Counting Probabilities ...

Subhojit Maji

14 December 2024

Email: bmat2434@isibang.ac.in

§1 Definitions and Notation

Nothing tricky here, just setting up some definitions and notations. I'll try to not be overly formal.

In combinatorics and other fields of mathematics, one technique to prove the existence of a mathematical object is to construct it randomly and show a positive probability of success.

Definition 1.1 (Complete Graph)**.** *A complete graph is a graph in which each pair of graph vertices is connected by an edge. The* **complete graph** *with n graph vertices is denoted K_n.*

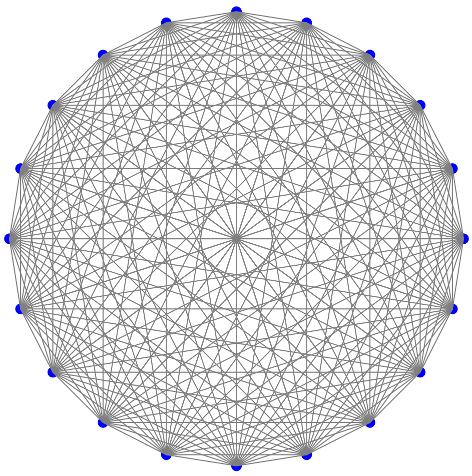

Figure 1: Here's an example of K_{20}

Remark. Some useless information –

- The K_n has $\frac{n(n+1)}{2}$ edges (a triangular number).
- The number of all distinct paths between a specific pair of vertices in K_{n+2} is given by $w_{n+2} = n!e_n = \lfloor en! \rfloor$, where e refers to Euler's constant, and

$$e_n = \sum_{k=0}^{n} \frac{1}{n!}$$

Definition 1.2 (Ramsey's Number)**.** *Let the* **Ramsey Number** $R(k, l)$ *be the smallest* n *such that if we color the edges of* K_n *(the complete graph on* n *vertices)red or blue, we always have a* K_k *that is all red or a* K_l *that is all blue.*

Basically, it asks – The minimum number of people you should invite in your party so that exactly k know each other or exactly l don't .

Definition 1.3 (Erdos-Renyi Random Graphs)**.** *A* **random graph** *is a graph in which properties such as the number of graph vertices, graph edges, and connections between them are determined in some random way. Here we only take into consideration of Erdos-Renyi Random Graphs which takes the appearance of edges into consideration with finite probabilities.*

Definition 1.4 (Chillax)**.** *It is the most useful one , defined for being chill and relaxed . Enjoy the beauty, even if it is ugly .*

Now as we discuss about Ramsey's Number , how do we clarify that it will be finite always ??

§2 Building Up ...

§2.1 Motivation using Classics...

Example 2.1 (3+3=6)

Suppose there is a gathering of 6 people such that every 2 person is a mutual enemy or mutual friend. We claim that there always exists a subset of 3 people who mutually friends or mutually enemies.

Proof. Let x be one of the six persons. Then by pigeon-hole principle, x has (at least) either $\lfloor \frac{5}{2} \rfloor = 3$ friends or 3 enemies and assume by W.L.O.G that x has three friends a, b and c. If a and b are mutual friends. Then $\{x, a, b\}$ forms the friends-group, otherwise if a and c are mutual friends then the same assertion holds true similiarly for b and c. If none of the above works then the group $\{a, b, c\}$ are neither mutual friends nor enemies which is not possible.

Now checking for 5 people cases using the similiar arguments we get that we can never form a friendship or an enemity triangle.

□

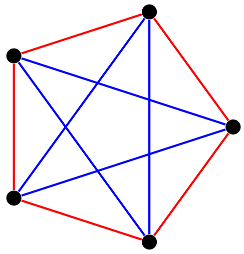

Figure 2: $R_{(3,3)} > 5$

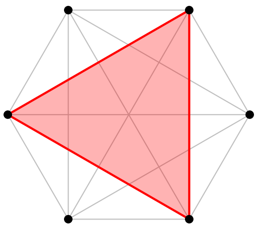

Figure 3: Demonstrating $R_{(3,3)}$

§2.2 Ramsey's Theorem

Theorem 2.2 (Ramsey)
For every integers k, l there exists a $R(k, l)$ and is finite.

Proof. So for proving that something is finite we must show that it is bounded from both sides, We know that the lower bound is naturally set at 0 (Although, we can better this lower boundary), what about the upper bound!!

Claim — For all $r, s \geq 2$; $R(r, s) \leq R(r-1, s) + R(r, s-1)$

Proof. Let us write $N = \mathcal{R}(m-1, n) + \mathcal{R}(m, n-1)$ for convenience. To prove the general upper bound, we must show that in any blue–red colouring of the edges of $\mathcal{K}_N$, there must exist either a blue K_m or a red K_n.

