Simple Constructions of Almost k-wise Independent Random Variables

Noga Alon* Oded Goldreich† Johan Håstad‡ René Peralta§

June 14, 1992

Abstract

We present three alternative simple constructions of small probability spaces on n bits for which any k bits are almost independent. The number of bits used to specify a point in the sample space is $(2+o(1))(\log\log n + k/2 + \log k + \log\frac{1}{\epsilon})$, where ϵ is the statistical difference between the distribution induced on any k bit locations and the uniform distribution. This is asymptotically comparable to the construction recently presented by Naor and Naor (our size bound is better as long as $\epsilon < 1/(k\log n)$). An additional advantage of our constructions is their simplicity.

Keywords: Probabilistic computation, removing randomness, shiftregister sequences, small probability spaces.

1 Introduction

In recent years, randomization has played a central role in the development of efficient algorithms. Notable examples are the massive use of randomness in computational number theory (e.g., primality testing [26, 28, 18, 1]) and in parallel algorithms (e.g. [21, 24]).

*Sackler Faculty of Exact Sciences, Tel Aviv University, Israel, and IBM Almaden Research Center, San Jose, CA 95120.

†Computer Science Dept., Technion, Haifa, Israel. Supported by grant No. 86-00301 from the United States - Israel Binational Science Foundation (BSF), Jerusalem, Israel.

‡Royal Institute of Technology, Stockholm, Sweden.

§Dept. of Electrical Engineering and Computer Science, University of Wisconsin, Milwaukee, WI, 53201. Supported by NSF grant No. CCR-8909657.

A randomized algorithm can be viewed as a two-stage procedure in which first a "sample point" is chosen at random and next a deterministic procedure is applied to the sample point. In the generic case the sample point is an arbitrary string of specific length (say n), the sample space consists of the set of all 2^n strings, and "choosing a sample at random" amounts to taking the outcome of n consecutive unbiased coin tosses. However, as observed by Luby [21], in many cases the algorithm "behaves as well" when the sample is chosen from a much smaller sample space. If points in the smaller sample space can be compactly represented and generated (i.e. reconstructed to their full length from the compact representation) then this yields a saving in the number of coin tosses required for the procedure. In some cases the required number of coin tosses gets so small that one can deterministically scan all possible outcomes (e.g. [21]).

To summarize, the construction of small sample spaces which have some randomness properties is of major theoretical and practical importance. A typical property is that the probability distribution, induced on every k bit locations in a string randomly selected in the sample space, should be uniform. Such a sample space is called *k-wise independent*.

Alon, Babai and Itai [5] presented an efficient construction of k-wise independent sample spaces of size approximately $n^{k/2}$, where n is (as above) the length of the strings in the sample space. This result is very close to best possible, in view of the lower bound of Chor. et. al. [11]. Hence, k-wise independent sample spaces of size polynomial in n are only possible for constant k. This fact led Naor and Naor to introduce the notion of *almost k-wise independent* sample spaces. Loosely speaking, the probability distribution induced on every k bit locations in the sample string is *"statistically close"* to uniform. Clearly, if an algorithm "behaves well" on points chosen from a k-wise independent sample space then it will "behave essentially as well" on points chosen from an almost k-wise independent sample space. In view of this property it is not surprising that these spaces can be used in many applications. Some applications are presented in [25], while more recent applications are given in [4] and [10]. Another, more detailed application, is that one can get an alternative (slightly nicer) proof of Lemma 8 (on page 8) of [15].

Naor and Naor presented an efficient construction of an almost k-wise independent sample space [25]. Points in their sample space are specified by $O(\log\log n + k + \log\frac{1}{\epsilon})$ bits, where ϵ is a bound on the statistical difference between the distribution induced on k bit locations and the uniform one. The heart of their construction is a sample space of size $\frac{n}{\epsilon^{O(1)}}$ for which the exclusive-or of any fixed bit locations, in the sample point, induces a 0-1 random variable with bias bounded by ϵ (i.e. the exclusive-or of these bits is 1 with probability $\frac{1}{2}(1 \pm \epsilon)$). The constant in the exponent depends, among other things, on the constants involved in an explicit construction of an expander (namely the degree and second eigenvalue of the expander). Using the best known expanders [22] this constant is slightly larger than 4.

We present three alternative constructions of sample spaces of size roughly $(\frac{n}{\epsilon})^2$ for which the exclusive-or of any fixed bit locations, in the sample point, induces a 0-1 random variable with bias bounded by ϵ. Another construction with similar parameters can be given [3] by applying the known properties of the duals of BCH codes (see [23], page 280). Our three constructions are so simple that they can be described in the three corresponding paragraphs below:

1. A point in the first sample space is specified by two bit strings of length $m \stackrel{\rm def}{=} \log(n/\epsilon)$ each, denoted $f_0 \cdots f_{m-1}$ and $s_0 \cdots s_{m-1}$, where $f_0 = 1$ and $t^m + \sum_{i=0}^{m-1} f_i \cdot t^i$ is an irreducible polynomial. The n-bit sample string, denoted $r_0 \cdots r_{n-1}$ is determined by $r_i = s_i$ for $i < m$ and $r_i = \sum_{j=0}^{m-1} f_j \cdot r_{i-m+j}$ for $i \geq m$.

2. A point in the second sample space is specified by a residue x modulo a fixed prime $p \geq (n/\epsilon)^2$. The n-bit sample string, denoted $r_0 \cdots r_{n-1}$, is determined by $r_i = 0$ if $x + i$ is a quadratic residue modulo p and $r_i = 1$ otherwise.

