T-79.5201 Discrete Structures

Autumn 2007: Probabilistic Combinatorics

Lectures: Pekka Orponen, Wed 12-14, TB353

Tutorials: ——"——, Wed 14-15, TB353

Textbook: N. Alon, J. H. Spencer,
The Probabilistic Method, 2nd Ed.
J. Wiley & Sons, 2000.

---

1. The Probabilistic Method

- First (?) used by Tibor Szele in 1943 to prove the existence of tournaments (orientations of complete graphs) with many Hamiltonian paths.

- Popularized by Paul Erdös & Alfred Rényi in the 1960's as a powerful nonconstructive method to prove the existence or nonexistence of all kinds of combinatorial structures.

- Presently of great significance in theoretical computer science e.g. because of the increased importance of randomised algorithms. ("Almost all structures S have interesting property P. Thus, let the algorithm first <u>guess</u> a random structure S. Then, assuming P holds...")

- Basic idea: consider a large family Ω_n of combinatorial objects of "size" n (e.g. graphs with n vertices, colourings of the complete graph K_n, permutations on $[n] = \{1, 2, \dots, n\}$ etc.)

  To prove the existence of some object $S \in \Omega_n$ with property P, define some probability distribution on Ω_n (usually the uniform distribution) and estimate

  $$\Pr_S(P_n) = \Pr(\text{random } S \in \Omega_n \text{ has property } P).$$

  If $\Pr_S(P_n) > 0$, then there is at least one $S \in \Omega_n$ with property P.

- This sounds like a fancy reformulation of a simple counting argument, but often:

  (a) the probabilistic view is more natural & transparent than counting; and

  (b) tools from probability theory can be brought to help in the counting (expectations, moment bounds, tail probability estimates, ...)

1.1 Simple Examples

- Example 1. Ramsey numbers

Consider a red-blue colouring of the edges of the complete graph K_n on vertex set $[n] = \{1, 2, \ldots, n\}$.

A set $V \subseteq [n]$ of size k (or the corresp. subgraph K_k) is monochromatic (briefly mono-χ) if all edges within V are either red or blue.

F.P. Ramsey proved (1930) that

$\forall k \, \exists n_k : n \geq n_k \rightarrow$ any red-blue colouring of K_n contains a mono-χ K_k.

For any k, the smallest such n_k is called the Ramsey number $R(k)$.

Ramsey numbers are notoriously hard to compute:

$R(2) = 2$ [trivial]
$R(3) = 6$ [easy]
$R(4) = 18$ [Greenwood & Gleason 1955]
$R(5) =$ unknown (best known bounds: $43 \leq R(5) \leq 49$)

- Proposition 1.1 (Erdös 1947)

$R(k) > 2^{k/2}$ for all $k \geq 4$.

[In fact the bound holds also for $k=3$, by direct calculation.]

Proof. For a given k, let us consider random two-colourings C of K_n and the probability

$$p_n = \Pr_C(K_n \text{ contains no mono-}\chi\ K_k \text{ under } C).$$

[Here "random" ≡ each edge coloured red/blue indep. w/ prob. $1/2$.]

- If for some n we can show that $p_n > 0$, then there is at least one two-colouring C of K_n with no mono-χ K_k, and so $R(k) > n$.

(In fact, there will be exactly $p_n \cdot 2^{\binom{n}{2}}$ two-colourings C of K_n with this property, but this is of no concern to us presently.)

- Now for given n, the prob. that a fixed k-set $V \subseteq [n]$ induces a mono-χ K_k (event "A_V" occurs) is:

$$\Pr(A_V) = 2 \cdot \left(\tfrac{1}{2}\right)^{\binom{k}{2}} = 2^{1-\binom{k}{2}}.$$

There are altogether $\binom{n}{k}$ k-sets $V \subseteq [n]$, so the prob. that some of them is mono-χ is:

$$q_n = \Pr\Big(\bigcup_{|V|=k} A_V\Big) \leq \sum_{|V|=k} \Pr(A_V) = \binom{n}{k} \cdot 2^{1-\binom{k}{2}}$$

- Since $p_n = 1 - q_n$, we have shown that if

$$(*) \quad \binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1,$$

then $R(k) > n$.

