Probabilistic Recurrence Relations

Richard M. Karp[1,2]

TR-91-019

March, 1991

Abstract

This paper is concerned with recurrence relations that arise frequently in the analysis of divide-and-conquer algorithms. In order to solve a problem instance of size x, such an algorithm invests an amount of work $a(x)$ to break the problem into subproblems of sizes $h_1(x), h_2(x),\ldots,h_k(x)$ and then proceeds to solve the subproblems. Our particular interest is in the case where the sizes $h_i(x)$ are random variables; this may occur either because of randomization within the algorithm or because the instances to be solved are assumed to be drawn from a probability distribution. When the h_i are random variables the running time of the algorithm on instances of size x is also a random variable $T(x)$. We give several easy-to-apply methods for obtaining fairly tight bounds on the upper tails of the probability distribution of $T(x)$ and present a number of typical applications of these bounds to the analysis of algorithms. The proofs of the bounds are based on an interesting analysis of optimal strategies in certain gambling games.

---

[1]Supported by NSF Grant No. CCR-9005448.

[2]University of California at Berkeley and International Computer Science Institute, Berkeley, California.

1 Introduction

This paper is concerned with a class of stochastic processes that arise frequently in the analysis of algorithms. Such a process can be described succinctly by a recurrence relation of the form $T(x) = a(x)+T(h(x))$, where x is a nonnegative real variable, $a(x)$ is a nonnegative real-valued function of x and $h(x)$ is a random variable ranging over $[0, x]$ and having expectation less than or equal to $m(x)$, where m is a nonnegative real-valued function.

Stochastic processes of this kind typically arise in connection with recursive algorithms that obey the following description: "To process an input of size x, invest an amount of computational effort $a(x)$ and then recursively solve a derived instance of the same problem, having size $h(x)$." Here $a(x)$ is a fixed quantity but $h(x)$, the size of the derived problem instance, is a random variable whose expectation is known to be less than or equal to $m(x)$. Such processes also arise in other settings; for example, in studying the number of cycles in a random permutation. We also consider a more general stochastic process corresponding to the probabilistic recurrence relation $T(x) = a(x) + \sum_{j=1}^{k} T(h_j(x))$, where $a(x)$ is a given function and the h_j are (not necessarily independent) random variables. Our analysis of this more general recurrence relation is based on an interesting gambling game related to games studied in [DuSa] and [Fr].

The purpose of this paper is to provide a simple "cookbook method" for obtaining fairly tight bounds on the upper tail of the distribution of $T(x)$. These bounds can be used to replace the bounds that are derived by *ad hoc* (and often imprecise) arguments throughout the computer science literature.

In the case of the probabilistic recurrence relation $T(x) = a(x) + T(h(x))$ these bounds are expressed in terms of the functions $a(x)$ and $m(x)$, and do not depend on any information about the distribution of $h(x)$. Throughout, we assume that $0 \leq m(x) \leq x$ for all x, and $a(x)$, $m(x)$ and $\frac{m(x)}{x}$ are nondecreasing functions. The equation $\tau(x) = a(x) + \tau(m(x))$ can be viewed as a deterministic counterpart of the given probabilistic recurrence relation, corresponding to the case where, for all x, the random variable $h(x)$ is deterministically equal to its expectation. Whenever this equation has a nonnegative solution it has a unique least nonnegative solution $u(x)$ (i.e., a nonnegative solution such that, for every nonnegative solution $v(x)$, $u(x) \leq v(x)$ for all x). The function $u(x)$ is given explicitly by the formula $u(x) = \sum_{i=0}^{\infty} a(m^{[i]}(x))$, where $m^{[i]}(x)$ is defined inductively by: $m^{[0]}(x) = x$; for $i = 1, 2, \ldots, m^{[i]}(x) = m(m^{[i-1]}(x))$. Throughout the paper, $u(x)$ denotes the least nonnegative solution of $\tau(x) = a(x) + \tau(m(x))$.

Theorem 1 *Suppose there is a constant d such that $a(x) = 0$, $x < d$ and $a(x) = 1, x \geq d$. Let $c_t = \min\{x | u(x) \geq t\}$. Then, for every positive real x and every positive integer w, $Prob[T(x) \geq u(x) + w] \leq (\frac{m(x)}{x})^{w-1} \frac{m(x)}{c_{u(x)}}$.*

We illustrate Theorem 1 in the following specific case, which comes up frequently in applications: $m(x) = px$, where p is a positive constant less than 1, $a(x) = 0$, $x < 1$, $a(x) = 1$, $x \geq 1$. Let b denote $1/p$. Then $u(x) = 0$, $x < 0$, $u(x) = \lfloor \log_b(x) \rfloor + 1$, $x \geq 1$ and $c_t = b^{t-1}$. Theorem 1 yields the following: for $x \geq 1$ and w a positive integer, $Prob[T(x) \geq \lfloor \log_{b(x)} \rfloor + w + 1] \leq p^{w-1} \frac{x}{b^{\lfloor \log_b(x) \rfloor + 1}}$.

