Approximating Probability Distributions Using Small Sample Spaces

YOSSI AZAR *
Computer Science Department
Tel Aviv University
Tel Aviv 69978, Israel

RAJEEV MOTWANI †
Computer Science Department
Stanford University
Stanford, CA 94305-2140

JOSEPH (SEFFI) NAOR ‡
Computer Science Department
Technion
Haifa 32000, Israel

Abstract

We formulate the notion of a "good approximation" to a probability distribution over a finite abelian group $\mathbb{G}$. The quality of the approximating distribution is characterized by a parameter ϵ which is a bound on the difference between corresponding Fourier coefficients of the two distributions. It is also required that the sample space of the approximating distribution be of size polynomial in $\log |\mathbb{G}|$ and $1/\epsilon$. Such approximations are useful in reducing or eliminating the use of randomness in certain randomized algorithms.

We demonstrate the existence of such good approximations to arbitrary distributions. In the case of n random variables distributed uniformly and independently over the range $\{0, \ldots, d-1\}$, we provide an efficient construction of a good approximation. The approximation constructed has the property that any linear combination of the random variables (modulo d) has essentially the same behavior under the approximating distribution as it does under the uniform distribution over $\{0, \ldots, d-1\}$. Our analysis is based on Weil's character sum estimates. We apply this result to the construction of a non-binary linear code where the alphabet symbols appear almost uniformly in each non-zero code-word.

*Most of this work was done while the author was at the Computer Science Department, Stanford University and supported by a Weizmann fellowship and Contract ONR N00014-88-K-0166.

†Supported by an Alfred P. Sloan Research Fellowship, an IBM Faculty Development Award, grants from Mitsubishi Electric Laboratories and OTL, NSF Grant CCR-9010517, and NSF Young Investigator Award CCR-9357849, with matching funds from IBM, Schlumberger Foundation, Shell Foundation, and Xerox Corporation.

‡Part of this work was done while the author was visiting the Computer Science Department, Stanford University, and supported by Contract ONR N00014-88-K-0166, and by Grant No. 92-00225 from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel.

1 Introduction

Recently a family of techniques has emerged to reduce or eliminate the use of random bits by randomized algorithms [ABI, BR, CG, KW, Lu1, Lu2, MNN, NN]. Typically, these techniques involve substituting independent random variables by a collection of dependent random variables which can be generated using fewer truly independent random bits. Motivated by this work, we formulate the notion of a *good approximation* to a *joint probability distribution* of a collection of random variables.

We consider probability distributions over a finite abelian group $\mathbb{G}$ and, in particular, over $\mathbb{Z}_d^n$ for any positive integers d and n. We measure the distance between two distributions over $\mathbb{G}$ by the distance, in the maximum norm, of their Fourier transforms over $\mathbb{G}$. Given an arbitrary distribution $\mathcal{D}$ over $\mathbb{G}$, a good approximation $\overline{\mathcal{D}}$ is a distribution with a small distance to $\mathcal{D}$, which is concentrated on a small subset of the sample space. Sampling from the approximating distribution requires significantly fewer random bits than sampling from the original distribution.

Before describing our work in detail, we briefly review some related work. Alon, Babai, and Itai [ABI] and Luby [Lu1] observed that certain algorithms perform as well using pairwise independent random bits, as on mutually independent bits. It turns out that n uniform k-wise independent bits can be generated using sample spaces of size $O(n^{\lfloor k/2 \rfloor})$; a lower bound of $\binom{n}{\lfloor k/2 \rfloor}$ on the minimum size of such a sample space is also known [ABI, CGHFRS]. Thus, these algorithms could be derandomized for constant k by an exhaustive search of the (polynomial-size) sample space. Unfortunately, this degree of independence is very restrictive and limits the applicability of the approach. Berger and Rompel [BR] and Motwani, Naor, and Naor [MNN] showed that several interesting algorithms perform reasonably well with only $(\log n)$-wise independence. The resulting sample space, while of super-polynomial size, could be efficiently searched via the *method of conditional probabilities*, due to Erdős and Selfridge [ES73] (cf. [AS, Chapter 15]), in time logarithmic in the size of the sample space. This led to the derandomization of a large class of parallel algorithms [BR, MNN].

An alternate approach was proposed by Naor and Naor [NN] based on the notion of the *bias* of a distribution due to Vazirani [Va].

Definition 1.1 *Let $X_1, \ldots, X_n$ be $\{0,1\}$ random variables. The* bias *of a subset S of the random variables is defined to be $|\Pr[\,\sum_{i\in S} X_i = 0] - \Pr[\,\sum_{i\in S} X_i = 1]|$, where the sum is taken modulo 2.*

For mutually independent and uniform random variables, the bias of each non-empty subset is zero. It is not hard to show that the converse holds as well. In an *ϵ-biased probability distribution*, each subset of the random variables has bias at most ϵ. Naor and Naor [NN] showed how to construct such a distribution, for any $\epsilon > 0$, such that the size of the sample space is polynomial in n and $1/\epsilon$. The ϵ-biased distribution can be viewed as an *almost* $(\log n)$*-wise independent distribution.* A result due to Peralta [Pe] implies a different construction of ϵ-biased probability distribution using the properties of *quadratic residues*; this and two additional constructions of two-valued ϵ-biased random variables are reported by Alon, Goldreich, Håstad, and Peralta [AGHP].

We formulate and study the notion of a "good approximation" to a joint probability distribution of (possibly multi-valued) random variables. Let $\mathcal{D}$ be any joint distribution of n random variables over the range $\{0, \ldots, d-1\}$. Informally, a good approximation $\overline{\mathcal{D}}$ to $\mathcal{D}$ satisfies the following properties: there is a uniform bound ϵ/d^{n-1} on the absolute difference between corresponding Fourier coefficients over the group $\mathbb{Z}_d^n$ of the two distributions; and, the sample space required for $\overline{\mathcal{D}}$ is of size polynomial in n, d, and $1/\epsilon$. We demonstrate the viability of such approximations by proving that for any distribution $\mathcal{D}$, there exists a good approximation $\overline{\mathcal{D}}$. In fact, this notion and the existence result extend to any probability distribution over a finite abelian group. The quality of the approximation can be further characterized by showing that the *variation distance* between the two distributions $\mathcal{D}$ and $\overline{\mathcal{D}}$ is bounded by the sum of the differences between their Fourier coefficients.

