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

December 12, 2003

Abstract

We say that a distribution over $\{0,1\}^n$ is (ϵ, k)-wise independent if its restriction to every k coordinates results in a distribution that is ϵ-close to the uniform distribution. A natural question regarding (ϵ, k)-wise independent distributions is how close they are to some k-wise independent distribution. We show that there exists (ϵ, k)-wise independent distribution whose statistical distance is at least $n^{O(k)} \cdot \epsilon$ from any k-wise independent distribution. In addition, we show that for any (ϵ, k)-wise independent distribution there exists some k-wise independent distribution, whose statistical distance is $n^{O(k)} \cdot \epsilon$.

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.

‡Research supported in part by a grant from the Israel Science Foundation.

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 and also leads to a 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., (ϵ,k)-wise distribution where every k bits are distributed ϵ-close to uniform). This is because (ϵ,k)-wise distributions can be generated using fewer random bits (i.e., $O(k+\log(n/\epsilon))$ bits suffice, where ϵ 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 (ϵ,k)-wise independent distributions, we could just re-analyze it for the case of k-wise independent distributions and apply a generic result. This would imply that the randomized algorithm can be further derandomized using an (ϵ,k)-wise independent distribution. 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 ϵ governing this measure of closeness is small enough. Of course, the issue is how small ϵ should be.

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

Our original hope was that $\delta = \text{poly}(2^k, n) \cdot \epsilon$ (or $\delta = \text{poly}(2^k, n) \cdot \epsilon^{1/O(1)}$). If this had been the case, we could have set $\epsilon = \text{poly}(2^{-k}, n^{-1}, \delta)$, and use an almost k-wise independent sample space of size $\text{poly}(n/\epsilon) = \text{poly}(2^k, n, \delta^{-1})$ (instead of size $n^{\Theta(k)}$ as for perfect k-wise independence). Unfortunately, the answer is that $\delta = n^{\Theta(k)} \cdot \epsilon$, and so this generic approach (which requires setting $\epsilon = \delta/n^{\Theta(k)}$ and using a sample space of size $\Omega(1/\epsilon) = n^{\Theta(k)}$) 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 $n^{\Theta(k)}$ 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 ϵ) 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 $\mathsf{Pr}[X = z] = D_X(z)$. 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

$$\Delta(X,Y) \stackrel{\text{def}}{=} \frac{1}{2} \cdot \sum_{z \in \{0,1\}^n} |\mathsf{Pr}[X=z] - \mathsf{Pr}[Y=z]|$$

$$= \max_{S\subset\{0,1\}^n} \{\mathsf{Pr}[X\in S] - \mathsf{Pr}[Y\in S]\}.$$

If $\Delta(X,Y) \leq \epsilon$ we say that X is ϵ-close to Y. (Note that $2\Delta(X,Y)$ equals to $\|D_X - D_Y\|_1 = \sum_{z\in\{0,1\}^n} |D_X(z) - D_Y(z)|$.)

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

$$\begin{aligned} \text{bias}_I(X) &\stackrel{\text{def}}{=} \mathsf{E}[(-1)^{\sum_{i\in I} X_i}] \\ &= \mathsf{Pr}[\oplus_{i\in I} X_i = 0] - \mathsf{Pr}[\oplus_{i\in I} X_i = 1] \;=\; 2\mathsf{Pr}[\oplus_{i\in I} X_i = 0] - 1. \end{aligned}$$

Clearly, for any (ϵ,k)-approximation X, the bias of the distribution X on every non-empty subset of size at most k is bounded above by 2ϵ. On the other hand, if X has bias at most ϵ on every non-empty subset of size at most k then X is a $(2^{k/2}\cdot\epsilon, k)$-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 $X = X_1....X_n$ 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 ϵ. Then X is $\delta(n,k,\epsilon)$-close to some k-wise independent distribution, where $\delta(n,k,\epsilon) \stackrel{\text{def}}{=} \sum_{i=1}^k \binom{n}{i}\cdot\epsilon \leq n^k\cdot\epsilon$.*

The proof appears in Section 3.1. It follows that any (ϵ,k)-approximation is $\delta(n,k,\epsilon)$-close to some $(0,k)$-approximation. (We note that our construction is not efficient both in its computation time and in the number of random bits, and its main purpose is to generate some k-wise independent distribution.) 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 ϵ such that $\epsilon > 2k^{k/2}/n^{(k/4)-1}$ there exists a distribution X over $\{0,1\}^n$ such that*

1. *The bias of X on any non-empty subset is at most ϵ.*
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 $\epsilon > 2k^{k/2}/n^{(k/4)-1}$), 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 $\delta(n,k,\epsilon) = f(n,k)\cdot\epsilon^{1/O(1)}$. 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, $\mathsf{Pr}[\oplus_{i\in I} X_i = 0] = (1+\epsilon)/2$. Then (for $p \approx \epsilon$ to be determined below), we define a new distribution $Y = Y_1 ... Y_n$ as follows.

