Sci. Cybernet.* SSC-4, (2) (1968) 100–107.
2. Nilsson, N., *Problem-Solving Methods in Artificial Intelligence* (McGraw-Hill, New York, 1971).
3. Chang, C. and Slagle, J., An admissible and optimal algorithm for searching and/or graphs, *Artificial Intelligence* 2 (1971) 117–128.
4. Harris, L., Heuristic search under conditions of error, *Artificial Intelligence* 5(3) (1974) 217–234.
5. Martelli, A., On the complexity of admissible search algorithms, *Artificial Intelligence* 8 (1977) 1–13.
6. Gelperin, D., On the optimality of A^*, *Artificial Intelligence* 8 (1977) 69–76.

7. Pohl, I., Practical and theoretical considerations in heuristic search algorithms, In: E. Elcock and D. Michie (Eds.), *Machine Intelligence 8* (Ellis Horwood Ltd., Chichester, 1977).
8. Gaschnig, J., Performance measurement and analysis of certain search algorithms, Dissertation, Department of Computer Science, Carnegie–Mellon University (1979).
9. Pohl, I., First results on the effect of error in heuristic search, In: B. Meltzer and D. Michie (Eds.), *Machine Intelligence* 5 (Edinburgh University Press, Edinburgh, 1970).
10. Munyer, J., Some results on the complexity of heuristic search in graphs, Techn. Rept. H.P.-76-2, Information Sciences Department, University of California, Santa Cruz (September 1976).
11. Vanderbrug, G., Problem representations and formal properties of heuristic search, *Information Sci.* 11(4) (1976).
12. Nilsson, N., *Principles of Artificial Intelligence* (Tioga Publishing Company, Palo Alto, CA, 1980).
13. Harris, T., *The Theory of Branching Processes* (Springer, Berlin 1963).
14. Feller, W., *An Introduction to Probability Theory and its Applications*, Vol. 1 (Wiley, New York, 1968) 286.
15. Pearl, J., The utility of precision in search heuristics, UCLA-ENG-CSL-8065, Cognitive Systems Laboratory, University of California, Los Angeles (October 1980).

Received July 1980; revised version received November 1980