k-wise Independence and ϵ-biased k-wise Indepedence

February 10, 1999 **Scribe: Felix Wu**

1 Definitions

Consider a distribution D on n bits $x = x_1 \cdots x_n$. D is *k-wise independent* iff for all sets of k indices $S = \{i_1, \ldots, i_k\}$,

$$\forall a_1, \ldots, a_k, \Pr[x_{i_1} \cdots x_{i_k} = a_1 \cdots a_k] = \frac{1}{2^k}.$$

The idea is that if we restrict our attention to any k positions in x, no matter how many times we sample from D, we cannot distinguish D from the uniform distribution over n bits.

We can get a Fourier interpretation of k-wise independence by viewing the distribution D as a function from $\mathcal{Z}_2^n$ to $\mathcal{R}$. Let $|x|$ denote the Hamming weight of x.

Claim 1 *D is k-wise independent (according to the definition above) iff $\hat{D}(y) = 0$ for all $y \neq 0 : |y| \leq k$.*

Proof: (sketch) For the forward direction, recall that $\hat{D}(y) = \sum_x (-1)^{y \cdot x} D(x)$. Let S_y be the set of ones in y, and let $X_{S_y} = \bigoplus_{i \in S_y} x_i$. Then,

$$\hat{D}(y) = \sum_{x : X_{S_y} = 0} D(x) - \sum_{x : X_{S_y} = 1} D(x) = \Pr[X_{S_y} = 0] - \Pr[X_{S_y} = 1].$$

For $|y| = k$, the definition of k-wise independence guarantees that if we look at just the k positions S_y, each of the 2^k possible values is equally likely. In particular, this means that the XOR of these bits is equally likely to be a 0 or a 1, and hence $\hat{D}(y) = 0$. Since k-wise independence implies $(k-1)$-wise independence, this same argument shows that $\hat{D}(y) = 0$ for any $y : 0 < |y| < k$.

Conversely, suppose all the low-order Fourier coefficients are 0. Since $\hat{D}(100 \cdots 0) = \Pr[x_1 = 0] - \Pr[x_1 = 1] = 0$, x_1 is equally likely to be 0 or 1, and similarly for the other x_i's. Now let $p_{a_1 a_2} = \Pr[x_1 x_2 = a_1 a_2]$. Since x_1, x_2, and $x_1 \oplus x_2$ are all unbiased, we get the equations $p_{00} + p_{01} = p_{00} + p_{10} = p_{00} + p_{11} = 1/2$. Furthermore, $p_{00} + p_{01} + p_{10} + p_{11} = 1$. From these equations, we deduce that all of the $p_{a_1 a_2}$'s are $1/4$, and similarly for any pair x_i and x_j. Proceeding inductively in this manner shows that D is k-wise independent. ■

Note that $\hat{D}(0) = \sum D(x)$. Hence, $\hat{D}(0) = 1$ iff D is a valid distribution.

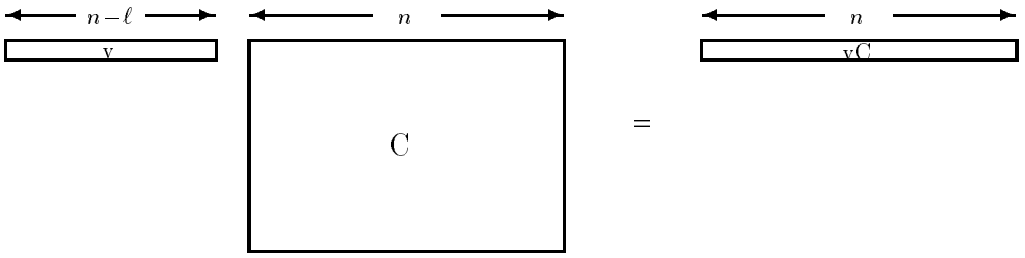
2 Constructing k-wise Independence

In general, our goal is to construct a k-wise independent distribution D which is uniform over as small a space $S \subseteq \mathcal{Z}_2^n$ as possible. Since we have a Fourier interpretation of k-wise independence, it seems natural to start from the Fourier coefficients and then transform back to a distribution.

