THE PROBABILISTIC METHOD IN PRACTICE

AGNIV SARKAR

ABSTRACT. The Probabilistic Method is a technique made popular by prominent mathematician Paul Erdős. In general, it is utilized in combinatorial and related fields to demonstrate the existence of objects with certain properties. Its strength comes from its ability to bypass constructive methods, ensuring existence through the pure idea of "nonzero probability of existing." The structure of this work is through introducing definitions and then playing around with the method, as it is simply the best way to learn it.

The Probabilistic Method is a method that can and should be taught to high school students due to the beautiful ideas present. This is inspired by The Probabilistic Method by Alon and Spencer [1], but written by and for high school students.

1. Basics of Probability

As the name suggests, the probabilistic method relies on probability. Here, we will use the prototypical example of rolling dice. Let us consider a standard 6-sided dice (Fig. 1).

FIGURE 1. 2 six-sided dice

For any possible output of the dice (1, 2, 3, 4, 5, 6), each one must have $\frac{1}{6}$ probability of appearing. This gives us our first insight. We will describe these outputs as "disjoint", as after rolling a dice once, you can never get two numbers.

Remark. Summing over the probability of all disjoint events must give us 1, as exactly one of these disjoint events must happen!

Furthermore, let us consider the unexpected—what is the probability of NOT getting a 1? Here it is clear that it is the probability of either getting a 2, 3, 4, 5, or 6. This is our second insight.

Remark. The probability of an event happening plus the probability of an event not happening is equal to one. This is the key idea behind "complementary counting."

It will become clear in later proofs why we care so much about complementary counting. We will end with some notation:

Definition. We say E is an event if it can happen in our probability space. Considering the dice roll, we can say E_1 is the event that we roll a one. Similarly, we can write E_{2k} in the event that we roll an even number. We can technically call E_7 the event that a 7 is rolled.

Definition. We also define the union and intersections of events. The union of two events is the event such that either of them happens, and the intersection of two events is the event such that both of them happen. Using our dice roll, consider E_{2k} and E_{3k}, where the first event is a rolling event and the second event is a rolling multiple of 3. We can then consider the event $E = E_{2k} \cup E_{3k}$, which is the event s.t. either E_{2k} OR E_{3k} happens, or we roll $2, 3, 4, 6$. Consider $E_{2k} \cap E_{3k}$, which corresponds to rolling a 6.

Definition. We let $P : E \to [0, 1]$ be the "probability function" on a probability space. Going back to the dice rolls and using the event examples from before, we have

$$\begin{aligned} P[E_1] &= \frac{1}{6} \\ P[E_{2k}] &= \frac{1}{2} \\ P[E_7] &= 0. \end{aligned}$$

If, instead, we were considering probabilities on a weighted dice (i.e. some numbers come out more than others), we would have a different probability function.

We have the following properties. Prove them yourself by drawing a Venn diagram!

Lemma 1. Let E_1, E_2 be events and P be our probability function. Then the following holds true, regardless of E_1, E_2 and P.

$$\begin{aligned} P[E_1 \cup E_2] &\leq P[E_1] + P[E_2] \\ P[E_1 \cap E_2] &\leq \min(P[E_1], P[E_2]) \\ P[E_1 \cup E_2] + P[E_1 \cap E_2] &= P[E_1] + P[E_2]. \end{aligned}$$

This is the key notation we will need for our work, but the author stresses that these definitions are fairly loose. There are more technical definitions of a probability space, but this one will be sufficient for now.

One last trick that the author believes important to know has to do with approximations, namely Stirling's.

Theorem 1 (Stirling's Approximation)**.** Let $n \in \mathbb{N}$. Then,

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

This looks nasty, and that is because it is. The proof will not be covered here, but its power lies in the ability to convert factorials into exponentials. This will be utilized later!

2. Beginning with Coloring

A general crash course in Graph Theory will unlock a lot of different modes of thinking for how we treat many problems in our day to day. This is not simply a

Definition. A graph $G = (V, E)$ is a collection of vertices V which are related by edges E. A vertex is a singular element, whereas an edge is a pair of vertices. A directed graph has the distinction that the edge set contains ordered pairs, similarly, an undirected graph has unordered pairs of edges. In general, we will talk about undirected graphs.

For example, consider the graph $G = (V, E)$ where

$$\begin{aligned} V &= \{1, 2, 3, 4, 5, 6\}, \\ E &= \{(4,6), (4,5), (3,4), (2,3), (2,5), (1,5), (1,2)\}. \end{aligned}$$

This is drawn in Fig. 2.

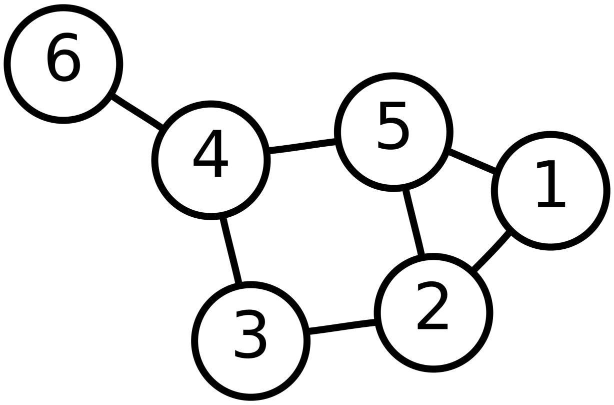

FIGURE 2. The graph G

Another example is K_7, which reads "the complete graph on 7" vertices. This simply means G has 7 elements in its vertex set and every pair of edges is contained in the edge set.

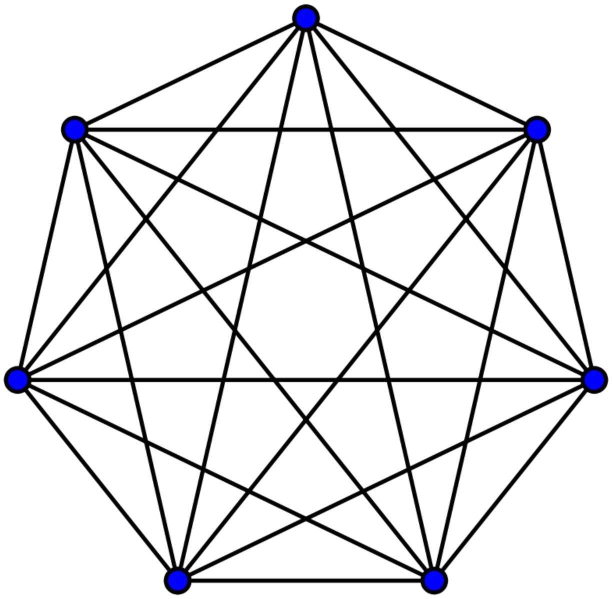

FIGURE 3. Complete graph on 7 vertices (all possible edges)

A nice question to ask is "how many graphs are there are n vertices?" The author leaves this for the reader to mull over.

