Probabilistic Method for Math 75SI

Aaryan Sukhadia

Winter 2024, Stanford University

1 Introductory Probability Theory

The basic idea of the probabilistic method is as follows: Take some object or set X. You are trying to figure out if a certain configuration of X that has a certain special property Y exists. If finding an explicit construction of the configuration is too hard, you consider a *random* configuration of X, and find the probability that is has property Y. If that probability is greater than 0, then there *must* exist a configuration with that property!

There are some important prerequisite definitions and facts we must cover first, however. First is the formal notion of what it means for an **event** to occur with some probability.

Definition. An **event space** is a set $U = \{A_1, A_2...\}$ (possibly infinite) together with a corresponding **probability function** $P : U \to [0, 1]$

Remark. The above definition is **not** the conventional definition for a formal event space. It's been simplified for the sake of this class, and you shouldn't really use it outside of this class. For a more detailed and correct resources, see Florian Herzog's 2013 Probability Theory slides.

Intuitively, each A_i is a thing that happens, and the output of the probability function $P(A_i)$ tells us the probability that the event A_i occurs. U is the set of all possible things that could happen (in the context of the probabilistic analysis you're doing). The following is useful notation:

- $P(A \wedge B)$; the probability of both events A **and** B occuring.
- $P(A \vee B)$; the probability of either A **or** B occuring (non-exclusively).
- $P(\overline{A})$; the probability of A **not** occuring.

Using this notation, we can express our first theorem on event probabilities.

Theorem 1.0.1. *If A and B are 2 events, then:*

- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$ *(Removing the overlap)*
- $P(A) + P(\overline{A}) = 1$ *(Law of Excluded Middle)*
- $P(\overline{A \wedge B}) = P(\overline{A} \vee \overline{B})$, $P(\overline{A \vee B}) = P(\overline{A} \wedge \overline{B})$ *(De Morgan's Laws)*

We will not prove this theorem, but hopefully all these properties should feel familiar/intuitive. Another notion that should feel intuitive is that of two **independent events**.

Definition. Two events A, B are **independent** if:

$$P(A \wedge B) = P(A) \times P(B)$$

Intuitively, two events are independent if the probability of one happening doesnt influence the probability of the other happening. For example, raining in Stanford today and me wearing a raincoat today might not be independent events, but raining in Stanford today and Jupiter's red-spot disappearing would be independent.

Using this, we can define a formal notion of what we mean for a variable or object to be **random**.

Definition. A **random variable** X is an object that takes on some values $\{a_1, a_2...\}$ with respective probabilities $\{p_1, p_2...\}$, and we write $P(X = a_i) = p_i$.

The event space for a random variable is $\{X = a_1, X = a_2, ...\}$. As you might know from experience or intution, the sum of all the probabilities for each random variable must be exactly one, i.e $\sum p_i = 1$. Using the definition of independent events, we can define **independent random variables** X_1 and X_2 as those such that $P(X_1 = a_1, X_2 = a_2) = P(X_1 = a_1) \times P(X_2 = a_2)$ for all values a_1, a_2.
Another heuristic notion we need to make formal is the notion of the average or **expected value** of a random variable.

Definition. If a random variable X takes on numerical values, then we can define the **expectation** $E[X]$ to be the *average value* of X as follows:

$$E[X] = \sum_k a_k p_k$$

Remark. The above definition assumes the sum *absolutely convergent*, else by Riemmann Rearrangement simply by reordering the terms we could get different expectations, which would make no sense. What this means is there may be cases where the expectation of a random variable *doesn't exist*! See the Cauchy random variable for more details.

Theorem 1.0.2 (Linearity of Expectation)**.** *If X_1, X_2 are two different random variables then:*

- $E[X_1 + X_2] = E[X_1] + E[X_2]$ *and* $E[aX] = aE[x]$

Proof. Follows easily from definition □

2 First examples

2.1 Probabilistic Pigeonhole

The Stanford Math Department (for a limited time only) is releasing a special edition cereal box, each coming with 2 action figures inside, either of Brian Conrad or Richard Taylor. However, due to Taylor's senority, they have decided to produce 2 action figures of Taylor for every one made of Conrad. They will only ever produce 30 of these.