1. With probability $1-p$, we let $Y = X$.

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

Then $\mathsf{Pr}[\oplus_{i\in I} Y_i = 0] = (1-p)\cdot((1+\epsilon)/2) + p\cdot 0$. Setting $p = \epsilon/(1+\epsilon)$, we get $\mathsf{Pr}[\oplus_{i\in I} Y_i = 0] = 1/2$ 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

$$\begin{aligned}\left|\mathsf{Pr}[\oplus_{i\in J} Y_i = 1] - \frac{1}{2}\right| &= \left|\left((1-p)\cdot \mathsf{Pr}[\oplus_{i\in J} X_i = 1] + p\cdot\frac{1}{2}\right) - \frac{1}{2}\right| \\ &= (1-p)\cdot\left|\mathsf{Pr}[\oplus_{i\in J} X_i = 1] - \frac{1}{2}\right|,\end{aligned}$$

and the theorem follows. ■

3.2 Proof of Theorem 2.2

On one hand, we know (cf., [2], following [5]) that there exists ϵ-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. (Recall that the support of a distribution is the set of points with non-zero probability.) 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 ϵ-bias distribution. Recall that a distribution Z has min-entropy m if m is the maximal real such that $\mathsf{Pr}[Z = \alpha] \leq 2^{-m}$ holds for every α. (Note that min-entropy is equivalent to $-\log_2 \max_{x\in\{0,1\}^n} D_Z(x)$.)

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

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 of size at most $2^m/2$. This follows because

$$\begin{aligned}\Delta(Y,X) &\geq \mathsf{Pr}[Y \in (\{0,1\}^n \setminus \mathrm{support}(X))] \\ &= 1 - \sum_{\alpha\in \mathrm{support}(X)} \mathsf{Pr}[Y = \alpha] \\ &\geq 1 - |\mathrm{support}(X)|\cdot 2^{-m} \geq \frac{1}{2}.\end{aligned}$$

Now, letting X be an ϵ-bias distribution (i.e., having bias at most ϵ 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 $Y = Y_1 \cdots Y_n$ be a k-wise independent distribution, and α be a string maximizing $\mathsf{Pr}[Y=\alpha]$. The key observation is that we may assume, without loss of generality (by XORing Y with α), that α is the all-zero string. Now, the lemma follows by applying a standard tail inequality for the sum of k-wise independent variables $Y_1, ..., Y_n$. Specifically, using the generalized Chebychev Inequality and defining $Z_i \stackrel{\text{def}}{=} Y_i - 0.5$, we have:

$$\begin{aligned}
\mathsf{Pr}[(\forall i)\ Y_i = 0] &= \mathsf{Pr}\left[\sum_{i=1}^{n} Y_i = 0\right] \\
&\leq \mathsf{Pr}\left[\left|\sum_{i=1}^{n} Z_i\right| \geq \frac{n}{2}\right] \\
&\leq \frac{\mathsf{E}\left[(\sum_i Z_i)^k\right]}{(n/2)^k} .
\end{aligned}$$

Next, we use a crude bound on the k-th moment of the sum of the Z_i's, which follows from their k-wise independence. Specifically, we first write

$$\mathsf{E}\left[\left(\sum_i Z_i\right)^k\right] = \sum_{i_1,...,i_k \in [n]} \mathsf{E}[Z_{i_1} \cdots Z_{i_k}] . \tag{1}$$

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 $\mathsf{E}[\prod_h Z_{i_h}] = \mathsf{E}[Z_{i_j}] \cdot \mathsf{E}[\prod_{h \neq j} Z_{i_h}] = 0$). All the remaining terms are such that each index appears at least twice. The number of these terms is smaller than $\binom{n}{k/2} \cdot (k/2)^k < (k/2)^k \cdot n^{k/2}$, and each contributes at most $1/2^k < 1$ to the sum (because each Z_i resides in $[\pm 0.5]$). Thus, the expression in Eq. (1) is strictly smaller than $(k/2)^k \cdot n^{k/2}$. The lemma follows (because $\mathsf{Pr}[(\forall i)\ Y_i = 0] \leq (k/2)^k \cdot n^{k/2}/(n/2)^k$). ■

References

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

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

[3] B. Chor, J. Friedmann, O. Goldreich, J. Håstad, S. Rudich and R. Smolensky. The bit extraction problem and t-resilient functions. In *26th FOCS*, pages 396–407, 1985.

[4] M. Luby. A Simple Parallel Algorithm for the Maximal Independent Set Problem. *SIAM J. on Computing*, Vol. 15, No. 4, pages 1036–1053, November 1986. (Preliminary version in *17th STOC*, 1985.)

[5] J. Naor and M. Naor. Small-bias Probability Spaces: Efficient Constructions and Applications. *SIAM J. on Computing*, Vol 22, 1993, pages 838–856. (Preliminary version in *22nd STOC*, 1990.)

[6] C. R. Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *J. Royal Stat. Soc.* 9, pages 128–139, 1947.

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