# Lecture 05: K-wise Independence and $\epsilon$- Bias

Topics in Complexity Theory and Pseudorandomness (Spring 2018)
Rutgers University
Swastik Kopparty
Scribes: Rahul Ilango and Neelesh Kumar

## 1 Review: $k$-wise independent

**Definition 1.** *A random variable $(x_1, \ldots, x_n) \in \Sigma^n$ is called **$k$-wise independent** if for all $S \subseteq [n]$ with $|S| \leq k$, we have that $(x_j)_{j \in S}$ are independent and uniformly distributed.*

Last class, we proved the following theorem.

**Theorem 2.** *We can sample an $n$ bit 2-wise independent random variable using $\log_2 n$ independent uniformly random bits.*

The construction is as follows. Let $t = \log_2 n$. Pick $y_1, \ldots, y_t$ uniformly independently at random. Output $(\bigoplus_{i \in V} y_i)$ where $V$ ranges over all nonempty subsets of $[t]$.

## 2 Application: derandomizing a MaxCut approximation algorithm

Given a graph $G$ on a vertex set $V$, **MaxCut** is the problem of finding a partition of $V$ so that the number of edges cut is as large as possible.

MaxCut has a simple random approximation algorithm. Given a graph $G = (V, E)$, output a random partition $(R, B)$ of $V$. Specifically, for each vertex $v \in V$, put $v$ in $R$ or $B$ with probability $\frac{1}{2}$.

**Theorem 3.** *The expected number of edges cut by this algorithm is at least $\frac{1}{2}|E|$, that is*

$$\mathbf{E}[\textit{number of edges cut}] \geq \frac{1}{2}|E|.$$

*Proof.* For an edge $e \in E$, define the random variable

$$Z_e = \begin{cases} 1, & \text{if } e \text{ is cut} \\ 0, & \text{if } e \text{ is not cut} \end{cases}$$

Then, we have

$$\mathbf{E}[\text{number of edges cut}] = \mathbf{E}[\sum_{e \in E} Z_e].$$

By linearity of expectation, we have

$$\mathbf{E}[\sum_{e \in E} Z_e] = \sum_{e \in E} \mathbf{E}[Z_e].$$

Fix some edge $e \in E$ which goes from vertex $a$ to vertex $b$. Recall, we partition $V$ into two sets $R$ and $B$. Since $a$ and $b$ are put into $R$ or $B$ independently with probability $\frac{1}{2}$, we have that

$$Pr[e \text{ is cut}] = Pr[a \in R \text{ and } b \in B] + Pr[a \in B \text{ and } b \in R] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

and

$$Pr[e \text{ is not cut}] = Pr[a, b \in R] + Pr[a, b \in B] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Thus

$$\mathbf{E}[Z_e] = \frac{1}{2}.$$

Hence we have

$$\mathbf{E}[\text{number of edges cut}] = \sum_{e \in E} \mathbf{E}[Z_e] = \sum_{e \in E} \frac{1}{2} = \frac{|E|}{2}$$

□

We want to minimize the amount of randomness we use in our algorithm. Currently, our algorithm requires deciding for each vertex $v \in V$ whether $v \in R$ or $v \in B$. Letting $n = |V|$, this means we need a $n$-bits of randomness. In the proof above, however, the only time we use the randomness of the bits is when we show that

$$Pr[e \text{ is cut}] = Pr[e \text{ is not cut}] = \frac{1}{2}.$$

Proving this only relied on the fact that for an edge $e$ from $a$ to $b$, the two bits that determine if $a$ and $b$ are in $R$ or $B$ are uniform and independent. Thus, if our algorithm uses $n$ bits which are 2-wise independent, we have the same guarantee that in expectation $\frac{|E|}{2}$ edges are cut.

