A Method for Obtaining Randomized Algorithms with Small Tail Probabilities

H. Alt *　　L. Guibas †　　K. Mehlhorn ‡　　R. Karp §
A. Wigderson ¶

26 September 1991

Abstract

We study strategies for converting randomized algorithms of the Las Vegas type into randomized algorithms with small tail probabilities.

1 Introduction

Let A be a randomized algorithm of the Las Vegas type, i.e., A's output is always correct and A's running time T_A is a random variable. Let $E_0 = E[T_A]$. Then $prob(T_A \geq t) \leq E_0/t$ for all t according to Markov's inequality. If no further information about the distribution of T_A is available, Markov's inequality is the best bound available for the tail probability. Consider now the following modified algorithm. It runs A for $t_1 = 2E_0$ time units. If A stops before the *threshhold* t_1 then the modified algorithm stops. If A does not stop before time t_1, then the modified algorithm restarts A and runs it again for $t_2 = 2E_0$ time units but with new random choices. In this way $prob(T_{mod} \geq k2E_0) \leq 2^{-k}$ for all $k \in \mathbb{N}$ or $prob(T_{mod} \geq t) \leq 2^{-\lfloor t/2E_0 \rfloor}$ for all $t \in \mathbb{R}$, where T_{mod} is the running time of the modified algorithm. The bound for the tail probability of the modified algorithm depends on the sequence $t_1, t_2, \ldots$ of threshholds chosen by the modified algorithm. What is an optimal sequence?

Let us first state the problem in more abstract terms. Let $X, X_1, X_2, \ldots$ be independent nonnegative random variables with common distribution function $f(x)$.

---

*Fachbereich Mathematik, Freie Universität Berlin, Arnimallee 2–6, W–1000 Berlin, Germany; supported by ESPRIT II Basic Research Actions Program of the EC under contract no. 3075(project ALCOM).

†Stanford University and DEC SRC, 130 Lytton Ave, Palo Alto, CA 94301.

‡Max–Planck–Institut für Informatik and Fachbereich Informatik, Universität des Saarlandes, Im Stadtwald, W–6600 Saarbrücken, Germany; supported by ESPRIT II Basic Research Actions Program of the EC under contract no. 3075(project ALCOM).

§International Computer Science Institute, Berkeley, CA 94704 and University of California at Berkeley; Research supported by NSF Grant No. CCR-9005448.

¶Department of Computer Science, Hebrew University and Department of Computer Science, Princeton University, Princeton, NJ 07974; partially supported by a Wolfson Research Award administered by the Israel Academy of Sciences and Humanities.

Let $Y_1, Y_2, \ldots$ be a sequence of nonnegative random variables (not necessarily independent) and let i_0 be the least i such that $X_i < Y_i$. Define random variable T by $T = Y_1 + Y_2 + \ldots + Y_{i_0-1} + X_{i_0}$. A *strategy* $\mathcal{S}$ is a distribution function for the Y's. A strategy $\mathcal{S}$ together with a distribution f for the X_i's induces a distribution for the random variable T. Let $b_{\mathcal{S},f}(t) = prob(T \geq t)$ and let

$$b_{\mathcal{S}}(t, E_0) = \sup\{b_{\mathcal{S},f}(t);\ f \text{ is a distribution with } \int_0^\infty xf(x)dx = E_0\},$$

i.e., $prob(T \geq t) \leq b_{\mathcal{S}}(t, E_0)$ for all distributions f for X with $E[X] = E_0$ and $b_{\mathcal{S}}(t, E_0)$ is the smallest such value. A strategy $\mathcal{S}$ is called *deterministic* if each Y_i can assume only a single value and *probabilistic* otherwise. Set $b_{\mathcal{S}}(t) = b_{\mathcal{S}}(t, 1)$.