Now, mathematics is not just technical mayhem—rather, it is art. So, a natural question to ask is if there are any ways to *color* the vertices of a graph such that each vertex is connected to different colors. Take the graph in Fig. 2, and try and color it (i.e. assign each vertex a color) with the smallest number of colors.

If you got 3 colors, you would be correct. In fact, there is a well-known upper bound on the smallest number of colors for a graph that is *planar*, or graphs that can be drawn in such a way that no edges (the lines) overlap with one another. For example, the graph given is planar, but the graph given in Fig. 3 is decidedly not planar, and requires 7 colors. For planar graphs, the bound is perhaps surprising for those who have never touched graph theory before:

Theorem 2 (4-Color Theorem)**.** Any planar graph G requires at most 4 colors to be colored such that each vertex that is connected by an edge has distinct colors.

Look at a map and count the number of colors used! What is surprising about this theorem is that it was conjectured in 1852, had some false proofs appear in 1879-80, and was finally proven with a computational method in 1977! This was one of the first major results to be proven with the help of a computer, and it was not until 2004 that the mathematical community had 100% certainty that this result was in fact proven true.

I quite like this theorem, and I hope you do too! Looking back at our graphs though, notice that the choice of 2 elements per edge is... a little arbitrary. This is a part of pure mathematics, wherein mathematicians generalize everything.

Definition. A hypergraph $H = (V, E)$ is a vertex set V and a hyper-edge set E, where an edge is a collection of vertices. The notion of undirected/directed is carried over from the graph definition. A n-uniform hypergraph H is one where each hyper-edge is composed of n vertices. For example, a 2-uniform hypergraph is a normal graph.

Another question naturally arises. Given n, k, is there a general formula for the number of k-uniform hypergraphs on n vertices? Mull over this.

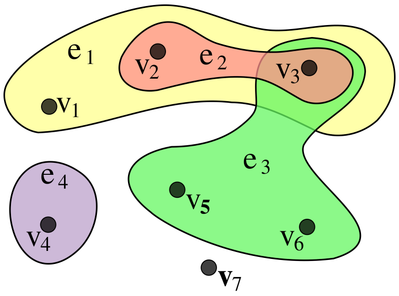

FIGURE 4. A hypergraph (Wikipedia)

As far as the author is aware, there are no explicit bounds on the number of colors needed to color a hypergraph such that each edge has distinct colors. Some more information is in [2].

So it's up to you now! Try and draw a 4-uniform hypergraph with 4 hyperedges. Try and color it such that each hyperedge has all of the colors you've chosen! This leads to our first problem.

Problem. Suppose $n \geq 4$ and let H be an n-uniform hypergraph with at most $\frac{4^{n-1}}{3^n}$ edges. Prove that there is coloring of the vertices of H by four colors so that in every edge, all four colors are represented.

Note the style of this problem—we are proving existence.

Proof. Let $n \geq 4$ and fix $H = (V, E)$ to be a hypergraph that is randomly colored on the vertices of a n-uniform hypergraph with 4 colors (each with equal probability of appearing), and having at max $\frac{4^{n-1}}{3^n}$ edges. To visualize this, one can imagine rolling a 4-sided dice, one for each vertex, and coloring it based on the number.

Define A_R to be the event for all edges $R \in E$ is not quad-chromatic, i.e. that it does not contain all four colors. We can then bound this probability as

$$P[A_R] < \frac{3^n}{4^{n-1}}.$$

The reason that this is true is the probability that an edge is without a color c is $\frac{3^n}{4^n}$. So, by adding it up for each color, you get an upper bound.

We know that $|E| \leq \frac{4^{n-1}}{3^n}$. Then,

$$P\left[\bigcup_{R \in E} A_R\right] < |E| \cdot \frac{3^n}{4^{n-1}} \leq \frac{4^{n-1}}{3^n} \cdot \frac{3^n}{4^{n-1}} = 1.$$

As the probability of any edge not containing every color is less than 1, there is a nonzero probability that every edge in E contains every color. As such, there is a coloring on H such that every color is represented on every edge. □

The key idea here is complementary counting!

This idea of coloring will be touched upon later, but before that, we will touch upon directed graphs at least once, as they come into practice a lot in the real world.

Definition. We say a directed graph $T = (V, E)$ is a tournament if T is a complete graph (each unordered pair of vertices is in E). Consider a round-robin tournament between 4 people, and draw a vertex from the winner to a loser (Fig. 5). This is why we call it a tournament. We call V the people!

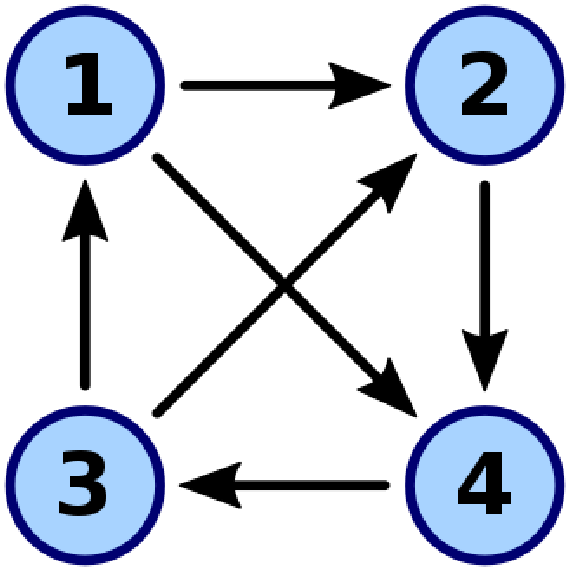

FIGURE 5. Tournament

Again, try and find the number of tournaments on n vertices. An interesting property is generally called S_k and is defined below.

Definition. Let T be a tournament. We say T has the S_k property if, for all k players, there exists another player that beats them. In terms of edges/vertices, this means for all sets of k vertices there exists another vertex that points to all of these vertices.

Try drawing a few tournaments of size 4 which have the S_2 property. Notice that all tournaments that satisfy S_k cannot have any player that wins all of the games!

It is an interesting problem to *bound* the size of tournaments such that they have the S_k property. Consider the following problem.

