Chapter 4

Limited independence

4.1 Lecture 13 (10/Nov): Pairwise independence and the Chebyshev inequality

4.1.1 Improvement to the proof of Shannon’s coding theorem, using linear codes

Very commonly, in Algorithms, we have a tradeoff between how much randomness we use, and efficiency.

But sometimes we can actually improve our efficiency by carefully eliminating some of the randomness we’re using. Roughly, the intuition is that some of the randomness is going not toward circumventing a barrier (especially, leaving the adversary in the dark about what we are going to do), but just into noise.

A case in point is the proof of Shannon’s Coding Theorem. In a previous lecture we proved the theorem as follows: we first built an encoding map $E : \{0, 1\}^k \rightarrow \{0, 1\}^n$ by sampling a uniformly random function; then, we had to delete up to half the codewords to eliminate all kinds of fluctuations in which codewords fell too close to one another.

It turns out that this messy solution can be avoided. The key observation is that our analysis depended only on pairwise data about the code—basically, pairwise distances between codewords. “Higher level” structure (mutual distances among triples, etc.) didn’t feature in the analysis. So the argument will still go through with a pairwise-independently constructed code. So we’ll do this now, and in the process we’ll see how this helps.

Sample $E$ from the following pairwise independent family of functions $\{0, 1\}^k \rightarrow \{0, 1\}^n$. Select $k$ vectors $v_1, \ldots, v_k \text{iid} \in_U \{0, 1\}^n$. Now map the vector $(x_1, \ldots, x_k)$ to $\sum_i x_i v_i$. This is, of course, a linear map:

$$
\begin{pmatrix}
\vdots & \vdots & \vdots & \vdots \\
- & - & v_1 & - \\
- & - & v_2 & - \\
\vdots & \vdots & \vdots & \vdots \\
- & - & v_k & - \\
\end{pmatrix}
= (\text{codeword})
$$

The message $\bar{0} \in \{0, 1\}^k$ is always mapped to the codeword $\bar{0} \in \{0, 1\}^n$, and every other codeword is uniformly distributed in $\{0, 1\}^n$. It is not hard to see that the images of messages are pairwise independent. (Including even the image of the $\bar{0}$ message.) Looking back at the analysis of the error probability on message $X$ in Section 3.2.2, it had two parts, in each of which we bounded the probability of one of the following two sources of error:

---

1 People who pack a tent, wind up spending the night on the mountain – a climbing instructor of mine
Bad 1: \( H(\mathcal{E}(X) + R, \mathcal{E}(X)) \geq (p + \delta)n \). That is to say, the error vector \( R \) has weight (number of 1’s) at least \((p + \delta)n \). This analysis is of course unchanged, and is independent of the choice of the code. As before the bound is \( \leq 2^{-D_2(p + \delta)p/n} \).

Bad 2: \( \exists X' \neq X : H(\mathcal{E}(X) + R, \mathcal{E}(X')) \leq (p + \delta)n \). For this, pairwise independence is enough to obtain an analysis similar to before. Specifically, for any pair \( X \neq X' \) and any \( R \), the rv (which now depends only on the choice of code) \( \mathcal{E}(X) + R - \mathcal{E}(X') \) is uniformly distributed in \( \{0, 1\}^n \) (because at least one of \( X \) and \( X' \) is not the zero string, so its image is uniform after conditioning on the other image and on \( R \) so, by the same union bound as before, we can bound the probability of this error event by \( 2^{-n(D_2(p/2) - \epsilon/2)} = 2^{-n\epsilon/2} \).

So, the analysis is unchanged from before insofar as bounding \( \Pr_{\mathcal{E}, R}(\text{Error on } X) \) above by \( 2^{1-cn} \) for some \( c > 0 \) that depends only on \( p, \epsilon \). (See Eqn. (3.5).) That is, for every \( X \),

\[
E_\mathcal{E}(M_X) \leq 2^{1-cn} \tag{4.1}
\]

where \( M_X = \Pr_R(\text{Error on } X|\mathcal{E}) \).

