Reminder: Random Variables

- A **random variable** is a function from the set of possible events to $\mathbb{R}$.

- **Example.** Say that we flip five coins.
  - We can define the random variable $X$ to be the number of coins that landed on heads.
  - We can define the random variable $Y$ to be the percentage of heads in the tosses.
  - Notice that $Y = 20X$. 
Reminder: Indicator Random Variables

- An *indicator random variable* is a random variable $X$ that is either 0 or 1, according to whether some event happens or not.

- **Example.** We toss a die.
  - We can define the six indicator variable $X_1, ..., X_6$ such that $X_i = 1$ iff the result of the roll is $i$.

Reminder: Expectation

- The *expectation* of a random variable $X$ is
  \[ E[X] = \sum_{\omega \in \Omega} X(\omega) \Pr[\omega]. \]
  - Intuitively, $E[X]$ is the expected value of $X$ in the long-run average value of repetitions of the experiment it represents.
Reminder: Expectation Example

• We roll a fair six-sided die.
  ◦ Let $X$ be a random variable that represents the outcome of the roll.

  $E[X] = \sum_{\omega \in \Omega} X(\omega) \Pr[\omega] = \sum_{i \in \{1, \ldots, 6\}} i \cdot \frac{1}{6} = 3.5$

Reminder: Linearity of Expectation

• If $X$ is a random variable, then $5X$ is a random variable with a value five times that of $X$

• Lemma. Let $X_1, X_2, \ldots, X_k$ be a collection set of random variables over the same discrete probability. Let $c_1, \ldots, c_k$ be constants. Then

  $E[c_1 X_1 + c_2 X_2 + \cdots + c_k X_k] = \sum_{i=1}^{k} c_i E[X_i]$. 
Independent Sets

- Consider a graph $G = (V, E)$. An independent set in $G$ is a subset $V' \subset V$ such that there is no edge between any two vertices of $V'$.
- Finding a maximum independent set in a graph is a major problem in theoretical computer science.
  - No polynomial-time algorithm is known.

Warm Up

- What are the sizes of the maximum independent sets in:

  - 2
  - 4
Large Independent Sets

- **Theorem.** A graph $G = (V, E)$ has an independent set of size at least
  \[ \sum_{v \in V} \frac{1}{1 + \deg v}. \]

- **Proof.** We uniformly choose an ordering for the vertices of $V = \{v_1, \ldots, v_n\}$.
  - The set of vertices that appear before all of their neighbors is an independent set.

  - We uniformly choose an ordering for the vertices of $V = \{v_1, \ldots, v_n\}$.
    - The set $S$ of vertices that appear before all of their neighbors is an independent set.
    - $X_i$ - indicator that is 1 if $v_i \in S$.
      \[ E[X_i] = \Pr[X_i = 1] = \frac{1}{1 + \deg v_i}. \]
    - $X$ – the random variable of the size of $S$.
      \[ E[X] = E\left[ \sum_{i=1}^{n} X_i \right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{1 + \deg v_i}. \]
    - There must exist an ordering for which $|S|$ has at least this value.

  - We uniformly choose an ordering for the vertices of $V = \{v_1, \ldots, v_n\}$.
    - The set $S$ of vertices that appear before all of their neighbors is an independent set.
    - $X_i$ - indicator that is 1 if $v_i \in S$.
      \[ E[X_i] = \Pr[X_i = 1] = \frac{1}{1 + \deg v_i}. \]
    - $X$ – the random variable of the size of $S$.
      \[ E[X] = E\left[ \sum_{i=1}^{n} X_i \right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{1 + \deg v_i}. \]
    - There must exist an ordering for which $|S|$ has at least this value.
Sum-free Sets

- Consider a set $A$ of positive integers. We say that $A$ is **sum-free** if for every $x, y \in A$, we have that $x + y \notin A$ (including the case where $x = y$).

