# Approximations of General Independent Distributions

Guy Even* Oded Goldreich† Michael Luby‡ Noam Nisan§ Boban Veličković¶

## Abstract

We describe efficient constructions of small probability spaces that approximate the independent distribution for general random variables. Previous work on efficient constructions concentrate on approximations of the independent distribution for the special case of uniform boolean-valued random variables. Our results yield efficient constructions of small sets with low discrepancy in high dimensional space and have applications to derandomizing randomized algorithms.

## 1 Introduction

The problem of constructing small sample spaces that "approximate" the independent distribution on $n$ random variables has received considerable attention recently (cf. [6, Chor Goldreich] [8, Karp Wigderson], [11, Luby], [1, Alon Babai Itai], [13, Naor Naor], [2, Alon Goldreich Håstad Peralta], [3, Azar Motwani Naor]). The primary motivation for this line of research is that random variables that are "approximately" independent suffices 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. The culmination of previous works are constructions of small sample spaces that approximate a constant amount of independence for general random variables (for a brief survey, see for example [11, Luby] or [1, Alon Babai Itai]) or that approximate complete independence for identically and uniformly distributed boolean-valued random variables [13, Naor Naor], [2, Alon Goldreich Håstad Peralta], [3, Azar Motwani Naor]). Although previous results are sufficient for some applications to derandomizing algorithms, in many applications what is needed is a small sample space that approximates more than a constant amount of independence for general random variables. In this paper we present constructions of small efficiently constructible sample spaces for this more general case.

### 1.1 Approximation definitions

The probability 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 is equal to 1. The $(i,v)$-entry $p_{i,v}$ specifies the probability that the $i^{th}$ random variable should take on value $v$. For all values of $l \in \{1,\ldots,n\}$, for all $I =< i_1,\ldots,i_l >$, where $1 \leq i_1 < \ldots < i_l \leq n$, and for all $V =< v_1,\ldots,v_l > \in \{0,\ldots,m-1\}^l$, let $p_{I,V} = \prod_{j=1}^{l} p_{i_j,v_j}$ be the probability that the subsequence of random variables indexed by $I$ should take on value $V$ if the random variables were truly independent.

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 the independent distribution for $\mathcal{P}_{n,m}$. All constructions for $S$ given in this paper are efficient in the sense that there is a deterministic algorithm which produces $S$ in time polynomial in the length of the description of $\mathcal{P}_{n,m}$ and in the length of the description of $S$. The description of each point $s \in S$ consists of, for each index $i$, a value $x_i(s) \in \{0,\ldots,m-1\}$ for the $i^{th}$ random variable. We view $S$ as a sample space that induces a distribution on $x_1,\ldots,x_n$ defined by choosing a point randomly and uniformly from $S$. We let $x_I$ be the subsequence of random variables indexed by $I$, we let $x_I = V$ denote the event that the subsequence $x_I$ takes on the value $V$, and we let $\Pr_S[x_I = V]$ be the probability that event $x_I = V$ occurs in the distribution induced by $S$.

*Computer Science Department, Technion, Haifa, Israel.

†Computer Science Department, Technion, Haifa, Israel, research partially supported by grant No. 89-00312 from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel.

‡International Computer Science Institute, 1947 Center Street, Berkeley, California 94704, research partially supported by NSF operating grant CCR-9016468 and by grant No. 89-00312 from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel.

§Department of Computer Science, Hebrew University, Jerusalem, Israel, Supported by the Wolfson research awards administered by the Israeli Academy of Sciences and Humanities and by USA-Israel BSF 89-00126

¶Department of Mathematics, U.C. Berkeley, research partially supported by NSF

Given a probability matrix $\mathcal{P}_{n,m}$, $S$ induces a distribution on $x_1, \ldots, x_n$ that is

- *independent* if for all $l$, $I =< i_1, \ldots, i_l >$ and $V =< v_1, \ldots, v_l >$, $\Pr_S[x_I = V] = p_{I,V}$.
- a *$k$-wise independent approximation* if for all $l \leq k$, $I =< i_1, \ldots, i_l >$ and $V =< v_1, \ldots, v_l >$, $\Pr_S[x_I = V] = p_{I,V}$.
- an *$\epsilon$-approximation* if for all $l$, $I =< i_1, \ldots, i_l >$ and $V =< v_1, \ldots, v_l >$, $|\Pr_S[x_I = V] - p_{I,V}| \leq \epsilon$.
- a *$(k, \epsilon)$-approximation* if for all $l \leq k$, $I =< i_1, \ldots, i_l >$ and $V =< v_1, \ldots, v_l >$, $|\Pr_S[x_I = V] - p_{I,V}| \leq \epsilon$.

