Probabilistic method notes

Shoham Letzter[*]

Contents

[*] Email: s.letzter@ucl.ac.uk

1 The basic method

In this section we apply the probabilistic method in its most basic form. We will see various examples where, in order to show that a construction with certain properties exists, we define a probability space and show that, with positive probability, the outcome has the desired properties. We will also see a few less direct applications of the probabilistic method.

1.1 Colouring hypergraphs

Our first example will be about 2-colourable hypergraphs. These two notions are defined next.

Definition 1.1 (Hypergraph)**.** A *hypergraph* is a pair (V, E) where the elements in V are called *vertices* and E is a set of subsets of V, called *edges*. We say that a hypergraph H is r*-uniform* if all its edges have size r (see Figure 1 for an example of a 3-uniform hypergraph).

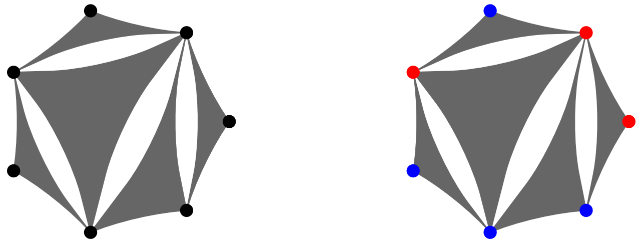


Figure 1: A 3-uniform hypergraph, and a 2-colouring of its vertices with no monochromatic edges

Definition 1.2 (2-colourability)**.** A hypergraph H is said to be 2-colourable (or have *property B*), if its vertices can be coloured by red and blue so that every edge contains both a red vertex and a blue one (equivalently, no edge is *monochromatic*, namely fully red or fully blue).

Example 1.3.

- *The hypergraph in Figure 1 is 2-colourable, as can be seen from the colouring on the right.*
- *The complete 3-uniform hypergraph on five vertices, namely $K_5^{(3)}$, is not 2-colourable. Indeed, given a red-blue colouring of its vertices, without loss of generality at least three are red, but any three vertices form an edge in this graph, so this means there is a monochromatic edge.*

Notice that a 2-uniform hypergraph is a graph, and that a graph G is 2-colourable if and only if it is bipartite, which is the case if and only if G has no odd cycles. There is no such characterisation for 2-colourable r-uniform hypergraphs, with $r \geq 3$. It is thus interesting to find sufficient conditions for 2-colourability. Here is an example of such a condition.

Proposition 1.4 (Erdős, 1963)**.** *Every r-uniform hypergraph with fewer than 2^{r-1} edges is 2-colourable.*

Proof. Let $H = (V, E)$ be an r-uniform hypergraph with fewer than 2^{r-1} edges. Colour each vertex red or blue *randomly*[1] and independently. For an edge e, let A_e be the event that e is monochromatic. Then $\mathbb{P}(A_e) = 2^{1-r}$. Thus, by the union bound,

$$\mathbb{P}\left(\bigcup_{e \in E} A_e\right) \leq \sum_{e \in E} \mathbb{P}(A_e) < 2^{r-1} \cdot 2^{1-r} = 1.$$

It follows that there is a red-blue colouring of V such that no edge is monochromatic, showing that H is 2-colourable. □

Remark 1.5. The result obtained above is not far from best possible. We will see later (see Theorem 1.19) that there are r-uniform hypergraphs on fewer than $c \cdot r^2 \cdot 2^r$ edges which are *not* 2-colourable (where c is a constant). The best known improvement on Proposition 1.4 shows that every r-uniform hypergraph with fewer than $c \cdot \sqrt{r} \cdot 2^r$ edges is 2-colourable.

Remark 1.6. This proof can, in fact, be phrased as a counting argument. Indeed, let $H = (V, E)$ be an r-uniform hypergraph on n vertices with fewer than 2^{r-1} edges. The total number of 2-colourings of V is 2^n, because each vertex has two possible colours. For a given edge e, the number of 2-colourings of V for which e is monochromatic is 2^{n-r+1}. Indeed, there are two ways to colour the vertices of e so that e is monochromatic (they can all be coloured red, or all blue) and 2^{n-r} ways to colour the remaining $n - r$ vertices that are not in e. Thus, the total number of 2-colourings of V for which at least one edge is monochromatic is at most $|E| \cdot 2^{n-r+1} < 2^{r-1} \cdot 2^{n-r+1} = 2^n$. This shows that there is a 2-colouring of V with no monochromatic edges, as desired.

In many cases a probabilistic proof could be converted into a counting proof. However, it is normally much more convenient and insightful to use the language of probability.

1.2 Ramsey numbers

Our next example will be about Ramsey numbers. Recall that K_n is the complete graph on n vertices.

Definition 1.7 (Ramsey numbers)**.** For positive integers s and t, the *Ramsey number* $r(s, t)$ of s and t is the minimum n such that in every red-blue colouring of the edges of K_n, there is either a red K_s or a blue K_t.

Notice that here, unlike in the previous section, we are colouring *edges*, not vertices. The notion of Ramsey numbers is called after Frank Ramsey, who in 1930 proved that $r(s, t)$ is finite for every s and

[1] By picking an element from U *randomly*, we mean that we pick exactly one element from U, with all elements having equal probability (of $1/|U|$) to be chosen

t. This initial study of Ramsey numbers has developed into a prominent branch of combinatorics, called *Ramsey theory.*

We first look at some examples of $r(s,t)$ with small s and t.

Example 1.8.

- *Let $s \geq 1$. Then $r(s,1) = 1$, because a red (or blue) K_1 is a single vertex.*
- *Let $s \geq 2$. Then $r(s,2) = s$. Indeed, the fully red K_{s-1} contains no blue K_2 (i.e. no blue edge) and no red K_s, showing $r(s,2) \geq s$. If we red-blue colour K_s, then either there is a blue edge, which is a blue K_2, or all edges are red, yielding a red K_s. Thus $r(s,2) \leq s$.*
- *It is not hard to see that $r(3,3) = 6$; see Figure 2 for a proof of $r(3,3) > 5$. It is also not very hard to see that $r(3,4) = 9$ and $r(4,4) = 18$.*

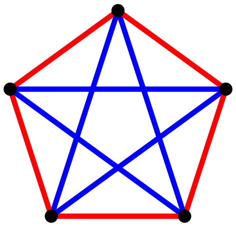


Figure 2: Example showing $r(3,3) > 5$

- *It is known that $43 \leq r(5,5) \leq 48$, with the upper bound established only in 2017.*

In the next proposition we give an upper bound on Ramsey numbers.

Proposition 1.9. *$r(s,t) \leq 2^{s+t}$ for $s,t \geq 1$.*

Proof. We prove the statement by induction on $s+t$. Notice that the statement trivially holds when $s = 1$ or $t = 1$. Thus, we may assume that $s,t \geq 2$. We need to show that every red-blue K_n, with $n = 2^{s+t}$, has either a red K_s or blue K_t. Fix a red-blue colouring of K_n. Consider a vertex v, and notice that v has $n-1$ edges incident to it. Without loss of generality, at least $\lceil \frac{n-1}{2} \rceil$ of them are red. Denote by U the set of vertices u such that uv is red. So $|U| \geq \lceil \frac{n-1}{2} \rceil \geq \left\lceil \frac{2^{s+t}-1}{2} \right\rceil \geq 2^{s+t-1}$. By induction, U contains either a red K_{s-1}, which together with v forms a red K_s, or a blue K_t, as required. □

The case where $s = t$ (known as *diagonal Ramsey numbers*) has received particular attention. Here we give upper and lower bounds for this case.

Proposition 1.10. *$(t-1)^2 < r(t,t) \leq 2^{2t}$ for $t \geq 1$.*

Proof. Proposition 1.9 implies the upper bound. For the lower bound, partition the vertices of a $K_{(t-1)^2}$ into $t-1$ sets of size $t-1$, colour the edges within each set red, and colour the edges between sets blue. □

This brings us to the second example of the probabilistic method presented in this module, vastly improving the lower bound from Proposition 1.10.

Theorem 1.11 (Erdős, 1947)**.** *$r(t,t) \geq 2^{t/2}$ for $t \geq 5$.*

Proof. Write $n = \lfloor 2^{t/2} \rfloor$. We need to show that there is a red-blue colouring of K_n that has no monochromatic K_t (monochromatic means that all its edges have the same colour). The colouring will be random. We colour each edge red with probability $1/2$, and blue otherwise, independently of other edges.

For a set of t vertices S, let A_S be the event that S forms a monochromatic clique, i.e. all its edges have the same colour. Notice that $\mathbb{P}(A_S) = 2 \cdot 2^{-\binom{t}{2}}$. Indeed, there are two choices for the colour of the edges in S, and for each choice we need all $\binom{t}{2}$ edges of S to pick that colour. Now, by the union bound,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_S A_S\right) \leq \sum_S \mathbb{P}(A_S) &= \binom{n}{t} \cdot 2 \cdot 2^{-\binom{t}{2}} \\
&\leq 2 \cdot \left(\frac{en}{t}\right)^t 2^{-\frac{t(t-1)}{2}} \\
&\leq 2 \cdot \left(\frac{e \cdot 2^{t/2} \cdot 2^{-(t-1)/2}}{t}\right)^t \\
&= 2 \cdot \left(\frac{e\sqrt{2}}{t}\right)^t < 1.
\end{aligned}$$

Here we used also the very useful bound $\binom{n}{t} \leq \left(\frac{en}{t}\right)^t$, and the assumption $t \geq 5$.

Since the probability of the event $\bigcup_S A_S$ is less than 1, this shows that with positive probability none of the events A_S hold, and so there is a red-blue colouring of K_n without a monochromatic K_t, as claimed. □

Remark 1.12. This is one of the earliest examples of the use of probabilistic methods in combinatorics. Erdős was the first to realise the great potential of the method, and has applied it to numerous combinatorial problems.

Remark 1.13. Both bounds $2^{t/2} \leq r(t,t) \leq 2^{2t}$ were improved only slightly since they became known in the 1940s. It is a major open problem to determine which of these two bounds in closer to the truth.