- For which values of n does $(*)$ hold? Roughly speaking

$$\binom{n}{k} \sim n^k, \quad 2^{\binom{k}{2}} \sim 2^{\frac{k^2}{2}},$$

so approximately $(*)$ holds when

$$n^k \cdot 2^{-k^2/2} \lesssim 1, \quad \text{i.e.} \quad n \lesssim 2^{k/2}.$$

- More precisely:

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

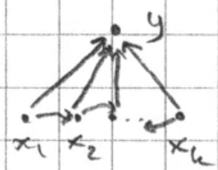
$$< 1 \quad \text{for } n = 2^{k/2},\ k \geq 4,$$

as can be verified. □

- <u>Example</u> 2. Tournaments.

A <u>tournament</u> of n players is an orientation of the edges of K_n. ($x \to y$ ≡ "x loses to y", "y beats x")

Tournament T has <u>property</u> S_k if for any set of k players there is one who beats them all.



- Can one for any fixed k arrange for a tournament with property S_k?

- <u>Theorem</u> 1.2 (Erdös 1963)

If $\binom{n}{k}(1-2^{-k})^{n-k} < 1$, then there is a tournament on K_n with property S_k.

<u>Proof</u> Consider a random tournament T on K_n.
[here "random" ≡ each edge oriented independently either way with prob. ½.]

For every fixed set $V \subseteq [n]$ of k vertices (players), the prob. that <u>no</u> other vertex beats them all is

$$\Pr_T(A_V) = (\underbrace{1-\underbrace{2^{-k}}_{\text{outside vertex } u \text{ beats all of } V}}_{\text{vertex } u \text{ does } \underline{\text{not}} \text{ beat all of } V})^{n-k} \longleftarrow \text{orientations to each outside vertex are independent}$$

- There are altogether $\binom{n}{k}$ k-sets $V \subseteq [n]$, so the prob. that <u>some</u> of them is not dominated by any outside vertex is

$$q_n = \Pr\Big(\bigcup_{|V|=k} A_V\Big) \leq \sum_{|V|=k} \Pr(A_V) = \binom{n}{k}\cdot(1-2^{-k})^{n-k}.$$

- Again, if $q_n < 1$,

$$p_n = 1 - q_n = \Pr_T(\underline{\text{all}}\ k\text{-sets are dominated})$$
$$= \Pr_T(T \text{ has property } S_k) > 0,$$

and so at least one T with property S_k exists. $\square$

- Denote

$T(k)$ = smallest n s.th. a tournament w/ S_k exists.

One can check that $T(1) = 3$, $T(2) = 7$.

Since $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ and $(1-2^{-k})^{n-k} < e^{-(n-k)/2^k}$, one obtains [Exercise]:

$$T(k) \leq k^2 \cdot 2^k \cdot (\ln 2)(1+o(1)).$$

It is also known that

$$T(k) \geq \text{const} \cdot k \cdot 2^k \qquad \text{[Szekeres]}$$

• Example 3. Dominating sets.

A dominating set of an undirected graph $G = (V, E)$ is a set $U \subseteq V$ s.th. every vertex $v \in V \setminus U$ has at least one neighbour in U.

Theorem 1.3 Let G have n vertices and min. degree δ. Then G has a dominating set of size at most

$$n \frac{\ln(\delta+1)+1}{\delta+1} \qquad \left(\sim n \frac{\ln \delta}{\delta} \text{ for large } \delta\right).$$

Proof. Fix some $p \in [0,1]$ (choice to be optimised later) and pick each of the vertices $v \in V$, randomly and independently w/ prob. p, into a set X. Let $Y = Y_X$ be the set of vertices in $V \setminus X$ with no neighbour in X.