For example, the strategy mentioned in the first paragraph is deterministic. We have $prob(Y_i = 2E_0) = 1$ for all i and $b_{\mathcal{S}}(t, E_0) \leq 2^{-\lfloor t/2E_0 \rfloor} \leq 2(2^{1/2})^{t/E_0}$. We show

Theorem 1 *For all strategies $\mathcal{S}$: $b_{\mathcal{S}}(t) \geq e^{-t}$ for all $t \geq 0$.*

Theorem 2 *There is a probabilistic strategy $\mathcal{S}$ with $b_{\mathcal{S}}(t) \leq e^{-(t-1)}$ for all $t \geq 0$.*

Theorem 3 *There is a deterministic strategy $\mathcal{S}$ with $b_{\mathcal{S}}(t) \leq e^{-t+O(\sqrt{t}\log t)}$ for all $t \geq 0$.*

Theorem 4 *There are positive constants c_1 and c_2 and a deterministic strategy $\mathcal{S}$ such that $b_{\mathcal{S}}(t, E) \leq e^{-c_1 t/(E(\ln E)^2) + \ln(c_2 t)}$ for all $t \geq 0$ and $E \geq 1$.*

Theorems 1, 2 and 3 imply that there are near–optimal probabilistic and deterministic strategies for the case of a known value of $E_0 = E[X]$, i.e., for the case where the strategy may depend on the value E_0. Note that, although these theorems are stated for the case $E_0 = 1$, simple scaling extends them to all values of E_0. Theorem 4 deals with the case of an unknown expectation $E[X]$. Of course, a lower bound has to be assumed for $E[X]$ to make the question meaningful. We prove an exponential bound for the tail probability but were not able to determine the optimal base of the exponential function.

All proofs are given in Section 2.

2 Proofs

2.1 The Proof of Theorem 1

We prove Theorem 1. Let $f(x) = e^{-x}$. Then $E[X] = \int_0^\infty xf(x)dx = 1$ and $prob(X \geq x) = \int_x^\infty f(z)dz = e^{-x}$. A strategy $\mathcal{S}$ is defined by a probability measure μ on $\Omega = (\mathbb{R}_{\geq 0})^\infty$, i.e., by a probability measure on the set of infinite sequences of nonnegative reals.

Let $t \in \mathbb{R}_{\geq 0}$ and let j_0 be the random variable defined by

$$Y_1 + \ldots + Y_{j_0-1} < t \leq Y_1 + \ldots + Y_{j_0}.$$

Then

$$prob(T \geq t) = \sum_{j \geq 1} prob(T \geq t \mid j_0 = j) prob(j_0 = j).$$

Let $\Omega_j = \{(y_1, y_2, \ldots); y_1 + \ldots + y_{j-1} < t \le y_1 + \ldots + y_j\}$. Then $prob(j_0 = j) = \mu\Omega_j$. Also, an element $(y_1, y_2, \ldots) \in \Omega_j$ contributes to the event $T \ge t$ if and only if $X_1 \ge y_1, X_2 \ge y_2, \ldots, X_{j-1} \ge y_{j-1}$, and $X_j \ge t - (y_1 + \ldots + y_{j-1})$, i.e., it contributes to the event $T \ge t$ with probability e^{-t}. Thus $prob(T \ge t \mid j_0 = j) = e^{-t}$ and hence $prob(T \ge t) = e^{-t}$. This proves Theorem 1.

2.2 The proof of Theorem 2

We prove Theorem 2. We first define a strategy $\mathcal{S}$. The random variables $Y_1, Y_2, \ldots$ are independent with common density function $g(y) = e^{-y}$. Let f be any distribution with $\int_0^\infty xf(x)dx = 1$ and let $b(t) = b_{\mathcal{S},f}(t)$ for all t. We will show $b(t) \le 1$ for $t \le 1$ and $b(t) \le e \cdot e^{-t}$ for $t \ge 1$. Consider some fixed t. Let $q = prob(X > t)$ be the probability that X exceeds the threshhold t, and for all x with $0 \le x \le t$ let $h(x) = prob(X \ge x \mid X \le t)$ be the conditional probability that $X \ge x$ given that $X \le t$. Then

$$m = E[X \mid X \le t] = \int_o^t h(x)dx \le \frac{1 - qt}{1 - q}$$

since

$$1 = E[X] = (1 - q)E[X \mid X \le t] + qE[X \mid X > t] \ge (1 - q)m + qt.$$

Also,

$$b(t) = q(e^{-t} + \int_0^t e^{-x}b(t - x)dx) + (1 - q)\int_0^t e^{-x}b(t - x)h(x)dx.$$