1.3 Tournaments

Our next example is about certain directed graphs called tournaments. We first define directed graphs and then tournaments.

Definition 1.14 (Directed graph)**.** A *directed graph* (or *digraph*) is a pair (V, E) where E is a set of *ordered pairs* (u, v) where u, v are distinct elements in V. As usual, we refer to elements in V as vertices and to elements in E as edges. We think of the edge (u, v) (often denoted simply as uv) as an edge directed from u to v, and if the edge uv exists we say that v is an *out-neighbour* of u and that u is an *in-neighbour* of v. (See Figure 3.)

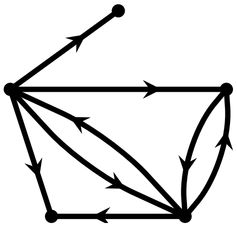


Figure 3: A directed graph

In particular, for any two vertices u, v in a digraph $D = (V, E)$, both edges uv and vu could be present, or none, or exactly one of the two.

Definition 1.15 (Tournament)**.** A *tournament* is a directed graph T where for any two distinct vertices u, v in T, exactly one of the pairs uv and vu is an edge (see Figure 4).

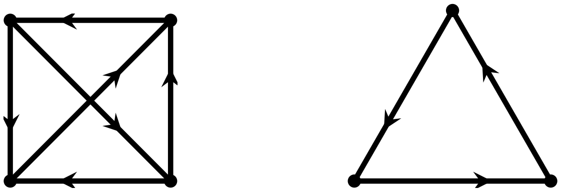


Figure 4: A tournament on four vertices and a directed triangle

One can think of a tournament as representing the results of a round robin tournament, with the vertices representing the players, and an edge uv signifying that u beat v (hence the name).

Definition 1.16 (Property S_k)**.** We say that a tournament has *property* S_k if for every k players, there is a player who beat all of them; in other words, every k vertices have a common in-neighbour.

Example 1.17. *The directed triangle (see Figure 4) is a tournament with property* S_1.

It seems hard to construct tournaments with property S_k, with $k \geq 2$, explicitly. A random construction, however, yields the property very easily.

Proposition 1.18 (Erdős, 1963)**.** *For every k there is a tournament (on at least k vertices) with property S_k.*

Proof. Let V be a set of n vertices, where n will be determined later. Form a random tournament by picking the direction of the edge between any two vertices u, v randomly and independently of other edges. For a set T of k vertices, denote by A_T the event that there is no vertex that beats all players in T. Then

$$\mathbb{P}(A_T) = \prod_{u \in V-T} \mathbb{P}\left(\{u \text{ does not beats all of } T\}\right) = \left(1 - 2^{-k}\right)^{n-k}.$$

because the events $\{u$ does not beats all of $T\}$, for $u \in V - T$, are independent (as they depend on pairwise disjoint sets of edges); moreover, the only scenario in which u does beat all of T is when all the edges between u and T are directed from u to T. By the union bound,

$$\mathbb{P}\left(\bigcup_T A_T\right) \leq \sum_T \mathbb{P}(A_T) = \binom{n}{k}\left(1 - 2^{-k}\right)^{n-k} < 1,$$

where n is picked to satisfy the last inequality (notice that $\lim_{n\to\infty} \binom{n}{k}\left(1 - 2^{-k}\right)^{n-k} = 0$, because exponential functions grow faster than polynomial, and so a suitably n can be found). This shows that there is a tournament in which none of the events A_T hold, i.e. the events $(A_T)^{\mathsf{C}} = \{$there is a vertex that beats all of $T\}$ all hold, i.e. property S_k is satisfied. □

1.4 Colouring hypergraphs – continued

Let $m(r)$ be the minimum possible number of edges in a r-uniform hypergraph that is not 2-colourable. We saw in Proposition 1.4 that $m(r) \geq 2^{r-1}$. The best known bound is only a bit better (by a factor of less than $\sqrt{r}$). The following theorem shows that, indeed, one cannot do much better than $m(r) \geq 2^{r-1}$.

Theorem 1.19 (Erdős, 1964)**.** *$m(r) \leq 8r^2 2^r$.*

Proof. Let V be a set of size $n = 8r^2$ and $m = n2^r = 8r^2 2^r$.

The idea is to 'turn the probability space on its head', meaning that we will pick the edges at random, rather than the colours. Let $e_1, \ldots, e_m$ be chosen randomly and independently among all subsets of V of size r (so we could have two or more of the edges be the same).

For a red-blue colouring χ of V let $A_{\chi,i}$ be the event that the edge e_i is monochromatic with respect to χ. Denote by b the number of vertices in V coloured blue by χ. Then

$$\begin{aligned}\mathbb{P}(A_{\chi,i}) &= \frac{\binom{b}{r} + \binom{n-b}{r}}{\binom{n}{r}} \geq \frac{2\binom{n/2}{r}}{\binom{n}{r}} \geq 2 \cdot \frac{\frac{(n/2-r)^r}{r!}}{\frac{n^r}{r!}} \\ &= 2 \cdot \left(\frac{n/2 - r}{n}\right)^r = 2^{-r} \cdot 2 \cdot \left(1 - \frac{2r}{n}\right)^r \geq 2^{-r} \cdot 2e^{-4r^2/n} = 2^{-r} \cdot 2e^{-1/2} \geq 2^{-r}.\end{aligned}$$

For the first inequality, we used the convexity of the function $f(x) := \binom{x}{r}$, which implies that $\frac{1}{2}(f(x)+f(y)) \geq f(\frac{x+y}{2})$ for all x, y; to get the inequality, take $x = b$ and $y = n - b$. For the second inequality we used the bounds $\frac{(n-r)^r}{r!} \leq \binom{n}{r} \leq \frac{n^r}{r!}$. The third inequality follows from $1 - x \geq e^{-2x}$ which holds for $x \in [0, 1/4]$.

Now, let B_χ be the event that none of the edges $e_1, \ldots, e_m$ are monochromatic. Then

$$\begin{aligned}\mathbb{P}(B_\chi) = \mathbb{P}\left(\bigcap_{1\leq i\leq m} A^{\mathsf{c}}_{\chi,i}\right) &= \prod_{1\leq i\leq m}\big(1 - \mathbb{P}(A_{\chi,i})\big) \\ &\leq (1-2^{-r})^m \leq \exp\big(-2^{-r}m\big) = e^{-n} < 2^{-n}.\end{aligned}$$

Here the second equality follows from the independence of the events $A_{\chi,i}$, and the first inequality follows from $1 - x \leq e^{-x}$.

Thus, by a union bound,

$$\mathbb{P}\left(\bigcup_\chi B_\chi\right) \leq \sum_\chi \mathbb{P}(B_\chi) < 1,$$

using that there are exactly 2^n red-blue colourings of V. This means that, with positive probability, there is a way to choose $e_1, \ldots, e_m$ such that for *every* red-blue colouring of V there is a monochromatic e_i. Take H to be the hypergraph with vertex set V and edges $\{e_1, \ldots, e_m\}$. So H is not 2-colourable, and has at most m edges (we get an upper bound, and not an exact number, for the number of edges because some e_i's could be the same). This shows $m(r) \leq m = 8r^2 2^r$. □

1.5 Set systems

We now present a less direct application of the probabilistic method. We denote by $[m]$ the set $\{1, \ldots, m\}$.

Theorem 1.20 (Bollobás, 1965). *Let $(A_i, B_i)_{i\in[m]}$ be a sequence of pairs of sets, such that*

- *$|A_i| = a$ and $|B_i| = b$ for $i \in [m]$,*
- *$A_i \cap B_i = \emptyset$ for $i \in [m]$,*
- *$A_i \cap B_j \neq \emptyset$ for all distinct $i, j \in [m]$*

Then $m \leq \binom{a+b}{a}$.

Proof. Let $S = \bigcup_{i\in[m]}(A_i \cup B_i)$. Let σ be a random ordering of the elements of S (namely, we pick one of $|S|!$ orderings randomly). Let E_i be the event that the elements of A_i precede those of B_i in the ordering σ. Notice that $\mathbb{P}(E_i) = \frac{1}{\binom{a+b}{a}}$. Indeed, we can think of the ordering σ as defined as follows: first, we determine the location of the elements in $S - (A_i \cup B_i)$, and then we determine the locations of the elements in $A_i \cup B_i$. Given any outcome of the first step, there are $(a + b)!$ ways to order $A_i \cup B_i$ in the remaining spots, and $a! \cdot b!$ ways to do it so that the elements of A_i precede those of B_i.

Next, we claim that the events E_i, with $i \in [m]$, are pairwise disjoint, i.e. no two of them can occur at the same time. Indeed, suppose that E_i occurs, meaning that the elements of A_i precede the elements of B_i in σ, and consider some $j \in [m] - \{i\}$. Let $a \in A_i \cap B_j$ and $b \in B_i \cap A_j$ (such elements exist by the third assumption of the theorem). Then a precedes b in σ because the elements of A_i precede those of B_i. But this shows that there is an element in B_j that precedes an element in A_j, showing that E_j does not occur.

It follows that

$$1 \geq \mathbb{P}\left(\bigcup_{i\in[m]} E_i\right) = \sum_{i\in[m]} \mathbb{P}(E_i) = \frac{m}{\binom{a+b}{a}}.$$

We used the disjointness of the events E_i in the first equality, and the value of $\mathbb{P}(E_i)$ in the second equality. This immediately gives $m \leq \binom{a+b}{a}$. □

Remark 1.21. In previous applications of the probabilistic method, we showed that a certain desirable outcome holds by showing that it occurs with positive probability. In this proof the probabilistic method is used indirectly, by leveraging the fact that the probability of the union of pairwise disjoint events is the sum of probabilities of the events.

Remark 1.22. The bound we obtained is tight: let S be a set of size $a + b$, let the A_i's be all subsets of S of size a, and let $B_i = S - A_i$.

