Lecture 7: ϵ-biased and almost k-wise independent spaces

Topics in Complexity Theory and Pseudorandomness (Spring 2013)
Rutgers University
Swastik Kopparty
Scribes: Ben Lund, Tim Naumovitz

Today we will see ϵ-biased spaces, almost k-wise independent spaces, and some applications.

1 Distributions

Let μ be a distribution on $\{0,1\}^n$.

Lemma 1. *μ is uniformly distributed on $\{0,1\}^n$ iff for each nonempty $S \subseteq [n]$*

$$\Pr_{x\in\mu}[\sum_{i\in S} x_i = 1] = \frac{1}{2}$$

(where the addition is mod 2).

Proof. Let χ_S: $\{0,1\}^n \to \mathbb{R}$ where

$$\chi_S(x) = (-1)^{\sum_{i\in S} x_i}.$$

Thus: $\chi_S(x) = \begin{cases} 1 & \text{if } \sum_{i\in S} x_i = 0 \\ -1 & \text{if } \sum_{i\in S} x_i = 1. \end{cases}$

We can treat μ as a real valued function μ: $\{0,1\}^n \to \mathbb{R}$ with $\mu(x) \geq 0$ for all $x \in \{0,1\}^n$ and $\sum_{x\in\{0,1\}^n} \mu(x) = 1$.

For nonempty S, consider the inner product:

$$\begin{aligned}
\langle \mu, \chi_S \rangle &= \sum_{x\in\{0,1\}^n} \mu(x)\chi_S(x) \\
&= \sum_{x\in\{0,1\}^n, \chi_S(x)=1} \mu(x) - \sum_{x\in\{0,1\}^n, \chi_S(x)=0} \mu(x) \\
&= \Pr_{x\in\mu}[\sum_{i\in S} x_i = 0] - \Pr_{x\in\mu}[\sum_{i\in S} x_i = 1] = 0 \text{ (by our hypothesis)}
\end{aligned}$$

Aside: Characters and Fourier analysis

The χ_S are called the characters of the group $(\mathbb{Z}_2^n, +)$.

We have the following important properties:

1. $\chi_S(x+y) = \chi_S(x) * \chi_S(y)$

2. For any nonempty S, we have $\sum_{x\in\{0,1\}^n} \chi_S(x) = 0$. This follows from the following calculation. Pick any $i \in S$. Then:

$$\sum_{x\in\{0,1\}^n} \chi_S(x) = -\sum_{x\in\{0,1\}^n} \chi_S(x+e_i) = -\sum_{x\in\{0,1\}^n} \chi_S(x),$$

   where e_i is the 0-1 vector with 1 only in the ith coordinate.

3. Furthermore, $\chi_S(x) * \chi_T(x) = (-1)^{\sum_{i\in S} x_i + \sum_{i\in T} x_i} = (-1)^{\sum_{i\in S\triangle T}} = \chi_{S\triangle T}(x)$. Sometimes we will also treat S, T as their characteristic vectors in $\mathbb{Z}_2^n$, and in this notation, $\chi_S(x)\cdot\chi_T(x) = \chi_{S+T}(x)$.

4. $\langle \chi_S, \chi_T\rangle = \sum_{x\in\{0,1\}^n} \chi_S(x)\chi_T(x) = \sum_{x\in\{0,1\}^n} \chi_{S\triangle T}(x) = \begin{cases} 0 & \text{if } S \neq T \neq 0 \\ 2^n & \text{if } S = T. \end{cases}$

   This means that $\{x_S\}$ are an orthogonal system, and since there are 2^n of them, they are a basis for $\mathbb{R}^{\{0,1\}^n}$.

Let $\hat{\mu}(S) = \langle \mu, \chi_S\rangle$: the S^{th} Fourier coefficient of μ.