Theorem 2 *Suppose that $a(x)$ is a nondecreasing, continuous function which is strictly increasing on $\{x | a(x) > 0\}$, and $m(x)$ is a continuous function. Then, for every positive real x and every positive integer w, $Prob[T(x) \geq u(x) + wa(x)] \leq (\frac{m(x)}{x})^w$.*

In formulating Theorems 1 and 2 we assumed that, in processing an instance of size x, the distribution of $h(x)$, the size of the derived problem instance, depended only on x. We can also consider the more general situation in which the distribution of the size of the derived problem instance may depend not just on the size of the original instance, but on the instance itself. To describe this more general situation, we assume that the function T has an arbitrary domain I (the set of *instances*), and that a function *size* from I into the nonnegative reals is given. We continue to assume that the size of the derived instance is nonnegative and bounded above by the size of the original instance, that the immediate work in processing an instance of size x is $a(x)$, and that, if x is the size of the original instance, then the expected size of the derived instance is less than or equal to $m(x)$; however, the distribution of the derived instance $h(z)$ may depend on the original instance z, not merely on its size. This more general situation can be represented by the probabilistic recurrence relation $T(z) = a(size(z)) + T(h(z))$, where z ranges over instances, $0 \leq size(h(z)) \leq size(z)$ and $E[size(h(z))] \leq m(size(z))$. Theorems 1 and 2 are easily extended to the more general setting, as follows.

Theorem 3 *Suppose there is a constant d such that $a(x) = 0$, $x < d$ and $a(x) = 1, x \geq d$. Let $c_t = \min\{x | u(x) \geq t\}$. Then, for every positive real x, every instance z of size x, and every positive integer w, $Prob[T(z) \geq u(x) + w] \leq (\frac{m(x)}{x})^{w-1} \frac{m(x)}{c_{u(x)}}$.*

Theorem 4 *Suppose that $a(x)$ is a nondecreasing, continuous function which is strictly increasing on $\{x | a(x) > 0\}$, and $m(x)$ is a continuous function. Then,*

for every positive real x, every instance z of size x and every positive integer w, Prob$[T(z) \geq u(x) + wa(x)] \leq (\frac{m(x)}{x})^w$.

Theorems 3 and 4 are proved in Section 3.

Our final main theorem concerns the probabilistic recurrence relation $T(z) = a(z) + \sum_{i=1}^{k} T(h_i(z))$. Here z ranges over an arbitrary set I of instances, and, for each z, the $h_i(z)$ are random variables ranging over I and having a joint distribution determined by z.

Theorem 5 *Suppose that, for all z, and for all possible joint values $(y_1, y_2, \ldots, y_k)$ of the k-tuple $(h_1(z), h_2(z), \ldots, h_k(z))$, $E[T(z)] \geq \sum_{i=1}^{k} E[T(y_i)]$. Then, for all z and all positive a, Prob*$[T(z) \geq (a+1)E[T(z)] < \exp(-a)]$.

Theorem 5 is proved in Section 4.

Note that this theorem is of a different character than the preceding ones, since it assumes that, for each instance z, the joint distribution of the derived instances $h_i(z)$ is given precisely, and thus the distribution of the execution time $T(z)$ is completely determined. The previous theorems did not assume a precise specification of the size of the derived instance, but merely required information about its expectation and range of possible values. The theorem requires the strong hypothesis that the total expected cost of processing the k instances derived from z can never exceed the expected cost of processing z; i.e., that the expected value of the amount of work remaining can never increase.

Although the proofs of the theorems differ in detail, there is a common point of view that underlies them. We illustrate this point of view with respect to the recurrence $T(x) = a(x) + T(h(x))$. This recurrence specifies a sequence $\{X_i\}$ of random variables, where X_0 is the size of the original problem and, X_i is the size of the ith derived problem. Thus $X_0 = x$, $0 \leq X_{i+1} \leq X_i$, and $E[X_{i+1}|X_0, X_1, \ldots, X_i] \leq m(X_i)$. The value of the random variable $T(x)$ is $\sum_{i=0}^{\infty} a(X_i)$; call this quantity the *payoff*. Now consider an *adversary* whose goal is to maximize the probability of achieving a payoff greater than or equal to some specific value V. Suppose the adversary has the freedom to specify the distributions of $X_1, X_2, \ldots$ sequentially, and gets to observe the values of $X_0, X_1, \ldots, X_i$ before specifying the distribution of X_{i+1}; However, this distribution must be restricted to the interval $[0, X_i]$ and must have expectation less than or equal to $m(X_i)$. In gambling terms, the adversary can choose the lottery he wishes to play at each step, knowing the outcomes of all his previous plays. Consider the step in which the adversary chooses the distribution of X_{i+1}. Let $V' = V - \sum_{j=1}^{i-1} a(X_j)$; this represents the payoff so far. The adversary's goal is to maximize the probability that the remaining payoff, $\sum_{j=i}^{\infty} a(X_j)$, is greater than or equal to V'. If $V' \leq u(X_i)$ then the adversary's task is trivial; he simply chooses the remaining random variables deterministically as follows: $X_{j+1} = m(X_j)$, for $j = i, i+1, \ldots$. We call this *timid play*. However, if $V' > u(X_i)$ then timid play will not work. It can be proven that in this case the adversary should resort to *bold play*, as follows.

- Case I: $u(x) \leq V' - a(x)$. The adversary chooses the following distribution: $X_{i+1} = x$ with probability $\frac{m(x)}{x}$ and 0 with probability $1 - \frac{m(x)}{x}$.
- Case II: $u(x) \geq V' - a(x)$. The adversary chooses the following distribution: $X_{i+1} = c_{V'-a(x)}$ with probability $\frac{m(x)}{c_{V'-a(x)}}$ and 0 with probability $1 - \frac{m(x)}{c_{V'-a(x)}}$.