We also consider the issue of an *efficient construction* of such an approximating distribution; specifically, for the uniform distribution over $\mathbb{Z}_d^n$. An efficient construction must determine $\overline{\mathcal{D}}$ in time polynomial in the description length of $\mathcal{D}$, and also in $1/\epsilon$; clearly, this bound must apply to the size of the sample space of the approximating distribution $\overline{\mathcal{D}}$. (Note that the description of a distribution $\mathcal{D}$ over $\mathbb{Z}_d^n$ may be of length as much as d^n.) We provide an efficient construction of a good approximation $\overline{\mathcal{U}}$ to the uniform distribution $\mathcal{U}$ on $\mathbb{Z}_d^n$, i.e., for the joint distribution of uniform and independent d-valued random variables $X_1, \ldots, X_n$. Since the construction must guarantee that the Fourier coefficients of $\overline{\mathcal{U}}$ are very close to those of $\mathcal{U}$, it is essentially an ϵ-biased distribution. This has the following natural interpretation in terms of linear combinations: for *any* vector $A = (a_1, \ldots, a_n)$, $\sum a_i X_i \pmod d$ has "almost" the same distribution in the case where the random variables $X_1, \ldots, X_n$ are chosen from $\mathcal{U}$, as in the case where they are chosen from $\overline{\mathcal{U}}$. The analysis of this construction is based on Weil's character sum estimates, and it generalizes the work of Peralta [Pe] to d-valued random variables. Our results hold for non-prime values of d as well[1].

This construction has found application in the work of Håstad, Phillips, and Safra [HPS]. They consider the approximability of the following algebraic optimization problem: given a collection of quadratic polynomials over $\mathbb{F}_q$, the field of order q, find a common root to the largest possible sub-collection of these polynomials. Our construction is used to show

[1]Following our work, Even [Ev] generalized one of the constructions of Alon, Goldreich, Håstad, and Peralta [AGHP] to the d-valued case when d is a prime.

that finding an approximate solution (to within a ratio of $d-\epsilon$) is as hard as finding an exact solution, and hence is NP-hard; this applies to polynomials over rationals and reals as well. The constructions of two-valued ϵ-biased random variables due to Naor and Naor [NN] and Alon, Goldreich, Håstad, and Peralta [AGHP] are insufficient for this purpose, and our construction of d-valued ϵ-biased random variables needs to be used.

We also show that the *variation distance* between two distributions can be bounded in terms of the differences in their Fourier coefficients. This allows us to demonstrate that our construction gives random variables which are "almost" $(\log_d n)$-wise independent. Our construction is optimal in this respect. We also explore some connections with the construction of linear codes. Our results provide a construction of a linear code over an arbitrary alphabet which has the property that for each non-zero codeword, the distribution of the alphabet symbols is almost uniform, and that the length of the codeword is polynomial (quadratic) in the dimension of the code. Previously, such codes were known only over $\mathbb{F}_2$.

The remaining sections are organized as follows: Section 2 provides some mathematical preliminaries; the existence of a good approximation to an arbitrary distribution and bounds on the variation distance are shown in Section 3; Section 4 studies the notions of bias and k-wise independence. Section 5 gives a construction of an ϵ-biased distribution; Section 6 studies the parameters of the construction; finally, in Section 7 our construction is applied to linear codes.

2 Preliminaries

2.1 Characters of Finite Abelian Groups. Our discussion here follows the exposition of Babai [Ba] and Ledermann [Le]. Let $\mathbb{T}$ denote the multiplicative group of complex numbers with unit modulus. A *character* of a finite abelian group $\mathbb{G}$ is a homomorphism $\chi : \mathbb{G} \to \mathbb{T}$. The characters of $\mathbb{G}$ form the *dual group* $\widehat{\mathbb{G}}$ under pointwise multiplication (for $\chi, \chi' \in \widehat{\mathbb{G}}$ we set $\chi\chi'(x) = \chi(x)\chi'(x)$). It is known that $\widehat{\mathbb{G}} \cong \mathbb{G}$ (cf. [Ba]). The identity element of $\widehat{\mathbb{G}}$ is the *principal character* χ_0 defined by setting $\chi_0(x) = 1$, for all $x \in \mathbb{G}$. The *order* of a character is its order as an element of $\widehat{\mathbb{G}}$.

Let $C(n)$ denote a cyclic group of order n, written multiplicatively. The characters of $\mathbb{G} = C(n)$ are constructed as follows. Let z denote a generator of $\mathbb{G}$.

Definition 2.1 *For $0 \le r \le n-1$, the* rth *character of $C(n)$, denoted by χ_r, is defined as follows:*

$$\chi_r(z^s) = \mathrm{e}\left(\frac{rs}{n}\right),$$

where $s = 0, \ldots, n-1$, and the function $\mathrm{e}(x)$ denotes $e^{2\pi i x}$ for $i = \sqrt{-1}$.

It follows that χ_r has order $n/\gcd(r,n)$.

We remark that in the case that $\mathbb{G}$ is the multiplicative group of a finite field $\mathbb{F}$, the characters are usually extended to all of $\mathbb{F}$ by setting $\chi(0) = 0$.

Let now $\mathbb{G}$ be an arbitrary finite abelian group, given as the direct product of cyclic groups: $\mathbb{G} = C(n_1) \times \ldots \times C(n_k)$. Each element $x \in \mathbb{G}$ can be uniquely expressed as

$$x = z_1^{a_1} z_2^{a_2} \cdots z_k^{a_k},$$

where z_i is a generator of $C(n_i)$ and $0 \leq a_i < n_i$. We can thus represent x by the k-tuple $(a_1, \ldots, a_k) \in \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_k}$. There is a character corresponding to each k-tuple $R = (r_1, \ldots, r_k) \in \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_k}$, defined as follows:

$$\chi_R(x) = \mathrm{e}\left(\sum_{i=1}^{k} \frac{a_i r_i}{n_i} \right).$$

We are particularly interested in the set of characters of the group $\mathbb{Z}_d^n$. In this case, the preceding formula simplifies to

$$\chi_R(a_1, \ldots, a_n) = \mathrm{e}\left(\frac{1}{d} \sum_{i=1}^{n} a_i r_i \right).$$

where $R = (r_1, \ldots, r_n) \in \mathbb{Z}_d^n$.

2.2 Discrete Fourier Transform. We give a brief overview of the basic concepts in discrete Fourier analysis; see [DM], [Kö], or [Ba] for more details.

As before, let $\mathbb{G}$ be a finite abelian group. The set $\mathbb{C}^{\mathbb{G}} = \{f : \mathbb{G} \to \mathbb{C}\}$ of complex functions over the group $\mathbb{G}$ forms a $|\mathbb{G}|$−dimensional vector space over $\mathbb{C}$. The inner product of two functions f and g is defined as follows:

$$< f, g > = \frac{1}{|\mathbb{G}|} \sum_{x \in \mathbb{G}} f(x) g(x)^*,$$

where $*$ denotes the complex conjugate operation. The characters of $\mathbb{G}$ form an orthonormal basis of $\mathbb{C}^{\mathbb{G}}$ with respect to the inner product $<,>$.

Any function $f \in \mathbb{C}^{\mathbb{G}}$ can be uniquely written as a linear combination of characters:

$$f = \sum_{\chi \in \widehat{\mathbb{G}}} \widehat{f}_\chi \chi.$$

The coefficients $\widehat{f}_\chi$ are called the *Fourier coefficients* of f, and are given by $\widehat{f}_\chi = < f, \chi >$. We use the term *principal Fourier coefficient* for $\widehat{f}_{\chi_0}$, the Fourier coefficient corresponding to the principal character.