3. A point in the third sample space is specified by two bit strings of length $m \stackrel{\rm def}{=} \log(n/\epsilon)$ each, denoted x and y. The n-bit sample string, denoted $r_0 \cdots r_{n-1}$, is determined by letting r_i equal the inner-product-mod-2 of the binary vectors x^i and y, where x^i is the $i^{\rm th}$ power of x when considered as an element of $GF(2^m)$.

The first construction may be viewed as an explanation for the popularity of using linear feedback shift registers for sampling purposes. We showed that *when both* the feedback rule and the starting sequence *are selected at random*, the resulting feedback sequence enjoys "almost independence" comparable to the length of the feedback rule (i.e., the sequence is "almost" $O(m)$-wise independent, where m is the length of the feedback rule). Similarly, an explanation is provided for the "random structure" of quadratic characters: a random subsequence of quadratic characters (mod p) enjoys "almost independence" comparable to the logarithm of the prime moduli (i.e. the sequence $\chi_p(x+i_0), ..., \chi_p(x+i_{n-1})$ is "almost" $O(\log p)$-wise independent when x is randomly selected).

2 Preliminaries

We will consider probability distributions on binary strings of length n. In particular, we will construct probability distributions which are uniform over some set $S \subseteq \{0,1\}^n$, called the *sample space*. The parameter that will be of interest to us is the "size of the probability space"; namely, the number of strings in the support (i.e. $|S|$). The aim is to construct "small" probability spaces which have "good" randomness properties. In particular we will be interested in k-wise independence.

Convention: A sample space which is contained in $\{0,1\}^n$ will be usually sub-indexed by n. The super-index will usually represent an upper bound on the logarithm (to base 2) of the cardinality of the sample space. Hence, S_n^m denotes a sample space of $\leq 2^m$ strings each of length n.

2.1 Almost k-wise Independence

Definition 1 (k-wise independence): *A sample space S_n is* k-wise independent *if when $X = x_1 \cdots x_n$ is chosen uniformly from S_n then for any k positions $i_1 < i_2 < \cdots < i_k$ and any k-bit string α, we have*

$$Pr[x_{i_1} x_{i_2} \cdots x_{i_k} = \alpha] = 2^{-k}.$$

In many applications it suffices that a bit sequence is "almost" k-wise independent. There are several standard ways of quantifying this condition (i.e. interpreting the phrase "almost"): cf. [9]. We use two very natural ways corresponding to the L_∞ and L_1 norms:

Definition 2 (almost k-wise independence): *Let S_n be sample space and $X = x_1 \cdots x_n$ be chosen uniformly from S_n.*

- (max-norm): S_n *is* (ϵ, k)-independent *(in max norm) if for any k positions $i_1 < i_2 < \cdots < i_k$ and any k-bit string α, we have*

$$|Pr[x_{i_1} x_{i_2} \cdots x_{i_k} = \alpha] - 2^{-k}| \leq \epsilon.$$

- (statistical closeness): S_n *is* ϵ-away (in L_1 norm) from k-wise independence *if for any k positions $i_1 < i_2 < \cdots < i_k$ we have*

$$\sum_{\alpha \in \{0,1\}^k} |Pr[x_{i_1} x_{i_2} \cdots x_{i_k} = \alpha] - 2^{-k}| \leq \epsilon.$$

Clearly, if S_n is (ϵ, k)-independent (in max norm) then it is at most $2^k\epsilon$-away (in L_1 norm) from k-wise independence, whereas if S_n is ϵ-away (in L_1 norm) from k-independence then it is (ϵ, k)-independent (in max norm).

2.2 Linear Tests

The heart of each of our constructions is a sample space which is very close to random with respect to "linear Boolean tests" (i.e., tests which take the exclusive-or of the bits in some fixed locations in the string). Following Naor and Naor [25], these sample spaces can be used in various ways to achieve almost k-wise independence.

Definition 3 :

- *Let $(\alpha, \beta)_2$ denote the inner-product mod 2 of the binary vectors α and β (i.e. $(\alpha_1 \cdots \alpha_n, \beta_1 \cdots \beta_n)_2 = \sum_{i=1}^{n} \alpha_i \beta_i \bmod 2$).*
- *A 0-1 random variable X is called ϵ-biased if*

$$|Pr[X = 0] - Pr[X = 1]| \leq \epsilon.$$

- *Let S_n be a sample space and $X = x_1 \cdots x_n$ be chosen uniformly from S_n. The sample space S_n is said to be ϵ-*biased with respect to linear tests *if for every $\alpha = \alpha_1 \cdots \alpha_n \in \{0,1\}^n - \{0\}^n$ the random variable $(\alpha, X)_2$ is ϵ-biased.*
- *The sample space S_n is said to be ϵ-*biased with respect to linear tests of size at most k *if for every $\alpha = \alpha_1 \cdots \alpha_n \in \{0,1\}^n - \{0\}^n$ such that at most k of the α_i are one, the random variable $(\alpha, X)_2$ is ϵ-biased.*

Clearly, the uniform distribution over all n-bit strings is unbiased (0-biased) with respect to all linear tests. A linear test can be interpreted as trying to refute the randomness of a probability space by taking a fixed linear combination of the bits in the sample.

The following lemma, attributed to Vazirani [29] (see also [30], [11]), links the ability to pass linear tests with almost independence.

