Journal of Computer and System Sciences **58**, 174–176 (1999)
Article ID jcss.1997.1551, available online at http://www.idealibrary.com on IDEAL®

A Note on the Influence of an ϵ-Biased Random Source

Amir Ben-Dor

Department of Computer Science, Technion, Haifa, Israel 32000

Anna Karlin

Department of Computer Science, FR-35, University of Washington, Seattle, Washington 98195

Nathan Linial*

Department of Computer Science, Hebrew University, Jerusalem, Israel 91904

and

Yuri Rabinovich

Department of Computer Science, University of Toronto, Ontario, Canada M5S 1A4
E-mail: amirbd@cs.technion.ac.il, karlin@geoduck.cs.washington.edu, nati@cs.huji.ac.il, yuri@cs.toronto.edu

Received November 22, 1995

An ϵ-biased random source is a sequence $X=(X_1, X_2, ..., X_n)$ of 0, 1-valued random variables such that the conditional probability $\Pr[X_i=1 \mid X_1, X_2, ..., X_{i-1}]$ is always between $\frac{1}{2}-\epsilon$ and $\frac{1}{2}+\epsilon$. Given a family $S \subseteq \{0, 1\}^n$ of binary strings of length n, its ϵ-enhanced probability $\Pr_\epsilon(S)$ is defined as the maximum of $\Pr_X(S)$ over all ϵ-biased random sources X. In this paper we establish a tight lower bound on $\Pr_\epsilon(S)$ as a function of $|S|$, n and ϵ. © 1999 Academic Press

1. INTRODUCTION

Following the definition of Santha and Vazirani [SV2], we consider in this paper the class of *semi-random* sources with *bias* ε, $0 \leqslant \varepsilon \leqslant \frac{1}{2}$. Such a source is a sequence $X=(X_1, X_2, ..., X_n)$ of 0, 1-valued random variables satisfying the condition

$$\tfrac{1}{2}-\varepsilon \leqslant \Pr[X_i=1 \mid X_1, X_2, ..., X_{i-1}] \leqslant \tfrac{1}{2}+\varepsilon$$

for all $i=1, ..., n$. Equivalently, n coins are flipped sequentially by an adversary who knows all previous coin flips and gets to choose the bias of each coin. Clearly, if the source is unbiased ($\varepsilon=0$), it is a perfect random source. On the other hand, if the source is completely biased ($\varepsilon=\frac{1}{2}$), the adversary has complete control over the outcome, and no randomness remains.

Let $S \subseteq \{0, 1\}^n$ be a set of length-n binary strings. A perfect source of randomness hits S with probability $|S|/2^n$, called the *density* of S. What happens if, instead of being perfect, our source is semi-random and the adversary who controls it aims to *maximize* the probability of hitting S? How large can the probability of hitting S be made if the bias is not exceed ε? Formally, the ε-*enhanced probability* $\Pr_\varepsilon(S)$ of S is defined as

$$\Pr_\varepsilon(S)=\max_X \Pr_X(S),$$

where X ranges over all ε-biased semi-random sources.

The question of establishing the optimal lower bound on $\Pr_\varepsilon(S)$ as a function of ε and the density d of $|S|$ (i.e., $d=|S|/2^n$) was raised in [SV1] in the context of bounding the influence of a semi-random source (first introduced in that paper). The authors claimed that the lower bound is attained a on certain explicitly constructed set, computed its value, and provided a short sketch outlining their proof. However, in the final version of their paper [SV2] this result was replaced by a different one (weaker, but still adequate for the paper's purposes), and the proof of the original claim never appeared in print. In subsequent papers discussing the circle of related problems [AR, BLS, H, P], the Santha–Vazirani claim was proven only in a special case when d is of the form $d=1-2^{-\ell}$ or $d=2^{-\ell}$.

In the present paper we amend this situation and prove the Santha–Vazirani claim for an arbitrary d in the range $[0, 1]$. The main technical contribution of the paper is the proof of Lemma 2.1, stated in [SV1] without a proof.

* Research supported by the Israel Academy of Science.

0022-0000/99 $30.00

2. THE LOWER BOUND

The following function ϕ_ε: $[0, 1] \to [0, 1]$ will play a key role in the following investigation. Recall that ε is between 0 and $\frac{1}{2}$.