The function $\widehat{f} : \widehat{\mathbb{G}} \to \mathbb{C}$ is the *Fourier transform* of f.

A *probability distribution* over $\mathbb{G}$ is a function $\mathcal{D} : \mathbb{G} \to \mathbb{R}$ such that for all $x \in \mathbb{G}$, $\mathcal{D}(x) \geq 0$, and $\sum_{x \in \mathbb{G}} \mathcal{D}(x) = 1$.

In our estimates of the distance between probability distributions over a finite abelian group $\mathbb{G}$, we shall make use of the Fourier transforms of these probability distributions.

As usual, for $1 \leq p \leq \infty$ we shall use $||f||_p$ to denote the L_p-norm of the function $f \in \mathbb{C}^{\mathbb{G}}$, i.e. for $p < \infty$ we set $||f||_p = (\sum_{x \in \mathbb{G}} |f(x)|^p)^{1/p}$; for $p = \infty$ we set $||f||_\infty = \max_{x \in \mathbb{G}} |f(x)|$. Note that for the L_2*-norm* this notion does *not* correspond to the inner product $<,>$.

3 Approximating arbitrary distributions

In this section we suggest an approach to approximating arbitrary distributions. Previous work concentrated on approximating the *uniform* distribution over two-valued random variables. Let $\mathcal{D}$ be a probability distribution over a finite abelian group $\mathbb{G}$. We will show that there exists a *small* probability space which approximates $\mathcal{D}$. The following is a somewhat strengthened version of our original theorem, based on an observation due to Mario Szegedy.

Let Γ be a sample space of size ℓ, and $\mathcal{D}$ be a probability distribution over Γ. In what follows, we will often represent such a distribution $\mathcal{D}$ by a (probability) vector D in ℓ dimensions.

Theorem 3.1 *Let M be an $\ell \times \ell$ matrix of complex numbers with entries of absolute value at most 1. For any probability distribution $\mathcal{D}$ represented by the vector D of length ℓ, and any $\epsilon > 0$, there exists a probability distribution $\mathcal{F}$ represented by a vector F with at most $O(\epsilon^{-2} \log \ell)$ non-zero entries, such that*

$$||D \cdot M - F \cdot M||_\infty \leq \epsilon.$$

Proof: We use the probabilistic method [ES74, AS] to demonstrate the existence of a sample space $\Omega \subset \Gamma$ such that a *uniformly* chosen sample point from Ω has a distribution approximating $\mathcal{D}$; thus, $\mathcal{F}$ is the uniform distribution over Ω. We choose $\Omega = \{\omega_1, \ldots, \omega_k\}$ as follows: pick each ω_i independently from Γ according to the distribution $\mathcal{D}$. Since the sample points $\omega_i \in \Gamma$ need not be distinct, in general, Ω will be a multi-set; if necessary, the repetitions can be eliminated by suitably modifying the probability measure.

We index the rows of M by $s \in S$. We claim that, provided k is large enough, for every $s \in S$ the probability that $|D \cdot M_s - F \cdot M_s| > \epsilon$ is less than $1/\ell$. Since the number of rows is ℓ, this implies that

$$\Pr[\exists\, s,\ |D \cdot M_s - F \cdot M_s| > \epsilon] < 1.$$

Note that the probability in the above expression is with respect to the random choice of Ω. Thus, it follows that there exists a choice of the elements $\omega_i \in \Gamma$, for $1 \leq i \leq k$, which will yield the probability space $(\Omega, \mathcal{F})$ as required.

It remains to prove the claim. Let us now concentrate only on the row indexed by a specific s. For $1 \le i \le k$, let w_i be the jth coordinate of the sth row, where j is the index of the element that was chosen as ω_i. It follows that

$$F \cdot M_s = \sum_{i=1}^{k} \frac{1}{k} w_i,$$

that is $F \cdot M_s$ is proportional to the sum of k independent random variables. In what follows, E and Pr denote expectation and probability with respect to the uniform measure on the (multi-set) sample space Ω. We have that $\mathrm{E}[w_i] = D \cdot M_s$ and

$$\mathrm{E}[F \cdot M_s] = \mathrm{E}\left[\frac{1}{k}\sum_{i=1}^{k} w_i\right] = \mathrm{E}[w_i] = D \cdot M_s.$$

To complete the proof we show that the sum of the w_i does not differ from its expected value by more than ϵk. Let S be sum of n independent variables, each of which has an absolute value of at most 1. By a version of the Chernoff bound [AS, p.240], for any $h \ge 0$,

$$\Pr[\,|S - \mathrm{E}[S]| \ \ge\ h\,] \ \le\ 2e^{-\Omega(h^2/n)}.$$

This bound implies that

$$\Pr\left[\left|\sum_{i=1}^{k} w_i - kD \cdot M_s\right| > \delta\right] \ \le\ 2e^{-\Omega(\delta^2/k)}.$$

In our case, the bound on the allowed deviation from the expected value is $\delta = \epsilon k$. We need to choose k such that $e^{-\Omega(\delta^2/k)} < 1/2\ell$. This is clearly true for $k = \Theta(\epsilon^{-2}\log \ell)$. $\square$

The following theorem shows the *existence* of a good approximation $(\Omega, \mathcal{F})$ to the distribution $\mathcal{D}$ such that the sample space Ω is small.

Theorem 3.2 *For any probability distribution $\mathcal{D}$ defined over a finite abelian group $\mathbb{G}$ and any $\epsilon \in [0,1]$, there exists a probability space $(\Omega, \mathcal{F})$, such that:*

1. $\|\widehat{\mathcal{F}} - \widehat{\mathcal{D}}\|_\infty \le \epsilon/|\mathbb{G}|$,

2. *the size of the probability space Ω is at most $O(\epsilon^{-2}\log|\mathbb{G}|)$.*

Proof: The proof is an immediate consequence of Theorem 3.1. We choose M to be the character table of the group $\mathbb{G}$, i.e., the rows are indexed by the characters, the columns by the elements of $\mathbb{G}$, and $M_{sx} = \chi_s(x)$. $\square$

The following Corollary shows the existence of a good approximation to the uniform distribution over $\mathbb{Z}_d^n$.

Corollary 3.3 *There exists a probability distribution $\mathcal{F}$ over $\mathbb{Z}_d^n$ of size $O(\epsilon^{-2} n \log d)$ such that the value of all of its Fourier coefficients (except for the principal coefficient) is at most ϵ/d^n.*

We now discuss the significance of Theorem 3.2.

Definition 3.4 *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two probability distributions over a finite abelian group $\mathbb{G}$. We define the* variation distance *between these two distributions as $\|\mathcal{D}_1 - \mathcal{D}_2\|_1$.*

The next theorem bounds the variation distance between $\mathcal{D}$ and $\mathcal{F}$ in terms of their Fourier coefficients.