Claim. *There must exist a box that contains 2 figures of Richard Taylor.*

Remark. The above result is a very basic application of the pigeonhole principle that a 6 year old with no knowledge of probability could deduce. However, for the sake of getting a handle on how the probabilistic method works, we work through this example using a sledgehammer, so to speak.

Probabilistic Proof. Note that in total there are 20 BC figurines and 40 RT figurines. Let X be a random variable representing the number of Richard Taylor in one box:

- $P(X = 0) = (20/60)(19/59) = 380/3540$
- $P(X = 1) = (20/60)(40/59) + (40/60)(20/59) = 1600/3540$
- $P(X = 2) = (40/60)(39/59) = 1560/3540$

Using the formula for expected value, we get:

$$E[X] = 0 \cdot \frac{380}{3540} + 1 \cdot \frac{1600}{3540} + 2 \cdot \frac{1560}{3540} = \frac{4720}{3540} > 1$$

Since the expected number of Richard Taylors per box is more than 1, there is no way for that to happen if every box has 1 or less Richard Taylor figurines. Thus, there must exist at least 1 box with 2 Richard Taylor figurines in it.

□

2.2 Covered points in the plane

We know that when we try to cover the plane in non-overlapping circles of the same size, we cannot cover every single point. Here's a somewhat inverse question:

Q:. *What's the minimum number of points n we can put on the plane such that they cannot be covered by non-overlapping unit disks?*

Claim. $n \geq 10$

Proof. Let us place 10 points in the plane in any configuration. Now, let us pack unit circles as best as we possibly can in a hexagonal packing, as shown in Figure 1.
Now let us shift this hexagonal configuration to a *random* position on the plane. What is the probability that this random positioning covers all of our 10 points? Let A_i denote the event that the i-th point is covered by a circle. We then get:

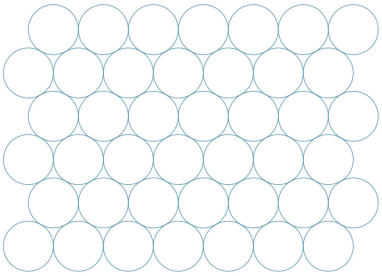

Figure 1: The best possible circle cover of the plane, with a density of $\eta < 0.907$

$$\begin{aligned} P(\text{all points covered}) &= P(A_1 \wedge A_2 \wedge ...A_{10}) \\ &= 1 - P(\overline{A_1 \wedge A_2 \wedge ...A_{10}}) \\ &= 1 - P(\overline{A_1} \vee \overline{A_2} \vee ...\overline{A_{10}}) \\ &\geq 1 - P(\overline{A_1}) - P(\overline{A_2}) - ...P(\overline{A_{10}}) \\ &= 1 - 10P(\overline{A_1}) \end{aligned}$$

Note that $P(A_1) = P(A_2) = ...P(A_{10})$, since we just chose some arbitrary collection of points. Since the packing covers more than 90.6% of the plane, the probability that a given point on the plane is *not* covered is at most 9.4%, i.e $P(\bar{1}) < 0.094 \implies 10P(\bar{1}) < 0.94$.

Thus, $P(\text{all points covered}) > 1 - 0.94 = 0.06$. Since they are all covered with positive probability, there *must* exist some random positioning that covers them all. Thus, there is no way to place 10 points on the plane such that they cannot be covered by non-overlapping unit disks.

□

3 The Tenure Game

Brian Conrad is attempting to promote one of his dear faculty to a tenure position at Stanford University, but the pernicious School Provost (let's call them SP) is standing in his way. To resolve this dispute, they are going to play a game over several years.
There are k faculty levels $1, 2...k$, with level 0 being tenure. Each level i has x_i people in it. Every year, Brian presents an arbitrary set of people S to SP, and SP gets to choose to either:

- Fire everyone in S, and promote everyone not in S
- Promote everyone in S, and fire everyone not in S

Here, promotion means moving someone from level i to $i - 1$, and firing means removing them entirely from the game. The game terminates once anyone reaches level 0, or once everyone has been fired (SP would rather fire the entire math department than give one more person tenure). As an exercise, show that this game must terminate in at most k years.

