Transformations of Biased Distributions

Abhijit Sahay *

Computer Science Division
U. C. Berkeley
Berkeley, CA 94720

Abstract

Most approaches to reducing (or eliminating) the use of randomness in probabilistic algorithms can be viewed as attempts to replace the probability space from which the algorithm samples by a much smaller one. The use of limited independence among the random choices of the algorithm has long been used as a technique for constructing small spaces that "approximate" the exponential-sized spaces resulting from complete independence. Biased sample spaces have recently been proposed as an alternative approximation scheme.

We present results that exhibit severe limitations in the extent to which biased spaces can approximate unbiased ones. Specifically, we show that the small bias of a biased space can be destroyed by extremely simple transformations, thereby rendering such spaces unsuitable for any randomized algorithm that employs such transformations. Our results partially explain the paucity of applications for small biased spaces as well as of tools (such as probability tail bounds) for analyzing these distributions. They also suggest that this state of affairs is likely to persist because of the inherent fragility of the notion of bias of a sample space.

*Supported by NSF Grant CCR-9005448

1 Introduction

The notion of biased sample spaces is best motivated by the following conceptual view of a randomized algorithm. Suppose that a randomized algorithm $\mathcal{A}$ tosses an unbiased coin n times independently on inputs of length l and is guaranteed to yield the correct answer with probability at least 1/2. We can think of $\mathcal{A}$ as sampling from the uniform distribution on $S = \{0,1\}^n$ whenever presented an input of length l and acting deterministically thereafter. The performance guarantee says that for each input of length l at least half the strings of S are "good" in that they lead $\mathcal{A}$ to the correct solution.

We might attempt to reduce the number of random bits required by $\mathcal{A}$ by replacing the sample space S with a suitably chosen small subset $S' \subset S$ and showing that the performance guarantee is not compromised. (Of course, this can only be done for specific algorithms since, in general, the omitted strings might all be good for some input x.) This approach has been used successfully for several algorithms ([ABI], [BR], [CG], [KR], [KW], [Lu1], [Lu2], [MNN]) by choosing a k-wise independent subset S' of S. Such a subspace need be no larger than $O(n^k)$ and hence $O(k \log n)$ bits suffice to sample uniformly at random from S'.

Naor & Naor [NN] proposed an alternative construction of small sample spaces to approximate $\mathcal{U}_n$, the uniform distribution on $S = \{0,1\}^n$. Their definition is based on the notion of the *bias* of a distribution (see [Va].) Given a set X of n random variables, each taking a value in $\{0,1\}$, the bias of a subset X' of X is the difference in the probabilities of even and odd parity for X'. It is clear that if the n random variables are distributed according to $\mathcal{U}_n$, the bias of each non-empty subset is zero; moreover, $\mathcal{U}_n$ is the only distribution with this property. [NN] showed how to construct sample spaces where the bias of every subset is at most ϵ and called such spaces *ϵ-biased* . Several simpler constructions were subsequently presented in [AGHP]. The size of these spaces is polynomial in n and $1/\epsilon$, implying great savings in randomness for any algorithm that can be shown to work using ϵ-biased spaces.

The notion of ϵ-approximations to arbitrary (i.e. non-uniform and non-binary) discrete distributions was developed by Azar, Motwani & Naor [AMN] as a generalization of ϵ-biased spaces, which they viewed as an ϵ-approximation to $\mathcal{U}_n$. These authors showed that every joint distribution of n variables admits an ϵ-approximation of size polynomial in n and $1/\epsilon$ and gave a construction for an ϵ-approximation to the uniform distribution on $\{0, \ldots, d-1\}^n$.

It would appear that ϵ-biased spaces with their "almost random" behavior on all subsets of the variables hold great promise for use in randomized algorithms. Indeed, an ϵ-biased space with $\epsilon = \frac{1}{n^{O(1)}}$ induces an "almost" $\log n$-wise independent distribution on a sample space of only polynomial size. Surprisingly, however, not many instances have been found of algorithms that work as well with these spaces as with (exponential size) unbiased spaces. This may be partly explained by the dearth of tools for analysing the probabilistic behavior of ϵ-biased distributions. On the other hand, the success of k-wise independence as a viable technique for reducing randomness suggests that total independence on small subsets of variables might be more useful than near-independence on large subsets.

