The Probabilistic Method in Combinatorics

Heidi Lei

August 18, 2020

1 Introduction

The probabilistic method is a method of attacking combinatorics problems using probabilistic tools. The general method works to show the existence of an object satisfying certain probabilities by showing that the probability for its existence in a certain probability space is positive.

In this paper, we follow the exposition of [1] and demonstrate the power of two simple probabilistic methods, alterations and the second moment method, through proving results in different combinatorial scenarios.

2 Alterations

In general, the alterations method entails showing the existence of an object that is fairly close to satisfying the requirements, and then"altering" it in some way to obtain an object of our desire.

2.1 Ramsey Numbers

The Ramsey number $R(k,l)$ is the minimum number n such that any two-coloring of the complete graph K_n contains either a red K_k or a blue K_l. Equivalently, if we have $R(k,l) > n$, we can find a two-coloring of K_n that contains neither a red K_k nor a blue K_l. Using the probabilistic method, we could prove the following lower bounds on Ramsey numbers.

Theorem 2.1. *For all integers n and $p \in [0,1]$*

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

Proof. Consider a random two-coloring of the complete graph K_n where each edge is colored red with probability p. Let X be the total number of red K_k and blue K_l subgraphs. Then by the linearity of expectation, we have

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

As a result, there exists a coloring of K_n with at most $\mathrm{E}(X)$ red K_k and blue K_l subgraphs. If we remove one vertex from each of these subgraphs, then we obtain a two-colored complete graph with neither a red K_k nor a blue K_l with $n - \mathrm{E}(X)$ vertices. □

In this proof, we first showed the existence of a graph with a certain number of undesirable cliques using the probabilistic method, and then modify the graph so that it no longer has those cliques, which is the crux of the alteration method.

2.2 Heibronn Triangle Problem

Next, we look at an example in combinatorial geometry, the Heilbronn triangle problem, where we try to arrange a set of points in a compact region of the plane such that the triangles formed by the points are as big as possible. More precisely, let S be a set of n points in a unit square, and let $T(S)$ denote the area of the smallest triangle formed by any three points in S. We are interested in the arrangements of points that maximize this area, so define $T(n)$ to be the supremum of $T(S)$ over all possible sets S with n points.

Heilbronn's original conjecture was $T(n) = O(1/n^2)$, which was disproved by Komlós, Pintz and Szemerédi, who showed that the lower bound is given by $T(n) = \Omega(\log n/n^2)$. Here we present and prove a simpler lower bound $T(n) = \Omega(1/n^2)$.

Theorem 2.2. *There is a set S of n points in the unit square U such that $T(S) \geq 1/\left(100n^2\right)$.*

Proof. We first obtain a probabilistic bound for the area of a random triangle. Let A, B, C be three points chosen uniformly at random from the unit square, and let $[ABC]$ denote its area. We are interested in calculating $\Pr([ABC] \leq \varepsilon)$.

If the distance $|AB| = x$, then the height of the triangle ABC from C to AB must be at most $2\varepsilon/x$, so C must lie in a strip with width $4\varepsilon/x$ and length at most $\sqrt{2}$. Since $\Pr(x \leq |AB| \leq x + dx) \leq \pi((x+dx)^2 - x^2) = 2\pi x dx$, we integrate to obtain

$$\Pr([ABC] \leq \varepsilon) \leq \int_0^{\sqrt{2}} \frac{4\varepsilon}{x} \cdot \sqrt{2} \cdot 2\pi x dx = 16\pi\varepsilon.$$

Then, we choose $2n$ points $P_1, \ldots, P_{2n}$ in the unit square uniformly and randomly. Let X be the number of triangles $P_iP_jP_k$ with area less than $1/(100n^2)$. By the linearity of expectation, we have

$$\mathrm{E}(X) = \binom{2n}{3} \Pr([P_iP_jP_k] \leq \varepsilon) \leq \frac{(2n)^3}{6} \cdot \frac{16\pi}{100n^2} < n.$$

Therefore, there exists a set of $2n$ points such that there form fewer than n triangles with area less than $1/100n^2$. After removing one point from each triangle, we have a set of at least n points with no small triangles. □

Erdős gave an ingenious explicit construction for the bound $T(n) \geq 1/2(n-1)^2$. Consider all the points (x, x^2) on the lattice $\mathbb{Z}_n \times \mathbb{Z}_n$. Note that no three points are collinear since a quadratic has at most two roots. Since the smallest area a lattice triangle can have is $1/2$, we have exhibited a set of n points in a $[0, n-1] \times [0, n-1]$ square with no triangle smaller than $1/2$. Scaling down by a linear factor of $n - 1$ results in the claimed bound.