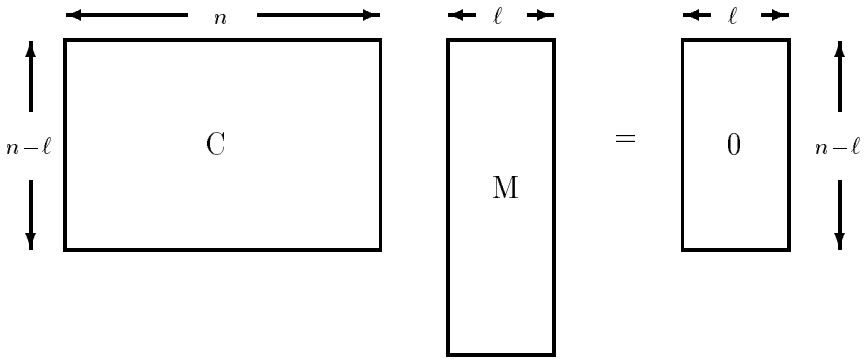
In particular, recall that if $\hat{D}$ is uniform on a subspace $C \subseteq \mathcal{Z}_2^n$, then D will be uniform on $C^\perp$. (We will take $\hat{D}(0) = 1$ to ensure that D is a valid distribution.) If we make C as large as possible, this will ensure that D is uniform over as small a space as possible.

Let C be generated by $c_1, \ldots, c_{n-\ell}$, i.e. $C = \left\{\sum_{x \in \mathcal{Z}_2^{n-\ell}} x_i c_i\right\}$. Then, $C^\perp = \{y : \forall c \in C,\ y \cdot c = 0\}$. In this case, $|C| = 2^{n-\ell}$ and $\left|C^\perp\right| = 2^\ell$. C must not contain any vector of Hamming weight less than or equal to k, since we want $\hat{D}$ to be 0 on all such vectors. Hence, C is a linear error correcting code with minimum distance at least $k+1$. It is known that such codes exist with ℓ about $k \log n$, which gives $\left|C^\perp\right|$ about n^k.

In more detail, let $C[n-\ell, n]$ be the generator matrix for a linear error correcting code. (Matrix dimensions are given in brackets.) Then, given a message $v[1, n-\ell]$, the encoding of v is $(vC)[1, n]$.

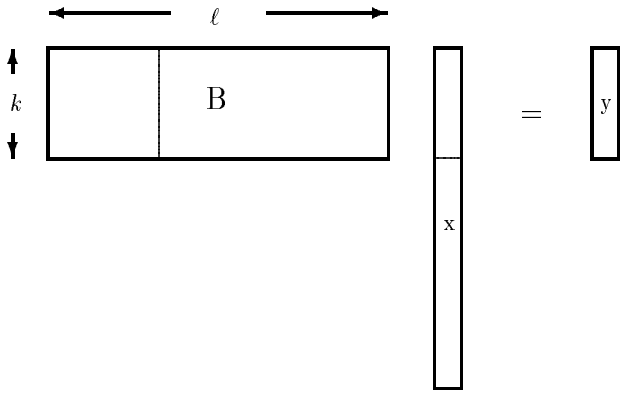


The parity check matrix $M[n, \ell] = (C^\perp)^T$ has the property that $C[n-\ell, n]M[n, \ell] = 0[n-\ell, \ell]$.



Hence, $wM = 0$ iff w is in the rowspace of C, i.e. w is a codeword. This means that any k rows of M must be linearly independent; otherwise, $wM = 0$ for a w with no more than k ones, contradicting the fact that every codeword has Hamming weight at least $k+1$.

Suppose we had a matrix $B[k, \ell]$ of rank k $(k \leq \ell)$. If we chose $x[\ell, 1]$ uniformly at random and computed $y = Bx$, y would also be uniformly random.



To see this, note that we can find k linearly independent columns in B, since it has rank k. Without loss of generality, let these be the first k columns. Then, for any way of fixing the latter $\ell - k$ bits of x, we get a bijection between the first k bits of x and the bits of y. Hence, choosing the first k bits of x uniformly guarantees that the bits of y will be chosen uniformly.

In our case, the parity check matrix M has the property that any k rows are independent. Hence, if we restrict our attention to k rows of M, we get a $k \times \ell$ matrix of rank k. Let $y = Mx$ with x chosen uniformly at random. Then, any k bits of y is the result of multiplying a $k \times \ell$ matrix of rank k by a uniformly chosen vector of length ℓ. By the argument above, such a set of k bits is uniformly random. Hence, y is k-wise independent.