Remark 1.23. There are various variants and extensions of the above theorem. For example, one can weaken the third condition to require that $A_i \cap B_j$ whenever $1 \leq i < j \leq m$, without changing the conclusion. The proof of this version is algebraic, and no probabilistic or combinatorial proof is known.

Next, we present another indirect application of the probabilistic method, providing an alternative proof to the following result of Sperner.

Theorem 1.24 (Sperner, 1928)**.** *Let $\mathcal{F}$ be a family of subsets of $[n]$, such that no two distinct sets A, B in $\mathcal{F}$ satisfy $A \subseteq B$ or $B \subseteq A$. Then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.*

Proof. Let $\sigma = (\sigma(1), \ldots, \sigma(n))$ be a random permutation of $[n]$. For a set $A \in \mathcal{F}$, let E_A be the event that A is a prefix of σ, namely that $\{\sigma(1), \ldots, \sigma(|A|)\} = A$. Then

$$\mathbb{P}(E_A) = \frac{|A|!(n - |A|)!}{n!} = \frac{1}{\binom{n}{|A|}} \geq \frac{1}{\binom{n}{\lfloor n/2 \rfloor}}. \tag{1}$$

Indeed, for the first equality note that the number of permutations in which A is a prefix of σ is $|A|!(n - |A|)!$, and the total number of permutations in $n!$. For the inequality, we use the following claim.

Claim 1.25. $\binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor}$ *for* $k \in \{0, 1, \ldots, n\}$.

Proof. First note the following for $k \leq (n-1)/2$.

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

where the inequality follows directly from $k \leq (n-1)/2$. This shows $\binom{n}{0} \leq \binom{n}{1} \leq \ldots \leq \binom{n}{\lfloor n/2 \rfloor}$, showing $\binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor}$ for $k \leq \lfloor n/2 \rfloor$. Using $\binom{n}{k} = \binom{n}{n-k}$, we get also $\binom{n}{k} \leq \binom{n}{\lceil n/2 \rceil} = \binom{n}{\lfloor n/2 \rfloor}$ for $\lceil n/2 \rceil \leq k \leq n$. Altogether, $\binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor}$ for all $k \in [n]$, as claimed. □

We now note that the events E_A, with $A \in \mathcal{F}$, are pairwise disjoint. Indeed, if $A, B \in \mathcal{F}$ are distinct, then E_A and E_B cannot simultaneously hold, as that would imply $A = \{\sigma(1), \ldots, \sigma(|A|)\}$ and $B = \{\sigma(1), \ldots, \sigma(|B|)\}$, showing that $A \subseteq B$ or $B \subseteq A$, a contradiction.

Using (1) and the disjointness of the events E_A, we get

$$1 \geq \mathbb{P}\left(\bigcup_{A \in \mathcal{F}} E_A\right) = \sum_{A \in \mathcal{F}} \mathbb{P}(E_A) \geq \frac{|\mathcal{F}|}{\binom{n}{\lfloor n/2 \rfloor}}.$$

This gives $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$, as required. □

Remark 1.26. A family $\mathcal{F}$ of sets containing no distinct elements A, B with $A \subseteq B$ is called an *antichain*. So the above theorem can be phrased as: every antichain of subsets of $[n]$ has size at most $\binom{n}{\lfloor n/2 \rfloor}$. This bound is tight: take $\mathcal{F}$ to be the family of all subsets of $[n]$ of size $\lfloor n/2 \rfloor$.

2 Linearity of expectation

In this section we will see several application of the probabilistic method, where the main tool will be the linearity of expectation. Recall that this means that for random variables $X_1, \ldots, X_n$ and reals $\alpha_1, \ldots, \alpha_n$, we have $\mathbb{E}\left(\sum_{i \in [n]} \alpha_i X_i\right) = \sum_{i \in [n]} \alpha_i \mathbb{E}(X_i)$.

2.1 Hamilton cycles in tournaments

This first example is about the number of Hamilton cycles in a tournament.

Definition 2.1 (Hamilton paths and cycles)**.** A *Hamilton path* in a (directed) graph G is a (directed) path through all the vertices in G (see Figure 5). Similarly, a *Hamilton cycle* in a (directed) graph G is a (directed) cycle through all the vertices in G.

It is a well known fact that every tournament has a Hamilton path. The following proposition gives a lower bound on the maximum possible number of Hamilton paths a tournament can have. It is considered to be the first application of the probabilistic method.

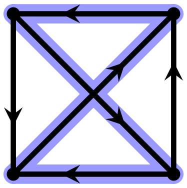


Figure 5: A Hamilton cycle in a tournament

Proposition 2.2 (Szele, 1943)**.** *There is a tournament on n vertices with at least $(n-1)! \cdot 2^{-n}$ Hamilton cycles.*

Proof. Let T be a random tournament on vertex set $[n]$, meaning that for any two vertices u, v the edge between them is directed randomly and independently. For a permutation σ of $[n-1]$, let X_σ be the indicator random variable of the event

$$\{(\sigma(1) \ldots \sigma(n-1)\, n) \text{ is a directed cycle}\}.$$

Write X for the random variable counting the number of Hamilton cycles in G. Then $X = \sum_\sigma X_\sigma$. Indeed, we can insist that the last vertex of the cycle is n, and then the each cycle corresponds to an ordering of $[n-1]$. Notice that $\mathbb{E}(X_\sigma) = \mathbb{P}(X_\sigma) = 2^{-n}$, because we need the n edges in $(\sigma(1) \ldots \sigma(n-1)\, n)$ to be directed the right way. Thus, by linearity of expectation,

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_\sigma X_\sigma\right) = \sum_\sigma \mathbb{E}(X_\sigma) = (n-1)! \cdot 2^{-n}.$$

It follows that there exists a tournament with at least $(n-1)! \cdot 2^{-n}$ Hamilton cycles. □

Remark 2.3. The bound given in the above proposition is almost tight; indeed, we will see that every tournament on n vertices has at most $O(\sqrt{n} \cdot n! \cdot 2^{-n})$ Hamilton cycles.

2.2 Max cut

Here is a quick application of the linearity of expectation, which you may have already encountered.

Proposition 2.4. *Let G be a graph with m edges. Then G contains a bipartite graph with at least $m/2$ edges.*

Proof. Let A be a random set of vertices, obtained by including each vertex of G with probability $1/2$, independently. Let $H = G[A, V(G) - A]$ (so H is the bipartite subgraph of G consisting of all edges in G with exactly one end in A).

Let X be the number of edges in H. Then $X = \sum_{e \in E(G)} X_e$, where X_e is the indicator function of the event $\{e \text{ is in } H\}$. Now, for an edge $e = uv$, we have $\mathbb{E}(X_e) = \mathbb{P}(\{e \text{ is in } H\}) = 1/2$, because for e to be in H we need either u to be in A and v to not be in A, or vice versa, and each of these outcomes occurs with probability $1/4$.

Thus,

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{e \in E(G)} X_e\right) = \sum_e \mathbb{E}(X_e) = m/2.$$

Here we used the linearity of expectation for for the second equality. In particular, $X \geq m/2$ with positive probability, showing that there is a choice of A for which H has at least $m/2$ edges. □

Remark 2.5. This proposition has an easy deterministic proof.

2.3 Number theory

The next example in this chapter is a more sophisticated example, about sum free sets of integers.

Definition 2.6 (Sum free sets)**.** A subset $A \subseteq \mathbb{Z} - \{0\}$ is *sum free* if there are no (not necessarily distinct) elements $a, b, c \in A$ such that $a + b = c$.

Example 2.7. *The set* $\{1, 3, 7\}$ *is sum free, the set* $\{2, 4\}$ *is not.*

Definition 2.8 (Sum free sets modulo p)**.** Recall that $\mathbb{Z}_p$ is the set $\{0, \ldots, p-1\}$ with addition and multiplication modulo p. As above, a subset $A \subseteq \mathbb{Z}_p$ is *sum free* if there are no elements $a, b, c \in A$ such that $a + b \equiv c \pmod{p}$.

The next result shows that every set of positive integers has a large sum free subset.

Theorem 2.9 (Erdős, 1965)**.** *Let A be a finite set of positive integers. Then there is a sum free subset $B \subseteq A$ of size larger than $|A|/3$.*

Proof. Let p be a prime which satisfies $p > 2\max_{a \in A} |a|$ and $p = 3k + 2$, where k is an integer.[2] Set I to be the interval $\{k+1, \ldots, 2k+1\}$, and let w be an integer chosen uniformly at random from $[p-1]$. Let B be the random subset of A, consisting of elements $a \in A$ such that $wa \pmod{p} \in I$.

Note that I is sum free when considered as a subset of $\mathbb{Z}_p$. Indeed, if $a, b \in I$ then, in $\mathbb{Z}$, they satisfy $2k + 2 \leq a + b \leq 4k + 2$. Thus, in $\mathbb{Z}_p$, either $a + b \in \{2k+2, \ldots, 3k+1\}$ or $a + b \in \{0, \ldots, k\}$. Either way, $a + b \notin I$.

We conclude that B is always sum free. Indeed, otherwise there are $a, b, c \in A$ and $w \in [p-1]$ such that $wa, wb, wc \pmod{p} \in I$ and thus $a + b = c$. But this implies $a + b \equiv c \pmod{p}$, and $wa + wb \equiv wc \pmod{p}$, a contradiction to I being sum free in $\mathbb{Z}_p$.

[2] It is not hard to see that such a number exists. Indeed, if not then the number of primes which are $-1 \pmod 3$ is finite; denote them by $p_1, \ldots, p_t$. Write $w = (p_1 \cdot \ldots \cdot p_t)^2 + 1$. Then w is not divisible by any of $p_1, \ldots, p_t$. Also, $w \equiv -1 \pmod 3$, so w has a prime divisor q which satisfies $q \equiv -1 \pmod 3$ but is not in $\{p_1, \ldots, p_t\}$, a contradiction.