The probability that $T(x) \geq V$ is bounded above by the probability that the adversary can gain a payoff of at least V by employing timid play when he has a "sure win," and bold play otherwise. Theorems 1 and 2 follow from the observation that this mixture of timid play and bold play is optimal for the adversary.

2 Applications

In this section we give a number of typical applications of our theorems. In all these applications the size of a problem instance is restricted to integer values, rather than general real values. Accordingly, we adopt the notational convention of using variables such as n or m to denote instance sizes, rather than using the symbol x (connoting a real quantity) as we do in developing the general theory.

2.1 Randomized List Ranking

In the analysis of a randomized parallel list ranking algorithm, the following iterative process occurs. At each step, one has a circular list of elements. If the list is of length 1 then the sole element in the list is deleted and the process terminates. If the list contains more than one element then each element independently chooses a sex uniformly at random from the set $\{Male, Female\}$, and each Female list element whose predecessor in the list is Male, is deleted. The process is iterated until the list becomes empty. At each iteration, the computational work can be taken to be the number of elements in the circular list. Let $T(n)$ denote the number of steps required to reduce an n-element list to an empty list, and let $T'(n)$ denote the total work in reducing an n-element list to the empty list, where the work at each iteration is taken to be the number of elements in the list at that iteration. Then $T(n) = 1 + T(h(n))$, with the initial condition $T(1) = 1$ and $T'(n) = n + T'(h(n))$, with the initial condition $T'(1) = 1$, where $m(n) = E[h(n)] = \frac{3n}{4}$.

The following bounds on the upper tails of the distributions of $T(n)$ and $T'(n)$ follow from Theorems 1 and 2:

- For every positive integer w, $Prob[T(n) \geq w + 1 + \lfloor \log_{4/3} n \rfloor] \leq (\frac{3}{4})^{w-1} \frac{n}{(\frac{4}{3})^{\lfloor \log_{4/3} n \rfloor + 1}}$.
- For every positive integer w, $Prob[T'(n) \geq (4 + w)n] \leq (\frac{3}{4})^w$.

2.2 Randomized Tree Contraction

The paper[MiRe] presents an algorithm called *Randomized Tree Contraction* that starts with an n-node tree representing an arithmetic expression and repeatedly applies a randomized operation called *contraction* which produces a new tree, representing a modified arithmetic expression. The process eventually reaches a tree with one node, and it terminates one step later. The work performed in a contraction step can be taken to be the number of nodes in the tree. Miller and Reif show that, when randomized tree contraction is applied to a tree with n nodes, the expected number of nodes in the resulting tree is at most $4n/5$; the distribution of the number of nodes in the resulting tree may depend on the particular n-node tree being considered. Define the size of a tree to be the number of nodes it contains. Since the distribution of the size of the derived problem depends on the particular original instance as well as its size, we need to apply Theorems 3 and 4, rather than Theorems 1 and 2. Let the random variables $T(z)$ and $T'(z)$ respectively denote the number of iterations and the total work when the process is applied to a tree z. We obtain:

For every positive integer n, every tree z of size n and every positive integer w

- $Prob[T(z) \geq w + 1 + \lfloor \log_{5/4} n \rfloor] \leq (\frac{4}{5})^{w-1} \frac{n}{(\frac{5}{4})^{\lfloor \log_{5/4} n \rfloor + 1}}$
- $Prob[T'(z) \geq (5 + w)n] \leq (\frac{4}{5})^w$.

2.3 Maximal Independent Set

The paper [Lu] gives a randomized parallel algorithm for constructing a maximal independent set of vertices in a graph. The algorithm is iterative. Each iteration deletes some of the edges of the current graph, and the iterative process is completed when all of the edges have been deleted. The work at each iteration can be taken to be the number of edges in the current graph. Luby showed that, at each iteration step, the expected number of edges deleted is at least one-eighth of the number of edges in the current graph. Let $T(G)$ be the number of iterations, and $T'(G)$, the amount of work, in executing Luby's algorithm on a graph G. Theorems 3 and 4, respectively, yield:

- For every n-edge graph G and every positive integer w, $Prob[T(G) \geq w + 1 + \lfloor \log_{8/7} n \rfloor] \leq (\frac{7}{8})^{w-1} \frac{n}{(\frac{8}{7})^{\lfloor \log_{8/7} n \rfloor + 1}}$.
- For every n-edge graph G and every positive integer w, $Prob[T'(G) \geq (8 + w)n] \leq (\frac{7}{8})^w$.