Problem. Prove that if $\binom{n}{k}(1-\frac{1}{2^k})^{n-k} < 1$, then there is a tournament on n vertices that has the property S_k.

We will prove this with the probabilistic method. Note this is also an existence proof!

Proof. Fix integers n and k such that $1 \leq k \leq n$. Let T be a random tournament on n vertices such that all edge's direction is chosen independently. For all sets of vertices, R with $|R| = k$ let A_R be the event such that S_k does not apply, i.e. no vertex beats every vertex in R. Then,

$$P[A_R] = (1 - 2^{-k})^{n-k}.$$

The reason this is true is that there are $(n-k)$ vertexes in T outside R, and each has a 2^{-k} probability of beating every vertex in R. So, as S_k not applying would mean that there exists an R such that A_R applies. So we can use the bound to find

$$P[T \text{ does not have } S_k] = P\left[\bigcup_{|R|=k} A_R\right] \leq \binom{n}{k}(1-2^{-k})^{n-k}.$$

So, if $\binom{n}{k}(1-2^{-k})^{n-k} < 1$, then there is a nonzero probability that S_k applies to T, and hence that there is a tournament on n vertices that has the property S_k. □

We again used the idea of complimentary counting! There is more on this problem and this specific property in [3], which goes further into other interesting properties.

This last problem is meant to be a little harder, so do not feel defeated if it seems a little more tricky!

Problem. A tournament is called *transitive* if the directed edges (i, j) and (j, k) imply the directed edge (i, k). (Equivalently, there is some ordering of the vertices such that the endpoint of every directed edge (i, j) satisfies $i < j$.) Show that there exists a tournament on n vertices, which does not contain any transitive subtournament on $2 + 2\log_2(n)$.

Proof. Let T be a random tournament on n vertices such that the probability of a vertex beating or losing to another node is $\frac{1}{2}$. For every sub-tournament R of T, let A_R be the probability that R is transitive. Then,

$$P[A_R] = s!2^{\frac{s(1-s)}{2}}.$$

This is because R being transitive is equivalent to R being a strict total ordering, and there are $s!$ different ways of ordering them and the edges must follow suit.

So, $P\left[T \text{ has at least one transitive subtournament}\right]$ can be written as

$$\begin{aligned} P\left[\bigcup_{|R|=s} A_R\right] &\leq \binom{n}{s}s!2^{\frac{s(1-s)}{2}} \\ &\leq \left(\frac{n}{k}\right)^s \cdot e\sqrt{n}\left(\frac{s}{e}\right)^s \cdot 2^{-(1+2\log_2(n))(1+\log_2(n))} \end{aligned}$$

If $s = 2 + 2\log_2(n)$, then

$$\begin{aligned} P\left[\bigcup_{|R|=2+2\log_2(n)} A_R\right] &= en^{s+\frac{1}{2}} 2^{-1-3\log_2(n)-2\log_2(n)\log_2(n)} \\ &= en^{s+\frac{1}{2}} \cdot \frac{1}{2} \cdot n^{-3} \cdot n^{-2\log_2(n)} \\ &= \frac{e}{2} \cdot n^{s-\frac{5}{2}-2\log_2(n)} \\ &\leq \frac{e}{2\sqrt{n}}. \end{aligned}$$

If there is a nonzero possibility for T to contain a transitive subtournament, the prior expression has to be less than 1. So,

$$\begin{aligned} \frac{e}{2\sqrt{n}} &< 1 \\ \frac{e}{2} &< \sqrt{n} \\ \frac{e^2}{4} &< n \\ 1.8 < \frac{e^2}{4} &< n \\ 1.8 &< n \end{aligned}$$

So, there exists a tournament on 2 or more vertices which does not contain any transitive subtournament on $2 + 2\log_2(n)$ vertices. □

3. Ramsey!

Frank Ramsey was an amazing man, and it is quite amazing that he has a theory named after him purely from lemmas that he developed for different theories. One of his most notable contributions is called the Ramsey number.

Definition. We define $R(s, t)$ to be the smallest number n such that any red-blue coloring of K_n must contain a red K_s or a blue K_t.

Ramsey's theorem guarantees $R(s, t)$'s existence, and the intuition is that for extremely large graphs, there are too many edges to not guarantee a K_s or a K_t.

The author asks the reader to compute $R(1, 1), R(2, 2)$, and begs them not to attempt $R(4, 4), R(5, 5), R(6, 6), \ldots$. There are just too many graphs, and it is actually unknown for $R(5, 5)$ and above. The most recent improvement was from the summer of 2023, and their proof utilizes some probabilistic methods. See [4] for more information!

This is why we find bounds.

Problem. Prove that if there exists a real number p, with $0 \leq p \leq 1$, such that

$$\binom{n}{s}p^{\binom{s}{2}} + \binom{n}{t}(1-p)^{\binom{t}{2}} < 1,$$

then $R(s,t) > n$.

The key idea here is we will color the vertices red with probability p. This will be more interesting a little later.

Proof. Let G be the random coloring of K_n where for each edge it is independently colored red with probability p and blue with probability $(1-p)$. For all subgraphs $R \in G$ such that $|R| = s$ let A_R be the event where all vertices in R are red. Then, $P[A_R] = p^{\binom{s}{2}}$, as there are $\binom{s}{2}$ edges within R. Similarly, for all subgraphs $B \in G$ such that $|B| = t$ let E_B be the event where B is monochromatically blue. So, $P[E_B] = (1-p)^{\binom{t}{2}}$. Then,

$$P\left[\left(\bigcup_{|R|=s} A_R\right) \cup \left(\bigcup_{|T|=t} E_B\right)\right] \leq \binom{n}{s}p^{\binom{s}{2}} + \binom{n}{t}(1-p)^{\binom{t}{2}}.$$

So, if $\binom{n}{s}p^{\binom{s}{2}} + \binom{n}{t}(1-p)^{\binom{t}{2}} < 1$, then there is a nonzero probability that $R(s,t) > n$. □

Ramsey numbers go extremely deep, and the probabilistic method was popularized through this.

4. Exploiting Expectation

Leaving graph theory land, let's depart into number theory! That's right *number theory.* This section assumes a little bit of modular arithmetic experience and begins to go into the expected value.

Let's go back into the dice-rolling situation for a second. What is the expected value for our dice roll?

Since each output 1-6 has a probability $\frac{1}{6}$, the expected value is

$$\sum_{i=1}^{6} \frac{i}{6} = 3.5$$

On a weighted dice, the expected would be different. But here we actually see the key properties of expected value.