1.1 Main Result and Implications

Our main result lends credence to this thesis by exhibiting a severe limitation of ϵ-biased spaces with respect to some natural transformations. Suppose we have random variables $X_1, \ldots, X_n$ with joint distribution $\mathcal{D}_X$. Let $Y_1, \ldots, Y_n$ with $Y_i = f_i(X_i)$ and let the induced joint distribution (on $Y_1, \ldots, Y_n$) be denoted $\mathcal{D}_Y$. It is clear that, regardless of the choice of f_i's, any (total) independence properties enjoyed by $\mathcal{D}_X$ is inherited by $\mathcal{D}_Y$; for example, $\mathcal{D}_Y$ is k-wise independent whenever $\mathcal{D}_X$ is. We show that the corresponding statement is false for ϵ-biased distributions even for very simple transformations. Specifically, we consider the following transformations:

- $f : \{0,1,2\}^n \to \{0,1\}^n$ given by

$$f(x_1, \ldots, x_n) = (y_1, \ldots, y_n) \text{ with } y_i = x_i \bmod 2$$

- $g : \{0,1\}^{2n} \to \{0,1\}^n$ given by

$$g(x_1, \ldots, x_{2n}) = (y_1, \ldots, y_n) \text{ with } y_i = x_{2i-1}x_{2i}$$

We show that for both transformations, we can find an ϵ-biased distribution $\mathcal{D}_X$ such that the bias of $\mathcal{D}_Y$ is $\Omega(c^n \epsilon)$ for some constant $c > 1$.

The bias of a distribution is thus shown to be highly fragile, incapable of withstanding even the simplest of transformations. This suggests that ϵ-biased spaces can only be used 'as is': if the algorithm that samples from such a space transforms the sampled values in any way, or if – as is often the case – the analysis of the algorithm rests on analysing the behavior of auxiliary random variables defined by such transformations, then the small bias of the distribution is of little help. As a useful contrast, observe (as noted above) that limited independence is a robust property, persisting through arbitrary point-wise transformations.

We venture to suggest that ϵ-biased spaces have found few applications precisely because of this limitation. An additional argument in favor of this view comes from observing that tools such as moment inequalities or probability tail bounds have not been developed for ϵ-biased distributions. Typically, such bounds are derived by studying some simple transformation of the original space and the discussion above precludes the possibility of using such an approach in establishing these bounds. It would appear then, that even though the notion of bias has an intuitive appeal in that it guarantees certain *global* randomness properties of the distribution, it is not the correct measure of reduced randomness from an algorithmic point of view. On the other hand, the ϵ-biased spaces of [NN], [AGHP] and [AMN] might have randomness properties other than small bias that make them appropriate for use by randomized algorithms and it would be extremely interesting to study these spaces more closely.

1.2 Outline of the Paper

Section 2 of this paper presents the basic definitions and develops the background necessary for understanding our proofs. Section 3 outlines a general framework for analysing transformations of sample spaces and studying their effect on bias. Section 4 contains proofs of our lower bounds for the two simple functions mentioned in the introduction. Some concluding remarks are made in Section 5.

2 Definitions and Background

2.1 ϵ-Biased Distributions

Let $X_1, \ldots, X_n$ be $\{0,1\}$-valued random variables with joint distribution $\mathcal{D}$.

Definition 2.1 *The* bias *(with respect to $\mathcal{D}$) of a subset $S \subset \{1, \ldots, n\}$ of indices is defined to be*

$$\mathrm{bias}_{\mathcal{D}}(S) = |\Pr_{\mathcal{D}}[\sum_{i \in S} X_i \pmod 2 = 0] - \Pr_{\mathcal{D}}[\sum_{i \in S} X_i \pmod 2 = 1]|$$

Definition 2.2 *A distribution $\mathcal{D}$ is called ϵ-biased if $bias_{\mathcal{D}}(S) \leq \epsilon$ for every non-empty subset S.*

Observe that the bias of the empty set is 1, irrespective of the distribution.

