Efficient Approximation of Product Distributions*

Guy Even[†] Oded Goldreich[‡] Michael Luby[§] Noam Nisan[¶] Boban Veličković [||]

February 1998

Abstract

We describe efficient constructions of small probability spaces that approximate the joint distribution of general random variables. Previous work on efficient constructions concentrate on approximations of the joint distribution for the special case of identical, uniformly distributed random variables.

*Preliminary version has appeared in the *Proceedings of the 24th ACM Symp. on Theory of Computing (STOC)*, pages 10–16, 1992.

†Dept. of Electrical Engineering–Systems, Tel–Aviv University, Ramat–Aviv, Tel–Aviv 69978, Israel. Email: guy@eng.tau.ac.il.

‡Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. Email: oded@wisdom.weizmann.ac.il. Research partially supported by grant No. 89-00312 from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel.

§International Computer Science Institute, Berkeley, CA 94704, USA. Email: luby@icsi.berkeley.edu. Research supported in part by National Science Foundation operating grants CCR-9304722 and NCR-9416101, and United States-Israel Binational Science Foundation grant No. 92-00226.

¶Institute for Computer Science, Hebrew University, Givat–Ram, Jerusalem, Israel. Email: noam@cs.huji.ac.il.

||Equipe de Logique, Université Paris VII, France. Email: boban@logique.jussieu.fr. Research partially done while visiting the International Computer Science Institute and while at Carnegie Mellon University.

1 Introduction

The problem of constructing small sample spaces that "approximate" the independent distribution on n random variables has received considerable attention recently (cf. [7, 1, 19, 2, 4]). The primary motivation for this line of research is that random variables that are "approximately" independent suffice for the analysis of many interesting randomized algorithm, and hence, constructing a *small probability space that "approximates" the independent distribution* yields a way to "derandomize" these algorithms, i.e., convert them to deterministic algorithms of reasonable complexity by using the deterministically constructed sample space in place of the "internal coin tosses" of the algorithm. A typical example of the use of this methodology has been provided by Luby in his work on the maximal independent set problem [17]. Surprisingly, it is often ignored that the random variables used in that work are *neither identically distributed nor uniformly distributed* over some sets, and furthermore that this is likely to be the case in many applications. In contrast, all general constructions (for limited independence), presented so far, apply to random variables *uniformly distributed over the same set* (in most cases the two-element set $\{0,1\}$). Hence, it is of primary importance to investigate the extent to which these constructions can be generalized to deal with the "k-wise approximation" of arbitrary stochastically independent events.

1.1 Definitions of Approximation

Throughout the paper we consider the approximation of product distributions; namely, distributions which are the product of many (say n) independent distributions. Without loss of generality, we assume that each of the individual distributions has a support which is a subset of $\{0, 1, ..., m-1\}$. Thus, a product distribution on n general m-valued random variables is described by a n-by-m probability matrix $\mathcal{P}_{n,m} = \{p_{i,v} : i \in \{1, \ldots, n\}, v \in \{0, \ldots, m-1\}\}$, which is a matrix of non-negative entries such that the sum of the entries in each row equals 1. We refer to this matrix as to the specification matrix. The (i, v)-entry, $p_{i,v}$, specifies the probability that the i^{th} random variable should take on value v.

From $\mathcal{P}_{n,m}$ we want to produce a finite set S that induces a distribution on n random variables $X_1, \ldots, X_n$ which approximates (in the sense defined below) the independent distribution for $\mathcal{P}_{n,m}$. We view S as a sample space that induces a distribution on $X_1, \ldots, X_n$ defined by choosing a sample point uniformly from S. That is, for each point $s \in S$ and each index i, we have $X_i(s) \in \{0, \ldots, m-1\}$, where $X_i(s)$ is the value of the random variable X_i on the sample point s. We let X_I denote the subsequence of random variables indexed by I, and $X_I = V$ denote the event that the subsequence X_I takes on the value sequence $V \in \{0, \ldots, m-1\}^{|I|}$. By $\Pr_S[X_I = V]$ we denote the probability that event $X_I = V$ occurs in the distribution induced by S.

- We say that S is perfect for $\mathcal{P}_{n,m}$ if it induces a distribution on $X_1, \ldots, X_n$ such that, for all $V =< v_1, \ldots, v_n >\in \{0, \ldots, m-1\}^n$,

$$\Pr_S[X_{\langle 1,\ldots,n\rangle} = V] = p_{\langle 1,\ldots,n\rangle, V}$$

  where $p_{I,V} \stackrel{\text{def}}{=} \prod_{j=1}^{\ell} p_{i_j, v_j}$ for $I = \langle i_1, ..., i_\ell \rangle$ and $V = \langle v_1, ..., v_\ell \rangle$.

- We say that S is a (k, ϵ)-approximation for $\mathcal{P}_{n,m}$ if for any subsequence I of size $\ell \leq k$ and for any set of possible values $V \in \{0, \ldots, m-1\}^\ell$,

$$|\Pr_S[X_I = V] - p_{I,V}| \ \leq \ \epsilon \tag{1}$$

We stress that Eq. (1) asserts a bound on the Max-Norm of the difference between X_I and the $p_{I,\cdot}$ vector. Bounds in other norms (e.g., Norm-1) can be easily derived from the Max-Norm bound. Specifically, we say that S is a (k,ϵ)-L1-approximation of $\mathcal{P}_{n,m}$ if for all subsequences I of size k,

$$\frac{1}{2} \cdot \sum_{V \in \{0,\ldots,m-1\}^k} |\Pr_S[X_I = V] - p_{I,V}| \;\leq\; \epsilon \tag{2}$$

- We say that S is an ϵ-approximation for $\mathcal{P}_{n,m}$ if the above holds for $k = n$ (i.e., Eq. (1) holds for all I's and all $V \in \{0,\ldots,m-1\}^{|I|}$).
- We say that S is a k-wise independent approximation for $\mathcal{P}_{n,m}$ if Eq. (1) holds with $\epsilon = 0$.

(Hereafter, the quantification "for $\mathcal{P}_{n,m}$" is omitted for brevity whenever $\mathcal{P}_{n,m}$ is clear from the context.)

All constructions for sample spaces considered in this paper are efficient in the sense that there is a deterministic algorithm which produces the sample space S in time polynomial in the length of the description of S, where the sample space is described as a list of sample points and each sample point is described by an n-ary sequence over $\{0,\ldots,m-1\}$.