3 The Second Moment

The second moment method is built upon Chebyshev's inequality, which concerns the variance of a random variable. The variance of X is defined to be

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2).$$

The variance is an indicator of how much X deviates from its expectation. To ease the notation, let μ denote the expectation and σ^2 denote the variance.

Theorem 3.1 (Chebyshev's inequality)**.** *For any positive λ,*

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

Proof.

$$\sigma^2 = \mathrm{E}((X - \mu)^2) \geq (\lambda\sigma)^2 \Pr(|X - \mu| \geq \lambda\sigma).$$

□

We have the following corollaries of Chebyshev's inequality when X is a nonnegative integral-valued random variable.

Corollary 3.2. $\Pr(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathrm{E}(X)}^2$.

Corollary 3.3. *If* $\mathrm{Var}(X) = o(E(X)^2))$*, then* $X > 0$ *and furthermore* $X \sim \mathrm{E}(X)$ *almost always.*

Next, we are interested in how variance behaves under addition. Let $X = \sum_{i=1}^{n} X_i$ be a sum of random variables X_i. Then by the linearity of expectation, the variance of X can be expanded to be

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

where the covariance of two random variables X_i and X_j is defined to be

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}(X_i X_j) - E(X_i)E(X_j).$$

If two random variables are independent, then their covariance is zero, so in the case that X is a sum of independent variables, the variance is linear.

Suppose further that each X_i is an indicator random variable for an event A_i where $X_i = 1$ with probability p_i and $X_i = 0$ with probability $1 - p_i$. Then it is simple to calculate that $\mathrm{E}(X_i) = p_i$ and $\mathrm{Var}(X_i) = p_i(1 - p_i) \leq \mathrm{E}(X_i)$. Therefore, we have

$$\mathrm{Var}(X) \leq \mathrm{E}(X) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

in the case then X_i are indicator random variables.

If X_i and X_j are not independent and $i \neq j$, we write $i \sim j$, and we have

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}(X_i X_j) - \mathrm{E}(X_i)\,\mathrm{E}(X_j) \leq E(X_i X_j) = \Pr(A_i \wedge A_j).$$

Therefore, we can rewrite the inequality to be

$$\mathrm{Var}(X) \leq \mathrm{E}(X) + \Delta$$

where

$$\Delta = \sum_{i \sim j} \Pr(A_i \wedge A_j).$$

Therefore, in this case we can rewrite Corollary 3.3 as the following.

Corollary 3.4. *If* $\mathrm{E}(X) \to \infty$ *and* $\Delta = o(E(X)^2))$*, then* $X > 0$ *and furthermore* $X \sim \mathrm{E}(X)$ *almost always.*

Furthermore, if $X_1, \ldots, X_n$ are symmetric, we have

$$\Delta = \sum_i \Pr(A_i) \sum_{j \sim i} \Pr(A_j | A_i),$$

and as a result of the symmetry, the second summand is independent of the index and we set it to be

$$\Delta^* = \sum_{j \sim i} \Pr(A_j | A_i),$$

then we have

$$\Delta = \Delta^* \sum_i \Pr(A_i) = \Delta^*\,\mathrm{E}(X).$$

We obtain another version of Corollary 3.3.