Next, just as before, we wish to remove \( \mathcal{E} \) from the randomization in the analysis. In order to do this it helps to consider the uniform distribution over messages \( X \) and derive from Eqn. 4.1 the weaker

\[
E_{X,\mathcal{E}}(M_X) \leq 2^{1-cn} \tag{4.2}
\]

The reason is that this weaker guarantee is maintained even if we now modify the decoding algorithm so that it commutes with translation by codewords. Specifically, no matter what the decoder did before, set it now so that \( \mathcal{D}(Y) \) is uniformly sampled among “max-likelihood” decodings of \( Y \), which is equivalent (thanks to the uniformity over \( X \) and to the noise \( R \) being independent of \( X \)) to those \( X \) which minimize \( H(\mathcal{E}(X), Y) \). For the uniform distribution, max-likelihood decoding minimizes the average probability of error, so this new decoder \( \mathcal{D} \) also satisfies 4.2. The new decoder has the commutation advantage that we promised: for any \( \mathcal{E} \),

\[
\mathcal{D}(\mathcal{E}(X) + R) = \mathcal{D}(\mathcal{E}(X)) + \mathcal{D}(R) \begin{cases} \text{commutes with} \\ \text{translation by code} \end{cases} \mathcal{D}(\mathcal{E}(X)) + \mathcal{D}(R) = X + \mathcal{D}(R) \begin{cases} \text{decoding correct} \\ \text{on codewords} \end{cases} \tag{4.3}
\]

As a consequence,

\[
\Pr_R(\text{Error on } X_1|\mathcal{E}) = \Pr_R(\text{Error on } X_2|\mathcal{E}).
\]

So we can define a variable \( M \) which depends on \( \mathcal{E} \),

\[
M = \Pr_R(\text{Error on } 0|\mathcal{E}) = \Pr_R(\text{Error on } X|\mathcal{E}) \text{ for all } X
\]

and we have

\[
E_\mathcal{E}(M) \leq 2^{1-cn}
\]

Since \( M \geq 0, \Pr_\mathcal{E}(M > 2^{2-cn}) < 1/2 \) and so if we just pick linear \( \mathcal{E} \) at random, there is probability \( \geq 1/2 \) that (using the already-described decoder \( \mathcal{D} \) for it), for all \( X \) the probability of decoding in error is \( \leq 2^{2-cn} \).

What is much more elegant about this construction than about the preceding fully-random-\( \mathcal{E} \) is that no \( X \)'s with high error probabilities need to be thrown away. The set of codewords is always just a subspace of \( \{0, 1\}^n \).
The code also has a very concise description, \(O(k^2)\) bits (recall \(n \in \Theta(k)\)); whereas the previous full-independence approach gave a code with description size exponential in \(k\).

One comment is that although picking a code at random is easy, checking whether it indeed satisfies the desired condition is slow: one can either do this in time exponential in \(n\), exactly, by exhaustively considering \(R\)'s, or one can try to estimate the probability of error by sampling \(R\), but even this will require time inverse in the decoding-error-probability of \(R\) until we see error events and can get a good estimate of the error probability of \(R\); in particular we cannot certify a good code this way in time less than \(2^{cn}\).

### 4.1.2 Variance and the Chebyshev inequality

Let \(X\) be a real-valued rv. If \(E(X)\) and \(E(X^2)\) are both well-defined and finite, let \(\text{Var}(X) = E((X - E(X))^2)\). Expanding and applying linearity of expectation, this is also \(E(X^2) - E(X)^2\).

Note that If \(c \in \mathbb{R}\) then since the variance is homogenous and quadratic, \(\text{Var}(cX) = c^2 \text{Var}(X)\).

**Lemma 52 (Chebyshev)** If \(E(X) = \theta\), then \(\Pr(|X - \theta| > \lambda \sqrt{\text{Var}(X)}) < 1/\lambda^2\).

**Proof:**

\[
\Pr \left( |X - \theta| > \lambda \sqrt{\text{Var}(X)} \right) = \Pr \left( (X - \theta)^2 > \lambda^2 \text{Var}(X) \right) < 1/\lambda^2
\]

by the Markov inequality (Lemma[12]).

A frequently useful corollary is:

**Corollary 53** Suppose \(X\) is a nonnegative rv. Then \(\Pr(X = 0) \leq \frac{\text{Var}(X)}{(E(X))^2}\).

### 4.1.3 Pairwise independence and the second-moment inequality

A common situation in which we use Chebyshev’s inequality is when we have many variables which are not fully independent, but are pairwise independent (or nearly so).

**Definition 54 (Pairwise and \(k\)-wise independence)** A set of rvs are pairwise independent if every pair of them are independent; this is a weaker requirement than that all be independent. Likewise, the variables are \(k\)-wise independent if every subset of size \(k\) is independent.

**Definition 55 (Covariance)** The covariance of two real-valued rvs \(X, Y\) is (if well-defined) \(\text{Cov}(X, Y) = E(XY) - E(X)E(Y)\).