1.2 Previous Work on Approximation

All previous work in this area deal with the approximation of *identical* random variables which are *uniformly* distributed over a finite set.[1] Let $\mathcal{U}_{n,m}$ be the probability matrix with all entries equal to $1/m$ that describes the special case of n identically and uniformly distributed m-valued random variables. Thus, $\mathcal{U}_{n,2}$ is the important subcase where all entries are $1/2$ (describing n identically and uniformly distributed boolean-valued random variables). It is easy to prove that S has to be of size at least 2^n to in order to be perfect for $\mathcal{U}_{n,2}$ (or any other joint distribution of n non-degenerate random variables). Previously known approximations to $\mathcal{U}_{n,m}$ are of three forms.

- *k-wise independent approximations*: constructions of sample spaces of size $\max\{n,m\}^k$ that are k-wise independent approximations for $\mathcal{U}_{n,m}$ are given in [7, 1].
- *ϵ-approximations*: constructions of sample spaces of size $\text{poly}(n/\epsilon)$ that are ϵ-approximations for $\mathcal{U}_{n,2}$ are given in [19, 2].
- *(k,ϵ)-approximations*: constructions of sample spaces of size $\text{poly}((k\log n)/\epsilon)$ that are (k,ϵ)-approximations for $\mathcal{U}_{n,2}$ can be derived from the above (via the reduction of [19]).

Efficient constructions of (k,ϵ)-approximations for general $\mathcal{P}_{n,m}$ with size $(\max\{n,k/\epsilon\})^k$, are implicit in many works (cf., [17, 1]): In fact, any k-wise independent approximation to $\mathcal{U}_{n,k/\epsilon}$ can be transformed into a (k,ϵ)-approximation for any $\mathcal{P}_{n,m}$ (by rounding the entries in $\mathcal{P}_{n,m}$ to integer multiples of ϵ/k).[2] For constant k and $\epsilon = \text{poly}(1/n)$, this yields a sample space of size polynomial in n. On the other hand, it has been shown that the sample space has to be of size at least $n^{\lfloor k/2 \rfloor}$ in order to be a k-wise approximation for $\mathcal{U}_{n,2}$ [6], and for non-constant k this is not polynomial in n.

[1] This sentence refers to work prior to the conference publication [9] of the current work. In addition to the discussion in Section 1.6, the reader is referred to [14, 11].

[2] This simple argument does not extend to (k,ϵ)-approximation of $\mathcal{U}_{n,k/\epsilon}$, because we need a bound (of ϵ) on the Norm-1 of the distance to $\mathcal{U}_{n,k/\epsilon}$ rather than a Max-Norm bound. Indeed, we may obtain such Norm-1 bound by using a $(k,\epsilon/m^k)$-approximation of $\mathcal{U}_{n,k/\epsilon}$, but then the sample space will be m^k times bigger.

For some applications the constructions described above suffice. For example, in the analysis of some of the randomized algorithms for graph problems presented in [17] (and in [1]), approximate pairwise independence of the random variables suffices. Thus, a construction of a sample space (of polynomial size) that is a pairwise independent approximation for general $\mathcal{P}_{n,m}$, can be used to convert these (polynomial-time) randomized algorithms into deterministic (polynomial-time) algorithms. In other applications (see [19]), approximations of identically and uniformly distributed boolean-valued random variables suffice. However, in the other applications the random variables are general and more than a constant amount of independence may be required in the analysis. Thus it is of primary importance to develop constructions for these cases.

1.3 New Results on Approximation

In this paper, we describe a construction of small sample spaces that are approximations of the independent distribution for any $\mathcal{P}_{n,m}$. The construction yields a sample space that is a (k, ϵ)-approximation, where the size of the sample space is polynomial in $\log(n)$, 2^k, and $1/\epsilon$. Previous results (cf., [1, 17]) that achieve the same kind of approximation yield a sample space of size polynomial in $\log(n)$ and $(k/\epsilon)^k$. In contrast to previous results, when $k = O(\log(n))$ and $\epsilon = \text{poly}(1/n)$, the size of the sample space in our construction is polynomial in n. This case is important to some applications (e.g., this construction improves the running time of some of the algorithms presented in [18]). Two natural examples where we obtain a significant improvement follow.

Example 1 *Suppose we wish to approximate n independently distributed 0-1 random variables, each assigned 1 with probability $\frac{1}{2} + 2\epsilon$ and 0 otherwise.* Using previously known techniques, one obtains a sample space of size polynomial in $\log(n)$ and $(k/\epsilon)^k$, which is a (k, ϵ)-approximation of the above. In contrast to previous results, our construction results in a sample space of size polynomial in $\log(n)$, $1/\epsilon$, and 2^k.

Example 2 *Suppose we wish to approximate n independently distributed random variables, where the i^{th} random variable is uniformly distributed over the set $\{1, 2, ..., m_i\}$ with the m_i's being arbitrary.* Using previously known techniques, one obtains a sample space of size polynomial in $(\max\{n, k/\epsilon\})^k$ that is a (k, ϵ)-approximation of the above. Again, our construction improves over the above by providing a sample space of size polynomial in $\log(n)$, $1/\epsilon$, and 2^k.

1.4 An overview of the main construction

Our main construction is described in this paragraph. For simplicity, we consider here the special case of approximating the joint distribution of n independent 0-1 random variables. Namely, each random variable X_i satisfies $X_i \in \{0, 1\}$. Let $p_i = \Pr(X_i = 0)$. To construct a (k, ϵ)-approximation of $X = X_1 \cdots X_n$, we use a $(\ell, \epsilon/2^{\ell+1})$-approximation of the uniform distribution over $\{0, 1\}^{n\ell}$, where $\ell \stackrel{\text{def}}{=} O(k + \log(1/\epsilon))$. We partition the latter $n\ell$ Boolean random variables into n consecutive blocks with ℓ random variables in each block. We interpret each block as the binary representation of an integer, and let B_i denote the integer represented by the i^{th} block. The approximation to X, denoted $Y = Y_1 \cdots Y_n$, is determined by letting $Y_i = 0$ if $B_i < p_i \cdot 2^\ell$, and $Y_i = 1$ otherwise.

If the B_i's are a $(k, \epsilon/2^{k\ell})$-approximation of the uniform distribution over $\{0, 1, ..., 2^\ell - 1\}^n$ then the Y_i's defined above would be a (k, ϵ)-approximation of the specification $\mathcal{P}_{n,2}$. This, however, requires the individual bits of all the B_i's to be a $(\ell \cdot k, \epsilon/2^{k\ell})$-approximation of the uniform distribution over $\{0, 1\}^{\ell n}$. However, we want to (and do) use a much weaker approximation; that is, a $(\ell, \epsilon/2^{\ell+1})$-approximation of the uniform distribution over $\{0, 1\}^{\ell n}$.