We now estimate the size of B. Let X_a be the indicator random variable of the event $\{a \in B\}$, for $a \in A$. Then

$$\begin{aligned}\mathbb{E}(X_a) = \mathbb{P}(a \in B) = \mathbb{P}\left(\bigcup_{b\in I}\{wa \equiv b \,(\mathrm{mod}\; p)\}\right) \\ = \sum_{b\in I}\mathbb{P}(wa \equiv b\,(\mathrm{mod}\; p)) \\ = \sum_{b\in I}\mathbb{P}\big(w = ba^{-1}\,(\mathrm{mod}\; p)\big) = \frac{|I|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}.\end{aligned}$$

For the third equality we used the disjointness of the events $\{wa \equiv b\,(\mathrm{mod}\; p)\}$, with $b \in I$. For the fourth equality we used that $a \not\equiv 0\,(\mathrm{mod}\; p)$, which follows from $p > 2\max_{a\in A}|a|$.

Finally, by linearity of expectation, we have

$$\mathbb{E}\big(|B|\big) = \mathbb{E}\left(\sum_{a\in A}X_a\right) = \sum_{a\in A}\mathbb{E}(X_a) > \frac{|A|}{3}.$$

So, with positive probability, $|B| > |A|/3$. Since B is always sum free, this shows that A has a sum free subset of size larger than $|A|/3$. □

Remark 2.10. The fraction 1/3 in the above theorem was shown to be tight by Eberhard, Green and Manners (2013).

2.4 Permanents

We now spend quite some time on an inequality on the *permanent* (defined below) of a $\{0,1\}$-matrix; see Theorem 2.14 below. The proof of the inequality is a clever and complicated application of linearity of expectation. We will later see an application to Hamilton cycles in tournaments.

Recall that for an $n \times n$ matrix A, its *determinant*, denoted $\det(A)$, is defined as

$$\det(A) = \sum_{\sigma\in S_n}\prod_{i\in[n]}(-1)^{\mathrm{sign}(\sigma)}A_{i,\sigma(i)},$$

where S_n is the set of all permutations of $[n]$, and $\mathrm{sign}(\sigma)$ is 0 if σ can be written as the product of an even number of transpositions (namely, swaps of two elements), and 1 otherwise. The permanent is defined similarly.

Definition 2.11 (Permanent)**.** The *permanent* of an $n \times n$ matrix A, denoted $\mathrm{per}(A)$, is defined as

$$\mathrm{per}(A) = \sum_{\sigma\in S_n}\prod_{i\in[n]}A_{i,\sigma(i)}. \tag{2}$$

We will focus on the permanent of $\{0,1\}$-matrices (i.e. matrices whose elements are 0's and 1's).

Example 2.12. *Let* $A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$. *Then* $\operatorname{per}(A) = 2$. *Indeed, the permutations* (312) *(illustrated in the left of* (3)*) and* (321) *(illustrated on the right of* (3)*) contribute* 1 *to the sum in* (2)*, the others contribute* 0.

$$\begin{pmatrix} 0 & 0 & \boxed{1} \\ \boxed{1} & 1 & 0 \\ 1 & \boxed{1} & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & \boxed{1} \\ 1 & \boxed{1} & 0 \\ \boxed{1} & 1 & 1 \end{pmatrix}. \tag{3}$$

Remark 2.13. Suppose that A is an $n \times n$, $\{0,1\}$-matrix. Then $\operatorname{per}(A)$ is the number of permutations $\sigma \in S_n$ such that $A_{i,\sigma(i)} = 1$ for every $i \in [n]$. Indeed, each permutation contributes 1 to the sum in (2) if all the elements $A_{i,\sigma(i)}$ are 1, and otherwise it contributes 0. In other words, $\operatorname{per}(A)$ is the number of ways to select exactly one 1 from each row and column.

We will spend some time proving the following inequality.

Theorem 2.14 (Brégman, 1973)**.** *Let* A *be an* $n \times n$ $\{0,1\}$*-matrix, and denote by* r_i *the number of* 1*'s in the* i^{th} *row. Then*

$$\operatorname{per}(A) \leq \prod_{i \in [n]} (r_i!)^{1/r_i}.$$

Remark 2.15. This inequality is tight in some cases. For example, this is true when A is the all-1 matrix of any dimension. Similarly, we can take A to consist of blocks of all-1 matrices, with 0's outside of the blocks; here is an example (the 1's are in squares for emphasis).

$$\begin{pmatrix} \boxed{1} & \boxed{1} & 0 & 0 & 0 & 0 & 0 \\ \boxed{1} & \boxed{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & \boxed{1} & \boxed{1} & \boxed{1} \\ 0 & 0 & 0 & \boxed{1} & \boxed{1} & \boxed{1} & \boxed{1} \\ 0 & 0 & 0 & \boxed{1} & \boxed{1} & \boxed{1} & \boxed{1} \\ 0 & 0 & 0 & \boxed{1} & \boxed{1} & \boxed{1} & \boxed{1} \end{pmatrix}.$$

2.5 Hamilton cycles – continued

We now use Theorem 2.14 about permanents of $\{0,1\}$-matrices to prove the following upper bound on the number of Hamilton cycles in a tournament.

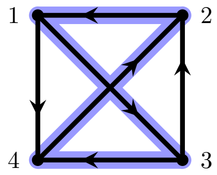
Theorem 2.16 (Alon, 1990)**.** *There is a constant* $c > 0$ *such that every tournament on* n *vertices has at most* $c\sqrt{n} \cdot \frac{n!}{2^n}$ *Hamilton cycles.*

Remark 2.17. This is close to best possible. Indeed, recall that there are tournaments with at least $(n-1)!2^{-n}$ tournaments; this is just a factor of $O(n^{3/2})$ off from the bound in Theorem 2.16.

Proof. Fix a tournament T on n vertices; for convenience, we assume that the vertex set of T is $[n]$. Define an $n \times n$ matrix A as follows (see Figure 6),

$$A_{i,j} = \begin{cases} 1 & \text{if } ij \text{ is a directed edge} \\ 0 & \text{otherwise (i.e. if } i = j \text{ or } ji \text{ is an edge).} \end{cases}$$

Given a Hamilton cycle $(v_1 \dots v_n)$, let σ be the permutation satisfying $\sigma(v_i) = v_{i+1}$ for $i \in [n]$ (in particular, $\sigma(v_n) = v_1$). Notice that such a permutation indeed exists, it defines the cycle uniquely, and it contributes 1 to the permanent of A (see Figure 6). Thus, the number of Hamilton cycles in T is at most $\mathrm{per}(A)$. Let r_i be the number of 1's in row i (this is the number of out-neighbours of vertex i). Then $\sum_{i\in[n]} r_i = \binom{n}{2}$. By Brégman's theorem (Theorem 2.14),



$$\begin{pmatrix} 0 & 0 & \boxed{1} & 1 \\ \boxed{1} & 0 & 0 & 0 \\ 0 & 1 & 0 & \boxed{1} \\ 0 & \boxed{1} & 0 & 0 \end{pmatrix}$$

Figure 6: An illustration of the correspondence between a tournament and a matrix and between a Hamilton cycle and a permutation

$$\mathrm{per}(A) \le \prod_{i\in[n]} (r_i!)^{1/r_i}. \tag{4}$$

We use the following claim to estimate the above expression.

Claim 2.18. *Let a, b be integers satisfying $1 \le a \le b-2$. Then*

$$(a!)^{\frac{1}{a}} \cdot (b!)^{\frac{1}{b}} < ((a+1)!)^{\frac{1}{a+1}} \cdot ((b-1)!)^{\frac{1}{b-1}}.$$

Proof. Define

$$f(a) = \frac{(a!)^{\frac{1}{a}}}{((a+1)!)^{\frac{1}{a+1}}}.$$

We will show that f is an increasing function on $\mathbb{N}$. This would imply $f(a) \le f(b-1)$, which would prove the claim. To see this, consider the following inequality with $a \geq 2$.

$$\begin{aligned}\left(\frac{f(a-1)}{f(a)}\right)^{a(a+1)(a-1)} &= \frac{((a-1)!)^{a(a+1)}((a+1)!)^{a(a-1)}}{(a!)^{2(a-1)(a+1)}}\\ &= ((a-1)!)^{a(a+1)+a(a-1)-2(a-1)(a+1)} \cdot a^{a(a-1)-2(a-1)(a+1)} \cdot (a+1)^{a(a-1)}\\ &= ((a-1)!)^2 \cdot a^{-a^2-a+2} \cdot (a+1)^{a^2-a}\\ &= (a!)^2 \left(\frac{a+1}{a}\right)^{a^2-a} a^{-2a}\\ &\leq 7a^{1/2} \left(\frac{a}{e}\right)^{2a} e^{a-1} a^{-2a}\\ &= \frac{7a^{1/2}}{e^{a+1}} \leq 1.\end{aligned}$$

For the penultimate inequality we used $n! \leq 7\sqrt{n}\left(\frac{n}{e}\right)^n$ (which holds for all $n \geq 1$) and $1 + x \leq e^x$ which holds for all x. The last inequality can be seen to hold for $a \geq 2$. This shows that $f(a-1) \leq f(a)$ for $a \geq 2$, as required. □

The claim implies that $\prod_{i\in[n]}(r_i!)^{1/r_i}$, with $\sum_{i\in[n]} r_i = \binom{n}{2}$, is maximised when every two r_i's differ by at most 1 (if say $r_i \leq r_j - 2$ then we can increase the value by increasing r_i by 1 and decreasing r_j by 1). Assuming n is odd (n even is similar but a bit more technical), the expression is maximised when $r_i = (n-1)/2$ for all i. Write $n = 2m + 1$ and suppose that m is large. Then

$$\begin{aligned}\mathrm{per}(A) \leq (m!)^{\frac{2m+1}{m}} &\leq \left(7\sqrt{m}\left(\frac{m}{e}\right)^m\right)^{2+\frac{1}{m}}\\ &\leq 49 \cdot m \cdot 2^{-2m} \cdot \left(\frac{2m}{e}\right)^{2m} \cdot (7\sqrt{m})^{\frac{1}{m}} \cdot \frac{m}{e}\\ &\leq 50 \cdot m^{3/2} \cdot 2^{-2m}(2m)!\\ &\leq 50 \cdot n^{3/2} \cdot 2^{-(n-1)}(n-1)! = 100\sqrt{n} \cdot \frac{n!}{2^n}.\end{aligned}$$

Here we used (for the second and fourth inequalities) the inequalities $\sqrt{n}\left(\frac{n}{e}\right)^n \leq n! \leq 7\sqrt{n}\left(\frac{n}{e}\right)^n$, which hold for all n. We also used that $\lim_{n\to\infty}(7\sqrt{n})^{1/n} = 0$, so for large n we have $(7\sqrt{n})^{1/n} \leq 1$, say. It follows that $\mathrm{per}(A) \leq 100\sqrt{n} \cdot \frac{n!}{2^n}$, showing that T has at most $100\sqrt{n} \cdot \frac{n!}{2^n}$ Hamilton cycles, as claimed. □

2.6 Permanents – continued

Before proving Brégman's theorem, we make a definition.

