

# Undergraduate Seminar: The Probabilistic Method

Chapter 1

# Topic

- This seminar will closely follow the book *The Probabilistic Method* by Alon and Spencer (which can be found for free online)
- This book covers the probabilistic method, a mathematical method for using probability to solve combinatorics problems
- Required background is a basic understanding of probability, for instance,
  - Random Variables
  - Expectation
  - Independence
  - Conditioning

# Logistics

- We will meet for 2 hours each week from 9:30-11:30 NYC time on Fridays.
- One student will present each week (therefore the number of times each student presents may depend on enrollment, however it should be no more than 3).
- Grading will be based on talk preparation and attendance (see syllabus for more details)

# Equations in Slides

If creating slides all equations are expected to be LaTeXed. There are several ways to do this, here are a few that I would recommend.

- Install the browser add on called math equations
  - [Here](#) is the link to download the extension.
  - Pro: it is probably the most efficient for small things like binomial coefficients since it works directly in Google Slides.
  - Con: Does not have the same versatility you would have in your own TeX document (It doesn't seem to allow you to use external packages such as amsthm, but let me know if you figure out how to do this!)
- Online TeX editor
  - [Here](#) is an example.
  - Pro: This has a math keyboard that will help if you are not so comfortable with LaTeX
  - Con: Not as efficient as the browser add on and also lacks the versatility
- Your own TeX editor
  - Pro: All the versatility you could dream of!
  - Con: Also not very efficient

# A Sample of the Method! (Section 1.1)

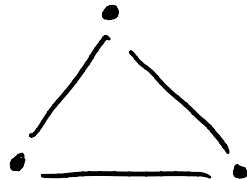
The general idea of the probabilistic method is using probability to prove that certain structures exist. We define probability spaces and show that such a structure occurs on this space with positive probability (and therefore, must be possible). To illustrate this, we will use a simple example, the Ramsey number!

# The Ramsey Number

The **Ramsey Number**  $R(j,k)$  is the smallest integer  $n$  such that in any 2-coloring of a complete graph on  $n$  vertices (here on denoted by  $K(n)$ ) by red or blue there must exist either a completely red  $K(j)$  or a completely blue  $K(k)$ .

Examples:

- $R(2,2) = 2$
- $R(2,3) = 3$



We will now use the method to prove the following proposition:

If  $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$  then  $R(k, k) > n$ . Thus  $R(k, k) > \lfloor 2^{k/2} \rfloor$  for all  $k \geq 3$

Proof. Consider a random 2-coloring of the edges of  $K(n)$  where each edge is colored independently with probability  $\frac{1}{2}$  of red and probability  $\frac{1}{2}$  of blue. For a set  $R$  of  $k$  vertices in  $K(n)$ , we define  $A(R)$  to be the event that the subgraph on  $R$  is either all red or all blue. Then we have,

$$Pr(A(R)) = 2^{1-\binom{k}{2}} \quad \left( \frac{1}{2} \right)^{\binom{k}{2}} + \left( \frac{1}{2} \right)^{\binom{k}{2}} = \frac{2^{\binom{k}{2}-1}}{2}$$

$\uparrow$   $R$                        $\uparrow$   $B$

To see this, note that there are  $\binom{k}{2}$  edges in the subgraph on  $R$ . We can write  $A(R)$  as the disjoint union of the event the subgraph on  $R$  is all red or all blue, both of which have probability  $\left(\frac{1}{2}\right)^{\binom{k}{2}}$ , so we get,  $Pr(A(R)) = 2\left(\frac{1}{2}\right)^{\binom{k}{2}} = 2^{1-\binom{k}{2}}$ . Since there are  $\binom{n}{k}$  choices for  $R$ , the probability that at least one of these events happens is at most

$$\binom{n}{k} 2^{1-\binom{k}{2}}$$

$$P\left(\bigcup A_i\right) \leq \sum P(A_i)$$

Now we have shown,

$$Pr[\mathbf{K}(n) \text{ has a complete red or blue subgraph of size } k] \leq \binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

Here we use the key insight of the probabilistic method. Since the probability of this not occurring is positive, there must exist a 2-coloring for which there is no complete red or blue subgraph of size  $k$ , which means that  $n$  cannot be the Ramsey number  $R(k,k)$ , and therefore  $R(k,k) > n$ .