Lemma 1 (Vazirani): *Let $S_n \subset \{0,1\}^n$ be a sample space that is ϵ-biased with respect to linear tests of size at most k. Then the sample space S_n is $((1-2^{-k})\epsilon, k)$-independent (in max norm), and $(2^k - 1)^{1/2}\epsilon$-away (in L_1 norm) from k-wise independence.*

In particular this implies:

Corollary 1 (Vazirani): *Let $S_n \subset \{0,1\}^n$ be a sample space that is ϵ-biased with respect to linear tests. Then, for every k, the sample space S_n is $((1-2^{-k})\epsilon, k)$-independent (in max norm), and $(2^k-1)^{1/2}\epsilon$-away (in L_1 norm) from k-wise independence.*

Remark: In the applications of the lemma we will use the bounds ϵ and $2^{k/2}\epsilon$ respectively, since the difference is minimal.

For completeness we give the proof of Lemma 1 in an appendix.

A more advantageous way of using ϵ-biased (w.r.t. linear tests) sample spaces, than just using Corollary 1, was suggested by Naor and Naor [25]: They combine the use of a sample space which is ϵ-biased w.r.t. linear tests with a "linear" k-wise independent sample space. A sample space is called *linear* if its elements are obtained by a linear transformation of their succinct representation (equivalently, the sample space is a linear subspace). For example, the construction of a k-wise independent sample space presented by Alon, Babai and Itai [5] is linear. To be more precise they construct a sample space on n bits, where $n = 2^t - 1$, which is generated by $td+1$ bits and is $(2d+1)$-wise independent. Naor and Naor observed that a sample space which is almost unbiased with respect to linear Boolean tests can be used to sample succinct representations of points in the linear k-wise independent space. The sample obtained can be shown to be ϵ-biased w.r.t. linear tests of size at most k. Using Lemma 1 we get.

Lemma 2 (Naor and Naor): *Let $S_n^m \subset \{0,1\}^n$ be a sample space of cardinality 2^m that is ϵ-biased with respect to linear tests. Let k be an integer and $L_N^n \subset \{0,1\}^N$ be a k-wise independent linear sample space of cardinality 2^n. Suppose L_N^n is defined by the linear map T. Then, the sample space R_N^m constructed by applying the linear map T to each sample point in S_n^m, is ϵ-biased with respect to linear tests of size at most k. Hence R_N^m is (ϵ, k)-independent (in max norm), and $2^{k/2}\epsilon$-away (in L_1 norm) from k-wise independence.*

Using the construction in [5] we get:

Corollary 2 (Naor and Naor): *Let $k < n$ be an odd integer and $N \le 2^{\lfloor \frac{2(n-1)}{k-1} \rfloor} - 1$. Given a sample space, S_n^m, as in Lemma 2, one can construct a sample space $R_N^m \subset \{0,1\}^N$ of cardinality 2^m such that R_N^m is (ϵ, k)-independent (in max norm), and $2^{k/2}\epsilon$-away (in L_1 norm) from k-wise independence.*

Hence, an (ϵ, k)-independent sample space on N bits can be constructed using $O(\log\log N + \log k + \log\frac{1}{\epsilon})$ random bits instead of $O(\log N + \log k + \log\frac{1}{\epsilon})$ random bits (as in direct application of Corollary 1).

In view of Corollary 2, the main part of the paper deals merely with the construction of small sample spaces which have small bias with respect to linear tests.

3 The LFSR Construction

Our first construction is based on linear feedback shift register (LFSR) sequences.

Definition 4 (linear feedback shift register sequences): *Given two sequences $\overline{s} = s_0, s_1, \ldots s_{m-1}$ and $\overline{f} = f_0, f_1, \ldots f_{m-1}$ of m bits each, the* shift register sequence generated *by the* feedback rule $\overline{f}$ *and the* start sequence $\overline{s}$ *is $r_0, r_1, \ldots r_{n-1}$ where $r_i = s_i$ for $i < m$ and $r_i = \sum_{j=0}^{m-1} f_j \cdot r_{i-m+j}$ for $i \geq m$.*

Our sample space will consist of all shift register sequences generated by "non-degenerate" feedback rules and any starting sequence.

Construction 1 (Sample Space A_n^{2m}): *The sample space A_n^{2m} is the set of all shift register sequences generated by a feedback rule $\overline{f} = f_0 f_1 \cdots f_{m-1}$ with $f_0 = 1$ and $f(t) \stackrel{\text{def}}{=} t^m + \sum_{j=0}^{m-1} f_j \cdot t^j$ being an irreducible polynomial (such a feedback rule is called* non-degenerate*). Namely, A_n^{2m} contains all sequences $\overline{r} = r_0 r_1 \cdots r_{n-1}$ such that there exists a non-degenerate feedback rule $\overline{f}$ and a start sequence $\overline{s}$ generating $\overline{r}$.*

Hence, the size of the sample space A_n^{2m} is at most 2^{2m} (actually, it is $\approx \frac{2^{2m}}{m}$). In view of Corollary 2 we now confine ourselves to evaluating the bias of this sample space with respect to linear Boolean tests.

Proposition 1 : *The sample space A_n^{2m} is $\frac{n-1}{2^m}(1 + O(2^{-m/2}))$-biased with respect to linear tests. Namely, for any nonzero α the random variable $(\alpha, r)_2$ is $(n-1)2^{-m}(1 + O(2^{-m/2}))$-biased when r is selected uniformly in A_n^{2m}.*

Proof: For the rest of this section we consider only polynomials over $GF(2)$. The number of irreducible monic polynomials ([19], p. 39) of degree m is

$$\frac{1}{m}\sum_{d|m}\mu(\frac{m}{d})2^d,$$

where μ is the ordinary Möbius function (i.e. $\mu(x) = (-1)^s$ where s is the number of primes that divide x if x is squarefree and $\mu(x) = 0$ otherwise). Since also $\mu(1) = 1$ the above expression is $(1+O(2^{-m/2}))\frac{2^m}{m}$. For the rest of this section we will, for notational simplicity, treat the number of irreducible monic polynomials of degree m as if it is exactly $\frac{2^m}{m}$. (The error introduced is absorbed in the error term.) Hence, with this convention we say that the size of A_n^{2m} is $\frac{2^{2m}}{m}$.