Our analysis uses the observation that, typically, each Y_i is determined by a few of the (most significant) bits of the corresponding B_i. Specifically, we show that only with small probability are the values of the Y_i's determined by more than $\ell = O(k + \log(1/\epsilon))$ bits in the representation of the B_i's. Thus, it suffices that the bit string, obtained by concatenating the binary representations of the B_i's, is an $(\ell, \epsilon/2^{\ell+1})$-approximation (of the uniform distribution over $\{0,1\}^{\ell n}$).

We end this overview by presenting an *alternative construction of unknown quality.* The problem of constructing (k, ϵ)-approximations to arbitrary product distributions, is reminiscent of the classic problem of generating arbitrary probability distributions by using a uniform probability distribution over binary strings (or in other words by using an unbiased coin). In particular, Knuth and Yao have extensively analyzed the expected number of coin tosses required in such schemes [13]. The input to such a scheme is a uniformly distributed binary sequence and the output is a sequence which approximates the desired distribution. A natural suggestion is to use one of these schemes to produce an a (k, ϵ)-approximation to the n-fold distribution, by feeding it as input a $(O(k), \epsilon')$-approximation to the uniform binary distribution, for some appropriate ϵ'. We do not know whether this alternative approach works; actually, we conjecture that, in general, it does not.

1.5 Omitted from this version

The conference version [9] of the current paper contains some material which is omitted here. This includes:

- Discussing the problem of constructing small sets with low discrepancy with respect to certain families of (axis-parallel) rectangles in high dimensional spaces. We comment that, in subsequent literature, the rectangles considered in [9] are referred to as *geometric rectangles*.

- Relating the problem of constructing approximations of product distributions to the problem of constructing small sets with low discrepancy (as above). In particular, the main construction presented in this write-up was presented in [9] using the terminology of low discrepancy sets.

- [9] contains two additional constructions of small sets with low discrepancy. The first such construction has been superseded by Chari et. al. [5] and by Armoni et. al. [3], whereas the second construction follows immediately from a theorem in [20]. Both constructions apply also to combinatorial rectangles (although the statement in [9] refers only to geometric rectangles).

In light of the above developments, we chose to omit all these results from the current write-up. In particular, we have omitted a result of [9] which has found further application in subsequent work (e.g., [5]). We refer to the fact that, *for any specification matrix $\mathcal{P}_{n,m}$, any k-wise independent approximation to $\mathcal{P}_{n,m}$ constitutes a $2^{-\Omega(k)}$-approximation to $\mathcal{P}_{n,m}$.* (A proof can be found in either [9, 10, 5].)

1.6 Subsequent Work

In the five years which have elapsed since the conference presentation of this work (cf., [9]), a few related works were done. We briefly describe the related results in [16, 5].

The work of Chari et. al. [5, Sec. 3] is most relevant to the current write-up. It presents constructions which match or improve over the sizes of all constructions presented in [9]. However, for some natural setting of the parameters, the size of the main construction of the current write-up is only matched by [5, Sec. 3]. Specifically, their construction has size polynomial in $\log(n)$, $1/\epsilon$, and $\min\{2^k, k^{\log(1/\epsilon)}\}$ (whereas our construction has size polynomial in $\log(n)$, $1/\epsilon$, and 2^k). Thus, [5, Sec. 3] yields no improvement when $\epsilon < 2^{-k}$, which is the typical case when one requires a bound on the approximation in the Norm-1 measure (rather than in Max-Norm as defined above).[3]

Linial et. al. [16] consider a one-sided version of the discrepancy problem. That is, rather than constructing sets which approximate the volume of all rectangles they construct sets which hit all sufficiently big rectangles. Their construction is polynomial in all relevant parameters (including the bound on the density of rectangles which must be hit).

The problem of constructing low discrepancy sets of polynomial size in all relevant parameters is still open. In particular, it is an open problem to construct sample spaces of size $\text{poly}(n, k, \epsilon^{-1})$ which (k, ϵ)-approximate any n-fold product distribution. We comment that if one drops the requirement that the sets be efficiently constructible, then sets (resp., spaces) of the desired sizes can be easily shown to exist using a random construction. This suggests the following (probably easier, definitely no harder) open problem: Using randomness to produce *certified* low discrepancy sets (i.e., a Las-Vegas rather than Monte-Carlo randomized construction).

2 Main Construction

Theorem 1 (general product approximator): *There is a deterministic algorithm, which on input a specification matrix $\mathcal{P}_{n,m} = \{p_{i,j} : i = 1, ..., n \ , \ j = 0, ..., m-1\}$, and parameters (k, ϵ), outputs a sample space of size $\text{poly}(2^k, \epsilon^{-1}, \log n)$ which constitutes a (k, ϵ)-approximation for $\mathcal{P}_{n,m}$. The algorithm works in time $\text{poly}(n, 2^k, \epsilon^{-1}, \log m)$.*

The above yields a $(k, m^k \cdot \epsilon)$-L1-approximation of $\mathcal{P}_{n,m}$.

We first present our construction for the special (yet interesting) case of approximating Boolean-valued random variables. We later generalize the construction to handle random variables ranging over arbitrary sets.

2.1 Special Case: Boolean-valued random variables

Assume we are given a Boolean specification matrix, $\mathcal{P}_{n,2} = \{p_{i,j} : i = 1, ..., n \ , \ j = 0, 1\}$ (i.e., $m = 2$). Clearly, it suffices to specify the probability that each of the n variables is to be assigned 0. Let $p_i \stackrel{\text{def}}{=} p_{i,0}$, for every $i \leq n$, and denote by $p_i(1), p_i(2), ...$ the bits in the binary expansion of p_i (i.e., $p_i = \sum_{j \geq 1} p_i(j) \cdot 2^{-j}$). We construct a (k, ϵ)-approximation of $\mathcal{P}_{n,2}$ as follows.

Let ℓ and t be integer parameters to be determined later (e.g., $\ell = t = 4(k + \log(2/\epsilon))$ will do). In our construction we use an arbitrary (efficiently constructible) $(t, (\epsilon/2^{t+1}))$-approximation of the uniform distribution over $\{0,1\}^{\ell n}$. Let us denote the 0-1 random variables in this approximation by $Z_1(1), ..., Z_1(\ell), Z_2(1), ..., Z_2(\ell), ..., Z_n(1), ..., Z_n(\ell)$. Intuitively, ℓ denotes the number of 0-1 random variables which may effect a single random variable in the result, and t denotes the total number of 0-1 variables which will be actually considered in the analysis.