## 1.2 Previous work

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 $1/2$ that describes $n$ identically and uniformly distributed boolean-valued random variables. It is fairly easy to prove that $S$ has to be of size at least $2^n$ to be independent even for $\mathcal{U}_{n,2}$. Constructions of sample spaces that are $k$-wise independent approximations for $\mathcal{U}_{n,m}$ of size $\max\{n, m\}^k$, and that are $(k, \epsilon)$-approximations for general $\mathcal{P}_{n,m}$ with size $(\max\{n, k/\epsilon\})^k$, are implicit in many works; a brief survey of some of these constructions can be found in either [11, Luby] or [1, Alon Babai Itai]. For constant $k$ and $1/\epsilon$ polynomial in $n$, this yields a sample space of size polynomial in $n$.

It has been recognized that in many other examples, what is needed is a sample space that has more than a constant amount of independence between the $n$ random variables; typically logarithmic in $n$ independence suffices. On the other hand, it has been shown that the sample space has to be of size at least $n^{k/2}$ in order to be a $k$-wise independent approximation for $\mathcal{U}_{n,2}$ [5, Chor Freidmann Goldreich Håstad Rudich Smolensky], and for non-constant $k$ this is not polynomial in $n$. [13, Naor Naor] introduced the idea of allowing the error parameter $\epsilon$ and gave an ingenious construction of a sample space that is an $\epsilon$-approximation for $\mathcal{U}_{n,2}$ where the size of the sample space is $O(n \log(n)/\epsilon^4)$. Simpler constructions with a sample space of size $O((n \log(n))^2/\epsilon^2)$ for $\mathcal{U}_{n,2}$ were subsequently presented in [2, Alon Goldreich Håstad Peralta]. These constructions can be extended to $\mathcal{U}_{n,m}$, basically using the same ideas, but in a slightly more complicated way (cf. [2, Alon Goldreich Håstad Peralta], [3, Azar Motwani Naor], [7, Even]), where the size of the resulting sample space is $O((n \log(n))^2/\epsilon^2)$.

## 1.3 New results

For some applications the constructions described in the previous subsection are quite useful. For example, in the analysis of some of the randomized algorithms for graph problems presented in [11, Luby] and [1, Alon Babai Itai], approximate pairwise independence of the random variables suffices. Thus, the 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 randomized algorithms into deterministic algorithms. In other applications (see [13, Naor Naor]), approximations of identically and uniformly distributed boolean-valued random variables suffice. However, in the more typical application the random variables are general and more than a constant amount of independence is required in the analysis, and thus it is of primary importance to develop constructions for these cases.

In this paper, we describe three constructions of small sample spaces that are approximations of the independent distribution for general $\mathcal{P}_{n,m}$; the first two constructions are new and the third is a construction based on a theorem in [14, Nisan]. The first construction yields a sample space that is a $(k, \epsilon)$-approximation, where the size of the sample space is polynomial in $\log(n)$, $2^k$ and $1/\epsilon$. Previous results that achieve the same kind of approximation result in a sample space of size polynomial in $\log(n)$ and $(k/\epsilon)^k$. In contrast to previous results, when $k = O(\log(n))$ and $1/\epsilon$ is polynomial in $n$ the size of the sample space in our construction is polynomial in $n$. This case is important to some applications, and in particular this construction improves the running time of some of the algorithms presented in [12, Luby Veličković].

The second and third constructions for $n$ general random variables yield samples spaces that are $\epsilon$-approximations for general $\mathcal{P}_{n,m}$, where the size of the sample space is polynomial in $(n/\epsilon)^{\log(1/\epsilon)}$ for the second construction and polynomial in $(n/\epsilon)^{\log(n)}$ for the third. In contrast, the previous bound on the sample space size, implicit in the classical work on discrepancy theory (see e.g. [4, Beck Chen] or [15, Niederreiter]), is polynomial in $n^n/\epsilon$. For interesting cases of $n$ and $\epsilon$, i.e. when $1/\epsilon$ is polynomial in $n$, the results presented here are dramatic improvements.

## 1.4 Discrepancies

Let $[0,1)^n$ be the $n$ dimensional unit cube, let $\mathcal{R}_n$ be the set of all axis parallel rectangles within $[0,1)^n$, and for each $R \in \mathcal{R}_n$, let $\mathrm{vol}(R)$ be the volume of $R$. For any finite set of points $S$ in $[0,1)^n$ and for any $R \in \mathcal{R}_n$, define the *discrepancy of $S$ on $R$* as $\mathrm{disc}_S(R) = |\mathrm{vol}(R) - |S \cap R|/|S||$. This quantity is the

absolute value of the difference between the probability that a randomly chosen point from $[0,1)^n$ falls in $R$ and the probability that a randomly chosen point from $S$ falls in $R$. For any $\mathcal{K}_n \subseteq \mathcal{R}_n$, the *discrepancy of $S$ on $\mathcal{K}_n$* is defined as $\Delta_S(\mathcal{K}_n) = \max_{R \in \mathcal{K}_n} \mathrm{disc}_S(R)$. Finding explicit constructions of sets with small discrepancy have a variety of applications, including applications to numerical integration. The discrepancy problem can be stated as follows: given $n$ and $\epsilon$, construct a set $S$ in $[0,1)^n$ with $\Delta_S(\mathcal{R}_n) \leq \epsilon$.

