77

BIASED COINS AND RANDOMIZED ALGORITHMS

N. Alon and M. O. Rabin

ABSTRACT

A slightly random source is a source of bits, where the bias of each bit, between $1/2 + \varepsilon$ and $1/2 - \varepsilon$ for some $\varepsilon > 0$, is fixed by an adversary who has a complete knowledge of all the previous bits. We study the properties of sequences of n consecutive bits generated by such a source. In particular we show that for most subsets S of half of the n-binary vectors, even a fixed bias $\varepsilon > 0$, and arbitrarily large, n will not enable the adversary (who knows S) to avoid it with probability approaching 1 as n tends to infinity. Also, for every n and every $S \subseteq \{0, 1\}, |S| = 2^{n-1}$, if $\varepsilon < 1/(2\sqrt{n})$ then the adversary cannot decrease the probability of landing in S below 1/6. These results mean that for randomized algorithms such as primality testing, even a fairly biased coin will produce good answers, without any change in the algorithm.

Advances in Computing Research, Volume 5, pages 499-507.

ISBN: 0-89232-896-7

1. INTRODUCTION

Several applications, such as randomized algorithms [Ra], require a source of fair coin flips. The available physical sources are imperfect. The simplest model of such an imperfect source of random bits is a coin whose flips are independent, and each has a fixed (and unknown) bias. von Neumann [vN] gave a simple algorithm for generating absolutely random independent bits from such a coin. Blum [Bl] (see also Elias [El]) generalized this algorithm to the case where the imperfect random source is an n-state Markov chain. This algorithm, however, is not useful for very large n since it produces bits only when states are repeated. A very general model of an imperfect source of randomness is considered by Santha and Vazirani [SV] and by Vazirani [V] (see also [CG, VV]). In this model, the next bit is an output of a coin whose bias (between $1/2 + \varepsilon$ and $1/2 - \varepsilon$ for some $0 < \varepsilon < 1/2$) is fixed by an adversary who has a complete knowledge of all the previous bits. Thus the previous bits can condition the next bit in an arbitrary bad way. Such a source is called a slightly random source in [SV], and as is explained in [Mu] it includes the known physical sources of randomness as, e.g., zener diodes. The algorithms of [SV, V], for extracting almost fair coin flips from such a model, use the existence of at least two *independent* slightly random sources. It is not clear at all that such an assumption is practical. The bad behavior of the sources might arise from the environment's influence and then the sources influence each other. On the other hand, it is trivial to show that no such algorithm that uses a single slightly random source exists. Thus it is interesting to check the properties of a single slightly random source. In this chapter we show that under reasonable assumptions n consecutive output bits of a single slightly random source form a "reasonable random" n-binary vector. In a typical randomized algorithm (such as the known primality test algorithms see [Ra]), we choose randomly an n vector and we succeed if this vector corresponds to a "witness". Suppose that the set of witnesses S forms a constant fraction c $(0 < c < 1)$ of all 2^n possible vectors. Our first observation is that if $\varepsilon(n) = d/\sqrt{n}$, then even an adversary who tries to avoid S and chooses the bias of every flip between $1/2 - \varepsilon(n)$ to $1/2 + \varepsilon(n)$ has a probability $f(c, d) > 0$ (independent of n) of getting an n vector in S. This result is sharp.

More surprising is our second result, which shows that under the (plausible) assumption that the set S of witnesses is a random set, even fixed bias $\varepsilon > 0$ and arbitrarily large n will not enable the adversary (who knows S) to avoid it with probability $\to 1$ as $n \to \infty$. Thus, for example, if $\varepsilon = 0.05$ and n is large, then for almost every subset S of the set of all 2^n binary vectors, an adversary who knows S and tries to avoid it by choosing the bias of each of his coin flips between $1/2 - \varepsilon$ to $1/2 + \varepsilon$ (taking into account the results of the previous flips), will get a vector of S with probability $> 1/4$. Therefore, for almost all sets S, a weakly random source is reasonable, even under adversary assumptions.