- What are large sum-free subsets of $S = \{1, 2, ..., N\}$?
  - We can take all of the odd numbers in $S$.
  - We can take all of the numbers in $S$ of size larger than $N/2$.

Large Sum-Free Sets Always Exist

- **Theorem.** For any set of positive integers $A$, there is a sum-free subset $B \subseteq A$ of size $|B| \geq \frac{1}{3} |A|$.

- **Proof.** Consider a prime $p$ such that $p > a$ for every $a \in A$.
  - From now on, calculations are mod $p$.
  - Notice that if $B$ is sum-free mod $p$, it is also sum-free under standard addition.
  - Thus, it suffices to find a large set that is sum-free mod $p$. 
Proof (cont.)

- The calculations are $\text{mod } p$.
- The set $S = \{[p/3], ..., [2p/3]\}$ is sum-free and $|S| \geq (p - 1)/3$.
- We uniformly choose $x \in \{1, 2, ..., p - 1\}$ and set $A_x = \{a \in A | ax \in S\}$.
- Consider $b, c \in A_x$. Since $bx, cx \in S$, we have that $(b + c)x = bx + cx \not\in S$. Thus, $(b + c) \not\in A_x$, and $A_x$ is sum-free.
- $X_a$ – indicator that is 1 if $a \in A_x$.

$$E[|A_x|] = E \left[ \sum_{a \in A} X_a \right] = \sum_{a \in A} E[X_a] = \sum_{a \in A} \Pr[X_a = 1].$$

- The set $S = \{[p/3], ..., [2p/3]\}$ is sum-free and $|S| \geq (p - 1)/3$.
- We uniformly choose $x \in \{1, 2, ..., p - 1\}$ and set $A_x = \{a \in A | ax \in S\}$. $A_x$ is sum-free.
- $X_a$ – indicator that is 1 if $a \in A_x$.
- $E[|A_x|] = \sum_{a \in A} \Pr[X_a = 1]$.
- Recall: If $x \not\equiv x' \text{ mod } p$ then $ax \not\equiv ax' \text{ mod } p$.
- We thus have $\Pr[X_a = 1] = |S|/(p - 1)$.

$$E[|A_x|] = \sum_{a \in A} \Pr[X_a = 1] = \frac{|A||S|}{p - 1} \geq |A| \frac{1}{3}.$$

- Thus, there exists an $x$ for which $|A_x| \geq |A| \frac{1}{3}$. 

Which Super Villain is a Mathematician?

Austin Powers’ Dr. Evil

Spiderman’s Dr. Octopus

Sherlock Holmes’ Professor Moriarty

More Ramsey Numbers

- In the previous class, we used a basic probabilistic argument to prove $R(p, p) > 2^{p/2}$.

- **Theorem.** For any integer $n > 0$, we have $R(p, p) > n - \binom{n}{p} 2^{1 - \binom{n}{p}}$. 

"Floyd, Evyn, David"
Proof

- Consider a random red-blue coloring of \( K_n = (V, E) \). The color of each edge is chosen uniformly and independently.

- For every subset \( S \subset V \) of size \( p \), we denote by \( X_S \) the indicator that \( S \) induces a monochromatic \( K_p \). We set \( X = \sum_{|S|=p} X_S \).

- \( E[X_S] = \Pr[X_S = 1] = 2^{1-(p/2)} \)

- By linearity of expectation, we have

\[
E[X] = \sum_{|S|=p} E[X_S] = \binom{n}{p} 2^{1-(p/2)}.
\]

Completing the Proof

- We proved that the in a random red-blue coloring of \( K_n \) the expected number of monochromatic copies of \( K_p \) is

\[
m = \binom{n}{p} 2^{1-(p/2)}.
\]

- There exist a coloring with at most \( m \) monochromatic \( K_p \)'s.