2.4 Random Permutations

By a *random permutation* of the elements $1, 2, \ldots, n$ we mean a permutation drawn from the uniform distribution over S_n, the set of all permutations of $\{1, 2, \ldots, n\}$. A *cycle of length* t in the permutation σ is a sequence $i_0, i_1, \ldots, i_{t-1}$ of distinct elements such that, for $j = 1, 2, \ldots, t$, $i_{(j+1) \bmod t} = \sigma(i_j)$. Any permutation of a finite set partitions the elements into disjoint cycles. Let the random variable $T(n)$ denote the number of cycles in a random permutation of n elements. It is known that $T(n)$ satisfies the probabilistic recurrence relation $T(n) = 1 + T(h(n))$, $T(0) = 0$, where $h(n)$ is uniformly distributed over $\{0, 1, \ldots, n-1\}$. We apply Theorem 1 with $m(x) = \frac{x-1}{2}$, $x \geq 1$, $m(x) = 0$, $x \leq 1$. In this case $u(x) = \lfloor \lg(x+1) \rfloor$ and $c_t = 2^t - 1$. Hence, $Prob[T(n) \geq \lfloor \lg(n+1) \rfloor + w] \leq (\frac{n-1}{2n})^{w-1} (\frac{n-1}{2^{\lfloor \lg(n+1) \rfloor + 1} - 1}) \leq 2^{-(w-1)}$. In particular, the probability that $T(n) \geq (a+1)\lg(n+1)$ is bounded above by $(n+1)^{-a}$. As might be expected, a sharper bound can be obtained by taking into account the precise distribution of $h(n)$, rather than merely its expectation and the fact that it lies between 0 and n. It can be shown that $Prob[T(n) \geq (a+1)\lg(n+1)]$ is bounded above by a function of the form $n^{-\Omega(a \ln a)}$.

2.5 A Selection Algorithm

Hoare has given the following algorithm for finding the kth-smallest element of an n-element set S:

Choose a random element $r \in S$;
by comparing each element of $S - \{r\}$ with r, partition $S - \{r\}$ into two sets, L and U, where $L = \{y \in S : y < r\}$ and $U = \{y \in S : y > r\}$;
if $|L| \geq k$ then, recursively, find the kth-smallest element of L;
if $|L| = k - 1$ then return r;
if $|L| < k - 1$ then, recursively, find the $k - 1 - |L|$th-smallest element of U.

The partitioning step requires $n - 1$ comparisons. It can be shown that, for all n and k, the expected size of the problem that remains after applying the partitioning step to an n-element set is at most $\frac{n+1}{2}$. Thus we can apply Theorem 4 with $m(x) = \frac{x+1}{2}$, giving $u(x) = 2(x - 1)$. Let the random variable $T(n, k)$ denote the number

of comparisons required for the algorithm to determine the kth-smallest element of an n-element set. Theorem 4 gives: for any positive integer w, $Prob[T(n,k) \geq 2n + w(n-1)] \leq (\frac{n+1}{2n})^w$.

2.6 A Greedy Clique Algorithm

The following algorithm can be used to construct a maximal clique in a graph with vertex set V.

begin $Q \leftarrow \phi$; $S \leftarrow V$;
while $S \neq \phi$ **do**
    begin
    choose a random element v from S;
    $Q \leftarrow Q \cup \{v\}$;
    delete from S the vertex v and all vertices in S that are not adjacent to v
    end
return Q
end

Consider the behavior of this algorithm when applied to a random graph, in which each edge independently is present with probability p. At a step in which the cardinality of S is m, the expected number of vertices that do not get deleted is $p(m-1)$. Let $T(n)$ denote the size of the clique obtained when the algorithm is applied to a n-vertex random graph. Then $T(n) = 1 + T(h(n)), T(1) = 1$, where $0 \leq h(n) \leq n-1$ and $m(n) = E[h(n)] = p(n-1)$. Let $T'(n)$ denote the number of steps in the algorithm, where each step consists of testing whether two given vertices are adjacent. Then $T'(n) = n - 1 + T'(h(n))$, where $m(n) = E[h(n)] = p(n-1)$. Let $b = 1/p$. Theorems 1 and 2 give: $Prob[T(n) \geq \lfloor \log_b(n(b-1)+1) \rfloor + w] \leq (\frac{p(n-1)}{n})^{w-1} \frac{(b-1)(n-1)}{b(b^{\lfloor (\log_b((b-1)n+1)) \rfloor} - 1)}$ and $Prob[T'(n) \geq \frac{n-1}{1-p} + w(n-1)] \leq (\frac{p(n-1)}{n})^w$. The first of these results is not tight; using more information about the structure of the problem, one can show that $Prob[T(n) \geq \lceil \log_b(x(b-1)+1) \rceil + w] \leq p^{2w}\Theta(log_b n)$.

2.7 Abstract Independence Systems

The paper [KaUpWi] gives two parallel randomized algorithms for finding a maximal independent set in an abstract independence system. The first algorithm uses an oracle for testing whether a given set is independent, and the second algorithm uses a more powerful oracle called a *rank oracle*. Let $T_1(n)$ denote the number of iterations required by the first algorithm when applied to an n-element independence system, and let $T_2(n)$ be the number of iterations required by the second algorithm when applied to such a system. It follows from results in [KaUpWi] that $T_1(n) = 1 + T_1(h_1(n))$, where $0 \leq h_1(n) \leq n$ and $m_1(n) = E[h_1(n)] \leq n - c\sqrt{n}$, where c is an absolute constant, and $T_2(n) = 1 + T_2(h_2(n))$ where $m_2(n) = E[h_2(n)] \leq n - \frac{n}{H_n}$; here H_n is the nth harmonic number. Let $u_1(n)$ be the minimal nonnegative solution of $u_1(n) = 1 + u_1(n - c\sqrt{n})$. It is easily checked that $u_1(n) \leq \frac{2\sqrt{n}}{c}$, and it follows from Theorem 2 that $Prob[T_1(n) \geq \frac{2\sqrt{n}}{c} + w] \leq (\frac{n-c\sqrt{n}}{n})^{w-1} \leq \exp(-c(w-1)n^{-1/2})$. Let $u_2(n)$ satisfy $u_2(n) = 1 + u_2(n - \frac{n}{H_n})$. Then, for any value $d > 1/2$, $u_2(n) \leq d\ln^2(n)$, when n is sufficiently large. It follows from Theorem 1 that, for n sufficiently large, $Prob[T_2(n) \geq d\ln^2 n + w] \leq (1 - \frac{1}{H_n})^w \leq \exp(-\frac{w}{H_n})$.

