Probabilistic Method: Making Life (sometimes) Easier

Paul Simanjuntak

16/11/2019

The problem of counting is as old as humankind itself. Combinatoric problems have been a staple of mathematical competitions, as they require nothing more than ingenious methods of counting. While these arguments are certainly beautiful, they are not the only ones. We will introduce a form of argument, first introduced by Érdös Pál and Rényi Alfréd. The basics statement is as follows:

> If, in a given set of objects, the probability that an object does not have a certain property is less than 1, then there must exist an object with this property.

In other words, we have an *existence* result. Generally, it might be (much) harder to find this object, but at least we know it exists. The method is less about large theorems than about casting everything in a new setting. In this talk, we will present four examples of how this could be done, all but one comes from the field of combinatorics.
The first two will be about colouring a finite set. First we give a definition:

Definition 0.1. *2-colourable* Let X be a finite set and $\mathcal{F} \subseteq 2^X$ so that each set in $\mathcal{F}$ is of size $d \geq 2$. We say that the family $\mathcal{F}$ is *2-colourable* if there exists a colouring of X with two colours such that in in every set in $\mathcal{F}$, both colours appear.

It is not hard to find a set that is not 2-colourable. For example, if we take all the 2-sets of a set of size 3, then no matter how we colour X, there must be a 2-subset of the same colour (in general, the collection of all d-sets in a set of size $2d-1$ is not 2-colourable). On the other hand, any sub-family of 2-colourable d-sets is again 2-colourable. It is then natural to ask the following question: what is the largest number $m = m(d)$ so that every family of d-sets with m members is 2-colourable?

Proposition 0.2. *Every family of at most 2^{d-1} d-sets is 2-colourable, i.e. $m(d) > 2^{d-1}$.*

Proof. Suppose $\mathcal{F}$ is a family of d-sets with at most 2^{d-1} members. Colour X randomly with 2 colours with equal probability (e.g. flip a coin each time). For each set $A \in \mathcal{F}$, let E_A be the event that all elements of A have the same colour. Since there are just two colours,

$$\mathbb{P}(E_A) = \left(\frac{1}{2}\right)^{d-1}$$

and hence $m = |\mathcal{F}| \leq 2^{d-1}$ (because otherwise the expected value of a monochrome set is greater than 1). The events E_A are not disjoint, so

$$\mathbb{P}\left(\bigcup_{A\in\mathcal{F}} E_A\right) < \sum_{A\in\mathcal{F}} \mathbb{P}(E_A) = m\left(\frac{1}{2}\right)^{d-1} \leq 1$$

Thus there exists some 2-colouring of X without an unicoloured d-set from $\mathcal{F}$, i.e. this family is 2-colourable.

□

This first example is quite simple: first we randomize the colouring, then compute the probability of the events of monochromatic colouring, which we found to be strictly less than 1. We'll need something a bit more for the next example.
This example comes from Erdös' original idea: to compute Ramsey's number. First the definition: consider the complete graph (graph where any two vertices is connected by an edge) K_N on N vertices. We say that K_N has property (m, n) if, no matter how we colour the edges of K_N (traditionally with red and blue), there is always a complete subgraph of m vertices with all red edges or any with n vertices with all blue edges. If K_N has this property, so does K_s for all $s \geq N$. This brings the following definition:

Definition 0.3. The smallest number N so that a complete graph K_N has property (m, n) is called the Ramsey number $R(m, n)$.

Let's begin with a simple observation: $R(m, 2) = m$ because either all of the edges of K_m are red or there is a blue edge, resulting in a blue K_2. By symmetry, we have $R(2, n) = n$. In fact we know all them are finite.

Lemma 0.4 (Ramsey 1929)**.** *For any integer k,* $R(k, k) \leq 2^{2k-3}$.

Now we come to the original proof of Erdös.

Theorem 0.5 (Erdös 1947)**.** *For all* $k \geq 2$*, then* $R(k, k) \geq \lfloor 2^{k/2} \rfloor$.