Fix the feedback rule (i.e. $\overline{f}$) and consider the distribution of $(\alpha, r)_2$ when we only vary the starting vector (i.e. $\overline{s}$). A key observation is that the r_i's are a linear combination of the s_j's (which are the only indeterminates as the f_i's were fixed). It is useful (and standard practice) to notice that in $GF(2)$, the reduction of t^j modulo $f(t)$ $(= t^m + \sum_{i=0}^{m-1} f_i \cdot t^i)$ is a linear combination of $t^0, t^1, \ldots t^{m-1}$ and that this linear combination is identical to the coefficients in the expression of r_i as a linear combination of the s_j's. Hence, a linear combination of the r_i's (which is exactly what $(\alpha, r)_2$ is) corresponds to a linear combination of the corresponding powers of t^i. This linear combination can be either identically zero or not. The first case means that the polynomial $f(t)$ divides the polynomial $g(t) \stackrel{\text{def}}{=} \sum_{i=0}^{n-1} \alpha_i \cdot t^i$; whereas in the second case $(\alpha, r)_2$ being a non constant combination of the s_i's is unbiased when the s_i's are uniformly selected. Hence the bias of $(\alpha, r)_2$, when r is uniformly selected in A_n^{2m} equals the probability that the polynomial $f(t)$ divides the polynomial $g(t)$. The latter probability is bounded by the fraction of irreducible monic polynomials of degree m which divide a specific polynomial of degree $n-1$. There are at most $\frac{n-1}{m}$ irreducible monic polynomials of degree m which divide a polynomial of degree $n-1$. Dividing by the number of irreducible monic polynomials of degree m (i.e. $\frac{2^m}{m}$) the proposition follows. ∎

4 The Quadratic Character Construction

Our second construction is based on Weil's Theorem regarding character sums (cf. [27], p. 43, Thm. 2C). A special case of this theorem is stated below.

Definition 5 (Quadratic Character): *Let p be an odd prime and x be an integer relatively prime to p. The* Quadratic Character of x mod p, *denoted $\chi_p(x)$, is 1 if x is a quadratic residue modulo p and -1 otherwise. For x a multiple of p we define $\chi_p(x) = 0$.*

Theorem 1 (Weil): *Let p be an odd prime. Let $f(t)$ be a polynomial over $GF(p)$ which is not the square of another polynomial and has precisely n distinct zeros. Then,*

$$\left| \sum_{x \in GF(p)} \chi_p(f(x)) \right| \leq (n-1)\sqrt{p}$$

The i^{th} bit in the j^{th} sample string, in our sample space, will be $\chi_p(i+j)$. A translation from ± 1 sequences to $\{0,1\}$ sequences can be easily effected. Theorem 1 will be used to analyze the bias of this sample space with respect to linear tests.

Construction 2 (Sample Space $B_n^{\log p}$): *The sample space $B_n^{\log p}$ consists of p strings. The x^{th} string, $x = 0, 1, ..., p-1$, is $\overline{r(x)} = r_0(x)r_1(x)\cdots r_{n-1}(x)$ where $r_i(x) = \frac{1-\chi_p(x+i)}{2}$, for $i = 0, 1, ..., n-1$. If $x + i = p$ then let $r_i = 1$.*

Hence, the size of the sample space $B_n^{\log p}$ is exactly p.

Proposition 2 : *The sample space $B_n^{\log p}$ is $\frac{n-1}{\sqrt{p}} + \frac{n}{p}$-biased with respect to linear tests. Namely, for any nonzero α the random variable $(\alpha, r)_2$ is $\frac{n-1}{\sqrt{p}} + \frac{n}{p}$-biased when r is selected uniformly in $B_n^{\log p}$.*

Proof: The bias of $(\alpha, r)_2$ equals the expectation of $(-1)^{(\alpha,r)_2}$ taken over all possible r's. Hence the bias is

$$\frac{1}{p} | \sum_{x \in GF(p)} (-1)^{\sum_{i=0}^{n-1} \alpha_i r_i(x)} |.$$

For most x's $(-1)^{r_i(x)} = \chi_p(x+i)$ and using $\chi_p(xy) = \chi_p(x)\chi_p(y)$, we get the following bound for the bias:

$$\frac{1}{p} | \sum_{x \in GF(p)} \chi_p(f(x)) | + \frac{n}{p}$$

where $f(x) = \Pi_{i=0}^{n-1}(x+i)^{\alpha_i}$ and the second term comes form the special x for which $x+i = p$ for some i with $\alpha_i = 1$. Using Theorem 1, our proposition follows. ■

5 The Powering Construction

Our third construction is based on arithmetic in finite fields. In particular, we will use $GF(2^m)$ arithmetic and also consider the field elements as binary strings.