It is not difficult to show that the bias of $\mathcal{D}$ is zero if and only if $\mathcal{D} = \mathcal{U}_n$, the uniform distribution on $\{0,1\}^n$. In view of this, one may think of an ϵ-biased distribution as an approximation to $\mathcal{U}_n$, with ϵ measuring the goodness of the approximation.

2.2 ϵ-Approximations of Arbitrary Distributions

An equivalent definition of ϵ-biased distributions can be made in terms of linear tests: each subset of random variables corresponds to a linear function of the variables with the coefficients being specified by the bits of the characteristic vector of the subset. The bias of a subset represents the efficacy of the corresponding function in distinguishing the given distribution from $\mathcal{U}_n$. Thus, an ϵ-biased distribution is precisely one which "passes" all linear tests (to within ϵ.)

In a more general context, we may consider linear tests that attempt to distinguish between distributions $\mathcal{D}$ and $\mathcal{D}'$ and call $\mathcal{D}'$ an ϵ-approximation to $\mathcal{D}$ if no such test achieves a success probability of more than ϵ. Azar, Motwani & Naor [AMN] elegantly formalized this notion using the theory of Fourier transformations of discrete functions. In the following subsection, some basic concepts and definitions of this theory are presented so as to motivate the definition of an ϵ-approximation to an arbitrary discrete distribution. For a more detailed exposition of the theory, see, for example [DM] or [Kö].

2.3 Fourier transformations

Let Z_d denote the set $\{0, \ldots, d-1\}$ and $\mathcal{C}$ the complex numbers. The set $\mathcal{V} = \{f : Z_d^n \to \mathcal{C}\}$ of complex-valued functions on the (additive) group Z_d^n forms a d^n-dimensional vector space.

Definition 2.3 *For $f, g \in \mathcal{V}$, the* inner product *of f and g, denoted $< f, g >$, is defined as*

$$< f, g > = \frac{1}{d^n} \sum_{x \in Z_d^n} f(x) g(x)^*$$

where a^* is the complex conjugate of a.

In what follows, we shall use $A = (a_1, \ldots, a_n)$ and $x = (x_1, \ldots, x_n)$ to denote arbitrary elements of Z_d^n. Given A and x, we let $A.x$ denote the value of $\sum_{i=1}^n a_i x_i \pmod d$. Finally, we let $\omega_d = e^{\frac{2\pi\iota}{d}}$.

Definition 2.4 *The* A-character *of* Z_d^n *is the function* $\chi_A \in \mathcal{V}$ *given by*

$$\chi_A(x) = \omega_d^{A.x}$$

Fact 1 *The set of characters of* Z_d^n, $\{\chi_A : A \in Z_d^n\}$ *forms an orthonormal basis for* $\mathcal{V}$.

Thus, $<\chi_A, \chi_A>= 1$ for each A, $<\chi_A, \chi_B>= 0$ for $A \neq B$ and every $f \in \mathcal{V}$ can be expressed (uniquely) as a linear combination of the characters:

$$f = \sum_{A \in Z_d^n} \hat{f}_A \chi_A$$

The coefficients $\{\hat{f}_A\}$ are called the Fourier coefficients of the function f and orthonormality of characters implies that

$$\hat{f}_A \ = <f, \chi_A>$$

Let $\mathcal{D}$ be a distribution on Z_d^n. Viewing $\mathcal{D}$ as an element of $\mathcal{V}$, we can write

$$\mathcal{D} = \sum_{A \in Z_d^n} \hat{\mathcal{D}}_A \chi_A$$

where the Fourier coefficient $\hat{\mathcal{D}}_A$ satisfies

$$d^n \hat{\mathcal{D}}_A = \sum_{x \in Z_d^n} \mathcal{D}(x) \chi_A(x)^*$$

$$= \sum_{j=0}^{d-1} \Pr_{\mathcal{D}}[A.x = j] * \omega_d^{-j}$$

Observe that if $d = 2$, $\text{bias}_{\mathcal{D}}(S) = 2^n |\hat{\mathcal{D}}_S|$ and that, more generally, $d^n |\hat{\mathcal{D}}_A|$ measures how well the linear test corresponding to A distinguishes $\mathcal{D}$ from the uniform distribution on Z_d^n. This observation motivates the following definition:

