THE PROBABILISTIC METHOD

ISAAC SUN

Abstract. We will discuss the fundamentals of the probabilistic method. Then, we proceed with some basic methods and some background material. Next we discuss different methods of the probabilistic method, including Linearity of Expectation, alterations, the second moment, and the Local Lemma. We conclude the paper with a brief conclusion of the importance of the probabilistic method.

1. Introduction

Fundamentally, we think about proofs very thoroughly. Often times, to prove a statement, we have to prove that it always works under certain conditions imposed, or to disprove something we find a counterexample to the statement. This usually makes proofs regarding probabilities difficult, because there are indeterminate variables. But what if we could prove the existence of something non-constructively? The fundamental idea of the probabilistic method is to non-constructively prove the existence of something by calculating the probability or expected value, and using it in a way to show that some sort of combinatorial structure must exist or not exist. Possibly the easiest way to understand the probabilistic method is to look at one of its most famous examples, the Ramsey numbers.

Definition 1.1 (Ramsey Number)**.** The Ramsey number $R(k, l)$ is the smallest integer n such that for every 2-coloring of the edges of a complete graph K_n (each pair of vertices is connected by an edge) red and blue, there exists a red K_k or a blue K_l.

In 1929, Ramsey proved that $R(k, l)$ is finite for any two integers (k, l). Let us find a lower bound for the diagonal Ramsey number $R(k, k)$

Theorem 1.2 (Erdős 1947)**.** *If* $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$*, then* $R(k, k) > n$*. Thus, for all* $k \geq 3$*,* $R(k, k) > \lfloor 2^{k/2} \rfloor$*.*

Proof. Let's take a look at a random 2-coloring of K_n where each edge is independently colored with equal probability. For any set R of k vertices on our probability space K_n, let A_R be the event that this subgraph is monochromatic (edges are all red or all blue). Since there are two colors and $\binom{k}{2}$ edges,

$$\Pr[A_R] = 2 \cdot \left(\frac{1}{2}\right)^{\binom{k}{2}} = 2^{1-\binom{k}{2}}.$$

By the union bound,

$$\Pr\left[\bigcup A_R\right] \leq \sum \Pr\left[A_R\right] = \binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$$

Date: July 2023.

Since the union of all possible A_R does not fill up the entire probability space,

$$\Pr\left[\bigcap \overline{A_R}\right] > 0.$$

With positive probability, none of the events A_R occurred. Hence, a 2-coloring of K_n without a monochromatic K_k exists, meaning $R(k,k) > n$. Note that if $k \geq 3$ and we let $n = \lfloor 2^{k/2} \rfloor$,

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

Thus, $R(k,k) > \lfloor 2^{k/2} \rfloor$ for all $k \geq 3$. □

The diagonal Ramsey numbers highlights the core of the probabilistic method, using probability as a way to prove this lower bound. However, a question worth noting is why we used probability instead of simply counting. Using the fact that the total number of 2-colorings of K_n is bigger than the number of monochromatic K_k would work just as well. And although most combinatorial problems deal with finite probability spaces, it is not always possible to replace our probability arguments with counting arguments, even in finite probability spaces. From this point onward, we'll be looking into the different applications and methods of the probabilistic method.

2. Basic Methods

Definition 2.1 (Tournament)**.** A tournament on n players is an orientation of the edges of K_n. A tournament T has the property S_k if for every set of k players there is a player who beats them all.

Theorem 2.2. *If* $\binom{n}{k}(1-2^{-k})^{n-k} < 1$*, there is a tournament on* n *vertices that has the property* S_k

Proof. Let's take a look at a random tournament on n vertices where each edge is oriented independently with equal probability. Let A_R be the event that there is no vertex that beats them all.

$$\begin{aligned} \Pr[A_R] &= \prod \Pr[\text{v does not beat them all}] \\ &= \prod (1 - \Pr[\text{v beats them all}]) \\ &= \prod (1 - 2^{-k}) \\ &= (1-2^{-k})^{n-k}. \end{aligned} \tag{2.1}$$

Now by the union bound,

$$\Pr\left[\bigcup A_R\right] \leq \sum \Pr[A_R] = \binom{n}{k} \cdot (1-2^{-k})^{n-k} < 1.$$

Thus,

$$\Pr\left[\bigcap \overline{A_R}\right] > 0.$$

With positive probability, none of the events A_R occur; hence, there exists a tournament on n vertices with the property S_k.

□

Let's look at one more example:

Definition 2.3 (Dominating Set)**.** A dominating set of a graph G is a set $U \subseteq V(G)$ such that every vertex in $V(G) \setminus U$ has a neighbor in U.

Theorem 2.4. *If G is a graph on n vertices with minimum degree $\delta > 1$, then G has a dominating set of at most $n \cdot \frac{1+\ln \delta+1}{\delta+1}$.*

Here, if we take a small random subset X of $V(G)$, we really don't know much about any dominating sets. Here, we use the idea of an alteration; instead of only looking at random subsets of our graph, we also add on the set of undominated vertices Y to set an upper bound on the dominating set.