Definition 2.19 (Geometric mean)**.** The *geometric mean* of a random variable X, denoted $\mathbb{G}(X)$, is defined as

$$\mathbb{G}(X) = e^{\mathbb{E}(\log X)}.$$

We will need two useful facts about the geometric mean.

Claim 2.20 ('Linearity of expectation')**.** *Let $X_1, \ldots, X_n$ be random variables. Then*

$$\mathbb{G}\left(\prod_{i\in[n]} X_i\right) = \prod_{i\in[n]} \mathbb{G}(X_i). \tag{5}$$

Proof.

$$\begin{aligned}\mathbb{G}\left(\prod_{i\in[n]} X_i\right) &= \exp\left(\mathbb{E}\left(\log\Big(\prod_{i\in[n]} X_i\Big)\right)\right) = \exp\left(\mathbb{E}\left(\sum_{i\in[n]} \log X_i\right)\right)\\ &= \exp\left(\sum_{i\in[n]} \mathbb{E}(\log X_i)\right) = \prod_{i\in[n]} e^{\mathbb{E}(\log X_i)}\\ &= \prod_{i\in[n]} \mathbb{G}(X_i),\end{aligned}$$

using (the usual notion of) linearity of expectation for the third equality. □

Recall that if X is a random variable and A is an event then the *conditional expectation* $\mathbb{E}(X \mid A)$ is defined as

$$\mathbb{E}(X \mid A) = \sum_x x \cdot \mathbb{P}(X = x \mid A).$$

Moreover, the law of total expectation asserts that, for random variables X and Y,

$$\mathbb{E}(X) = \sum_y \mathbb{E}(X \mid Y = y) \cdot \mathbb{P}(Y = y).$$

(This is easy to verify.) Next, we prove a version of this for the geometric mean.

Claim 2.21 ('Law of total expectation')**.** *Let X and Y be random variables. Then*

$$\prod_y \mathbb{G}(X \mid Y = y)^{\mathbb{P}(Y=y)} = \mathbb{G}(X). \tag{6}$$

Proof. We get

$$\begin{aligned}\prod_y \mathbb{G}(X \mid Y = y)^{\mathbb{P}(Y=y)} &= \prod_y \exp\Big(\mathbb{E}\big(\log X \mid Y = y\big) \cdot \mathbb{P}(Y = y)\Big)\\ &= \exp\left(\sum_y \mathbb{E}\big(\log X \mid Y = y\big) \cdot \mathbb{P}(Y = y)\right)\\ &= \exp(\mathbb{E}(\log X)) = \mathbb{G}(X),\end{aligned}$$

using linearity of expectation for the second equality and the law of total expectation for the third. □

We now prove Brégman's theorem.

Proof of Theorem 2.14. Let S be the family of permutations that contribute 1 in the sum defining the permanent, namely the permutations corresponding to a choice of exactly one 1 from each row and column. Recall that S_n is the collection of all permutations of $[n]$. Let σ be chosen uniformly from S, and let τ be chosen uniformly from S_n, and independently from σ.

We define matrices $A_1, \ldots, A_n$ and numbers $R_1, \ldots, R_n$ as follows. Define $A_1 = A$. Let $R_{\tau(1)}$ be the number of 1's in row $\tau(1)$ of A_1, and let A_2 be the matrix obtained by removing row $\tau(1)$ and column $\sigma(\tau(1))$ from A_1. We continue similarly: $R_{\tau(2)}$ is the number of 1's in row $\tau(2)$ in A_2 (we keep the same numbering of rows as in A, so that if row 2 was removed in the first step, then now we have rows $1, 3, 4, \ldots, n$), and A_3 is obtained from A_2 by removing row $\tau(2)$ and column $\sigma(\tau(2))$. In general, A_i is the matrix obtained by removing rows $\tau(1), \ldots, \tau(i-1)$ and columns $\sigma(\tau(1)), \ldots, \sigma(\tau(i-1))$ from A, and $R_{\tau(i)}$ is the number of 1's in row $\tau(i)$ of A_i. Finally, set

$$L = \prod_{i \in [n]} R_{\tau(i)}.$$

Example 2.22. *Consider the case where A, σ and τ are as follows (the 1's corresponding to σ are marked in A).*[3]

$$A = \begin{pmatrix} 1 & 0 & \boxed{1} & 1 \\ 1 & \boxed{1} & 0 & 1 \\ 0 & 1 & 1 & \boxed{1} \\ \boxed{1} & 1 & 1 & 0 \end{pmatrix} \qquad \tau = (3124) \qquad \sigma = (3241).$$

Then $A_1 = A$, and $R_{\tau(1)} = R_3$ is the number of 1's in row 3 of A_1, so $R_3 = 3$. Next, notice that $\sigma(\tau(1)) = \sigma(3) = 4$, i.e. the marked 1 in row 3 is in column 4. Thus A_2 is obtained by removing row $\tau(1) = 3$ and column $\sigma(\tau(1)) = 4$, namely,

$$A_2 = \begin{pmatrix} 1 & 0 & \boxed{1} & \\ 1 & \boxed{1} & 0 & \\ & & & \\ \boxed{1} & 1 & 1 & \end{pmatrix}.$$

Now $R_{\tau(2)} = R_1$ is the number of 1's in row 1 of A_2, so $R_1 = 2$. Next, since $\sigma(\tau(2)) = \sigma(1) = 3$,

[3] Recall that the notation $(\sigma_1, \ldots, \sigma_n)$, where $\{\sigma_1, \ldots, \sigma_n\} = [n]$, refers to the permutation σ of $[n]$ that sends i to σ_i (equivalently, $\sigma(i) = \sigma_i$) for $i \in [n]$.

we remove row 1 *and column* 3 *to get*

$$A_3 = \begin{pmatrix} 1 & \boxed{1} & \\ & & \\ \boxed{1} & 1 & \end{pmatrix}.$$

Thus $R_{\tau(3)} = R_2 = 2$. *Finally,*

$$A_4 = \begin{pmatrix} & & \\ & & \\ \boxed{1} & & \end{pmatrix},$$

and $R_{\tau(4)} = R_4 = 1$. *Collecting all the terms, we have* $L = R_1 \cdot \ldots \cdot R_4 = 2 \cdot 2 \cdot 3 \cdot 1 = 12$.

One can think of L as a lazy estimate for $\mathrm{per}(A)$. Indeed, a correct way to calculate $\mathrm{per}(A)$ would be to, at step 1, sum the permanents of all matrices obtained by removing row $\tau(1)$ and a column with a 1 in row $\tau(1)$. Instead, we consider just one such matrix, obtained by removing row $\tau(1)$ and column $\sigma(\tau(1))$, and pretend that the permanents of the other matrices are all the same, so we multiply the result by the number of 1's in row $\tau(1)$. Because of how σ is chosen, the more ways the matrix obtained by removing column i and row $\tau(1)$ has of choosing exactly one 1 from each row and column, the more likely it is to choose to remove column i, for all i such that $A_{\tau(1),i} = 1$. Thus, intuitively, L is likely to be at least as large as $\mathrm{per}(A)$. We formalise this intuition in Claim 2.23 below.

Claim 2.23. $\mathrm{per}(A) \leq \mathbb{G}(L)$.

We can calculated $\mathbb{G}(R_i)$ precisely for all i, as follows.

Claim 2.24. $\mathbb{G}(R_i) = (r_i!)^{1/r_i}$ *for every* $i \in [n]$.

End of lecture 5

Proof of Claim 2.23. We prove that $\mathrm{per}(A) \leq \mathbb{G}(L \,|\, \tau)$, for every fixed permutation τ. This would suffice, as by (6), we would get

$$\mathbb{G}(L) = \prod_{\tau} \mathbb{G}(L \,|\, \tau)^{\mathbb{P}(\tau)} \geq \prod_{\tau} \mathrm{per}(A)^{\mathbb{P}(\tau)} = \mathrm{per}(A).$$

Fix τ. We assume that $\tau(1) = 1$, write $r = r_1$, and assume that the first r elements in row 1 are 1's. (This can be justified by noticing that the permanent does not change by swapping rows or columns.) The proof will be by induction on the dimension of A; it is easy to see that this holds for 1×1 matrices.

We set some notation: let $\rho = \mathrm{per}(A)$ and let ρ_j be the permanent of the matrix obtained from A by removing row 1 and column j, for $j \in [r]$. Then, by induction,

$$\mathbb{G}(L \,|\, \tau \text{ and } \sigma(1) = j) = r \cdot G(R_2 \cdot \ldots \cdot R_n \,|\, \tau \text{ and } \sigma(1) = j) \geq r\rho_j. \tag{7}$$

Indeed, we always have $R_1 = r$ because row 1 is always first to be removed, hence the factor r (using 'linearity of expectation' (5)), and the rest follows by induction.