Construction 3 (Sample Space C_n^{2m}): *Let* $\mathrm{bin} : GF(2^m) \mapsto \{0,1\}^m$ *be a one-to-one mapping satisfying* $\mathrm{bin}(0) = 0^m$ *and* $\mathrm{bin}(u + v) = \mathrm{bin}(x) \oplus \mathrm{bin}(y)$*, where* $\alpha \oplus \beta$ *means the bit-by-bit xor of the binary strings* α *and* β*. (The standard representation of* $GF(2^m)$ *as a vector space satisfies the above conditions.) A string in the sample space* C_n^{2m} *is specified using two field elements,* x *and* y*. The* i^{th} *bit in this string is the inner-product of* x^i *and* y*. More precisely, the* i^{th} *bit of the sample point is* $(\mathrm{bin}(x^i), \mathrm{bin}(y))_2$*.*

Hence, the size of the sample space C_n^{2m} is 2^{2m}. We now evaluate the bias of this sample space with respect to linear Boolean tests.

Proposition 3 : *The sample space* C_n^{2m} *is* $\frac{n-1}{2^m}$*-biased with respect to linear tests. Namely, for any nonzero* α *the random variable* $(\alpha, r)_2$ *is* $(n-1)2^{-m}$*-biased when* r *is selected uniformly in* C_n^{2m}*.*

Proof: Let $r(x,y) = r_0(x,y) \cdots r_{n-1}(x,y)$ denote the sample point specified by the field elements x and y. Note that

$$(\alpha, r(x,y))_2) = \sum_{i=0}^{n-1} \alpha_i (\mathrm{bin}(x^i), \mathrm{bin}(y))_2$$

which equals $(\mathrm{bin}(\sum_{i=0}^{n-1} \alpha_i x^i), \mathrm{bin}(y))_2$. Let $p_\alpha(t) = \sum_{i=0}^{n-1} \alpha_i t^i$ be a polynomial over $GF(2)$. We are interested in the distribution of $(\mathrm{bin}(p_\alpha(x)), \mathrm{bin}(y))_2$ when $x \in GF(2^m)$ and $y \in GF(2^m)$ are chosen uniformly. As in the proof of Proposition 1, we analyze this probability by fixing x and going through all possible y's. There are two cases to consider. If x is not a zero of the polynomial $p_\alpha(t)$ then $\mathrm{bin}(p_\alpha(x)) \neq 0^m$ and $(\mathrm{bin}(p_\alpha(x)), \mathrm{bin}(y))_2$ is unbiased when selecting y uniformly. If on the other hand x is a zero of $p_\alpha(t)$ then $(\mathrm{bin}(p_\alpha(x)), \mathrm{bin}(y))_2 = 0$ for all y's, but $p_\alpha(t)$ has at most $n-1$ zeros. Hence, the proposition follows. ■

Remark: Construction 3 actually constructs linear feedback shiftregister sequences, but with a different distribution compared to construction 1. The interested reader is invited to check this. The minimal polynomial for x will give the feedback rule.

Remark: It is possible to get slightly more bits without affecting the bias of linear tests. Let v_1, $v_2 \ldots v_m$ be a basis of $GF[2^m]$ over GF[2]. Then we can extract nm bits by letting $b_{ij} = (bin(v_j x^i), bin(y))$. The bias of any xor is still $n2^{-m}$. The proof of this fact is almost identical to the present proof. The only difference is that we get a polynomial over $GF[2^m]$ instead of a polynomial over $GF[2]$.

6 Main Theorems for Almost k-wise independence

Let us put the pieces together. All three constructions use at most $2m$ bits to get n bits with $n2^{-m}$-bias with respect to linear tests. Combining this with Corollary 2 we get:

Theorem 2 *Let $N = 2^t - 1$ and let k be an odd integer. Then it is possible to construct N bits which are (ϵ, k)-independent (in max norm) using $2\left(\lceil \log \frac{1}{\epsilon} + \log\left(1 + \frac{(k-1)t}{2}\right)\rceil\right)$ bits.*

This is roughly $2 \log\left(\frac{k \log N}{2\epsilon}\right)$ bits.

Theorem 3 *Let $N = 2^t - 1$ and let k be an odd integer. Then we can construct N bits which are ϵ-away (in L_1 norm) from k-wise independence using $2\left(\lceil \frac{k}{2} + \log \frac{1}{\epsilon} + \log\left(1 + \frac{(k-1)t}{2}\right)\rceil\right)$ bits.*

This is roughly $k + 2\log\left(\frac{k \log N}{2\epsilon}\right)$ bits.

7 The smallest possible ϵ-bias spaces

The constructions above lead naturally to the problem of studying how close to optimal these are in terms of the size of the sample spaces. Here we briefly comment on this problem. We note that tight bounds for the related quantity which is the minimum possible size of a sample space in which there are n random variables which are k-wise independent are given in [11] and in [5].

For an integer n and for a real $\epsilon < 1/2$, let $m(n, \epsilon)$ denote the minimum m such that there exists a sample space of size m and n $(0, 1)$-random variables embedded in it, such that for any nontrivial linear combination over $GF(2)$ of the random variables, the probability that it is 0 is between $1/2 - \epsilon$ and $1/2 + \epsilon$.

Our objective is to study the function $m(n, \epsilon)$. A very similar function is studied in [6], and most of the techniques applied there can be used in our case as well, as we briefly describe below. Besides these techniques, we need a new result, stated in proposition 4 below.

As mentioned in [25] the problem of estimating $m(n, \epsilon)$ can be best formulated as a problem about error correcting codes. Indeed, suppose there is a sample space of size m and n $(0,1)$-random variables as above over it. Let $a_{ij} \in \{0, 1\}$ denote the value of the i^{th} random variable in the j^{th} point of the sample space, and let A be the n by m matrix given by: $A = (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}$. Then A is the generating matrix of a linear code of dimension n and length m over $GF(2)$ in which all the distances are between $(1/2 - \epsilon)m$ and $(1/2 + \epsilon)m$. Conversely, from any linear code as above we can obtain a sample space and random variables with the corresponding properties.