Proof. Let's pick each vertex independently with $p \in [0, 1]$. $\Pr[v \in Y] \leq (1-p)^{\delta+1}$, since G has minimum degree δ. By Linearity of Expectation,

$$E[|Y|] = \sum \Pr[v \in Y] \leq n(1-p)^{\delta+1}.$$

Since $(1-p) \leq e^{-p}$,

$$\begin{aligned} E[|X| + |Y|] &= E[|X|] + E[|Y|] \\ &\leq np + n(1-p)^{\delta+1} \\ &\leq n(p + e^{-p(\delta+1)}) \end{aligned} \tag{2.2}$$

Taking the derivative of this expression with respect to p and setting it to zero gives us that this expression is minimized at $p = \frac{\ln \delta+1}{\delta+1}$. Hence, there exists some X such that

$$|X| + |Y| \leq E[|X| + |Y|] \leq n \cdot \frac{1 + \ln \delta + 1}{\delta + 1},$$

and $X \cup Y$ is a dominating set of G as desired. □

The two key takeaways from this proofs is the application of *Linearity of Expectation*, and the addition of some set Y, which is a method called an *Alteration*, where a random outcome is altered to gain the desired outcome.

3. Preliminaries

Before we proceed, let us briefly go over some probability theory.

Definition 3.1 (Probability Space)**.** A *probability space* is a triple (Ω, Σ, P), where Ω is a sample space, Σ is a σ-algebra on Ω, and P is a probability measure on Σ. Elements A_i on Σ are *events*, and $\Pr[A]$ for some event A is the *probability* of A (Note that $\Pr[]$ and $P()$ both denote probability).

For example, given a finite graph where you randomly choose edges to color white or black, Ω would be the set of all possible colorations, and Σ would be the set of all subsets of Ω so that every event is measurable. The probability measure could give an equal probability to every coloring of the graph.

Lemma 3.2 (Union bound)**.** *For a countable set of events $A_1, A_2, \ldots$*

$$\Pr\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \Pr[A_i].$$

Proof. One of the axioms of a probability space states that for disjoint events $B_1, B_2, \ldots$,

$$\Pr\left[\bigcup B_i\right] = \sum \Pr[B_i].$$

Now to modify A_i such that they are disjoint, let

$$B_i = A_i - \bigcup_{j=i}^{i-1} A_j.$$

Note that if

$$x \in \bigcup_{i=1}^{\infty} B_i$$

then $x \in B_k$ for some k, and

$$B_k = A_k - \bigcup_{j=1}^{k-1} A_j,$$

so $x \in A_k$, meaning

$$\bigcup_{i=1}^{\infty} A_i \subset \bigcup_{i=1}^{\infty} B_i.$$

Similarly, if

$$x \in \bigcup_{i=1}^{\infty} A_i,$$

then $x \in A_k$ for some minimum k such that $i < k$, which means $x \notin A_i$. Thus,

$$x \in B_k = A_k - \bigcup_{j=1}^{k-1} A_j,$$

and so,

$$\bigcup_{i=1}^{\infty} B_i \subset \bigcup_{i=1}^{\infty} A_i \implies \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

Since $B_i \subset A_i$, $\Pr[B_i] \leq \Pr[A_i]$, and

$$\Pr\left[\bigcup A_i\right] = \Pr\left[\bigcup B_i\right] = \sum \Pr[B_i] \leq \sum \Pr[A_i].$$

□

Definition 3.3 (Random Variable)**.** A random variable on a probability space (Ω, Σ, P) is a function $X : \Omega \to \mathbb{R}$ that is $\mathcal{F}$-measurable. Almost all of the random variables will be discrete in this paper, meaning that they only take up a countable number of outcomes.

Definition 3.4 (Expected Value (Expectation))**.** The expected value of a random variable X is denoted as $E[X]$. More formally, for random variables on finite probability spaces,

$$E[X] = \sum_{\omega \in \Omega} \Pr[\omega] X(\omega).$$

With this out of the way, let's look at some methods of the probabilistic method.

4. Linearity of Expectation

Theorem 4.1 (Linearity Of Expectation)**.** *Given a discrete random variable $X = \sum_i c_i X_i$ for random variables X_i,*

$$E[X] = \sum_i c_i E[X_i].$$

Proof. Let us prove that $E[X+Y] = E[X] + E[Y]$ for random variables X and Y, and extend the result through induction. By definition,

$$\begin{aligned}
E[X+Y] &= \sum_x \sum_y [(x+y) \cdot P(X=x, Y=y)] \\
&= \sum_x \sum_y [x \cdot P(X=x, Y=y)] + \sum_x \sum_y [y \cdot P(X=x, Y=y)] \\
&= \sum_x x \sum_y P(X=x, Y=y) + \sum_y y \sum_x P(X=x, Y=y) \\
&= \sum_x x \cdot P(X=x) + \sum_y y \cdot P(Y=y) \\
&= E[X] + E[Y].
\end{aligned}$$

Note note that each step is repeatable for each given variable added, which concludes our proof. □

Although it may not be intuitive, this holds even if X_i are not independent! Often times we will be showing that there must exist some X on our probability space such that $X \leq E[X]$ or $X \geq E[X]$.

