CCS Discrete II	Professor: Padraic Bartlett
Lecture 8: The Probabilistic Method	
Week 8	*UCSB 2015*

In last week's lectures, we defined the Ramsey numbers $R(k, l)$, and found upper bounds on them! This week, we will study one particularly powerful technique for finding **lower bounds** on Ramsey numbers: the **probabilistic method in combinatorics**!

1 The Probabilistic Method

1.1 Background and Definitions

The probabilistic method in combinatorics first arose in 1947, when Erdös used it to prove the following claim:

Theorem 1. $R(k, k) > \lfloor 2^{k/2} \rfloor$.

This result at the time was far better than anything we currently knew; indeed, our best bounds to the present day without using the probabilistic method are roughly $O(k^3)$, which is far smaller than $2^{k/2}$ for any remotely large value of k!

To go through his methods, however, we first need an introduction to the basics of probability! We do this here.

Definition. A **finite probability space** consists of two things:

- A finite set Ω.
- A **measure** Pr on Ω, such that $Pr(\Omega) = 1$. In case you've forgotten, saying that Pr is a measure is a way of saying that Pr is a function $\mathcal{P}(\Omega) \to \mathbb{R}^+$, such that the following properties are satisfied:
  - $Pr(\emptyset) = 0$.
  - For any collection $\{X_i\}_{i=1}^{\infty}$ of subsets of Ω, $Pr\left(\bigcup_{i=1}^{\infty} X_i\right) \leq \sum_{i=1}^{n} Pr(X_i)$.
  - For any collection $\{X_i\}_{i=1}^{\infty}$ of **disjoint** subsets of Ω, $Pr\left(\bigcup_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{n} Pr(X_i)$.

For example, one probability distribution on $\Omega = \{1, 2, 3, 4, 5, 6\}$ could be the distribution that believes that $Pr(\{i\}) = 1/6$ for each individual i, and more generally that $Pr(S) = |S|/6$ for any subset S of Ω. In this sense, this probability distribution is capturing the idea of rolling a fair six-sided die, and seeing what comes up.

This sort of "fair" distribution has a name: namely, the **uniform** distribution!

Definition. The **uniform** distribution on a finite space Ω is the probability space that assigns the measure $|S|/|\Omega|$ to every subset S of Ω. In a sense, this measure thinks that any two elements in Ω are "equally likely;" think about why this is true!

We have some useful notation and language for working with probability spaces:

Definition. An **event** S is just any subset of a probability space. For example, in the six-sided die probability distribution discussed earlier, the set $\{2, 4, 6\}$ is an event; you can think of this as the event where our die comes up as an even number. The probability of an event S occurring is just $Pr(S)$; i.e. the probability that our die when rolled is even is just $Pr(\{2, 4, 6\}) = 3/6 = 1/2$, as expected.

Notice that by definition, as Pr is a measure, for any two events A, B, we always have $Pr(A \cup B) \leq Pr(A) + Pr(B)$. In other words, given two events A, B, the probability of either A or B happening (or both!) is at most the probability that A happens, plus the probability that B happens.

Definition. A real-valued **random variable** X on a probability space Ω is simply any function $\Omega \to \mathbb{R}$.

Given any random variable X¡ we can talk about the **expected value** of X; that is, the "average value" of X on Ω, where we use Pr to give ourselves a good notion of what "average" should mean. Formally, we define this as the following sum:

$$\sum_{\omega in \Omega} Pr(\omega) \cdot X(\omega).$$

For example, consider our six-sided die probability space again, and the random variable X defined by $X(i) = i$ (in other words, X is the random variable that outputs the top face of the die when we roll it.)

The expected value of X would be

$$\sum_{\omega in \Omega} Pr(\omega) \cdot X(\omega) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{21}{6} = \frac{7}{2}.$$

In other words, rolling a fair six-sided die once yields an average face value of 3.5.

With this notation established, we can use it to solve some problems in graph theory!

1.2 Applications to Graphs

Theorem 2. $R(k, k) > \lfloor 2^{k/2} \rfloor$.

Proof. Fix some value of n, and consider a random uniformly-chosen 2-coloring of K_n's edges: in other words, let us work in the probability space $(\Omega, Pr) =$ (all 2-colorings of K_n's edges, $Pr(\omega) = 1/2^{\binom{n}{2}}$.)

