1 Overview

In this lecture, we:

1. Review some simple notions concerning linear codes.
2. Introduce the concept of approximate $k$-wise independence.
3. See some applications of approximate $k$-wise independence, including how to use the concept in conjunction with Chernoff bounds.

2 Linear codes

For this whole lecture, let $q$ be an arbitrary prime number, and let $F_q$ be the field of non-negative integers modulo $q$.

2.1 Review

Definition 1 A linear code is simply represented by a matrix $C \in F_q^{n \times l}$. Encoding a message $m \in F_q^l$ is done by multiplying $C$ with $m$, yielding $Cm \in F_q^n$.

$C$ thus defines an application $F_q^l \rightarrow F_q^n$, and $Im(C)$ is a subset of points of $F_q^n$.

Definition 2 A linear code $C$ has distance $d$ if $\forall x, y \in F_q^l, \Delta(Cx, Cy) > d$, where $\Delta$ is the Hamming distance function.

By linearity, the previous definition is equivalent to the following:

Definition 3 A linear code $C$ has distance $d$ if $\forall x \in F_q^l, \Delta(Cx, 0) > d$.

If a linear code $C$ has distance $d$, then it can correct up to $\lfloor \frac{d}{2} \rfloor$ errors. Given a message $m'$, simply find the closest point to $m'$ in $Im(C)$.

This motivates us to find a linear code with as large as possible a distance. It turns out that this isn’t very hard to do:

Theorem 4 If we pick $C$ uniformly at random in $F_q^{n \times l}$, then $C$ has a nearly optimal distance.
Note that if $q \geq n$, then we can pick $C$ to be the Vandermonde matrix of $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_q$. This is called a Reed-Solomon code, and has distance $l - 1$.

**Definition 5** Let $C \in \mathbb{F}_q^{n \times l}$ be a linear code. $M \in \mathbb{F}_q^{n \times l}$ is a parity-check matrix of $C$ if:

$$\forall w \in \mathbb{F}_q^n, \quad w^T M = 0 \iff \exists x \in \mathbb{F}_q^l / w = Cx$$

In other words, if $\text{Ker}(M^T) = \text{Im}(C)$.

**Claim 6** Let $C \in \mathbb{F}_q^{n \times l}$ be a linear code, $M$ a parity-check matrix of $C$, and $d \geq 0$. If $C$ has distance $d$, then any $d$ columns of $M$ are linearly independent.

**Proof** If $d$ columns are linearly dependent, then $\exists w \in \mathbb{F}_q^n$ such that $\Delta(w, 0) \leq d$ and $w^T M = 0$. Therefore $\exists x \in \mathbb{F}_q^l$ such that $\Delta(Cx, 0) \leq d$, and $C$ has distance less than $d$. ■

### 2.2 Application

Linear codes interest us particularly because of their connection to $k$-wise independence.

Let $x_1, \ldots, x_l$ be $l$ independent random variables in $\mathbb{F}_q$. Let $M \in \mathbb{F}_q^{n \times l}$ be the parity-check matrix of a linear code with distance $k$. Let $x = (x_1, \ldots, x_l) \in \mathbb{F}_q^l$, and $y = Mx \in \mathbb{F}_q^n$.

**Fact 7** $y_1, \ldots, y_n$ are $k$-wise independent random variables in $\mathbb{F}_q$.

**Definition 8** Let the sample space of $M$ be:

$$\text{Im}(M) = M \times l \left( \begin{array}{c} \mathbb{F}_q^l \\ \mathbb{F}_q \end{array} \right) = n \left( \begin{array}{c} \mathbb{F}_q^l \\ \text{Im}(M) \end{array} \right)$$

where we take $\mathbb{F}_q^l$ to be the horizontal concatenation of its vectors.

Therefore in order to simulate $k$-wise independent random variables, one can just sample uniformly at random from $\text{Im}(M)$.

### 3 Approximate $k$-wise independence

Let $X_1, \ldots, X_n \in \Omega$ be $n$ random variables, and let $k \leq n$. Recall that they are said to be $k$-wise independent if $\forall \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ and $\forall \alpha \in \Omega^k$:

$$\Pr [X_{i_1} = \alpha_1, \ldots, X_{i_k} = \alpha_k] = \prod_{j=1}^k \Pr [X_{i_j} = \alpha_j]$$

We covered the notion of $k$-wise independence in previous lectures, and showed how many of our proofs did not require full independence to go through. Instead we could weaken the requirement on many random variables to $k$-wise independence, where $k$ varied from problem to problem.
Still, we usually throw away constants in most of our calculations in search for asymptotic bounds, so why not weaken the requirement even further? Instead of requiring exact equality, we would like to use:

$$\Pr \left[ X_{i_1} = \alpha_1, \ldots, X_{i_k} = \alpha_k \right] \approx \prod_{j=1}^{k} \Pr \left[ X_{i_j} = \alpha_j \right]$$