As we describe (and as also has been describe before by others, e.g. [15, Niederreiter]), there is an close connection between discrepancy and approximating independent distributions of $n$ general random variables. Our primary interest in sets with small discrepancy is that they are *universal* for the problem of constructing sample spaces that are good approximations for general distributions. For example, a set $S$ in $[0,1)^n$ for which $\Delta_S(\mathcal{R}_n) \leq \epsilon$ can be viewed as *universal* in the following sense: There is a simple efficient algorithm that given $S$ and *any* $\mathcal{P}_{n,m}$ computes a sample space that is an $\epsilon$-approximation for $\mathcal{P}_{n,m}$. To be of interest in the problem of approximating random variables, it is crucial that the size of $S$ be small in terms of both parameters $\epsilon$ and $n$.

Classical work on the discrepancy problem concentrates on minimizing the size of $S$ primarily as a function of $1/\epsilon$ and then secondarily as a function of $n$ [4, Beck Chen], [15, Niederreiter], i.e. the dimension $n$ is thought of as arbitrary but fixed and the goal is to find a set $S$ with size as small as possible as a function of $1/\epsilon$. Although classical work shows that there are explicit constructions of $S$ with size smaller than that implied by a random construction for fixed $n$, the bounds are exponential in $n$ and say nothing non-trivial for values of $n$ and $\epsilon$ interesting for the case of approximating general distributions. i.e. when $n$ and $1/\epsilon$ are comparable.

The constructions presented here give new results for the discrepancy problem. For any constant $\beta < 1$ let $\mathcal{R}_n^{[0,\beta)} \subset \mathcal{R}_n$ be the set of rectangles $R$ such that in each dimension $i$ the length of $R$ is either 1 or else it is in the range $[0,\beta)$. The first construction yields, for any constant $\beta < 1$, a set $S$ in $[0,1)^n$ of size polynomial in both $n$ and $1/\epsilon$ such that $\Delta_S(\mathcal{R}_n^{[0,\beta)}) \leq \epsilon$. The second construction yields a set $S$ in $[0,1)^n$ of size polynomial in $(n/\epsilon)^{\log(1/\epsilon)}$ such that $\Delta_S(\mathcal{R}_n) \leq \epsilon$, and the third construction yields a set $S$ with the same properties of size $(n/\epsilon)^{\log(n)}$. In contrast to these new results, the previous known bounds from classical discrepancy theory on the size of an explicitly constructible set $S$ in $[0,1)^n$ with small discrepancy are exponential in $n$ [4, Beck Chen], [15, Niederreiter].

It is easy to see that a random set of points $S$ in $[0,1)^n$ of size $cn\log(n/\epsilon)/\epsilon^2$ for some constant $c > 1$ has the property that $\Delta_S(\mathcal{R}_n) \leq \epsilon$ with high probability. The crucial property missing from this proof of existence is efficient constructibility. We leave this as an open question, i.e. the problem of finding an explicit construction of a set $S$ in $[0,1)^n$ with $\Delta_S(\mathcal{R}_n) \leq \epsilon$ and with $|S|$ polynomial in both $n$ and $1/\epsilon$. As stated above, a solution to this problem would yield a universal set $S$ of size polynomial in both $n$ and $1/\epsilon$ that for all $\mathcal{P}_{n,m}$ can be interpreted as a sample space that is an $\epsilon$-approximation for $\mathcal{P}_{n,m}$.

# 2 Linking discrepancy and approximation

In this section we provide the (straightforward) link between sets $S$ with small discrepancy and sample spaces that approximate the independent distribution on $n$ random variables.

DEFINITION 2.1 (classes of rectangles) *The $n$ dimensional unit cube is $[0,1)^n$. Let $R = \prod_{i \in \{1,\ldots,n\}} r_i$ be an axis-parallel rectangle in $[0,1)^n$, where each $r_i = [a_i, b_i)$ is a subinterval of $[0,1)$. We say $R$ is* trivial *in dimension $i$ if $r_i = [0,1)$. Without loss of generality, we restrict attention to those rectangles for which there is no $i \in \{1,\ldots,n\}$ with $a_i = b_i$. The volume of $R$ is* $\mathrm{vol}(R) = \prod_{i \in \{1,\ldots,n\}} b_i - a_i$.