End Aside

Proof (continued) We have $\hat{\mu}(S) = 0$ for each nonempty S. Thus $\mu = \alpha * \chi_\emptyset$ for some α, and since μ is a distribution, we have $\mu = \frac{1}{2}^n \cdot \chi_\emptyset$, and this is the uniform distribution. ☐

Lemma 2. *Suppose that for all nonempty $S \subseteq [n]$,*

$$\Pr_{x\in\mu}\left[\sum_{i\in S} x_i = 1\right] \in ((1-\epsilon)/2, (1+\epsilon)/2).$$

Then μ is ϵ close to the uniform distribution in L_2, and thus μ is $\epsilon 2^{n/2}$ close the the uniform distribution in L_1 and μ is ϵ close the the uniform distribution in L_∞.

Proof. The hypothesis gives us that $|\hat{\mu}(S)| \leq \epsilon$
We also have $\hat{\mu}(\emptyset) = \langle \mu, \chi_\emptyset\rangle = 1$

Aside: Parseval's Identity

Lemma 3. *Take any $f : \{0,1\}^n \to \mathbb{R}$. We have*

$$f = \frac{1}{2^n}\sum_S \langle f, \chi_S\rangle \cdot \chi_S = \frac{1}{2^n}\sum_S \hat{f}(S)\cdot\chi_S.$$

Then

$$\sum_x f(x)^2 = \frac{1}{2^n}\sum_S \hat{f}(S)^2$$

Proof. This follows from the orthogonality of the χ_S:

$$\begin{aligned}\sum_x f(x)^2 &= \langle f, f\rangle \\ &= \frac{1}{4^n}\langle \sum_S \hat{f}(S)\chi_S, \sum_T \hat{f}(T)\chi_T\rangle \\ &= \frac{1}{4^n}\sum_{S,T} \hat{f}(S)\hat{f}(T)\langle \chi_S, \chi_T\rangle \\ &= \frac{1}{2^n}\sum_S \hat{f}(S)^2.\end{aligned}$$

$\square$

<u>*End Aside*</u>
Proof (cont'd): Let $U = \chi_\emptyset * \frac{1}{2^n}$ be the uniform distribution.
We have $\hat{U}(\emptyset) = 1$, and $\hat{U}(S) = 0$ for all nonempty S. Thus $\mu - U$ (which is what we want to show is small in the L_2 norm), has the following Fourier coefficients:

$$\widehat{\mu - U}(S) = \begin{cases} 0 & \text{if } S = \emptyset \\ \hat{\mu}(S) & \text{otherwise} \end{cases}$$

Now using the Parseval identity, we get that

$$\|\mu - U\|_2^2 = \frac{1}{2^n}\sum_{S\neq\emptyset} \hat{\mu}(S)^2 \le \frac{1}{2^n}\epsilon^2(2^n - 1) \le \epsilon^2.$$

This completes the proof, and the result for L_1 and L_∞ follow from Cauchy-Schwarz and trivially.

$\square$

Thus if a distribution fools linear functions really well, it is almost uniform.

1.1 Notes on the L_1 distance

There are many distances between probability distributions. But the L_1 distance has special status when we are interested in pseudorandomness.

Lemma 4 (Data processing inequality)**.** *Suppose μ, ν are distributions over the same domain D. Let f be a function defined on D. Pick $x \in \mu$, $y \in \nu$.*

$$\|f(x) - f(y)\|_{L_1} \le \|\mu - \nu|_{L_1}.$$

This is closely related to the following simple characterization of L_1 distance:
$\frac{1}{2}||\mu - \nu||_{L_1} \ge \epsilon$ if and only if there exists a distinguishing test $T : D \to \{0, 1\}$ such that

$$\|T(\mu) - T(\nu)\|_{L_1} = \Pr_{x\in\mu}[T(x) = 1] - \Pr_{x\in\nu}[T(x) = 1] \ge \epsilon.$$

Sketch of Proof: Graph the distributions μ and ν over their domain (They have the same domain). $\frac{1}{2}||\mu - \nu||_{L_1}$ is the area between the curves where μ is above ν. Choose T to output 1 on sets where $\mu > \nu$. This gives us $\Pr_{x\in\mu}[D(x) = 1] - \Pr_{x\in\nu}[D(x) = 1]$ is the area between μ and ν on these sets, which is $\frac{1}{2}||\mu - \nu|| \ge \epsilon$.