Therefore, the known bounds in the theory of error correcting codes can be used to estimate the function $m(n, \epsilon)$. The Gilbert-Varshamov bound (in fact, with a slight modification, as here we need a code in which all code words have weight which is very close to $0.5m$) implies that for any n and ϵ:

$$m(n, \epsilon) \leq O(\frac{n}{\epsilon^2}).$$

The same bound can be easily proved by a probabilistic argument as well. It is trivially true that for any positive ϵ, $m(n, \epsilon) \leq 2^n$ since there is a code of length $m = 2^n$ and dimension n in which the weight of each code-word is precisely $0.5m$. Combining this with the modification in the remark of Construction 3 we conclude that for any n and ϵ:

$$m(n, \epsilon) \leq O(MIN\{\frac{n}{\epsilon^2}, 2^n, \frac{n^2}{\epsilon^2(\log_2(n/\epsilon))^2}\}). \tag{1}$$

A lower bound for $m(n, \epsilon)$ can be derived- (as is also mentioned in [6] for the case of fixed ϵ)- from the McEliece-Rodemich-Rumsey-Welch bound (see [23], page 559). Although the proof of this bound, as described, e.g., in [23] is given only for the case of a fixed ϵ (when the length of the code tends to infinity), the same proof can be extended to a more general case, by studying the asymptotic behavior of the smallest roots of the corresponding Krawtchouk polynomials. This gives here that there exists a fixed (small) $\delta > 0$, such that for every n and $\epsilon \geq 2^{-\delta n}$:

$$m(n, \epsilon) \geq \Omega(\frac{n}{\epsilon^2 \log 1/\epsilon}). \tag{2}$$

What happens when $\epsilon < 2^{-\delta n}$? A lower bound which is sharp in this range for $m(n,\epsilon)$ is given in the following result from [3].

Proposition 4 *For every n and ϵ:*

$$m(n,\epsilon) \geq \Omega(MIN\{\frac{1}{\epsilon^2},\ 2^n\}).$$

Combining Proposition 4 and inequality (2) we conclude that for every n and ϵ:

$$m(n,\epsilon) \geq \Omega(MIN\{\frac{n}{\epsilon^2 \log(1/\epsilon)},\ 2^n\}). \tag{3}$$

Note that the upper and the lower bounds for $m(n,\epsilon)$ given in (1) and in (3) coincide (up to a constant factor) when $\epsilon = 2^{-\Theta(n)}$ and give that in this range $m(n,\epsilon) = \Theta(\frac{1}{\epsilon^2})$. Note also that in this range our third explicit construction (the improvement in the remark) gives also a bound of the form $O(1/\epsilon^2)$, which is, thus, tight in this case.

In general, our three explicit constructions all give

$$m \leq \frac{n^2}{\epsilon^2(\log(n/\epsilon))^\delta},$$

where in the first construction $\delta = 1$, in the second $\delta = 0$ and in the third $\delta = 2$. The construction that can be obtained using the BCH-codes, also gives $\delta = 2$ [3].

8 Concluding Remarks

All our three constructions admit fast translation of the succinct representation into the full length sample point. In fact, given succinct representation s and bit location i, the i^{th} bit of the sample determined by s can be computed in $\mathcal{NC}$.

All three constructions can be generalized to d-ary strings, for any prime d. The generalized constructions have small bias with respect to linear tests (which compute a linear combination mod d of the d-ary values considered as elements of Z_d). The first such generalization is due to Azar, Motwani and Naor [7] (extending the characters construction). The second such generalization is due to Guy Even [14] (extending the LFSR construction). Our third construction can be easily generalized as well. However, if one is interested in distributions over d-ary sequences which are statistically close to k-wise independent (d-ary) distributions then these construction do not offer any improvement in efficiency (over the trivial construction which uses a binary construction) (cf. [7],[14]).

An issue to be addressed is the "semi-explicit" presentation of all three constructions. To be fully specified, the first construction requires a list of irreducible polynomials of degree m over $GF(2)$, the second construction requires a prime p (of size $\approx 2^{2m}$), whereas the third construction assumes a representation of $GF(2^m)$ (which amounts to an irreducible polynomial of degree m over $GF(2)$). The reader may wonder whether these requirements can be met in the applications (in which the sample spaces are used). In the rest of this section we answer this question in the affirmative.

In some applications we are allowed to use a preprocessing stage of complexity comparable to the size of the sample space. Two notable examples follow

- The sample space S_n^s is used for *deterministic simulation* of a randomized algorithm. In such a case the overall complexity of the simulation is 2^s times the cost of one call to the randomized algorithm. Hence, adding a preprocessing stage of complexity 2^s does not increase the overall complexity.
- The sample space S_n^s contains strings of length comparable to 2^s (i.e. $s = O(\log n)$). This is the case, for example, when m is selected such that the sample space is ϵ-away from $\log n$-wise independent, for some fixed ϵ (or $\epsilon = n^{-O(1)}$) (cf. [25]).

In a preprocessing stage (of complexity 2^s), we may enumerate all monic polynomials of degree s and discard those which have non-trivial divisors. Similarly, to find a prime larger than M, we can test all integers in the interval $[M, 2M]$ for primality.