Very recently, Vazirani and Vazirani [VV] (see also [CG] for some extensions) have found a clever algorithm that works for *every* set S of $c \cdot 2^n$ witnesses in the following sense. In the algorithm, a single slightly random source is used to produce a large polynomial number of n-vectors, at least one of which belongs to S with probability $f(c) > 0$ (independent of n).

In the present chapter we do not discuss possible algorithms to obtain witnesses with high probability, but rather study the properties of the bits produced directly by a single weakly random source. We believe that this supplies a better understanding of the,behavior of such a source. Moreover, in our approach (unlike in the more sophisticated algorithms of [VV, CG]) we need only n slightly random bits to produce an n bit number, and we do not need any extra space.

The chapter is organized as follows. In Section 2 we find, for every bias $0 < \varepsilon < 1/2$, for every n, and for every $0 \leqslant k \leqslant 2^n$, the "worst possible" set S of n-vectors of cardinality k. In Section 3 we consider random sets S. Section 4 contains some concluding remarks.

2. THE EXTREMAL CASE

We begin with some notation. For $n \geqslant 1$ let $N = N(n)$ denote the set of all binary vectors of length n. For $0 \leqslant \varepsilon \leqslant 1/2$, let $F(n, \varepsilon)$ be the following set of strategies F for choosing a binary vector $(x_1, x_2, \ldots, x_n) \in N$. $x_1 \in \{0, 1\}$ is chosen according to the probability distribution $\mathrm{Prob}(x_1 = 0) = \rho_1 = \rho_1(F)$ where $1/2 - \varepsilon \leqslant \rho_1 \leqslant 1/2 + \varepsilon$. (The value of ρ_1 is determined by the strategy F.) For every given binary values of $x_1, \ldots, x_{i-1}$, $x_i \in \{0, 1\}$ is chosen according to the probability distribution $\mathrm{Prob}(x_i = 0) = \rho_i$, where $\rho_i = \rho_i(F, x_1, \ldots, x_{i-1})$ satisfies $1/2 - \varepsilon \leqslant \rho_i \leqslant 1/2 + \varepsilon$.

Let S be a set of binary vectors of length n. Define $P(S, \varepsilon) = \min_{F \in F(n,\varepsilon)} \operatorname{Prob}\{(x_1, x_2, \ldots, x_n) \in S; (x_1, x_2,, \ldots, x_n)$ is chosen according to $F\}$. Thus $P(S, \varepsilon)$ is the minimal possible probability of a binary vector to be in S if it is chosen according to one of the strategies in $F(n, \varepsilon)$. [That is, according to biased coin flips, each in the range $(1/2 - \varepsilon, 1/2 + \varepsilon)$, where the bias is chosen by an adversary who knows the previous flips results, knows S, and tries to avoid it.]

Define a linear order on the set of all binary vectors of length n as follows: If $u = (u_1, u_2, \ldots, u_n)$, $v = (v_1, v_2, \ldots, v_n)$ then $u \leqslant v$ iff $\sum_{i=1}^n u_i < \sum_{i=1}^n v_i$ or $\sum_{i=1}^n u_i = \sum_{i=1}^n v_i$ and $\sum_{i=1}^n u_i 2^i < \sum_{i=1}^n v_i 2^i$. A set S of binary vectors is called *compressed* if $v \in S$ and $u \geqslant v \to u \in S$. It is easy to check that if S is compressed then it contains all vectors with at most j 0's and possibly some vectors with precisely $j + 1$ 0's, where $0 \leqslant j < n$ satisfies

$$\sum_{i=0}^{j} \binom{n}{i} \leqslant |S| \leqslant \sum_{i=0}^{j+1} \binom{n}{i}. \tag{2.1}$$

Finally, for a set $S \subseteq N$ we denote by CS the unique compressed set of cardinality $|S|$.