Q:. *When (if ever) does Brian not have a winning strategy?*

Claim. *If $\sum_i x_i 2^{-1} < 1$, then SP has a winning strategy no matter what Brian does.*

Proof. As always, let SP play *randomly*. In other words, they flip a coin, and if it's heads they fire everyone in the set Brian presents, and if it's tails they promote everyone that Brian presents.

For each person p, let I_p be the ***indicator variable*** that person p reaches level 0 and attains tenure. This means $I_p = 1$ if p reaches level 0, and $I_p = 0$ otherwise.

Lemma 3.0.1. *If X is an indicator variable for an event A that occurs with probability p, then $E[X] = p$*

Proof.

$$E[X] = 0 \cdot (1 - p) + 1 \cdot p = p$$

□

As we will come to see, indicator variables come in very handy when working with the probabilistic method. We define $X := \sum_p I_p$ to be the random variable representing the total number of people reaching tenure.

Consider an arbitrary person p in level i. Every year, Brian may have chosen $p \in S$ or $p \notin S$, but since SP is playing randomly, p is moved up a level with probability $1/2$. To reach level 0, p needs to be moved up a level i times. If at any step it is *not* moved up a level, that means it is fired and out of the game. Thus, p achieves tenure if and only if it is promoted i times consecutively, which happens with probability 2^{-i}.

Thus, $E[I_p] = 2^{-i}$, and by linearity of expectation we get that $E[X] = \sum_p E[I_p] = \sum_{i=1}^{k} x_i 2^{-i}$. Note if $E[X] < 1$, then everyone is fired with positive probability, which means SP must win. However, if SP can play randomly and potentially no matter what Brian does, then that means there is a sequence of choices SP can make to secure their victory, which means they have a winning strategy. □

4 Combinatorial Basics and the Erdös-Ko-Rado Theorem

Combinatorics is essentially the study of counting things; the number of ways they can be arranged, subdivided, the number of symmetries an object might have, the number of ways we can color things, and so on. One of the most basic things we count are **combinations**, or how many ways there are to pick a subset of a set. Although the previous example of the probabilistic method may have seemed trivial, this problem will show a way of applying it that is most definitely not.

Definition. If we have a set A of n elements, then $\binom{n}{k}$ (pronounced "n choose k") represents the number of distinct unordered subsets of A that have exactly k elements.

For example, $\binom{4}{2} = 6$ and $\binom{n}{1} = n$ for any n.

Theorem 4.0.1 (Basic Results)**.** *The binomial coefficient $\binom{n}{k}$ has all the following properties:*

- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- $\binom{n}{k} = \binom{n}{n-k}$
- $\frac{n+1}{k+1}\binom{n}{k} = \binom{n+1}{k+1}$

Proof. Exercise! □

Suppose $\mathcal{F}$ is a set of sets. A common genre of combinatorics questions involves counting or boudning the possible number of overlaps between sets inside $\mathcal{F}$.

Definition. $\mathcal{F}$ is **intersecting** if for all $A, B \in \mathcal{F}$, it holds that $A \cap B \neq \emptyset$.

For example, $\{\{1,2\},\{2,3\},\{1,3\}\}$ is an intersecting family of sets but $\{\{1,2\},\{2,3\},\{3,4\}\}$ is not.

Theorem 4.0.2 (Erdös-Ko-Rado)**.** *Suppose $n \geq 2k$, and consider the set $[n] = \{1, ..., n\}$. Let $\mathcal{F}$ be a family of some k-element subsets of $[n]$, then $|\mathcal{F}| \leq \binom{n-1}{k-1}$.*

Th original proof of this theorem used induction and was significantly more involved than the one we are about to show. The key idea here is to *randomly* generate a member of $\mathcal{F}$ and compute its probability of not intersecting anything else in $\mathcal{F}$.