Definition. We call E our expectation function. We have for random variable X,

$$E[X] = \sum_{\substack{x \text{ is a value} \\ \text{that } X \text{ can achieve}}} P[x]x.$$

Lemma 2. Expectation is a linear operator. Namely, for two random variables X, Y with constants a, b, we have

$$E[aX + bY] = aE[X] + bE[Y].$$

The author asks the reader to prove this to themselves. Now, we will proceed with a proof.

Definition. An element x is a residue mod n^2 if $x \in \mathbb{Z}/n^2\mathbb{Z}$.

I'd like to preface this problem with how zany it is. Most number theory results are.

Problem. Let A be any set of n residues mod n^2. Show that there is a set B of n residues mod n^2 such that at least half of the residues mod n^2 can be written as $a + b$ with $a \in A$ and $b \in B$.

Play around with this before you read the proof. When the author first saw this, it was... astonishing.

Proof. Fix n to be some positive integer, and then fix A to be any set of n residues mod n^2. Pick B uniformly from the collection of all sets of n residues mod n^2. Then, denote X to be the random variable $|A + B|$ and X_r to be the random indicator variable for the event that an element $r \in \mathbb{Z}_2$ such that $r \in A + B$. Then, we should find the probability of $r \in A + B$.

For r to be in $A + B$, there are exactly n residues mod n^2 that could be in B for that to be true, namely $\{r\} - A$. So, $P[r \in A + B] = 1 - P[r \notin A + B]$. So, $P[r \notin A + B]$ is the number of possible B's without those values over the number of all possible B's. This would be $\frac{\binom{n^2-n}{n}}{\binom{n^2}{n}}$. As such,

$$P[r \in A + B] = 1 - \frac{\binom{n^2-n}{n}}{\binom{n^2}{n}}.$$

We can then simplify this expression to be

$$P[r \notin A + B] = \frac{\binom{n^2-n}{n}}{\binom{n^2}{n}} = \frac{(n^2 - n)!^2}{(n^2 - 2n)!(n^2)!}.$$

Now we must introduce a lemma.

Lemma 3. For every nonzero n and some nonnegative k such that $k < n^2$,

$$\frac{n^2 - n}{n^2} \geq \frac{n^2 - n - k}{n^2 - k}$$

In the author's opinion, this should hopefully seem intuitive, as we're making the number overall smaller on the right-hand side. Work backward on paper before you read the proof.

Proof. Fix some nonzero n and k such that k is nonnegative and $k < n^2$. As k is nonnegative,

$$\begin{aligned} k &\geq 0 \\ n^3 - n^2 - nk + k &\geq n^3 - n^2 - nk \\ (n^2 - k)(n-1) &\geq n(n^2 - n - k) \\ \frac{n^2 - n}{n^2} &\geq \frac{n^2 - n - k}{n^2 - k}. \end{aligned}$$

□

So, with Lemma 3 as well as the fact that $1 - x \leq e^{-x}$,

$$\begin{aligned} \frac{(n^2-n)!^2}{(n^2-2n)!(n^2)!} &= \frac{(n^2-n)!}{(n^2)!} \frac{(n^2-n)!}{(n^2-2n)!} \\ &= \frac{(n^2-n)(n^2-n-1)\dots(n^2-2n+1)}{(n^2)(n^2-1)\dots(n^2-n+1)} \cdot \frac{(n^2-2n)!}{(n^2-n)!} \cdot \frac{(n^2-n)!}{(n^2-2n)!} \\ &= \underbrace{\left(\frac{n^2-n}{n^2}\right)\left(\frac{n^2-n-1}{n^2-1}\right)\cdots\left(\frac{n^2-2n+1}{n^2-n+1}\right)}_{n \text{ distinct terms}} \\ &\leq \left(1 - \frac{1}{n}\right)^n. \\ &\leq \left(e^{-\frac{1}{n}}\right)^n \\ &= \frac{1}{e} < \frac{1}{2} \end{aligned}$$

Then,

$$E[X] = \sum_{r=0}^{n^2-1} X_r = n^2 \cdot \left(1 - \frac{\binom{n^2-n}{n}}{\binom{n^2}{n}}\right) \geq \frac{n^2}{2}.$$

So, there exists a B such that $|A+B| \geq \frac{n^2}{2}$.

□

5. Incredible Independent Sets

Returning to graphs, we need to define a notion of degree.

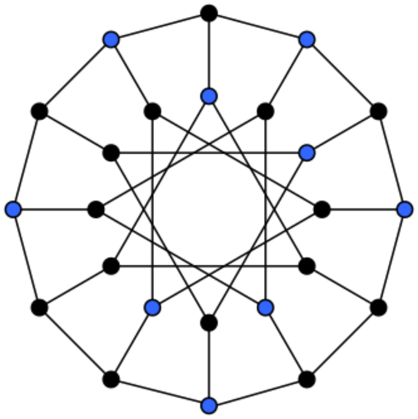

Figure 6. The blue vertices are the Independent Set on the graph.

Definition. Let G be a graph and v a vertex in G. The degree of v in G is the number of edges that contain v, or, the number of nodes v is connected to by an edge in G.

It is currently unclear how this leads to our next definition.

Definition. Let $G = (V, E)$ be a graph, and let G' be a subgraph of G, i.e. its vertex set and edge set are subsets of G's vertex and edge set such that G' is also a graph. We say that G' is an independent set if there are no possible edges in E that connect any two vertices in G'.

Do complete graphs have independent sets? If so, how large are they? An example of another graph whose independent set is highlighted is below in Fig. 6.

It is a surprisingly nontrivial question to find bounds on the size of independent sets for general graphs. We can use the probabilistic method to find a lower bound.

Problem. Let G be a graph on n vertices and let $d(v)$ denote the degree of vertex v. Prove that G has an independent set containing at least $\sum_{i=1}^{n} \frac{1}{d(v_i)+1}$.

Try and convince yourself of why this might be the case.

Proof. Fix $G = (V, E)$ to be a graph on $n \in \mathbb{N}$ vertices and let $d(v)$ be the degree of the vertex v. Construct a random independent set I as follows. Create a random subgraph of G by first looking at the vertices of G in a random order, and add this to I. Then, add each next vertex in the ordering to I if no neighbors of that vertex are in I. Let the random variable X as $X = |I|$, and for all vertex v let X_v be a random indicator variable when $v \in I$. Then, $X = \sum_{v \in V} X_v$.