PROPOSITION 2.1. (i) *For every $0 \leqslant \varepsilon \leqslant 1/2$ and for every set S of binary vectors*

$$P(S, \varepsilon) \geqslant P(CS, \varepsilon).$$

(ii) *Suppose $0 \leqslant \varepsilon \leqslant 1/2$ and S, j satisfy* (2.1). *Put $r = |S| - \sum_{i=0}^{j} \binom{n}{i}$. Then*

$$P(CS, \varepsilon) = \sum_{i=0}^{j} \binom{n}{i} (1/2 + \varepsilon)^i (1/2 - \varepsilon)^{n-i}$$
$$+ r \cdot (1/2 + \varepsilon)^{j+1} (1/2 - \varepsilon)^{n-j-1}.$$

That is, the best adversary's strategy to avoid CS is to bias each flip, as strongly as he can, toward 0.

Proof. The set N of all 2^n binary vectors can be naturally represented by the set of all leaves of a rooted binary tree of height n. Each left edge represents a zero and each right edge a 1. A leaf corresponds to the vector arising from the edges of the unique path from the root to the leaf. Any strategy $F \in F(n, \varepsilon)$ is an assignment of a pair of probabilities $1/2 - \varepsilon \leqslant \rho$, $q \leqslant 1/2 + \varepsilon$, $\rho + q = 1$ to each pair of edges that emanates from a common parent. It is easy to check that we can assume that the adversary always chooses, for each pair of probabilities, either $\rho = 1/2 - \varepsilon$ or $\rho = 1/2 + \varepsilon$. Indeed, from each parent he will prefer to go with the highest possible probability to the child from whom he has more chances to avoid S. Thus, we can assume that each flip is as biased as possible.

For a vector $v = (v_1, v_2, \ldots, v_n) \in N$ put $\rho(v) = (1/2 - \varepsilon)^{|\{i: v_i = 1\}|} (1/2 + \varepsilon)^{|\{j: v_j = 0\}|}$. If each flip is as biased as possible then the sequence of probabilities of the leaves of our tree is clearly some permutation of the numbers $\{\rho(v) : v \in N\}$. The total probability of vectors in S is thus at least the sum of the $|S|$ smallest numbers in the sequence $\{\rho(v) : v \in N\}$. These numbers are, however, precisely those whose sum is given in part (ii) of the proposition, and if S is compressed the strategy of always preferring 0 achieves this bound. This completes the proof. □

Since a binomial distribution can be approximated by a normal one, one can get a very good estimate for the bound supplied by Proposition 2.1. Thus, for example, it implies that for all fixed c, $d > 0$ there exists an $f = f(c, d) > 0$ such that if $S \subseteq N$, $|S| = c \cdot 2^n$ and $\varepsilon = \varepsilon(n) = d/\sqrt{n}$ then $P(S, \varepsilon) \geqslant P(CS, \varepsilon) > f$. On the other hand if $\varepsilon = \varepsilon(n) = dg(n)/\sqrt{n}$ where $g(n) \to \infty$ arbitrarily slowly it is easy to check that $\lim_{n \to \infty} P[CS, \varepsilon(n)] = 0$.

As a special case we mention that if $|S| = 1/2 \cdot 2^n$, $\varepsilon = \varepsilon(n) = 1/(2\sqrt{n})$ then the normal approximation gives that $P(S, \varepsilon) \geqslant 1/6$.

Remark 2.2. The assertion of Proposition 2.1 can be easily generalized to the case of a random "dice" ($t > 2$ possible results at each flip). This can be used to improve some of the results of [TRV]. We omit the details.

3. THE RANDOM CASE

In this section we show that for a random subset S of binary vectors of length n, even a fixed bias $\varepsilon > 0$ and arbitrarily large n will not enable the adversary, who knows S, to avoid it with probability $\to 1$ as $n \to \infty$.

Let S be a random subset of N. That is, each $v \in N$ is in S with probability 1/2, independently. For each $0 \leqslant \varepsilon \leqslant 1/2$, $P(S, \varepsilon)$ is now a random variable (on the space of all possible 2^{2^n} subsets S). Let $E = E_{n,\varepsilon} = E[P(S, \varepsilon)]$ and $\sigma = \sigma_{n,\varepsilon} = \sigma[P(S, \varepsilon)]$ denote the expected value and the standard deviation of $P(S, \varepsilon)$.