Relationship to pseudorandom generators When we studied PRGs, the goal was to find a simple μ s.t. $\forall$ small circuits C,

$$\Pr_{x\in\mu}[C(x)=1] - \Pr_{x\in U}[C(x)=1] \le \epsilon.$$

The only difference between this condition and the condition for small L_1 distance is the complexity constraint that C is small. *THIS MAKES ALL THE DIFFERENCE IN THE WORLD!*

By the above discussion, we could try to show that μ is a PRG by showing the stronger condition that μ is ϵ-close to the uniform distribution in L_1. This strengthening ruins the approach: there cannot be a μ which is generated using a small seed that is close to the uniform distribution in L_1 distance:

$$\|\mu - U\|_{L_1} \le \epsilon \Rightarrow \mathsf{support}(\mu) \ge (1-\epsilon)\cdot 2^n.$$

Try to show this.

2 ϵ-biased distributions and k-wise independence

Definition 5. *μ is ϵ-biased if for all nonempty $S \subseteq [n]$,*

$$\Pr[\sum_{i\in S} x_i = 1] \in [(1-\epsilon)/2, (1+\epsilon)/2].$$

Note:

1. μ is 0-biased $\iff$ μ is uniform.
2. ϵ-biased $\Rightarrow$ μ is $(\epsilon, \epsilon 2^{n/2})$-close to uniform in (L_2, L_1).
3. ϵ-biased $\Leftarrow$ μ is ϵ-close to uniform in L_1.

When is μ k-wise independent?

μ is k-wise independent $\iff \forall S\ 1 \le |S| \le k$, we have $\hat{\mu}(S) = 0$.

Proof: ($\Rightarrow$) Take such an S. Since μ is k-wise independent, we have $\Pr_{x\in\mu}[\sum_{i\in S} x_i = 1] = \frac{1}{2}$. By definition of $\hat{\mu}$, this yields $\hat{\mu}(S) = 0$.
($\Leftarrow$) Let S be as stated. Look at $\mu|_S$. $\forall T \subseteq S$, $T \ne \emptyset$, we know that $\Pr_{x\in\mu|_S}[\sum_{i\in S} x_i = 1] = \frac{1}{2}$, so $\widehat{\mu|}_S(T) = 0$, meaning $\mu|_S$ is uniform. This implies that μ is k-wise independent. □

When is δ-almost μ k-wise independent?

Suppose $\forall S \subseteq [n]$, $|S| \le k$, $S \ne \emptyset$ we have $|\hat{\mu}(S)| \le \epsilon$, then μ is δ-almost k-wise independent in L_1 for $\delta = \epsilon 2^{k/2}$.

Take any $S \subset [n]$ $|S| = k$. Look at $\nu = \mu|_S$, a distribution on $\{0,1\}^S$. $\forall T \subseteq S$, since $\hat{\nu}(T) = \hat{\mu}(T)$, we have $|\hat{\nu}(T)| \le \epsilon$. This implies that ν is $\epsilon \cdot 2^{k/2}$ close to uniform on $\{0,1\}^S$.

How much randomness is needed to generate these spaces?

For k-wise independence: we saw that $k(\log n)$ bits suffices. In fact, $\Omega(k \log n)$ bits are needed (you will prove this in the homework).

For ϵ-biased spaces: First let us show that there exist "simple" ϵ-biased spaces; we will later see how to get them explicitly. We try to get an ϵ-biased μ which is uniform on some $K \subseteq \{0,1\}^n$ with $|K|$ small. Then $\log |K|$ bits suffice to generate a sample from μ.

We use the probabilistic method: Choose K at random as follows: pick $y_1, \ldots, y_m$ in $\{0,1\}^n$ uniformly. We want that for all linear functions $h : \mathbb{Z}_2^n \to \mathbb{Z}_2$,

$$\Pr_{i \in [m]}[h(y_i) = 1] \in ((1-\epsilon)/2, (1+\epsilon)/2).$$

Fix h. Then:

$$\Pr_{y_1, \cdots, y_m}[y_1, \ldots, y_m \text{ are bad for h }] \leq e^{-\Omega(\epsilon^2 m)},$$

by a Chernoff bound (This is because for a random $y \in \mathbb{Z}_2^n$, $\Pr[h(y) = 1] = \frac{1}{2}$).