- By removing a vertex from each of these copies, we obtain a coloring of \( K_{n-m} \) with no monochromatic \( K_p \).
Recap

- How we used the probabilistic method:
  - Our first applications were simply about making random choices and showing that we obtain some **property with non-zero probability**.
  - We moved to more involved proofs, where we use **linearity of expectation** to talk about the “expected” result.
  - In the previous proof, we used a **two step method** – first we randomly choose an object, and then we alter it. This method is called **the alternation method**.

Transmission Towers

- **Problem.** A company wants to establish **transmission towers** in its large compound.
  - Each tower must be on top of a building and each building must be covered by at least one tower.
  - We are given the pairs of buildings such that a tower on one covers the other.
  - We wish to minimize the number of towers.
Building a graph

- We build a graph $G = (V, E)$.
  - A **vertex** for every building.
  - An **edge** between every pair of buildings that can cover each other.
  - We need to find the minimum subset of vertices $V' \subseteq V$ such that every vertex of $V$ has at least one vertex of $V'$ as a neighbor.

Dominating Sets

- Consider a graph $G = (V, E)$. A **dominating set** of $G$ is a subset $V' \subseteq V$ such that every vertex of $V$ has at least one neighbor in $V'$.
- It is not known whether there exists a polynomial-time algorithm for finding a minimum dominating set in a graph.
Warm Up

- **Problem.** Let $G = (V, E)$ be a graph with maximum degree $k$. Give a lower bound for the size of any dominating set of $G$.

- **Answer.**
  - Every vertex covers itself and at most $m$ other vertices, so any dominating set is of size at least
  
  $$\frac{|V|}{k + 1}.$$  

The Case of a Minimum Degree

- **Theorem.** Let $G = (V, E)$ be a graph with minimum degree $k$. Then there exists a dominating set of size at most $n \cdot \frac{1 + \lg k}{k + 1}$.

- **Proof.** We consider a random subset $S \subset V$ by independently taking each vertex of $V$ with probability $p = \frac{\lg k}{k + 1}$.

  - Let $T \subset V \setminus S$ be the vertices that have no neighbors in $S$.
  - $S \cup T$ is a dominating set.
Proof (cont.)

- $S \subset V$ – a random subset formed by independently taking each vertex of $V$ with probability $p = \frac{\lg(k+1)}{k+1}$.
- $T \subset V \setminus S$ – the vertices with no neighbors in $S$.

- $S \cup T$ is a dominating set.

- $E[|S|] = \sum_{v \in V} \Pr[v \in S] = \sum_v p = p|V|$.
- A vertex is in $T$ if it is not in $S$ and none of its neighbors are in $S$. The probability for this is at most $(1 - p)^{k+1}$.

- $E[|T|] = \sum_{v \in V} \Pr[v \in T] \leq \sum_v (1 - p)^{k+1} = (1 - p)^{k+1}|V|$.

Completing the Proof

- $p = \frac{\lg(k+1)}{k+1}$.

- Famous inequality. $1 - p \leq e^{-p}$ for any positive $p$.

- Thus, $(1 - p)^{k+1} \leq e^{-p(k+1)} = \frac{1}{k+1}$.

- We proved

  $$E[|S| + |T|] = E[|S|] + E[|T|] \leq \left(p + (1 - p)^{k+1}\right)|V| = \frac{\lg(k + 1) + 1}{k + 1}|V|.$$  

- There must exist a dominating set of size.
How to Choose the Probability?

- In the previous problem, we knew to choose \( p = \frac{\log(k+1)}{k+1} \). But how?
  - When you solve a question and the choice is not uniform, first mark the probability as \( p \).
  - At the end of the analysis you will obtain some expression with \( p \) in it. Choose the value of \( p \) that optimizes the expression.

The End: Professor Moriarty

- Professor Moriarty is a mathematician.
  - “At the age of twenty-one he wrote a treatise upon the binomial theorem”.
  - So a combinatorist?!
- Dr. Octopus is a nuclear physicist.
- Dr. Evil is a medical doctor.