Construction 1 *Let $Z_1(1), ..., Z_1(\ell), Z_2(1), ..., Z_2(\ell), ..., Z_n(1), ..., Z_n(\ell)$ be a $(t, \epsilon/2^{t+1})$-approximation of the uniform distribution over $\{0,1\}^{\ell n}$. For every i, if the string $Z_i(1) \cdots Z_i(\ell)$ is smaller (in lexicographic order) than the string $p_i(1) \cdots p_i(\ell)$ then set $Y_i = 0$, otherwise set $Y_i = 1$.*

[3] Recall that an (k, ϵ)-approximation (in Max-Norm) to, say, $\mathcal{U}_{n,2}$ yields variation distance (i.e., Norm-1 approximation) of at most $2^k \epsilon$ over windows of size k. Thus, to derive a meaningful result for Norm-1, one needs to have $\epsilon < 2^{-k}$.

In other words, the sample space S (over which the $Z_i(j)$'s are defined) also serves as the probability space for the Y_i's. Each sample point in S, denoted by $z_1(1), ..., z_1(\ell), ..., z_n(1), ..., z_n(\ell)$, is mapped to a point, $y_1, ..., y_n$, in the new sample space (where $y_i = 0$ iff $z_i(1)\cdots z_i(\ell) < p_i(1)\cdots p_i(\ell)$).

Our analysis of the above construction is somewhat analogous to the proof of Theorem 3 in [18]. We fix k variables $Y_{i_1}, ..., Y_{i_k}$ out of $Y_1, ..., Y_n$, and consider the quality of the approximation which they provide. Without loss of generality, we consider the random variables $Y_1, ..., Y_k$. Rather than bounding the Max-Norm performance of the approximation (as required in the theorem), we will actually bound the variation distance:

$$\frac{1}{2} \cdot \sum_{\sigma_1 \cdots \sigma_k} \left| \Pr(Y_1 \cdots Y_k = \sigma_1 \cdots \sigma_k) - \prod_{i=1}^{k} p_{i,\sigma_i} \right| \tag{3}$$

Consider the following cards game. The values of the random variables $Z_i(j)$ are written on cards that are placed face down. We turn the cards over, one by one, starting with the card holding $Z_1(1)$, and continue turning the cards over in the first block until the value of Y_1 is determined (which happens when the sequence of $Z_1(j)$'s deviates from the binary expansion of p_1). Then, we skip to the next block, and continue in the same fashion. Our goal is to prove that if $t = \ell = 4(k+\log_2(2/\epsilon))$, then the probability that we will turn over more than t cards is bounded by $\epsilon/2$. This will enable us to take advantage of variables $Z_i(j)$ that are a $(t, \epsilon/2^{t+1})$-approximation of $n\ell$ perfectly random bits. Our proof is divided into two stages: First, we assume that the bits $Z_i(j)$ are perfectly random, and we bound the probability of having to turn over more than t cards. Then, we add the error term due to the fact that the $Z_i(j)$'s are not perfectly random.

We model this cards game by considering an infinite random walk in a labeled infinite binary tree as follows.

1. Each node has two children, one reachable by an edge labeled 0 and the other reachable by an edge labeled 1. We associate to each path the binary string obtained by the edge labels along the string. Moreover, this binary string is interpreted as the binary representation of fraction $0.b_1b_2\ldots$, where the most significant bit appears closer to the root. The random walk starts at the root and at each node a random step is made so that each of the two children is reached with equal probability (of one-half). A step in the random walk corresponds to turning a card over in the cards game.

2. The nodes in the tree are labeled by pairs of the form (i, σ), where $i \in \{1, .., k\}$ and $\sigma \in \{0, 1, *\}$. The i component in a node-label signifies the block that we are now dealing with. The σ component signifies the status of the block as follows: $\sigma \in \{0, 1\}$ means that Y_i has been determined (i.e., Y_i set to σ), and $\sigma = *$ means that Y_i is not determined yet, and that we need to reveal more bits from the i'th block.

3. We now describe how the node labels are defined. We start by describing how the labels corresponding to Y_1 are "hung" from the root. The root is labeled $(1, *)$. Consider the infinite path, denoted by $path(root, p_1) = p_1(1), p_1(2), ...$, starting from the root, that corresponds to the binary representation of p_1. All the nodes along $path(root, p_1)$ are labeled $(1, *)$. All the nodes that are exactly an edge away from $path(root, p_1)$ are labeled either $(1, 0)$ or $(1, 1)$ according to the following rule: Consider the fractions corresponding to the labels along paths, and consider the order induced by this correspondence on paths that have the same starting point. A deviation from $path(root, p_1)$ that defines a path "smaller" than $path(root, p_1)$ ends with a node label $(1, 0)$, and a deviation that defines a path "greater" than $path(root, p_1)$

ends with a node label $(1,1)$. For example, suppose that $p_1(j)=0$, then the node reached by the path $p_1(1),...,p_1(j-1),1$ has label $(1,1)$. This completes the description of the node labels corresponding to Y_1.

4. We continue to label the tree with node labels corresponding to Y_2 by "hanging" them from subtrees rooted at nodes labeled $(1,0)$ or $(1,1)$. Given a node v labeled $(1,\sigma)$, where $\sigma \in \{0,1\}$, let $path(v,p_2) = p_2(1), p_2(2), ...$ denote the infinite path starting at v corresponding to the binary representation of p_2. All the nodes along $path(v,p_2)$ (except v) are labeled $(2,*)$. Deviations from $path(v,p_2)$ are labeled $(2,0)$ or $(2,1)$ according to whether they are "smaller" or "greater" than $path(v,p_2)$. We continue in this fashion until the nodes labels of $Y_1,\ldots,Y_k$ are given.

5. We define a node to be *complete* if it has the label $(k,0)$ or $(k,1)$. For sake of simplicity, we label by (k,σ) the children of any node labeled (k,σ) for $\sigma \in \{0,1\}$. Recall that reaching a complete node via a random walk results in setting values to all the (relevant) variables $Y_1,\ldots,Y_k$.

The following claims are easily verified.