Proof. From previous observation, we know that $R(2, 2) = 2$. The lemma above says $R(3, 3) \leq 6$, and the example of the pentagon forces $R(3, 3) = 6$.
Now consider when $k \geq 4$. Suppose $N < 2^{k/2}$ and colour each edge red or blue with equal probability. All colourings are equally likely, each with probability $2^{-\binom{N}{2}}$. Let A be a set composed of k vertices. The probability of the 'all-red' event A_R is then $2^{-\binom{k}{2}}$. It then follows that the probability for some k-set to be 'all red' is bounded by

$$p_R := \mathbb{P}\left(\bigcup_{|A|=k} A_R\right) \leq \sum_{|A|=k} \mathbb{P}(A_R) = \binom{N}{k} 2^{-\binom{k}{2}}$$

Because

$$\binom{N}{k} = \frac{N(N-1)\cdots(N-k+1)}{k!} \leq \frac{N^k}{2^{k-1}}$$

we then have

$$\binom{N}{k} 2^{-\binom{k}{2}} \leq \frac{N^k}{2^{k-1}} 2^{-\binom{k}{2}} < 2^{\frac{k^2}{2} - \binom{k}{2} - k + 1} = 2^{-\frac{k}{2}+1} \leq \frac{1}{2}$$

for $k \geq 4$. Hence $p_R < \frac{1}{2}$. By symmetry, $p_B < \frac{1}{2}$, where p_B is the probability of some 'all blue' k vertices. Hence $p_R + p_B < 1$ for $N < 2^{k/2}$, so there must be a colouring of K_N that fails the property (k, k). This proves the theorem.

□

The next statement is interesting for two reasons. First, it actually has a very elegant non-probabilistic proof. Second, the probabilistic proof doesn't even need involve us computing some probability.

Theorem 0.6. *Let $v_1, \cdots, v_n \in \mathbb{R}^n$, all $\|v_i\|_2 = 1$. Then there exists $\varepsilon_1, \cdots, \varepsilon_n = \pm 1$ so that*

$$\|\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n\|_2 \leq \sqrt{n}$$

and also there exists $\varepsilon_1, \cdots, \varepsilon_n = \pm 1$ so that

$$\|\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n\|_2 \geq \sqrt{n}$$

Prüfchen. (Without coin toss)
The proof is by iteration. Start by any choice on ε_1. For any $2 \leq i \leq n$, compute first the partial sum $w = \varepsilon_1 v_1 + \cdots + \varepsilon_{i-1} v_{i-1}$. If we want the sum to be small, choose $\varepsilon_i \in \{-1, 1\}$ so that $\varepsilon_i v_i$ makes an obtuse (or right) angle with w. For the opposite case, choose so that the angle is acute (or right). In the extreme case where we always right angles, then the complete sum w will has norm $\sqrt{n}$, otherwise it is strictly less or larger than $\sqrt{n}$, as desired.

□

Proof. (Probabilistic proof)
Let $\varepsilon_1, \cdots, \varepsilon_n$ be chosen independently and uniformly from $\{-1, +1\}$. Set the random variable

$$X = \|\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n\|_2^2$$

This is a finite sum, which we can rewrite as

$$X = \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \varepsilon_j (v_i \cdot v_j)$$

Thus

$$\mathbb{E}(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j \, \mathbb{E}(\varepsilon_i \varepsilon_j)$$

Now note that because the ε_i's are chosen independently, $\mathbb{E}(\varepsilon_i \varepsilon_j) = \mathbb{E}(\varepsilon_i)\mathbb{E}(\varepsilon_j) = 0$ whenever $i \neq j$. On the other hand, $\varepsilon_i^2 = 1$, so $\mathbb{E}(\varepsilon_i^2) = 1$. Thus the sum collapses to

$$\mathbb{E}(X) = \sum_{i=1}^{n} v_i \cdot v_i = n$$

Because X is not constant, there exists some choices of $\varepsilon_1, \cdots, \varepsilon_n = \pm 1$ so that either $X \geq n$ or $X \leq n$. Take square root to get the theorem. □

In fact, we can give a slightly better bound if we allow the v_i's to be in the disk, not just the sphere.