Definition 2.5 *Let* $\mathcal{D}$ *and* $\mathcal{D}'$ *be distributions on* Z_d^n. *The* Fourier distance *between* $\mathcal{D}$ *and* $\mathcal{D}'$ *is defined as:*

$$d^n \max_{A \in Z_d^n} |\hat{\mathcal{D}'}_A - \hat{\mathcal{D}}_A|$$

We say that $\mathcal{D}'$ *is an* ϵ*-approximation of* $\mathcal{D}$ *if the Fourier distance between* $\mathcal{D}$ *and* $\mathcal{D}'$ *is at most* ϵ.

It is clear that this definition extends that of ϵ-biased distributions in that the latter are ϵ-approximations to $\mathcal{U}_n$.

It was shown in [AMN] that, unless d is large compared to n, every distribution on Z_d^n admits an ϵ-approximation of size polynomial in n and $\frac{1}{\epsilon}$. However, explicit constructions of such approximations are known only for uniform distributions.

3 Transformations and Bias

Several interesting distributions can be obtained by natural transformations of uniform ones: for example, a coin which has probability 1/4 of coming up heads can be simulated by tossing a fair coin twice and declaring the outcome to be heads if and only if both tosses resulted in a head. More generally, we can consider transforming an n-bit string to an m-bit string according to some transformation $f : \{0,1\}^n \rightarrow \{0,1\}^m$. Given such a transformation, any distribution $\mathcal{D}$ on n-bit strings induces a distribution $f(\mathcal{D})$ on m-bit strings.

In the context of ϵ-approximations, it is natural to ask whether success against linear tests is preserved by such transformations. Indeed, an affirmative answer to this question would immediately yield explicit constructions of ϵ-approximations to non-uniform distributions: we could simply apply the appropriate transformation to an ϵ-biased distribution. We shall now develop a framework for analysing the effect of transformations on the bias of distributions.

Let $\mathcal{D}$ be a distribution on Z_d^n and $\mathcal{D}'$ an ϵ-approximation to $\mathcal{D}$. (Observe that there are many ϵ-approximations to $\mathcal{D}$; we will later choose one that yields the strongest lower bound in our proof.) Let $f : Z_d^n \rightarrow Z_c^m$ be a function and let $\mathcal{E} = f(\mathcal{D}), \mathcal{E}' = f(\mathcal{D}')$ be the induced distributions on Z_c^m given by:

$$\mathcal{E}(y) = \sum_{x \in f^{-1}(y)} \mathcal{D}(x)$$

$$\mathcal{E}'(y) = \sum_{x \in f^{-1}(y)} \mathcal{D}'(x)$$

We wish to study how well $\mathcal{E}$ is approximated by $\mathcal{E}'$. To this end, consider a fixed $B \in Z_c^m$ and define

$$\Delta_B = c^m(\hat{\mathcal{E}}_B - \hat{\mathcal{E}}'_B)$$

Clearly, $\Delta_0 = 0$ and, in general, $|\Delta_B|$ is a measure of the efficacy of the linear test corresponding to B in distinguishing between the distributions $\mathcal{E}$ and $\mathcal{E}'$. By definition,

$$\Delta_B = \sum_{y \in Z_c^m} (\mathcal{E} - \mathcal{E}')(y)\chi_B(y)^*$$

Defining $b = \chi_B \circ f$ we can rewrite this as

$$\Delta_B = \sum_{x \in Z_d^n} (\mathcal{D} - \mathcal{D}')(x)b(x)^*$$

The following lemma allows us to express Δ_B in terms of the Fourier differences $\hat{\mathcal{D}}_A - \hat{\mathcal{D}}'_A$:

Lemma 2 *Suppose $a, b \in \mathcal{V}$. Then*

$$\sum_{x \in Z_d^n} a(x)b(x)^* = d^n \sum_{A \in Z_d^n} \hat{a}_A \hat{b}^*_A$$

Proof: Write $a = \sum_{A_1 \in Z_d^n} \hat{a}_{A_1} \chi_{A_1}, b = \sum_{A_2 \in Z_d^n} \hat{b}_{A_2} \chi_{A_2}$ and use orthonormality of characters. □