- *Define $\mathcal{R}_n$ to be the set of all axis parallel rectangles in $[0,1)^n$.*
- *For constant $\beta < 1$, define $\mathcal{R}_n^{[0,\beta)}$ to be the subset of $\mathcal{R}_n$ consisting of all rectangles $R = \prod_{i \in \{1,\ldots,n\}} r_i$ such that for each $i$ either $R$ is trivial in dimension $i$ or else $r_i = [a_i, b_i)$ with $b_i - a_i \leq \beta$.*
- *For positive integer $k$, define $\mathcal{R}_n^k$ to be the subset of $\mathcal{R}_n$ consisting of all rectangles $R$ such that $R$ is trivial for all but at most $k$ dimensions.*
- *For any $\mathcal{K}_n \subseteq \mathcal{R}_n$ and any positive integer $m \geq 2$, define $\mathcal{K}_{n,m}$ as the set of rectangles $R = \prod_{i \in \{1,\ldots,n\}} r_i \in \mathcal{K}_n$ such that each $r_i$ is of the form $[a_i/m, b_i/m)$ for integers $a_i$ and $b_i$ satisfying $0 \leq a_i < b_i \leq m$. For example, $\mathcal{R}_{n,2}$ is the subset of $\mathcal{R}_n$ consisting of all rectangles $R = \prod_{i \in \{1,\ldots,n\}} r_i$ such that for each $i \in \{1,\ldots,n\}$, $r_i = [0,1/2)$ or $r_i = [1/2,1)$ or $r_i = [0,1)$.*

DEFINITION 2.2 (projection sample space) *Let $S$ be a finite subset of points from $[0,1)^n$ and let $\mathcal{P}_{n,m}$ be a probability matrix. $S$ can be viewed as the* projection sample space for $\mathcal{P}_{n,m}$, *inducing a distribution on random variables $x_1, \ldots, x_n$ as follows. For all $i \in \{1,\ldots,n\}$*

*let interval* $r_{i,0} = [0, p_{i,0})$ *and for all* $v \in \{1, \ldots, m\}$ *let interval* $r_{i,v} = [a_{i,v}, b_{i,v})$, *where* $a_{i,v} = \sum_{0 \le w < v} p_{i,w}$ *and* $b_{i,v} = a_{i,v} + p_{i,v}$. *Random variable* $x_i$ *at a point* $s =< s_1, \ldots, s_n > \in S$ *takes on the unique value* $v$ *that satisfies* $s_i \in r_{i,v}$.

A set $S \subset [0,1)^n$ as just described is *universal* in the sense that it can be interpreted in a straightforward way as a sample space for *any* $\mathcal{P}_{n,m}$. The interpretation has the property that it is *coordinate independent* in the sense that the value given to $x_i$ at sample point $s \in S$ depends only on the $i^{th}$ coordinate of $s$ and on the $i^{th}$ row of $\mathcal{P}_{n,m}$.

The crucial links between discrepancies and approximations are the following observations.

1. If $\Delta_S(\mathcal{R}_n) \le \epsilon$ then for any $\mathcal{P}_{n,m}$ the projection sample space of $S$ for $\mathcal{P}_{n,m}$ is an $\epsilon$-approximation.
2. If $\Delta_S(\mathcal{R}_n^k) \le \epsilon$ then for any $\mathcal{P}_{n,m}$ the projection sample space of $S$ for $\mathcal{P}_{n,m}$ is a $(k, \epsilon)$-approximation.
3. If $\Delta_S(\mathcal{R}_{n,2}) \le \epsilon$ then the projection sample space of $S$ for $\mathcal{U}_{n,2}$ is an $\epsilon$-approximation.

From this discussion it is clear that in order to produce sample spaces which approximate the independent distribution for $n$ general random variables it suffices to produce small finite sets $S \subset [0,1)^n$ with small discrepancy.

DEFINITION 2.3 (the natural mapping to $[0,1)^n$) *We can view a sample space* $S$ *that induces a distribution on* $x_1, \ldots, x_n$ *for* $\mathcal{U}_{n,m}$ *in a natural way as a finite set of points in* $[0,1)^n$, *where the* $i^{th}$ *coordinate of* $s \in S$ *is* $x_i(s)/m$.

From this the converse of observation 3 follows, i.e. it is not hard to verify that if $S$ is an $\epsilon$-approximation for $\mathcal{U}_{n,2}$ then $\Delta_S(\mathcal{R}_{n,2}) \le \epsilon$ when sample points in $S$ are mapped to $[0,1)^n$ in the natural way. The converses of observations 1 and 2 are not so obvious. For example, it is true that if $S$ is a sample space that is a $(k, \epsilon)$-approximation for $\mathcal{U}_{n,m}$ then $\Delta_S(\mathcal{R}_{n,m}^k) \le \epsilon m^k$ when points in $S$ are mapped to $[0,1)^n$ in the natural way, but this is too weak of a bound for most purposes.