Next, notice that ρ_j is the number of permutations $\sigma \in S$ for which $\sigma(1) = j$. Hence, $\rho = \sum_j \rho_j$, and the probability that $\sigma(1) = j$ is ρ_j over the total number of permutations, i.e. $\mathbb{P}(\sigma(1) = j) = \frac{\rho_j}{\rho}$. Using this, (7) and the law of total expectation (6),

$$\begin{aligned}\mathbb{G}(L \,|\, \tau) &= \prod_{j \in [r]} \mathbb{G}\big(L \,|\, \tau \text{ and } \sigma(1) = j\big)^{\mathbb{P}(\sigma(1)=j)} \\ &\geq \prod_{j \in [r]} (r\rho_j)^{\rho_j/\rho} \\ &= r \cdot \Big(\prod_{j \in [r]} \rho_j^{\rho_j}\Big)^{1/\rho} \geq \rho = \mathrm{per}(A).\end{aligned} \tag{8}$$

Indeed, for the equality on the first line we used $\sum_{j \in [r]} \rho_j = \rho$. For the inequality in the last line we used that the function $f(x) = x \log x$ is convex for $x \geq 1$, and thus

$$\sum_{j \in [r]} f(\rho_j) \geq r \cdot f\left(\frac{\sum_j \rho_j}{r}\right) = rf(\rho/r) = \rho \log(\rho/r).$$

Taking e to the power of the left- and right-hand sides, we get

$$\prod_{j \in [r]} \rho_j^{\rho_j} \geq (\rho/r)^{\rho},$$

as needed for (8). Notice that (8) completes the proof. □

Proof of Claim 2.24. By symmetry, it suffices to prove this for $i = 1$. We prove that $\mathbb{G}(R_1 \,|\, \sigma) = (r_1!)^{1/r_1}$ for every $\sigma \in S$. By (6), this would prove $\mathbb{G}(R_1) = (r_1!)^{1/r_1}$, as required.

We assume that $\sigma(1) = 1$ and that the first $r = r_1$ elements in row 1 are 1's (as before, this is fine because we can change the order of columns and rows).

Notice that column j is removed at time i exactly when $\tau(i) = \sigma^{-1}(j)$. Thus, the order upon which columns $1, \ldots, r$ are removed is the order of $\sigma^{-1}(1), \ldots, \sigma^{-1}(r)$ within τ. Since this order is uniformly random, it follows that columns $1, \ldots, r$ are removed in a uniformly random order. In particular, the probability that column 1 is the j^{th} to be removed among columns $1, \ldots, r$ is $1/r$.

Observe that R_1 is the number of columns from $1, \ldots, r$ that remain right before row 1 is removed; since $\sigma(1) = 1$, this is the number of columns from $1, \ldots, r$ that remain right before column 1 is removed. By the previous paragraph, this implies $\mathbb{P}(R_1 = j \mid \sigma) = 1/r$. Thus,

$$\mathbb{G}(R_1 \mid \sigma) = \exp\left(\sum_{j\in[r]} \log j \cdot \mathbb{P}(R_1 = j \mid \sigma)\right) = \exp\left(\frac{1}{r} \cdot \sum_{j\in[r]} \log j\right) = \left(\prod_{j\in[r]} j\right)^{1/r} = (r!)^{1/r}.$$

This proves the claim, as $r = r_1$. □

By the two claims and 'linearity of expectation' (5),

$$\mathrm{per}(A) \leq \mathbb{G}(L) = \mathbb{G}\left(\prod_{i\in[n]} R_i\right) = \prod_{i\in[n]} \mathbb{G}(R_i) = \prod_{i\in[n]} (r_i!)^{1/r_i},$$

thus proving the theorem. □

3 Alterations

In this section we continue to apply linearity of expectation, but with a twist: the structure resulting from an initial experiment will need to be altered to fix a small amount of badness.

3.1 Ramsey numbers

Recall that in Theorem 1.11 we showed that the Ramsey number $r(t,t)$ satisfies $r(t,t) \geq 2^{t/2}$ (a more careful analysis of the inequalities there would give $r(t,t) \gtrsim \frac{t\,2^{t/2}}{\sqrt{2}e}$). Here we improve this a bit, using the method of alterations.

We write $g(n) = o(f(n))$ if $\lim_{n\to\infty} \frac{g(n)}{f(n)} = 0$. In particular, $o(1)$ denotes any function $f(n)$ that goes to 0 as $n \to \infty$.

Theorem 3.1. $r(t,t) \geq (1 + o(1))\frac{t\,2^{t/2}}{e}$.

Proof. Let n be an integer, to be determined later. Colour each edge of a complete graph K_n red or blue, randomly and independently. For a set T of t vertices, let X_T be the indicator random variable for the event $\{T \text{ is monochromatic}\}$, and let X be the random variable counting the number of sets of t vertices that are monochromatic. Then $\mathbb{E}(X_T) = 2 \cdot 2^{-\binom{t}{2}}$, and, by linearity of expectation,

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_T X_T\right) = \sum_T \mathbb{E}(X_T) = \binom{n}{t} \cdot 2 \cdot 2^{-\binom{t}{2}}.$$

Fix a colouring with $X \leq \mathbb{E}(X)$, and remove one vertex from each monochromatic clique of size t. The resulting graph has at least $n - \mathbb{E}(X) = n - \binom{n}{t} \cdot 2 \cdot 2^{-\binom{t}{2}}$ vertices and has no monochromatic cliques of size t, showing that

$$r(t,t) \geq n - \binom{n}{t} \cdot 2 \cdot 2^{-\binom{t}{2}}, \tag{9}$$

for every n.

Take $n = e^{-1} \cdot t \cdot 2^{t/2}$. Then

$$\binom{n}{t} \cdot 2 \cdot 2^{-\binom{t}{2}} \leq 2\left(\frac{en}{t}\right)^t 2^{-\binom{t}{2}} \leq 2 \cdot 2^{t^2/2} \cdot 2^{-\binom{t}{2}} = 2 \cdot 2^{t/2}.$$

Plugging in this value of n into (9), we get

$$r(t,t) \geq \frac{t \cdot 2^{t/2}}{e}\left(1 - \frac{2e}{t}\right) = (1 + o(1)) \cdot \frac{t \cdot 2^{t/2}}{e}.$$

□

3.2 Turán theorem

We will now see an easy proof of a weaker version of Turán's theorem, a well known result in extremal graph theory.

Definition 3.2 (Independent sets and cliques)**.** A set of vertices U in a graph G is called *independent* if no two vertices of U are joined by an edge in G (see Figure 7). Similarly, U is called a *clique* if every two of its edges are joined by an edge in G.

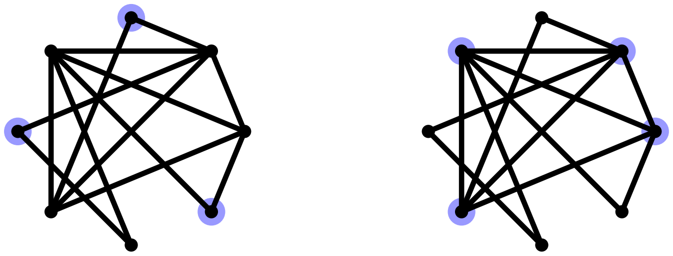


Figure 7: An independent set of size 3 (on the left) and a clique of size 4 (on the right).

Theorem 3.3. *Let G be a graph on n vertices with $\frac{nd}{2}$ edges, where $d \geq 1$. Then G has an independent set of size at least $\frac{n}{2d}$.*

Proof. Let S be a random set of vertices of G, obtained by including each vertex with probability p, independently, where p will be determined later. Let X be the number of vertices in S, and let Y be the number of edges with both ends in S. Write X_v for the indicator random variable for the event $\{v \in S\}$ and Y_{uv} for the indicator random variable for the event $\{u, v \in S\}$. Then $\mathbb{E}(X_v) = \mathbb{P}(v \in S) = p$ and $\mathbb{E}(Y_{uv}) = \mathbb{P}(u, v \in S) = p^2$. Thus

$$\begin{aligned}\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = \mathbb{E}\left(\sum_{v \in V(G)} X_v\right) - \mathbb{E}\left(\sum_{uv \in E(G)} Y_{uv}\right) \\ = \sum_{v \in V(G)} \mathbb{E}(X_v) - \sum_{uv \in E(G)} \mathbb{E}(Y_{uv}) = np - \frac{nd}{2} \cdot p^2 = np\left(1 - \frac{pd}{2}\right).\end{aligned}$$

Pick $p = 1/d$, which maximises the expression above (note that $p \leq 1$). Let S be an outcome for which $X - Y \geq \mathbb{E}(X - Y) = nd/2$. Let S' be a subset of S obtained by removing one vertex from each edge with both ends in S. Then S' is an independent set satisfying $|S'| \geq X - Y \geq nd/2$. □

Remark 3.4. The bound given by Theorem 3.3 is tight up to a factor of about 2. Indeed, if d is an integer and $d+1$ divides n, then the graph that consists of $\frac{n}{d+1}$ pairwise disjoint copies of K_{d+1} has average degree d and its largest independent set has size $\frac{n}{d+1}$.

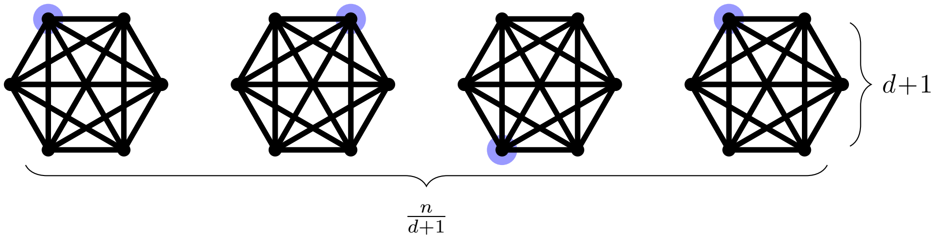


Figure 8: A graph on n vertices with average degree d whose largest independent set (see example) has size $\frac{n}{d+1}$

Turán theorem, asserts that this bound is best possible (Turán's theorem is usually phrased for complete graphs rather than independent sets, but the two version can be seen to be equivalent, by taking the complement graph.)

3.3 Domination

Our next application will give an upper bound on the size of the smallest dominating set in a graph with given minimum degree.