2.8 Quicksort

To sort an n-element set S, where $n > 1$, Quicksort selects a random element $\hat{x}$ and compares each of the other elements with $\hat{x}$, thereby partitioning S into L, the set of elements less than $\hat{x}$, U, the set of elements greater than $\hat{x}$, and the singleton set $\{\hat{x}\}$. Then, recursively, Quicksort completes the sorting process by sorting L and U. Let the random variable $T(n)$ denote the number of comparisons required by Quicksort to sort a set of n elements; we may use this notation, which does not mention the particular n-element set, because the distribution of the number of comparisons depends only on the cardinality of the set. Then $T(n)$ satisfies the probabilistic recurrence relation $T(n) = n - 1 + T(h_1(n)) + T(h_2(n))$, where $h_1(n)$ denotes the cardinality of L and $h_2(n)$ denotes the cardinality of U. Here $h_1(n)$ and $h_2(n)$ each have the uniform distribution over $\{0, 1, \ldots, n-1\}$, and they satisfy the equation $h_1(n) + h_2(n) = n - 1$. It is known that $E[T(n)] = 2((n+1)(H_n - 1) - (n-1))$, where H_n is the nth harmonic number. The conditions of Theorem 5 are satisfied, and thus we conclude that $Prob[T(n) \geq (a+1)E[T(n)]] \leq \exp(-a)$.

3 Proofs of Theorems 3 and 4

Lemma 1 *Let X be a random variable ranging over some interval $[0, x]$. Let f be a nonnegative real-valued function over the nonnegative reals such that there is a real number b with the property that*

- *On the interval $[0, b]$, $\frac{f(x)}{x}$ is a nondecreasing function;*
- *For all $y \geq b$, $f(y) = 1$;*
- *$E[X] \leq \min(b, x)$.*

Then $E[f(X)] \leq \frac{E[X]f(\min(x,b))}{\min(x,b)}$.

Proof: Let X have the cumulative distribution function F. Then $E[f(X)] = \int_0^x f(y)dF(y)$. But, for all y,

the hypotheses imply that $f(y) \leq \frac{yf(\min(b,x))}{\min(b,x)}$. Hence $E[f(X)] \leq \frac{f(\min(b,x))}{min(b,x)} \int_0^x y dF(y) = \frac{E[x]f(\min(b,x))}{\min(b,x)}$. □

We shall now prove Theorem 3 and a slight generalization of Theorem 4. Theorems 1 and 2 will follow as corollaries. Theorems 3 and 4 are concerned with the probabilistic recurrence relation $T(z) = a(size(z))+T(h(z))$, where z denotes a generic instance, $size(z)$ denotes its size, $0 \leq size(h(z)) \leq size(z)$ and $E[size(h(z)] \leq m(size(z))$. We assume that $a(x)$ and $\frac{m(x)}{x}$ are nonnegative nondecreasing functions.

The following theorem is a restatement of Theorem 3.

Theorem 6 *Suppose there is a constant d such that $a(x) = 0$, $x < b$ and $a(x) = 1, x \geq b$. Let $c_t = \min\{x|u(x) \geq t\}$. Then, for every positive integer r and every instance z, $Prob[T(z) \geq r] \leq D_r(size(z))$, where: $D_1(x) = 1$ if $x \geq d$ and $D_1(x) = 0$ if $x < d$; for $r = 2, 3, \ldots$, $D_r(x) = 1$ if $u(x) \geq r$ and $D_r(x) = (\frac{m(x)}{x})^{r-1-u(x)} \frac{m(x)}{c_{u(x)}}$ if $u(x) < r$.*

Proof: We shall verify that, over the interval $[0, c_r]$, $\frac{D_r(x)}{x}$ is a nondecreasing function. Note that $D_r(x)$ is a nondecreasing function and that $D_r(x) < 1$ if and only if $x < c_r$. Consider x_1 and x_2 such that $0 \leq x_1 \leq x_2 \leq c_r$. We need to show that $\frac{m(x_1)^{r-u(x_1)}}{x_1^{r-u(x_1)} c_{u(x_1)}} \leq \frac{m(x_2)^{r-u(x_2)}}{x_2^{r-u(x_2)} c_{u(x_2)}}$. We shall use the facts that $0 \leq m(x) \leq x$ and $c_r = m(c_{r+1})$. Since $\frac{m(x_1)}{x_1} \leq \frac{m(x_2)}{x_2}$ it suffices to prove that $(\frac{m(x_1)}{x_1})^{(u(x_2)-u(x_1)} \leq \frac{c_{u(x_1)}}{c_{u(x_2)}}$. This clearly holds when $u(x_1) = u(x_2)$. When $u(x_1) < u(x_2)$, $c_u(x_1) \leq m(c_u(x_2))$ and we have $(\frac{m(x_1)}{x_1})^{(u(x_2)-u(x_1)} \leq (\frac{m(x_1)}{x_1}) \leq \frac{m(c_{u(x_2)})}{c_{u(x_2)}} \leq \frac{c_{u(x_1)}}{c_{u(x_2)}}$, completing the verification.