One can formalize this idea in the following manner:

**Definition 9** Let \( \epsilon \geq 0 \). \( X_1, \ldots, X_n \) are \( \epsilon \)-biased \( k \)-wise independent if for all \( \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\} \) and \( \forall \alpha \in \Omega^k \):

$$\left| \Pr \left[ X_{i_1} = \alpha_1, \ldots, X_{i_k} = \alpha_k \right] - \prod_{j=1}^{k} \Pr \left[ X_{i_j} = \alpha_j \right] \right| \leq \epsilon$$

In practice, this notion is hard to use, as the additive errors sum up quickly through union bounding. This leads us to the notion of approximate independence, which is more immediately useful.

**Definition 10** Let \( \delta \geq 0 \). \( X_1, \ldots, X_n \) are \( \delta \)-approximate \( k \)-wise independent if for all event \( E \) depending only on \( k \) variables out of \( n \),

$$|\Pr[E] - \Pr[E \text{ if } X_1, \ldots, X_n \text{ were } k\text{-wise independent}]| \leq \delta$$

Note that if \( X_1, \ldots, X_n \in F_q \) are \( \epsilon \)-biased \( k \)-wise independent, then they also are \( q^k \epsilon \)-approximate \( k \)-wise independent. Indeed, there are at most \( q^k \) elementary settings for any \( E \) depending on \( k \) variables.

**Lemma 11** There exist \( n \) \( \delta \)-approximate \( k \)-wise independent random variables which are created from \( O\left(k + \log n + \log(1/\delta)\right) \) independent random variables.

**Sketch of Proof** Let \( C \) be a linear code with distance \( k \), and let \( M \) be a parity-check matrix of \( C \). Instead of taking our sample space to be:

$$Im(M) = M \times \left( \begin{array}{c} q^l \\ \mathbb{F}_q^l \end{array} \right)$$

Let \( C' : \mathbb{F}_q^{\log \frac{l}{\epsilon}} \rightarrow \mathbb{F}_q^l \) be a linear code with distance \( \frac{l}{2} - \epsilon \) [it is possible to create such a code]. Then our sample space is:

$$MC' = M \times \left( \begin{array}{c} l \log \frac{l}{\epsilon} \\ C' \end{array} \right)$$

**Remark** If \( k = \log n, \delta = \frac{1}{\text{poly}(n)} \), then the lemma states that it is possible to create \( n \frac{1}{\text{poly}(n)} \)-approximate \( \log n \)-wise independent random bits using only \( O(\log n) \) bits.

As we will see in the next section, this works very nicely with proofs using Chernoff bounds and ”with high probability”. Additionally, the sample space is reduced to \( \text{poly}(n) \).
4 Chernoff bounds and limited independence

The classical way to prove Chernoff bounds is the following:

- Observe that by convexity, $\forall t > 0$:
  \[
  \Pr \left[ \sum X_i < a \right] = \Pr \left[ e^{t \sum X_i} < e^{ta} \right] = \Pr \left[ \prod e^{tX_i} < e^{ta} \right]
  \]

- Use Markov’s inequality:
  \[
  \Pr \left[ \prod e^{tX_i} < e^{ta} \right] \leq \frac{\Exp \left[ \prod e^{tX_i} \right]}{e^{ta}}
  \]

- Then by independence:
  \[
  \Pr \left[ \prod e^{tX_i} < e^{ta} \right] \leq \prod \frac{\Exp \left[ e^{tX_i} \right]}{e^{ta}}
  \]

- Finally, minimize the resulting function of $t$ using functional analysis.

Unfortunately, this proof does not carry through if the $X_i$’s are not fully independent. In this section, we will see another more combinatorial proof of Chernoff bounds that will also work with limited independence instead of full independence.