Claim 2 *Consider a random infinite path going down the tree and set $\widetilde{Y}_i=\sigma_i$ if and only if the path goes through a node labeled (i,σ_i), where $\sigma_i \in \{0,1\}$. Then for every $\alpha=\sigma_1\cdots\sigma_k \in \{0,1\}^k$,*

$$\Pr(\widetilde{Y}_1\cdots\widetilde{Y}_k = \alpha) = \prod_{i=1}^{k} p_{i,\sigma_i}$$

Proof: The process by which the $\widetilde{Y}_i$'s are set is identical to independently and uniformly selecting (real numbers) r_i's in the interval $[0,1]$ and setting $\widetilde{Y}_i=0$ if $r_i<p_i$ (and $\widetilde{Y}_i=1$ if $r_i>p_i$). □

Claim 3 *The number of non-complete nodes at level t is $\sum_{i=0}^{k-1}\binom{t}{i} \ll 2^{\frac{3}{4}t+k}$.*

Proof: Consider an incomplete node v and the path p from the root to v. Let i denote the maximum index for which a label $(i,0)$ or $(i,1)$ exists along the path p. The path p is uniquely determined by the $i-1$ levels that contain nodes with labels $(i',0)$ or $(i',1)$ (where $i' \le i$) along p. The reason is that, between two such levels, the path p follows the edge labels equal to the binary representation of the fraction $p_{i'}$, where $(i'-1,\sigma)$ is the last non-* node label reached. Thus, the number of non-complete nodes at level t equals $\sum_{i=0}^{k-1}\binom{t}{i}$. Clearly, for $t \le 4k$ this expression is bounded by $2^t \le 2^{\frac{3}{4}t+k}$. For $t>4k$, the expression is bounded by $2^{H_2(k/t)\cdot t} < 2^{\frac{3}{4}t+k}$ (since $H_2(\alpha) < \frac{3}{4}+\alpha$ for all $\alpha<1$, where H_2 is the binary entropy function). □

Claim 4 *Consider a random path of length t going down the tree and set $\widetilde{Y}_i=\sigma_i$ if the path goes through a node labeled (i,σ_i), where $\sigma_i \in \{0,1\}$. In case the path does not go through any node labeled (i,σ) (with $\sigma \in \{0,1\}$), set $\widetilde{Y}_i$ arbitrarily. Then the variation distance between the distribution of $\widetilde{Y}_1\cdots\widetilde{Y}_k$ and the specification $\mathcal{P}_{n,2}$ is bounded by $2^{-(\frac{t}{4}-k)}$.*

Proof: By Claim 2, the variation distance is due to non-complete nodes in the t^{th} level. Using Claim 3, such nodes are reached with probability bounded by $2^{-\frac{t}{4}+k}$. □

The definition of the $\widetilde{Y}_i$'s in Claim 4 differs from the setting of the Y_i in Construction 1 only in the independence requirements. Specifically, the proof of Claim 4 assumes that each path in the tree is equally likely, whereas in Construction 1, the $Z_i(j)$ are a $(t, \epsilon/2^{t+1})$-approximation of the uniform distribution over $\{0,1\}^{\ell n}$. We show below that using the $Z_i(j)$'s to define the probabilities of taking the paths down the tree merely adds an error term bounded by $\epsilon/2$. Hence, we get

Proposition 5 *Let $\ell = t = 4(k + \log_2(2/\epsilon))$. Then Y_i's presented in Construction 1 constitute a (k,ϵ)-L1-approximation of $\mathcal{P}_{n,2}$.*

Proof: Our aim is to establish an upper bound on the variation distance of Expression (3) for the Y_i's presented in Construction 1. By Claim 4 and the setting of t, it follows that for the $\widetilde{Y}_i$'s, as defined in Claim 4, we have

$$\frac{1}{2} \cdot \sum_{\alpha} \left| \Pr(\widetilde{Y}_1 \cdots \widetilde{Y}_k = \alpha) - \prod_{i=1}^{k} p_{i,\sigma_i} \right| \; < \; 2^{-(\frac{t}{4} - k)} \; = \; \frac{\epsilon}{2} \tag{4}$$

However, we claim that the $\widetilde{Y}_i$'s are set exactly as they would have been set in the construction if the $Z_i(j)$'s were to be a t-wise independent approximation of the uniform distribution over $\{0,1\}^{\ell n}$. Intuitively, the $\widetilde{Y}_i$'s are defined by limiting the length of the random walk down the tree to t levels, and the same should hold even when the moves down the tree are governed by a non-disjoint sequence of random variables as long as the random variables along each path are independent and uniformly distributed in $\{0,1\}$.

Claim: Suppose that the $Z_i(j)$'s are a t-wise independent approximation of the uniform distribution over $\{0,1\}^{\ell n}$. Then Y_i's presented in Construction 1 constitute a $(k, \epsilon/2)$-L1-approximation of $\mathcal{P}_{n,2}$.
Proof: Recall that the values of the $Z_i(j)$'s determine the random walk down the tree. Specifically, the steps taken until node $(1,\sigma)$ (with $\sigma \in \{0,1\}$) is reached are determined by the random variables in the first block

$$Z_1(1), Z_1(2), .., Z_1(j_1),$$

where $j_1 \le t \le \ell$ is the number of such steps. The next steps, taken until node $(2,\sigma)$ is reached, are determined by the random variables $Z_2(1), Z_2(2), .., Z_2(j_2)$, where $j_2 \le t$ is the number of such steps, and so on. It is important to observe that, although different paths make us consider (or reveal) different $Z_i(j)$'s, the hypothesis that the $Z_i(j)$'s constitute a t-wise independent approximation (to the uniform distribution over $\{0,1\}^{\ell n}$) implies that each node in level t is reached with probability 2^{-t}.

This important observation follows from the fact that we reach a *specific* node v in level t if and only if we reveal t specific values on t specific $Z_i(j)$'s. The cards we reveal from the i'th block are determined by the subpath from the root to v the nodes of which are labeled (i,σ), and the values that we reveal in the i'th block are the edge labels along this subpath. Thus, as claimed, each node in level t is reached with probability 2^{-t}, and so $(Y_1, .., Y_k)$ is distributed identically to $(\widetilde{Y}_1, .., \widetilde{Y}_k)$. Using Eq. (4), the claim follows. $\square$

Turning to the actual $Z_i(j)$'s which are "only" a $(t, \epsilon/2^{t+1})$-approximations of the uniform distribution over $\{0,1\}^{\ell n}$, we conclude that if the walk on the tree is determined by the actual $Z_i(j)$'s, then, for every node v in level t, the probability of reaching v is in the interval $[2^{-t} - \epsilon/2^{t+1}, 2^{-t} + \epsilon/2^{t+1}]$. Thus, the Y_i's deviate from the $\widetilde{Y}_i$'s by at most $2^t \cdot 2^{-(t+1)} \cdot \epsilon$, and the proposition follows. $\blacksquare$

Comment 1: Proposition 5 holds also when setting $\ell = \log_2(4k/\epsilon)$ and $t = 4(k + \log_2(4/\epsilon)$. This can be accomplished by rounding the probabilities to $O(\log(k/\epsilon))$ bits of precision, and modifying the tree labeling so that paths labeled $(i,*)$ are finite. However, the gain is negligible since the value of ℓ has only a poly-log effect on the size of the sample space for the $Z_i(j)$'s (specifically, the size is exponential in k and logarithmic in ℓ, and so reducing ℓ from $O(k + \log(1/\epsilon))$ to $O(\log(k/\epsilon))$ has little impact).