Theorem 3.5 *Let the probability distributions $\mathcal{D}$ and $\mathcal{F}$ be defined over a finite abelian group $\mathbb{G}$. Then,*

$$\|\mathcal{D} - \mathcal{F}\|_1 \leq |\mathbb{G}| \cdot \|\widehat{\mathcal{D}} - \widehat{\mathcal{F}}\|_2 \leq |\mathbb{G}| \cdot \|\widehat{\mathcal{D}} - \widehat{\mathcal{F}}\|_1.$$

Proof: The right inequality is immediate. Let $\mathcal{H} = \mathcal{D} - \mathcal{F}$. Using the Cauchy-Schwarz Inequality and Parseval's Equality, we conclude that

$$\|\mathcal{H}\|_1 \leq \sqrt{|\mathbb{G}|} \cdot \|\mathcal{H}\|_2 = |\mathbb{G}| \cdot \|\widehat{\mathcal{H}}\|_2.$$

□

Let $X_1, \ldots, X_n$ be random variables taking values from $\mathbb{Z}_d$. Let $\mathcal{D} : \mathbb{Z}_d^n \to \mathbb{R}$ denote their joint probability distribution. Let $S \subseteq \{1, \ldots n\}$ be of cardinality k. For any $x \in \mathbb{Z}_d^n$, let $x|_S$ denote the projection of the vector x specified by S. We define $\mathcal{D}_S$, the *restriction* of $\mathcal{D}$ to S, by setting

$$\mathcal{D}_S(X_S = y) = \sum_{x \in \mathbb{Z}_d^n,\, x|_S = y} \mathcal{D}(x)$$

for all $y \in \mathbb{Z}_d^k$.

We first observe the following relation between the Fourier coefficients of $\mathcal{D}$ and $\mathcal{D}_S$. Let $\mathcal{A} \subset Z_d^n$ denote the set of elements $(a_1, \ldots, a_n)$ in Z_d^n for which $a_i = 0$ for all $i \notin S$.

Lemma 3.6 *For all $A \in \mathcal{A}$,*

$$d^{n-k} \cdot u = v$$

where u is the Fourier coefficient of $\mathcal{D}$ corresponding to A, and v is the Fourier coefficient of $\mathcal{D}_S$ corresponding to $A|_S$.

Proof: The proof follows directly by substituting appropriate values into the definition of Fourier coefficients. □

Corollary 3.7 *Let $\mathcal{D}$ and $\mathcal{F}$ be probability distributions defined over $\mathbb{Z}_d^n$ such that $\|\widehat{\mathcal{D}} - \widehat{\mathcal{F}}\|_\infty \leq \epsilon/d^n$ for some $0 \leq \epsilon \leq 1$. Then, for any subset S of cardinality k of the random variables,*

$$\|\mathcal{D}_S - \mathcal{F}_S\|_1 \leq \epsilon d^k.$$

Proof: Applying Theorem 3.5 and Lemma 3.6, we conclude that

$$\begin{aligned} \|\mathcal{D}_S - \mathcal{F}_S\|_1 &\leq d^k \cdot \|\widehat{\mathcal{D}}_S - \widehat{\mathcal{F}}_S\|_1 \\ &\leq d^{2k} \cdot \|\widehat{\mathcal{D}}_S - \widehat{\mathcal{F}}_S\|_\infty \\ &\leq d^{2k} \cdot d^{n-k} \cdot \|\widehat{\mathcal{D}} - \widehat{\mathcal{F}}\|_\infty \\ &\leq d^{n+k} \cdot \frac{\epsilon}{d^n} = \epsilon d^k, \end{aligned}$$

which completes the proof. □

If ϵ is chosen to be polynomially small, then Corollary 3.7 implies that: for any distribution $\mathcal{D}$, there exists a distribution $\mathcal{F}$ over a polynomial size sample space such that any subset S of the random variables is distributed in $\mathcal{F}$ "almost" as in $\mathcal{D}$, provided that $|S| = O(\log_d n)$.

4 Bias and k-wise near-independence

In this section we define the notion of a ϵ-biased distribution. (This distribution has been studied earlier [NN, AGHP] for the case $d = 2$). Generalized ϵ-biased distributions represent a convenient formalization of the concept of "good" approximation to the uniform distribution. Our main result here is a theorem that bounds the Fourier coefficients of a probability distribution over $\mathbb{Z}_d^n$ in terms of the bias of the distribution. We also give a bound on the variation distance of a distribution from the uniform distribution in terms of the Fourier coefficients.

We first generalize the definition of ϵ-biased distributions to the case of multi-valued random variables. Let $X = (X_1, \ldots, X_n)$ be a random variable over a set $\Omega \subseteq \mathbb{Z}_d^n$. We define the bias of X with respect to any $A \in \mathbb{Z}_d^n$ as follows.

Definition 4.1 *Let $A = (a_1, \ldots, a_n)$ be any vector in $\mathbb{Z}_d^n$ and let $g = \gcd(a_1, \ldots, a_n, d)$. The bias of A is defined to be*

$$\mathrm{bias}(A) = \frac{1}{g} \max_{0 \leq k < \frac{d}{g}} \left| \Pr\left[\sum_{i=1}^{n} a_i X_i \equiv kg \pmod{d} \right] - \frac{g}{d} \right|.$$

We introduce g in this definition because, regardless of the distribution of the random variables, the only values that $\sum_{i=1}^{n} a_i X_i \pmod{d}$ can take are multiples of g.

Definition 4.2 *Let $0 \leq \epsilon \leq 1$ and let $\Omega \subseteq \mathbb{Z}_d^n$. A probability space $(\Omega, \mathcal{P})$ is said to be ϵ-biased if the corresponding random variable $X = (X_1, \ldots, X_n)$ has the following properties.*

1. *For $1 \leq i \leq n$, X_i is uniformly distributed over $\mathbb{Z}_d$.*
2. *For all vectors $A \in \mathbb{Z}_d^n$, $\mathrm{bias}(A) \leq \epsilon$.*

We first note that Theorem 3.1 implies that an ϵ-biased probability space of small size exists. In Section 5 we provide an explicit construction which is somewhat weaker.

Corollary 4.3 *There exists a probability distribution $\mathcal{F}$ over $\mathbb{Z}_d^n$ of size $O(\epsilon^{-2} n \log d)$ such that for all $A = (a_1, \ldots, a_n) \in \mathbb{Z}_d^n$, $\text{bias}(A) \leq \epsilon$.*

Proof: The proof follows immediately from Theorem 3.1 by the following choice of matrix M. Let the columns of M correspond to the elements of $\mathbb{Z}_d^n$, and the rows of M correspond to all pairs (A, k) such that $A = (a_1, \ldots, a_n) \in \mathbb{Z}_d^n$, $0 \leq k < d/g$, where $g = \gcd(a_1, \ldots, a_n, d)$. Let $X = (X_1, \ldots, X_n) \in \mathbb{Z}_d^n$. We define:

$$M((A,k), X) = \begin{cases} 1 & \text{if } \sum_{i=1}^n a_i X_i \equiv kg \pmod{d} \\ 0 & \text{otherwise} \end{cases}$$

In order to apply Theorem 3.1, we transform matrix M into a square matrix by adding zero rows. $\square$

Let $\mathcal{D}$ be an ϵ-biased distribution. We now relate the bias and the Fourier coefficient for any $A \in \mathbb{Z}_d^n$ as follows.