Thus, we have a new algorithm. Given a graph $G$ with $n$ vertices, generate $n$ 2-wise independent random bits $x_1, ..., x_n$, and use them to partition $G$. By Theorem 2, this algorithm requires only $\log_2 n$ bits of randomness, and by the same proof as before, we have that the expected number of edges cut is at least $\frac{|E|}{2}$.

Finally, we are ready to make this algorithm deterministic. Since we use only $\log_2 n$ bits of randomness, there are only $2^{\log_2 n} = n$ possibilities for $x_1, \ldots, x_n$. Try all of them, and in each case find the number of edges cut. Return the partition which cuts the most edges. Since the expected number of edges cut for the previous algorithm is at least $\frac{|E|}{2}$. There is at least one partition which cuts at least $\frac{E}{2}$ edges, so this deterministic algorithm returns a partition which cuts at least $\frac{E}{2}$ edges.

## 3 Finite Fields

**Fact 4.** *Fix a prime $p$. The $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$ with the operations $+$, addition mod $p$, and $\times$ multiplication mod $p$ is a field.*

A field $\mathbb{F}$ satisfies some properties:

- It has some elements named 0 and 1

- $0 \times a = 0$ for any $a \in \mathbb{F}$
- $0 + a = a$ for any $a \in \mathbb{F}$
- $1 \times a$ for any $a \in \mathbb{F}$
- $a \times b = 0 \implies a = 0$ or $b = 0$
- If $a \in \mathbb{F}$ and $a \neq 0$, then there exists a $b \in \mathbb{F}$ such that $a \times b = 1$

**Fact 5.** *$\mathbb{F}_p[x] = \{$polynomials with coefficients in $\mathbb{F}_p\}$ is a field.*

**Theorem 6** (Division Theorem for $\mathbb{F}_p[x]$)**.** *For all $a(x), b(x) \in \mathbb{F}_p[x]$ there exists $q(x), r(x) \in \mathbb{F}_p[x]$ such that*

$$a(x) = b(x)q(x) + r(x)$$

*where degree(r) < degree(b).*

**Theorem 7** (Remainder Theorem for $\mathbb{F}_p[x]$)**.** *Suppose $p(x) \in \mathbb{F}_p[x]$ and $a \in \mathbb{F}_p$. Then $p(a) = 0$ if and only if $(x - a)$ divides $p(x)$.*

**Theorem 8** (Unique factorization for $\mathbb{F}_p[x]$)**.** *Every polynomial in $\mathbb{F}_p[x]$ is a product of a unique set of irreducible polynomials (polynomials which cannot factor).*

**Theorem 9** (At most degree zeroes for $\mathbb{F}_p[x]$)**.** *Suppose $p(x) \in \mathbb{F}_p[x]$ has degree $d$. Then*

$$|\{a \in \mathbb{F}_p : p(a) = 0\}| \leq d > 0.$$

**Fact 10.** *For any prime $p$ and $m \geq$, there is a set $\mathbb{F}_{p^m}$ with operations $+, \times$ such that*

- $|\mathbb{F}_{p^m}| = p^m$
- *$\mathbb{F}_{p^m}$ is a field.*

## 4 Constructing a $k$-wise independent random variable

We describe an algorithm to sample a $k$-wise independent random variable. Fix $\mathbb{F}$ be a finite field, and fix $n$ distinct elements $a_1, \ldots, a_n \in \mathbb{F}$.

Now, we start to use randomness. Pick $b_0, b_1, \ldots, b_d \in \mathbb{F}$ uniformly at random. Define

$$Q(x) = b_0 + b_1 x + \cdots + b_d x^d.$$

**Theorem 11.** *$(Q(a_1), \ldots, Q(a_n))$ are $d + 1$-wise independent.*

*Proof.* To gain some intuition, we first consider the $d = 1$ case. We must show that the random variables are uniform and independent.