DEFINITION 2.1. Let $0 \leqslant x \leqslant 1$ be a number with a (finite or infinite) binary expansion $x = \sum_k 2^{-\alpha_k}$, where $0 \leqslant a_1 < a_2 < \cdots$ is an increasing sequence of nonnegative integers. Define $\phi_\varepsilon(x)$ as

$$\phi_\varepsilon(x) = \sum_i \left(\tfrac{1}{2} - \varepsilon\right)^{i-1} \left(\tfrac{1}{2} + \varepsilon\right)^{a_i - i + 1}.$$

It is a routine matter to verify that $\phi_e(x)$ is well defined on $[0, 1]$ (even though some x have two distinct binary representations). Furthermore, ϕ_ε is monotone increasing and continuous on this interval.

For example, $\phi_\varepsilon(0) = 0$, $\phi_\varepsilon(1) = 1$, $\phi_\varepsilon(\frac{1}{2}) = \frac{1}{2} + \varepsilon$. The emergence of the above ϕ_ε, as well as some of its properties (i.e., monotonicity), might, perhaps, be clarified by the following construction. Let k be a number between 0 and 2^n. Define recursively the set $S(k, n) \subseteq \{0, 1\}^n$ as follows:

If $k = 0$ then $S(k, n) = \varnothing$. If $k = 2^n$ then $S(k, n)$ is all of $\{0, 1\}^n$. Otherwise, if $k < 2^{n-1}$, let $S(k, n) = 1 \times S(k, n-1)$ (this is a subset of $1 \times \{0, 1\}^{n-1}$). Finally, if $k \geqslant 2^{n-1}$, let $S(k, n)$ be the union of $1 \times \{0, 1\}^{n-1}$ and $0 \times S(k - 2^{n-1}, n-1)$.

The set $S(n, k)$ comes up in the study of isoperimetric problems in combinatorics, because of the following extremal property that it has: its edge-boundary is the smallest among all sets of k points in $\{0, 1\}^n$ (see, e.g., [Bo]).

CLAIM 2.1. $\Pr_\varepsilon(S(k, n)) = \phi_\varepsilon(k/2^n)$.

Proof. It is easy to see that the adversary, aiming at maximizing the hitting probability of S, should always bias the source towards 1, making its probability $\frac{1}{2} + \varepsilon$. The reason for this is that for any (binary) prefix $(b_1, ..., b_i)$, the cardinality of the intersection $|S(k, n) \cap b_1 \times \cdots \times b_i \times 0 \times \{0, 1\}^{n-i-1}|$ is always smaller than $|S(k, n) \cap b_1 \times \cdots \times b_i \times 1 \times \{0, 1\}^{n-i-1}|$. (In fact, $S(k, n)$ is an initial segment in the lexicographic ordering of $\{0, 1\}^n$.) Therefore,

$$\Pr_\varepsilon(S(k, n)) = \begin{cases} \left(\tfrac{1}{2} + \varepsilon\right) \Pr_\varepsilon(S(k, n-1)), \\ \quad \text{if } k < 2^{n-1}, \\ \left(\tfrac{1}{2} + \varepsilon\right) + \left(\tfrac{1}{2} - \varepsilon\right) \Pr_\varepsilon(S(k - 2^{n-1}, n-1)), \\ \quad \text{otherwise.} \end{cases}$$

Notice also that $\Pr_\varepsilon(S(2^i, i)) = 1$ and $\Pr_\varepsilon(S(0, i)) = 0$. Expanding the expression for $\Pr_\varepsilon(S(k, n))$ according to the above identities leads precisely to the definition of $\phi_\varepsilon(d)$. The easy verification is omitted. ∎

The main result of this present paper says that $\phi_\varepsilon(d)$ is, in fact, the smallest ε-enhanced of any set S of density d. The proof is based on the following lemma.

LEMMA 2.1. *ϕ_ε satisfies the inequality*

$$\left(\frac{1}{2} - \varepsilon\right) \phi_\varepsilon(a) + \left(\frac{1}{2} + \varepsilon\right) \phi_\varepsilon(b) \geqslant \phi_\varepsilon\left(\frac{a+b}{2}\right),$$

where $0 \leqslant a \leqslant b \leqslant 1$.

Proof. Let us first list for future use the following four simple properties of ϕ_ε:

(a) $\phi_\varepsilon(x/2) = (\frac{1}{2} + \varepsilon)\, \phi_\varepsilon(x)$ for all $0 \leqslant x \leqslant 1$.

(b) $\phi_\varepsilon(x + \frac{1}{2}) = (\frac{1}{2} + \varepsilon) + ((1 - 2\varepsilon)/(1 + 2\varepsilon))\, \phi_\varepsilon(x)$ for all $0 \leqslant x \leqslant \frac{1}{2}$.