Some further useful observations are:

4. $\Delta_S(\mathcal{R}_{n,4n/\epsilon}) \le \epsilon/2$ implies that $\Delta_S(\mathcal{R}_n) \le \epsilon$. This is because for any rectangle $R \in \mathcal{R}_n$ there are rectangles $R^-, R^+ \in \mathcal{R}_{n,4n/\epsilon}$ such that $R^- \subseteq R \subseteq R^+$ and such that $\mathrm{vol}(R^+) - \mathrm{vol}(R^-) \le \epsilon/2$.
5. By similar reasoning to that used in observation 4, $\Delta_S(\mathcal{R}_{n,4k/\epsilon}^k) \le \epsilon/2$ implies that $\Delta_S(\mathcal{R}_n^k) \le \epsilon$.
6. $\Delta_S(\mathcal{R}_n^k) \le \epsilon$ implies that $\Delta_S(\mathcal{R}_n^{[0,\beta)}) \le \epsilon + \beta^k$. This follows because $\mathrm{vol}(R) \le \beta^k$ for any rectangle $R \in \mathcal{R}_n^{[0,\beta)}$ which is non-trivial in more than $k$ dimensions. This shows for constant $\beta$ and for $k = O(\log(1/\epsilon))$ that $\Delta_S(\mathcal{R}_n^k) \le \epsilon/2$ implies that $\Delta_S(\mathcal{R}_n^{[0,\beta)}) \le \epsilon$.

# 3 The constructions

All of the results are stated in terms of constructions of sets with small discrepancy. The constructions of small sample spaces that approximate the independent distribution follow from the observations of the previous section.

## 3.1 From boolean to general

**Theorem 1** *There is an explicitly constructible finite set* $S \subset [0,1)^n$ *with* $\Delta_S(\mathcal{R}_n^k) \le \epsilon$ *such that* $|S|$ *is polynomial in* $\log(n)$, $2^k$ *and* $1/\epsilon$.

PROOF:

Let $x_1^1, \ldots, x_1^l, \ldots, x_n^1, \ldots, x_n^l$ be $n$ blocks of $l$ boolean-valued random variables each, where $l$ is a positive integer whose value is determined later. For each $i \in \{1, \ldots, n\}$ we let random variable $x_i = .x_i^1 \cdots x_i^l$ be a binary fraction where $x_i^j$ is the $j^{th}$ most significant bit. We show, the event $< x_1, \ldots, x_n > \in R$ occurs with probability within $\epsilon$ of $\mathrm{vol}(R)$ for every rectangle $R \in \mathcal{R}_n^k$ when these random variables have the properties we develop below.

Without loss of generality, fix a rectangle $R = \prod_{i \in \{1,\ldots,n\}} [a_i, b_i) \in \mathcal{R}_n^k$ such that the first $k$ dimensions are the non-trivial ones, i.e. for all $i = k+1, \ldots, n$, $[a_i, b_i) = [0,1)$. For simplicity of presentation, for all $i = 1, \ldots, k$, we restrict the $i^{th}$ interval to be of the special form $[0, b_i)$. (The analysis for the case when the interval is of the general form $[a_i, b_i)$ is no more difficult technically, just not as clean.) To determine if the event $< x_1, \ldots, x_n > \in R$ occurs, it is enough to determine if the $k$ subevents $x_1 \in [0, b_1), \ldots, x_k \in [0, b_k)$ all occur simultaneously.

We think of determining the outcomes of the $k$ subevents starting with subevent $x_1 \in [0, b_1)$ and ending with $x_k \in [0, b_k)$. Let $b_1^j$ be the $j^{th}$ bit in the binary expansion of $b_1$. We compare the bits $x_1^1, \ldots, x_1^l$ with $b_1^1, \ldots, b_1^l$ one at a time, starting with the most significant bit and working down, stopping as soon as $x_1 \in [0, b_1)$ or $x_1 \notin [0, b_1)$ has been determined. Note that if $x_1^1 \neq b_1^1$ then the outcome of the first subevent is determined one way or the other, i.e. if $x_1^1 = 0$ and $b_1^1 = 1$ then $x_1 \in [0, b_1)$, whereas if $x_1^1 = 1$ and $b_1^1 = 0$ then $x_1 \notin [0, b_1)$. In this case, we move on to determine

the outcome of the second subevent. On the other hand, if $x_1^1 = b_1^1$ then the outcome of the first subevent hasn't been determined and we next compare $x_1^2$ with $b_1^2$, etc.

Determining the outcomes of the $k$ subevents can be viewed as a complete binary tree labeled with the boolean-valued random variables. The root of the tree is labeled with $x_1^1$, the left edge out of the root corresponds to $x_1^1 = b_1^1$ and the right edge corresponds to $x_1^1 \neq b_1^1$. At each subsequent node of the tree, the node is labeled with the boolean-valued random variable that is considered next; e.g. the left child of the root is labeled $x_1^2$ and the right child label is $x_2^1$.