Furthermore, if we take  $k \geq 3$  and we set  $n = \lfloor 2^{k/2} \rfloor$  then we have,

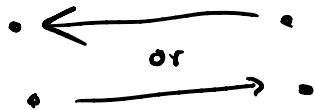
$$\binom{n}{k} 2^{1 - \binom{k}{2}} < \frac{2^{1+k/2}}{k!} \frac{n^k}{2^{k^2/2}} < 1$$

So, in particular this tells us that  $R(k,k) > \lfloor 2^{k/2} \rfloor$ .

# Some Remarks

- Could this have been done using a counting argument?
  - Yes, but it would probably be way more annoying, and later on we will see proofs where the probability is not so easily replaced.
  
- Can we actually construct a 2-color on  $K(n)$  that has no monochromatic  $K(k)$ ?
  - The proof is non-constructive but does lead us towards a very efficient randomized algorithm!
  - This is much better than an exhaustive search since there are  $2^{\binom{n}{2}}$  possible 2-colorings of  $K(n)$ .

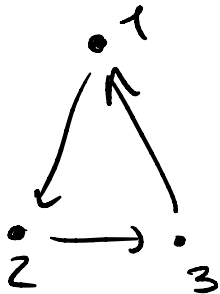
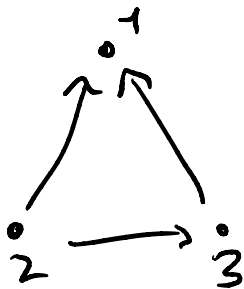
## 1.2 Graph Theory



A **tournament**  $T$  on a set  $V$  of  $n$  players is an orientation on the edges of the complete graph on the set of  $V$  vertices. We can think of  $T$  as being a directed graph on the vertex set  $V$  for which exactly one of the edges  $(v,w)$  or  $(w,v)$  is included for every pair  $v,w$  in  $V$ .

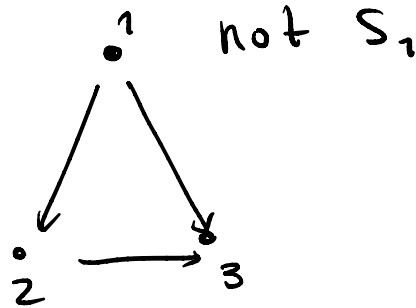
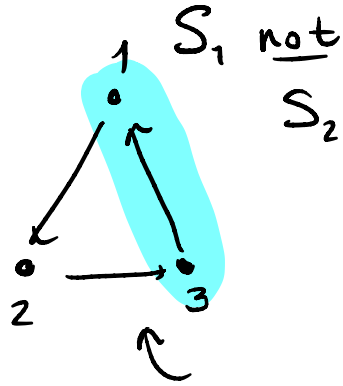
We call this a tournament because you can think of the edge  $(v,w)$  as “ $v$  beats  $w$ ”.

Examples:



We say a tournament  $T$  has the **property  $S_k$**  if for every set of  $k$  vertices in  $V$  there exists a vertex that beats them all. That is, for all  $A \subset V$ ,  $|A| = k$ , there is some  $v$  in  $V$  such that  $(v, a)$  is in  $T$  for all  $a$  in  $A$ .

Examples:



Question: Is it true that for every finite  $k$  there is a tournament with the property  $S_k$ ?

Construct random tournaments:

For each pair of vertices  $v, w$  in  $V$ , we choose between the edges  $(v, w)$  and  $(w, v)$  with probability  $\frac{1}{2}$  each. As a result all  $2^{\binom{n}{2}}$  are equally likely (i.e. the probability space is symmetric, which is often the case when applying the method).

We use this probability space to prove the following theorem:

If  $\binom{n}{k} (1 - 2^{-k})^{n-k} < 1$  then there is a tournament on  $n$  vertices with the property  $S_k$ .

If  $\binom{n}{k}(1 - 2^{-k})^{n-k} < 1$  then there is a tournament on  $n$  vertices with the property  $S_k$ .