Let $S_r(z) = Prob[T(z) \geq r]$. We shall prove by induction on r that, for all instances z and all positive integers r, $S_r(z) \leq D_r(size(z))$. This will complete the proof of the theorem. The case $r = 1$ is immediate. Assuming the result holds for r, we prove it for $r + 1$. If $u(size(z)) \geq r + 1$ then $D_{r+1}(size(z)) = 1$ and the result is immediate. Assume that $u(size(z)) \leq r + 1$. Then $S_{r+1}(z) = E[S_r(h(z))]$. By the induction hypothesis, $S_r(y) \leq D_r(size(y))$ for all instances y. It follows that $S_{r+1}(z) \leq E[D_r(size(h(z)))]$. By Lemma 1, together with the facts that $0 \leq size(h(z)) \leq size(z)$ and $E[size(h(z))] \leq m(size(z))$ we obtain

1. If $c_r \leq size(z)$ then $S_{r+1}(z) \leq \frac{m(x)}{c_r}$.
2. If $size(z) < c_r$ then $S_{r+1}(z) \leq \frac{m(size(z))}{size(z)} D_r(size(z))$.

In each of the two cases, the right-hand-side of the expression is equal to $D_{r+1}(size(z))$. □.

The following theorem concerns the case where $a(x)$ is a nondecreasing, continuous function which is strictly increasing on $\{x|a(x) > 0\}$, and $m(x)$ is a continuous function. Then $u(x)$ is a nondecreasing, continuous function which is strictly increasing on $\{x|a(x) > 0\}$, and thus the inverse function $u^{-1}(t)$ is well defined for all positive t in the range of u. The theorem is a slight extension of Theorem 4.

Theorem 7 *For all $r > 0$ and all instances z, let $s_r(z) = Prob[T(z) \geq r]$. Then $s_r(z) \leq d_r(size(z))$, where:*

- *if $u(x) \geq r$ then $d_r(x) = 1$;*
- *if $u(x) < r$ then $d_r(x) = (\frac{m(x)}{x})^{\lceil \frac{r-u(x)}{a(x)} \rceil} \frac{x}{u^{-1}(r-a(x)\lceil \frac{r-u(x)}{a(x)} \rceil)}$.*

Proof: Clearly $s_r(z) = E[s_{r-a(size(z))}(h(z))]$. We introduce a nonnegative integer iteration index i and define the sequence of functions $\{s_r^i\}$ as follows: $s_r^0(z) = 1$ if $r \leq u(size(z))$ and 0 if $r > u(size(z))$; for $i = 0, 1, \ldots, s_r^{i+1}(z) = E[s_{r-a(size(z))}^i(h(z))]$. Then $s_r(z) = \sup_i s_r^i(z)$ and it suffices to prove that, for all r, i and z, $s_r^i(z) \leq d_r(size(z))$. The proof will be by induction on i.

In preparation for the inductive proof we observe that, in the range $0 \leq x \leq u^{-1}(r)$, the function $\frac{d_r(x)}{x}$ is nondecreasing. To see this, first note that, since $u(x)$ and $a(x)$ are increasing functions, $\frac{r-u(x)}{a(x)}$ is a decreasing function. Thus, the function $\lceil \frac{r-u(x)}{a(x)} \rceil$ is a nonincreasing function, and the interval $[0, u^{-1}(r)]$ can be decomposed into subintervals such that, within each subinterval, $\lceil \frac{r-u(x)}{a(x)} \rceil$ is constant. Consider the subinterval within which $\lceil \frac{r-u(x)}{a(x)} \rceil$ is equal to the constant k. Within this subinterval the function $\frac{d_r(x)}{x}$ is equal to $\left(\frac{m(x)}{x}\right)^{k+1} \frac{1}{u^{-1}(r-ka(x))}$, and it is clear, since the functions $\frac{m(x)}{x}$, $a(x)$ and $u_{-1}(x)$ are all nondecreasing, that the function $\frac{d_r(x)}{x}$ is nondecreasing. To complete the proof it suffices to show that the function $\frac{d_r(x)}{x}$ is continuous at the boundary points between subintervals. Such a boundary point x has the property that $\frac{r-u(x)}{a(x)}$ is equal to a positive integer k. In the subinterval to the left of x the function $\frac{d_r(x)}{x}$ is equal to $\left(\frac{m(x)}{x}\right)^{k+1} \frac{1}{u^{-1}(r-ka(x))}$, and in the subinterval to the right of x, the function $\frac{d_r(x)}{x}$ is equal to $\left(\frac{m(x)}{x}\right)^{k} \frac{1}{u^{-1}(r-(k-1)a(x))}$. But, noting that $u(x) = r - ka(x)$ we see that $u^{-1}(r - ka(x)) = x$ and $u^{-1}(r-(k-1)a(x)) = m(x)$. Thus the two expressions are equal at x, establishing continuity.