Suppose for now that $x_1^1, \ldots, x_1^l, \ldots, x_n^1, \ldots, x_n^l$ are independently and uniformly distributed. A random setting of the variables defines a random path down the tree, and it is easy to see that if a random path is taken down this tree (and $l$ is infinite) then the probability that the $k$ subevents all simultaneously occur is exactly vol($R$). Furthermore, on average the values of two boolean-valued random variables are examined to determine the outcome of each subevent, and thus on average we examine $2k$ boolean-valued random variables to determine the outcomes of all $k$ subevents. Consider the probability that the outcomes of all $k$ subevents are not determined by the time the first $k'$ boolean-valued random variables are examined. This probability is exactly the same as the probability that there are less than $k$ "heads" in $k'$ tosses of a fair coin. By a standard analysis, when $k'$ is set to a value that is $O(k + \log(1/\epsilon))$ it can be easily shown that this probability is at most $\epsilon/2$. This shows the probability a random path down the tree to depth $k'$ doesn't determine the outcomes of all $k$ subevents is at most $\epsilon/2$. Consequently if $x_1^1, \ldots, x_1^l, \ldots, x_n^1, \ldots, x_n^l$ are $k'$-wise independent and uniformly distributed then the probability that all $k$ subevents occur simultaneously is within $\epsilon/2$ of vol($R$).

There are only $2^{k'}$ paths down to depth $k'$ in the tree. Thus, if, for every path down to depth $k'$ in the tree, the actual probability of the path is within $\epsilon/2^{k'+1}$ of $1/2^{k'}$ then the analysis shows that the probability all $k$ subevents occur simultaneously is within $\epsilon$ of vol($R$). From this it follows that any distribution on $x_1^1, \ldots, x_1^l, \ldots, x_n^1, \ldots, x_n^l$ that is a $(k', \epsilon')$-approximation for $\mathcal{U}_{nl,2}$ (with $k' = O(k + \log(1/\epsilon))$ and $\epsilon' = \epsilon/2^{k'+1}$ and $l = k'$) has the property that the probability that all $k$ subevents occur simultaneously is within $\epsilon$ of vol($R$) for all $R \in \mathcal{R}_n^k$. For these values of $l$, $k'$ and $\epsilon'$ we can use [13, Naor Naor] or [2, Alon Goldreich Håstad Peralta] to construct a sample space $S$ which induces a distribution on $x_1^1, \ldots, x_1^l, \ldots, x_n^1, \ldots, x_n^l$ that is a $(k', \epsilon')$-approximation for $\mathcal{U}_{nl,2}$ with $|S|$ polynomial in $\log(n)$, $2^k$ and $1/\epsilon$. ■

It should be noted that this analysis uses components of analysis for "Discrete Distribution Generating tree" described in [9, Knuth Yao] and also component of an analysis presented in [12, Luby Veličković].

Setting $k = O(\log(1/\epsilon))$ and using observation 6 from section 2 shows that for any constant $\beta < 1$ there is a set $S$ of size polynomial in $1/\epsilon$ and $\log(n)$ with $\Delta_S(\mathcal{R}_n^{[0,\beta)}) \leq \epsilon$.

### 3.1.1 Special case alternative

For the special case of $\mathcal{U}_{n,p}$ where $p$ is a small prime (e.g. $p = 3$) there is a smaller sample space that is a $(k, \epsilon)$-approximation for $\mathcal{U}_{n,p}$; the construction is a generalization of the construction for $\mathcal{U}_{n,2}$, and can be found in [2, Alon Goldreich Håstad Peralta], [3, Azar Motwani Naor] and [7, Even].

## 3.2 Using inclusion-exclusion

Let $k = O(\log(1/\epsilon))$ and let $S$ be a sample space that is a $k$-wise independent approximation for $\mathcal{U}_{n,4n/\epsilon}$. In this subsection, we show this implies $\Delta_S(\mathcal{R}_n) \leq \epsilon$ when points in $S$ are mapped in the natural way to $[0,1)^n$. Using standard constructions (see for example [11, Luby] or [1, Alon Babai Itai]), there is a constructible set $S$ with these properties of size polynomial in $(n/\epsilon)^{\log(1/\epsilon)}$.

The first easy observation is that $\Delta_S(\mathcal{R}_{n,4n/\epsilon}^k) = 0$ when points in $S$ are mapped in the natural way to $[0,1)^n$. Then, we use the theorem described below to show that $\Delta_S(\mathcal{R}_{n,4n/\epsilon}^k) = 0$ and $k = O(\log(1/\epsilon))$ implies that $\Delta_S(\mathcal{R}_{n,4n/\epsilon}) \leq \epsilon/2$. Finally, observation 4 from section 2 shows that $\Delta_S(\mathcal{R}_n) \leq \epsilon$.