Lemma 4.4 *For all non-zero $A = (a_1, \ldots, a_n) \in \mathbb{Z}_d^n$, we have that*

$$|\widehat{\mathcal{D}}_A| \leq \frac{\text{bias}(A)}{d^{n-1}}.$$

Proof: Let $g = \gcd(a_1, \ldots, a_n, d)$. By the definition of a Fourier coefficient,

$$\begin{aligned} \widehat{\mathcal{D}}_A &= <\mathcal{D}, \chi_A> \\ &= \frac{1}{d^n} \sum_x \mathcal{D}(x) \chi_A(x)^* \\ &= \frac{1}{d^n} \sum_x \mathcal{D}(x) \left(\mathrm{e}\left(\frac{1}{d} \sum_{i=1}^n a_i x_i \right) \right)^* \\ &= \frac{1}{d^n} \sum_x \mathcal{D}(x)\ \mathrm{e}\left(-\frac{1}{d} \sum_{i=1}^n a_i x_i \right). \end{aligned}$$

Taking absolute values, we have that

$$\begin{aligned} |\widehat{\mathcal{D}}_A| &= \frac{1}{d^n} \left| \sum_x \mathcal{D}(x)\ \mathrm{e}\left(-\frac{1}{d} \sum a_i x_i \right) \right| \\ &= \frac{1}{d^n} \left| \sum_{k=0}^{\frac{d}{g}-1} \mathrm{e}\left(\frac{-kg}{d} \right) \Pr\left[\sum a_i x_i \equiv kg \pmod{d} \right] \right|. \end{aligned}$$

The probability is with respect to a random choice of $x \in \mathbb{Z}_d^n$ with the distribution $\mathcal{D}$.

Define $P_{kg} = \Pr[\sum a_i x_i \equiv kg \pmod{d}]$. Then,

$$\begin{aligned} |\widehat{\mathcal{D}}_A| &= \frac{1}{d^n}\left|\sum_{k=0}^{\frac{d}{g}-1} \mathrm{e}\left(\frac{-kg}{d}\right) P_{kg}\right| \\ &= \frac{1}{d^n}\left|\frac{g}{d}\sum_{k=0}^{\frac{d}{g}-1} \mathrm{e}\left(\frac{-kg}{d}\right) + \sum_{k=0}^{\frac{d}{g}-1} \mathrm{e}\left(\frac{-kg}{d}\right)\left(P_{kg} - \frac{g}{d}\right)\right| \end{aligned}$$

Note that $\sum \mathrm{e}\left(\frac{-kg}{d}\right) = 0$ since the (d/g)th roots of unity sum to zero. We then conclude that

$$\begin{aligned} |\widehat{\mathcal{D}}_A| &= \frac{1}{d^n}\left|\sum_{k=0}^{\frac{d}{g}-1} \mathrm{e}\left(\frac{-kg}{d}\right)\left(P_{kg} - \frac{g}{d}\right)\right| \\ &\leq \frac{1}{d^n}\sum_{k=0}^{\frac{d}{g}-1}\left|\mathrm{e}\left(\frac{-kg}{d}\right)\right|\left|P_{kg} - \frac{g}{d}\right| \\ &\leq \frac{1}{d^n}\cdot\frac{d}{g}\cdot(g\cdot\mathrm{bias}(A)) \\ &= \frac{\mathrm{bias}(A)}{d^{n-1}}, \end{aligned}$$

where the last inequality follows from the definition of the bias as well as the fact that $|\,\mathrm{e}(-kg/d)\,| = 1$. □

The following theorem is a generalization of a result due to Vazirani [Va]. It relates the biases of an arbitrary distribution to its variation distance from the uniform distribution.

Theorem 4.5 *Let $\mathcal{D}$ be an arbitrary probability distribution defined on $\mathbb{Z}_d^n$, and let $\mathcal{U}$ denote the uniform distribution on $\mathbb{Z}_d^n$. Then,*

$$||\mathcal{D} - \mathcal{U}||_1 \leq d\sum_A \mathrm{bias}(A),$$

where the bias is defined with respect to the distribution $\mathcal{D}$.

Proof: We first evaluate $\widehat{\mathcal{D}}_{\vec{0}}$,

$$\widehat{\mathcal{D}}(\vec{0}) = <\mathcal{D}, \chi_{\vec{0}}> = \sum_{x\in\mathbb{Z}_n^d} \frac{\mathcal{D}(x)}{d^n} = \frac{1}{d^n}.$$

The variation distance is,

$$||\mathcal{D} - \mathcal{U}||_1 = \sum_{x\in\mathbb{Z}_n^d}\left|\mathcal{D}(x) - \frac{1}{d^n}\right| = \sum_{x\in\mathbb{Z}_n^d}\left|\sum_A \widehat{\mathcal{D}}_A\chi_A(x) - \frac{1}{d^n}\right|.$$

Since $\widehat{\mathcal{D}}_{\vec{0}} = \frac{1}{d^n}$,

$$\begin{aligned}
\sum_{x \in \mathbb{Z}_n^d} \left| \sum_A \widehat{\mathcal{D}}_A \chi_A(x) - \frac{1}{d^n} \right| &= \sum_{x \in \mathbb{Z}_n^d} \left| \sum_{A \neq \vec{0}} \widehat{\mathcal{D}}_A \chi_A(x) \right| \\
&\leq \sum_{x \in \mathbb{Z}_n^d} \sum_{A \neq \vec{0}} \left| \widehat{\mathcal{D}}_A \right| |\chi_A(x)| \\
&= d^n \sum_{A \neq \vec{0}} \left| \widehat{\mathcal{D}}_A \right| \\
&\leq d^n \sum_A \frac{d}{d^n} \text{bias}(A),
\end{aligned}$$

where the last inequality follows from Lemma 4.4. Thus,

$$||\mathcal{D} - \mathcal{U}||_1 \leq d \sum_{A \neq \vec{0}} \text{bias}(A).$$

□

Corollary 4.6 *For $\epsilon = 0$, an ϵ-biased distribution is the same as the uniform distribution.*

The following definition is similar to that of Naor and Naor [NN] and Ben-Nathan [Be].

Definition 4.7 *Let $X_1, \ldots, X_n$ be random variables taking values from $\mathbb{Z}_d$. Let $\mathcal{D} : \mathbb{Z}_d^n \to \mathbb{R}$ denote their joint probability distribution. For any $x \in \mathbb{Z}_d^n$, let $x|_S$ denote the projection of the vector x specified by S. Let $\mathcal{D}_S$ denote the restriction of $\mathcal{D}$ to S, by setting*

$$\mathcal{D}_S(X_S = y) = \sum_{x \in \mathbb{Z}_d^n, x|_S = y} \mathcal{D}(x)$$

for all $y \in \mathbb{Z}_d^k$. We say that the variables $X_1, \ldots, X_n$ are k-wise δ-dependent if for all subsets S such that $|S| \leq k$,

$$||\mathcal{D}(S) - \mathcal{U}(S)||_1 \leq \delta,$$

where $\mathcal{U}$ denotes the uniform distribution.