Corollary 3.5. *If* $\mathrm{E}(X) \to \infty$ *and* $\Delta^* = o(E(X))$*, then* $X > 0$ *and furthermore* $X \sim \mathrm{E}(X)$ *almost always.*

3.1 Random Graphs

Informally, a random graph $G(n,p)$ is the probability space that consists of graphs with n vertices where edges are formed uniformly and independently between each pair of vertices with probability p. For a graph property A, if there exists a function $r(n)$ such that if $p(n) \ll r(n)$ then $G(n,p)$ almost always does not satisfy A, and if $p(n) \gg r(n)$ then $G(n,p)$ almost always satisfies A, then $r(n)$ is a *threshold function* of property A. We prove some results on threshold functions for certain graph properties using the second moment method.

Let $\omega(G)$ denote the maximal size of a complete subgraph of G.

Theorem 3.6. *A threshold function for* $\omega(G) \geq 4$ *is* $n^{-2/3}$.

Proof. Since there are 6 edges in K_4, the probability of any four vertices in G forming a clique is p^6. By the linearity of expectation, the expected number of 4-cliques in G is given by

$$\mathrm{E}(X) = \binom{n}{4} p^6 \sim \frac{n^4 p^6}{24}.$$

Therefore, when $p \ll n^{-2/3}$, we have $\mathrm{E}(X) = o(1)$ and $X = 0$ almost always.

Conversely, when $p \gg n^{-2/3}$ and thus $\mathrm{E}(X) \to \infty$, we calculate Δ^*. Let S and T be sets of 4 vertices and X_S and X_T denote the indicator random variables of S and T forming a 4-clique. Then $S \sim T$ if and only if $|S \cup T| = 2$ or 3. For a given set S, there are $O(n^2)$ sets T such that T and S share a pair of vertices. And in this case, if S forms a 4-clique, T must already contain an edge, so $\Pr(A_T|A_S) = p^5$. Similarly, there are $O(n)$ sets T that share three vertices with S and $\Pr(A_T|A_S) = p^3$. Since $p \gg n^{-2/3}$, we have

$$\Delta^* = O(n^2p^5) + O(np^3) = o(n^4p^6) = \mathrm{E}(X).$$

Therefore, Corollary 3.5 implies $X > 0$, so G contains a 4-clique almost surely. □

Using the same method, we can prove a more general result on balanced subgraphs, not just 4-cliques.

Definition 3.7. The density $\rho(G)$ of a graph G with v vertices and e edges is given by $\rho(G) = e/v$. A graph H is *balanced* if $\rho(H) \geq \rho(H')$ for every subgraph H'.

Note that complete graphs are balanced since $\rho(K_n) = (n-1)/2$.

Theorem 3.8. *A threshold function for G containing a subgraph H with v vertices and e edges is $n^{-v/e}$ if and only if H is balanced.*

Proof. We first consider the case when H is balanced. Let S be a set with v vertices in G, and A_S be the event that $G|_S$ contains H, then

$$p^e \leq \Pr(A_S) \leq v!p^e$$

since there are potentially $v!$ arrangements for H. Let X_S be the indicator random variable for A_S and $X = \sum_{|S|=v} X_S$. By linearity of expectations,

$$\mathrm{E}(X) = \binom{n}{v} \Pr(A_S) = \Theta(n^v p^e).$$

Thus, if $p \ll n^{-v/e}$, $\mathrm{E}(X) = o(1)$, and $X = 0$ always surely.

If $p \gg n^{-v/e}$, $\mathrm{E}(X) \to \infty$, and we calculate

$$\Delta^* = \sum_{i=2}^{v-1} \sum_{|T \cup S| = i} \Pr(A_T|A_S)$$

since $S \sim T$ if and only if S and T share between 2 to $v-1$ vertices. Fixing S, for each number of shared vertices i, there are $O(n^{v-i})$ choices for T each with $O(1)$ possible copies of H on it. Since H is balanced with density e/v, there are at most ie/v edges with both vertices in S, and thus at least $e - ie/v$ edges not in the induced subgraph of S. Therefore,

$$\Pr(A_T|A_S) = O(p^{e-ie/v})$$

and

$$\begin{aligned}\Delta^* &= \sum_{i=2}^{n-1} O(n^{v-i}p^{e-ie/v}) \\ &= \sum_{i=2}^{n-1} O((n^v p^e)^{1-i/v}) \\ &= \sum_{i=2}^{n-1} o(n^v p^e) \\ &= o(E(X)).\end{aligned}$$

By Corollary 3.5, $X > 0$ almost surely.