For example, consider the following HMMT Problem:

Question 4.2 (HMMT 2006)**.** *At a nursery,* 2006 *babies sit in a circle. Suddenly each baby randomly pokes either the baby to its left or to its right. What is the expected value of the number of unpoked babies?*

Solution 4.3. *Number the babies arbitrarily* $1, 2, \ldots, 2006$*. Let X_i be the indicator variable such that*

$$X_i = \begin{cases} 1 & \text{if baby } i \text{ is unpoked} \\ 0 & \text{otherwise.} \end{cases}.$$

Now note that $E[X_i] = \left(\frac{1}{2}\right)^2$ for each i since there is a $\frac{1}{2}$ chance that a baby misses and each baby has two babies on either side.

$$E[X_1 + \ldots X_{2006}] = E[X_1] + \ldots E[X_{2006}] = 2006 \cdot \frac{1}{4} = \frac{1003}{2}.$$

Definition 4.4 (Hamiltonian Path)**.** A Hamiltonian path in a tournament T is a directed path that includes all vertices of T.

Theorem 4.5 (Szele 1943)**.** *There exists a tournament T with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths for all positive integers n.*

Proof. Let X be the number of Hamiltonian paths in T, and for every permutation σ of $[n]$, let X_σ be the indicator random variable for whether $\sigma(1)\sigma(2)\dots\sigma(n)$ is a Hamiltonian path. By Linearity of Expectation,

$$E[X] = \sum E[X_\sigma] = n!2^{-(n-1)}.$$

Thus there exists some T such that T has at least $n!2^{-(n-1)}$ Hamiltonian paths.

□

This proof is such a powerful idea; we don't know anything about our tournament, yet we can exploit the randomness to calculate the expected value to show the existence of some lower bound of the number of Hamiltonian paths. For the curious, Alon (1990) proved that the upper bound for the number of Hamiltonian paths in a tournament with n players to be $\frac{n!}{(2-\sigma(1))^n}$.

Definition 4.6 (Bipartite Graph)**.** A bipartite graph G is a graph whose vertices can be divided into two disjoint and independent sets U and V such that every edge connects a vertex from U to one in V.

Theorem 4.7. *Every graph G has a bipartite subgraph with at least $\frac{|E(G)|}{2}$ edges, where $E(G)$ denotes the number of edges of G.*

Proof. Let T be a random subset of the vertices of T where each vertex is chosen independently with a probability of $1/2$. Let H be the graph with the same vertices of G and edges that only contain one vertex in T. Note that H is a bipartite subgraph of G. Since each edge has a $1/2$ probability of being chosen,

$$\Pr[e \in E(H)] = \frac{1}{2}.$$

The probability that some edge $\{x, y\}$ is in $E(H)$ is the sum of the probability that $x \in T$ and $y \notin T$ plus the probability that $x \notin T$ and $y \in T$, which is just $2 \cdot \left(\frac{1}{2}\right)^2$. By Linearity of Expectation,

$$E[|E(H)|] = \sum_{e \in E(G)} \Pr[e \in E(H)] = \sum_{e \in E(G)} \frac{1}{2} = \frac{|E(G)|}{2}$$

as desired. □

Another example of using Linearity of Expectation is actually with vectors!

Theorem 4.8 (Vector Balancing)**.** *Let $v_1, \dots, v_n \in \mathbb{R}^n$, with all $|v_i| = 1$. There exist $\epsilon_1, \dots, \epsilon_n = \pm 1$ such that*

$$|\epsilon_1 v_1 + \dots + \epsilon_n v_n| \leq \sqrt{n}$$

and there also exist $\epsilon_1, \dots, \epsilon_n = \pm 1$ such that

$$|\epsilon_1 v_1 + \dots + \epsilon_n v_n| \geq \sqrt{n}.$$

Proof. Let $\epsilon_1, \dots, \epsilon_n$ be chosen independently and uniformly, and let $X = |\epsilon_1 v_1 + \dots + \epsilon_n v_n|^2 = \sum_{i=1}^n \sum_{j=1}^n \epsilon_i \epsilon_j v_i \cdot v_j$. Thus

$$E[X] = \sum_{i=1}^n \sum_{j=1}^n v_i \cdot v_j E[\epsilon_i \epsilon_j] = \sum_{i=1}^n \sum_{j=1}^n (v_i v_j) E[\epsilon_i \epsilon_j].$$

When $i \neq j$,

$$E[\epsilon_i \epsilon_j] = E[\epsilon_i]E[\epsilon_j] = 0,$$

and when $i = j$,

$$E[\epsilon_i^2] = 1.$$

Thus,

$$E[X] = \sum_{i=1}^{n} |v_i|^2 = n$$

which means that there exists ϵ_i such that $X \leq n$ and $X \geq n$. Taking the square root finishes our proof. □

There's also a pretty nice generalization of balancing weighted vectors as follows:

Theorem 4.9 (Balancing Weighted Vectors)**.** *Let $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$, with all $|v_i| \leq 1$. Let $p_1, \ldots, p_n \in [0,1]$ and $w = p_1 v_1 + \ldots + p_n v_n$. Then there exist $\epsilon_1, \ldots, \epsilon_n \in \{0,1\}$ such that*

$$|w - v| \leq \frac{\sqrt{n}}{2},$$

where $v = \epsilon_1 v_1 + \ldots + \epsilon_n v_n$.

Proof. Remark: $p_i = 1/2$ for $1 \leq i \leq n$ is equivalent to the theorem above.