The next Corollary follows from Theorem 4.5 and Corollary 3.7.

Corollary 4.8 *If the random variables $X_1, \ldots, X_n$ taking values from $\mathbb{Z}_d$ are ϵ-biased, then they are also k-wise δ-dependent, for $\delta = \epsilon d^k$. In particular, they are $(\log_d n)$-wise $(1/poly(n))$-dependent with a polynomially small ϵ.*

5 Constructing an ϵ-biased probability distribution

In this section we show how to approximate a uniform probability distribution over $\mathbb{Z}_d^n$. We present an explicit construction of ϵ-biased random variables such that the sample space Ω

has size which is bounded by a small polynomial in n, d and $1/\epsilon$. This implies that we have an explicit construction for random variables which are almost $(\log_d n)$-wise independent, such that the corresponding sample space is of polynomial size, where ϵ is polynomially small.

We describe the ϵ-biased probability distribution implicitly by specifying an algorithm for choosing a random sample point. In what follows, we assume that the prime power q and a character $\mathcal{X}_r$ of $\mathbb{F}_q^*$ are chosen such that $d = (q-1)/r$ is the order of the character; further, we assume that $q - 1 \geq n$. Let z be a generator for $\mathbb{F}_q^*$. Let $b_1, b_2, \ldots, b_n$ be some fixed distinct elements in $\mathbb{F}_q$.

Random Sampling Algorithm:

1. Choose the value of the random variable Y from $\mathbb{F}_q^*$ uniformly at random. For $1 \leq i \leq n$, let $Y_i = Y + b_i$.
2. For $1 \leq i \leq n$:
   (a) Let $Z_i = \begin{cases} Y_i & \text{if } Y_i \neq 0 \\ b_i & \text{otherwise} \end{cases}$
   (b) Let s_i be such that $Z_i = z^{s_i}$.
   (c) Choose $X_i = s_i \bmod d$.

In Step 2(a), we take care of the case where one of the random variables $Y_i = Y + b_i$ is zero and, therefore, not in $\mathbb{F}_q^*$. This guarantees that each Z_i is uniformly distributed over $\mathbb{F}_q^*$.

Let Θ be a primitive dth root of unity and let $\mathcal{X}_r(x) = \Theta^{\log x}$, where $\log x$ denotes the discrete log of x to the base z. Notice that for all i, provided each Y_j is non-zero,

$$\mathcal{X}_r(Y + b_i) = \mathcal{X}_r(Z_i) = \mathcal{X}_r(z^{s_i}) = \Theta^{s_i} = \Theta^{X_i}. \tag{1}$$

We will establish that these random variables have the desired properties via Weil's character sum estimates (see Schmidt [Sc, page 43, Theorem 2C]). Let f be a polynomial over a field $\mathbb{F}$. Let k be the greatest common divisor of the multiplicities of the roots of f over the algebraic closure of $\mathbb{F}$. We shall say that k is the *greatest common multiplicity* of f.

Theorem 5.1 (Weil's Theorem) *Let $\mathbb{F}$ be a finite field of order q and let $\mathcal{X}$ be a multiplicative character of order d. Let $f \in \mathbb{F}[x]$ be a polynomial with n distinct zeros in the algebraic closure of $\mathbb{F}$. Suppose d does not divide the greatest common multiplicity of f. Then*

$$\left| \sum_{x \in \mathbb{F}} \mathcal{X}(f(x)) \right| \leq (n-1)\sqrt{q}.$$

To analyze the properties of our construction, we need the following corollary.

Corollary 5.2 *Let $\mathbb{F}$ be a finite field of order q and let Θ be a primitive dth root of unity. Let $f \in \mathbb{F}[x]$ be a polynomial with n distinct roots in the algebraic closure of $\mathbb{F}$. Assume that the greatest common multiplicity of f is relatively prime to d. Define r_k to be the number of solutions $x \in \mathbb{F}$ to the equation $\chi(f(x)) = \Theta^k$. Then,*

$$\left| r_k - \frac{q}{d} \right| \leq (n-1)\sqrt{q}.$$

Proof: The definition of r_j implies that for $0 \leq \ell \leq d-1$,

$$\sum_{x \in \mathbb{F}} (\chi(f(x)))^\ell = \sum_{j=0}^{d-1} r_j \Theta^{\ell j}. \tag{2}$$

(Here for $\ell = 0$ we set $0^\ell = 0$.) Denoting the number of distinct roots of f in $\mathbb{F}$ by ν ($\nu \leq n$), it follows that

$$\begin{aligned} q - \nu + \sum_{\ell=1}^{d-1} \sum_{x \in \mathbb{F}} (\chi(f(x)))^\ell &= \sum_{\ell=0}^{d-1} \sum_{x \in \mathbb{F}} (\chi(f(x)))^\ell \\ &= \sum_{\ell=0}^{d-1} \sum_{j=0}^{d-1} r_j \Theta^{\ell j} \\ &= d r_0 + \sum_{j=1}^{d-1} r_j \sum_{\ell=0}^{d-1} \Theta^{j\ell} \\ &= d r_0. \end{aligned}$$

Hence,

$$|d r_0 - q| \leq \nu + \left| \sum_{\ell=1}^{d-1} \sum_{x \in \mathbb{F}} (\chi(f(x)))^\ell \right| \leq \nu + \sum_{\ell=1}^{d-1} \left| \sum_{x \in \mathbb{F}} (\chi(f(x)))^\ell \right|.$$

The order of the character χ^ℓ is $d' = d/\gcd(d, \ell)$ which is greater than 1 for $0 < \ell < d$ and it is relatively prime to the greatest common multiplicity of f, hence we may apply Theorem 5.1 to each term on the right hand side. We obtain

$$|d r_0 - q| \leq \nu + \sum_{\ell=1}^{d-1} (n-1)\sqrt{q} \leq \nu + (d-1)(n-1)\sqrt{q} < d(n-1)\sqrt{q}.$$

This implies that

$$\left| r_0 - \frac{q}{d} \right| \leq (n-1)\sqrt{q}.$$

To conclude the proof, observe that if both sides of (2) are multiplied by $\Theta^{-k\ell}$, then the same result is obtained for any r_k. □

We now analyze the properties of the random variables defined above.