Let v be any vector in V. Then, let S be the set of all neighbors of v and v itself, meaning that $|S| = d(v)+1$. So, v has a $\frac{1}{d(v)+1}$ chance of being the first element of our random ordering of V. In this case, $|S \cap I| = 0$, so $v \in I$. Thus, $P[v \in I] = P[v \text{ is first in the sequence}] + P[v \text{ is not first, but } v \in I]$. Thus, $P[v \in I] \geq P[v \text{ is first}] = \frac{1}{d(v)+1}$. So,

$$E[X] = \sum_{i=1}^{n} E[X_{v_i}] = \sum_{i=1}^{n} P[v_i \in I] \geq \sum_{i=1}^{n} P[v_i \text{ is first}] = \sum_{i=1}^{n} \frac{1}{d(v_i)+1}$$

There then must be some independent set with size at least $\sum_{i=1}^{n} \frac{1}{d(v_i)+1}$. □

I think the utilization of the linearity of expectation to convert it into the sum of probabilities is quite powerful.

For this next problem, read it and draw it out first.

Problem. Let $v_1, \ldots v_n$ be unit vectors in $\mathbb{R}^d$. Prove that it is possible to assign weights $\mathcal{E}_i \in \{\pm 1\}$ such that the vector $\sum_i \mathcal{E}_i v_i$ has Euclidean norm less than or equal to $\sqrt{n}$.

The author begs the reader to not turn the page and attempt constructive/inductive proof. The page space is yours.

Proof. Fix n to be a non-negative integer. Then, fix a set of n unit vectors $v_1, \dots v_n$. Let W be the random set of size n such that each element is either 1 or -1 with equal probability $\frac{1}{2}$. Denote the ith value of W as $\mathcal{E}_i$.

Then, let V be $\sum_i \mathcal{E}_i v_i$, e.g. the weighted sum of all the vectors. Then,

$$E\left[|V|^2\right] = E\left[\left(\sum \mathcal{E}_i v_i\right)\left(\sum \mathcal{E}_i v_i\right)\right] = \sum_i \sum_j E\left[\mathcal{E}_i \mathcal{E}_j\right] v_i v_j.$$

Note that $E[\mathcal{E}_i \mathcal{E}_j] = 0$ if $i \neq j$. If $i = j$, then $E[\mathcal{E}_i^2] = 1$. So,

$$\sum_i \sum_j E\left[\mathcal{E}_i \mathcal{E}_j\right] v_i v_j = \sum_i v_i \cdot v_i = n.$$

This means that there is a set of weights such that $|V|^2 \leq n$. This would mean that for this set of weights, $|V| \leq \sqrt{n}$. So, there exists a set of weights such that it satisfies the problem. □

Isn't that insane?! The proof is almost immediate.

6. Alterations

This is now one of the more technical applications of the probabilistic method. Currently, we have seen random events, random orderings, expectations, etc. Now, however, we consider random processes to have some more control over the outcome.

Problem. Prove that every k-uniform hypergraph with n vertices and $m \geq \frac{n}{3}$ edges contains an independent set (i.e., a set of vertices containing no edges) of size at least

$$\frac{2n^{\frac{3}{2}}}{3\sqrt{3m}}.$$

This bound is much cleaner than the one we proved before, and as such, requires this idea of an alteration. Note it is also a generalization to a hypergraph.

Proof. Fix a k-uniform hypergraph $G = (V, E)$ with $|V| = n$ and $|E| = m \geq \frac{n}{k}$. Choose a random subset $V' \subseteq V$ with the following process. Choose a set of vertices $S \subseteq V$ each independently with probability p. Let $|S| = X$. If there is an edge in S, then remove one of the vertices connected to the edge so that S is eventually independent. Denote the number of vertices that were removed from the original S as the random variable Y. This final set is now V'.

So, we can then write $E[|V'|] = E[X] - E[Y]$ by linearity of expectation. We can see that $E[X] = np$, and we can also figure out $E[Y]$. Within the V', the probability of an edge e being chosen is $P[e \in V'] = p^k$, as there are k vertices in an edge. Then, an upper bound on $E[Y]$ is mp^k, as there are m edges, and we may overcount some vertices being removed. So,

$$E\left[|V'|\right] = E\left[X - Y\right] \geq np - mp^k.$$

We then can take the derivative of the function on the left with respect to p in order to find the maximum value of the function.

$$\frac{d}{dp}\left[np - mp^k\right] = n - kmp^{k-1} = 0$$

$$p = \left(\frac{n}{km}\right)^{\frac{1}{k-1}}.$$

Then, we can plug this in to get

$$E[|V'|] \geq n\left(\frac{n}{km}\right)^{\frac{1}{k-1}} - m\left(\frac{n}{km}\right)^{\frac{k}{k-1}}.$$

For the problem, if you plug in $k = 3$, it simplifies rather nicely into the expected solution. □

Here's the secret trick! From before, we didn't really utilize p that much beyond a variable. Here, we actively utilize it through the derivative to maximize our expected value. It's a neat trick!

7. Additional Alterations

Instead of independence, we'll look at another subgraph property.

Definition. A dominating set for a graph $G = (V, E)$ is a subset V' of V such that any vertex of G is either in V or has an edge with an element in D.

Again, let's think about a complete graph. What's the smallest dominating set? The largest?

The largest in general is the whole set, and it is generally not too interesting. However, what about the smallest bound?

Problem. Let $G = (V, E)$ be a graph on n vertices with minimum degree $\delta > 1$. Prove that G has a dominating set containing at most $n\frac{1+\ln(\delta+1)}{\delta+1}$ vertices by choosing a random subset $S \subseteq V$ (independently including each vertex with probability p) and considering the size of the dominating set formed by adding to S any vertices not already covered by S.

Proof. Fix $G = (V, E)$ to be a graph on n vertices each with minimum degree $\delta > 1$. Then, let $S \subseteq V$ be a random subset created through the following process. Choose a random subset of $V' \subseteq V$ where each point has independent probability p of being chosen. Let $X = |V'|$. Then, if there are any points outside of V' that are not connected to any of the vertices inside the subset, e.g., the points that prevent the subset from being dominating, add them to V'. Denote the number of vertices that you add to the subset as the random variable Y. Then the final subset becomes S.

Then, we can see that $E[|S|] = E[X] + E[Y]$. We have $E[X] = np$, and we can figure out $E[Y]$ as well. There is a $(1 - p)$ probability for two points to be disconnected (do not share an edge). Then, $E[Y] \leq n(1 - p)^{\delta+1}$, as this is the probability that not choosing a vertex that is disconnected (does not share an edge) with S.