**Exercise:** Show that if \(X\) and \(Y\) are independent then \(\text{Cov}(X, Y) = 0\), but that the converse need not be true.

**Exercise:** If \(X = \sum_i X_i\), \(\text{Var} X = \sum_i \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)\).

**Corollary 56** If \(X_1, \ldots, X_n\) are pairwise independent real rvs with well-defined variances, then \(\text{Var}(\sum X_i) = \sum \text{Var}(X_i)\). If in addition they are identically distributed and \(\overline{X} = \frac{1}{n} \sum X_i\), then \(E(\overline{X}) = E(X_1)\) and \(\text{Var}(\overline{X}) = \frac{1}{n} \text{Var}(X_1)\).

**Exercise:** Apply the Chebyshev inequality to obtain: 58
Lemma 57 (2nd moment inequality) If $X_1, \ldots, X_n$ are identically distributed, pairwise-independent real rvs with finite 1st and 2nd moments then $P(|X - E(X)| > \lambda \sqrt{\text{Var}(X)/n}) < 1/\lambda^2$.

Corollary 58 (Weak Law) Pairwise independent rvs obey the weak law of large numbers. Specifically, if $X_1, \ldots, X_n$ are identically distributed, pairwise-independent real rvs with finite variance then for any $\varepsilon$, $\lim_{n \to \infty} P(|X - E(X)| > \varepsilon) = 0$.

So we see that the weak law holds under a much weaker condition than full independence. When we talk about the cardinality of sample spaces, we’ll see why pairwise (or small $k$-wise) independence has a huge advantage over full independence, so that it is often desirable in computational settings to make do with limited independence.

4.1.4 Preview: Threshold for $K_4$ in $G(n, p)$

Working with low moments of random variables can be incredibly effective, even when we are not specifically looking for limited-independence sample spaces. Here is a prototypical example. “When” does a 4-clique (written $K_4$) show up in a random graph selected from the distribution $G(n, p)$? We have in mind that we are “turning the knob” on $p$. When $p = 0$, of course with probability 1 there is no subgraph isomorphic to $K_4$. When $p = 1$, with probability 1 there is such a subgraph, in fact, $\binom{n}{4}$ of them. In between, for any finite $n$, the probability is finite. But we won’t take $n$ finite, we will take it tending to $\infty$.

So the question is can we identify a function $\pi(n)$ such that in the model $G(n, p(n))$, with $[K_4]$ denoting the event that there is a $K_4$ in the random graph $G$,

(a) If $p(n) \in o(\pi(n))$, then $\lim_n \Pr([K_4]) = 0$.

(b) If $p(n) \in \omega(\pi(n))$, then $\lim_n \Pr([K_4]) = 1$.

Such a function $\pi(n)$ is known as the threshold for appearance of $K_4$. It follows from work of Bollobas & Thomason [16] that monotone events—events that hold in $G'$ if they hold in $G$ and $G \subseteq G'$—always have a threshold function.

(A related but incomparable statement: for a monotone graph property, i.e., one invariant under vertex permutations, for any $\varepsilon > 0$ there is a $p(n)$ such that $\Pr_{p(n)}(\text{property}) \leq \varepsilon$ and $\Pr_{p(n)+O(1/\log n)}(\text{property}) \geq 1 - \varepsilon$. See [34].)

---

2 Recall $p(n) \in o(\pi(n))$ means that $\lim \sup p(n)/\pi(n) = 0$, and $p(n) \in \omega(\pi(n))$ means that $\lim \sup \pi(n)/p(n) = 0$. 

59
4.2 Lecture 14 (15/Nov): Second moments and applications of pairwise and nearly-pairwise independence

4.2.1 Most pairs independent: Threshold for $K_4$ in $G(n, p)$

Let $S \subseteq \{1, \ldots, n\}$, $|S| = 4$. Let $X_S$ be the event that $K_4$ occurs as a subgraph of $G$ at $S$—that is, when you look at those four vertices, all the edges between them are present.\footnote{More generally the method we are studying can be used to establish the probability of any fixed graph $H$ occurring as a subgraph in $G$, that is, there is an injection of the vertices of $H$ into the vertices of $G$ such that every edge of $H$ is present in $G$. This is different from asking that $H$ occur as a induced subgraph of $G$, in which case one also demands that the non-edges of $H$ be non-edges in $G$. That is an interesting question too but different in an essential way: the event is not monotone in $G$.} Conflating $X_S$ with its indicator function and letting $X$ be the number of $K_4$’s in $G$, we have

$$X = \sum_S X_S$$

and

$$E(X) = \binom{n}{4} p^6.$$

We are interested in $\Pr(X > 0)$. Let $\pi(n) = n^{-2/3}$.