2.2 The General Case

The construction for the general case extends Construction 1 in the obvious manner. That is, let $\mathcal{P}_{n,m} = \{p_{i,j} : 1 \le i \le n, 0 \le j \le m-1\}$ be a specification matrix. For $i = 1, ..., n$ and $j = 0, 1 ..., m$, let $q_{i,j} \stackrel{\rm def}{=} \sum_{v=0}^{j-1} p_{i,v}$ (and $q_{i,0} = 0$). Denote by $q_{i,j}(1), q_{i,j}(2), ...$ the bits in the binary expansion of $q_{i,j}$. Let ℓ and t be integers to be determined later (e.g., $\ell = t = 4(2k + \log(2/\epsilon))$ will do).

Construction 2 *Let $Z_1(1), ..., Z_1(\ell), Z_2(1), ..., Z_2(\ell), ..., Z_n(1), ..., Z_n(\ell)$ be a $(t, (\epsilon/2^{t+1}))$-approximation of the uniform distribution over $\{0,1\}^{\ell n}$. For every i, if the string $Z_i(1) \cdots Z_i(\ell)$ is (in lexicographic order) between the string $q_{i,j}(1) \cdots q_{i,j}(\ell)$ and the string $q_{i,j+1}(1) \cdots q_{i,j+1}(\ell)$ then set $Y_i = j$. In case $Z_i(1) \cdots Z_i(\ell) = q_{i,j}(1) \cdots q_{i,j}(\ell)$, set $Y_i = j$.*

Extending the argument used in the previous subsection we can easily evaluate the quality of approximation provided by Construction 2. Again, we consider without loss of generality, the variables $Y_1, ..., Y_k$. This time, we upper bound for each $\alpha = \sigma_1 \cdots \sigma_k \in \{0, 1, ..., m-1\}^k$, the absolute difference

$$\left| \Pr(Y_1 \cdots Y_k = \alpha) - \prod_{i=1}^{k} p_{i,\sigma_i} \right| \tag{5}$$

The cards game we play this time depends on $\sigma_1 \cdots \sigma_k$, and the purpose of the game is to decide whether $Y_i = \sigma_i$, for $i = 1, ..., k$. This means that we turn over cards from each block until we can decide whether $Y_i = \sigma_i$, or not. In case $Y_i \ne \sigma_i$, we do not care about the exact value of Y_i.

The labeling of the binary tree, that models the cards game, depends on $\sigma_1 \cdots \sigma_k$ and is described below. For each $i = 1, ..., k$, we are interested in the binary expansions of both q_{i,σ_i} and q_{i,σ_i+1}. Each node in the tree is labeled by a pair of the form (i, τ), where $i = 1, ..., k$ and $\tau \in \{+, -, *\}$. Intuitively, a node labeled $(i, +)$ corresponds to a setting $Y_i = \sigma_i$, a node labeled $(i, -)$ corresponds to a setting $Y_i \ne \sigma_i$, and a node labeled $(i, *)$ indicates that Y_i is yet to be set. (Indeed, in the binary case setting $Y_i \ne \sigma_i$ means $Y_i = \sigma_i \oplus 1$.)

The root is labeled $(1, *)$, and there are two infinite paths going down from the root with all nodes on it labeled $(1, *)$. These are the paths corresponding to the binary expansion of q_{1,σ_1} and q_{1,σ_1+1}, respectively. All the nodes reached by following such a path up to some node and then taking a single step away from the path are labeled $(1, \tau)$, where $\tau \in \{+, -\}$. Specifically, if $q_{1,\sigma_1}(j) = 0$ then the path going down from the root following the edge labeling $\langle q_{1,\sigma_1}(1), ..., q_{1,\sigma_1}(j-1), 1\rangle$ reaches a node labeled $(1, +)$. In case $q_{1,\sigma_1}(j) = 1$, the path going down from the root following the edge labeling $\langle q_{1,\sigma_1}(1), ..., q_{1,\sigma_1}(j-1), 0\rangle$ reaches a node labeled $(1, -)$. Similarly, if $q_{1,\sigma_1+1}(j) = 0$ (resp., $q_{1,\sigma_1+1}(j) = 1$) then the path going down from the root following the edge labeling $\langle q_{1,\sigma_1+1}(1), ..., q_{1,\sigma_1+1}(j-1), 1\rangle$ (resp., labeling $\langle q_{1,\sigma_1+1}(1), ..., q_{1,\sigma_1+1}(j-1), 0\rangle$) reaches a node labeled $(1, -)$ (resp., labeled $(1, +)$). Recall, when a random walk reaches a node labeled $(1, +)$ (resp., labeled $(1, -)$), the random variable Y_1 is set to σ_1 (resp., to $\sigma' \ne \sigma_1$).

From each node labeled (i, τ), with $i < k$ and $\tau \in \{+, -\}$, there are two infinite paths going down the tree with all nodes labeled $(i+1, *)$. These are the paths corresponding to the binary expansion of $q_{i+1,\sigma_{i+1}}$ and $q_{i+1,\sigma_{i+1}+1}$, respectively. The nodes reached from a node labeled (i,τ) by following the "$q_{i+1,\sigma_{i+1}}$-expansion path" (resp., "$q_{i+1,\sigma_{i+1}+1}$-expansion path") up to some node and then taking a single step away from the path are labeled $(i+1,\tau')$, where $\tau' \in \{+,-\}$. Again, if $q_{i+1,\sigma_{i+1}}(j) = 0$ (resp., $q_{i+1,\sigma_{i+1}}(j) = 1$) then the path going down from the root following the edge labeling $\langle q_{i+1,\sigma_{i+1}}(1), ..., q_{1,\sigma_{i+1}}(j-1), 1\rangle$ (resp., labeling $\langle q_{i+1,\sigma_{i+1}}(1), ..., q_{1,\sigma_{i+1}}(j-1), 0\rangle$) reaches a node labeled $(i+1,+)$ (resp., labeled $(i+1,-)$). Also, if $q_{i+1,\sigma_{i+1}+1}(j) = 0$ (resp., $q_{i+1,\sigma_{i+1}+1}(j) = 1$) then the path going down from the root following the edge labeling $\langle q_{i+1,\sigma_{i+1}+1}(1), ..., q_{1,\sigma_{i+1}+1}(j-1), 1\rangle$ (resp., labeling $\langle q_{i+1,\sigma_{i+1}+1}(1), ..., q_{1,\sigma_{i+1}+1}(j-1), 0\rangle$) reaches a node labeled $(i+1,-)$ (resp., labeled $(i+1,+)$). Reaching a node labeled $(i+1,+)$ (resp., labeled $(i+1,-)$), sets the random variable Y_{i+1} to σ_{i+1} (resp., to $\sigma_{i+1} \pm 1$). We define a node to be *complete* if it has the label $(k,+)$ or $(k,-)$. For sake of simplicity, we label by (k,τ) the children of any node labeled (k,τ) for $\tau \in \{+,-\}$. Intuitively, reaching a complete node via a random path from the root means that we can determine for every $i \in [1..k]$ whether $Y_i = \sigma_i$.