Let  $T$  be a random tournament on the set  $V = \{1, \dots, n\}$ . For every fixed subset  $K \subset V$  of size  $k$ , let  $A_K$  be the event that there is no vertex that beats them all. Then we have that

$$Pr(A_k) = (1 - 2^{-k})^{n-k}$$

This is because for any fixed vertex  $v$  in  $V - K$ , the probability that  $v$  does not beat all of the vertices in  $K$  is  $1 - 2^{-k}$  and these events are independent for distinct vertices.

Therefore we obtain the bound,

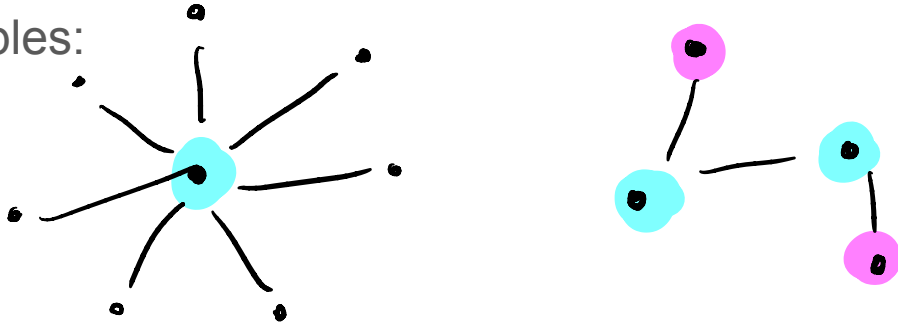
*"at least one subset does not have a  $v \in V$  that beats them all"*  $\Leftrightarrow$  *Our tourn. is not  $S_k$*

$$Pr(\bigcup_{K \subset V, |K|=k} A_k) \leq \sum_{K \subset V, |K|=k} Pr(A_k) = \binom{n}{k}(1 - 2^{-k})^{n-k}$$

By assumption in the theorem, this bound is  $< 1$ . Therefore with nonzero probability this tournament has the property  $S_k$ , implying that such a tournament exists.

A **dominating set** of an undirected graph  $G = (V, E)$  is a set  $U \subseteq V$  such that every vertex  $v$  in  $V$  has at least one neighbor in  $U$ .

Examples:



We can use the probabilistic method to prove the following bound for a dominating set.

Let  $G = (V, E)$  be a graph on  $n$  vertices, with minimum degree  $\delta > 1$ .

Then  $G$  has a dominating set of at most  $n \frac{1 + \ln(\delta + 1)}{\delta + 1}$  vertices.

Let  $G = (V, E)$  be a graph on  $n$  vertices, with minimum degree  $\delta > 1$ .

Then  $G$  has a dominating set of at most  $n \frac{1+\ln(\delta+1)}{\delta+1}$  vertices.

We will randomly generate a vertex set  $X$  by including each vertex  $v$  independently with probability  $p$  (where  $p$  will be specified later). We can compute the expected value of  $|X|$ :

$$\mathbb{E}[|X|] = \mathbb{E}\left[\sum_{v \in V} \mathbf{1}_{v \in X}\right] = np \sum_{v \in V} \mathbb{E}[\mathbf{1}_{v \in X}]$$

Define the set  $Y_X = \{v \text{ in } V: v \text{ and its neighbors are not in } X\}$ . Then we can bound the expected value of  $|Y_X|$ :

$$\mathbb{E}[|Y_X|] = \sum_{v \in V} \mathbb{E}[\mathbf{1}_{v \in Y_X}] \leq n(1-p)^{\delta+1}$$

Where the last inequality comes from computing the probability that  $v$  and its neighbors are all not in  $X$  in the worst case scenario where  $v$  has degree  $\delta$ .



So now we have,  $|X| + |Y_X|$

$$\mathbb{E}[|X \cup Y_X|] = \mathbb{E}[|X|] + \mathbb{E}[|Y_X|] \leq np + n(1-p)^{\delta+1}$$

Therefore there must exist at least one  $X$  such that  $|X \cup Y_X| \leq np + n(1-p)^{\delta+1}$  (this follows from the principle that if  $X$  is a random variable with mean  $m$ , then  $X$  cannot be strictly larger than  $m$ ).