As for the idea behind this proof, we want to choose ϵ_i independently such that

$$\Pr[\epsilon_i = 1] = p_i, \Pr[\epsilon_i = 0] = 1 - p_i.$$

Then, let $X = |w - v|^2$ and show that $E[X] \leq \frac{n}{4}$, then square root to finish it off like before. □

5. Alterations

Aside from Linearity of Expectation, one of the most powerful techniques of the probabilistic method is the use of alterations. The idea is pretty simple: alter a random outcome to get the desired object.

Theorem 5.1 (Better Ramsey Number Lower Bound)**.** *For any positive integer n,*

$$R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}$$

Proof. Consider a random 2-coloring of K_n. Let X be the number of monochromatic K_k. Then,

$$E[X] = \binom{n}{k} 2^{1-\binom{k}{2}}$$

so there exists some two coloring for which $X \leq E[X]$. Now, we fix this coloring by removing one vertex from each monochromatic k-set from K_n. The coloring on the set of $n - \binom{n}{k} 2^{1-\binom{k}{2}}$ vertices now has no monochromatic $k - set$, giving us our desired, better lower bound. □

But what about off-diagonal Ramsey Numbers?

Theorem 5.2 (Off-Diagonal Ramsey Numbers)**.** *If there exists $p \in [0,1]$ such that*

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

then $R(k,l) > n$.

Equivalently, with the same previous alteration for diagonal Ramsey Numbers, we can get this better bound as follows:

Theorem 5.3. *For all $p \in [0,1]$ and positive integers n,*

$$R(k,l) > n - \left(\binom{n}{k}p^{\binom{k}{2}} + \binom{n}{l}(1-p)^{\binom{l}{2}}\right)$$

Proof. In both theorems, consider a random 2-coloring of K_n by coloring each edge independently red with probability p. Let X be the number of red k-sets plus the number of blue l-sets. By Linearity of Expectation,

$$E[X] = \binom{n}{k}p^{\binom{k}{2}} + \binom{n}{l}(1-p)^{\binom{l}{2}}.$$

For theorem 3.2, $E[X] < 1$, meaning there exists a two-coloring with $X = 0$. And with the same logic of removing points from "bad" sets, we can get a lower bound with $n - E[X]$ points with no bad sets. □

Theorem 5.4 (Turán's Theorem)**.** *If G is a K_r-free graph on n vertices, then*

(a)$|E(G)| \leq (1 - \frac{1}{r-1})\binom{n}{2}$,

(b) max achieved by the unique complete $(r-1)$-partite graph with part sizes nearly equal.

Although we will not prove the entire theorem, we can prove a weak form of part a:

Theorem 5.5 (Half of Turán's)**.** *Let $G = (V, E)$ have n vertices and $\frac{nd}{2}$ edges, for $d \geq 1$. Then*

$$\alpha(G) \geq \frac{n}{2d},$$

where $\alpha(G)$ denotes the independence number.

Proof. Let S be a random subset of V where each vertex is in S independently with probability p, and let $X = |S|$ and $Y = |E(G[S])|$. Thus,

$$E[X] = np$$

and

$$\Pr[e \in G[S]] = p^2.$$

By Linearity of Expectation,

$$E[Y] = \sum_{e \in E(G)} \Pr[e \in G[S]] = |E(G)|p^2 = \frac{nd}{2}p^2.$$

Now applying Linearity of Expectation again, we get

$$E[X - Y] = np - \frac{nd}{2}p^2.$$

Letting $p = \frac{1}{d}$ maximizes

$$E[X - Y] = \frac{n}{2d}.$$

Hence, there exists X with $X - Y \geq E[X - Y] = \frac{n}{2d}$. Deleting one end of each edge in $G[S]$ gives

$$\alpha(G) \geq X - Y \geq \frac{n}{2d}.$$

Remark:

$$|E(\overline{G})| \geq \frac{n^2}{4\alpha(\overline{G})} \geq \frac{n^2}{4(r-1)}$$

□

Definition 5.6 (Packing)**.** Let C be a bounded measurable subset of $\mathbb{R}^d$ and $B(x)$ be the cube $[0, x]^d$ of side x. A *packing* of C into $B(x)$ is a family of pairwise disjoin copies of C, all lying inside $B(x)$. Let $f(x)$ denote the size of a largest such family.

The *packing constant*

$$\delta(C) = \mu(C) \lim_{x \to \infty} \frac{f(x)}{x^d},$$

where $\mu(C)$ is the measure of C.

Theorem 5.7. *Let C be bounded ($\exists w, |c| \leq w \forall c \in C$), convex (if $c_1, c_2 \in C$, then $pc_1 + (1-p)c_2 \in C \forall p \in [0,1]$), and centrally symmetric (if $c_1 \in C$, then $-c_1 \in C$). Then*

$$\delta(C) \geq \frac{1}{2^{d+1}}.$$

Proof. Let $P_1, \ldots, P_n$ be independently and uniformly selected from $B(x)$, and let X be the number of pairs $\{i, j\}$ with $(C+P_i) \cap (C+P_j) \neq \emptyset$. We now calculate $\Pr[(C+P_i) \cap (C+P_j) \neq \emptyset]$ as follows:

If $(C + P_i) \cap (C + P_j) \neq \emptyset$, then there exist $c_1, c_2 \in C$ with

$$c_1 + P_i = c_2 + P_j,$$

or equivalently

$$P_j - P_i = c_1 - c_2 = 2\frac{c_1 - c_2}{2}.$$

Since $-c_2 \in C$, $\frac{c_1}{2} + \left(\frac{-c_2}{2}\right) \in C$ by convexity,

$$P_j \in P_i + 2C.$$

Hence

$$\Pr[(C + P_i) \cap (C + P_j) \neq \emptyset] \leq \Pr[P_j \in P_i + 2C] \leq \frac{\mu(2C)}{x^d} = \frac{2^d \mu(C)}{x^d}.$$

Now by Linearity of Expectation,

$$E[X] \leq \frac{n^2}{2} \cdot \frac{2^d \mu(C)}{x^d}$$

and there exists an X such that $X \leq E[X]$. Removing one copy from each intersecting pair leaves

$$n - X \geq n - n^2 2^{d-1} x^{-d} \mu(C)$$

pairwise disjoint copies. Letting $n = 2^{-d}x^d/\mu(C)$ maximizes this expression giving

$$2^{-(d+1)}x^d/\mu(C)$$

pairwise disjoint copies.These copies may not all lie inside of $B(x)$, but they all lie inside of $B(x+2w)$, where $w = \max_{c\in C}|c|$.
Thus,

$$f(x+2w) \geq 2^{-(d+1)}x^d/\mu(C),$$

meaning that

$$\delta(C) \geq \lim_{x\to\infty} \frac{\mu(C)f(x+2x)}{(x+2w)^{-d}} \geq 2^{-(d+1)}$$

□

6. Hypergraphs

Definition 6.1 (Hypergraph)**.** A hypergraph $G = (V, E)$ is a graph such that edges can join any number of vertices, whereas edges in graphs can only join two vertices.

Definition 6.2 (Property B)**.** A hypergraph $H = (V, E)$ has *property B*, also known as 2-colorable, if there is a 2-coloring of V such that no edge is monochromatic. Let $m(n)$ denote the minimum number of edges in an $n-$uniform hypergraph that does not have property B.

Lower bounds: $m(n) \geq$

- 2^{n-1} (Erdős 1963)
- $\Omega(2^n n^{\frac{1}{3}})$ (Beck 1978)
- $\Omega(2^n\sqrt{n/\ln n})$ (Radhakrishnan and Srinivasan 2000)

Let's present a proof by Cherkashin and Kozik (2015). To prove this theorem, we must first define a conflicting pair.

Definition 6.3. For a $n-$uniform hypergraph (all hyperedges have size n), if σ is an ordering of $V(H)$ and $e, f \in E(H)$, then (e, f) is a conflicting pair under σ if the last vertex of e is the first vertex of f.

Lemma 6.4. *If there exists an ordering σ of $V(H)$ with no conflicting pairs, then H is 2-colorable.*

Proof. Color a vertex red if it is the last vertex of an edge under σ, and blue otherwise. Then there are no monochromatic blue edges because the last coloring is always red, and there are no monochromatic red edges because there are no conflicting pairs. □

Theorem 6.5. *If there exists $p \in [0, 1]$ with*

$$k(1-p)^n + k^2 p < 1,$$

then $m(n) > 2^{n-1}k$.

Proof. Let H be a hypergraph with $|E(H)| = 2^{n-1}k$ with probability p that satisfy the condition. Now to create a random ordering of $V(H)$, $\forall v \in V(H)$, let $x_v \in [0, 1]$ be chosen independently and uniformly (note that x_v are all distinct with probability 1, which creates

the ordering of $V(H)$). Now let us split $[0,1]$ into 3 different intervals to check for conflicting pairs:

$$L = \left[0, \frac{1-p}{2}\right], M = \left(\frac{1-p}{2}, \frac{1+p}{2}\right), R = \left[\frac{1+p}{2}, 1\right].$$

Then for $e \in E(H)$,

$$\Pr[e \in L] = \Pr[e \in R] = \left(\frac{1-p}{2}\right)^n.$$

Thus,

$$\sum_{e \in E(H)} \Pr[e \in L \text{ or } e \in R] = 2^{n-1}k \cdot 2 \cdot \left(\frac{1-p}{2}\right)^n = k(1-p)^n.$$

which is the probability that no edge is entirely in the left or entirely in the right. Now, the only way to get a conflicting pair is such that the intersection of $e \cap f = v$ is in the middle, and the rest of e is before v, and the rest of f is after v.

$\forall e, f \in E(H)$ with $e \cap f = v$, let

$$A_{e,f} = v \in M \text{ and } (e,f) \text{ is conflicting.}$$

$$\Pr[A_{e,f}] = p x_v^{n-1}(1-x_v)^{n-1} \leq p\left(\frac{1}{4}\right)^{n-1}.$$

Thus,

$$\sum_{e,f \in E(H)} \Pr[A_{e,f}] \leq (2^{n-1}k)^2 p\left(\frac{1}{4}\right)^{n-1} = k^2 p.$$

Hence,

$$\Pr[\exists \text{ conflicting pair}] \leq k(1-p)^n + k^2 p < 1.$$

□

Corollary 6.6. $m(n) = \Omega(2^n\sqrt{n/\ln n})$

Proof. Once again, by bounding $(1-p) \leq e^{-p}$, we minimize $ke^{-pn} + k^2 p$. Taking the derivative, we get $-kne^{-pn} + k^2$, which is 0 at $p = \frac{\ln(n/k)}{n}$. Substituting gives

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

and this expression is less than 1 when $k = O(\sqrt{n/\ln n})$. □

7. Second Moment

Definition 7.1 (k-th moment)**.** The *k−th moment* of a random variable X is $E[X^k]$.