THEOREM 3.1. *For every $\varepsilon > 0$ that satisfies $1/2 + 2\varepsilon^2 + 2\varepsilon < 1$ and for every n:*

$$E_{n,\varepsilon} \geqslant \frac{1}{2}\left(1 - \frac{\sqrt{2\varepsilon}}{1 - \sqrt{1/2 + 2\varepsilon^2 + 2\varepsilon}}\right)$$

and

$$\sigma_{n,\varepsilon} \leqslant 1/2(1/2 + 2\varepsilon^2 + 2\varepsilon)^{n/2}.$$

Thus, for example, if $\varepsilon = 0.05$ then $E_{n,\varepsilon} \geqslant 1/3$ and $\sigma_{n,\varepsilon} \leqslant (0.78)^n$. Hence, by Chebyshev's inequality [F, p. 219], for random $S \subseteq N(n)$, the probability that $P(S, \varepsilon)$ is smaller than 0.3 is at most $(0.03)^{-2} \cdot (0.78)^{2n}$. That is, almost for every, S, $P(S, \varepsilon) \geqslant 0.3$.

To prove our theorem we need some preparations and a probabilistic lemma.

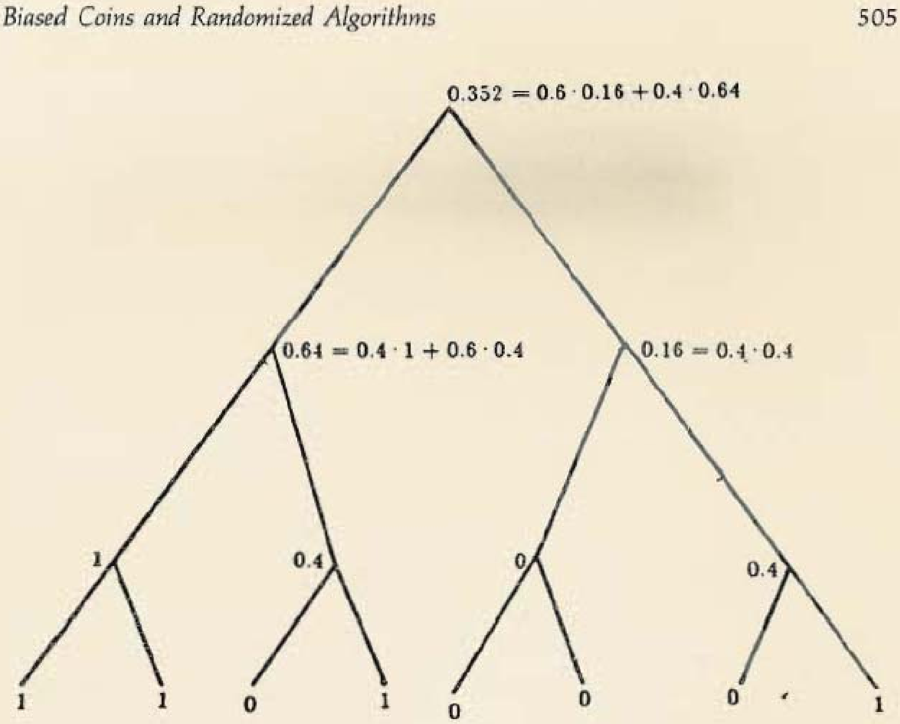
For a given set S and a given ε, one can easily convince himself that $P(S, \varepsilon)$ can be computed as follows: Let T be a binary tree of depth n whose leaves correspond naturally to the binary vectors of length n. Label a leaf corresponding to a vector v by 0 if $v \notin S$ and by 1 if $v \in S$. Now label, recursively, each parent f of the already labeled children s_1, s_2 with the following real number: Suppose s_i is labeled by r_i, then the label of f is $(1/2 + \varepsilon)\min(r_1, r_2) + (1/2 - \varepsilon)\max(r_1, r_2)$. One can check that the label of the root is $P(S, \varepsilon)$. In Figure 1 we have an example of $n = 3$, $S = \{000, 001, 011, 111\}$, $\varepsilon = 0.1$). Here $P(S, \varepsilon) = 0.352$.