For some fixed set R of k vertices in $V(K_n)$, let A_R be the event that the induced subgraph on R is monochrome. Then, we have that

$$Pr(A_R) = 2 \cdot \left(2^{\binom{n}{2} - \binom{k}{2}}\right) / 2^{\binom{n}{2}} = 2^{1 - \binom{k}{2}}.$$

Thus, we have that the probability of at least one of the A_R's occuring is bounded by

$$Pr(\bigcup_{|R|=k} A_R) \leq \sum_{R \subset \Omega, |R|=k} Pr(A_R) = \binom{n}{k} 2^{1 - \binom{k}{2}}.$$

If we can show that $\binom{n}{k}2^{1-\binom{k}{2}}$ is less that 1, then we know that with nonzero probability there will be some 2-coloring $\omega \in \Omega$ in which none of the A_R's occur! In other words, we know that there is a 2-coloring of K_n that avoids both a red and a blue K_k.

Solving, we see that

$$\binom{n}{k}2^{1-\binom{k}{2}} < \frac{n^k}{k!} \cdot 2^{1+(k/2)-(k^2/2)} = \frac{2^{1+k/2}}{k!} \cdot \frac{n^k}{2^{k^2/2}} < 1$$

whenever $n = \lfloor 2^{k/2} \rfloor, k \geq 3$. So we're done!

□

So: why did we do this? In other words, what did using probabilistic methods gain us?

The answer, essentially, is that the probabilistic method allows us to work with graphs that are both *large* and *unstructured*! When using constructive methods, we can rarely (if at all) do this! I.e.:

- If you're trying to construct a large graph by gluing together pieces of smaller graphs, you are almost always inducing a lot of structure into your larger graph; consequently, your construction will usually be a highly atypical graph! For example, try constructing a graph of both girth and chromatic number greater than 6 – you'll quickly find that it's stunningly difficult to avoid introducing structure in any building method that won't create small cycles or small chromatic numbers. Yet, using the probabilistic method we can easily show that there are graphs of arbitrarily high girth and chromatic number! – in fact, that almost all sufficiently large graphs are such things.

- Conversely, suppose that you're trying to avoid such problems, and have decided to simply check by hand all of the cases for some reasonably small number of vertices – say, 20. But there are $2^{\binom{20}{2}} = 2^{190} \approx 1.5 * 10^{57}$ such graphs! Even with stunningly powerful supercomputers, there's no hope. Yet, with the probabilistic method, we will routinely create counterexamples with $> 10^{10}$ vertices in them! – things we could never hope to find in any deterministic search.

We give another simple example here:

Theorem 3. *If G is a graph, then G contains a bipartite subgraph with at least $E/2$ edges.*

Proof. Pick a subset of G's vertices, T, uniformly at random (i.e. select T by flipping a coin for each of G's vertices, and placing vertices in T iff our coin comes up heads.) Let $B = V(G) \setminus T$.

Call an edge $\{x, y\}$ of $E(G)$ **crossing** iff exactly one of x, y lie in T, and let X be the random variable defined by

$$X(T) = \text{number of crossing edges for } T.$$

Then, we have that

$$X(T) = \sum X_{x,y}(T),$$

where $X_{x,y}(T)$ is the 0-1 random variable defined by $X_{x,y}(T) = 1$ if $\{x, y\}$ is an edge of G that's crossing, and 0 otherwise.

The expectation $\mathbb{E}(X_{x,y})$ is clearly $1/2$, because we chose x and y to be in T at random. Thus, by the linearity of expectation, we have that

$$\mathbb{E}(X) = \sum \mathbb{E}(X_{x,y}) = E/2.$$

so the expected number of crossing edges for a random subset of G is $E/2$. Thus, there must be some $T \subset V(G)$ such that $X(T) \geq E/2$; taking the collection of crossing edges this set creates then gives us a bipartite graph (B, T) with $\geq E/2$ edges in it. $\square$

1.3 A Trickier Example: Graphs with Arbitrarily High Girth and Chromatic Number

In past lectures, we used the Mycielski construction to create triangle-free graphs with arbitrarily high chromatic number; on the HW, we looked at a different construction that created C_3, C_4, C_5-free graphs with arbitrarily high chromatic number. A natural question you might ask is whether we can generalize this to find graphs with arbitrarily large girth[1] and chromatic number!

In this section, we will develop Markov's inequality, a remarkably useful tool, and show how its application can give us (with relatively little effort, especially when compared to constructive methods!) graphs of arbitrarily high girth and chromatic number.