Definition 3.5 (Dominating set)**.** A set of vertices U in a graph G is *dominating* if every vertex in $V(G) - U$ has a neighbour in U (see Figure 9).

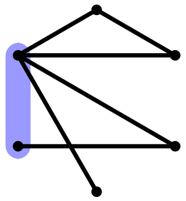


Figure 9: A dominating set in a graph

Recall that log refers to the logarithm with base e.

Theorem 3.6. *Let G be a graph on n vertices with minimum degree δ. Then G has a dominating set of size at most $\frac{(1+\log(\delta+1))n}{\delta+1}$.*

Proof. Write $V := V(G)$, and let $p \in [0,1]$ to be determined later. Let X be a random set of vertices, obtained by including each vertex of G with probability p, independently. Let Y be the set of vertices in $V - X$ that do not have a neighbour in X. Notice that $X \cup Y$ is a dominating set. Indeed, every vertex in $V - (X \cup Y)$ has a neighbour in X, by choice of Y.

To complete the proof, we evaluate the expectation of $|X \cup Y|$, and find a p that minimises the expectation. First, note that $\mathbb{E}(|X|) = np$, which can be seen using linearity of expectation.

Now let us evaluate $\mathbb{E}(|Y|)$. Denote by ψ_v the indicator random variable of the event $\{v \in Y\}$. Observe that v is in Y if and only if both it and all its neighbours are not in X. Thus $\mathbb{P}(\psi_v) = (1-p)^{d(v)} \le (1-p)^{\delta+1}$, where $d(v)$ is the degree of v in G, and we used the assumption that G has minimum degree at least δ. Using linearity of expectation, we get

$$\mathbb{E}(|Y|) = \mathbb{E}\left(\sum_{v \in V} \psi_v\right) = \sum_{v \in V} \mathbb{P}(\psi_v) \le n(1-p)^{\delta+1}.$$

Thus, using linearity of expectation one more time,

$$\mathbb{E}(|X \cup Y|) = \mathbb{E}(|X| + |Y|) = \mathbb{E}(|X|) + \mathbb{E}(|Y|) \le n\left(p + (1-p)^{\delta+1}\right) \le n\left(p + e^{-p(\delta+1)}\right), \qquad (10)$$

where for the last step we used the inequality $1 - p \le e^{-p}$.

Write $f(p) = p + e^{-p(\delta+1)}$. The derivative of f is given by $f'(p) = 1 - (\delta+1)e^{-p(\delta+1)}$. Solving $f'(p) = 0$, we get

$$\begin{aligned} e^{-p(\delta+1)} &= \frac{1}{\delta+1} \\ -p(\delta+1) &= \log\left(\frac{1}{\delta+1}\right) \\ p &= -\log\left(\frac{1}{\delta+1}\right) \cdot \frac{1}{\delta+1} = \frac{\log(\delta+1)}{\delta+1}. \end{aligned}$$

Since $f'(p)$ is increasing, the function $f(p)$ is minimised at $p = \frac{\log(\delta+1)}{\delta+1}$, so we pick this value for the parameter p. Plugging in this value in (10), we find that

$$\mathbb{E}(|X \cup Y|) \le n\left(\frac{\log(\delta+1)}{\delta+1} + e^{-\log(\delta+1)}\right) \le n\left(\frac{\log(\delta+1)}{\delta+1} + \frac{1}{\delta+1}\right) = \frac{(1+\log(\delta+1))n}{\delta+1}.$$

In particular, there is a dominating set of size at most $\frac{(1+\log(\delta+1))n}{\delta+1}$, as claimed. □

Remark 3.7. We could have also guessed a value of p that works (like we did in class) instead of explicitly doing the optimisation. In real life we might not know the target value and then optimising might be the only way to go.

3.4 Dependent Random Choice

For the purpose of this section, given a set of vertices U let $\Gamma(U)$ denote the *common neighbourhood* of the set U, which is the set of vertices joined to all vertices in U (see Figure 10).

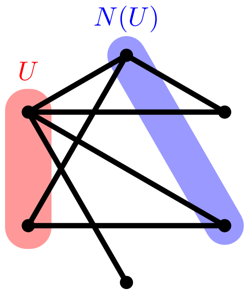


Figure 10: A set U and its common neighbourhood $N(U)$

The following lemma is from a paper called Dependent random choice. It and its variants have many applications, two of which we will see here.

Lemma 3.8 (Fox–Sudakov, 2010)**.** *Let a, m, n, r be positive integers, let $d > 0$, and suppose that t is a positive integer satisfying*

$$\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq a.$$

Then for every graph G on n vertices and with average degree d, there is a set of vertices $U \subseteq V(G)$ such that $|U| \geq a$ and every r vertices in U have at least m common neighbours.

Proof. Let $v_1, \ldots, v_t$ be vertices in G, chosen randomly and independently (i.e. they are chosen *with repetition*, meaning that it could happen that say $v_1 = v_3$). Let A be the common neighbourhood $\Gamma(\{v_1, \ldots, v_t\})$. Then, writing $V := V(G)$ and $d(v)$ for the degree of a vertex v,

$$\mathbb{E}(X) = \sum_{u \in V} \mathbb{P}(u \in A) = \sum_{u \in V} \mathbb{P}(v_1, \ldots, v_t \in N(u)) = \sum_{u \in V} \left(\frac{d(u)}{n}\right)^t \geq n \cdot \left(\frac{d}{n}\right)^t = \frac{d^t}{n^{t-1}}.$$

Here we used linearity of expectation for the first equality, and convexity for the inequality. (Recall that a function $f(x)$ is *convex* if the segment between any two points on the graph $(x, f(x))$ lies above the graph. If a function f satisfies $f''(x) > 0$ then it is convex. Finally, if f is convex then $f(\frac{1}{n}\sum_{i\in[n]} x_i) \leq \frac{1}{n}\sum_{i\in[n]} f(x_i)$ for all $x_1, \ldots, x_n$.)

Now let Y be the number of subsets of A of size r that have fewer than m common neighbours. Then

$$\mathbb{E}(Y) = \sum_{S} \mathbb{P}(S \subseteq A) = \sum_{S} \mathbb{P}\big(v_1, \ldots, v_t \in N(S)\big) = \sum_{S} \left(\frac{|N(S)|}{n}\right)^t < \binom{n}{r}\left(\frac{m}{n}\right)^t,$$

where the sum is over all sets S of size r with $|N(S)| < m$, and the inequality follows from there being a total of $\binom{n}{r}$ sets of r vertices.

Finally, by linearity of expectation,

$$\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) \geq \frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq a.$$

Thus there exists a choice of $v_1, \ldots, v_t$ such that $X - Y \geq a$. For each subset of A of size r with fewer than m common neighbours, remove one of its vertices from A, to obtain a set U which satisfies: $|U| \geq |A| - Y = X - Y \geq a$, and all sets of r vertices from U have at least m common neighbours. □

3.4.1 Turán number of bipartite graphs

The first application of Lemma 3.8 is related to Turán numbers of bipartite graphs.

Definition 3.9 (Turán numbers)**.** The *Turán number* of a graph H, denoted $\text{ex}(n, H)$, is the maximum number of edges in a graph on n vertices which does not contain a copy of H.

Example 3.10.

- $\text{ex}(n, K_2) = 0$ *(every graph with at least one edge has a copy of* K_2*),*
- *Denote by* P_n *the path on* n *vertices. Then* $\text{ex}(n, P_3) = \left\lfloor \frac{n}{2} \right\rfloor$ *(a* P_3*-free graph is a matching, with possibly some isolated vertices).*
- $\text{ex}(n, K_3) = \left\lfloor \frac{n^2}{4} \right\rfloor$ *(this was first proved by Mantel in 1907; the extremal example is the balanced complete bipartite graph* $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$*.*

Turán numbers are important parameters in extremal graphs theory. They are known, at least up to a small error term, for all non-bipartite graphs. For bipartite graph, much less is known. Here is a classical example.

Theorem 3.11 (Kövari–Sós–Turán, 1954)**.** *Let* $r \leq s$ *be positive integers. Then there exists a constant* $c = c(r, s)$ *such that* $\text{ex}(n, K_{r,s}) \leq cn^{2-1/r}$*.*

The following theorem generalises Kövari–Sós–Turán's theorem.

Theorem 3.12 (Alon–Krivelevich–Sudakov, 2003)**.** *Let* H *be a bipartite graph with bipartition* $\{A, B\}$*, such that vertices in* B *have degree at most* r*. Then there is a constant* $c = c(H)$ *such that* $\text{ex}(n, H) \leq cn^{2-1/r}$*.*

Before proving the theorem, we state and prove the following lemma, about embedding bipartite graphs as above.

Lemma 3.13. *Let H be a bipartite graph with bipartition $\{A, B\}$, where $a = |A|$ and $b = |B|$, and the vertices in B have degree at most r. Suppose that G is a graph containing a set of vertices U of size a whose every subset of size r has at least $a + b$ common neighbours. Then G contains a copy of H.*

Proof. We define an injective function $f : V(H) \to V(G)$ as follows. First, map each vertex of A to a different vertex in U (where U is as in the statement). Enumerate B as $\{v_1, \ldots, v_b\}$. For $i \in [b]$, suppose that $f(v_1), \ldots, f(v_{i-1})$ have been defined; we will show how to define $f(v_i)$. Write $S = \{f(u) : u \in N_H(v_i)\}$ (so S is the set of vertices in U corresponding to neighbours of v_i). Then $|S| \leq r$, because vertices in B have degree at most r in H. Thus, by assumption, S has at least $a + b$ common neighbours in G, and so there is a vertex u which is a common neighbour of S and is not the image of a vertex in $A \cup \{v_1, \ldots, v_{i-1}\}$. Define $f(v_i) = u$. It is easy to check that the resulting f, obtained by running the above procedure for $i = 1, \ldots, b$, is an injective function that sends edges of H to edges of G (meaning that $f(u)f(v)$ is an edge in G if uv is an edge in H), showing that G contains a copy of H. □

We now prove the theorem.

