Almost \( k \)-wise independence versus \( k \)-wise independence

Noga Alon*  
Sackler Faculty of Exact Sciences  
Tel Aviv University  
Ramat-Aviv, ISRAEL.  
nogaa@post.tau.ac.il

Oded Goldreich†  
Department of Computer Science  
Weizmann Institute of Science  
Rehovot, ISRAEL.  
oded@wisdom.weizmann.ac.il

Yishay Mansour  
School of Computer Science  
Tel Aviv University  
Ramat-Aviv, ISRAEL.  
mansour@post.tau.ac.il

July 31, 2002

Abstract

We say that a distribution over \( \{0,1\}^n \) is almost \( k \)-wise independent if its restriction to every \( k \) coordinates results in a distribution that is close to the uniform distribution. A natural question regarding almost \( k \)-wise independent distributions is how close they are to some \( k \)-wise independent distribution. We show that the latter distance is essentially \( n^{\Theta(k)} \) times the former distance.

Keywords: Small probability spaces, \( k \)-wise independent distributions, almost \( k \)-wise independent distributions, small bias probability spaces.

*Research supported in part by a USA Israeli BSF grant, by a grant from the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

†Supported by the MINERVA Foundation, Germany.
1 Introduction

Small probability spaces of limited independence are useful in various applications. Specifically, as observed by Luby [4] and others, if the analysis of a randomized algorithm only relies on the hypothesis that some objects are distributed in a k-wise independent manner then one can replace the algorithm’s random-tape by a string selected from a k-wise independent distribution. Recalling that k-wise independent distributions over \( \{0,1\}^n \) can be generated using only \( O(k \log n) \) bits (see, e.g., [1]), this yields a significant saving in the randomness complexity as well as to derandomization in time \( n^{O(k)} \). (This number of random bits is essentially optimal; see [3], [1].)

Further saving is possible whenever the analysis of the randomized algorithm can be carried out also in case its random-tape is only “almost k-wise independent” (i.e., every k bits are distributed almost uniformly). The reason being that the latter distributions can be generated using fewer random bits (i.e., \( O(k + \log(n/e)) \) bits suffice, where \( e \) is the variation distance of these k-projections to the uniform distribution): See the work of Naor and Naor [5] (as well as subsequent simplifications in [2]).

Note that, in both cases, replacing the algorithm’s random-tape by strings taken from a distribution of a smaller support requires verifying that the original analysis still holds for the replaced distribution. It would have been nicer, if instead of re-analyzing the algorithm for the case of almost k-wise independent distributions, we could just re-analyze it for the case of k-wise independent distributions and apply a generic result. Such a result may say that if the algorithm behaves well under any k-wise independent distribution then it would behave essentially as well also under any almost k-wise independent distribution, provided that the parameter \( e \) governing this measure of closeness is small enough. Of course, the issue is how small should \( e \) be.

A generic approach towards the above question is to ask what is the statistical distance \( \delta \) between any almost k-wise independent distribution and some k-wise independent distribution. Specifically, how does this distance \( \delta \) depend on \( n \) and \( k \) (and on the parameter \( e \)). Note that we will have to set \( e \) sufficiently small so that \( \delta \) will be small (e.g., \( \delta = 0.1 \) may do).

Our original hope was that \( \delta = \text{poly}(2^k, n) \cdot e \) (or \( \delta = \text{poly}(2^k, n) \cdot e^{1/O(\log k)} \)). If this were the case, we could have set \( e = \text{poly}(2^{-k}, n^{-1}, \delta) \), and use an almost k-wise independent sample space of size \( \text{poly}(n/e) = \text{poly}(2^{k}, n, \delta^{-1}) \) (instead of size \( n^{O(k)} \) as for perfect k-wise independence). Unfortunately, the answer is that \( \delta = n^{O(k)} \cdot e \), and so this generic approach does not lead to anything better than just using an adequate k-wise independent sample space. In fact we show that every distribution with support less than \( n^{O(k)} \) has large statistical distance to any k-wise independent distribution.