Proof. Consider the set $A_s := \{s, s+1, ..., s+k-1\}$, where the numbers are taken modulo n. For example, if $k = 4$ and $n = 9$, then $A_7 = \{7, 8, 9, 1\}$.

Lemma 4.0.3. *The family $\mathcal{F}$ contains at most k elements of the form A_s.*

Proof. Suppose $A_s \in \mathcal{F}$. Note that the A_t that intersect it are all of the form $A_{s\pm i}$ for $1 \leq i \leq k-1$. We can partition them into $k-1$ pairs of the form $\{A_{s-i}, A_{s+k-i}\}$. Note that these two don't intersect, so $\mathcal{F}$ can contain at most one of them. Thus, by the pigeonhole principle $\mathcal{F}$ contains at most k distinct elements of the form A_s, as desired. □

Now consider an arbitrary $A \subseteq [n]$. Instead of generating it concretely, let's do it *randomly*. Take a random permutation σ of $[n]$ and a random $i \in [n]$. Then we consider:

$$A = \{\sigma(i), \sigma(i+1), ..., \sigma(i+k-1)\}$$

again, taking addition modulo n. If we fix the permutation σ, then by the same argument as with the sets A_s, the family $\mathcal{F}$ can contain at most k such sets, for k distinct choices of i. What then, is the probability that A, conditioned on σ, is in $\mathcal{F}$? By the lemma, we can bound it by k/n. However,

there was nothing special about our choice of σ. Thus, if we randomly generated a k-element subset $A \subseteq [n]$, we would have:

$$P(A \in \mathcal{F} \leq \frac{k}{n}$$

Since A was generated uniformly at random, we can exactly describe the probability that it is in $\mathcal{F}$ to begin with: it is simply the ratio of the size of $\mathcal{F}$ to the total possible number of k-element subsets. In other words:

$$P(A \in \mathcal{F}) = \frac{|\mathcal{F}|}{\binom{n}{k}}$$

Substituting into our inequality, above, we get:

$$\frac{|\mathcal{F}|}{\binom{n}{k}} \leq \frac{k}{n} \implies |\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$$

which is exactly the inequality we wanted. □

5 Large number of paths in Tournaments

Definition. A **graph** is a set of points (or *vertices*) V, along with a set of edges E consisting of pairs of vertices (v_1, v_2). Two vertices with an edge between them are said to be **adjacent**.

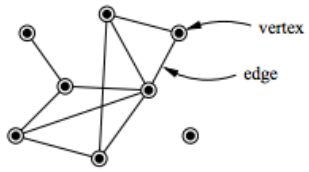

Figure 2: Example of a graph

Definition. A **complete graph** on n vertices is one where every vertex is adjacent to every other vertex. A **tournament** is a complete graph where every edge is directed in one way or the other

Definition. A **Hamiltonian path** along the directed edges is one that traverses every vertex exactly once.

We can ask how many different hamiltonian paths a tournament can have. Does it grow polynomially? Exponentially? Even faster?

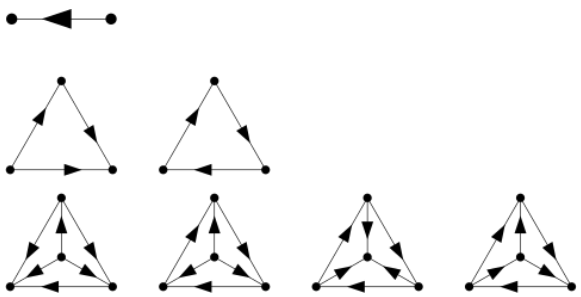

Figure 3: Examples of Tournaments on 2, 3, 4 vertices

Q:. *Can we always find a tournament on n vertices such that there are at least $n!2^{1-n}$ distinct Hamiltonian paths?*

The answer is yes, and we show this using probability. Consider a *random* tournament on n vertices $v_1, v_2 ... v_n$, where we orient each edge in either direction with probability 1/2.