Theorem 5.3 *Let the random variables $X_1, \ldots, X_n$ be defined by the Random Sampling Algorithm. Then the following two conditions hold:*

1. *For $1 \leq i \leq n$, X_i is uniformly distributed over $\mathbb{Z}_d$.*

2. *For all $A \in \mathbb{Z}_d^n$, $\mathrm{bias}(A) \leq 2n/\sqrt{q}$.*

Proof: For each i, distinct values of $Y \in \mathbb{F}_q^*$ yield distinct values for $Z_i \in \mathbb{F}_q^*$. Since Y is chosen uniformly at random from $\mathbb{F}_q^*$, it follows that Z_i is uniformly distributed over $\mathbb{F}_q^*$. Since $\mathbb{F}_q^*$ is cyclic, we conclude that the random variable s_i is uniformly distributed over $\{0, 1, \ldots, q-2\}$. By our choice of q, we have $d|q-1$, and this implies that $X_i \equiv s_i \pmod d$ is uniform over the set $\mathbb{Z}_d$, thereby establishing the first part of the theorem.

Let $A = (a_1, \ldots, a_n)$ be any vector in $\mathbb{Z}_d^n$ and let $g = \gcd(a_1, \ldots, a_n, d)$. Assume first that $g = 1$. We define the polynomial $f_A(x)$ as follows:

$$f_A(x) = \prod_{i=1}^{n} (x + b_i)^{a_i}.$$

Let us now restrict ourselves to the case where all values Y_j are non-zero. By (1),

$$\chi_r(f_A(Y)) = \prod_{i=1}^{n} [\chi_r(Y + b_i)]^{a_i} = \prod_{i=1}^{n} (\Theta^{X_i})^{a_i} = \Theta^{\sum_{i=1}^{n} a_i X_i}.$$

The number of values of $Y \in \mathbb{F}_q^*$ such that $\sum_{i=1}^{n} a_i X_i \equiv j \pmod d$ is exactly equal to the value of r_j defined in Corollary 5.2 for the polynomial $f_A(x)$. However, we are only considering the case where all values Y_j are non-zero. This can create at most an additive error of n in the bounds given in Theorem 5.1 and Corollary 5.2. It then follows from the definition of the bias that

$$\mathrm{bias}(A) \leq \max_j \frac{|r_j - q/d| + n}{q}.$$

The assumption $g = 1$ means the greatest common multiplicity of f is relatively prime to d. From Corollary 5.2 it follows that

$$\mathrm{bias}(A) \leq \frac{n}{\sqrt{q}} + \frac{n}{q} \leq \frac{2n}{\sqrt{q}}.$$

Consider now the case $g > 1$, and let $\beta_i = a_i/g$. Let $g' = \gcd(a_1, \ldots, a_n)$ and $c_i = a_i/g'$. By the preceding argument, the bias of the vector $C = (c_1, \ldots, c_n)$ is bounded by $2n/\sqrt{q}$. For $0 \leq j \leq d/g - 1$, the number of vectors X that satisfy the equation

$$\sum_{i=1}^{n} a_i X_i \equiv jg \pmod d.$$

is equal to the number of vectors X that satisfy

$$\sum_{i=1}^{n} c_i X_i \equiv j + \frac{dl}{g} \pmod d,$$

where $0 \le l \le g-1$. Since g'/g is relatively prime to d/g, the number of such vectors is also equal to the number of vectors X that satisfy

$$\sum_{i=1}^{n} c_i X_i \equiv j' + \frac{dl'}{g} \pmod{d}$$

where $j \equiv j'(g'/g) \pmod{d/g}$. By Definition 4.3,

$$\text{bias}(A) = \frac{1}{d} \cdot d \cdot \text{bias}(G) = \text{bias}(G),$$

which establishes the second part of the theorem. □

The parameters of our construction are described in the following theorem. Let $q(d,k)$ denote the smallest prime power such that $d|q-1$ and $d \ge k$.

Theorem 5.4 *For any $\epsilon > 0$, $n \ge 2$, and $d \ge 2$, the probability space $(\Omega, \mathcal{P})$ defined by the Random Sampling Algorithm generates n random variables over $\mathbb{Z}_d$ which are ϵ-biased, and the size of the sample space is $|\Omega| = q(d, 4n^2\epsilon^{-2}) - 1$.*

Proof: Generating the random variables X only requires choosing $Y \in \mathbb{F}_q^*$ uniformly at random. Hence the sample space is $\Omega = \mathbb{F}_q^*$ where q is a prime such that $d|q-1$ and $q \ge 4n^2\epsilon^{-2}$. Choose the smallest prime power satisfying these constraints. □

6 Estimates for $q(d,k)$

In this section we review results from number theory relevant to estimating $q(d,k)$. Let $p(d,k)$ denote the smallest prime such that $d|p-1$ and $d \ge k$. Clearly $q(d,k) \le p(d,k)$.

For any d and k, the quantity $p(d,k)$ can be estimated using *Linnik's Theorem* establishing the existence of small primes in arithmetic progressions: among the integers $\equiv t \pmod{m}$ (where $\gcd(m,t) = 1$) there exists a prime $p = O(m^C)$. Heath-Brown [H] proves $C \le 11/2$. Note that this result does *not* depend on any hypothesis. Under the Extended Riemann Hypothesis, Bach and Sorenson [BS] prove that p can be chosen to be $\le 2(m \ln m)^2$, hence $C \le 2 + o(1)$.

Let now $p(d)$ denote the smallest prime such that $d|p-1$. Let further $m(d,k)$ be the smallest integer m such that $d|m$ and $m \ge k$. Note that $m(d,k) < d+k$. Note further that $p(d,k) \le p(m(d,k))$. Summarizing, there exist absolute constants c and C such that

$$p(d,k) < c(d+k)^C. \tag{3}$$

Here C is the exponent in Linnik's Theorem discussed above.

The exponent C can be reduced to 1 if d is small compared to k. For fixed d we have

$$p(d,k) < (1+o(1))k. \tag{4}$$

Moreover, for any constant $c > 0$ and for any $d \leq \log^c k$ we have

$$p(d,k) < c_1 k, \tag{5}$$

where the constant c_1 depends only on c. These bounds follow from results that in this range, the primes are nearly uniformly distributed among the $\bmod d$ residue classes which are relatively prime to d (prime number theorem for arithmetic progressions, cf. [Da, pp. 132-133]).

In conclusion we summarize the bounds obtained for the size of the sample space.

Theorem 6.1 *For any $\epsilon > 0$, $n \geq 2$, and $d \geq 2$, the probability space $(\Omega, \mathcal{P})$ defined by the Random Sampling Algorithm generates n random variables over $\mathbb{Z}_d$ which are ϵ-biased, and the size of the sample space is $|\Omega| < c_0(d + n^2\epsilon^{-2})^C$ where C is the constant in Linnik's Theorem. Moreover, if $d \leq \log^c(n^2\epsilon^{-2})$ then we have $|\Omega| < c_1 n^2 \epsilon^{-2}$ where c_1 depends on c only. For constant d we have $|\Omega| < (1+o(1))n^2\epsilon^{-2}$.*

Note that the bounds obtained in the above theorem are not the best possible, compare with Corollary 4.3. Theorem 6.1 together with Corollary 4.8 imply that we can construct $(\log_d n)$-wise $(1/poly(n))$-dependent random variables over $\mathbb{Z}_d^n$ using a polynomially large sample space. Also, Theorem 6.1 together with Lemma 4.4 imply that we can approximate the Fourier coefficients of the uniform distribution on $\mathbb{Z}_d^n$ within ϵ/d^n with a sample space of size $O(\epsilon^{-2}n^2d^2)$ for small d. This construction may not be the best possible since Corollary 3.3 guarantees the existence of an approximating sample space whose size is $O(\epsilon^{-2} n \log d)$.