So, by using the fact that $1 - x \leq e^{-x}$,

$$E[|S|] = E[X] + E[Y] \leq np + n(1 - p)^{\delta+1} \leq np + ne^{-p(\delta+1)}$$

We can take the derivative of the function on the left with respect to p in order to find the maximum of the function. Then,

$$\frac{d}{dp}\left[np + ne^{-p(\delta+1)}\right] = n - ne^{-p(\delta+1)}(\delta + 1) = 0$$

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

For this p,

$$E[|S|] \leq n\frac{\ln(1 + \delta)}{1 + \delta} + ne^{-(\delta+1)\frac{\ln(1+\delta)}{1+\delta}} = n\frac{1 + \ln(1 + \delta)}{1 + \delta}.$$

This means there exists a dominating subgraph in G containing at most $n^{\frac{1+\ln(1+\delta)}{1+\delta}}$ vertices. □

8. Asymmetric Alterations

We return to Ramsey numbers.

Problem. Show that $R(s,t) > n - \binom{n}{s}p^{\binom{s}{2}} - \binom{n}{t}(1-p)^{\binom{t}{2}}$ for all positive integers n, s, and t, and for all $p \in [0,1]$.

This should look somewhat similar to one of the first few problems we did; flip back to it!

Proof. Fix $n, s, t \in \mathbb{N}$, and some $p \in [0,1]$. Then, fix a random coloring of the edges of K_n, constructed in the following way. Color an edge red with probability p or blue with probability $(1-p)$. Then, for each complete K_s that is all red within this graph, remove one vertex. Denote the number of vertices to be removed as X. Similarly, for each complete K_t that is all blue within this graph, remove a vertex. Denote the number of vertices to be removed as Y. Then, let G be the resultant graph. Denote the random variable S to be the number of vertices in G. This can be written as

$$E[S] = n - E[X] - E[Y]$$

Note that $E[X] \leq \binom{n}{s}p^{\binom{s}{2}}$, as for each set of vertices of size s in the original K_n coloring, it has probability $p^{\binom{s}{2}}$ of being monochromatically red. This means removing a vertex, and it is less than or equal to as we may be counting a removed vertex multiple times. Similarly, $E[Y] \leq \binom{n}{t}(1-p)^{\binom{t}{2}}$. So,

$$\begin{aligned} E[S] &= n - E[X] - E[Y] \\ &\geq n - \binom{n}{s}p^{\binom{s}{2}} - \binom{n}{t}(1-p)^{\binom{t}{2}}. \end{aligned}$$

Note that $R(s,t) > S$, as G does not contain any red K_s or blue K_t. So, in particular,

$$R(s,t) > E\left[|S|\right] \geq n - \binom{n}{s}p^{\binom{s}{2}} - \binom{n}{t}(1-p)^{\binom{t}{2}}.$$

□

9. Acute Subsets of $\mathbb{R}^d$

Going away from graphs again (by now it should be clear that the TPM is almost field-less, i.e. it works in many of them), let's consider sets in the real numbers.

Definition. We say a set $S \subseteq \mathbb{R}^d$ is acute if, for any three elements in S, they form an acute angle.

Consider acute subsets in $\mathbb{R}^2$! Can you get more than 3?

Problem. State and prove a lower bound on the maximum size of an acute subset of $\mathbb{R}^d$ by using alterations! You should again start by randomly choosing n vertices of the d-dimensional hypercube uniformly and independently ... And again, try to find an explicit bound in terms of d so that you have a full appreciation of what you've accomplished.

This proof assumes a little bit of linear algebra, namely an understanding of the dot product.

Proof. Fix $d, n \in \mathbb{N}, n \leq d$. Then, let P be a randomly chosen set of points in $\mathbb{R}^d$ with the following process. Uniformly randomly choose n points from a d-dimensional hypercube. Then, for all triples of points that form a right angle, remove one point, and let the random variable X be the number of points removed. The resulting point set will then only have acute angles, and this will be P. Denote the random variable S to be the size of P. Then, $E[S] = n - E[X]$. For $n < 3$, this is simply n, but we are interested in much higher dimensions.

So, if we have three ordered independently chosen points, $A, B, C \in P$, what is the probability that they can form a right angle? This can be done with the dot product. By comparing the vectors formed with $(A - B)$ and $(C - B)$. Then, $(A - B) \cdot (C - B) = 0$ if and only if $\angle ABC = 90°$.

So, we can look at this component-wise. Let $A = (a_1, a_2, \ldots a_d), B = (b_1, b_2, \ldots b_d), C = (c_1, c_2, \ldots c_d)$. Then, the expression from before can be written as

$$\sum_{i=1}^{d} (a_i - b_i)(c_i - b_i) = 0.$$

We know that each component of the points must be 0 or 1, so we can figure out the probability for this to be true. Note that if $b_i = 0$, then the expression $(a_i - b_i)(c_i - b_i)$ must evaluate to a nonnegative, and if $b_i = 1$, then note that the function must return either 1 or 0. So, this term is always nonnegative, meaning that we must focus on when $(a_i - b_i)(c_i - b_i) = 0$. This means that at least one of the terms is 0, which has a $\frac{3}{4}$ probability of happening. Because we want it for each component, then, we can finally get

$$P[\angle ABC = 90°] = \left(\frac{3}{4}\right)^d.$$

So, there are $\frac{{}^nP_3}{2}$ unique angles (as order matters, but reversing order does not change the angle). Then,

$$E[X] \leq \frac{{}^nP_3}{2} \cdot \left(\frac{3}{4}\right)^d = \frac{n!}{2(n-3)!} \cdot \left(\frac{3}{4}\right)^d \leq \frac{n^3}{2} \cdot \left(\frac{3}{4}\right)^d.$$

We can then plug this into the original expression to get

$$E[S] \geq n - \frac{n^3}{2} \cdot \left(\frac{3}{4}\right)^d.$$

We can take the derivative of this with respect to n to find the maximum.

$$\frac{d}{dn}\left[n - \frac{n^3}{2} \cdot \left(\frac{3}{4}\right)^d\right] = 1 - 2^{-2d-1}3^{d+1}n^2 = 0$$

$$n = \sqrt{\frac{2^{2d+1}}{3^{d+1}}}$$

So, we can plug this back into the original function to get

$$E[S] \geq \sqrt{\frac{2^{2d+3}}{3^{d+3}}}.$$

This is then a lower bound on the size of an acute subset in $\mathbb{R}^d$.