We then union bound over all h to get

$$\Pr_{y_1, \cdots, y_m}[\exists h : y_1, \ldots, y_m \text{ are bad for } h] \leq 2^n \cdot e^{-\Omega(\epsilon^2 m)}.$$

Now choose $m = O(\frac{n}{\epsilon^2})$, so that this is less than 1. We thus get an ϵ-biased space which can be generated using $\log n + 2 \log \frac{1}{\epsilon} + O(1)$ bits of true randomness. It turns out that this seed length is near optimal. Later this class we will explicitly construct ϵ-biased spaces with seed length $O(\log n + \log \frac{1}{\epsilon})$.

For δ-almost k-wise independent spaces: We know that ϵ-biased spaces are automatically δ-almost k-wise independent for some δ. It turns out that this already gives us almost k-wise independent spaces using smaller seed length than what is needed for pure k-wise independence. Indeed, if we take $\epsilon = 2^{-k/2} \cdot \delta$ and take an explicit ϵ biased space as mentioned above, then this is δ-almost k-wise independent, and has a seed length of $O(\log n + \log \frac{1}{\epsilon}) = O(\log n + k + \log \frac{1}{\delta})$.

3 Efficient construction of δ-almost k-wise independent spaces

One way to get k-wise independence is to multiply a random seed y by a matrix M:

$$y \mapsto y^T M.$$

In this construction, y is a vector with $O(k \log n)$ elements chosen uniformly at random, and M has n columns. Each element of $y^T M$ is $\langle y, a_i \rangle$, where a_i is a column of M.

Claim: The output $y^T M$ will be k-wise indpendent if and only if every k rows of M are indepenent.

Proof: Let $S \subset [n]$. Suppose $\sum_{i \in S} \langle a_i, y \rangle$ is not uniformly distributed. This is equivalent to $\langle y, \sum_{i \in S} a_i \rangle$ is not uniformly distributed. But, this implies that $\sum_{i \in S} a_i = 0$, since $\langle y, b \rangle$ is uniform for any fixed $b \neq 0$.

Claim: If, instead of taking y from a uniformly random distribution, we take y from an ϵ-biased distribution, $y^T M$ will still be $\epsilon 2^{k/2}$-almost k-wise independent.

Proof: Suppose not; then there exists $S \in [n], S \neq \emptyset, |S| \leq k$ such that bias $\left(\langle \sum_{i \in S} a_i, y \rangle\right) \geq \epsilon$. Since each set of k columns of M are independent, we know that $\sum_{i \in S} a_i \neq 0$. In addition, y is chosen from an ϵ-biased space. So, we've reached a contradiction.

How many bits of randomness will we need to generate an n bit sample from a δ-almost k-wise independent space using this procedure? We need a $\delta 2^{-k/2}$-biased sample of length $O(k \log n)$. Since $\log m + 2\log(\frac{1}{\epsilon})$ bits of uniform randomness are needed for m bits of ϵ-biased randomness, we need $O(\log k + \log\log n) + O(\log(\frac{1}{\delta}) + k) = O(k + \log\log n + \log(\frac{1}{\delta}))$ total bits of randomness.

4 Applications of δ-almost k-wise independent distribution

4.1 k-universal sets

A k universal set $S \subseteq \{0,1\}^n$ has the property that the projection of S onto any k indexes contains all 2^k possible patterns. We can use δ-almost k-wise independent distribution to construct k-universal sets. If $\delta = \frac{1}{10} \cdot \frac{1}{2^k}$, then any δ-almost k-wise independent distribution has k-universal support.

The size of the k-universal set we get out of this is $2^{O(k + \log\log n + \log(\frac{1}{\delta}))} = \left(2^k \cdot \log n\right)^{O(1)}$, and is nearly optimal. (Note the surprisingly tiny dependence on n!)