First, we show that $Q(a_i)$ is uniform over $\mathbb{F}$ for an arbitrary $i$, that is for every $c \in \mathbb{F}$

$$Pr[Q(a_i) = c] = \frac{1}{|\mathbb{F}|}.$$

Fix $i$ and $c$. For a uniformly random $b_0, b_1 \in \mathbb{F}$, we have

$$Pr[Q(a_i) = c] = Pr[b_0 + b_1 a_i = c] = Pr[b_0 = c - b1a_i].$$

Since $b_0$ is chosen uniformly at random,

$$Pr[b_0 = c - b1a_i] = \frac{1}{\mathbb{F}}$$

as desired.

Next, we must show independence, that is for $i \neq j$ and any $c_i, c_j \in \mathbb{F}$, we must show that

$$Pr[Q(a_i) = c_i \wedge Q(a_j) = c_j] = \frac{1}{|\mathbb{F}|^2}.$$

Fix $i, j, c_i$, and $c_j$.

$$Pr[Q(a_i) = c_i \wedge Q(a_j) = c_j] = Pr[b_0 + b_1 a_i = c_i \wedge b_0 + b_1 a_j = c_j]$$

If $b_0 + b_1 a_i = c_i$ and $b_0 + b_1 a_j = c_j$, we have that (since $a_i, a_j$ are distinct)

$$b_1 = \frac{c_i - c_j}{a_i - a_j}, \qquad b_0 = c_i - a_i \frac{c_i - c_j}{a_i - a_j}.$$

Since $a_i, a_j, c_i$, and $c_j$ are all fixed, and $b_0$ and $b_1$ are picked uniformly at random from $\mathbb{F}$, this means that

$$Pr[Q(a_i) = c_i \wedge Q(a_j) = c_j] = Pr[b_0 + b_1 a_i = c_i \wedge b_0 + b_1 a_j = c_j] = \frac{1}{|\mathbb{F}|^2}$$

as desired.

Now, we consider the case for general $d > 1$. A similar proof to $d = 1$ case shows that $Q(a_i)$ is uniform over $\mathbb{F}$ for all $i$. It remains to show the independence portion. Let $S \subset [n]$. We must show that for any choice of $|S|$ elements $c_j \in \mathbb{F}$ indexed by $j \in S$

$$Pr[Q(a_j) = c_j \text{ for all } j \in S] = \frac{1}{|\mathbb{F}|^{|S|}}.$$

It suffices to prove the case when $|S| = d + 1$, because if independence holds for all $S'$ of size $d + 1$ and $S < d + 1$, then we have

$$\begin{aligned} Pr[Q(a_j) = c_j \text{ for all } j \in S] &= \sum_{\substack{c_j \in \mathbb{F} \\ \text{indexed by} \\ j \in (S' \setminus S)}} Pr[Q(a_j) = c_j \text{ for all } j \in S'] \\ &= \sum_{\substack{c_j \in \mathbb{F} \\ \text{indexed by} \\ j \in (S' - S)}} \frac{1}{|\mathbb{F}|^{d+1}} \\ &= \frac{1}{|\mathbb{F}|^{|S|}} \end{aligned}$$

where we pick some $S' \subseteq [n]$ with $S \subseteq S'$ and $|S'| = d+1$. Thus is suffices to consider $S$ of size $d+1$.

Fix $S \subseteq [n]$ of size $d+1$. And fix $|S| = d+1$ elements $(c_j)_{j \in S}$ of $\mathbb{F}$.

$$Pr[Q(a_j) = c_j \text{ for all } j \in S] = \frac{\text{number of polynomials } Q(x) \text{ such that } Q(a_j) = c_j \text{ for all } j \in S}{|\mathbb{F}|^{d+1}}.$$

Since

$$\prod_{j \in S} Pr[Q(a_j) = c_j] = \frac{1}{\mathbb{F}^{d+1}},$$

we must show that there is exactly one such $Q(x)$, which follows from the Fact 12. □