**Theorem 2** *Let $\mathcal{P}_{n,2}$ be a general probability matrix for $n$ boolean-valued random variables $x_1, \ldots, x_n$. Then,*

$$|\Pr_D[\bigwedge_{i \in \{1,\ldots,n\}} x_i = 0] - \prod_{i \in \{1,\ldots,n\}} p_{i,0}| \leq 2^{-\Omega(k)}$$

*for any probability space $D$ that induces a distribution on $x_1, \ldots, x_n$ that is a $k$-wise independent approximation for $\mathcal{P}_{n,2}$. (Note that if $x_1, \ldots, x_n$ are independently distributed for $\mathcal{P}_{n,2}$, then the event $\bigwedge_{i \in \{1,\ldots,n\}} x_i = 0$ has probability exactly $\prod_{i \in \{1,\ldots,n\}} p_{i,0}$.)*

PROOF: The idea is to use the inclusion-exclusion formula. Fix $D$ to be any space that induces a distribution on $x_1, \ldots, x_n$ that is a $k$-wise independent approximation for $\mathcal{P}_{n,2}$. Define $T_0 = 1$ and for all $j = 1, \ldots, k$ define

$$T_j = \sum_{I=<i_1,\ldots,i_j>} \prod_{l=1,\ldots,j} p_{i_j,1},$$

i.e. $T_j$ is the $j^{th}$ term of the inclusion-exclusion formula. Then, for all even values of $j$, $\sum_{l=0,\ldots,j} (-1)^j T_j$ is an upper bound on $\Pr_D[\bigwedge_{i \in \{1,\ldots,n\}} x_i = 0]$, this quantity is

a lower bound for all odd values of $j$ and $T_k$ is an upper bound on $|\Pr_D[\bigwedge_{i\in\{1,\ldots,n\}} x_i = 0] - \prod_{i\in\{1,\ldots,n\}} p_{i,0}|$.

Define $\alpha = \sum_{i\in\{1,\ldots,n\}} p_{i,1}$. There are two cases to the proof, depending on whether $\alpha \leq \frac{k}{2e}$ or $\alpha > \frac{k}{2e}$. (Where $e = 2.718\ldots$) Suppose that $\alpha \leq \frac{k}{2e}$. We show this implies $T_k \leq 2^{-k}$, which finishes the proof for the first case. This inequality holds because, subject to the restriction that $\sum_{i\in\{1,\ldots,n\}} p_{i,1} = \alpha$, $T_k$ is maximized when, for all $i \in \{1,\ldots,n\}$, $p_{i,1} = \alpha/n$. Thus, $T_k \approx (\frac{\alpha}{ek})^k \leq 2^{-k}$. Now suppose that $\alpha > \frac{k}{2e}$. Consider the first $n' < n$ random variables such that $\frac{k}{2e} - 1 < \sum_{i\in\{1,\ldots,n'\}} p_{i,1} \leq \frac{k}{2e}$ and let $\alpha' = \sum_{i\in\{1,\ldots,n'\}} p_{i,1} \approx \frac{k}{2e}$. We first show this implies $\prod_{i\in\{1,\ldots,n'\}} p_{i,0} \leq 2^{-\Omega(k)}$ and then we show how to finish the proof from this for the second case. Subject to the restriction that $\sum_{i\in\{1,\ldots,n\}} p_{i,1} = \alpha'$, $\prod_{i\in\{1,\ldots,n'\}} p_{i,0}$ is maximized when, for all $i \in \{1,\ldots,n'\}$, $p_{i,1} = \alpha'/n'$. Thus, because $\alpha' \approx \frac{k}{2e}$, $\prod_{i\in\{1,\ldots,n'\}} p_{i,0} \leq (1-\alpha'/n')^{n'} = 2^{-\Omega(k)}$. From the same proof as used in the first case, noting that $\alpha' \leq \frac{k}{2e}$, $\Pr_D[\bigwedge_{i\in\{1,\ldots,n'\}} x_i = 0] \leq \prod_{i\in\{1,\ldots,n'\}} p_{i,0} + 2^{-k}$. Because $\prod_{i\in\{1,\ldots,n'\}} p_{i,0} \leq 2^{-\Omega(k)}$ and because $\Pr_D[\bigwedge_{i\in\{1,\ldots,n\}} x_i = 0] \leq \Pr_D[\bigwedge_{i\in\{1,\ldots,n'\}} x_i = 0]$, this implies that $0 \leq \Pr_D[\bigwedge_{i\in\{1,\ldots,n\}} x_i = 0] \leq 2^{-\Omega(k)}$. This and $0 \leq \prod_{i\in\{1,\ldots,n\}} p_{i,0} \leq \prod_{i\in\{1,\ldots,n'\}} p_{i,0}$ finishes the proof of the second case. ■

The obvious corollary to this theorem we use to prove the result stated at the beginning of this section is that $\Delta_S(\mathcal{R}^k_{n,m}) = 0$ implies that $\Delta_S(\mathcal{R}_{n,m}) \leq 2^{-\Omega(k)}$.