The following (analogous to the above) claims are easily verified

Claim 6 *Let $\alpha = \sigma_1 \cdots \sigma_k \in \{0,1\}^k$, and consider a random infinite path going down the tree in which nodes are labeled according to α* (as described above). *Suppose we set $\widetilde{Y}_i = \sigma_i$ if and only if the path goes through a node labeled $(i,+)$. Then*

$$\Pr(\widetilde{Y}_1 \cdots \widetilde{Y}_k = \alpha) = \prod_{i=1}^{k} p_{i,\sigma_i}$$

Claim 7 *The number of nodes at level t which are not complete is $\sum_{i=0}^{k-1} \binom{t}{i} \cdot 2^{i+1} \ll 2^{\frac{3}{4}t+2k}$.*

The extra 2^{i+1} factor (compared to Claim 3) is due to the fact that incomplete paths are not fully determined by the levels in which the path is not marked by $*$ (as before): From each vertex marked (j,τ), with $j \le i$ and $\tau \in \{+,-\}$, there are two paths (rather than one) marked by $(j+1,*)$, and so an incomplete path is determined by both the non-$*$ levels and the identity of one of the two corresponding possible paths.

Claim 8 *Let $\alpha = \sigma_1 \cdots \sigma_k \in \{0,1\}^k$, and consider a random path of length t going down the tree in which nodes are labeled according to α. Suppose we set $\widetilde{Y}_i = \sigma_i$ if the path goes through a node labeled $(i,+)$, and set $\widetilde{Y}_i \neq \sigma_i$ if the path goes through a node labeled $(i,-)$. In case the path does not go through any node labeled (i,τ), with $\tau \in \{+,-\}$, we set $\widetilde{Y}_i$ arbitrarily. Then*

$$\left| \Pr(\widetilde{Y}_1 \cdots \widetilde{Y}_k = \alpha) - \prod_{i=1}^{k} p_{i,\sigma_i} \right| < 2^{-(\frac{t}{4}-2k)}.$$

The definition of the $\widetilde{Y}_i$'s in Claim 8 corresponds to the setting of the Y_i's in Construction 2, provided that the $Z_i(j)$'s constitute a t-wise independent approximation of the uniform distribution over $\{0,1\}^{\ell n}$, and that $\ell \ge t$. As in the proof of Proposition 5, setting $\ell = t$ and allowing the $Z_i(j)$'s to constitute a $(t, (\epsilon/2^{t+1}))$-approximation of the uniform distribution over $\{0,1\}^{\ell n}$, merely adds an error term bounded by $\epsilon/2$. Hence, we get

Proposition 9 *Let $\ell = t = 4 \cdot (2k + \log_2(2/\epsilon))$. Then Y_i's presented in Construction 2 constitute a (k,ϵ)-approximation of $\mathcal{P}_{n,m}$.*

The reason that we are able to prove (k,ϵ)-L1-approximation in the binary case and only (k,ϵ)-approximation in the general case is due to the fact that the tree labeling in the general case depends on the values $\sigma_1 \cdots \sigma_k$, whereas in the binary case the tree labeling depends only on the probed indices (which for simplicity were chosen to be $1,\ldots,k$).

Proof of Theorem 1: Using the known results on (t, ϵ')-approximation of the uniform distribution over $\{0,1\}^{\ell n}$, Theorem 1 follows. Specifically, we need a $(t, 2^{-(t+1)} \cdot \epsilon)$-approximation of the uniform distribution over $\{0,1\}^{\ell n}$, where $\ell = t = 4 \cdot (2k + \log_2(2/\epsilon))$. By results of [2], such approximations can be efficiently constructed having sample space of size

$$\begin{aligned} \left(\frac{t \cdot \log_2(\ell n)}{2^{-t-1}\epsilon}\right)^2 &= \frac{2^{2t+2} \cdot (t \log_2(\ell n))^2}{\epsilon^2} \\ &= \widetilde{O}\left(\frac{2^{16k} \cdot (\log n)^2}{\epsilon^{10}}\right) \end{aligned}$$

where $\widetilde{O}(x) = \text{poly}(\log x) \cdot x$. The theorem follows. ■

Comment 2: The complexity of our construction is determined by the complexity of the (t, ϵ')-approximation to the uniform distribution over $\{0,1\}^{\ell n}$, for $\epsilon' = \epsilon/2^{t+1}$. The overhead added by our construction is merely performing the easily computed mapping of ℓn-bit long sequences (i.e., points in the latter sample space) to n-sequences over $\{0, \ldots, m-1\}$ (i.e., points in the former sample space). Recall that this mapping amounts to comparisons of ℓ-bit strings.

Comment 3: Construction 2 corresponds to a small set of low discrepancy with respect to (axis-parallel) (geometric) rectangles with at most k non-degenerate coordinates. This claim follows from the fact that the underlying sample space (i.e., the random variables $Z_1(1), ..., Z_n(\ell)$) do not depend on the specification matrix $\mathcal{P}_{n,m}$. Thus, for any specification matrix $\mathcal{P}_{n,m}$, the approximation $Y_1, ..., Y_n$ is determined by the same $Z_i(j)$'s (using the actual specification matrix $\mathcal{P}_{n,m}$). Furthermore, it is important that each Y_i is determined only by the random variables $Z_i(j)$'s, $j = 1, ..., \ell$, and that the setting of Y_i (under any specification matrix) corresponds to settings of $Z_i(1), ..., Z_i(\ell)$ which correspond to intervals in $[0,1]$. For details see [9]. Actually, to derive such low discrepancy sets, it suffices to use our construction while setting $m = 3$.