2 Formal Setting

We consider distributions and random variables over \( \{0,1\}^n \), where \( n \) (as well as \( k \) and \( e \)) is a parameter. A distribution \( D_X \) over \( \{0,1\}^n \) assigns each \( z \in \{0,1\}^n \) a value \( D_X(z) \in [0,1] \) such that \( \sum_z D_X(z) = 1 \). A random variable \( X \) over \( \{0,1\}^n \) is associated with a distribution \( D_X \) and randomly selects a \( z \in \{0,1\}^n \), where \( \Pr[X = z] = D_X(z) \). Throughout the paper we use interchangeably the notation of a random variable and a distribution. The statistical distance, denoted \( \Delta(X,Y) \), between two random variables \( X \) and \( Y \) over \( \{0,1\}^n \) is defined as

\[
\Delta(X,Y) \overset{\text{def}}{=} \frac{1}{2} \cdot \sum_{z \in \{0,1\}^n} |\Pr[X = z] - \Pr[Y = z]| \\
= \max_{S \subseteq \{0,1\}^n} \{\Pr[X \in S] - \Pr[Y \in S]\}
\]
If $\Delta(X, Y) \leq \epsilon$ the we say that $X$ is $\epsilon$-close to $Y$. (Note that $2\Delta(X, Y)$ is equivalent to $\|D_X - D_Y\|_1$, where $\|w\|_1 = \sum |w_i|$.)

A distribution $X = X_1 \cdots X_n$ is called an $(\epsilon, k)$-approximation if for every $k$ (distinct) coordinates $i_1, \ldots, i_k \in \{1, \ldots, n\}$ it holds that $X_{i_1} \cdots X_{i_k}$ is $\epsilon$-close to the uniform distribution over $\{0,1\}^k$. An $(0,k)$-approximation is sometimes referred to as a $k$-wise independent distribution (i.e., for every $k$ (distinct) coordinates $i_1, \ldots, i_k \in \{1, \ldots, n\}$ it holds that $X_{i_1} \cdots X_{i_k}$ is uniform over $\{0,1\}^k$).

A related notion is that of having bounded bias on (non-empty) sets of size at most $k$. Recall that the bias of a distribution $X = X_1 \cdots X_n$ on a set $I$ is defined as

$$\text{bias}_I(X) \overset{\text{def}}{=} E[(-1)\sum_{i \in I} X_i] = \Pr[\oplus_{i \in I} X_i = 0] - \Pr[\oplus_{i \in I} X_i = 1] = 2\Pr[\oplus_{i \in I} X_i = 0] - 1$$

Clearly, for any $(\epsilon, k)$-approximation $X$, the bias of the distribution $X$ on every non-empty subset of size at most $k$ is bounded above by $\epsilon$. On the other hand, if $X$ has bias at most $\epsilon$ on every non-empty subset of size at most $k$ then $X$ is an $(2^{k/2} \cdot \epsilon, k)$-approximation (see [7] and the Appendix in [2]).

Since we are willing to give up on $\exp(k)$ factors, we state our results in terms of distributions of bounded bias.

**Theorem 2.1** (Upper Bound): Let $X = (X_1, \ldots, X_n)$ be a distribution over $\{0,1\}^n$ such that the bias of $X$ on any non-empty subset of size up to $k$ is at most $\epsilon$. Then $X$ is $\delta(n,k,\epsilon)$-close to some $k$-wise independent distribution, where $\delta(n,k,\epsilon) \overset{\text{def}}{=} \sum_{i=1}^{k} \binom{n}{i} \cdot \epsilon \leq n^k \cdot \epsilon$.

The proof appears in Section 3.1. It follows that any $(\epsilon, k)$-approximation is $\delta(n,k,\epsilon)$-close to some $(0,k)$-approximation. We show that the above result is nearly tight in the following sense.