3 An Alternate Construction

This construction, due to Alon, Babai, and Itai [ABI86], gives a k-wise independent distribution over a sample space of size about $n^{\lfloor k/2 \rfloor}$, saving a square root factor over the previous construction.

Assume $n = 2^r - 1$, and work over $GF(2^r)$. Let $a_1, \ldots, a_{2^r-1}$ be the nonzeroes of $GF(2^r)$. Then, the van der Monde matrix

$$M_1[n,k] = \begin{bmatrix} 1 & a_1 & \cdots & a_1^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & a_n & \cdots & a_n^{k-1} \end{bmatrix}$$

has the property that any k rows are independent. Now write the elements of $GF(2^r)$ in their standard representation as r 0/1 polynomial coefficients. This gives a matrix $M_2[n, kr]$ in which it is still true that any k rows are independent (since polynomial addition is componentwise). At this point, we again have a space of size n^k.

Note that we can delete the first $r-1$ columns of M_2, since these are all columns of zeros. To save the factor of 2 in the number of columns, we delete all even powers in M_1 (except the first column). Call this new matrix M_1'.

The claim is that any k rows of M_1' are still independent. To see this, suppose we had k linearly dependent rows in M_1'. As noted above, these rows must be independent in M_1; hence, there must be some even column in which the rows do not sum to 0, that is, $\sum_{i \in S} a_i^{2t} \neq 0$. But since the field has characteristic 2, $\sum a_i^{2t} = (\sum a_i^t)^2$, so $\sum a_i^t \neq 0$. By repeating the argument until t is odd, we get that if there is a nonzero even column, then there must be a nonzero odd column, contradicting the fact that these rows were supposed to be dependent in M_1'.

4 A Lower Bound

In this section, we give an essentially tight lower bound for the size of the sample space of a k-wise independent distribution. (See Chor et. al. [CFG+85].)

Claim 2 *Let D be a k-wise independent distribution on the non-constant random variables $x_1, \ldots, x_n$, over a sample space S. (D need not be uniform over S.) Then, $|S| = \Omega(n^{\lfloor k/2 \rfloor})$.*

In this claim, $x_1, \ldots, x_n$ need not be boolean. k-wise independence over general random variables is defined as follows:

$$\forall S = \{i_1, \ldots, i_k\},\ \forall a_1, \ldots, a_k,\ \Pr[\bigwedge x_{i_j} = a_j] = \prod \Pr[x_{i_j} = a_j].$$

Proof: Let

$$m(n,k) = \begin{cases} \sum_{j=0}^{k/2} \binom{n}{j} & \text{if } k \text{ even} \\ (\sum_{j=0}^{(k-1)/2} \binom{n}{j}) + \binom{n-1}{(k-1)/2} & \text{if } k \text{ odd} \end{cases}$$

When k is even, the function m counts the number of subsets of size no greater than $k/2$. When k is odd, m counts the number of subsets which contain no more than $(k-1)/2$ of the elements other than the first. Let J be the family of subsets counted by m. In either case, J has the property that the union of any two elements of J has size no greater than k. To prove the claim, we will show that $|S| \geq m(n,k)$.

Define the random variables $z_i = x_i - E[x_i]$. For any $T \subseteq [n]$, define $\alpha_T = \prod_{i \in T} z_i$. Then, for $|T| \leq k$,

$$E[\alpha_T^2] = E[\prod z_i^2] = \prod E[z_i^2] > 0,$$

since the x_i's are not constant. For $T \neq T'$, $|T \cup T'| \leq k$,

$$E[\alpha_T \alpha_{T'}] = (\prod_{i \in T \triangle T'} E[z_i])(\prod_{i \in T \cup T'} E[z_i^2]) = 0,$$

since $E[z_i] = 0$.

For any $T \subseteq [n]$, consider the $|S|$-dimensional vector $(\alpha_T(x))_{x \in S}$, with inner product $\alpha_T \cdot \alpha_{T'} = \sum_x D(x)\alpha_T(x)\alpha_{T'}(x) = E[\alpha_T \alpha_{T'}]$.