We now carry out the inductive proof. The case $i = 0$ is immediate. For the induction step, recall

that $s_r^{i+1}(z) = E[s^i_{r-a(size(z))}(h(z))]$. By induction hypothesis $s^i_{r-a(z)}(h(z)) \leq d_{t-a(z)}(size(h(z))$. Hence $s_r^{i+1}(z) \leq E[d_{r-a(size(z))}size(h(z))]$. By Lemma 1, $E[d_{r-a(size(z))}(h(z))] \leq d_r(size(z))$. This completes the induction step. □

4 A Gambling Game

The proof of Theorem 5 hinges on the following one-person gambling game. The player starts with a fortune of 0 and an investment capital of 1 unit. As the game progresses, he gradually invests his capital and his fortune increases. His goal is to acquire a fortune of at least $1 + a$, where a is a nonnegative real constant. At any given step the player has an amount of capital c remaining, and he opts to increase his fortune by r, where $0 \leq r \leq c$. His capital at the next step is a random variable c', where $0 \leq c' \leq c$ and $E[r + c'] = c$. Thus, at each step, the expected decrease in the player's capital is equal to the amount by which his fortune increases. Our main result is as follows: the probability that the player can achieve his goal is less than $\exp(-a)$. Related theorems can be found in [DuSa] and [Fr].

We formalize the game in terms of a stochastic process defined by two sequences $\{X_n\}$ $\{Y_n\}$ of nonnegative random variables. The random variable X_n is called the *fortune after step n*, and Y_n is called the *capital remaining after step n*. Y_0 is called the *initial capital*. We require the following properties.

- The sequence $\{X_n\}$ is monotone nondecreasing.
- The sequence $\{Y_n\}$ is monotone nonincreasing.
- $X_0 = 0$
- X_n is completely determined by $X_0, Y_0, X_1, Y_1, \ldots, X_{n-1}, Y_{n-1}$.
- For all n, $E[Y_{n+1}|X_0, Y_0, X_1, Y_1, \ldots, X_n, Y_n, X_{n+1}] \leq Y_n + X_n - X_{n+1}$.

Let X, the *eventual fortune*, be defined as $\sup_{n\to\infty} X_n$.

Theorem 8 *For all $a > 0$, $Prob[X \geq (a + 1)Y_0] \leq \exp(-a)$.*

Proof: For $a \geq 0$ and all nonnegative integers n, let $R_n(a)$ be the supremum, over all stochastic processes satisfying the above conditions, of the probability that $X_n \geq (1+a)Y_0$. We shall prove by induction on n that, for all n and a, $R_n(a) \leq \exp(-a)$. The case $n = 0$ is immediate, since $X_0 = 0$. The induction step is based on the following recurrence, in which the variable r ranges over the positive reals, and the variable F ranges over cumulative distribution functions of probability distributions over $[0, 1]$ with mean $(1 - r)$:

$$R_{n+1}(a) = sup_{(r,F)} \quad \int_0^{(a+1-r)} R_n\left(\tfrac{a+1-r-x}{x}\right) dF(x) + \int_{(a+1-r)}^1 dF(x)$$

The recurrence is justified as follows. We can assume without loss of generality that $Y_0 = 1$. The quantity r denotes a possible choice of X_1 and F denotes a possible choice for the distribution of Y_1, given that $X_1 = r$. Consider the case where $Y_1 = x$. Then X_{n+1}, the cumulative reward after $n + 1$ steps, will be greater than or equal to the goal of $a + 1$ if and only if the cumulative reward received at steps 2 through $n + 1$ is at least $a + 1 - r$. Also, the capital available at the second step is x. Thus we are interested in the supremum, over all policies, of the probability of gaining a cumulative reward of $a + 1 - r$ in n steps, starting with a capital of x. This quantity is equal to 1 when $\frac{a+1-r}{x} \leq 1$, and is equal to $R_n(\frac{a+1-r-x}{x})$ when $\frac{a+1-r}{x} \geq 1$. With these observations, the recurrence follows from the theorem of total probability (also known as the unconditioning principle). Using the inductive hypothesis that $R_n(a) \leq \exp(-a)$ we obtain

$$R_{n+1}(a) \leq sup_{r,F} \quad (\int_0^{a+1-r} \exp(-\tfrac{a+1-r-x}{x}) dF(x) + \int_{a+1-r}^1 dF(x)).$$

for a given choice of r, we wish to determine the cumulative distribution function F that maximizes the objective functional $\int_0^1 I(x)dF(x)$ subject to $\int_0^1 xdF(x) = 1 - r$, where $I(x) = \exp(-\frac{a+1-r-x}{x})$ when $x \leq a + 1 - r$ and $I(x) = 1$ when $x \geq a + 1 - r$. Note that I(x) is strictly increasing in $[0, a + 1 - r]$ and constant in $[a + 1 - r, 1]$. Applying Lemma 1 we find that the maximum value of the objective functional is $\frac{1-r}{a+1-r}$ if $r \geq a$ and $(1 - r)\exp(-(a - r)$ if $r \leq a$. It is an easy exercise (using the inequality $\exp(y) \geq 1 + y$ for all real y) to show that, for all nonnegative r, this maximum value is less than or equal to $\exp(-a)$. □

Theorem 9 *For all $a > 0$, $Prob[X \geq a+1] \leq \exp(-a)$*

Proof: Assume for contradiction that there is a policy for which $Prob[X \geq a + 1] > \exp(-a)$. Then there exists an n such that $Prob[X_n \geq a + 1] > \exp(-a)$, contradicting Theorem 2. □

This result cannot be strengthened, as the following example illustrates. Let a be given, and let r be a positive constant. Let the joint distributions of the random variables X_n and Y_n, $n = 0, 1, \ldots$ be defined as follows: $X_0 = 0$, $Y_0 = 1$; for all n, one of the following situations holds.