**Theorem 2.2** (Lower Bound): For every $n$, every even $k$ and every $\epsilon$ such that $\epsilon > 2^{k/2}/n^{(k/4) - 1}$ there exists a distribution $X$ over $\{0,1\}^n$ such that

1. The bias of $X$ on any non-empty subset is at most $\epsilon$.

2. The distance of $X$ from any $k$-wise independent distribution is at least $\frac{1}{2}$.

The proof appears in Section 3.2. In particular, setting $\epsilon = n^{-k/5}/2$ (which, for sufficiently large $n \gg k \gg 1$, satisfies $\epsilon > 2^{k/2}/n^{(k/4) - 1}$), we obtain that $\delta(n,k,\epsilon) \geq 1/2$, where $\delta(n,k,\epsilon)$ is as in Theorem 2.1. Thus, if $\delta(n,k,\epsilon) = f(n,k) \cdot \epsilon$ (as is natural and is indeed the case in Theorem 2.1) then it must hold that

$$f(n,k) \geq \frac{1}{2\epsilon} = n^{-k/5}$$

A similar analysis holds also in case $\delta(n,k,\epsilon) = f(n,k) \cdot \epsilon^{1/\Omega(1)}$. We remark that although Theorem 2.2 is shown for an even $k$, a bound for an odd $k$ can be trivially derived by replacing $k$ by $k - 1$.

## 3 Proofs

### 3.1 Proof of Theorem 2.1

Going over all non-empty sets, $I$, of size up to $k$, we make the bias over these sets zero, by augmenting the distribution as follows. Say that the bias over $I$ is exactly $\epsilon > 0$ (w.l.o.g., the bias is positive); that is, $\Pr[\oplus_{i \in I} X_i = 0] = (1 - \epsilon)/2$. Then (for $p \approx \epsilon$ to be determined below), we define a new distribution $Y = Y_1 \cdots Y_n$ as follows.
1. With probability $1 - p$, we let $Y = X$.

2. With probability $p$, we let $Y$ be uniform over the set \( \{ \sigma_1 \cdots \sigma_n \in \{0,1\}^n : \oplus_i \sigma_i = 1 \} \).

Then \( \Pr[\oplus_i Y_i = 0] = (1 - p) \cdot ((1 + \epsilon)/2) + p \cdot 0 \). Setting \( p = \epsilon/(1 + \epsilon) \), we get \( \Pr[\oplus_i Y_i = 0] = 1/2 \) as desired. Observe that \( \Delta(X, Y) \leq p < \epsilon \) and that we might have only decreased the biases on all other subsets. To see the latter, consider a non-empty \( J \neq I \), and notice that in Case (2) \( Y \) is unbiased over \( J \). Then

\[
\left| \Pr[\oplus_i Y_i = 1] - \frac{1}{2} \right| = \left| \left( (1 - p) \cdot \Pr[\oplus_i X_i = 1] + p \cdot \frac{1}{2} \right) - \frac{1}{2} \right|
\]

\[
= (1 - p) \cdot \left| \Pr[\oplus_i X_i = 1] - \frac{1}{2} \right|
\]

The theorem follows. □

### 3.2 Proof of Theorem 2.2

On one hand, we know (cf., [2], following [5]) that there exists \( \epsilon \)-bias distributions of support size \((n/\epsilon)^2\). On the other hand, we will show (in Lemma 3.1) that every \( k \)-wise independent distribution, not only has large support (as proven, somewhat implicitly, in [6] and explicitly in [3] and [1]), but also has a large min-entropy bound. It follows that every \( k \)-wise independent distribution must be far from any distribution that has a small support, and thus be far from any such \( \epsilon \)-bias distribution. Recall that a distribution \( Z \) has min-entropy \( m \) if \( \Pr[Z = \alpha] \leq 2^{-m} \) holds for every \( \alpha \). (Note that min-entropy is equivalent to \( \| \log_2 \|D_Z\|_\infty \|, \) where \( \| \|_\infty = \max_i |v_i| \).)

**Lemma 3.1** For every \( n \) and every even \( k \), any \( k \)-wise independent distribution over \( \{0,1\}^n \) has min-entropy at least \( -\log_2(k^k n^{-k/2}) \).