This can be seen as follows. Define random variable T' by $Y_2 + \ldots + Y_{i_0 - 1} + X_{i_0}$ if $i_0 \ge 2$ and by $T' = 0$ if $i_0 = 1$. If $X_1 > t$ the event $T \ge t$ occurs iff either $Y_1 \ge t$ or Y_1 assumes a value x between 0 and t and $T' \ge t - x$. If $X_1 \le t$ then the event $T \ge t$ occurs iff Y_1 assumes a value x between 0 and t, $X_1 \ge Y_1$ and $T' \ge t - x$. Next observe that $prob(T' \ge t - x \mid X_1 \ge Y_1) = b(t - x)$ since the random variables $X_1, X_2, \ldots, Y_1, Y_2, \ldots$ are independent. Make the substitution $Q(t) = e^t b(t)$. Then

$$Q(t) = q(1 + \int_0^t Q(t - x)dx) + (1 - q)\int_0^t Q(t - x)h(x)dx.$$

We will show that $Q(t) \le e$ for $t \ge 1$ and $Q(t) \le e^t$ for $t < 1$. The case $t < 1$ is immediate. For $t \ge 1$ it suffices to plug this inequality into the right-hand-side and show that it holds for the left-hand-side. The right-hand-side is bounded above by

$$q(1 + et - 1) + (1 - q)em \le qte + e(1 - qt) \le e.$$

This completes the proof.

2.3 The Proof of Theorem 3

We prove Theorem 3. For any integers n and i with $1 \le i \le n$ define

$$t_i(n) = \frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{n - i + 1}.$$

Note that $\sum_{1 \le i \le n} t_i(n) = n$. Let $s(n)$ be the sequence $t_1(n), t_2(n), \ldots, t_n(n)$ and let the strategy $\mathcal{S}$ be obtained by concatenating together $s(1), s(2), s(3), \ldots$. For $1 \le i \le n$ let $p_i(n) = prob(X \ge t_i(n))$. The following Lemma is crucial for the analysis of strategy $\mathcal{S}$.

Lemma 1 *For all integers n, $\prod_{1 \le i \le n} p_i(n) \le \frac{n!}{n^n}$.*

Proof: Let $t_0(n) = 0$ and $p_{n+1}(n) = 0$. Then

$$\begin{aligned} 1 = E[X] &\ge \sum_{1 \le i \le n} (p_i(n) - p_{i+1}(n)) t_i(n) \\ &= \sum_{1 \le i \le n} p_i(n)(t_i(n) - t_{i-1}(n)) \\ &= \sum_{1 \le i \le n} p_i(n)/(n - i + 1). \end{aligned}$$

Let $\overline{p} = (\overline{p_1}, \overline{p_2}, \ldots, \overline{p_n}) \in \mathbb{R}^n$ be the n-tuple which maximizes the product function $P(p_1, p_2, \ldots, p_n) = \prod_{1 \le i \le n} p_i$ subject to the constraint $\sum_{1 \le i \le n} p_i/(n - i + 1) \le 1$. Clearly, $\sum_{1 \le i \le n} \overline{p_i}/(n-i+1) - 1 = 0$. Let $g(p_1, p_2, \ldots, p_n) = \sum_{1 \le i \le n} p_i/(n-i+1) - 1$. The Lagrange multiplier rule [Erw64, Theorem 66] implies the existence of a constant λ such that

$$\frac{\partial P}{\partial p_i}(\overline{p}) - \lambda \frac{\partial g}{\partial p_i}(\overline{p}) = 0$$

for all i, i.e., $P(\overline{p})/\overline{p_i} = \lambda/(n-i+1)$ or $\overline{p_i} = C(n-i+1)$ for some constant C. The constraint $g(\overline{p}) = 0$ implies $C = 1/n$. Thus $\prod_{1 \le i \le n} p_i(n) \le P(\overline{p}) = n!/n^n$. ∎

We now bound $b_{\mathcal{S}}(t)$. Consider a t that lies between the binomial coefficients $\binom{n+1}{2}$ and $\binom{n+2}{2}$ and let $t_0 = \binom{n+1}{2}$. Since $\sum_{1 \le i \le k} t_i(k) = k$, we have $\sum_{1 \le i \le k \le n} t_i(k) = t_0 \le t$ and therefore $b_{\mathcal{S}}(t) \le \prod_{1 \le i \le k \le n} p_i(k) \le \prod_{1 \le k \le n} k!/k^k$. By Stirling's approximation [Kn73, page 111], $k!/k^k \le \sqrt{2\pi k} e^{-k}(k+1)/k$ and hence $b_{\mathcal{S}}(t) \le (2\pi n)^{n/2} e^{-t_0}(n+1) \le e^{-t + O(\sqrt{t} \log t)}$, completing the proof.