Applying the lemma to the expression for Δ_B we get

$$\Delta_B = d^n \sum_{A \in Z_d^n} (\hat{\mathcal{D}}_A - \hat{\mathcal{D}}'_A) \hat{b}_A^*$$

In general, each term in this sum is a product of complex numbers but by an appropriate choice of the ϵ-approximation $\mathcal{D}'$, we can ensure that for some constant α,

$$|\Delta_B| \geq \alpha\epsilon \sum_{A \in Z_d^n, A \neq 0} |\hat{b}_A|$$

The worst-case bias of the induced distribution against the linear test B can thus be determined by studying the Fourier coefficients of the composition $b = \chi_B \circ f$ defined only in terms of the test vector B and the transformation f. In the next section, we shall do this for two simple transformations and demonstrate exponential lower bounds for Δ_B.

4 Simple Transformations with Exponential Bias

4.1 The Pairing Transformation

The pairing transformation $g : Z_2^{2n} \to Z_2^n$ is defined by

$$g(x_1, \ldots, x_{2n}) = (y_1, \ldots, y_n) \text{ where } y_i = x_{2i-1} x_{2i}.$$

Thus, $g(\mathcal{U}_{2n})$ is the distribution on n-bit strings corresponding to n independent tosses of a $\{0,1\}$-coin which has probability $1/4$ of coming up 1.

Following the notation of the previous section, let $B = (b_1, \ldots, b_n)$ be a non-zero element of Z_2^n and define $b : Z_2^{2n} \to \{+1, -1\}$ by $b = \chi_B \circ g$. We evaulate $\sum_{A \in Z_2^{2n}, A \neq 0} |\hat{b}_A|$ by explicitly evaluating each Fourier coefficient $\hat{b}_A$.

Let $A = (a_1, \ldots, a_n) \neq 0$ be a fixed vector in Z_2^{2n}. By definition, $\hat{b}_A = \frac{1}{2^{2n}} \sum_{x \in Z_2^{2n}} b(x) \chi_A(x)^*$

$$= \frac{1}{2^{2n}} [\sum_{x \in b^{-1}(1)} \chi_A(x)^* - \sum_{x \in b^{-1}(-1)} \chi_A(x)^*]$$

We distinguish between the following cases:

Case 1: There is some index $i \in \{1, \ldots, n\}$ such that $b_i = 0$ and either $a_{2i-1} = 1$ or $a_{2i} = 1$.

Without loss of generality, assume that $a_{2i} = 1$. It is clear that $b(x)$ is independent of the $2i$-th bit of x whereas $\chi_A(x)$ is not. But then each partial sum in the expression above must evaluate to zero since whenever $x = (x_1, \ldots, x_{2i}, \ldots, x_{2n})$ is in $b^{-1}(k)$ so is $x' = (x_1, \ldots, 1 - x_{2i}, \ldots, x_{2n})$ and $\chi_A(x) + \chi_A(x') = 0$.

Case 2: $b_i = 0 \Rightarrow a_{2i-1} = a_{2i} = 0$.
Without loss of generality, we may assume that $b_1 = \ldots = b_l = 1; b_{l+1} = \ldots = b_n = 0$. The evaluation of $\hat{b}_A$ is facilitated by some auxiliary definitions. Let

- $S^k(p,q;v_1,v_2) = \{x \in Z_2^{2k} : x_{2k-1} = v_1, x_{2k} = v_2, \sum_{i=1}^{2k} a_i x_i = q, \sum_{i=1}^{k} b_i x_{2i-1} x_{2i} = p\}$
- $S^k(p,q) = \cup_{v_1,v_2} S^k(p,q;v_1,v_2)$
- $N^k(p,q;v_1,v_2) = |S^k(p,q;v_1,v_2)|$
- $N^k(p,q) = |S^k(p,q)|$

(In the definitions above and in the recurrences below, all the arithmetic is modulo 2.)