Let V and E denote the set of vertices and edges, respectively, of $\mathcal{K}_N$, and consider any blue–red colouring of the edges of $\mathcal{K}_N$. Choose any $v \in V$, and partition the set $V \setminus \{v\}$ into sets

$$B = \{x \in V : xv \in E \text{ and is coloured blue}\}$$

and

$$R = \{x \in V : xv \in E \text{ and is coloured red}\}.$$

Then $|B| + |R| = N - 1 = \mathcal{R}(m-1, n) + \mathcal{R}(m, n-1) - 1$, so that $|B| < \mathcal{R}(m-1, n)$ and $|R| < \mathcal{R}(m, n-1)$ is not simultaneously possible. Therefore, at least one of $|B| \geq \mathcal{R}(m-1, n)$ and $|R| \geq \mathcal{R}(m, n-1)$ must hold.

Consider the case $|B| \geq \mathcal{R}(m-1, n)$; the parallel case $|R| \geq \mathcal{R}(m, n-1)$ can be argued by replacing the role of blue with red. Since the subgraph K_B of $\mathcal{K}_N$ has at least $\mathcal{R}(m-1, n)$ vertices, K_B must contain either a blue K_{m-1} or a red K_n by the definition of Ramsey number $\mathcal{R}(m-1, n)$. If the first of these cases hold, then the vertex v together with those of K_{m-1} forms a blue K_m by construction of B. Thus, in any case, $\mathcal{K}_N$ must contain either a blue K_m or a red K_n. This completes the assertion that

$$\mathcal{R}(m, n) \leq \mathcal{R}(m-1, n) + \mathcal{R}(m, n-1)$$

for $m, n \geq 2$. □

HomeWork — Try to refine this upper bound for $R(k, k)$. □

§2.3 Bounds on Ramsey's Number

Lemma 2.3. For any positive integers n and k,

$$\binom{n}{k} \leq \frac{1}{e}\left(\frac{en}{k}\right)^k.$$

Here $e \approx 2.718\ldots$ is Euler's constant.

Proof. Do $\binom{n}{k} \leq \frac{n^k}{k!}$ and then use calculus to prove that $k! \geq e(k/e)^k$. Specifically,

$$\ln 1 + \ln 2 + \cdots + \ln k \geq \int_{x=1}^{k} \ln x \, dx = k \ln k - k + 1$$

whence exponentiating works. □

Algebra isn't much fun, but at least it's easy. Let's get back to the combinatorics.

Theorem 2.4 (Lower Bound)
Let n and k be integers with $n \leq 2^{k/2}$ and $k \geq 3$. Then it is possible to color the edges of the complete graph on n vertices each either red or blue with the following property: one cannot find k vertices for which the $\binom{k}{2}$ edges among them are monochromatic.

Remark. In the language of Ramsey numbers, prove that $R(k, k) > 2^{k/2}$.

Solution. Again we just randomly color the edges and hope for the best. We use a coin flip to determine the color of each of the $\binom{n}{2}$ edges. Let's call a collection of k vertices *bad* if all $\binom{k}{2}$ edges are the same color. The probability that any collection is bad is

$$\left(\frac{1}{2}\right)^{\binom{k}{2}-1}.$$

The number of collections in $\binom{n}{k}$, so the expected number of bad collections is

$$\mathbb{E}[\text{number of bad collections}] = \frac{\binom{n}{k}}{2^{\binom{k}{2}-1}}.$$

We just want to show this is less than 1. You can check this fairly easily using Lemma 2.3; in fact, we have a lot of room to spare. □

§3 Random Graphs

Example 3.1

Suppose you are again at a party and there are n people, but there is a condition in meeting someone, you have a coin and you toss if it lands at heads both of you greet and if it is tails you walk away and never meet with that person.

Note :- Coin Toss is independent of previous trials

A simple observation shows that it is a Binomial Random Variable $Bin(n, \frac{1}{2})$. But why does this problem come up all of a sudden ? **Chillax !!**

You can clearly observe the connection beautiful connection between the 2 concepts **(Ramsey and Random Graphs)**.

Remark (**Bits of Info**).

- There is also something called as adjacency matrix such that every element $a_{ij} = 1$ if the vertex {i, j} are connected and 0 otherwise.
- Now suppose I just add the condition that the probability of $a_{ij} = 1$ is p. This is basically what Random Graph is ... Every element occurs with some probability just like a coin toss.

Remark (Notations)**.** A random graph with n vertices is denoted by $\mathcal{G}_{n,p}$.