(c) $\phi_\varepsilon(x + \frac{1}{4}) = (\frac{1}{2} + \varepsilon)^2 + ((1 - 2\varepsilon)/(1 + 2\varepsilon))\, \phi_\varepsilon(x)$ for all $0 \leqslant x \leqslant \frac{1}{4}$.

(d) $\phi_\varepsilon(x + \frac{1}{4}) = (\frac{1}{2} + \varepsilon) - (\frac{1}{2} + \varepsilon)^2 + \phi_\varepsilon(x)$ for all $\frac{1}{4} \leqslant x \leqslant \frac{1}{2}$.

The verification of the above identities is straightforward and is omitted.

Since ϕ_ε is continuous, it is enough to prove the lemma when both a and b have finite binary representations. The proof will proceed by induction on the (max of) the lengths of the binary representations of a, b.

In the base case $a, b \in \{0, 1\}$, and the lemma is verified directly.

Assume inductively that it holds for any a, b with binary expansions of length $\leqslant l$. In order to extend the lemma to length $l + 1$, we need to consider the following three cases:

Case 1. $A = \frac{1}{2}a$, $B = \frac{1}{2}b$, where $a \leqslant b$.

Case 2. $A = \frac{1}{2} + \frac{1}{2}a$, $B = \frac{1}{2} + \frac{1}{2}b$, where $a \leqslant b$.

Case 3. $A = \frac{1}{2}a$, $B = \frac{1}{2} + \frac{1}{2}b$,

where a, b always have an expansion of length $\leqslant l$. We shall deal with each case separately.

Case 1. By (a), we have $\phi_\varepsilon(A) = (\frac{1}{2} + \varepsilon)^{-1}\, \phi_\varepsilon(a)$, $\phi_\varepsilon(B) = (\frac{1}{2} + \varepsilon)^{-1}\, \phi_\varepsilon(b)$, and $\phi_\varepsilon((A + B)/2) = (\frac{1}{2} + \varepsilon)^{-1}\, \phi_\varepsilon((a + b)/2)$. Since by the inductive assumption the lemma is true for a, b, it must be true for A, B as well.

Case 2. Similar to Case 1, using (b) and (a) to express $\phi_\varepsilon(A)$ and $\phi_\varepsilon(B)$ in terms of $\phi_\varepsilon(a)$ and $\phi_\varepsilon(b)$.

Case 3. Requires a more involved analysis. Let $x = \frac{1}{2}a$, $y = \frac{1}{2}b$. Our goal is to show that the inequality holds for $\frac{1}{2} + x$; y. Namely,

$$\left(\frac{1}{2} + \varepsilon\right) \phi_\varepsilon\left(\frac{1}{2} + x\right) + \left(\frac{1}{2} - \varepsilon\right) \phi_\varepsilon(y)$$
$$\geqslant \phi_\varepsilon\left(\frac{x + y}{2} + \frac{1}{4}\right).$$

Equivalently, applying (b) to the left-hand side, one need to show that

$$\left(\frac{1}{2}+\varepsilon\right)^2+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(y)+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(x) \geqslant \phi_\varepsilon\left(\frac{x+y}{2}+\frac{1}{4}\right), \tag{1}$$

where $0 \leqslant x, y \leqslant \frac{1}{2}$. Without loss of generality, we assume in that follows $x \leqslant y$. Arguing as in Case 1, we see that

$$\left(\frac{1}{2}+\varepsilon\right)\phi_\varepsilon(y)+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(x) \geqslant \phi_\varepsilon\left(\frac{x+y}{2}\right).$$

The discussion splits now in two, according to the value of $x+y$.

First case: $x+y \leqslant \frac{1}{2}$. Expanding the right-hand side of the last inequality according to (a), and using $\frac{1}{2}+\varepsilon \geqslant \frac{1}{2}-\varepsilon$, we conclude that

$$\phi_\varepsilon(x)+\phi_\varepsilon(y) \geqslant \phi_\varepsilon(x+y).$$

Therefore,

$$\begin{aligned}\left(\frac{1}{2}+\varepsilon\right)^2&+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(y)+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(x)\\ &\geqslant \left(\frac{1}{2}+\varepsilon\right)^2+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(y+x)\\ &= \left(\frac{1}{2}+\varepsilon\right)^2+\frac{1-2\varepsilon}{1+2\varepsilon}\phi_\varepsilon\left(\frac{y+x}{2}\right).\end{aligned}$$

By (c), the rightmost expression is equal to $\phi_\varepsilon((x+y)/2+\frac{1}{4})$, implying (1).