Now take any permutation of the vertices $v_{p(1)}v_{p(2)}...v_{p(n)}$. What's the probability that this sequence of vertices forms a Hamiltonian path? First of all, the edge between $v_{p(1)}$ and $v_{p(2)}$ would have to be directed the correct way (towards $v_{p(2)}$), and that happens with probability 1/2. The same must be true for every edge in the sequence, as only then will a Hamiltonian path exist.

Since all of the edges are oriented independently with probability 1/2, the probability that *all* of them in the path are oriented correctly will be $(1/2)^{n-1} = 2^{1-n}$.

Let X_p be the ***indicator variable*** of whether the permutation p induces a Hamiltonian path. In other words, $X_p = 0$ if the permutation p doesn't induce a Hamiltonian path, and $X_p = 1$ if it does. The expected number of hamiltonian paths is then given by:

$$\begin{aligned} E[X] &= \sum_{\text{permutations } p} E[X_p] \\ &= \sum_{\text{permutations } p} 2^{1-n} \\ &= n!2^{1-n} \end{aligned}$$

Since the expected value is $n!2^{1-n}$, there must exist a random configuration that has *at least* that many Hamiltonian paths. By **Stirling's Approximation**, we know that:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Thus the number of expected possible Hamiltonian paths grows (at least!) exponentially with the number of vertices on a tournament. In fact, it is known that the maximum number of Hamiltonian paths on an n-vertex tournament is at most $\frac{n!}{(2-\epsilon)^n}$, so our result is pretty much optimal. This result was also proved probabilistically by Noga Alon!

6 Bipartite Subgraphs

Given a graph $G = (V, E)$, a **subgraph** of G is what it intuitively sounds like: a subset of vertices along with a subset of edges between them. We now introduce the notion of **bipartedness**, which is an extremeley important graph-theoretic property.

Definition. A graph G is **bipartite** if its vertex set V can be partitioned into two disjoint sets A, B (i.e $V = A \sqcup B$) such that any edge in G connects a vertex in A and a vertex in B. In other words, there are no edges within A or B.

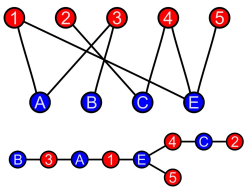

Figure 4: Two examples of bipartite graphs

Definition. A **vertex coloring** of a graph G is an assignment of a color to each vertex $v \in V$ such that if $(v_1, v_2) \in E$, then v_1, v_2 cannot have the same color. In other words, two vertices connected by an edge should be different colors. The **chromatic number** of a graph is the minimum number of distinct colors required for a vertex coloring.

Lemma 6.0.1. *A graph being bipartite is equivalent to it having chromatic number 2*

Proof. By definition, we can partition the vertex sets into two sets A, B such that any edge only connects A to B. Then, coloring everything in A one color and everything in B another color gives a valid 2-coloring of the graph. □

Bipartite graphs are very important to understanding larger graph theory results, but they also pop up outside of math in all sorts of places, from coding theory to (see Tanner graphs for an example) to systems modelling (e.g matching patients to hospitals). As such, a graph theorist might care about something like a maximal bipartite subgraph of any given graph.

Theorem 6.0.2. *Suppose $G = (V, E)$ has n vertices and e edges. Then G contains a bipartite subgraph with at least $e/2$ edges.*

Proof. Again, the goal here will be to think probabilistically. Instead of trying to find a partition of our vertex set V, what if we just partitioned it *randomly*?
For every $v \in V$, let $v \in A$ with probability $1/2$, and define $B = V \setminus A$. Let us call an edge

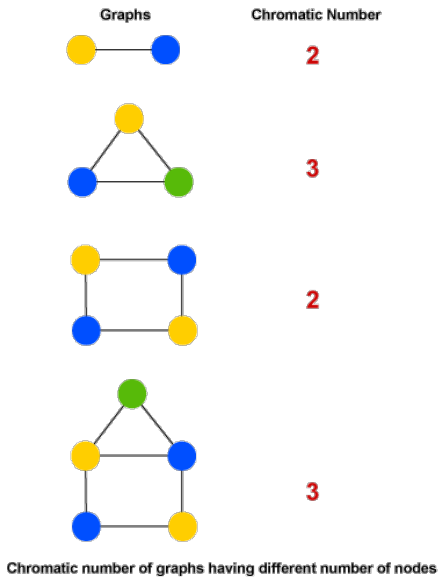

Figure 5: Examples of graphs and their chromatic numbers (check for yourself!)