Note that  $X \cup Y_X$  is clearly a dominating set. To obtain the bound from the statement of the theorem, we simply optimize over  $p$ .

Optimization:

$$np + \underbrace{n(1-p)^{\delta+1}}$$

1. Apply the bound  $1 - p \leq e^{-p}$  to bound the formula by  $np + \underbrace{ne^{-p(\delta+1)}}$
2. Take the derivative and set it equal to 0:

$$n + -n(\delta + 1)e^{-p(\delta+1)} = 0$$

3. Solve the equation to find a minimum at

$$p = \frac{\ln(\delta+1)}{\delta+1}$$

# Remarks

1. The method did not immediately give the bound, we had to choose  $p$  later. Additionally, we weakened our bound to get a cleaner result.
2. Note the use of expectation! Particularly the mean principle.
3. On page 5 there is a summary of an interesting algorithmic proof of this theorem that is interesting.
4. There is also a discussion of using dominating sets to determine edge-connectivity.

## 1.3 Combinatorics

A **hypergraph** is a pair  $H = (V, E)$  where  $V$  is a finite set whose elements are called vertices and  $E$  is a family of subsets of  $V$ , called edges. It is **n-uniform** if each of its edges has exactly  $n$  elements (note that a normal graph is a 2-uniform hypergraph). We say that  $H$  has **property B**, or that it is **2-colorable** if there is a 2-coloring of  $V$  such that no edge is monochromatic.

Let  $m(n)$  be the minimum number of edges of an  $n$ -uniform hypergraph that does not have property B.

Every  $n$  – uniform hypergraph with less than  $2^{n-1}$  edges has property  $B$ , therefore  $m(n) > 2^{n-1}$ .

Let  $H = (V, E)$  be a  $n$ -uniform hypergraph with less than  $2^{n-1}$  <sup>edges</sup>. We randomly color  $V$  by 2 colors, coloring each vertex ~~identically~~ <sup>independently</sup> with probability  $\frac{1}{2}$  of each color. For each edge  $e$ , let  $A_e$  be the event that  $e$  is colored monochromatically. Clearly,

$$Pr(A_e) = 2^{1-n} \quad \left(\frac{1}{2}\right)^n + \left(\frac{1}{2}\right)^n$$

So we can bound the probability that this random 2 coloring has at least one monochromatic edge by,

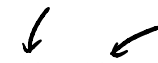
$$Pr(\text{Monochromatic edge}) \leq \sum_{e \in E} Pr(A_e) < \underbrace{2^{n-1}}_{\text{edges}} 2^{1-n} = 1$$

Therefore there must exist a 2-coloring that has no monochromatic edges.

$\Rightarrow H$  has property  $B$

We can obtain a better bound using a slightly better probability space! Here the probability space will be random over the edges of  $H$  rather than the coloring.

Let  $V$  be a set with  $v$  points ( $v$  will later be optimized). Let  $\chi$  be a coloring of  $V$  by 2 colors, with  $a$  red points and  $b = v - a$  blue points. Let  $S$  be a random  $n$ -set uniformly selected from all  $\binom{v}{n}$  possibilities. Then,

$$Pr(S \text{ is monochromatic under } \chi) = \frac{\binom{a}{n} + \binom{b}{n}}{\binom{v}{n}}$$


Assuming that  $v$  is even for convenience, this expression is minimized at  $a = b$ . Therefore we have that,

$$Pr(S \text{ is monochromatic under } \chi) \geq \frac{2\binom{v/2}{n}}{\binom{v}{n}}$$

Letting  $S_1, \dots, S_m$  be uniformly and independently chosen  $n$ -sets, with  $m$  to be determined later.

For each coloring  $\chi$  let  $A_\chi$  be the event that none of the  $S_i$  are monochromatic under  $\chi$  (i.e. that  $\chi$  is a valid 2-coloring).

$$Pr(A_\chi) \leq \left(1 - \frac{2^{\binom{v/2}{n}}}{\binom{v}{n}}\right)^m$$

There are  $2^v$  possible colorings, so,

*1 - probability not valid on a given edge*

$$Pr(\text{At least one coloring is valid a 2-coloring of } S_1, \dots, S_m) \leq \sum_{\chi} Pr(A_\chi) \leq 2^v \left(1 - \frac{2^{\binom{v/2}{n}}}{\binom{v}{n}}\right)^m < 1$$

Clearly if this quantity is less than 1,  $m(n) < m$ . Therefore we want to optimize for  $m$  as small as possible such that this quantity is  $< 1$ .

