CS395T Pseudorandomness and Combinatorial Constructions
Lecture 7, February 6, 2001
Lecturer: Prof. David Zuckerman, Scribe: Peiyush Jain

1 Construction of ϵ-biased Random Variables

Notation: Let $m = \lceil \log_2 \frac{n}{\epsilon} \rceil$ and $\mathbb{F} = \mathbb{F}_{2^m}$.
Now we show how to construct a space of size $O(\frac{n^2}{\epsilon^2})$.

Choose $y,\ z$ uniformly at random from $\mathbb{F}$.
Define $X_i \ = \ y^i \cdot z \quad \pmod 2$, $\forall i$ in $\{0, 1, ..., n-1\}$.
Note that the '$\cdot$' is the dot product, viewing elements in $\mathbb{F}$ as vectors in $\{0,1\}^m$.
Since we are choosing two elements y, z at random from $\mathbb{F}$, the size of the probability space is $(2^m)^2 \ = \ O(\frac{n^2}{\epsilon^2})$.

Now we show that the X_i's are ϵ-biased.

Proposition 1 $X_0, X_1, ..., X_{n-1}$ *defined as above are* ϵ*-biased.*

Proof:
Consider:
$X_{i_1} \oplus X_{i_2} \oplus ... \oplus X_{i_l} \ = \ (y^{i_1} \cdot z) \oplus (y^{i_2} \cdot z) ... \oplus (y^{i_l} \cdot z) \ = (y^{i_1} + y^{i_2} + ... + y^{i_l}) \cdot z.$

Note that the '+' in the last expression is the same as the '+' of $\mathbb{F}$. This is because the '+' of $\mathbb{F}$ is the same as $\oplus$ when the terms are viewed as vectors in $\{0,1\}^m$.

Now, $y^{i_1} + y^{i_2} + ... + y^{i_l} = P(y)$ is a polynomial of degree less than n. So, $\Pr[P(y) = 0] < \frac{n}{2^m} \leq \epsilon$. Also note that for any fixed v in $\{0,1\}^m$, $v \neq 0$, $\Pr[v \cdot z = 1] = 1/2$, if z is chosen uniformly at random.

Therefore,
$\Pr[X_{i_1} \oplus X_{i_2} \oplus ... \oplus X_{i_l} = 1] = \Pr[P(y) \cdot z = 1] = \Pr[P(y) \neq 0] \Pr[P(y) \cdot z = 1] \geq (1-\epsilon).\frac{1}{2}.$

Also $\Pr[P(y) \neq 0] \Pr[P(y) \cdot z = 1] \leq 1/2$. Therefore,

$\frac{1}{2} \geq \Pr[X_{i_1} \oplus X_{i_2} \oplus ... \oplus X_{i_l} = 1] \geq \frac{1-\epsilon}{2}.$

Hence the X_i's are ϵ-biased.

$\square$

Recall from the previous lecture that ϵ-biased implies ϵ-close to k-wise independence in the max norm.

2 Construction of an $O(\frac{k^2 \log^2 n}{\epsilon^2})$ Space for Almost k-Wise Independence

Assume we have a $0-1$ matrix M with l rows and n columns such that any k columns of M are linearly independent over $\mathbb{F}_2$.(We show later how to obtain such a matrix,).

Choose a vector $X = [x_1\ x_2\ ...\ x_l]$, each $x_i \in \mathbb{F}_2$.

Let $XM = [Y_1\ Y_2\ ...\ Y_n]^T$.

Then we have the following proposition.

Proposition 2 *If X is chosen uniformly random, then the random variables Y_i's defined as above are k-wise independent.*

Proof:
Consider any $Y_{i_1}, Y_{i_2}, ..., Y_{i_k}$.
Let M' be the matrix formed by selecting the columns $i_1, i_2, ..., i_k$ of M.

We have that $XM' = [Y_{i_1}\ Y_{i_2}\ ...\ Y_{i_k}] = Y$ say.

Now for a given Y, the number of X such that $XM' = Y$ is the size of the kernel of the linear transformation M', which is equal to 2^{l-k}.

Thus, $\Pr[XM' = Y] = \frac{2^{l-k}}{2^l} = 2^{-k}$.

Also, for any given Y, there is an X such that $XM' = Y$.
This is because the k columns of M' are linearly independent.