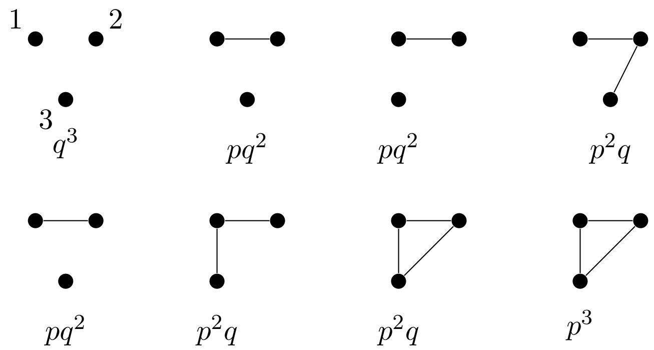

Figure 4: Random graphs on three vertices and their probabilities $\mathcal{G}_{n,p}$.

Example

If $p = \frac{1}{2}$, then for a $\mathcal{G}_{n,\frac{1}{2}}$

- The number of edges has the distribution $Bin(\binom{n}{2}, \frac{1}{2}) \sim X$
- The $\mathbb{E}[X] = \frac{n(n-1)}{4}$
- The $var[X] = \frac{n(n-1)}{8}$

Remark. If you have any doubts about **Small and Big O Notations** please refer to this reference.

§4 Calling Old Guy...

Now let's recall some Probability Stuffs ...

§4.1 Markov's Inequality

Theorem 4.1 (Markov's Inequality)
If X is a nonnegative random variable and $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

§4.2 Chebyshev's Inequality

Theorem 4.2 (Chebyshev's Theorem)
If X is a random variable and $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, then for all λ we have

$$\mathbb{P}(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

Lemma. For any non-negative random variable X ,

$$\mathbb{P}(X = 0) \leq (\frac{\sigma}{\mu})^2$$

Proof. Observe that the max value of the $\mathbb{P}(X = 0)$ is $\mathbb{P}(|X - \mu| \geq \mu)$. So ,

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mu| \geq \mu)$$

Now just apply the Chebyshev's Inequality .

□

Lemma. If X is a random variable such that $\mathbb{E}[X] > 0$, then the following holds true :

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \epsilon\, \mathbb{E}[X]) \leq (\frac{\sigma}{\epsilon\, \mathbb{E}[X]})^2$$

In particular, if $var(X) = o(\mathbb{E}[X]^2)$, then $X \sim \mathbb{E}[X]$ with high probability

Direct application of Chebyshev's Inequality.

§5 Grand Finale

§5.1 Random Graphs and Triangles

Now our question is how does the distribution of the triangles look like ?

Lemma 5.1. The $\mathbb{E}[\texttt{\#Triangles}]$ in $\mathcal{G}_{n,p}$ is $\binom{n}{3}p^3 \sim O(n^3p^3)$.

Proof : Let us declare a Random Variable X_n such that :

$$X_n = \sum_{\substack{i,j,k\in[n]\\ \text{distinct}}} X_{ij}X_{jk}X_{ik}.$$

where,

$$X_{ij} = \begin{cases} 0 & \text{When the vertex i and j are not connected} \\ 1 & \text{When the vertex i and j are connected} \end{cases}$$

So we can now finish the computation:
By linearity of expectation, each term is p^3, so $\mathbb{E}[X_n] = \binom{n}{3}p^3$.

Lemma 5.2. The $var[\texttt{\#Triangles}]$ in $\mathcal{G}_{n,p} \sim O(n^3p^3 + n^4p^5)$

Proof : The variance is a bit harder, and we're mostly worried about the covariance term: when do those cross-terms come up?
Well, given a pair of triples T_1, T_2 of vertices, we can find the covariance for those triangles. If there is at most one vertex of overlap, no edges overlap, so there is no covariance. The others are a bit harder, but we use $\text{cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$:

$$\text{cov}[X_{T_1}, X_{T_2}] = \begin{cases} 0 & |T_1 \cap T_2| \le 1 \\ p^5 - p^6 & |T_1 \cap T_2| = 2 \\ p^3 - p^6 & T_1 = T_2 \end{cases}$$

So we can now finish the computation:

$$var(X_n) = \binom{n}{3}(p^3 - p^6) + \binom{n}{2}(n-2)(n-3)(p^5 - p^6) \sim O(n^3p^3 + n^4p^5),$$

Theorem 5.3
The $\mathbb{P}\,[\texttt{\#Triangles=k}]$ in $\mathcal{G}_{n,p} \sim \mathcal{N}(\mu, \sigma^2)$

Proof : Let us denote $\mathcal{S}$ as the set of the set of all combinations of the 3 vertices $\mathcal{E}_j$ by the event that the j^{th} set forms a triangle. Then,

$$\mathbb{P}\left(\bigcap_{j=1}^{k} \mathcal{E}_j\right) = \mathbb{P}(\texttt{\#Required}) = \mathbb{P}(\texttt{X}_n = k)$$

Now, this is indeed tough to calculate. So we have a better idea.
We know that

$$\frac{X_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

This was shown by Rucinski by calculating the k^{th} moment, i.e. $\mathbb{E}\left(\frac{X_n-\mu}{\sigma}\right)^k$ and that turns out to be the same as the k^{th} moment of the $\mathcal{N}(0,1)$.