(a) For $p(n) \in o(\pi(n))$, $E(X) \in o(1)$, so $\Pr([K_4]) \in o(1)$ and therefore $\lim_n \Pr([K_4]) = 0$.

(b) For $1 > p(n) \in \omega(\pi(n))$, $E(X) \in \omega(1)$. We’d like to conclude that likely $X > 0$ but we do not have enough information to justify this, as it could be that $X$ is usually $0$ and occasionally very large. We will exclude that possibility by studying the next moment of the distribution.

Before carrying out this calculation, though, we have to make one important note. Since the event $[K_4]$ is monotone, $[p \leq p'] \Rightarrow [\Pr_{G(n, p)}[K_4] \leq \Pr_{G(n, p')}[K_4]]$. (An easy way to see this is by choosing reals iid uniformly in $[0, 1]$ at each edge, and placing the edge in the graph if the rv is above the $p$ or $p'$ threshold.) This means that it is enough to show that $K_4$ “shows up” slightly above $\pi$. This is useful because some of our calculations break down far above $\pi$, not because there is anything wrong with the underlying statement but because the inequalities we use are not strong enough to be useful there and a direct calculation would need to take account of further moments.

To simplify our remaining calculations, then, let $p = n^{-2/3}g(n)$ for any sufficiently small $g(n) \in \omega(1)$; we’ll see how this is helpful in the calculations.

By an earlier exercise,

$$\Var(X) = \sum_S \Var(X_S) + \sum_{S \neq T} \Cov(X_S, X_T)$$

$X_S$ is a coin (or Bernoulli rv) with probability $p^6$ of coming up “heads”. The variance of such an rv is $p^6(1 - p^6)$. So

$$\Var(X_S) > 0 \text{ for } 0 < p < 1$$

The covariance terms are more interesting.

1. If $|S \cap T| \leq 1$, no edges are shared, so the events are independent and $\Cov(X_S, X_T) = 0$.

2. If $|S \cap T| = 2$, one edge is shared, and a total of 11 specific edges must be present for both cliques to be present. A simple way to bound the covariance is (since $E(X_S), E(X_T) \geq 0$) that

$$\Cov(X_S, X_T) = E(X_S X_T) - E(X_S) E(X_T) \leq E(X_S X_T) = p^{11}.$$

3. If $|S \cap T| = 3$, three edges are shared, and a specific 9 edges must be present for both cliques to be present. Similarly to the previous case, $\Cov(X_S, X_T) \leq p^9$. 

\[ \text{Var}(X) \leq \left( \frac{n}{4} \right) p^6 (1 - p^6) + \left( \frac{n}{2 \cdot 2 \cdot 2} \right) p^{11} + \left( \frac{n}{3 \cdot 1 \cdot 1} \right) p^9 \]
\[ \in O(n^4 p^6 + n^6 p^{11} + n^5 p^9) \]
\[ = O\left(n^4 p^6 + n^6 p^{11} + n^5 p^9\right) \]
\[ = O\left(g^6\right) \]

This gives us the key piece of information:
\[ \frac{\text{Var}(X)}{(E(X))^2} \in \frac{O(g^6)}{\Theta((n^4 p^6)^2)} = \frac{O(g^6)}{\Theta(g^{12})} = O(g^{-6}) \subseteq o(1) \]

and we have only to apply the Chebyshev inequality (Cor. 53) to conclude that \( \Pr(X = 0) \in o(1) \) and so \( \lim_n \Pr(\|K_4\|) = 1. \)

**Exercise:** Show that the threshold for appearance of the graph with 5 edges and 4 vertices is \( n^{-4/5} \).

**Comment:** In general the threshold is determined not by the ratio of edges to vertices, but by the maximum of this ratio over induced subgraphs. See [6].

### 4.2.2 Most pairs nearly independent: Turan’s proof of the concentration of the number of prime factors of a random integer

Now for an application of near-pairwise independence in number theory.