Corollary 0.7. *Let $v_1, \cdots, v_n \in \mathbb{R}^n$ so that $\|v_i\|_{\leq} 1$ for $1 \leq i \leq n$. Let $p_1, \cdots, p_n \in [0, 1]$ be arbitrary and set $w = p_1 v_1 + \cdots + p_n v_n$. Then there exists $\varepsilon_1, \cdots, \varepsilon_n \in \{0, 1\}$ so that, if $v = \varepsilon_1 v_1 + \cdots + v_n \varepsilon_n$,*

$$\|w - v\|_2 \leq \frac{\sqrt{n}}{2}$$

This last example shows that we can get close to a point in an n-simplex by some combination of the same points. If we restrict ourselves to convex combination and normalize the lenght, we may consider it to be some sort of approximation of an old theorem of Caratheodory.

Theorem 0.8 (Caratheodory 1907)**.** *If $x \in \mathbb{R}^d$ lie in the convex hull of a set P (which may be discrete), then x can be written as the convex combination of at most $d+1$ point in P.*

The proof of this statement is purely linear algebraic, but if getting close is good enough, we can give a probabilistic proof.

Theorem 0.9 (Approximate to Caratheodory's)**.** *Consider a set $T \subseteq \mathbb{R}^n$ whose diameter is bounded by 1. Then, for every x in the convex hull of T and every integer k there are points $x_1, \cdots, x_k \in T$ (not necessarily distinct) so that*

$$\left\| x - \frac{1}{k} \sum_{j=1}^{k} x_j \right\|_2 \leq \frac{1}{\sqrt{k}}$$

Proof. By translation, we may assume the radius of T is also bounded by 1, i.e. $\|y\|_2 \leq 1$ for all $y \in T$.
Fix a point x in the convex hull and express it as a convex combination of $z_1, \cdots, z_m \in T$ with coefficients $\lambda_1, \cdots, \lambda_m$. Now we'll interpret this probabisticially: define a random variable Z so that

$$\mathbb{P}(Z = z_i) = \lambda_i \qquad i = 1, \cdots, m$$

Then

$$\mathbb{E}\, z = \sum_{i=1}^{m} \lambda_i z_i = x$$

Now consider independent copies $Z_1, Z_2, \cdots$ of Z. By the law of large number,

$$\frac{1}{k} \sum_{j=1}^{k} Z_j \to x \text{ a.s. as } k \to \infty$$

To get the claimed rate of convergence, we'll compute the variance of this average. Because $\mathbb{E}(Z_i - x) = 0$,

$$\mathbb{E} \left\| x - \frac{1}{k} \sum_{j=1}^{k} Z_j \right\|_2^2 = \frac{1}{k^2} \sum_{j=1}^{k} \left\| Z_j - x \right\|_2^2$$

It then remains to bound the variance of each Z_j.

$$\mathbb{E} \left\| Z_j - x \right\|_2^2 = \mathbb{E}\, \|Z\|_2^2 - \|\mathbb{E}\, Z\|_2^2 \leq \|Z\|_2^2 \leq 1$$

where the last inequality uses the centering of T. Hence,

$$\mathbb{E} \left\| x - \frac{1}{k} \sum_{j=1}^{k} Z_j \right\|_2^2 \leq \frac{1}{k}$$

So there must be a combination of the Z_j's so that

$$\left\| x - \frac{1}{k} \sum_{j=1}^{k} Z_j \right\|_2^2 \leq \frac{1}{k}$$

Because by definition Z_j takes value in T, there are some points $x_1, \cdots x_k \in T$ so that the wanted conclusion holds.

□