Definition 7.2 (Variance)**.** The *variance* of a random variable X is

$$\mathrm{Var}[X] = E[(X - E[X])^2],$$

and the square root of the variance is called the *standard deviation*. It is standard notation to let μ denote the expectation and σ denote the standard deviation.

Theorem 7.3 (Markov's Inequality (for first moment of random variables))**.** *If $X \geq 0$ is a random variable, then $\forall a > 0$,*

$$\Pr[X \geq a] \leq \frac{E[X]}{a}.$$

Proof.

$$\begin{aligned} E[X] &= \sum_i i \Pr[X = i] \\ &\geq \sum_{i \geq a} i \Pr[X = i] \\ &\geq a \Pr[X \geq a]. \end{aligned}$$

□

The second moment involves using the following inequality:

Theorem 7.4 (Chebyshev's Inequality)**.** *For all $\lambda > 0$,*

$$\Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

Proof. Since $(X - \mu)^2 \geq 0$, by Markov's

$$\begin{aligned} \Pr[|X - \mu| \geq \lambda\sigma] &= \Pr[(X - \mu)^2 \geq \lambda^2\sigma^2] \\ &\leq \frac{E[(X - \mu)^2]}{\lambda^2\sigma^2} \\ &= \frac{\sigma^2}{\lambda^2\sigma^2} \\ &= \frac{1}{\lambda^2} \end{aligned}$$

□

Essentially, Chebyshev's states that the probability decreases quadratically in the number of standard deviations, but when X is the sum of nearly independent random variables, exponentially decreasing bounds are obtainable.

Definition 7.5 (Covariance)**.** The *covariance* of two random variables X and Y is

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

If X and Y are independent, then $\mathrm{Cov}[X, Y] = 0$.

Proposition 7.6. *If $X = X_1 + \ldots + X_n$,*

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j].$$

If X_i are indicator random variables with probability p_i, then

$$\mathrm{Var}[X_i] = p_i(1 - p_i) \leq p_i = E[X_i],$$

and thus

$$\mathrm{Var}[X] \leq E[X] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j].$$

Now let's jump into some examples in number theory. In 1920, Hardy and Ramanujan proved that "almost all" numbers n have "very close to" $\ln \ln n$ prime factors. More concretely, they proved the following:

Theorem 7.7 (Hardy and Ramanujan (1920))**.** *Let $\omega(n) \to \infty$ arbitrariliy slowly. Then the number of $x \in [n]$ such that*

$$|v(x) - \ln \ln n| > \omega(n)\sqrt{\ln \ln n}$$

is $o(n)$, where $v(n)$ denotes the number of primes dividing n (without multiplicity).

Proof. Let x be randomly chosen from $[n]$. For p prime, let X_p be the following indicator variable:

$$X_p = \begin{cases} 1 & \text{if } p \mid x \\ 0 & \text{otherwise} \end{cases}.$$

Let $M = n^{1/10}$ and $X = \sum_{p \le M} X_p$. As x has at most 10 prime factors that are greater than M,

$$v(x) - 10 \le X(x) \le v(x).$$

Note that

$$E[X_p] = \frac{\lfloor n/p \rfloor}{n} = \frac{1}{p} + O(1/n).$$

By Linearity of Expectation,

$$E[X] = \sum_{p \le M} \left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right) = \ln \ln n + O(1).$$

As for the variance,

$$\mathrm{Var}[X] = \sum_{p \le M} \mathrm{Var}[X_p] + \sum_{p \ne q} \mathrm{Cov}[X_p, X_q].$$

Note that

$$\mathrm{Var}[X_p] = \frac{1}{p}\left(1 - \frac{1}{p}\right) + O\left(\frac{1}{n}\right).$$

Thus,

$$\sum_{p \le M} \mathrm{Var}[X_p] = \left(\sum_{p \le M} \frac{1}{p} \right) + O(1) = \ln \ln n + O(1).$$

As for covariance, when $p \ne q$,

$$X_p X_q = 1 \iff p|x \text{ and } q|x \iff pq|x.$$

Thus,

$$\begin{aligned} \mathrm{Cov}[X_p, X_q] &= E[X_p X_q] - E[X_p]E[X_q] \\ &= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n} \\ &= O\left(\pm \frac{1}{n}\right). \end{aligned}$$

And so,

$$\sum_{p \ne q} \mathrm{Cov}[X_p, X_q] = M^2 O\left(\pm \frac{1}{n}\right) = O(\pm n^{-8/10}) = o(1).$$

Thus,

$$\mathrm{Var}[X] = \ln\ln n + O(1).$$

By Chebyshev's Inequality,

$$\Pr[|X - \ln\ln n| > \lambda\sqrt{\ln\ln n}] < \frac{1}{\lambda^2} + o(1).$$

□

Something interesting to note is that if $X \geq 0$ and $E[X] \to 0$, then $X = 0$ almost always (essentially with probability 1). However, if $E[X] \to \infty$, it is not always the case that $X > 0$, which is extremely unintuitive. Consider the following:

$$X = \begin{cases} n^2 & \text{with probability } 1/n \\ 0 & \text{otherwise} \end{cases}.$$

However, if the variance is small enough, we can conclude that $X > 0$ almost always.