Let \( m(k) \) be the number of primes dividing \( k \). Hardy and Ramanujan showed that for large \( k \), this number is almost always close to \( \log \log k \). Specifically, let \( k \in \mathbb{U}[n] \), and let \( M \) be the RV
\[ M = m(k). \]

Always \( M \leq \log k \). But usually this is a vast overestimate:

**Theorem 59 (Hardy & Ramanujan)**
\[ \Pr(\ |M - \log \log k| > \lambda \sqrt{\log \log k}) < \frac{1 + o(1)}{\lambda^2} \]

**Proof:**
We show an elegant proof due to Turan.

Before we begin the proof in earnest let’s simplify things. The function \( \log \log \), besides being monotone, is so slowly growing that it hardly distinguishes between \( n \) and \( \sqrt{n} \). Specifically, \( \log \log n = \log 2 + \log \log \sqrt{n} \), so for \( k \geq \sqrt{n} \), we have
\[ |M - \log \log n| + \log 2 \geq |M - \log k| \]
\[ \log \log k \geq \log \log n - \log 2 \]

Consequently:
\[ \Pr(\ |M - \log k| > \lambda \sqrt{\log \log k}) \]
\[ \leq \Pr(k \leq \sqrt{n}) \]
\[ + \Pr(\ |M - \log \log n| + \log 2 > \lambda \sqrt{\log \log n - \log 2}) \]
\[ \leq \frac{1}{\sqrt{n}} \]
\[ + \Pr(\ |M - \log \log n| > (1 - o(1))\lambda \sqrt{\log \log n}) \]
Since $M \leq \lg n$, the probability on the LHS in the Theorem is 0 for $\lambda > (\lg n)/\sqrt{\log \log n}$, so it remains only to prove the theorem for $\lambda \leq (\lg n)/\sqrt{\log \log n}$. In this range, the $1/\sqrt{n}$ term in (4.4) is dominated by the $\sim 1/\lambda^2$ bound we need to show for the second term, so it is enough to show the latter bound. It remains to show, then:

**Proposition 60**

$$\Pr( |M - \log \log n| > \lambda \sqrt{\log \log n} ) < \frac{1 + o(1)}{\lambda^2}.$$  

Now to show the proposition. For prime $p$ let $[p|k]$ be the indicator for $p$ dividing $k$. Note $M = \sum_p [p|k]$. 

$$E([p|k]) = \lfloor n/p \rfloor / n$$ 

So 

$$1/p - 1/n \leq E([p|k]) \leq 1/p$$  

$$-1 + \sum_{\text{prime } p \leq n} \frac{1}{p} \leq E(M) \leq \sum_{\text{prime } p \leq n} \frac{1}{p} \quad (4.4)$$

For $k \geq 1$ let $\pi(k) = |\{p : p \leq k, p \text{ prime}\}|$. We remind ourselves of the

**Theorem 61 (Prime number theorem)** $\pi(k) \in (1 + o(1))k/\log k.$

We use the following corollary (proof omitted in class):

**Lemma 62** \( \sum_{\text{prime } p \leq n} 1/p \in (1 + o(1)) \log \log n. \)

So, from Eqn. (4.4) and Lemma 62 we know that 

$$E(M) \in (1 + o(1)) \log \log n.$$  

Now for the variance of $M$. The proposition will follow from showing 

$$\Var(M) = (1 + o(1)) \log \log n \quad (4.5)$$

and an application of the Chebyshev inequality.

To show Eqn (4.5), as always we can write 

$$\Var(M) = \sum_{\text{prime } p \leq n} \Var([p|k]) + \sum_{\text{primes } p \neq q \leq n} \Cov([p|k], [q|k]) \quad (4.6)$$

What we will discover is that the sum of covariances is very small and so the bound on $\Var(M)$ almost as if we had pairwise independence between the events $[p|k]$. 

We already noted in a previous lecture that for $\{0,1\}$-valued rvs $Y$, $\Var(Y) = E(Y)(1 - E(Y)) \leq E(Y)$. 

Applying this we have 

$$\sum_{\text{prime } p \leq n} \Var([p|k]) \leq \sum_{\text{prime } p \leq n} E([p|k]) \in (1 + o(1)) \log \log n.$$ 

Now to handle the covariances. Observe that for primes $p \neq q$, $[p|k][q|k]$ is the indicator rv $[pq|k]$. 