2.4 The Proof of Theorem 4

We prove Theorem 4. We first define the strategy $\mathcal{S}$. For integers i and j let $m_{ij} = \lfloor e^{j-i}/i^2 \rfloor$. Let $\mathcal{S}_j$ be the sequence consisting of m_{1j} copies of e^1, followed by m_{2j} copies of e^2, followed by m_{3j} copies of $e^3, \ldots$ Let $\mathcal{S}$ be obtained by catenating $\mathcal{S}_1, \mathcal{S}_2, \ldots$ We now bound $prob(T \ge t)$ for $t \in \mathbb{R}$. Let $i_0 \in \mathbb{N}$ be such that $e^{i_0 - 2} < E_0 = E[X] \le e^{i_0 - 1}$, set $M_j = \sum_{i \ge 1} m_{ij} e^i$, and let j_0 be such that $\sum_{j \le j_0} M_i \le t < \sum_{j \le j_0 + 1} M_j$.

Lemma 2 (a) $j_0 \ge \ln(6t(e-1)/\pi^2 e^2)$,

(b) $prob(T \ge t) \le e^{-\sum_{j \le j_0} m_{i_0 j}}$,

(c) $\sum_{j \le j_0} m_{i_0 j} \ge \frac{c_1 t}{E_0 (\ln E_0)^2} - \ln(c_2 t)$.

Proof:

(a) Note first that $M_j = \sum_i m_{ij} e^i \leq \sum_i e^j/i^2 = \pi^2 e^j/6$ and hence $\sum_{j \leq j_0} M_j \leq \sum_{j \leq j_0} \pi^2 e^j/6 \leq \pi^2 e^{j_0+2}/(6(e-1))$. Thus $t < \pi^2 e^{j_0+2}/(6(e-1))$ and therefore $j_0 \geq \ln(6t(e-1)/(\pi^2 e^2))$.

(b) It follows from the definition of $\mathcal{S}$ and j_0 that the event $T \geq t$ implies the occurrence of $\sum_{j \leq j_0} m_{i_0 j}$ events of the form $X \geq e^{i_0}$. But $prob(X \geq e^{i_0}) \leq 1/e$ according to Markov's inequality and the fact that $E[X] \leq e^{i_0 - 1}$.

(c)

$$\begin{aligned}
\sum_{1 \leq j \leq j_0} m_{i_0 j} &= \sum_{1 \leq j \leq j_0} \left\lfloor \frac{e^{j-i_0}}{i_0^2} \right\rfloor \\
&\geq \frac{1}{i_0^2} \sum_{1 \leq j \leq j_0} e^{j-i_0} - j_0 \\
&\geq \frac{e^{j_0} - 1}{i_0^2 e^{i_0+1}(e-1)} - j_0 \\
&\geq \frac{6t(e-1)/(\pi^2 e^2) - 1}{i_0^2 e^{i_0+1}(e-1)} - \ln \frac{6t(e-1)}{\pi^2 e^2} \\
&\geq \frac{c_1 t}{E_0 (\ln E_0)^2} - \ln(c_2 t)
\end{aligned}$$

for some constants c_1 and c_2. Here, the first inequality follows from the definition of m_{ij}, the fourth inequality follows from part (a), and the last inequality follows from the fact that $E_0 \geq e^{i_0 - 2}$.

∎

Theorem 4 is now a direct consequence of parts (b) and (c) of the preceding Lemma.

References

[Erw64] F. Erwe. *Differential- und Integralrechnung.* Bibliographisches Institut Mannheim, 1964.

[Kn73] D.E. Knuth. *The Art of Computer Programming, Volume I.* Addison-Wesley, 1973