**Fact 12.** *Let $\mathbb{F}$ be a field. Let $e_1, \ldots, e_{d+1} \in \mathbb{F}$ all be distinct. Let $c_1, \ldots, c_{d+1} \in \mathbb{F}$. There is exactly one polynomial $Q(x) \in \mathbb{F}[x]$ of degree at most $d$ such that $Q(e_i) = c_i$ for $i = 1, \ldots, d+1$.*

*Proof.* Proof of uniqueness is simple. Suppose $Q(x), R(x) \in \mathbb{F}[x]$ are polynomials of degree at most $d$ with $Q(e_i) = R(e_i) = c_i$ for all $i$. Let $S(x) = Q(x) - R(x)$. Then $S(x)$ is a polynomial of degree at most $d$, and

$$S(e_i) = Q(e_i) - R(e_i) = 0.$$

Thus $S(x)$ is a polynomial of degree at most $d$ with $d+1$ zeroes, so by Theorem 9, $S(x) = 0$, so $Q(x) = R(x)$.

Existence is a bit tougher. One can construct a polynomial explicitly using Lagrange polynomials, which we will not describe here, but one can look up.

Alternatively, we can do a linear algebra proof. One can view a polynomial evaluation as a matrix evaluation. Let $f(x) = \sum_{i=0}^{d} b_i x^i$. Then

$$F(x) = \begin{pmatrix} x^d & x^{d-1} & \ldots & x & 1 \end{pmatrix} \begin{pmatrix} b_d \\ b_{d-1} \\ \vdots \\ b_1 \\ b_0 \end{pmatrix}$$

Thus, finding a polynomial $Q(x) = \sum_{i=0}^{d} (b_i x^i)$ with $Q(e_i) = c_i$ for all $i = 1, \ldots, d+1$. is equivalent to finding a vector $(b_0, \ldots, b_d)$ such that

$$\begin{pmatrix} e_1^d & e_1^{d-1} & \ldots & e_1 & 1 \\ e_2^d & e_2^{d-1} & \ldots & e_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e_d^d & e_d^{d-1} & \ldots & e_d & 1 \\ e_{d+1}^d & e_{d+1}^{d-1} & \ldots & e_{d+1} & 1 \end{pmatrix} \begin{pmatrix} f_d \\ f_{d-1} \\ \vdots \\ f_1 \\ f_0 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_d \\ c_{d+1} \end{pmatrix}$$

Let $E$ be the $d+1 \times d+1$ matrix above. Showing that there is exactly one such polynomial is equivalent to showing that $E$ is invertible. $E$ is matrix of a special type, a Vandermonde matrix, and has a special formula for its determinant

$$det(E) = \prod_{1 \leq i < j \leq n} (e_j - e_i)$$

which is non-zero because the $e_i$ are distinct, so $E$ is invertible. □

Thus, letting $k = d + 1$ and $n = |\mathbb{F}|$, we can get $n$ $\mathbb{F}$-valued $k$-wise independent random variable using $k \log_2 n$ bits of randomness.

In fact, we cannot hope to do better than with $k \log n$ bits of randomness. The question now is how can we use the above bits to create k-wise independent bits. One proposal is to take a prime $p > n$ and create k-wise independent $\mathbb{F}_p$ valued $x_1 ... x_n$, each uniform in $\{0 ... p - 1\}$. The issue here is that we cannot map from field of odd size to the set of $\{0, 1\}$ without bias. There will always be more number of elements in either of the split. To resolve the problem, we force the field to be of even size. Hence, instead of considering field $\mathbb{F}_p$, we consider the field $\mathbb{F}_{p^{2n}}$. So we create k-wise independent $\mathbb{F}_{p^{2n}}$ valued $x_1 ... x_n$. Write each $x_i$ as a vector of length $m$ for $m > \log n$ , this gives $m.2^m$ k-wise independent bits and uses $k \log n$ bits of randomness.