Just as for primes, $E([pq|k]) = \lfloor n/pq \rfloor / n \leq 1/pq$. So 

$$\Cov([p|k], [q|k]) = E([pq|k]) - E([p|k])E([q|k])$$

$$\leq \frac{1}{pq} - \left( \frac{1}{p} - \frac{1}{n} \right) \left( \frac{1}{q} - \frac{1}{n} \right)$$

$$\leq \frac{1}{n} \left( \frac{1}{p} + \frac{1}{q} \right)$$
This is a very low covariance, which is crucial to the theorem.

\[
\sum_{\text{primes } p \neq q \leq n} \text{Cov}(\{[p|k]\}, \{[q|k]\}) \leq \sum_{\text{primes } p \neq q \leq n} \frac{1}{n} \left( \frac{1}{p} + \frac{1}{q} \right)
\]

\[
= (1 + o(1)) \frac{2}{n} \pi(n) \sum_{p \leq n} \frac{1}{p} \quad \text{prime number theorem}
\]

\[
= (1 + o(1)) \frac{2}{n} \pi(n) \log \log n \quad \text{Lemma 62}
\]

\[
= (1 + o(1)) \frac{2 \log \log n}{\log n}
\]

This is dominated by the first term of \text{Var}(M) in 4.6 (it is even tending to 0), so we have established Eqn (4.5). \qed
4.3 Lecture 15 (17/Nov): 4-wise independent random walk; 2-wise independent vertex selection

4.3.1 Lower tail bound on random walk using 4th moment. Application to Gale-Berlekamp

In an earlier lecture we used a strong hammer, the CLT, to conclude that the value of the Gale-Berlekamp game is $\Omega(n^{3/2})$. Specifically we applied the CLT to show that for a symmetric random walk of length $n$, $X = \sum_{i=1}^{n} X_i$ with $X_i \in \{1, -1\}$, $E(|X|) \in \Omega(n^{1/2})$. Now we will show this from first principles—and more importantly, using only information about the 2nd and 4th moments.

This is not only of methodological interest. It makes the conclusion more robust, specifically the conclusion holds for any 4-wise independent space, and therefore implies a poly-time deterministic algorithm to find a Gale-Berlekamp solution of value $\Omega(n^{3/2})$, because there exist $k$-wise independent sample spaces of size $O(n^{\lfloor k/2 \rfloor})$, as we will show in a later lecture.

**Theorem 63** Let $X = \sum_{i}^{n} X_i$ where the $X_i$ are 4-wise independent and $X_i \in \{1, -1\}$. Then $E(|X|) \in \Omega(n^{1/2})$.

**Proof:** We start with two calculations. These calculations are made easy by the fact that for any product of the form $X_{i_1}^{b_1} \cdots X_{i_q}^{b_q}$, with $i_1, \ldots, i_q$ distinct and $b_i \geq 0$ integer,

$$E(X_{i_1}^{b_1} \cdots X_{i_q}^{b_q}) = \begin{cases} 0 & \text{if any } b \text{ is odd} \\ 1 & \text{otherwise} \end{cases}$$

So now

$$E(X^2) = \sum_{i,j} E(X_i X_j) = \sum_i E(X_i^2) = n$$

$$E(X^4) = 3 \sum_{i,j} E(X_i^2 X_j^2) - 2 \sum_i E(X_i^4) = 3n^2 - 2n.$$

One is tempted to apply Chebyshev’s inequality to the rv $X^2$, because we know both its expectation and its variance. Unfortunately, the numbers are not favorable! $\text{Var}(X^2) = 3n^2 - 2n - n^2 = 2n^2 - 2n > n^2 = E(X^2)^2$.

So we must focus directly on the rv $|X|$. Following [10], we start with an elementary inequality.

**Lemma 64** $y > \frac{3^{3/2}}{2} (y^2 - y^4)$ for all $y > 0$.

**Proof:** The quartic has simple roots at $-2/\sqrt{3}, 0$, and a double root at $1/\sqrt{3}$. See Fig. 4.3.1.

Now to obtain a lower bound on $E(|X|)$ we use a trick analogous to that we used when obtaining Chernoff bounds: we introduce an adjustable parameter, obtain a generic bound, and then optimize over the parameter. In this case we use a parameter $a > 0$ and use the previous lemma to conclude that

$$E(|X|/a) \geq \frac{3^{3/2}}{2} (X^2/a^2 - X^4/a^4).$$

Substituting our known moments,

$$E(|X|/a) \geq \frac{3^{3/2}}{2} \left( \frac{n}{a^2} \frac{3n^2 - 2n}{a^4} \right)$$
and using \( a = 3\sqrt{n} \) gives

\[
E(|X|) \geq \frac{\sqrt{3}}{2\sqrt{n}} \left( n - \frac{3n - 2}{9} \right) \\
\geq \sqrt{n/3}.
\]

\[\Box\]  