Optimizing  $2^v (1 - \frac{2^{\binom{v/2}{n}}}{\binom{v}{n}})^m$

Let  $p = \frac{2^{\binom{v/2}{n}}}{\binom{v}{n}}$ . We again apply the bound  $1 - p \leq e^{-p}$ . When  $m = \lceil \frac{v \ln 2}{p} \rceil$  then,

$$2^v (1 - p)^m < 2^v e^{-pm} \leq 1 \quad e^{-pm} = e^{-\frac{v \ln 2}{p} p} = e^{-v \ln 2} = \frac{1}{2^v}$$

So now we want to choose  $v$  that minimizes this expression for  $m$ , i.e., that minimizes  $v/p$ . See page 8 in the book for details. Ultimately it yields the bound,

$$m(n) < (1 + o(1)) \frac{e \ln 2}{4} n^2 2^n$$

We will skip theorem 1.3.3 in the notes for time. See page 8 if curious.



## 1.4 Combinatorial Number Theory

A subset  $A$  of the integers is called **sum-free** if there does not exist any  $a, b, c$  in  $A$  such that  $a + b = c$ .

$$(A + A) \cap A = \emptyset$$

Example: Odd integers

Every set  $B = \{b_1, \dots, b_n\}$  of nonzero integers contains a sum-free subset  $A$  of size  $|A| > \frac{n}{3}$

Let  $p = 3k + 2$  be a prime number which satisfies  $p > 2 \max_i |b_i|$ . Set  $C = \{k + 1, k + 2, \dots, 2k + 1\}$ . Then note that  $C$  is a sum free subset of  $\mathbb{Z}/p\mathbb{Z}$ , and that  $\frac{|C|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}$

$$k+i + k+j = 2k + (i+j) > 2k+1$$

Choose at random an element  $x$  from  $\{1, 2, \dots, p - 1\}$  and set  $d_i = x b_i \pmod{p}$ . As  $x$  ranges over all possible values,  $d_i$  ranges over all elements of  $\mathbb{Z}/p\mathbb{Z}$ . Therefore

$$Pr(d_i \in C) > \frac{1}{3} \quad \begin{array}{l} \text{randomly choose } x \text{ unit from} \\ \{1, \dots, p-1\} \end{array}$$

$$d_i = x b_i \pmod{p} \in \{1, \dots, p-1\}$$

Thus the expected value of the number of  $i$  such that  $d_i$  is in  $C$  is  $n/3$ . Therefore there must be an  $x$  such that at least  $n/3$  of the  $d_i$  are in  $C$ . If  $\{b_{i(1)}, \dots, b_{i(k)}\}$  are such that  $d_{i(j)}$  are all in  $C$ , then the set is sum-free. Thus by the expectation principle

sum free  
 $\{b_{i_1}, \dots, b_{i_n}\} \rightsquigarrow \{d_{i_1}, \dots, d_{i_n}\}$   
 $\uparrow$   
 $C$

there must exist a sum free set of size greater than  $n/3$ .

$\frac{1}{p}$

$$n \neq 0 \quad n, 2n, 3n, \dots \leftarrow$$

$$\{n, 2n, \dots, (p-1)n, pn=0\} = \{0, 1, \dots, p-1\} \text{ in } \mathbb{Z}/p\mathbb{Z}$$

$$\begin{aligned} n(b_{i_1} + b_{i_2}) &= (b_{i_1})n \\ x b_{i_1} + b_{i_2} &= b_{i_1} x \\ d_{i_1} + d_{i_2} &= d_{i_1} \pmod{p} \end{aligned}$$

# Summary

- Some new techniques we saw today:
  - Using the expectation to bound and instance of a random variable
  - Bounds and approximations, loosening the bounds to get cleaner results
  - Different options for what we could take to be random,
- We saw the method used in a number of different ways