In case such a preprocessing is too costly we either omit it or replace it by a randomized preprocessing stage of complexity $m^{O(1)}$. In the 2^{nd} and 3^{rd} constructions all that is needed is one "element" (either a prime in $[M, 2M]$ or a irreducible polynomial of degree m). Such an element can be found by sampling the strings of length ℓ (ℓ equals m or $1 + \log M$, respectively). In both cases the density of good elements is $\approx \frac{1}{\ell}$. A straightforward algorithm to achieve this will require ℓ^2 independently selected ℓ-bit strings, meaning that we use ℓ^3 coin flips in the precomputation (which dominates the $O(\ell)$ coin flips used to select a sample point in the sample space). An alternative procedure is suggested below (for sake of clarity we consider the problem of finding an irreducible polynomial of degree m).

Construction 4 (sample space for irreducible polynomials):

- *Use pairwise-independent sampling to specify m monic polynomials of degree m. With probability at least $\frac{1}{2}$, at least one of these polynomials is irreducible. The pairwise independent sampling requires $2m$ bits (cf. [12]). Call the resulting sample space P_m.*

- *Use an expander-path of length $O(m)$ to specify $O(m)$ points in the sample space P_m. This is done by using $O(m)$ bits to specify a starting point and then using $O(m)$ bits to choose a path of length $O(m)$ starting at this point. With probability at least $1 - 2^{-m}$, at least one of these points specifies a sequence of m polynomials containing at least one irreducible polynomial (cf. [2, 13, 20, 17, 8]). This sampling requires $O(m)$ bits. Call the resulting sample space E_m.*

- *A sample point in E_m specifies $O(m^2)$ polynomials and with overwhelming probability at least one of them is irreducible. Say we* output *the first irreducible polynomial among these m^2 polynomials.*

Remark 1 *When using an expander graph in this construction, it is important to note that there are explicit constructions of expander graphs which do not use a large prime or anything else that might be hard to find deterministically. An example of such a construction is the construction by Gabber and Galil [16].*

Construction 4 suffices as a randomized preprocessing for Constructions 2 and 3, and for a modification of Construction 1 (sketched below). However, for Construction 1 (as appearing in Section 3) we need the ability to select a random irreducible polynomial (and not merely to find and fix one). Construction 4 does get "close" to that goal: although the output *does not* specify a uniformly selected irreducible polynomial, it is easy to see that the probability that a particular polynomial appears in the output is bounded above by $O(m^2 \frac{1}{N})$, where N denotes the number of irreducible polynomials (and hence $\frac{1}{N}$ is the probability that a particular one is picked when we select with uniform probability). Thus, the probability that the polynomial selected by Construction 4 divides a fixed n degree polynomial is bounded above by $m^2 \cdot \frac{n}{2^m}$. Hence, the implementation of Construction 1 in which the feedback rule is selected using Construction 4 yields a sample space of size $2^{O(m)}$ which is $\frac{nm^2}{2^m}$-biased with respect to linear tests.

Finally, we sketch a modification of Construction 1, suggested by Y. Azar. Fix a "non-degenerated" feedback rule f (i.e. an irreducible polynomial of degree m). The sample point specified by a start sequence $\overline{s} = s_0, s_1, ..., s_{m-1}$ and an integer $k < 2^m$ (called the *gap*) is $r_0, r_k, \ldots r_{(n-1)k}$ where $r_i = s_i$ for $i < m$ and $r_i = \sum_{j=0}^{m-1} f_j \cdot r_{i-m+j}$ for $i \geq km$. It can be shown that for any fixed irreducible polynomial $f(t)$ (of degree m) and any degree-n polynomial $g(t)$, when k is uniformly chosen in $\{1, 2, ..., 2^m - 1\}$, the probability that $f(t)$ divides $g(t^k)$ is bounded above by $\frac{n}{2^m}$. Using the argument of Proposition 1 it follows that the modified sample space is $\frac{n}{2^m}$-biased with respect to linear tests. As in Construction 3 we again get linear feedback shiftregister sequences. This time with a third distribution.

Acknowledgment: Noga Alon wishes to thank Ronny Roth for stimulating conversations. Oded Goldreich wishes to thank Guy Even for collaboration in the early stages of this research. Johan Håstad wishes to thank Avi Wigderson for inviting him to Israel. René Peralta wishes to thank Eric Bach for referring him to Schmidt's book.

References

[1] Adleman, L.M., and M-D.A. Huang, "Recognizing Primes in Random Polynomial Time", *Proc. 19th STOC*, 1987, pp. 462–470.

[2] M. Ajtai, J. Komlos, E. Szemerédi, "Deterministic Simulation in LOGSPACE", *Proc. 19th STOC*, 1987, pp. 132–140.

[3] N. Alon, *Unpublished manuscript.*

[4] N. Alon, "A parallel algorithmic version of the local lemma", *Random Structures and Algorithms*, Vol 2, 1991, pp. 367–378. Also *32nd FOCS*, 1991, pp 586–593.

[5] N. Alon, L. Babai, and A. Itai, "A fast and Simple Randomized Algorithm for the Maximal Independent Set Problem", *J. of Algorithms*, Vol. 7, 1986, pp. 567–583.

[6] N. Alon, J. Bruck, J. Naor, M. Naor and R. Roth, "Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs", to appear in *IEEE Trans. on Information theory.*

[7] Y. Azar, R. Motwani and J. Naor, "An efficient construction of a multiple value small bias probability space", to appear.

[8] M. Bellare, O. Goldreich, and S. Goldwasser "Randomness in Interactive Proofs", *31st FOCS*, 1990, pp. 318–326.