### 4.3.2 Maximal Indep. Set in NC

**Parallel complexity classes**

\[ L = \text{log-space} = \text{problems decidable by a Turing Machine having a read-only input tape and a read-write work tape of size (for inputs of length } n) \mathcal{O}(\log n). \]

\[ NC = \bigcup_k NC^k, \text{ where } NC^k = \text{languages s.t. } \exists c < \infty \text{ s.t. membership can be computed, for inputs of size } n, \text{ by } n^c \text{ processors running for time } \log^k n. \]

\[ RNC = \text{same, but the processors are also allowed to use random bits. For } x \in L \text{ Pr( error )} \leq 1/2, \text{ for } x \notin L \text{ Pr( error )} = 0. \]

\[ L \subseteq NC^1 \subseteq \ldots \subseteq NC \subseteq RNC \subseteq RP. \]

**P-Complete** = problems that are in P, and that are complete for P w.r.t. reductions from a lower complexity class (usually, log-space).

**Maximal Independent Set**

MIS is the problem of finding a Maximal Independent Set. That is, an independent set that is not strictly contained in any other. This does not mean it needs to be a big, let alone a maximum cardinality set. (It is NP-complete to find an independent set of maximum size. This is more commonly known as the problem of finding a maximum clique, in the complement graph.)

There is an obvious sequential greedy algorithm for MIS: list the vertices \( \{1, \ldots, n\} \). Use vertex 1. Remove it and its neighbors. Use the smallest-index vertex which remains. Remove it and its neighbors, etc.
The independent set you get this way is called the Lexicographically First MIS. Finding it is P-complete w.r.t. L-reductions [18]. So it is interesting that if we don’t insist on getting this particular MIS, but are happy with any MIS, then we can solve the problem in parallel, specifically, in NC².

We’ll see an RNC, i.e., randomized parallel, algorithm of Luby [58] for MIS. (The ideas we discuss in this and the next lecture were developed in this and in the papers [49, ?]). Then, we’ll see how to derandomize the algorithm.

Notation: $D_v$ is the neighborhood of $v$, not including $v$ itself. $d_v = |D_v|$.

Luby’s MIS algorithm:
Given: a graph with $n$ vertices.
Start with $I = \emptyset$.
Repeat until the graph is empty:

1. Mark each vertex $v$ of degree 0; mark each remaining vertex independently with probability $\frac{1}{2d_v}$.
2. For each doubly-marked edge, unmark the vertex of lower degree (break ties arbitrarily).
3. For each marked vertex $v$, append $v$ to $I$ and remove the vertices $v \cup D_v$ (and of course all incident edges) from the graph.

An iteration can be implemented in parallel in time $O(\log n)$ with a processor per pair of vertices. (One can be more careful about the number of processors.)

We’ll show that an expected constant fraction of edges is removed in each iteration (and then we’ll show that this is enough to ensure expected logarithmically many iterations).

Definition 65 A vertex $v$ is good if it has $\geq \frac{d_v}{3}$ neighbors of degree $\leq d_v$. (Let $G$ be the set of good vertices, and $B$ the remaining ones which we call bad.) An edge is good if it contains a good vertex.

Lemma 66 If $d_v > 0$ and $v$ is good then $\Pr(\exists$ marked $w \in D_v$ after step 1) $\geq 1 - e^{-1/6}$.

Proof: It’s enough to argue that one of the low-degree neighbors is likely to be marked. Markings are independent, so $\Pr(\text{none of the low-degree neighbors is marked}) \leq (1 - \frac{1}{2d_v})^{d_v/3} \leq e^{-1/6}$. (We will not use anything about the value $e^{-1/6}$ except that it is strictly less than 1. This will be useful later on.)

Lemma 67 If $v$ is marked then the probability it is unmarked in step 2 is $\leq 1/2$.

Proof: It is unmarked only if a weakly-higher-degree neighbor is marked. Each of these events happens with probability at most $\frac{1}{2d_v}$. Apply a union bound.

Corollary 68 The probability that a good vertex is removed in step 3 is at least $(1 - e^{-1/6})/2$.

Proof: Immediate from the previous two lemmas.

Now for our measure of progress.