## 5 A theoretical application of K-wise Independence

Most of the application of K-wise independence follow a general approach to de-randomize an algorithm. We analyze a randomized algorithm and notice that we are not using full independence of variables. We then devise an algorithm that uses partial independence and use K-wise independence to reduce the randomness. Reduction in randomness allows us to enumerate and evaluate solutions over all settings of randomness. Hence we obtain a fast deterministic algorithm. In this section, we will follow this template to solve (almost) a common problem associated with Ramsey Graphs.

**Theorem 13** (Ramsey's Theorem)**.** *If $G$ is a graph on $n$ vertices, then $G$ contains either a clique or an independent set of size $> \Omega(\log n)$.*

The proof of the theorem is based on induction and pigeon-hole principle and is an interesting read in itself. A closely related statement was given by Erdos:

**Theorem 14** (Erdos)**.** *There exists graph $G$ on $n$ vertices such that $G$ does not contain a clique or an independent set of size $> O(\log n)$.*

*Proof.* Consider the graph $G(n, 0.5)$ which is a random graph on $n$- vertices where each edge appears randomly with probability 0.5. This is same as picking one adjacent matrix uniformly at random from amongst all $2^{\binom{n}{x}}$ possible adjacency matrices. We wish to analyze the probability of the bad event that it has a clique or independent set and want to show that it is small. We fix a subset $S \in V$ of size $t$, then the probability that $S$ is a clique or an independent set is given by:

$$Pr[S \textit{ is clique or independent set}] = 2\left(\frac{1}{2}\right)^{\binom{t}{2}} \tag{1}$$

A union bound over all such subsets $S \in V$ of size $t$ results in a total probability:

$$Pr[\exists\ S \in V \textit{ s.t. } |S| = t \textit{ and } S \textit{ is clique or independent set}] \leq 2\binom{n}{t}2^{-\binom{t}{2}} \tag{2}$$

This probability is really small. For eg. for $t = 10 \log n$, this would be $\leq 2e^{10 \log n} n^{10 \log n} 2^{-100.\log^2 n/4} = o(1)$ Hence we can conclude that with high probability, $G$ has Erdos property. □

Note that the Erdos statement is a probabilistic statement doesn't really produce such a graph. The problem that we are interested in is to give an explicit construction of a graph $G$ on $n$-vertices with no clique or independent set of size $O(\log n)$.

### 5.1 How to get an explicit graph

We notice that in the proof of the Erdos statement, we did not use full independence of the graph, rather we only used independence related to clique size.

**Main Idea**: Instead of taking $G$ from a uniform distribution, we pick $G$ from $\binom{t}{2}$-wise independent distribution. This requires smaller amount of randomness but the proof would still hold true, and we will be able to enumerate over all possible settings of the randomness because of reduced randomness.

More formally, pick the edges of $G$ in a $\binom{t}{2}$-wise independent manner. This needs $O(t^2 \log(n^2))$ bits of randomness. For $t = 10 \log n$, this reduces to $O(\log^3 n)$.
Going over all $2^{O(\log^3 n)}$ settings of these bits and checking which of these graphs have big cliques or independent sets (this takes time $2^{O(\log^2 n)}$), we finally output the graphs that have no clique or independent sets.
Note that checking for a clique in a graph is NP-hard, but checking for a big clique (of size $\log n$) can be done in the time stated before. The over-all time complexity of this procedure is $2^{O(\log^3 n)}$ which is much better than $2^{n^2}$ which is the time it would have taken apriori. This is pretty close to being a polynomial (also called quasipolynomial).
A disadvantage of this approach is that eventually we will have to check for existence of clique in the graphs which cannot be done in time better than $2^{O(\log^2 n)}$). Hence we cannot hope to achieve a polynomial time algorithm using this approach.

## 6 $\epsilon$- Bias

Distributions have properties that make them look like totally uniform distribution from certain point of view. K-wise independence is one such property that looks totally random from the point of view of looking at only small number of coordinates at a time. $\epsilon$- bias is also another notion of pseudo randomness that looks at it from the point of view of certain properties involving parities.