Theorem 7.8.

$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{E[X]^2}.$$

Proof. Setting $\lambda = \frac{\mu}{\sigma}$ in Chebyshev's Inequality, we get

$$\Pr[X = 0] \leq \Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2}.$$

□

Corollary 7.9. *If* $\mathrm{Var}[X] = o(E[X]^2)$*, then* $X > 0$ *almost always.*

Even better,

Corollary 7.10. *If* $\mathrm{Var}[X] = o(E[X]^2)$*, then* $X \sim E[X]$ *almost always.*

Essentially, the idea of the second moment method is applying variances and covariances to often times show that X is nonnegative almost always or even prove that it's very close to what we expect almost always.

8. Local Lemma

Most of our methods have only showed that an outcome happens with high probability but doesn't cover the entire probability space. Now what if we wanted to actually show that an outcome happens with a low positive probability? One common event in which this occurs is the intersection of independent events. For example, consider $A_1, \ldots, A_n$ independent events with $\Pr[A_i] = p > 0$. Then,

$$\Pr\left[\bigcap_{i=1}^{n} A_i\right] = p^n > 0.$$

To see if this still works if A_i are mostly independent, let's formalize being "mostly independent".

Definition 8.1 (Mutual Independence)**.** An event A in a probability space is mutually independent of a set $\mathscr{B}$ of other events if for each $S \subseteq \mathscr{B}$,

$$\Pr\left[A| \bigcap_{B_i \in S} B_i\right] = \Pr[A].$$

Note that an event may be pairwise independent to each event in $\mathscr{B}$ but not mutually independent. Now the *Lovász Local Lemma* shows that if a set of undesirable events that are mostly mutually independent and happen with low probability, then with positive probability none of them happen.

There are two versions of the Local Lemma:

(1) Symmetric version - probabilities and dependencies are uniformly bounded
(2) General version- probabilities and dependencies may vary

Lemma 8.2 (Lovász Local Lemma (Symmetric))**.** *Let $\mathscr{A} = \{A_1, \ldots, A_n\}$ be a set of events in a probability space. If $\forall i \in [n]$:*

- *$\exists D_i \subset \mathscr{A}$ with $|D_i| \leq d$ such that A_i is mutually independent of $\mathscr{A} \setminus D_i$*
- *$\Pr[A_i] \leq p$*

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

$$\Pr\left[\bigcap_{i=1}^{n} \overline{A_i}\right] > 0.$$

Proof. If $d = 0$, the events are independent and our result immediately follows.

Now when $d \geq 1$, $\forall i \in [n]$, let

$$x_i = \frac{1}{d+1}.$$

$\forall d \geq 1$,

$$\left(1 - \frac{1}{d+1}\right)^d \geq \frac{1}{e}$$

Thus

$$x_i \prod_{A_j \in D_i} (1 - x_j) \geq \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d \geq \frac{1}{e(d+1)} \geq p \geq \Pr[A_i].$$

By the general version,

$$\Pr\left[\bigcap_{i=1}^{n} \overline{A_i}\right] \geq \left(1 - \frac{1}{d+1}\right)^n > 0.$$

□

Lemma 8.3 (Lovász Local Lemma (General))**.** *Let $\mathscr{A} = \{A_1, \ldots, A_n\}$ be a set of events in a probability space. If $\forall i \in [n]$:*

- *$\exists D_i \subset \mathscr{A}$ such that A_i is mutually independent of $\mathscr{A} \setminus D_i$*
- *$x_i \in [0, 1)$ such that*

$$\Pr[A_i] \leq x_i \prod_{A_j \in D_i} (1 - x_j)$$

and $ep(d+1) \leq 1$*, then*

$$\Pr\left[\bigcap_{i=1}^{n} \overline{A_i}\right] \geq \prod_{i=1}^{n}(1-x_i) > 0.$$

Claim 8.4. *For* $T \subseteq [n]$*, let*

$$\mathrm{Not}(T) = \bigcap_{j \in T} \overline{A_j}.$$

$\forall i \in [n]$ *and* $S \subseteq [n] \setminus i$*,*

$$\Pr[A_i | \mathrm{Not}(S)] \leq x_i.$$

Proof. Let's do induction on $|S|$.