So, the probability essentially becomes

$$\mathbb{P}(X_n = k) \approx \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(k-\mu)^2}{2\sigma^2}\right)$$

Remark. We can even condition on the value of np_n.

- If $np_n \to 0$ then there is a high probability that the graph is triangle-free.
- If $np_n \to \infty$ then there is a high probability that the graph has atleast 1 triangle also the `#Triangles` $\sim \mathbb{E}[X_n]$.
- If $np \to c$, it turns out that the `#Triangles` $\sim \text{Poisson}(c)$.

§5.2 Pushing towards Probability ...

This article was motivated by the following problem, given at the 55th International Mathematical Olympiad.

Example 5.4 (IMO 2014/6)

A set of lines in the plane is in *general position* if no two are parallel and no three pass through the same point. A set of lines in general position cuts the plane into regions, some of which have finite area; we call these its *finite regions*. Prove that for all sufficiently large n, in any set of n lines in general position it is possible to colour at least $\sqrt{n}$ lines blue in such a way that none of its finite regions has a completely blue boundary.

Note: Results with $\sqrt{n}$ replaced by $c\sqrt{n}$ will be awarded points depending on the value of the constant c.

Remark. We are not going to solve this exactly, which I would leave as an Excercise to gather your thoughts on the problem .

We'll show the bound $c\sqrt{n}$ for any $c < \frac{2}{3}$.
First, we need to bound the number of triangles.

Claim — There are at most $\frac{1}{3}n^2$ triangles.

Proof. Consider each of the $\binom{n}{2}$ intersection of two lines. One can check it is the vertex of at most two triangles (**??**)

Since each triangle has three vertex, this implies there are at most $\frac{2}{3}\binom{n}{2} < \frac{1}{3}n^2$ triangles. □

It is also not hard to show there are at most $\frac{1}{2}n^2$ finite regions[1].

[1] Say, use $V - E + F = 2$ on the graph whose vertices are the $\binom{n}{2}$ intersection points and whose edges are the $n(n-2)$ line segments.

Now color each line blue with probability p. The expected value of the number of lines chosen is

$$\mathbb{E}[\text{lines}] = np.$$

The expected number of completely blue triangles is less than

$$\mathbb{E}[\text{bad triangles}] < \frac{1}{3}n^2 \cdot p^3.$$

For the other finite regions, of which there are at most $\frac{1}{2}n^2$, the probability they are completely blue is at most p^4. So the expected number of completely blue regions here is at most

$$\mathbb{E}[\text{bad polygons with 4+ sides}] < \frac{1}{2}n^2 \cdot p^4.$$

Note that the expected number of quadrilaterals (and higher) is really small compared to any of the preceding quantities; we didn't even bother subtracting off the triangles that we already counted earlier. It's just here for completeness, but we expect that it's going to die out pretty soon.

Now we do our alteration – for each bad, completely blue region, we un-blue one line. Hence the expected number of lines which are blue afterwards is

$$np - \left(\frac{n^2}{3} \cdot p^3\right) - \left(\frac{n^2}{2} \cdot p^4\right) = np\left(1 - \frac{np^2}{3} - \frac{np^3}{2}\right).$$

Ignore the rightmost $\frac{np^3}{2}$ for now, since it's really small. We want $p = c/\sqrt{n}$ for some c; the value is roughly $c \cdot (1 - c^2/3)$ at this point, so an optimal value of p is $p = n^{-1/2}$ (that is, $c = 1$); this gives

$$\sqrt{n} \cdot \left(\frac{2}{3} - \frac{27}{16}\frac{1}{\sqrt{n}}\right) = \frac{2}{3}\sqrt{n} - \frac{81}{32}.$$

For n sufficiently large, this exceeds $c\sqrt{n}$, as desired.

How does it feel to solve a recent IMO Problem 6 ?
But we are just ending it right now ... Chillax

If you like to check-out the full power of today's theorems, then do learn about Threshold-Functions, Clique-Number, Turan's Theorem and Turan's Graph.