□

I don't know about you, but the fact that we can approach this problem and find bounds on the size of acute sets is incredibly interesting and powerful. More on this type of work is contained in [5], which was done by another high school researcher and generalizes the definition of acute to other angles.

10. The Symmetric Local Lemma

Remember way back when, I said "we don't need much more technical probability jargon than this?"

I didn't lie, but we will need the following lemma.

Lemma 4 ((Symmetric) Lovász Local Lemma)**.** Let $E_1, \ldots, E_n$ be events such that $P[E_i] \leq p$ for $i = 1, \ldots n$ and that each event is independent (i.e. the probability of an event is not influenced by another) of all the other events except for at most d of them. Then, if

$$ep(d+1) \leq 1,$$

there is a nonzero probability that none of these events occur.

The proof of this is omitted. We will refer to this as the LLL in later proofs.

Problem. Let $G = (V, E)$ be a graph with maximum degree d and let $V_1, V_2, \ldots, V_s$ be a collection of s pairwise disjoint subsets of V with $|V_i| \geq 2ed$ for all i. Prove that there is an independent set of G containing precisely one vertex from each V_i.

Pause for a moment and consider how interconnected each vertex subset is with one another, and how this connects to the LLL.

Proof. Fix $G = (V, E)$ to be a graph with maximum degree d. Fix s pairwise disjoint subsets of V to be $V_1, V_2, \dots, V_s$, such that $|V_i| \geq \lceil 2ed \rceil \geq 2ed$ for all $1 \leq i \leq s$. For each i, let V_i' to be an arbitrarily chosen subset of V_i of size $\lceil 2ed \rceil$.

Form a random set S of vertices by uniformly and independently selecting one vertex from each V_i'. For each edge f, let A_f be the event that both endpoints of f are in S. Then, $P[A_f] \leq \frac{1}{\lceil 2ed \rceil^2}$, so let the p, which we will use for the LLL, be $p = \frac{1}{\lceil 2ed \rceil^2}$.

Note that A_f is mutually independent of all $A_{f'}$, except those for which f and f' both have an endpoint in some V_i'. As there are $2\lceil 2ed \rceil$ vertices in the set $V_i' \cup V_j'$, which are the sets that f has endpoints in. Each of these vertices has maximum degree d, meaning that there are at most $2d\lceil 2ed \rceil$ such f'. So, we can take $a = 2d\lceil 2ed \rceil - 1$. We subtract 1 because otherwise, we double-count the edge f.

Because $ep(a+1) = \frac{2ed\lceil 2ed \rceil}{\lceil 2ed \rceil^2} = \frac{2ed}{\lceil 2ed \rceil} \leq 1$, by the LLL, it is possible to choose S such that it contains one vertex from $V_1, V_2, V_3 \dots$ such that S is an independent set in G. □

Let's return to a more artistic problem.

Problem. Let $H = (V, E)$ be a hypergraph in which every edge has at least k elements, and suppose that each edge of H intersects at most d other edges. If $e(d+1) \leq 2^{k-1}$, then H is two-colorable.

Yep, that's the LLL's doing all right!

Proof. Fix $H = (V, E)$ to be a hypergraph such that every edge has at least k elements and each edge of H intersects at most d other edges.

Let S be a uniformly independently randomly coloring on V with the two colors red and blue. Then, for every edge f, let the event A_f such that the edge f is monochromatic. Then, $P[A_f] \leq 2^{1-k} = p$.

Note that A_f is mutually independent of all $A_{f'}$, except for those for which $|f \cap f'| > 0$. The set of events $A_{f'}$ such that $|f \cap f'| > 0$ and $f \neq f'$ is of size at most d, as each edge intersects at most d other edges.

Then $e(d+1) \leq 2^{k-1}$, so by the LLL, the H is two colorable. □

11. More LLL Practice.

Up until now, we have talked about graph theory, some number theory, some linear algebra, etc. This next problem truly demonstrates TPM's applicability.

Problem. Satisfiability is an important problem in Computer Science, asking the following. Given a logical formula—a collection of booleans combined with AND's ($\wedge$), OR's ($\vee$), and

NOT's (¬)—is there an assignment of TRUE or FALSE to each variable so that the formula is true.

In one formulation of this problem (called k-SAT), the formula takes the form of an AND or clauses, where each clause is an OR of exactly k variables or their negation. For example

$$\overbrace{\underbrace{x_1 \vee \neg x_3 \vee \neg x_{23} \vee \ldots \vee x_{101}}_{k \text{ terms}} \wedge \underbrace{\neg x_2 \vee \ldots \vee x_{13}}_{k \text{ terms}} \wedge \ldots}^{\text{any number of clauses}}.$$

State a condition on formulae of this form that guarantees that a satisfying variable assignment exists, then use the (symmetric) LLL to prove it!

Proof. You want each clause to be true in order for the AND statements to collapse to a true output. Fix $k \geq 1$. Fix a formula F with n clauses to look something like:

$$\overbrace{\underbrace{x_1 \vee \neg x_3 \vee \neg x_{23} \vee \ldots \vee x_{101}}_{k \text{ terms}} \wedge \underbrace{\neg x_2 \vee \ldots \vee x_{13}}_{k \text{ terms}} \wedge \ldots}^{n \text{ number of clauses}}.$$

For each term x_i in this formula, randomly pick it to be true with probability half, and false with probability half. Then, for a clause C, denote the event A_C as the event that C is false. Then, $P[A_C] = 2^{-k}$, as this is the probability that every element in the clause is false. Note that this is not the case if terms are repeated within the clause.

Then, $P[\text{A set of terms exist that evaluate the formula to true}] = 1 - P[\forall C \in F, A_C] \leq 1 - n2^{-k}$. If this probability is nonzero, then

$$\begin{aligned} 0 < 1 - n2^{-k} &\leq 1 \\ n2^{-k} < 1 &\leq 1 + n2^{-k} \\ n < 2^k &\leq 2^k + n. \end{aligned}$$

This means if there are $n \leq 2^k$ clauses in F where there exists a solution set that evaluates it to true.

In order for the expression from the LLL to evaluate to true, we would need $d \leq \frac{2^k}{e} - 1$, where d represents how many other clauses different from C contain at least one of the same terms. This then tells us how many times we can use a term in order to still see the statement evaluated to true. □

12. Matching Sums

Note we have only ever had to use expectation, probability, and independence until now. But as any statistics student will know, variance is lurking in the corner.