Suppose now that S is a random set of vectors in $N(n)$. We have to estimate the expected value and the standard deviation of the random variable $P(S, \varepsilon)$. We need the following lemma.

LEMMA 3.2. *Let X_1, X_2 be two independent random variables, each with expected value E and standard deviation σ. Put*

$$Y = (1/2 + \varepsilon)\min(X_1, X_2) + (1/2 - \varepsilon)\max(X_1, X_2)$$

$$= \frac{X_1 + X_2}{2} - \varepsilon|X_1 - X_2|.$$



Then

$$E(Y) \geqslant E - \sqrt{2}\varepsilon\sigma \tag{3.1}$$

$$\sigma(Y) \leqslant \sqrt{1/2 + 2\varepsilon^2 + 2\varepsilon}\,\sigma. \tag{3.2}$$

Figure 1.

Proof. By Jensen's inequality $E(|X_1 - X_2|)^2 \leqslant (E|X_1 - X_2|)^2$. However, $E(|X_1 - X_2|)^2 = E[(X_1 - X_2)^2] - [E(X_1 - X_2)]^2 = \mathrm{Var}(X_1 - X_2) = 2\sigma^2$. Hence $E|X_1 - X_2| \leqslant \sqrt{2}\sigma$ and (3.1) follows. To prove (3.2) we compute $\sigma^2(Y) = \mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2$.

$$\begin{aligned}\mathrm{Var}(Y) &= 1/4\,\mathrm{Var}(X_1 + X_2) + \varepsilon^2\,\mathrm{Var}|X_1 - X_2| \\ &\quad + \varepsilon\{E(X_1 + X_2)E|X_1 - X_2| - E[(X_1 + X_2)|X_1 - X_2|]\} \\ &\leqslant 1/2\sigma^2 + 2\varepsilon^2\sigma^2 + \varepsilon\{E(X_1 + X_2)E|X_1 - X_2| \\ &\quad - E[(X_1 + X_2)|X_1 - X_2|]\}.\end{aligned}$$

For every two random variables Z, T, $|E(Z)E(T) - E(ZT)| \leqslant \sqrt{\mathrm{Var}\,Z}\sqrt{\mathrm{Var}\,T}$. [This is the well-known fact that the correlation constant is, in absolute value, at most 1, or can be derived by applying Cauchy–Schwarz inequality to obtain $\{E[(Z-EZ)(T-ET)]\}^2 \leqslant E(Z-EZ)^2 E(T-ET)^2$.] Applying this to $Z = X_1 + X_2$, $T = |X_1 - X_2|$ we conclude that

$$E(X_1 + X_2)\,E|X_1 - X_2| - E[(X_1 + X_2)|X_1 - X_2|]$$

$$\leqslant \sqrt{\operatorname{Var}(X_1 + X_2) \cdot \operatorname{Var}|X_1 - X_2|} \leqslant \sqrt{4\sigma^4} = 2\sigma^2.$$

Hence $\operatorname{Var}(Y) \leqslant (1/2 + 2\varepsilon^2 + 2\varepsilon)\sigma^2$ and (3.2) follows. □

