CS 6815 Pseudorandomness and Combinatorial Constructions **Fall 2019**

Lecture 3: September 05

Lecturer: Eshan Chattopadhyay *Scribe: Yue Guo*

3.1 k-Wise Independence

Definition 3.1 (k-wise independence) *$X_1, \ldots, X_n$ is k-wise independent if $\forall T \subseteq [n], |T| = k$, $\{X_j\}_{j \in T}$ are independent random variables.*

Lemma 3.2 *Let $\{X_1, \ldots, X_n\}$ be k-wise independent random variables in the range $[0,1]$. Assume k is even. Let $X = \sum_{i=1}^n X_i$, $\mu = \mathbb{E}[X]$. Then $\forall t > 0$, $\Pr[|X - \mu| > t] \leq (O(\frac{\sqrt{2ne}}{t}))^k$.*

Proof:

With Markov inequility, we have $\Pr[(X-\mu)^k > t^k] \leq \frac{\mathbb{E}[(X-\mu)^k]}{t^k}$. Moreover, $\mathbb{E}[(X-\mu)^k] = \mathbb{E}[(\sum_{i=1}^n (X_i - \mu_i))^k]$ where $\mu_i = \mathbb{E}[X_i]$, as X_i's are k-wise independent. Let $Y_i = X_i - \mu_i$. We have $\mathbb{E}[Y_i] = 0$. Then the expectation can be written as $\mathbb{E}[(Y_1 + Y_2 + \ldots + Y_n)^k]$. Expand it, we have

$$\mathbb{E}[(X-\mu)^k] = \sum_{j_1, \ldots, j_k \in [n]} \mathbb{E}\left[\prod_{i=1}^k Y_{j_i}\right]$$

where there might be duplicated j_i. If we assume that X_i is binary (or chosen within $[0,1]$), each item in the above sum is less than or equal to 1.

Now we show that some of the items $\mathbb{E}\left[\prod_{i=1}^k Y_{j_i}\right]$ are zero. Combine all of the duplicated j_i's together and rewrite the expectation as $\mathbb{E}\left[Y_{j_1}^{k_1} \cdot \ldots \cdot Y_{j_l}^{k_l}\right]$ where each j_i are distinct and the sum of $k_1, \ldots, k_l$ is k. Now we show that if $l > k/2$,

$$\begin{aligned} \mathbb{E}\left[Y_{j_1}^{k_1} \cdot \ldots \cdot Y_{j_l}^{k_l}\right] &= \mathbb{E}[Y_{j_1}^{k_1}] \cdot \mathbb{E}[Y_{j_2}^{k_2}] \cdot \ldots \cdot \mathbb{E}[Y_{j_l}^{k_l}] && (Y_i\text{'s are } k\text{-wise independent}) \\ \text{if } l > k/2, \quad &= 0 && (\exists i \in [l] \text{ s.t. } k_i = 1 \text{ and } \mathbb{E}[Y_i] = 0\ \forall i \in [n]) \end{aligned} \tag{3.1}$$

Thus we need only consider the items with $l \leq k/2$. The number of items with $l \leq k/2$ is upper bounded by $\binom{n}{\frac{k}{2}} \cdot k^{\frac{k}{2}}$, which can be considered as the number of choices, choosing $n/2$ different j_i's from $[n]$, and then choosing k_i to be either 0 or $2, \ldots, k$ for all $i \in \{1, \ldots, \frac{k}{2}\}$. As each item is less than or equal to 1, we have

$$\mathbb{E}[(X-\mu)^k] \leq \binom{n}{\frac{k}{2}} \cdot k^{\frac{k}{2}} \cdot 1 \leq \left(\frac{2ne}{k}\right)^{\frac{k}{2}} \cdot k^{\frac{k}{2}} = (2ne)^{\frac{k}{2}} = \sqrt{2ne}^k$$

■

Lemma 3.3 (Construction of k-wise independent random variables) *For any prime p and $k > 0$, there is a construction of k-wise independent random variables $X_1, \ldots, X_p$ using $k\lceil \log p \rceil$ bits of randomness.*

Proof: Fix a finite field $\mathbb{F}_p$ where p is a prime. Sample a uniform vector $\overrightarrow{\alpha} = \{\alpha_0, \alpha_1, \ldots, \alpha_{k-1}\} \in \mathbb{F}_p^k$. Let $h_{\overrightarrow{\alpha}}(y) := \sum_{i=0}^{k-1} \alpha_i \cdot y^i$. We can construct a set of p random variables $\{X_i = h_{\overrightarrow{\alpha}}(i)\}_{i=0}^{p-1}$.