$|S| = 0$ holds true since

$$\Pr[A_i] \leq x_i \prod_{A_j \in D_i} (1-x_j) \leq x_i.$$

Now for $S \neq \emptyset$, let

$$S_1 = \{j \in S : A_j \in D_i\}, S_2 = S \setminus S_1.$$

Then,

$$\Pr[A_i | \mathrm{Not}(S)] = \frac{\Pr[A_i \cap \mathrm{Not}(S_1) | \mathrm{Not}(S_2)]}{\Pr[\mathrm{Not}(S_1) | \mathrm{Not}(S_2)]}$$

by conditional probability. Note that

$$\Pr[A_i \cap \mathrm{Not}(S_1) | \mathrm{Not}(S_2)] \leq \Pr[A_i | \mathrm{Not}(S_2)],$$

and since A_i is mutually independent of S_2,

$$\Pr[A_i | \mathrm{Not}(S_2)] = \Pr[A_i] \leq x_i \prod_{A_j \in D_i} (1-x_j).$$

Now suppose $S_1 = \{j_1, \ldots, j_r\}$. Then

$$\begin{aligned}
\Pr\left[\mathrm{Not}\left(S_1\right) \mid \mathrm{Not}\left(S_2\right)\right] &= \prod_{k=1}^{r} \Pr\left[\overline{A_{j_k}} \mid \mathrm{Not}\left(\{j_\ell : \ell < k\} \cup S_2\right)\right] \\
&= \prod_{k=1}^{r} \left(1 - \Pr\left[A_{j_k} \mid \mathrm{Not}\left(\{j_\ell : \ell < k\} \cup S_2\right)\right]\right), \\
&\geq \prod_{k=1}^{r} (1 - x_{j_k}) \geq \prod_{A_j \in D_i} (1-x_j)
\end{aligned}$$

where we used induction with $|\{j_\ell : \ell < k\} \cup S_2| < |S|$. Thus,

$$\Pr[A_i | \mathrm{Not}(S)] \leq \frac{x_i \prod_{A_j \in D_i}(1-x_j)}{\prod_{A_j \in D_i}(1-x_j)} = x_i$$

□

Now to prove the general lemma:

Proof.

$$\Pr\left[\bigcap_{i=1}^{n}\overline{A_i}\right] = \prod_{i=1}^{n}\Pr\left[\overline{A_i} \mid \text{Not}(\{j : j < i\})\right]$$
$$= \prod_{i=1}^{n}\left(1 - \Pr\left[A_i \mid \text{Not}(\{j : j < i\})\right]\right)$$

By the claim,

$$\Pr\left[A_i \mid \text{Not}(\{j : j < i\})\right] \leq x_i.$$

And thus,

$$\Pr\left[\bigcap_{i=1}^{n}\overline{A_i}\right] \geq \prod_{i=1}^{n}(1 - x_i) > 0.$$

□

Now let's look again at a problem regarding hypergraphs. Recall that a hypergraph H has *Property B* (2-colorable) if there exists a 2-coloring of $V(H)$ such that no edge is monochromatic.

Theorem 8.5. *Let H be a hypergraph. If $\forall e \in E(H), e$ has size at least k and intersects at most d other edges, and*

$$e(d+1) \leq 2^{k-1},$$

then H has Property B.

Proof. Consider a random 2-coloring of $V(H)$. $\forall f \in E(H)$, let $A_f =$ event f is monochromatic. Note that

$$\Pr[A_f] = \frac{2}{2^{|f|}} \leq 2^{1-k},$$

which we will define to be p in our local lemma.

Now let $D_f = \{f' \in E(H) : f' \cap f \neq \emptyset\}$. By assumption, $|D_f| \leq d$, and yet A_f is mutually independent of $\bigcup_e A_e \setminus D_f$. By the Local Lemma, since $ep(d+1) \leq 1$, with positive probability none of the events A_f hold.

Thus, there exists a 2-coloring of H with no monochromatic edge. □

Lastly, let's do one final lower bound for the Ramsey Numbers.

Theorem 8.6 (Even Better Lower Bound for Diagonal Ramsey Numbers)**.** *If $e\binom{k}{2}\binom{n-2}{k-2}2^{1-\binom{k}{2}} < 1$, then $R(k,k) > n$.*

Proof. Consider a random 2-coloring of $E(K_n)$. $\forall S \in \binom{V(K_n)}{k}$, let A_S be the event S is monochromatic. Note that

$$\Pr[A_s] = 2^{1-\binom{k}{2}},$$

and let this be the p in our local lemma. Now let $D_S = \left\{T \in \binom{V(K_n)}{k} : |S \cap T| \geq 2\right\}$. Note that $|D_S| < \binom{k}{2}\binom{n-2}{k-2}$, which we will let be d, and yet A_S is mutually independent of $U_T A_T \backslash D_S$. By the Lovász Local Lemma, since $ep(d+1) \leq 1$, with positive probability none of the events A_S hold.

Some more calculations actually gives:

$$R(k,k) > \frac{\sqrt{2}}{e}(1+o(1))k2^{k/2},$$

which is a factor of two in terms of improvement over the basic lower bound. This will be our final altercation with the Ramsey Numbers. □

9. Conclusion

The essence of the probabilistic method is quite simple: the use of probability to unconventionally prove the existence of objects non-constructively. The coolest part about the probabilistic method is actually highlighted through the Ramsey numbers: as we apply more and more methods, we can further better our bound more and more!

Acknowledgements

The author would like to thank Dr. Simon Rubinstein-Salzedo for the opportunity to write this paper, as well as Andrew Lin for his constant support and expertise on this topic.

References

[AS04] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, New York, second edition, 2004.
[Spe83] Joel Spencer. Ramsey theory and ramsey theoreticians. *J. Graph Theory*, 7(1):15–23, 1983.
[Spe85] Joel Spencer. Probabilistic methods. *Graphs Comb.*, 1(1):357–382, 1985.