The following relationships are readily verified:

- $N^k(i,j;0,0) = N^{k-1}(i,j)$
- $N^k(i,j;0,1) = N^{k-1}(i,j-a_{2k})$
- $N^k(i,j;1,0) = N^{k-1}(i,j-a_{2k-1})$
- $N^k(i,j;1,1) = N^{k-1}(i-b_k,j-a_{2k-1}-a_{2k})$

Thus we may write

$$N^k(i,j) = N^{k-1}(i,j) + N^{k-1}(i,j-a_{2k}) + N^{k-1}(i,j-a_{2k-1}) + N^{k-1}(i-b_k,j-a_{2k-1}-a_{2k})$$

Finally, we observe that

$$\hat{b}_A = \frac{1}{2^{2n}}(N^n(0,0) - N^n(0,1) - N^n(1,0) + N^n(1,1))$$

Using our recurrence repeatedly, we get $\hat{b}_A = 2^{-l}$ or $\hat{b}_A = -2^{-l}$. (Which of these holds is determined by whether the number of pairs $a_{2i-1}a_{2i} = 11$ is even or odd.)

This yields
$$\sum_{A \in Z_2^{2n}, A \neq 0} |\hat{b}_A| = \sum_{A \in Z_2^{2l}, A \neq 0} 2^{-l} = (2^l - 2^{-l})$$

We have thus proved the following:

Theorem 3 *Let $\mathcal{D}$ be a distribution on Z_2^{2n} and let g denote the pairing transformation. Then there exists an ϵ-approximation $\mathcal{D}'$ of $\mathcal{D}$ such that the Fourier distance between $g(\mathcal{D})$ and $g(\mathcal{D}')$ is $\Omega(\epsilon 2^n)$.*

4.2 The Modular Transformation

The modular transformation $g : Z_3^n \to Z_2^n$ is defined by

$$f(x_1, \ldots, x_n) = (y_1, \ldots, y_n) \text{ where } y_i = x_i \bmod 2.$$

Thus, if $\mathcal{U}$ denotes the uniform distribution on Z_3^n, $f(\mathcal{U})$ is the distribution on n-bit strings corresponding to n independent tosses of a $\{0,1\}$-coin which has probability $1/3$ of coming up 1.

As before, let $B \in Z_2^n, B \neq 0$ and define $b : Z_3^n \rightarrow \{+1, -1\}$ by $b(x) = \chi_B(f(x))$. We wish to evaluate

$$\sum_{A \in Z_3^n, A \neq 0} |\hat{b}_A|$$

The Fourier coefficient $\hat{b}_A = \frac{1}{3^n} \sum_{x \in Z_3^n} b(x) \chi_A(x)^*$

$$= \frac{1}{3^n} [\sum_{x \in b^{-1}(1)} \chi_A(x)^* - \sum_{x \in b^{-1}(-1)} \chi_A(x)^*]$$

Again, we consider two cases:

Case 1: For some index $i, b_i = 0, a_i \neq 0$.

For such A, note that if $x = (x_1, \ldots, x_i, \ldots, x_n)$ is in $b^{-1}(k)$, then so are $x' = (x_1, \ldots, x_i + 1, \ldots, x_n)$ and $x'' = (x_1, \ldots, x_i + 2, \ldots, x_n)$ (the additions being modulo 3.) Moreover, $\chi_A(x) + \chi_A(x') + \chi_A(x'') = 0$ from which it follows that each partial sum in the expression for $\hat{b}_A$ is zero.

Case 2: For each index $i, b_i = 0 \Rightarrow a_i = 0$.
By a counting argument similar to that used in the previous subsection, we can show the following:

Lemma 4 *Suppose that $b_1 = \ldots = b_l = 1; b_{l+1} = \ldots = b_n = a_{l+1} = \ldots = a_n = 0$. Let r denote the number of indices j such that $a_j \neq 0$. Then $\hat{b}_A = \frac{2^r}{3^l}$.*

Using the lemma it is easy to see that whenever B has l non-zero indices,

$$|\Delta_B| \geq \alpha \frac{\epsilon}{3^l}(5^l - 1)$$

yielding the following:

Theorem 5 *Let $\mathcal{D}$ be a distribution on Z_3^n and let f denote the modular transformation. Then there exists an ϵ-approximation $\mathcal{D}'$ of $\mathcal{D}$ such that the Fourier distance between $f(\mathcal{D})$ and $f(\mathcal{D}')$ is $\Omega(\epsilon(5/3)^n)$.*

5 Summary and Conclusions

We have shown that the bias of a distribution – or, more generally, its security against linear tests – is not a robust attribute: very simple transformations can bring about an exponential deterioration in the bias. It seems unlikely, therefore, that ϵ-biased distributions will find general application in reducing the use of randomness in probabilistic algorithms since most algorithms (or their performance analyses) employ transformations of the sample space.