**Definition 15** ($\epsilon$- Bias). *A collection of random variables $(x_1, \ldots, x_n) \in \{0,1\}^n$ is called $\epsilon$- biased if $\forall S \subset [n], S \neq \phi$, we have:*

$$Pr[\bigoplus_{j \in S} x_j = 0] \in \left[\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}\right] \tag{3}$$

In other words, we take subsets of bits, look at their parity and look at the resulting distribution. The distribution should be $\epsilon$ close to uniform. Smaller the $\epsilon$, more like uniform the distribution looks. To remove the complications arising out of the use of set, we have an alternative definition that uses vector instead:

**Definition 16.** *A collection of random variables* $(x_1, \ldots, x_n) \in \{0,1\}^n$ *is called* $\epsilon$*- biased if* $\forall y \in \{0,1\}^n, y \neq \phi$*, we have:*

$$Pr[< y, x >= 0] \in \left[\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}\right] \tag{4}$$

**Fact 17.** *Zero bias implies uniform distribution.*

**Theorem 18.** *There exists* $\epsilon$*- bias distributions on* $\{0,1\}^n$ *which are uniform distributions over a multi-set of size* $O(n/\epsilon^2)$*.*

In other words, there are sets of size $O(n/\epsilon^2)$ such that the uniform distribution over such sets is $\epsilon$ biased. If you knew this set and you could index these elements, that gives you a very efficient way to sample a distribution which is $\epsilon$ biased. For a set of size *poly*$(n)$, we need $O(\log n)$ bits to specify an element. Instead of sampling $n$ uniformly random bits and getting a zero biased distribution, we can get away with just $O(\log n)$ bits of randomness to get $\epsilon$ biased distribution for some small $\epsilon$. Though, the theorem doesn't give an explicit construction of such a set.

*Proof.* Take a parameter $m$ to be fixed later. Pick $z_1, ..., z_m \in \{0,1\}^n$ independently and uniformly at random. We want to show that with high probability over the choice of $z_1, ...z_m$, we have $\forall y \in \{0,1\}^n, y \neq \phi$,

$$|\{i \in [m] s.t. < y, z_i >= 0\}| \in \left[\frac{1-\epsilon}{2}m, \frac{1+\epsilon}{2}m\right] \tag{5}$$

We fix $y \in \{0,1\}^n$, $y \neq 0$. Fixing $y$ induces a split on the space of $\{0,1\}^n$ where one of the partition corresponds to $z_i$ such that $< z_i, y >= 0$ and the other corresponds to $z_i$ such that $< z_i, y >= 1$. These two partitions should be roughly equal. We wish to bound the bad event that these partitions are uneven, i.e. there are either too many or too few $z_i^s$ in the partition of $z_i$ such that $< z_i, y >= 0$. More formally we bound the following probability:

$$Pr[|\#(\{i \in [m] s.t. < y, z_i >= 0\}) - m/2| > \frac{\epsilon m}{2}] \tag{6}$$

We can use Chernoff bound to bound the above probability:

$$Pr[|\#(\{i \in [m] s.t. < y, z_i >= 0\}) - m/2| > \frac{\epsilon m}{2}] < e^{\Omega(\epsilon^2 m)} \tag{7}$$

Next we take a union bound over all choices of $y \neq 0$:

$$Pr[\exists y \in \{0,1\}^n s.t. |\#(\{i \in [m] s.t. < y, z_i >= 0\}) - m/2| > \frac{\epsilon m}{2}] \leq 2^n e^{\Omega(\epsilon^2 m)} \tag{8}$$

This is $\leq e^{-\Omega(n)}$ if $m = \Omega(n/\epsilon^2)$. We have shown that the probability of this bad event is exponentially small. □

In the next class, we will study construction of $\epsilon$-biased sets as well as their applications in derandomization.