Consider now the random variable $P(S,\varepsilon)$ when S is chosen randomly. Define a sequence of random variables $X_0, X_1, \ldots, X_n$ as follows. $\operatorname{Prob}(X_0 = 0) = 1/2$, $\operatorname{Prob}(X_0 = 1) = 1/2$. For $i \geqslant 0$, X_{i+1} is obtained from X_i as follows: let Z_1, Z_2 be two independent random variables having the probability distribution of X_i and put $X_{i+1} = (1/2 + \varepsilon)\min(Z_1, Z_2) + (1/2 - \varepsilon)\max(Z_1, Z_2)$. Clearly X_n is the random variable $P(S,\varepsilon)$. Since $E(X_0) = \sigma(X_0) = 1/2$ repeated application of Lemma 3.2 implies

$$E(X_n) \geqslant E(X_0) - \sqrt{2}\varepsilon \frac{1}{1 - \sqrt{1/2 + 2\varepsilon^2 + 2\varepsilon}}\sigma(X_0)$$

$$= 1/2\{1 - \sqrt{2}\varepsilon/[1 - (1/2 + 2\varepsilon^2 + 2\varepsilon)^2]\}$$

$$\sigma(X_n) \leqslant (1/2 + 2\varepsilon^2 + 2\varepsilon)^{n/2}\sigma(X_0) = 1/2(1/2 + 2\varepsilon^2 + 2\varepsilon)^{n/2}.$$

This proves Theorem 3.1. □

It is worth noting that we can slightly improve our bounds to show that $E > 0$ provided $1/2 + 2\varepsilon^2 + 2\varepsilon < 1$. We omit the details.

4. CONCLUDING REMARKS

We have shown that under reasonable assumptions the output bits of a single weakly random source are reasonably random. Thus, e.g., by the observation of Section 2, a $1/2 \pm (1/2\sqrt{900}) = 1/2 \pm 160$ biased coin is reasonably good, even under adversary assumptions, for checking primality of a 900-bit number using the randomized algorithm of [Ra]. Under the (plausible) assumption that the set of witnesses is random, even a much worse coin is sufficient, by the results of Section 3.

It would be interesting to decide if $E_{n,\varepsilon}$ defined in Section 3 is bounded away from 0 for every fixed $\varepsilon > 0$.

ACKNOWLEDGMENTS

We would like to thank Johan Hastad for fruitful discussions. Research supported in part by Allon Fellowship and by the Fund for Basic Research administered by the Israel Academy of Sciences and in part by NSF Grant CCR-8704513 at Harvard University.

REFERENCES

[Bl] M. Blum, "Independent unbiased coin clips from a correlated biased source: A finite state Markov chain," *Proc. 25th FOCS, Florida* 425–433 (1984).

[CG] B. Chor and O. Goldreich, "Unbiased bits from weak sources of randomness," *Proc. 26th FOCS, Portland, Oregon* 429–442 (1985).

[El] P. Elias, "The efficient construction of an unbiased random sequence," *Ann. Math. Statist.* 43: 865–870 (1972).

[F] W. Feller, *An Introduction to Probability Theory and Its Applications*. 2nd ed., Vol. 1. John Wiley, New York, 1965.

[Mu] H. F. Murray, "A general approach for generating natural random variables," *IEEE Trans. Comput.* C-19: 1210–1213 (1970).

[Ra] M. Rabin, "Probabilistic algorithms," In *Algorithms and Complexity* (J. Traub, ed.), pp. 21–39 Academic Press, New York, 1976.

[SV] M. Santha and U. V. Vazirani, "Generating quasi-random sequences from slightly random sources," *Proc. 25th FOCS, Florida* 434–440 (1984).

[TRV] D. Tyger, M. Rabin, and V. V. Vazirani, in preparation.

[V] U. V. Vazirani, "Towards a strong communication complexity theory or generating quasi-random sequences from two communication slightly random sources," *Proc. 17th STOC, Providence, RI* 366–378 (1985).

[vN] J. von Neumann, "Various techniques used in connection with random digits," Notes by G. E. Forsythe, National Bureau of Standards, *Appl. Math. Ser.* 12: 36–38 (1951). Reprinted in von Neumann's collected works, Pergamon Press, New York, 1963, pp. 768–770.

[VV] U. V. Vazirani and V. V. Vazirani, "Random polynomial time is equal to slightly random polynomial time," *Proc. 26th FOCS, Portland, Oregon* 417–428 (1985).