Our technique also makes explicit the abstract properties required of bias-preserving transformations, viz. that the sum of the *absolute values* of the Fourier coefficients of certain functions determined by the transformation be small. Most natural transformations will fail to satisfy this property. The stringency of this requirement stems directly from the fact that an ϵ-biased space constrains only the *magnitude* of the bias against linear tests and not the *direction*. An adversary can exploit this freedom to ensure that the transformed distribution has a large bias.

It is interesting to note that the ϵ-biased distributions constructed by [NN] have a positive bias against each linear test. One may wonder if the transformations considered here have more success when applied to these distributions. Unfortunately, it turns out that the bias of the transformed space can still be exponentially large in the worst case.

The techniques developed here have been used in [Sa] to investigate how moments of simple functions of n binary random variables are affected by introducing a small bias in the distribution. The finding there is as pessimistic : the bias of the distribution interacts strongly with the Fourier coefficients of the functions and could, in the worst case, cause the moments to deviate by an exponentially large amount from the unbiased moments.

References

[AGHP] N. Alon, O. Goldreich, J. Håstad and R. Peralta, "Simple Constructions of Almost k-wise Independent Random Variables," *Proceedings of the 31st Annual Symposium on the Foundations of Computer Science, pp. 544-553.*

[ABI] N. Alon, L. Babai and A. Itai, "A fast and simple randomized parallel algorithm for the maximal independent set problem," *Journal of Algorithms*, vol. 7 (1986), pp. 567-583.

[AMN] Y. Azar, R. Motwani, J. Naor, "Approximating Distributions Using Small Sample Spaces," manuscript, 1991.

[BR] B. Berger and J. Rompel, "Simulating ($\log^c n$)-wise independence in NC," *Proceedings of the 30th Annual Symposium on the Foundations of Computer Science,* (1989), pp. 2-7.

[CG] B. Chor and O. Goldreich, "On the power of two-point based sampling," *Journal of Complexity*, vol. 5 (1989), pp. 96-106.

[DM] H. Dym and H. P. McKean, *Fourier Series and Integrals,* Academic Press (1972).

[KR] H. Karloff and P. Raghavan, "Randomized algorithms and pseudorandom numbers," *Proceedings of the 20th annual ACM Symposium on Theory of Computing*, 1988, pp. 310-321.

[KW] R. M. Karp and A. Wigderson, "A Fast Parallel Algorithm for the Maximal Independent Set Problem," *JACM*, vol. 32 (1985), pp. 762-773.

[Kö] T. W. Körner, *Fourier Analysis,* Cambridge University Press (1988).

[Lu1] M. Luby, "A simple parallel algorithm for the maximal independent set," *SIAM J. on Computing*, vol. 15 (1986), pp. 1036-1053.

[Lu2] M. Luby, "Removing randomness in parallel computation without a processor penalty," *29th Annual Symposium on Foundations of Computer Science*, 1988, pp. 162-173.

[MNN] R. Motwani, J. Naor and M. Naor, "The probabilistic method yields deterministic parallel algorithms," *Proceedings of the 30th Annual Symposium on the Foundations of Computer Science,* (1989), pp. 8-13.

[NN] J. Naor and M. Naor, "Small-bias probability spaces: efficient constructions and applications," *Proceedings of the 22nd annual ACM Symposium on Theory of Computing*, 1990, pp. 213-223.

[Sa] A. Sahay, "Moment Analysis Using Fourier Transformations," manuscript, 1992.

[Va] U. Vazirani, *Randomness, adversaries and computation*, Ph.D. Thesis, University of California, Berkeley (1986).