Let us first see how to prove Theorem 2.2, using Lemma 3.1. First we observe, that a distribution \( Y \) that has min-entropy \( m \) must be at distance at least \( 1/2 \) from any distribution \( X \) that has support \( 2^m/2 \). This follows because

\[
\Delta(Y, X) \geq \Pr[Y \in (\{0,1\}^n \setminus \text{support}(X))] = 1 - \sum_{\alpha \in \text{support}(X)} \Pr[Y = \alpha] \geq 1 - |\text{support}(X)| \cdot 2^{-m} \geq \frac{1}{2}
\]

Now, letting \( X \) be an \( \epsilon \)-bias distribution (i.e., having bias at most \( \epsilon \) on every non-empty subset) of support \((n/\epsilon)^2\) and using Lemma 3.1 (while observing that \( \epsilon > 2k^{k/2}/n^{(k/4)-1} \) implies \((n/\epsilon)^2 < 2^m/2 \) for \( m = \log_2(n^{k/2}/k^k) \)), Theorem 2.2 follows. In fact we can derive the following corollary.

**Corollary 3.2** For every \( n \), every even \( k \), and for every \( k \)-wise independent distribution \( Y \), if distribution \( X \) has support smaller than \( n^{k/2}/2k^k \) then \( \Delta(X, Y) \geq \frac{1}{2} \).

**Proof of Lemma 3.1:** Let \( Y \) be a \( k \)-wise independent distribution, and \( \alpha \) be a string maximizing \( \Pr[Y = \alpha] \). Assume (w.l.o.g., by shifting/XORing \( Y \) by \( \alpha \)) that \( \alpha \) is the all-zero string. We consider the \( k \)-th moment of \( Y \); i.e., \( E[(\sum_i (Y_i - 0.5))^k] \).
Upper bound: Following standard manipulation, we let \( Z_i = Y_i - 0.5 \), (note that \( \mathbb{E}[Z_i] = 0 \)) and write
\[
\mathbb{E} \left[ \left( \sum_i Z_i \right)^k \right] = \sum_{i_1, \ldots, i_k \in [n]} \mathbb{E}[Z_{i_1} \cdots Z_{i_k}] .
\] (1)

Observe that all (r.h.s) terms in which some index appears only once are zero (i.e., if for some \( j \) and all \( h \neq j \) it holds that \( i_j \neq i_h \) then \( \mathbb{E} [\prod_{h} Z_{i_h}] = \mathbb{E}[Z_{i_j}] \cdot \mathbb{E}[\prod_{h \neq j} Z_{i_h}] = 0 \)). All the remaining terms are such that each index appears at least twice. The number of these terms is bounded above by \( \binom{n}{k/2} \cdot (k/2)^k < (k/2)^k \cdot n^{k/2} \), and each contributes at most 1 to the sum. Thus, Eq. (1) is strictly smaller than \((k/2)^k \cdot n^{k/2}\).

Lower bound: We write the formal expression for expectation (of the l.h.s of Eq. (1)).
\[
\mathbb{E} \left[ \left( \sum_i Z_i \right)^k \right] = \mathbb{E} \left[ \left( \left( \sum_i Y_i \right) - (n/2) \right)^k \right]
\]
\[
\quad = \sum_{\sigma_1 \cdots \sigma_n \in \{0,1\}^n} \Pr[(\forall i) \ Y_i = \sigma_i] \cdot \left( \left( \sum_i \sigma_i \right) - (n/2) \right)^k
\]
\[
\quad \geq \Pr[(\forall i) \ Y_i = 0] \cdot (-n/2)^k
\]

where we use the fact that all terms are non-negative (because \( k \) is even).

Combining the two bounds on Eq. (1), we infer that \((n/2)^k \cdot \Pr[Y = 0^n] < (k/2)^k \cdot n^{k/2}\), and we get \(\Pr[Y = 0^n] < ((k/2)^k \cdot n^{k/2})/(n/2)^k = k^k n^{-k/2}\). The lemma follows.

\[ \square \]

References