4.2 Ramsey graphs

Pick the edges $(x_{ij})_{i<j} \in \{0,1\}^{\binom{n}{2}}$ from a δ-almost k-wise independent space; we can interpret $x_{ij} = 1$ as an edge between vertex i and vertex j, and $x_{ij} = 0$ as the absence of an edge. Fix $S \in [n], |S| = k$. By the data processing inequality,

$$\Pr[x_{ij} \text{ are all 0 or all 1 for all } i, j \in S] \leq 2 \cdot 2^{-\binom{k}{2}} + \delta.$$

Taking a union bound,

$$\Pr[\exists S, |S| = k, \text{ such that } S \text{ is a clique or independent set}] \leq \binom{n}{k}(2 \cdot 2^{-\binom{k}{2}} + \delta)$$

By setting $\delta = 2^{-\binom{k}{2}}$ and $n = 2^{k/10}$ in the above inequality, we ensure that the probability that there is a clique or independent set of size k is less than 1. Thus, we've described an explicit family of $2^{O(\log^2 n)}$ graphs on n vertexes, at least one of which is $O(\log n)$ Ramsey. As a corollary, we can construct an $O(\log n)$ Ramsey graph in $2^{O(\log^2 n)}$ time.

5 Construction of ϵ-biased spaces

5.1 Finite extension field review

The construction described here will use the finite field $\mathbb{F}_{2^n}$. This is an n-dimensional vector space over $\mathbb{F}_2$. Addition is the same as for $\mathbb{F}_2^n$; multiplication is a bilinear map. A polynomial $P(x) \in \mathbb{F}_{2^n}[x]$ of degree d has at most d roots.

5.2 Properties needed from an ϵ-biased space

Suppose $y_1 \dots y_m$ is an ϵ-biased space; let G be the n by m matrix with columns $y_1 \dots y_m$. Pick $x \in \{0,1\}^n$. Consider $x^T G$. If $x \neq 0$, it must have $((1-\epsilon)/2, (1+\epsilon)/2)$ fraction of 1s. Pick any $x, y \in \{0,1\}^n$, and consider the Hamming distance $\Delta(x^T G, y^T G)$; this is the number of 1s in $x^T G - y^T G = (x-y)^T G$, which is in the range $((1-\epsilon)/2, (1+\epsilon)/2)$. Thus, the image of $x^T G$ is a linear space in $\{0,1\}^m$ of dimension n, such that any two vectors in the space have distance of $\frac{1}{2} \pm \epsilon/2$.

5.3 Construction and proof of correctness

A point from the ϵ-biased space is calculated from two uniformly chosen elements $\alpha, \beta \in \mathbb{F}_{2^n}$. The point is calculated as

$$(\alpha, \beta) \mapsto [\langle 1, \beta \rangle, \langle \alpha, \beta \rangle, \dots \langle \alpha^N, \beta \rangle].$$

where $\langle \cdot, \cdot \rangle$ is the inner product over $\mathbb{F}_2^n$, when elements of $\mathbb{F}_{2^n}$ are represented in some basis over $\mathbb{F}_2$.

If N is set to $\epsilon 2^n$, then this construction needs $2\log(N/\epsilon) = 2\log N + 2\log(\frac{1}{\epsilon})$ bits of randomness.

We need to show that the sample space described is ϵ-biased. Pick (α, β) uniformly from $\mathbb{F}_{2^n}$, and take any $S \subseteq [N]$. We need to show that

$$\Pr_{\alpha,\beta}\left[\sum_{i \in S} \langle \alpha^i, \beta \rangle = 0\right] \in ((1-\epsilon)/2, (1+\epsilon)/2).$$

We have $\sum_{i \in S} \langle \alpha^i \beta \rangle = \langle \sum_{i \in S} \alpha^i, \beta \rangle$. There are two cases to consider; either α is a root of $P(x) = \sum_{i \in S} x^i$, or it is not. If α is not a root of $P(x)$, then $\Pr_\beta\left[\langle \sum_{i \in S} \alpha^i, \beta \rangle = 0\right] = \frac{1}{2}$. If α is a root of $P(x)$,then certainly $\langle \sum_{i \in S} \alpha^i, \beta \rangle = 0$. However, there are at most N roots of $P(x)$, so the probability that α is a root is at most $\frac{N}{2^n} = \epsilon$. Thus, the space is ϵ-biased.