3 Alternative construction for a special case

In the special case where all the entries in the specification matrix, $\mathcal{P}_{n,m}$, are rationals of the form $\frac{q}{p}$, for some small prime p or prime power (e.g., $p = 3$), better (k, ϵ)-approximation schemes can be constructed. In this case, a (k, ϵ)-approximation of $\mathcal{P}_{n,m}$ is constructed (in the obvious manner) using a (k, ϵ)-approximation of the uniform distribution over $GF(p)^n$. Recall that (k, ϵ)-approximation of the uniform distribution over $GF(p)^n$ can be constructed using support of the same cardinality as in the construction of such approximations for the uniform binary distribution [2, 4, 8] (cf., [10, Chap. 3]).[4] We get

Theorem 10 *Let p be a prime and $\mathcal{M}^p_{n,m}$ denote the set of all n-by-m specification matrices in which every entry is an integer multiple of $\frac{1}{p}$. Then, there is a deterministic algorithm, which on input a prime p, a specification matrix $\mathcal{P}_{n,m} = \{p_{i,j} : i = 1, ..., n \ , \ j = 0, ..., m-1\} \in \mathcal{M}^p_{n,m}$, and*

[4] We stress that when constructing a small-bias probability space over $GF(p)^n$, the size of the space does not depend on p, provided small-bias is defined in terms of an upper bound on the Fourier coefficients. The transformation from max-norm in the Fourier basis to max-norm in the (standard) pointwise basis preserves this upper bound. (Note that the Fourier basis is not normal, and while normalizing you gain a $\sqrt{p^k}$ factor, lost in the cosequetive basis transformation.) For details see [10, Chap. 3].

parameters (k,ϵ), outputs a sample space of size $(k \cdot \epsilon^{-1} \cdot \log n)^2$ which constitutes a $(k, f_p(\mathcal{P}_{n,m}) \cdot \epsilon)$-approximation for $\mathcal{P}_{n,m}$, where

$$f_p(\mathcal{P}_{n,m}) \stackrel{\rm def}{=} \max_{I:|I|\le k} \left\{ \prod_{i\in I} \left(p \cdot \max_{j=0,\ldots,m-1} \{p_{i,j}\} \right) \right\}$$

Proof: We use an efficient construction of a (k,ϵ)-approximation of the uniform distribution over $GF(p)^n$. Such a construction has size $(k \cdot \epsilon^{-1} \cdot \log n)^2$. We transform each sample point $(s_1, ..., s_n)$ in the above space to a sample point $(r_1, ..., r_n)$ in our space by setting $r_i = j$ if $q_{i,j} < s_i \le q_{i,j+1}$, where $q_{i,j} \stackrel{\rm def}{=} \sum_{v=0}^{j-1} p_{i,v}$. Clearly, the value sequence $\langle v_1, ..., v_\ell \rangle \in \{0, \ldots, m-1\}^\ell$, $\ell \le k$, (as an assignment to a specific sequence $\langle i_1, ..., i_\ell \rangle$ of coordinates) deviates from the specification by at most $(\prod_{j=1}^{\ell} (p \cdot p_{i_j,v_j})) \cdot \epsilon$, and the theorem follows. ■

The alternative construction can be extended to the case all entries in the specification matrix $\mathcal{P}_{n,m}$ are well approximated by rationals of the form $\frac{q}{p}$, for some small prime p. By "good approximation" we mean that each entry is approximated up to additive error $\frac{\epsilon}{2k}$, where ϵ is the approximation parameter desired for the final construction. Hence, this approach is applicable only if the specification matrix has good approximation by a rational matrix with relatively small common denominator.

4 Acknowledgments

We thank Josef Beck, Nati Linial, Emo Welzl and Avi Wigderson for helpful discussions. We are also grateful to two annoymous referees for their useful comments.

References

[1] Alon, N., Babai, L., Itai, A., "A Fast and Simple Randomized Parallel Algorithm for the Maximal Independent Set Problem", *Journal of Algorithms*, 7, pp. 567–583, 1986.

[2] Alon, N., Goldreich, O., Håstad, J., Peralta, R., "Simple Constructions of Almost k-wise Independent Random Variables", *Journal of Random structures and Algorithms*, Vol. 3, No. 3, (1992), pp. 289–304.

[3] R. Armoni, M. Saks, A. Wigderson, and S. Zhou, "Discrepancy Sets and Pseudorandom Generators for Combinatorial Rectangles", *Proc. 37th FOCS*, pages 412–421, 1996.

[4] Y. Azar, R. Motwani, and J. Naor, "Approximating arbitrary probability distributions using small sample spaces", to appear in *Combinatorica.*

[5] S. Chari, P. Rohatgi, and A. Srinivasan, "Improved Algorithms via Approximation of Probability Distributions", in *Proceedings of the 26th ACM Symp. on Theory of Computing (STOC)*, pages 584–592, 1994. A revised version has appeared as TR97-01 of DIMACS, 1997.

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

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

[8] Even, G., "Construction of Small Probabilistic Spaces for Deterministic Simulation", M. Sc. (in Computer Science) Thesis, submitted to the Senate of the Technion – Israel Institute of Technology, Aug. 1991. (In Hebrew, abstract in English).

[9] G. Even, O. Goldreich, M. Luby, N. Nisan, and B. Veličković, "Approximations of General Independent Distributions", in *Proceedings of the 24th ACM Symp. on Theory of Computing (STOC)*, pages 10–16, 1992.

[10] G. Even, O. Goldreich, M. Luby, N. Nisan, and B. Veličković, "Approximations of General Independent Distributions", see `http://theory.lcs.mit.edu/~oded/eglnv.html`, Jan. 1997.

[11] D.R. Karger and D. Koller, "(De)randomized Construction of Small Sample Spaces in NC", In *35th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 252-263, 1994.

[12] Karp, R., Wigderson, A., "A Fast Parallel Algorithm for the Maximal Independent Set Problem", *Journal of the ACM*, Vol. 32, No. 4, pp. 762–773, October 1985.

[13] Knuth, D., Yao, A., "The complexity of non uniform random number generation", in Algorithms and Complexity, Ed. J. Traub, AC Press, New York, pp. 357-428, 1976.

[14] D. Koller and N. Megiddo, "Constructing Small Sample Spaces Satisfying Given Constraints", In *Proceedings of the 25th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 268-277, 1993.

[15] Linial, N., Nisan, N., "Approximate Inclusion-Exclusion", *Combinatorica* 10 (4), 1990, pp. 349-365.

[16] Linial, N., Luby, M., M. Saks, and D. Zuckerman, "Efficient Construction of a Small Hitting Set for Combinatorial Rectangles in High Dimension", *Combinatorica* (to appear). Preliminary version in *25th STOC*, 1993.

[17] Luby, M., "A Simple Parallel Algorithm for the Maximal Independent Set Problem," *SIAM J. on Computing*, November 1986, Vol. 15, No. 4, pp. 1036-1053.

[18] Luby, M., Veličković, B., "On Deterministic Approximation of DNF", *Algorithmica* (special issue devoted to randomized algorithms), Vol. 16, No. 4/5, October/November 1996, pp. 415–433.

[19] Naor, J., Naor, M., "Small-bias Probability Spaces: Efficient Constructions and Applications", *SIAM J. on Computing*, Vol 22, 1993, pp. 838–856.

[20] Nisan, N., "Pseudo-random Generators for Space-Bounded Computation", *Combinatorica*, Vol. 12, No. 4, pages 449–461, 1992.