Then $E[|X|] = np$ and

indicator random variable

$$E[|Y|] = E\Big[\sum_{v \in V} I[v \notin X \;\&\; v \text{ has no neighbours in } X]\Big]$$

$$= \sum_{v \in V} E[I[v \notin X \;\&\; v \text{ has no neighbours in } X]]$$

(linearity of expectation!)

$$= n \cdot \Pr(v \notin X \;\&\; v \text{ has no neighbours in } X)$$

$$\le n \cdot (1-p)^{\delta+1}$$

Now $U = X \cup Y_X$ is a (random) dominating set for G of expected size

$$E[|U|] = E[|X| + |Y_X|]$$

$$= E[|X|] + E[|Y|] \quad \text{(linearity of expectation!)}$$

$$(*) \qquad \le n\left(p + (1-p)^{\delta+1}\right).$$

Thus, for any choice of p there is always at least one dom. set of size at most $(*)$.

- For a given δ, what value of p gives the tightest bound in $(*)$?

- Before differentiating, simplify $(*)$ a bit: $1-p \leq e^{-p}$ for all $p \geq 0$, so

$$(*) \leq n\left(p + e^{-p(\delta+1)}\right)$$

- Now the r.h.s. is minimized by

$$\hat{p} = \frac{\ln(\delta+1)}{\delta+1}.$$

- Thus, a dominating set of size at most

$$n\left(\hat{p} + e^{-\hat{p}(\delta+1)}\right)$$

$$= n\left(\frac{\ln(\delta+1)}{\delta+1} + \frac{1}{\delta+1}\right)$$

$$= n\,\frac{\ln(\delta+1)+1}{\delta+1}$$

always exists. $\square$

• Example 4. The Erdös–Ko–Rado Theorem.

Let $\mathcal{F}$ be a family of k-subsets of the set $[n] = \{1, 2, \dots, n\}$.
(i.e. a "k-uniform hypergraph" on vertex set $[n]$.)

$\mathcal{F}$ is intersecting if $A \cap B \neq \emptyset$ for any $A, B \in \mathcal{F}$.

Suppose $n \geq 2k$. How big can an intersecting family $\mathcal{F}$ be?

By taking $\mathcal{F} = \{$all sets containing a given vertex (say 1)$\}$, one obtains $|\mathcal{F}| = \binom{n-1}{k-1}$.

Erdös–Ko–Rado: this is maximum.

• Lemma. For $1 \leq s \leq n$, denote $A_s = \{s, s+1, \dots, \overset{\text{mod } n}{s+k-1}\}$.
Then an intersecting family $\mathcal{F}$ contains at most k of the A_s.

Proof. Fix some $A_s \in \mathcal{F}$. The other A_t intersecting A_s can be grouped into $k-1$ pairs $\{A_{s-i}, A_{s-i+k}\}$. The members of each pair are disjoint, so $\mathcal{F}$ can contain at most one of each. □

Proof (EKR). Choose a permutation σ of $[n]$ and an index $s \in \{1, 2, \dots, n\}$ uniformly & independently at random.
Consider set $A = \{\sigma(s), \sigma(s+1), \dots, \sigma(s+k-1)\}$.

Conditioned on any fixed σ, the Lemma yields (for an intersecting family $\mathcal{F}$): $\Pr_s(A \in \mathcal{F} \mid \sigma) \leq k/n$.
Thus, also for random σ:

$$\Pr_{\sigma, s}(A \in \mathcal{F}) \leq \frac{k}{n}.$$

But a unif. indep. random selection of σ, s yields a unif. random selection over all possible k-sets.
Thus, $\frac{k}{n} \geq \Pr(A \in \mathcal{F}) = |\mathcal{F}| / \binom{n}{k}$, and so

$$|\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}. \quad \square$$