Definition. The variance of a real random variable X is defined to be

$$\text{Var}(x) = E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

The standard deviation of X is written as σ_X and is the square root of the variance.

Variance is sadly not a linear operator. In general, it measures how wide of a spread X can achieve. We will also define the covariance.

Definition. The covariance of real random variables X, Y is defined to be

$$\mathrm{Cov}(X, Y) = E[(X - E[X](Y - E[Y])] = E[XY] - E[X]E[Y].$$

The covariance of a real random variable with itself is its own variance.

Note that if X, Y are independent, then their covariance is zero, but the opposite is not true. So, covariance is somewhat of a measure of how related to variables are. Sadly, variance is not linear, but we have a formula for it in terms of covariance. Use the expectation definitions of Var + Cov to prove the following lemma:

Lemma 5. Let X_i be random variables for $i = 1, \ldots n$. Then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j).$$

Finally, we will use a final probability lemma that is sadly not taught in high school courses like AP Statistics, even though it can be built using machinery from those classes.

Lemma 6 (Chebyshev Inequality)**.** Let X be a random variable. Then, for any $t > 0$, we have

$$P[|X - E[x]| \leq t] \leq \frac{\mathrm{Var(X)}}{t^2}.$$

Proof. We have the short proof from the definitions:

$$\mathrm{Var}[X] = E[(X - E[X])^2] \geq t^2 P[|X - E[X]| \geq t].$$

□

We end on a similar problem as the linear algebra one from before.

Problem. Let $v_1 = (x_1, y_1), \ldots, v_n = (x_n, y_n)$ be a collection of n vectors in $\mathbb{Z}^2$, where each x_u and each y_i is an integer with absolute value less than $m = \frac{2^{n/2}}{6\sqrt{n}}$.

Use the probabilistic method to show that there exist two distinct sets $I, J \subseteq \{1, \ldots, n\}$ such that

$$\sum_{i \in I} v_i = \sum_{j \in J} v_j.$$

Proof. Fix n vectors in $\mathbb{Z}^2$ $v_1 = (x_1, y_1), \ldots, v_n = (x_n, y_n)$ such that each x_i and y_i are integers with absolute value less than $m = \frac{2^{n/2}}{6\sqrt{n}}$.

Randomly choose a subset of the vectors to be S by independently randomly choosing each element with probability $\frac{1}{2}$. Let V be the sum of vectors within S. Let X and Y be the x and y component of V. Let X_i be the random variable that v_i is in S, and if it is, then it has value x_i, otherwise, it has value 0.

By Chebyshev,

$$\begin{aligned} P\left[|X-\mu_X| \geq k\right] &\leq \frac{\sigma_X^2}{k^2} \\ P\left[|X-\mu_X| \leq k\right] &\geq 1-\frac{\sigma_X^2}{k^2} \\ \sigma_X &= \sqrt{\mathrm{Var}(X)} \\ \mathrm{Var}(X) &= \mathrm{Var}\left(\sum X_i\right) \\ &= \sum_i \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j) \\ &= \sum_i \frac{x_i^2}{4} \leq \frac{nm^2}{4} = \frac{2^n}{144} \\ \sigma_X &\leq \frac{2^{n/2}}{12} \end{aligned}$$

We can do the same process for Y and then combine the two to get

$$P\left[|X-\mu_X| \leq k \cup |Y-\mu_Y)| \leq k\right] \geq 1-\frac{2^{n+1}}{144k^2}.$$

This is a lower bound. Assume that all the subsets of S sum up to different lattice points. Then the probability of any specific lattice point being in the set of sums of subsets of the vectors is either 0 or $\frac{1}{2^n}$. Then, an upper bound on the probability that the sum falls within k of the mean in both the x and y directions is $\frac{(2k+1)^2}{2^n}$, as there are at most $(2k+1)^2$ lattice points within range k, and it is an upper bound because not all of these points are necessarily inside the set of the sums of subsets of vectors. If there exists a value of k such that

$$\frac{(2k+1)^2}{2^n} < 1-\frac{2^{n+1}}{144k^2},$$

then there exists two distinct sets of vectors such that their sums are equivalent, as this would contradict the assumption that all the sums of subsets of vectors are distinct.

Whenever

$$k = \frac{2^{(n+1)/2}}{6},$$

this is true. This simplifies the expression into

$$\begin{aligned} \frac{\left(\frac{2^{(n+1)/2}}{3}+1\right)^2}{2^n} &< 1 - \frac{2^{n+1}}{2^{n+3}} = \frac{3}{4} \\ \left(\frac{2^{(n+1)/2}}{3}+1\right)^2 &< 3\cdot 2^{n-2} \\ 2^{(n+1)/2}+3 &< 2^{(n-2)/2}3^{3/2} \\ 6 &< 2^{n/2}(3^{3/2}-2^{3/2}) \\ 2 &< 3^{3/2}-2^{3/2} \\ 6 &< 2^{n/2+1} \end{aligned}$$

This is always true for all $n \geq 3$. When $n = 1, 2$, each v_i can only be the 0 vector, which is the trivial case of this problem. So we can always find two distinct subsets such that their sum is the same.

□

13. Conclusion.

Alon and Spencer put it best when they said "Paul Erdős was a searcher, a searcher for mathematical truth." I personally believe that the probabilistic method is not just a mindset but a *mindset*, that sometimes seemingly unfathomable problems simply require a little bit of chance to make them more approachable.

There are of course some problems with purely probabilistic solutions and those with purely constructive ones. The author suggests the reader search out for these and then ponder on the why. TPM is clearly pervasive, and the author personally believes that many more problems will be solved, bounds tightened, and existences proved through good utilization of this method.

References

[1] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[2] Carl Georg Heise, Konstantinos Panagiotou, Oleg Pikhurko, and Anusch Taraz. Coloring d d-embeddable k k-uniform hypergraphs. *Discrete & computational geometry*, 52:663–679, 2014.

[3] Hans Ulrich Simon. Minimum tournaments with the strong *s_k*-property and implications for teaching. *arXiv preprint arXiv:2205.08357*, 2022.

[4] Marcelo Campos, Simon Griffiths, Robert Morris, and Julian Sahasrabudhe. An exponential improvement for diagonal ramsey. *arXiv preprint arXiv:2303.09521*, 2023.

[5] Tanya Otsetarova and Zhulin Li. Boundaries on the number of points in acute sets, 2017.

Bay Area, CA
Email address: agnivsarkar@proofschool.org