(v_1, v_2) a *crossing edge* if one of the vertices is in A and the other is in B. To create a bipartite subgraph, we can simply keep every vertex and remove every edge that isn't a crossing edge. If X is a random variable representing the number of crossing edges after this process, let us calculate $E[x]$, the expected number of crossing edges. We note:

$$X = \sum_{(v,w)\in E} X_{v,w}$$

where $X_{v,w}$ is the *indicator variable* (see previous section) for whether an edge is crossing or not. Then by linearity of expectation:

$$\begin{aligned} E[X] &= \sum_{(v,w)\in E} E[X_{v,w}] \\ &= \sum_{(v,w)\in E} 1/2 \\ &= e/2 \end{aligned}$$

Convince yourself that the second line is true, i.e $P(X_{v,w} = 1) = 1/2$. Note that now we have our result: since the expected number of crossings for a random partition of vertices is $e/2$, there must exist one partition with *at least* that many. □

There are of course lots of ways to prove similar results by simply altering the evetn space. For example, when the graph has an even number of edges we can actually do a teensy bit better on our lower bound.

Theorem 6.0.3. *A graph G with $2n$ vertices and e edges contains a bipartite subgraph with at least $e\frac{n}{2n-1}$ edges*

Proof. Before we chose our set A out of all possible subsets of our vertex sets. Now suppose we choose A out of all possible n-element vertex sets. Convince yourself that $P(X_{v,w} = 1) = \frac{n}{2n-1}$, and the proof works exactly the same as the previous one to arrive at the conclusion. □

7 Maximum number of points without large angles

Q:. *How many points can we pick from $\mathbb{R}^n$ such that all angles determined by three points in that set are strictly less than $90°$?*

In 1962, Danzer conjectured that the cardinality of such a set was upper bounded by $2n - 1$ (i.e linear in the dimension). However, a result of Erdös more than 20 years later found an exponential lower bound, using a purely probabilistic construction.

Theorem 7.0.1. *For every $n \geq 1$ there is a set of at least $\lfloor \frac{1}{2}\left(\frac{2}{\sqrt{3}}\right)^n \rfloor$ points in $\mathbb{R}^n$ such that any three points create an angle of less than $90°$.*

Proof. Let C_n be the set of points that are the vertices of an n-dimensional hypercube. These are exactly the elements of $\{0, 1\}^n$, i.e ordered n-tuples of 0 or 1 values. We use a useful bijection:

Lemma 7.0.2. *There exists a bijection between vertices of an n-dimensional hypercube and subsets of $[n] = \{1, ..., n\}$*

Proof. Assign the vector $\vec{a} = (a_1, a_2, ...a_n)$ to the set $S_{\vec{a}} := \{i \in [n] : a_i = 1\}$. Check that this is a bijection. □

Note that within a hypercube, any 3 points must either form a right angle, or something less than a right angle (check this trigonometrically if you wish, but it should hopefully be intuitive).

Lemma 7.0.3. *Three vertices a, b, c of the hypercube, corresponding to sets $A, B, C \subset [n]$ in the above bijection, determine a right angle at c if and only if $A \cap B \subseteq C \subseteq A \cup B$*

Proof. Hint: Use the inverse pythagoras theorem and consider how to find the lengths of ac and bc, and what that corresponds to in terms of the sets A, B, C. □

Define $m := \lfloor \frac{1}{2}\left(\frac{2}{\sqrt{3}}\right)^n \rfloor$, and randomly and independently choose $2m$ vertices of the hypercube, such that each coordinate is chosen to be 0 or 1 with equal probability. Given an arbitary three vertices $\vec{a}, \vec{b}, \vec{c}$, we observe that satisfying the set inclusion of the above lemma is equivalent to, for every i-th coordinate, we want $\vec{a}_i = \vec{b}_i = 0, \vec{c}_i = 1$ and $\vec{a}_i = \vec{b}_i = 1, \vec{c}_i = 0$ to *not* occur. Thus the property that 3 randomly chosen vertices of the hypercube form a right angle occurs with $(3/4)^n$.

