The Probabilistic Method

Isaac Sun

Euler Circle

July 2023

Background

- Usually proofs in mathematics are very concrete and thorough
- The probabilistic method is a **non-constructive way** to prove the existence of some combinatorial object
- Proofs generally consist of proving an object has a positive probability of occurring, or using the expected value of some random variable.

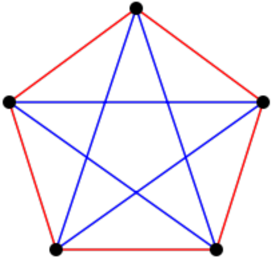
Ramsey Numbers

Ramsey Numbers

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 by red and blue, there exists a red K_k or a blue K_l.

Ramsey Numbers

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 by red and blue, there exists a red K_k or a blue K_l.



This diagram above shows a 2-coloring of K_5 with no monochromatic K_3.
Thus $R(3, 3) > 5$.

Ramsey Numbers

In 1929, Ramsey proved that $R(k, l)$ is finite for any two integers (k, l). Then in 1947, Erdős proved a lower bound for the diagonal Ramsey numbers $R(k, k)$

Erdős 1947

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

Ramsey Numbers

In 1929, Ramsey proved that $R(k, l)$ is finite for any two integers (k, l). Then in 1947, Erdős proved a lower bound for the diagonal Ramsey numbers $R(k, k)$

Erdős 1947

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

Our proof idea with the probabilistic method here is to show that there is some positive probability that there is no monochromatic K_k, meaning that $R(k, k)$ must be greater than $\lfloor 2^{k/2} \rfloor$, since $R(k, k)$ should have probability 0 of having no monochromatic K_k. We will be considering all possible colorings with equal probability.

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,

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

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 (the union of some probabilities is less than or equal to their sum, consider a venn diagram),

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

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

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

Proof

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 $k \geq 3$.

Core Idea

The Ramsey numbers is a great example of the essence of the probabilistic method: Using probability to determine the existence of combinatorial objects *non-constructively*. Note how we never explicitly found a coloring that didn't have a monochromatic K_k in our proof.

Core Idea

The Ramsey numbers is a great example of the essence of the probabilistic method: Using probability to determine the existence of combinatorial objects *non-constructively*. Note how we never explicitly found a coloring that didn't have a monochromatic K_k in our proof.

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.

Core Idea

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. You can check out my paper for more details regarding why we choose probability over counting.

Linearity of Expectation

One of the common methods of the probabilistic Method is to use Linearity of Expectation.

Linearity of Expectation

One of the common methods of the probabilistic Method is to use Linearity of Expectation.

Linearity Of Expectation

Given a discrete random variable $X = \sum_i c_i X_i$ for random variables X_i,

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

- Holds true regardless of independence between X_i
- We can now say that there exists some point X in the probability space X such that $X \geq E[X]$ or $X \leq E[X]$

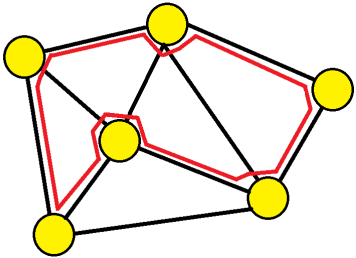
Hamiltonian Paths

Tournament

A tournament on n players is an orientation of the edges of a complete graph K_n.

Hamiltonian Path

A Hamiltonian path in a tournament T is a directed path that includes all vertices of T.



Hamiltonian Paths

Szele 1943

There exists a tournament T with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths.

Hamiltonian Paths

Szele 1943

There exists a tournament T with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths.

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.

Hamiltonian Paths

Szele 1943

There exists a tournament T with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths.

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 $\sigma(1)\sigma(2)\ldots\sigma(n)$ is a Hamiltonian path.

By Linearity of Expectation, since there are $n-1$ edges and $\frac{1}{2}$ chance that it is in the right direction,

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

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

Dominating Sets

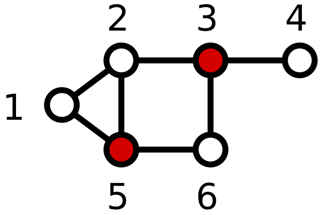
Dominating Set

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.

Dominating Sets

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.



The set consisting of points $\{3, 5\}$ is a dominating set.

Dominating Sets

Theorem

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

Dominating Sets

Theorem

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. Now 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, making it a dominating set.

Proof

Let's pick each vertex independently with $p \in [0, 1]$ this time to try to get the best possible bound. Note that $\Pr[v \in Y] \leq (1-p)^{\delta+1}$, since G has minimum degree δ (anything greater than δ would give a smaller chance of being in Y). By Linearity of Expectation,

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

Proof

Let's pick each vertex independently with $p \in [0, 1]$ this time to try to get the best possible bound. Note that $\Pr[v \in Y] \le (1 - p)^{\delta+1}$, since G has minimum degree δ (anything greater than δ would give a smaller chance of being in Y). By Linearity of Expectation,

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

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

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

Proof

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.

Conclusion

The key takeaways are as follows:

- The probabilistic method is a **non-construtive** method to prove that certain combinatorial objects exist.
- Common methods include using Linearity of Expectation or Alterations to give our desired object.
- Some other important theorems that can be proven by the probabilistic method include Turán's theorem and the Weierstrass Approximation theorem

If you would like to learn more, check out my paper!

Acknowledgements

Thanks for attending my talk!