Conversely, if H is not balanced, there exists a subgraph $H' \leq H$ with v' vertices and e' edges such that $e'/v' > e/v$. Let α be a number such that $v'/e' < \alpha < v/e$, then by a similar argument as above, if $p = n^{-\alpha} \ll n^{-v'/e'}$, then $E(X) = o(1)$ where X is the number of copies of H', so G does not contain H' almost surely. Since H' is a subgraph of H, G does not contain H almost surely as well. □

With a few additions and modifications to the proof above, we can arrive at the following more generalized result.

Theorem 3.9. *Let H be a balanced grpah with v vertices, e edges, and a automorphisms. Then the number of copies of H in a random graph G where $p \gg n^{-v/e}$ is almost always*

$$X \sim \frac{n^v p^e}{a}.$$

3.2 Distinct Sums

A set of positive integers $S = \{x_1, \ldots, x_n\}$ is said to have distinct sums if the sums $\sum_{x \in R} x$ of each subset $R \subset S$ are distinct. We are interested in the maximal size $f(n)$ of a set $S \subset [n]$ with distinct sums.

Since the set $S = \{2^k : 0 \leq k \leq \lfloor \log_2 n \rfloor\}$ has distinct sums, we must have $f(n) \geq \lfloor \log_2 n \rfloor + 1$. It remains an open problem whether $f(n) \leq \log_2 n + O(1)$, i.e., we cannot do much better than the set of powers of two. We can obtain an upper bound through a simple counting argument. Since there are $2^{f(n)}$ total subsets, and the maximal subset sum is $nf(n)$, we must have $2^{f(n)} \leq nf(n)$ since the sums are distinct, which gives the bound

$$f(n) \leq \log_2 n + \log_2 \log_2 n + O(1).$$

With the second moment method, we can obtain a slightly better bound.

Theorem 3.10. $f(n) \leq \log_2 n + \frac{1}{2}\log_2 \log_2 n + O(1)$.

Proof. Let the set $S \subset [n]$ be $\{x_1, \dots, x_k\}$. Define the random variable

$$X = \varepsilon_1 x_1 + \varepsilon_2 x_2 + \cdots + \varepsilon_k x_k$$

where ε_i is chosen randomly and independently from $\{0, 1\}$. Thus, by linearity of expectations, we have $\mathrm{E}(X) = \frac{1}{2}\sum_{i=1}^{k} x_i$ and since $x_i^2 \leq n^2$, we can bound the variance

$$\mathrm{Var}(X) = \frac{1}{4}\sum_{i=1}^{k} x_i^2 \leq \frac{n^2 k}{4}.$$

By the Chebyshev's inequality, we have for $\lambda > 1$,

$$\Pr(|X - \mu| \geq \frac{\lambda n\sqrt{k}}{2}) \leq \frac{1}{\lambda^2}.$$

Reversing the probability gives

$$\Pr(|X - \mu| < \frac{\lambda n\sqrt{k}}{2}) \geq 1 - \frac{1}{\lambda^2}.$$

Since S has distinct sums, each value X can potentially achieve appears with possibility 0 or $1/2^k$ since there are 2^k total choices for all the ε_i. Therefore, we have an upper bound

$$\Pr(|X - \mu| < \frac{\lambda n\sqrt{k}}{2}) \leq \frac{1}{2^k}(\lambda n\sqrt{k} + 1).$$

Combining the two bounds, we have

$$1 - \frac{1}{\lambda^2} \leq \frac{1}{2^k}(\lambda n\sqrt{k} + 1),$$

which rearranges to

$$n \geq \frac{2^k(1 - \lambda^{-2}) - 1}{\lambda\sqrt{k}}.$$

Any $\lambda > 1$ gives the desired asymptotic bound.

□

References

[1] N. Alon & J. Spencer: *The Probabilistic Method*, Third edition, Wiley-Interscience 2008.