Lemma 69 At least half the edges in a graph $(V, E)$ are good.
Proof: Direct each edge from lower to higher degree vertex; now we have in-degrees $d_v^{\text{in}}$ and out-degrees $d_v^{\text{out}}$. A bad vertex has $> 2d_v/3$ neighbors with degree $> d_v$, so its out-degree is $> 2d_v/3$. In particular

$$d_v/3 \leq d_v^{\text{out}} - d_v^{\text{in}}$$

For two sets of vertices $V_1, V_2$ let $E(V_1, V_2)$ be the edges directed from $V_1$ to $V_2$. (In particular $E = E(V, V)$. Note that $E(B, B)$ is the set of bad edges.

Now (on the LHS) we’ll count the pairs $(v, \{v, w\})$ where $v$ is a bad vertex and $\{v, w\}$ is an undirected edge.

$$2|E(B, B)| + |E(B, G)| + |E(G, B)| = \sum_{v \in B} d_v$$

$$\leq 3 \sum_B (d_v^{\text{out}} - d_v^{\text{in}})$$

$$= 3(|E(B, G)| - |E(G, B)|)$$

So $|E(B, B)| \leq |E(B, G)| - 2|E(G, B)|$.

$E(B, G)$ is a subset of the good edges. So $|E(B, B)| \leq |E|/2$.

Due to the corollary, each good edge is removed with probability at least $(1 - e^{-1/6})/2$. Of course the edge-removals are correlated, but in any case, the expected fraction of edges removed is at least $(1 - e^{-1/6})/4$.

Now we analyze how long it takes this kind of descent process to terminate. The state of the process is a nonnegative integer; the process terminates at 0. At $n > 0$, you sample a random variable $X$ from a distribution $\mu_n$ on $\{0, \ldots, n\}$, then transition to state $n - X$, where you start over. The question is, how many iterations $T$ does it take you to hit 0? Write $E_n(T)$ for the expectation of $T$ starting from state $n$.

Lemma 70 Let $g(i) = \min_{i \leq m \leq n} E_{\mu_m}(X)$. (Note $g$ is nondecreasing.) Then $E_n(T) \leq \sum_i 1/g(i)$.

Proof: By induction. For $n = 1$ this is just the expectation of a geometric distribution, which is to say, the expected time until a biased coin comes up heads. Note $g(1) = E_{\mu_1}(X) = \Pr_{\mu_1}(X = 1)$.

$$E_1(T) = 1 + \Pr_{\mu_1}(X = 0)E_1(T)$$

$$1 = E_1(T)(1 - \Pr_{\mu_1}(X = 0)) = E_1(T)\Pr_{\mu_1}(X = 1) = E_1(T)g(1)$$

For $n > 1$ we proceed by induction.

$$E_n(T) = 1 + E_{X \sim \mu_n}(E_{n-X}(T))$$

$$= 1 + \sum_{i=0}^n \Pr(X = i)E_{n-i}(T)$$
\[
\Pr(X > 0)E_n(T) = 1 + \sum_{i=1}^{n} \Pr(X = i)E_{n-i}(T) \\
\leq 1 + \sum_{i=1}^{n} \Pr(X = i) \sum_{j=1}^{n-i} \frac{1}{g(j)} \quad \text{induction} \\
= 1 + \sum_{i=1}^{n} \Pr(X = i) \left( \sum_{j=1}^{n} \frac{1}{g(j)} - \sum_{j=n-i+1}^{n} \frac{1}{g(j)} \right) \\
= 1 + \Pr(X > 0) \sum_{j=1}^{n} \frac{1}{g(j)} - \sum_{i=1}^{n} \Pr(X = i) \sum_{j=n-i+1}^{n} \frac{1}{g(j)} \\
\leq 1 + \Pr(X > 0) \sum_{j=1}^{n} \frac{1}{g(j)} \\
- \frac{1}{g(n)} \sum_{i=1}^{n} \Pr(X = i)i \quad \text{g monot. nondecr.} \\
= \Pr(X > 0) \sum_{j=1}^{n} \frac{1}{g(j)} + 1 - \frac{1}{g(n)}E_{\mu_n}(X) \\
\]

Now \( g \) was defined in such a way that \( g(n) = E_{\mu_n}(X) \), so 

\[
\ldots = \Pr(X > 0) \sum_{j=1}^{n} \frac{1}{g(j)} 
\]

\( \square \)

As a consequence, the expected number of iterations until the algorithm terminates is \( \leq \sum_{i=1}^{\lvert E \rvert} \frac{4}{i(1-e^{-1/6})} \in O(\log \lvert E \rvert) \in O(\log n) \). This is an RNC algorithm for MIS.