The number of possible angles formed with $2m$ points is given by every possible triplet of vertices, multiplied by 3, since the angle could be at any of those 3 vertices. Let X be the random variable for the total number of right angles. It's expectation will be given by:

$$E[X] = \sum_{\text{possible angles}} (3/4)^n = \binom{2m}{3} \cdot 3 \cdot (3/4)^n$$

Check that, by the definition of m, $E[X] \leq m$. In other words, we can pick a set of size $2m$ where the number of right angles is at most m. Each of those right angles occurs at a certain vertex. Deleting those vertices, we can get a set of size *at least* m with no right angles (and thus only containing angles strictly smaller than $90°$), which is what we wanted to show. □

8 Sum-Free Sets

Definition. A set $S \subset \mathbb{N}$ is said to be **sum-free** if you cannot add two elements of the set to get a third.

Example 8.0.1. *The set* $\{1, 2, 3\}$ *is not sum-free since* $1 + 2 = 3$ *and* $1 + 1 = 2$*, but* $\{1, 3\}$ *is sum-free*

Given any set $A \subset \mathbb{N}$, can you always have a "large" sum-free subset?

Theorem 8.0.2. *Every finite* $A \subset \mathbb{N}$ *has a sum-free subset* B *with* $|B| > |A|/3$.

To prove this constructively, or to find some sort of 'algorithm' to generate B given A is unknown. But probability comes to our aid.

Proof. Take $A = \{a_1, a_2...a_n\}$, and WLOG $a_1 < a_2 < ...a_n$. We use a fact that involves a special case of *Dirichlet's Theorem.*

Fact. *There are infinitely many primes of the form* $3k + 2$.

Choose some prime p of the form $3k + 2$, such that $p > 2a_n$. Let us define a new set $C := \{k + 1, k + 2...2k + 1\}$. Observe that C is sum-free.

Claim. *Let* $B = \{b_1, b_2...b_m\} \subseteq A$*. Take some number* r*, and if* $B' = \{rb_i \pmod p\}$ *consists of distinct numbers in* C*, then* B *is sum-free.*

The proof of the claim is as follows: if B is not sum free, then we have $b_i + b_j = b_k$. Multiplying by r, we would get $rb_i + rb_j = rb_k$, and since these are all distinct numbers in C, it would imply C is not sum-free, which is a contradiction.

For each possible value $r \in \{1, 2...p\}$, we consider the set $A_r := \{rb_1, rb_2...rb_m\} \pmod p$. By the claim, we note that $A_r \cap C$ is a sum-free set. Thus, we now have to try and see if we can find an r.

Our plan? Let us *randomly* select r, where each element of $\mathbb{Z}/p\mathbb{Z}$ is chosen with probability $1/p$. Now comes an intuitive but important point: ***since multiplication by r is an injective map, the probability that $ra_i = x$ is $1/p$ for all $x \in \mathbb{Z}/p\mathbb{Z}$.***

Note that $|C| > |\mathbb{Z}/p\mathbb{Z}|/3$. Thus, for an arbitrary element a_i, we can create an *indicator variable* X_i that is 1 if $ra_i \in C$ and 0 if $ra_i \notin C$. Note that $E[X_i] = P(ra_i \in C)$. If Y is a random variable for the size of $A_r \cap C$, then by linearity of expectation we have:

$$\begin{aligned} E[Y] &= \sum_{i}^{n} E[X_i] \\ &= \sum_{i}^{n} P(ra_i \in C) \\ &\geq \sum_{i}^{n} 1/3 \\ &= |A|/3 \end{aligned}$$

Since the expected value of $|A_r \cap C|$ for some *random* r is at least $|A|/3$, then there must exist some r such that $|A_r \cap C| \geq |A|/3$, and this A_r will be sum-free set.

□