7 Linear codes

In this section we observe that the ϵ-biased distribution can also be looked upon as a construction of a nearly uniform linear code. The linear code that we obtain has a large distance and the interesting property that each non-zero codeword has roughly the same number of occurrences of each possible symbol in the alphabet, or the field, over which the code is defined. Also, the length of the codewords is only polynomial (quadratic) in the dimension of the code and thus the code is relatively dense.

A *code* C is called an $[n,k]$ code if it transforms words of length k into codewords of length n. The dimension of C is defined to be k. A *linear code* C is a linear subspace of $\mathbb{F}^n$, for some field $\mathbb{F}$. A *generator matrix* G for a linear code C is a $k \times n$ matrix whose rows form a basis for C. If G is a generator matrix for C, then the code can be defined as

$$C = \{a \cdot G \mid a \in \mathbb{F}^k\}.$$

The distance between two codewords is defined to be their *Hamming* distance. The *weight* of a codeword is the number of non-zero symbols that it contains.

We may interpret the sample space of an ϵ-biased distribution as the generator matrix G of a particular linear code C_ϵ. Let q be a prime power chosen in accordance with Theorem 5.4; the generator matrix G is of dimension $n \times q$ and every column in G is a possible assignment to the random variables $X_1, \ldots, X_n$. Let $N(c, k)$ denote the number of occurrences of the letter k in codeword c. The following corollary is a consequence of Theorem 5.3.

Corollary 7.1 *For every codeword $c \in C_\epsilon$ and letter $k \in \{0, \ldots, d-1\}$ where d is a prime,*

$$\left| N(c, k) - \frac{q}{d} \right| \leq q\epsilon = 2n\sqrt{q}.$$

It is well known that for linear codes, the minimum distance between any two codewords is equal to the minimum (positive) weight among all codewords. It follows from the above theorem that a codeword can contain at most $q(\epsilon + 1/d)$ zero entries and hence, the minimum distance of C_ϵ is $q(1 - \epsilon - 1/d)$.

We note that a construction of a code which has the property that for *every* codeword, the distribution of the alphabet symbols is almost uniform and that the length of the codeword is polynomial in the dimension has been known for the case of a binary alphabet. The dual code of a binary BCH code has this property and the proof follows from Weil's Theorem (see MacWilliams and Sloane [MS, pages 280–282]).

8 Open Problems

An important direction for further work is to efficiently construct (in time polynomial in the number of random variables n) probability distributions that approximate special types of joint distributions. In particular, can we construct in time polynomial in n a good approximation to the joint distribution where each random variable independently takes value 1 with probability p and 0 with probability $1 - p$? Note that this is only known for the case where $p = 1/2$.

It is also not clear that our construction of an ϵ-biased distribution on n d-valued random variables is the best possible. Theorem 3.2 guarantees the existence of such a distribution using a smaller sample space (by a factor of n). Can this be achieved constructively?

Acknowledgements

The authors would like to thank Noga Alon and Moni Naor for several helpful discussions. We are also grateful to Laci Babai for simplifying the proof of Corollary 5.2 and to Mario Szegedy for his remarks concerning Theorem 3.1. Special thanks go to the anonymous referees and the editor Laci Babai for efforts above and beyond the call of duty towards improving the quality of this article.

References

[AGHP] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple Constructions of Almost k-wise Independent Random Variables. *Random Structures and Algorithms*, 3 (1992), pp. 289–304.

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

[AS] N. Alon and J. Spencer. *The Probabilistic Method.* John Wiley (1992).

[Ba] L. Babai. *Fourier Transforms and Equations over Finite Abelian Groups.* Lecture Notes, University of Chicago (1989).

[BS] E. Bach and J. Sorenson. Explicit Bounds for Primes in Residue Classes. *Mathematics of Computation*, **65** (1996), pp. 1717–1735.

[Be] R. Ben-Nathan. On dependent random variables over small sample spaces. *M.Sc. Thesis*, Hebrew University, Jerusalem, Israel (1990).

[BR] B. Berger and J. Rompel. Simulating $(\log^c n)$-wise independence in NC. *Journal of the ACM*, 38 (1991), pp. 1026–1046.

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

[CGHFRS] B. Chor, O. Goldreich, J. Håstad, J. Friedman, S. Rudich, and R. Smolensky. t-Resilient functions. In *Proceedings of the 26th Annual Symposium on the Foundations of Computer Science* (1985), pp. 396–407.

[Da] H. Davenport. *Multiplicative Number Theory.* 2nd Edition, Springer Verlag (1980).

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

[ES73] P. Erdős and J. Selfridge, On a combinatorial game. *J. Combinatorial Theory*, Ser. B, 14 (1973), pp. 298–301.

[ES74] P. Erdős and J. Spencer. *Probabilistic Methods in Combinatorics.* Akadémiai Kiadó, Budapest (1974).

[Es] T. Estermann. *Introduction to Modern Prime Number Theory.* Cambridge University Press (1969).

[Ev] G. Even. Construction of small probability spaces for deterministic simulation. *M.Sc. Thesis*, Technion, Haifa, Israel (1991).

[HPS] J. Håstad, S. Phillips, and S. Safra. A Well Characterized Approximation Problem. In *Proceedings of the 2nd Israel Symposium on Theory and Computing Systems* (1993), pp. 261–265.

[H] D.R. Heath-Brown. Zero-Free Regions for Dirichlet L-Functions and the Least Prime in an Arithmetic Progression. *Proc. London Math. Soc.* **64** (1991), 265-338.

[KW] R.M. Karp and A. Wigderson. A Fast Parallel Algorithm for the Maximal Independent Set Problem. *Journal of the ACM*, 32 (1985), pp. 762–773.

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

[Le] W. Ledermann. *Introduction to Group Characters.* Cambridge University Press (1987, 2nd edition).

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

[Lu2] M. Luby. Removing randomness in parallel computation without a processor penalty. In *Proceedings of the 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. *Journal of Computer and System Sciences*, 49 (1994), pp. 478-516.

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

[NN] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM Journal on Computing*, 22 (1993), pp. 838–856.

[Pe] R. Peralta. On the randomness complexity of algorithms. CS Research Report TR 90-1, University of Wisconsin, Milwaukee (1990).

[Sc] W.M. Schmidt. *Equations over Finite Fields: An Elementary Approach.* Lecture Notes in Mathematics, v. 536, Springer-Verlag (1976).

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