Second case: $\frac{1}{2} \leqslant x+y \leqslant 1$. Since $y \leqslant \frac{1}{2}$ and ϕ_ε is monotone increasing,

$$\phi_\varepsilon(y) \leqslant \phi_\varepsilon(\tfrac{1}{2}) = \tfrac{1}{2}+\varepsilon.$$

Therefore, since the equation is true for x, y, one has

$$\begin{aligned}&\left(\frac{1}{2}+\varepsilon\right)^2+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(y)+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(x)\\ &\quad=\left(\frac{1}{2}+\varepsilon\right)^2+\left[\left(\frac{1}{2}+\varepsilon\right)\phi_\varepsilon(y)+\left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(x)\right]-2\varepsilon\phi_\varepsilon(y)\\ &\quad\geqslant\left(\frac{1}{2}+\varepsilon\right)^2+\phi_\varepsilon\left(\frac{x+y}{2}\right)-2\varepsilon\left(\frac{1}{2}+\varepsilon\right)\\ &\quad=\left(\frac{1}{2}+\varepsilon\right)-\left(\frac{1}{2}+\varepsilon\right)^2+\phi_\varepsilon\left(\frac{x+y}{2}\right)\\ &\quad=\phi_\varepsilon\left(\frac{x+y}{2}+\frac{1}{4}\right),\end{aligned}$$

where the last equality follows from (d). Thus (1) is true in this case as well.

This concludes the proof of the lemma. ∎

THEOREM 2.1. *Let S be a subset of $\{0,1\}^n$ with density $d=|S|/2^n$. Let $\frac{1}{2} \geqslant \varepsilon \geqslant 0$ be the bias of the source. Then $\Pr_\varepsilon(S) \geqslant \phi_\varepsilon(d)$.*

Proof. The proof is by induction on n. For $n=1$ the theorem is verified directly. Assume now that the theorem holds for every subset of $\{0,1\}^{n-1}$. Given $S \subseteq \{0,1\}^n$ as above, let $S=S_0 \cup S_1$ be a partition of S according to the value of the first coordinate. Let d_0 and d_1 denote the densities of S_0 and S_1, respectively, whence $d=(d_0+d_1)/2$.

Without loss of generality, we may assume that $d_0 \leqslant d_1$. Since the adversary can bias the first bit to be 1 with probability $\frac{1}{2}+\varepsilon$, it holds that

$$\Pr_\varepsilon(S) \geqslant (\tfrac{1}{2}-\varepsilon)\Pr_\varepsilon(S_0)+(\tfrac{1}{2}+\varepsilon)\Pr_\varepsilon(S_1).$$

By the induction hypothesis, $\Pr_\varepsilon(S_0) \geqslant \phi_\varepsilon(d_0)$ and $\Pr_\varepsilon(S_1) \geqslant \phi_\varepsilon(d_1)$. Combining this with Lemma 2.1, we obtain the desired lower bound:

$$\begin{aligned}\Pr_\varepsilon(S) &\geqslant \left(\frac{1}{2}-\varepsilon\right)\phi_\varepsilon(d_0)+\left(\frac{1}{2}+\varepsilon\right)\phi_\varepsilon(d_1)\\ &\geqslant \phi_\varepsilon\left(\frac{d_0+d_1}{2}\right)=\phi_\varepsilon(d).\end{aligned}$$ ∎

REFERENCES

[AR] N. Alon and M. O. Rabin, Biased coins and randomized algorithms, *Adv. in Comput. Res.* **5** (1987), 499–507.

[BLS] M. Ben-Or, N. Linial, and M. Saks, Collective coin flipping and other models of imperfect randomness, *in* "Colloq. Math. Soc. Janos Bolyai," Vol. 52, Combinatorics, Edger, Hungary, 1987.

[Bo] B. Bollobas, "Combinatorics," Cambridge Univ. Press, Cambridge, 1986.

[H] J. Hastad, A simpler proof that it is not possible to extract a good random bit from one slightly random source, unpublished manuscript.

[P] B. Pinkas, "Cryptography and Models of Weak Randomness," Masters thesis, Technion, 1991.

[SV1] M. Santha and U. V. Vazirani, Generating quasi-random sequences from semi-random sources, *in* "Proc. 25th Annual Symposium of Foundation of Computer Science, 1984," pp. 434–440.

[SV2] M. Santha and U. V. Vazirani, Generating quasi-random sequences from semi-random sources, *J. Comput. System. Sci.* **33** (1986), 75–87.