**Theorem 12** Let $p \in [0, 1]$, $k = 4np$ and $X_1, \ldots, X_n \in \{0, 1\}$ $n$ $k$-wise independent random variables such that $\forall i \in \{1, \ldots, n\}, \Exp [X_i] = p$. Then the following inequality holds:

\[
\Pr \left[ \sum_{i=1}^{n} X_i \geq 4np \right] < \exp(-np)
\]

**Proof** First, observe the following:

\[
\Pr \left[ \sum_{i=1}^{n} X_i \geq 4np \right] = \Pr \text{[at least } k \text{ } X_i \text{’s = 1]}
= \Pr \left[ \exists S \subseteq \{1, \ldots, n\} \text{ s.t. } |S| = k \text{ and } \forall i \in S, X_i = 1 \right]
= \sum_{S \subseteq \{1, \ldots, n\}, |S| = k} \Pr [\forall i \in S, X_i = 1]
\]

And $k$-wise independence gives us that $\forall S \subseteq \{1, \ldots, n\}, |S| = k$:

\[
\Pr [\forall i \in S, X_i = 1] = p^k
\]

Summing over all $\binom{n}{k}$ different subsets of size $k$ of $\{1, \ldots, n\}$:

\[
\Pr \left[ \sum_{i=1}^{n} X_i \geq 4np \right] = \binom{n}{k} p^k \leq \left( \frac{ne}{k} \right)^k p^k = \left( \frac{np}{k} \right)^k = \left( \frac{e}{4} \right)^{4np} = \exp(-np)
\]

Note that this theorem still holds for $X_1, \ldots, X_n \in [0, 1]$ [that can be proven through a convexity argument]. Here however, for simplicity, we only consider the case where $X_1, \ldots, X_n \in \{0, 1\}$.
Theorem 13 Let \( p \in [0, 1] \), and \( X_1, \ldots, X_n \in \{0, 1\} \) \( n \) \( k \)-wise independent random variables such that \( \forall i \in \{1, \ldots, n\}, \text{Exp}[X_i] = p \).

Let \( \delta > 0, \ l = (1 + \delta)np \), and suppose \( k \leq \frac{\delta l^2}{4} \). Then the following inequality holds:

\[
\Pr \left[ \sum_{i=1}^{n} X_i \geq l \right] = (1 + \delta)np < \exp(-np)
\]

Proof As before, observe that:

\[
\Pr \left[ \sum_{i=1}^{n} X_i \geq l \right] = \Pr \left[ \text{at least } l \text{ } X_i \text{'s } = 1 \right]
\]

\[
= \Pr \left[ \exists S \subseteq \{1, \ldots, n\}, \text{ s.t. } |S| = l \text{ and } \forall i \in S, X_i = 1 \right]
\]

Now, if such a subset \( S \) exists, then there also exist \( \binom{l}{k} \) subsets \( S_j \subseteq S \) of size \( k \) (\( k \)-subsets of \( S \)), and \( \forall j, \forall i \in S_j, X_i = 1 \).

Thus the probability that \( S \) exists is at most equal to the probability that there exist \( \binom{l}{k} \) such \( k \)-subsets of \( \{1, \ldots, n\} \):

\[
\Pr \left[ \sum_{i=1}^{n} X_i \geq l \right] \leq \Pr \left[ \# \text{ } k \text{-subsets of } \{1, \ldots, n\} \text{ which are } "\text{all 1s}" \geq \binom{l}{k} \right]
\]

We can now apply Markov’s inequality and derive the following:

\[
\Pr \left[ \sum_{i=1}^{n} X_i \geq l \right] \leq \frac{\text{Exp}[\# \text{ } k \text{-subsets of } \{1, \ldots, n\} \text{ which are } "\text{all 1s}" ]}{\binom{l}{k}}
\]

\[
\leq \sum_{S \subseteq \{1, \ldots, n\}, |S| = k} \Pr \left[ \forall i \in S, X_i = 1 \right]
\]

Just as in the previous proof, we then use \( k \)-wise independence and finish off with some calculations:

\[
\Pr \left[ \sum_{i=1}^{n} X_i \geq l \right] \leq \binom{n}{k} p^k \leq \frac{n \ldots (n-k+1)}{l \ldots (l-k+1)} p^k \leq \left( \frac{n}{l-k} \right)^k p^k = \left( \frac{np}{l-k} \right)^k \leq \left( \frac{1}{1+\delta/2} \right)^k = e^{-k\log(1+\delta/2)} \leq \exp(-k\delta)
\]

In particular, if \( X_1, \ldots, X_n \) are \( \log n \)-independent, then the Chernoff bound holds with probability at most \( \exp(-\delta \log n) = \frac{1}{n^\delta} \), i.e. with high probability. The same proof also goes through if \( X_1, \ldots, X_n \) are \( \frac{1}{\text{poly}(n)} \)-approximate \( \log n \)-independent, such as we showed can be built using only \( O(\log n) \) bits.

Thus, the previous algorithms we covered, which we proved with Chernoff bounds, can be made to use only \( O(\log n) \) bits instead of \( O(n) \) bits. If randomness can be viewed as a resource, with an associated complexity, in the same manner as time and space, this shows we can get an exponential reduction in randomness requirement.