Theorem 4. *(Markov's Inequality) For a random variable X and any positive constant a, $Pr(|X| > a) \leq \mathbb{E}(|X|)/a$.*

Solution. So: let $I_{(|X|\geq a)}$ be the indicator random variable defined by

$$I_{(|X|\geq a)}(\omega) = \begin{cases} 1 & |X(\omega)| \geq a, \\ 0 & \text{otherwise.} \end{cases}$$

Then, for any ω in Ω, we trivially have that $a \cdot I_{(|X|\geq a)}(\omega) \leq |X(\omega)|$; consequently, we have

$$\begin{aligned}
\mathbb{E}(a \cdot I_{(|X|\geq a)}) &= \sum_{\omega\in\Omega} a \cdot I_{(|X|\geq a)} Pr(\omega) \\
&= \sum_{\omega\in\Omega:|X(\omega)|\geq a} aPr(\omega) \\
&\leq \sum_{\omega\in\Omega:|X(\omega)|\geq a} |X(\omega)|Pr(\omega) \\
&\leq \sum_{\omega\in\Omega} |X(\omega)|Pr(\omega) \\
&= \mathbb{E}(X).
\end{aligned}$$

[1]A graph has **girth** k if it contains a cycle of length k, and no smaller cycles. Cycle-free graphs have"infinite" girth.

However, on the other hand, we have

$$\begin{aligned}\mathbb{E}(a \cdot I_{(|X|\geq a)}) &= \sum_{\omega \in \Omega} a \cdot I_{(|X|\geq a)} Pr(\omega) \\ &= a \cdot \sum_{\omega \in \Omega : |X(\omega)| \geq a} Pr(\omega) \\ &= a \cdot Pr(|X| \geq a);\end{aligned}$$

combining, we have $Pr(|X| > a) \leq \mathbb{E}(|X|)/a$.

So: with this theorem under our belt, we now have the tools to resolve the following graph theory question (which otherwise is fairly hard to surmount:)

Theorem 5. *There are graphs with arbitarily high girth and chromatic number.*

Proof. So: let $G_{n,p}$ denote a random graph on n vertices, formed by doing the following:

- Start with n vertices.
- For every pair of vertices $\{x, y\}$, flip a biased coin that comes up heads with probability p and tails with probability $1 - p$. If the coin is heads, add the edge $\{x, y\}$ to our graph; if it's tails, don't.

Our roadmap, then, is the following:

- For large n and well-chosen p, we will show that $G_{n,p}$ will have relatively "few" short cycles at least half of the time.
- For large n, we can also show that G will have high chromatic number at least half the time.
- Finally, by combining these two results and deleting some vertices from our graph, we'll get that graphs with both high chromatic number and no short cycles exist in our graph.

To do the first: fix a number l, and let X be the random variable defined by $X(G_{n,p}) =$ the number of cycles of length $\leq l$ in $G_{n,p}$.

We then have that

$$X(G_{n,p}) \leq \sum_{j=3}^{l} \sum_{all j-tuples x_1 \ldots x_j} N_{x_1 \ldots x_j},$$

where $N_{x_1 \ldots x_j}$ is the event that the vertices $x_1 \ldots x_j$ form a cycle.

Then, we have that

$$\begin{aligned}
\mathbb{E}(X) &\leq \sum_{j=3}^{l} \sum_{j-tuples\ x_1 \ldots x_j} Pr(N_{x_1 \ldots x_j}) \\
&= \sum_{j=3}^{l} \sum_{j-tuples\ x_1 \ldots x_j} p^j \\
&= \sum_{j=3}^{l} n^j p^j.
\end{aligned}$$

To make our sum easier, let $p = n^{\lambda - 1}$, for some $\lambda \in (0, 1/l)$; then, we have that

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{j=3}^{l} n^j p^j \\
&= \sum_{j=3}^{l} n^j n^{j\lambda - j} \\
&= \sum_{j=3}^{l} n^{j\lambda} \\
&< \frac{n^{\lambda(l+1)}}{n^{\lambda} - 1} \\
&= \frac{n^{\lambda l}}{1 - n^{-\lambda}}
\end{aligned}$$

We claim that this is smaller than n/c, for any c and sufficiently large n. To see this, simply multiply through; this gives you that

$$\begin{aligned}
&\frac{n^{\lambda l}}{1 - n^{-\lambda}} < n/c \\
\Leftrightarrow & n^{\lambda l} < n/c - n^{1-\lambda}/c \\
\Leftrightarrow & n^{\lambda l} + n^{1-\lambda}/c < n/c
\end{aligned}$$

,

which, because both λl and $1 - \lambda$ are less than 1, we know holds for large n.