- $Y_n = 0$. In this case, for all $j > n$, $(X_j, Y_j) = (X_n, Y_n)$, and $X = X_n$.
- $X_n < a$, $Y_n = 1$. in this case, $X_{n+1} = X_n + r$; with probability r, $Y_{n+1} = 0$ and, with probability $1-r$, $Y_{n+1} = 1$.
- $X_n \geq a$, $Y_n = 1$. In this case, $X_{n+1} = 1 + a$.

It is easy to check that $Prob[X \geq 1 + a] = (1-r)^{\lceil \frac{a}{r} \rceil}$. This probability approaches $\exp(-a)$ as r tends to zero.

We now indicate how Theorem 7 can be applied to probabilistic recurrence relations. Consider a probabilistic recurrence relation of the form $T(z) = A(z) + \sum_{i=1}^{k} T(h_i(z))$, where

- $A(z)$ is a function;
- for each instance z the $h_i(z)$ are random variables whose joint distribution is completely determined by z;
- for all z, and all possible joint values $y_1, y_2, \ldots, y_k$ of the random variables $h_1(z), h_2 z), \ldots, h_k(z)$, $\sum_{i=1}^{k} E[T(y_i)] \leq E[T(z)]$.

The last condition expresses the strong hypothesis that the expected cost of processing the k instances derived from an instance z can never exceed the expected cost of proceeding z.

Corollary 1 $Prob[T(z) \geq (a+1)E[T(z)]] \leq \exp(-a)$.

Proof: The process of sampling from the distribution of $T(z)$ by numerically "unwinding" the given recurrence can be regarded as an instance of the gambling game. Initially, the gambler's fortune is zero and his capital is $E[T(z)]$. At the first step his fortune is increased by $A(z)$. To determine the reduction in his capital he draws a sample $(y_1, y_2, \ldots, y_k)$ from the joint distribution of $h_1(z), h_2(z), \ldots, h_k(z)$, his capital becomes $\sum_{i=1}^{k} E[T(y_i)]$ and the arguments $y_1, y_2, \ldots, y_k$ are placed on a stack for later evaluation. A general step corresponds to taking an argument y off the stack and drawing a sample $w_1, w_2, \ldots, w_k$ from the joint distribution of $h_1(y), h_2(y), \ldots, h_k(y)$. The gambler's fortune increases by $A(y)$ and his capital is decreased by the (nonnegative) amount $E[T(y)] - \sum_{i=1}^{k} E[T(w_i)]$. The result now follows by noting that all the hypotheses about the gambling game are satisfied by this stochastic process. □

5 Open Questions

Theorem 7, which concerns the probabilistic recurrence relation $T(z) = A(z) + \sum_{i=1}^{k} T(h_i(z))$, requires the hypothesis that the expected remaining work can never increase; i.e., that for every possible assignment $y_1, y_2, \ldots, y_k$ of joint values to the random variables $h_1(z), h_2(z), \ldots, h_k(z)$, $E[T(z)] \geq \sum_{i=1}^{k} E[T(y_i)]$. This hypothesis is violated by many probabilistic recurrence relations that arise in computational geometry and in the study of data structures. In order to weaken or eliminate this hypothesis it will be necessary to study a version of the gambling game of Section 4 in which the gambler's capital need not decrease, but is permitted to increase in a constrained way.

It would also be of interest to study probabilistic recurrence relations of the form $T(z) = A(z) + \max_{i=1}^{k} T(h_i(z))$. Here the random variable $T(z)$ can be interpreted as the execution time of a parallel divide-and-conquer algorithm having the following description: to process instance z, spend $A(z)$ steps to split the problem into derived instances $h_1(z), h_2(z), \ldots, h_k(z)$, and then solve these derived instances in parallel.

Acknowledgement: The problems considered in this paper first suggested themselves in connection with the author's joint research with Eli Upfal and Avi Wigderson, and a related theorem is presented in the Appendix to [KaUpWi]. The writer wishes to thank Michael Luby, Gary Miller, Jim Pitman, Raymond Seidel, Eli Upfal, Avi Wigderson and Wolf Zimmerman for their extremely helpful comments.

References

[DuSa] L.E. Dubins, L.J. Savage, *How to Gamble If You Must*, McGraw-Hill, New York, 1965.

[Fr] D. Freedman, "Another Note on the Borel-Cantelli Lemma and the Strong Law, with the Poisson Approximation as a By-Product," *The Annals of Probability*, Vol. 1, No. 6, pp. 910–925, 1973.

[KaUpWi] R.M. Karp, E. Upfal and A. Wigderson, "The Complexity of Parallel Search," *J. Comput. System. Sci.*, Vol. 36(2), pp. 225–253, 1988.

[Lu] M. Luby, "A Simple Parallel Algorithm for the Maximal Independent Set Problem," *SIAM J. Comput.*, Vol. 15, pp. 1036–1053, 1986.

[MiRe] G.L. Miller and J.H. Reif, "Parallel Tree Contraction and its Application," *Proc. 26th Annual IEEE Symp. on Foundations of Computer Science*, pp. 478–489, 1985.