As noted above, the family of subsets J has the property that the union of any two elements of J has size no greater than k. Then, by the calculation above, for any $T, T' \in J$, $T \neq T'$, the vectors $(\alpha_T(x))$ and $(\alpha_{T'}(x))$ are orthogonal. Since there are m such mutually orthogonal vectors, each vector must have dimension at least m, so $|S| \geq m$. ■

5 ϵ-bias

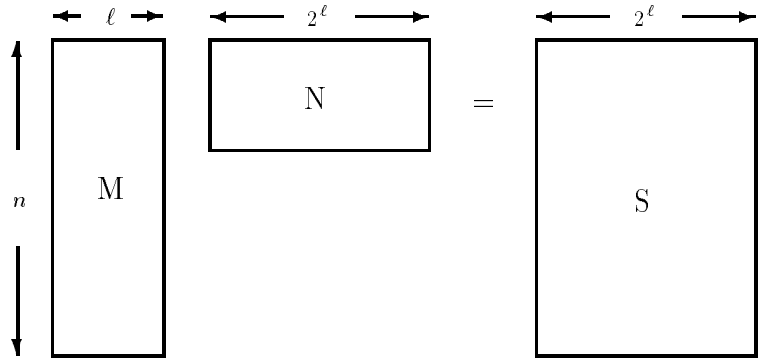
If we are willing to settle for a distribution which is *almost* k-wise independent, we can actually make do with a sample space which is exponentially smaller (in n) than the one needed for perfect k-wise independence. To make the notion of "almost" independence precise, we say that a distribution D on boolean random variables $x_1, \ldots, x_n$ is ϵ-biased k-wise independent if for all $S : |S| = k$ and all a as above,

$$\left| \Pr[x_{i_1} \cdots x_{i_k} = a_1 \cdots a_k] - \frac{1}{2^k} \right| \leq \epsilon.$$

As before, we can give an equivalent Fourier definition: D is ϵ-biased k-wise independent iff

$$\left| \hat{D}(y) \right| \leq \epsilon \ \ \forall\, |y| \leq k, \ y \neq 0.$$

Recall that our k-wise independent distribution from above was defined by a matrix $M[n, \ell]$, where $\ell = k \log n$. We can view the sample space generated by M as the columns of a matrix $S[n, 2^\ell] = M[n, \ell]N[\ell, 2^\ell]$, where the matrix N contains every column vector of length ℓ.



The k-wise independence property then states that if we sum any k rows of S, the resulting row vector has exactly half 0s and half 1s.

When we allow an ϵ-bias, we can replace N by an error correcting code $C[\ell, m]$, where $m = \ell/\epsilon^{O(1)}$, with minimum distance $m(1-\epsilon)/2$. Let $T \subseteq [n]$ be a set of size k, and let w_T be the characteristic vector of T. Then, summing the corresponding rows of $S[n, m]$ is equivalent to calculating $w_T S = w_T MC$.

As noted before, M has the property that $w_T M \neq 0$, since w_T has Hamming weight no greater than k. Hence, $(w_T M)C$ is a nonzero linear combination of rows of C. Since C has minimum distance $m(1-\epsilon)/2$, the number of 0s and the number of 1s in $w_T MC$ are each bounded by $m(1-\epsilon)/2$. Hence, $\left|\hat{D}(y)\right| = \left|\Pr[X_{S_y} = 0] - \Pr[X_{S_y} = 1]\right| \leq |(1-\epsilon)/2 - (1+\epsilon)/2| = \epsilon$.

The size of the sample space here is $m = k \log n/\epsilon^{O(1)}$. We are thinking in terms of $\epsilon \sim \frac{1}{2^k}$, so we get $m = O(2^k \log n)$. Relative to our earlier upper and lower bounds on the size of a k-wise independent sample space, this space is exponentially smaller in n. The preceding construction is due to Naor and Naor [NN90]. For further constructions, see also the paper by Alon et. al. [AGHP90].

References

[ABI86] N. Alon, L. Babai, and A. Itai, "A Fast and Simple Randomized Algorithm for the Maximal Independent Set Problem," *J. of Algorithms*, 1986, pp. 567-583.

[AGHP90] N. Alon, O. Goldreich, J. Hastad, R. Peralta, "Simple Constructions of Almost k-wise Independent Random Variables," *Proceedings of the 31st FOCS*, 1990, pp. 544-553.

[CFG+85] B. Chor, J. Freidmann, O. Goldreich, J. Hastad, S. Rudich, R. Smolensky, "The bit extraction problem and t-resilient functions," *Proceedings of the 26th FOCS*, 1985, pp. 396-407.

[NN90] J. Naor and M. Naor, "Small-bias Probability Spaces: Efficient Constructions and Applications," *Proceedings of the 22nd STOC*, 1990, pp. 213-223.