Hence the Y_i's are k-wise independent.

$\square$

A similar result holds when X is ϵ-biased.

Proposition 3 *If X is ϵ-biased, then for any $i_1, i_2, ..., i_k$, $Y_{i_1}, Y_{i_2}, ..., Y_{i_k}$ is ϵ-biased.*

Proof:
Take $\{j_1, j_2, ..., j_s\} \subseteq \{i_1, i_2, ..., i_k\}$, where $s \leq k$.
Now,
$Y_{j_1} \oplus Y_{j_2} \oplus ... \oplus Y_{j_s} = X.(M[\cdot, j_1] \oplus ... \oplus M[\cdot, j_s]) = X.V$ say.
where $M[\cdot, t]$ is the t^{th} column of M.

Now V is non-zero as any s columns of M are linearly independent. Suppose $V = [v_1\ v_2\ ...\ v_l]^T$.
Then,
$XV = \bigoplus_{i:v_i \neq 0} X_i$.

Hence the Y_{i_k}'s are ϵ-biased. $\square$

Corollary 1 *$\{Y_1, Y_2, ..., Y_n\}$ as constructed above is ϵ-close to k-wise independence.*

3 Constructing the Matrix M

Recall that we assumed that any k columns of M were independent. In this section, we show how to construct such an M.

In order to construct M, we first construct a matrix M' with entries in a larger field $\mathbb{F}_q$, and then transform M' to give M.

Assume $q > n$.

Consider the polynomial $P(z) = p_0 + p_1 z + ... + p_{k-1} z^{k-1}$.

Let $\mathbb{F}_q = \{0, a_1, a_2, ..., a_{q-1}\}$.

We evaluate the polynomial $P(z)$ at each of a_i.
This we can write in matrix form as

$$\begin{bmatrix} p_0 & p_1 & \dots & p_{k-1} \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 \\ a_1 & a_2 & \dots & a_{q-1} \\ {a_1}^2 & {a_2}^2 & \dots & {a_{q-1}}^2 \\ \vdots & \vdots & \vdots & \vdots \\ {a_1}^{k-1} & {a_2}^{k-1} & \dots & {a_{q-1}}^{k-1} \end{bmatrix} = \begin{bmatrix} P(a_0) & P(a_1) & \dots & P(a_{q-1}) \end{bmatrix}$$

or simply,

$$\begin{bmatrix} p_0 & p_1 & \dots & p_{k-1} \end{bmatrix} M' = \begin{bmatrix} P(a_0) & P(a_1) & \dots & P(a_{q-1}) \end{bmatrix}$$

Taking any k-columns of M', we get a Vandermonde matrix, which is invertible. Alternately, we can see that the sub-matrix is invertible because a polynomial of degree $k-1$ is uniquely determined by k values taken by it. Hence the k columns of M' are linearly independent.

Now from M' we construct the matrix M whose elements are in $\mathbb{F}_2$.

Let q be a power of 2, say $q = 2^s$. Then we can view each element of $\mathbb{F}_q$ as a vector of length s over $\mathbb{F}_2$.

For each element ${a_i}^j$, $j \neq 0$, stick in M' the column corresponding to this element of $\mathbb{F}_q$. We now have a matrix M with $(k-1)s+1$ rows and $q-1$ columns, with elements in $\mathbb{F}_2$.

We now show that M is the matrix we want.

Proposition 4 *Any k columns of M as defined above, are linearly independent.*

Proof:
If some k columns of M are dependent(i.e there are some columns of M which sum to 0), then we have that

$\exists i_1, i_2, ..., i_t, t \leq k$

such that

$M[\cdot, i_1] + M[\cdot, i_2] + ... + M[\cdot, i_t] = 0.$

Now in $\mathbb{F}_q$ adding the vector representations is same as the addition of the elements. Thus, the corresponding columns in M' are dependent.

Hence any k columns of M are linearly independent.

$\square$

We want $q - 1 = n$, i.e $s \approx \log n$.

so M has about $k \log n$ rows and n columns.
It can be shown that we can reduce the number of rows further to $\frac{k \log n}{2}$ by considering only the odd rows of M'.

We deduce:

Theorem 1 *There is a construction of a probability space of size $O(\frac{k^2 \log^2 n}{\epsilon^2})$ for n almost k-wise independent random variables with ϵ-error in max norm.*