This theorem is interesting in its own right and should be contrasted with the main theorem of [10, Linial Nisan]. Loosely stated, the above theorem says that the leading $O(\log(1/\epsilon))$ terms of the inclusion-exclusion formula completely determines the probability of the union of the $n$ events to within an error $\epsilon$ if these terms correspond to the leading terms of the *independent* distribution on the $n$ events. Loosely stated, one direction of the main theorem in [10, Linial Nisan] says that, for any $l$ sufficiently smaller than $\sqrt{n}$, the leading $l$ terms of the inclusion-exclusion formula doesn't even determine the probability of the union of the $n$ events to within a constant amount if these terms only correspond to the leading terms of *some* probability distribution on the $n$ events.

One application of the result is to deterministic approximation of the number of satisfying truth assignments to a disjunctive normal form boolean formula [12, Luby Veličković]. A more philosophical application is that the result says that the probability of unions of events that are somewhat independent and the probability of unions of events that are totally independent are not very different. This gives some partial justification for modeling "real world" events, which are somewhat independent but not totally so, by events that are totally independent, without drastically affecting the probability of their union.

### 3.3 Using hashing

The third construction uses the results given in section 5 of [14, Nisan] and observation 4 of section 2. The result is that there is an efficiently constructible set $S$ of size polynomial in $(n/\epsilon)^{\log(n)}$ such that $\Delta_S(\mathcal{R}_n) \leq \epsilon$.

## 4 Open Problems

One open problem motivated by this work can be found at the end of subsection 1.4. An even harder problem, which was motivation for this work, is the following generalization of that problem: Find an efficiently constructible set $S$ of size polynomial in $n$, $m$ and $1/\epsilon$ such that for any union of at most $m$ rectangles in $n$ dimensional space, the fraction of points in $S$ that fall in their union is within $\epsilon$ of the volume of their union. A positive solution to this problem would provide an efficient deterministic approximation algorithm for the DNF counting problem.

In subsection 3.2 it is shown that, for $k = O(\log(1/\epsilon))$, if $S$ is a sample space that is a $k$-wise independent approximation for $\mathcal{U}_{n,4n/\epsilon}$ then $\Delta_S(\mathcal{R}_n) \leq \epsilon$. It can be easily seen that if $S'$ is any probability distribution that is within statistical distance $\delta$ of the distribution induced by $S$ then $\Delta_{S'}(\mathcal{R}_n) \leq \epsilon + \delta$. Hence, it is important to find out whether it is possible to construct a polynomial size sample space $S'$ that induces a distribution within statistical distance $\delta$ of some distribution that is a $k$-wise independent approximation. A natural suggestion is to try to use a sample space $S'$ that is a $(k,\delta')$-approximation, for some $\delta'$ that is polynomial in $\delta$. Unfortunately, all *we know* is that for any $(k,\delta')$-approximation, for say $\mathcal{U}_{n,2}$, the statistical distance is $\binom{n}{k} \cdot \delta'$ from $k$-wise independence. A better bound, say polynomial in $n$, $2^k$ and $1/\delta'$, would be very useful, but it *seems* that the upper bound presented above is close to is optimal. Providing tight bounds on the statistical distance from a distribution that is a $(k,\delta')$-approximation to the closest distribution that is a $k$-wise independent approximation is an interesting open problem.

## 5 Acknowledgments

We thank Emo Welzl for discussions which led to the understanding of the connections between approximations of distributions and discrepancy theory. We thank Nati Linial and Avi Wigderson for a number of helpful technical discussions. We thank Josef Beck for sharing with

us his enthusiasm for this work and his knowledge about discrepancies.

# 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", *Proc. 31st FOCS*, 1990.

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

[4] Beck, J., Chen, W., "Irregularities of distribution", Cambridge University Press, 1987.

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

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

[7] 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) in Aug. 1991. (In Hebrew, abstract in English).

[8] Karp, R., Wigderson, A., "A Fast Parallel Algorithm for the Maximal Independent Set Problem", proceedings of $16^{\text{th}}$ ACM Symposium on Theory of Computing, 1984.

[9] 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.

[10] Linial, N., Nisan, N., "Approximate Inclusion-Exclusion", $22^{nd}$ *STOC*, 1990.

[11] Luby, M., "A Simple Parallel Algorithm for the Maximal Independent Set Problem," $17^{\text{th}}$ *STOC*, May 6-8 1985, pp. 1-10, *SIAM J. on Computing*, November 1986, Volume 15, No. 4, pp. 1036-1053

[12] Luby, M., Veličković, B., "On Deterministic Approximation of DNF", *Proc. 23rd STOC*, 1991, pp. 430–438.

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

[14] Nisan, N., "Pseudo-random Generators for Space-Bounded Computation", $22^{nd}$ *STOC*, May 14-16 1990, pp. 204-212.

[15] Niederreiter, H.,
"Constructions of Low-Discrepancy Point Sets and Sequences", manuscript and lecture notes.