[9] R. Ben-Natan, "On Dependent Random Variables Over Small Sample Spaces", M.Sc. Thesis, Computer Science Dept., Hebrew University, Jerusalem, Israel, Feb. 1990.

[10] J. Boyar, G. Brassard, and R. Peralta, "Subquadratic Zero-Knowledge", *32nd FOCS*, 1991, pp. 69–78.

[11] B. Chor, J. Friedmann, O. Goldreich, J. Hastad, S. Rudish, and R. Smolensky, "The bit extraction problem and t-resilient functions", *Proc. 26th FOCS*, 1985, pp. 396–407

[12] B. Chor and O. Goldreich, "On the Power of Two–Point Based Sampling," *Jour. of Complexity*, Vol 5, 1989, pp. 96–106.

[13] A. Cohen and A. Wigderson, "Dispensers, Deterministic Amplification, and Weak Random Sources", *30th FOCS*, 1989, pp. 14–19.

[14] G. Even, private communication, May 1990, to be included in his M. Sc. thesis to be submitted in CS Dept., Technion, Israel.

[15] U. Feige, S. Goldwasser, L. Lovasz, S. Safra, and M. Szegedy, "Approximating Clique is almost NP-complete", *32nd FOCS*, pp. 2–12, 1991.

[16] O. Gabber, Z. Galil, "Explicit Constructions of Linear Size Superconcentrators", *JCSS*, **22** (1981), pp. 407-420.

[17] O. Goldreich, R. Impagliazzo, L.A. Levin, R. Venkatesan, D. Zuckerman, "Security Preserving Amplification of Hardness", *31st FOCS*, 1990, pp. 318-326.

[18] S. Goldwasser, and J. Kilian, "Almost All Primes Can Be Quickly Certified", *Proc. 18th STOC*, 1986, pp. 316–329.

[19] S. W. Golomb, *Shift Register Sequences*, Aegean Park Press, Revised edition, 1982.

[20] R. Impagliazzo, and D. Zuckerman, "How to Recycle Random Bits", *30th FOCS*, 1989, pp. 248-253.

[21] M. Luby, "A simple parallel algorithm for the maximal independent set problem", *Proc. 17th STOC*, 1985, pp. 1–10.

[22] A. Lubotzky, R. Phillips, P. Sarnak, "Explicit Expanders and the Ramanujan Conjectures", *Proc. 18th STOC*, 1986, pp. 240-246.

[23] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North Holland, Amsterdam, 1977.

[24] K. Mulmuley, U.V. Vazirani and V.V. Vazirani, "Matching is as easy as Matrix Inversion", *Proc. 19th STOC*, 1987, pp. 345–354.

[25] J. Naor and M. Naor, "Small-bias Probability Spaces: Efficient Constructions and Applications", *22nd STOC*, 1990, pp. 213–223.

[26] M.O. Rabin, "Probabilistic Algorithms for Testing Primality", *J. of Num. Th.*, 12, pp. 128–138, 1980.

[27] W. M. Schmidt, Equations over finite fields, an elementary approach, Lecture Notes in Mathematics, Vol. 536, Springer Verlag (1976).

[28] R. Solovay, and V. Strassen, "A Fast Monte-Carlo Test for Primality", *SIAM J. of Comp.*, 6, pp. 84–85, 1977.

[29] U.V. Vazirani, "Randomness, Adversaries and Computation", Ph.D. Thesis, EECS, UC Berkeley, 1986.

[30] U.V. Vazirani, and V.V. Vazirani, "Efficient and Secure Pseudo-Random Number Generation," *Proc. 25th FOCS*, 1984, pp. 458–463.

A Proof of Lemma1

Without loss of generality let us look at the variables $x_1, x_2 \dots x_k$. For $\alpha \in \{0,1\}^k$ let p_α be the probability that $x_i = \alpha_i$ for all $1 \le i \le k$. For $\beta \in \{0,1\}^k$ let ϕ_β be the function defined by $\phi_\beta(\alpha) = (-1)^{\sum_{i=1}^k \alpha_i\beta_i}$. Then the discrete Fourier transform of the sequence p_α is defined by

$$c_\beta = \sum_\alpha \phi_\beta(\alpha) p_\alpha.$$

If $\beta \ne 0$ this is exactly the same as the bias of the linear test given by β, and hence in this case $|c_\beta| \le \epsilon$, while $c_0 = 1$.

By standard Fourier analysis we have

$$p_\alpha = 2^{-k} \sum_\beta \phi_\beta(\alpha) c_\beta$$

and

$$\sum_\alpha p_\alpha^2 = 2^{-k} \sum_\beta c_\beta^2.$$

Now we have

$$|p_\alpha - 2^{-k}| = 2^{-k} |\sum_{\beta \neq 0} \phi_\beta(\alpha) c_\beta| \leq (1 - 2^{-k})\epsilon,$$

which proves the first part of the lemma. To see the second part let $p'_\alpha = p_\alpha - 2^{-k}$ and let c'_β be the Fourier transform of the p'-sequence. Then $c'_0 = 0$, while $c'_\beta = c_\beta$ for $\beta \neq 0$. Hence by Cauchy-Schwarz inequality we have

$$\sum_\alpha |p_\alpha - 2^{-k}| \leq 2^{k/2} \left(\sum_\alpha (p_\alpha - 2^{-k})^2 \right)^{1/2} =$$

$$2^{k/2} \left(2^{-k} \sum_{\beta \neq 0} c_\beta^2 \right)^{1/2} \leq (2^k - 1)^{1/2} \epsilon.$$

This finishes the proof of the lemma. ■