First we observe that X_i for all $i \in \{0, \ldots, p-1\}$ is uniform over $\mathbb{F}_p$. Now we prove that the random variables $\{X_i\}$ are k-wise independent. In other words, we need to prove the following claim:

Claim 3.4 *For any $T \subseteq \{0, \ldots, p-1\}$, $|T| = k$, denote $T = \{i_0, \ldots, i_{k-1}\}$, for any $\overrightarrow{\beta} = (\beta_0, \ldots, \beta_{k-1})$,*

$$\Pr[X_{i_j} = \beta_j \ \forall j \in \{0, \ldots, k-1\}] = \frac{1}{p^k}$$

Proof: We can write the event in the following way,

$$\Pr[X_{i_j} = \beta_j \ \forall j \in \{0, \ldots, k-1\}] = \Pr\left[\begin{pmatrix} 1 & i_0 & i_0^2 & \ldots & i_0^{k-1} \\ 1 & i_1 & i_1^2 & \ldots & i_1^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & i_{k-1} & i_{k-1}^2 & \ldots & i_{k-1}^{k-1} \end{pmatrix} \overrightarrow{\alpha} = \overrightarrow{\beta}\right]$$

The matrix (denoted with M) in the equation above is a Vandermonde's matrix, the determinant of which is non-zero. Thus $M\overrightarrow{\alpha} = \overrightarrow{\beta}$ has single solution. ■

As the sampling of vector $\overrightarrow{\alpha}$ uses $k \cdot \lceil \log p \rceil$ random bits, the lemma is proven. ■

Error Reduction Comparison Assume that some algorithm $\mathcal{A}$ uses R bits of randomness with running time T has success probability $\frac{2}{3}$. The table shows running time overhead and number of random bits needed with different kinds of random variables to amplify the success probability of $\mathcal{A}$ to $1 - \epsilon$,

	Running Time	Random bits
Independent r.v.s	$O(T \cdot \log(\frac{1}{\epsilon}))$	$O(R \cdot \log(\frac{1}{\epsilon}))$
2-wise independent r.v.s	$O(T \cdot \frac{1}{\epsilon})$	$2R + 2\log(\frac{1}{\epsilon}) + O(1)$
k-wise independent r.v.s	$O((\frac{1}{\epsilon})^{\frac{2}{k}} \cdot T)$	$kR + 2 \cdot \log(\frac{1}{\epsilon}) + O(k)$

3.2 Probabilistic Method

This is a general technique to show the existence of objects using probabilistic arguments. As an example, we prove the existence of Ramsey graphs using this technique.

Definition 3.5 (k-Ramsey Graphs) *$G = (V, E)$ is a k-Ramsey Graph if it is an undirected graph on n vertices (i.e., $|V| = n$) and the largest independent set and the largest clique in G are of size not larger than k.*

Claim 3.6 (Erdös 1947) *There exists* $(2\log n + O(1))$*-Ramsey graphs on* n *vertices.*

Proof: Pick a random graph $G(n, \frac{1}{2})$, i.e., there are n vertices and each edge is presented with probability $\frac{1}{2}$. Let k be a parameter which will be determined later. Let $T \subseteq [n]$ be any set of indices such that $|T| = k$. Then we have

$$\Pr[G_T \text{ is a clique or an independent set}] \leq 2 \cdot 2^{-\binom{k}{2}}$$

where G_T denotes the induced subgraph in G by T. For succinctness, we denote the event "G_T is a clique or an independent set" with $\mathcal{E}_T$. Then

$$\begin{aligned}
\Pr[G \text{ is not } k-\text{Ramsey}] &\leq \Pr\left[\bigcup_{T\subseteq[n],|T|=k} \mathcal{E}_T\right] \\
\text{(union bound)} \ &\leq \sum_{T\subseteq[n],|T|=k} \Pr[\mathcal{E}_T] \leq \binom{n}{k} \cdot 2 \cdot 2^{-\binom{k}{2}} \qquad (3.2) \\
&\leq \left(\frac{ne}{k}\right)^k \cdot 2^{-\frac{k(k-1)}{2}} \cdot 2
\end{aligned}$$

Pick $k = 2\log n + O(1)$ such that

$$\left(\frac{ne}{k}\right)^k \cdot 2^{-\frac{k(k-1)}{2}} \cdot 2 < 1 \qquad (3.3)$$

which means there must exist some graph G that is k-Ramsey. ∎