So: to recap: we've shown that

$$\mathbb{E}(|X|) < n/4.$$

So: what happens if we apply Markov's inequality? Well: we get that

$$Pr(|X| \geq n/2) \leq \frac{\mathbb{E}(|X|)}{n/2} < \frac{n/4}{n/2} = 1/2;$$

in other words, that more than half of the time we have relatively "few" short cycles! So this is the first stage of our theorem.

Now: we seek to show that the chromatic number of our random graphs will be "large," on average. Doing this directly, by working with the chromatic number itself, would be rather ponderous: as week two of our class showed, the chromatic number of a graph can be a rather mysterious thing to calculate.

Rather, we will work with the **independence number** $\alpha(G)$ of our graph, the size of the independent set of vertices[2] in our graph. Why do we do this? Well, in a proper k-coloring of a graph, each of the colors necessarily defines an independent set of vertices, as there are no edges between vertices of the same color; ergo, we have that

$$\chi(G) \geq \frac{|V(G)|}{\alpha(G)},$$

for any graph G.

So: to make the chromatic number large, it suffices to make $\alpha(G)$ small! So: look at $Pr(\alpha(G) \geq m)$, for some value m. We then have the following:

$$\begin{aligned} Pr(\alpha(G) \geq m) &= Pr(\text{there is a subset of } G \text{ of size } m \text{ with no edges in it}) \\ &\leq \sum_{S \subset V, |S|=m} Pr(\text{there are no edges in } S\text{'s induced subgraph}) \\ &= \binom{n}{m} \cdot (1-p)^{\binom{m}{2}} \end{aligned}$$

,

as there are $\binom{n}{m}$-many such subsets, and in order for there to be no edges in S's subgraph we need merely for all of the coin-flips that go into creating S to come up tails: i.e. this happens $(1-p)^{\binom{m}{2}}$ of the time.

So: note the following useful inequalities:

- $\binom{n}{m} \leq n^m$. To see this, expand the binomial coefficient into the form $\frac{n \cdot \ldots (n-m+1)}{m!}$, discard the $m!$, and bound the m terms in the numerator by n^m. It's a awful bound, but one that is really easy to work with!

- $(1-p) < e^{-p}$. There are a number of proofs of this, which can be regarded as a simplified form of Bernoulli's inequality; elementary calculus methods should suffice to give this to you!

Applying both of these inequalities, we have that

$$\begin{aligned} Pr(\alpha(G) \geq m) &< n^m \cdot e^{-p \cdot \binom{m}{2}} \\ &= n^m \cdot e^{-p \cdot m \cdot (m-1)/2}. \end{aligned}$$

[2]Recall: a set of vertices is called **independent** if their induced subgraph has no edges. in it

So: motivated by a desire to make the above simple, let $m = \left\lceil \frac{3}{p} \ln(n) \right\rceil$. This then gives us that

$$\begin{aligned} Pr(\alpha(G) \geq m) &< n^m \cdot e^{-p \cdot \left\lceil \frac{3}{p} \ln(n) \right\rceil \cdot (m-1)/2} \\ &= n^m \cdot n^{-3(m-1)/2}, \end{aligned}$$

which goes to 0 as n gets large. So, in particular, we know that for large values of n and any m, we have

$$Pr(\alpha(G) \geq m) < 1/2.$$

So: let's combine our results! In other words, we've successfully shown that for large n,

$$Pr(G \text{ has more than } (n/2)\text{-many cycles of length } \leq l, \text{ or} \alpha(G) \geq m) < 1.$$

So: for large n, there is a graph G so that neither of these things happen! Let G be such a graph. G has less than $n/2$-many cycles of length $\leq l$; so, from each such cycle, delete a vertex. Call the resulting graph G'.

Then, we have the following:

- By construction, G' has girth $\geq l$.
- Also by construction, G' has at least $n/2$ many vertices, as it started with n and we deleted $\leq n/2$.
- Deleting vertices doesn't decrease the independence number of a graph! (If you're not sure why this is, look at the definition and work it out!) Thus, we have that

$$\begin{aligned} \chi(G') &\geq \frac{|V(G')|}{\alpha(G')} \\ &\geq \frac{n/2}{\alpha(G)} \\ &\geq \frac{n/2}{3\ln(n)/p} \\ &= \frac{n/2}{3n^{1-\lambda}\ln(n)} \\ &= \frac{n^\lambda}{6\ln(n)}, \end{aligned}$$

  which goes to infinity as n grows large.

So, for large n, this graph has arbitrarily large girth and chromatic number! □