Proof of Theorem 3.12. Write $a = |A|$, $b = |B|$, $m = a + b$. Let c satisfy $c \geq \max\{a^{1/r}, \frac{e(a+b)}{r}\}$. Suppose that G is a graph on n vertices with at least $cn^{2-1/r}$ edges. Then G has average degree at least $2cn^{1-1/r}$; write $d = 2cn^{1-1/r}$. Then

$$\frac{d^r}{n^{r-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^r \geq \frac{(2c)^r n^{r-1}}{n^{r-1}} - \left(\frac{en}{r}\right)^r \left(\frac{a+b}{n}\right)^r = (2c)^r - \left(\frac{e(a+b)}{r}\right)^r \geq c^r \geq a.$$

Here we used $\binom{n}{r} \leq \left(\frac{en}{r}\right)^r$ and the choice of c. It follows from Lemma 3.8 that G contains a set U of size at least a whose every subset of size r has at least $m = a + b$ common neighbours. Thus, by Lemma 3.13, G contains a copy of H. This proves that every graph on n vertices with at least $cn^{2-1/r}$ edges contains a copy of H, as required. □

3.4.2 Ramsey number of the hypercube

Next, we consider the Ramsey number of the hypercube. Similarly to Definition 1.7, the *Ramsey number* of a graph H, denoted $r(H)$, is the minimum n such that every red-blue edge-colouring of K_n contains a monochromatic copy of H.

Definition 3.14 (Hypercube)**.** The *hypercube* of dimension r is the graph whose vertices are $\{0, 1\}$-sequences of length r, whose edges join two sequences that differ in exactly one coordinate (see Figure 11).

The next theorem proves an upper bound on the Ramsey number of the hypercube (recall that our previous Ramsey theory results were upper bounds on the Ramsey number of the complete graph).

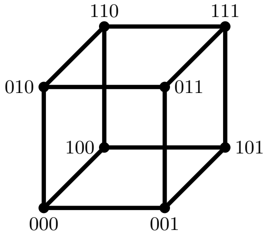


Figure 11: The hypercube Q_3

Theorem 3.15. *The Ramsey number of the hypercube Q_r satisfies $r(Q_r) \leq 2^{3r}$.*

Proof. Write $n = 2^r$ and $N = n^3$. Consider a red-blue colouring of the complete graph on N vertices; we need to show that it contains a monochromatic copy of Q_r. Without loss of generality, there are at least as many red edges as there are blue ones; denote the subgraph of red edges by G. Then $e(G) \geq \frac{1}{2}\binom{N}{2}$. Let d be the average degree of G, then $d \geq \frac{2e(G)}{N} = \frac{N-1}{2}$. Let $t = \frac{3r}{2}$, $m = n$, and $a = 2^{r-1}$. Then

$$\binom{N}{r}\left(\frac{m}{N}\right)^t \leq \left(\frac{en^3}{r}\right)^r \left(\frac{n}{n^3}\right)^{3r/2} = \left(\frac{e}{r}\right)^r < 1,$$

if $r \geq 3$. Also,

$$\frac{d^t}{N^{t-1}} = \left(\frac{d}{N}\right)^t N \geq \left(\frac{N-1}{2N}\right)^t N = 2^{-t} N \left(\left(1 - \frac{1}{N}\right)^N\right)^{t/N} \geq \frac{1}{2} \cdot n^{3/2} \geq n.$$

Where we used the fact that the sequence $\left(1 - \frac{1}{N}\right)^N$ tends to $1/e$ as N tends to infinity, and t/N tends to 0. Altogether,

$$\frac{d^t}{N^{t-1}} - \binom{N}{r}\left(\frac{m}{N}\right)^t \geq n - 1 \geq a.$$

It follows from Lemma 3.8 that there is a set U of a vertices in G whose every subset of size r has at least m common neighbours. By Lemma 3.13, G contains a copy of Q_r. □

Remark 3.16. Notice that what we showed is that any graph on N vertices with at least $\frac{1}{2}\binom{N}{2}$ edges contains a hypercube Q_r. Such a result is known as a *density result*.

Remark 3.17. A famous conjecture asserts that $r(Q_r) \leq c \cdot 2^r$, for some constant $c > 0$. The best known upper bound is $c \cdot 2^{(2-\varepsilon)r}$, for a (small) constant $\varepsilon > 0$ and a constant $c > 0$.

4 The second moment

So far, we have often used that, given a random variable X, it satisfies $X \geq \mathbb{E}(X)$, with positive probability. As the expectation is sometimes referred to as the *first moment*, this method is called the *first moment method*. In this section we will consider the *second moment*, namely $\mathbb{E}(X^2)$, to be able to argue that X is close to its expectation *with high probability* (namely, with probability close to 1).

4.1 Chebyshev's inequality

We first recall the definition of variance and two important inequalities.

Definition 4.1 (Variance)**.** The *variance* of a random variable X, denoted $\mathrm{Var}(X)$, is defined as

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

Equivalently, $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Notice that $\mathrm{Var}(X) \geq 0$ always. We sometimes denote a variance as σ^2 (where $\sigma \geq 0$).

Proposition 4.2 (Markov's inequality)**.** *Suppose that X is a non-negative random variable, and let $\lambda > 0$. Then*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}.$$

Proof.

$$\mathbb{E}(X) = \sum_x x \cdot \mathbb{P}(X = x) \geq \sum_{x \geq \lambda} \lambda \cdot \mathbb{P}(X = x) = \lambda \cdot \mathbb{P}(X \geq \lambda),$$

where we used the non-negativity of X for the second inequality. The desired inequality follows. □

Proposition 4.3 (Chebyshev's inequality)**.** *Let X be a random variable, and let $\lambda > 0$. Then*

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq \lambda\big) \leq \frac{\mathrm{Var}(X)}{\lambda^2}.$$

Proof. Write $\mu = \mathbb{E}(X)$.

$$\mathbb{P}\big(|X - \mu| \geq \lambda\big) = \mathbb{P}\big((X - \mu)^2 \geq \lambda^2\big) \leq \frac{\mathbb{E}\big((X - \mu)^2\big)}{\lambda^2} = \frac{\mathrm{Var}(X)}{\lambda^2},$$

where we used Markov's inequality for the inequality. □

It is often convenient to use the following corollary of Chebyshev's inequality.

Corollary 4.4. *Let X be a random variable. Then*

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2}.$$

Proof.

$$\mathbb{P}(X = 0) \leq \mathbb{P}\big(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)\big) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2},$$

where for the last inequality we used Chebyshev's inequality (with $\lambda = \mathbb{E}(X)$). □

Remark 4.5. A typical way of using the above inequalities is as follows. Suppose that X is a random variable counting certain objects. Then if $\mathbb{E}(X)$ is small, by Markov's inequality, the probability that $X \geq 1$ is also small, and so the probability that $X = 0$ is large. In the other direction, if $\frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2}$ is small, then the probability that $X = 0$ is also small, by the above corollary of Chebyshev's inequality.

Recall the definition of covariance.

Definition 4.6 (Covariance)**.** The *covariance* of random variables X and Y, denoted $\mathrm{Cov}(X, Y)$, is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

We recall a few useful facts about variance and covariance (without proof).

Proposition 4.7. *Let $X, Y, X_1, \ldots, X_n$ be random variables.*

(a) *If $X = \sum_{i \in [n]} X_i$, then*

$$\mathrm{Var}(X) = \sum_{i,j \in [n]} \mathrm{Cov}(X_i, X_j).$$

(b) *If X, Y are independent, then* $\mathrm{Cov}(X, Y) = 0$.

(c) *If $X_1, \ldots, X_n$ are pairwise independent and $X = \sum_{i \in [n]} X_i$, then*

$$\mathrm{Var}(X) = \sum_{i \in [n]} \mathrm{Var}(X_i).$$

4.2 Threshold for containing K_4

Our first example in this chapter will be about random graphs, defined as follows.

Definition 4.8 (Random graph)**.** The *Erdős–Rényi random graph* (or *random graph* in short) $G(n, p)$, is the graph on vertex set $[n]$, where each pair of edges is joined with probability p, independent.

Recall that we have seen applications of random graphs when thinking about Ramsey numbers.

We will say that a sequence of events (A_n) holds *with high probability* (or w.h.p. in short), if $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. A typical question in the study of random graphs is: for which $p = p(n)$ does a given property holds with high probability? For example: for which p is it true that $G(n, p)$ is connected, with high probability? For which p does $G(n, p)$ contain a copy of a specific graph H, with high probability? In the next proposition we investigate the case $H = K_4$.

Proposition 4.9.

(a) *If $n^2 p^3 \to 0$ as $n \to \infty$ then, with high probability, $G(n, p)$ has no copy of K_4.*

(b) *If $n^2p^3 \to \infty$ as $n \to \infty$ then, with high probability, $G(n,p)$ has a copy of K_4.*

Proof. For a set S of four vertices, denote by X_S the indicator random variable of the event $\{S \text{ is a clique}\}$. Write $X = \sum_S X_S$, where the sum is over all sets of four vertices. For (a), we need to show that if $n^2p^3 \to 0$ then, with high probability, $X = 0$; and, for (b), we need to show that if $n^2p^3 \to \infty$, then, with high probability, $X \geq 1$.

First, we calculate the expectation. Notice that $\mathbb{E}(X_S) = \mathbb{P}(\{S \text{ is a clique}\}) = p^6$. Thus,

$$\mathbb{E}(X) = \sum_S \mathbb{E}(X_S) = \binom{n}{4}p^6.$$

This allows us to complete the first task. Indeed, suppose that $n^2p^3 \to 0$. Then, by Markov's inequality (Proposition 4.2) (observing that $X \geq 0$),

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}(X) = \binom{n}{4}p^6 \leq n^4p^6 = (n^{2/3}p)^6 \to 0.$$

Equivalently, with high probability, $X = 0$, as required.

Next, we wish to calculate the variance of X. With this in mind, notice that for two sets S and T of four vertices,

$$\mathrm{Cov}(X_S, X_T) = \mathbb{E}(X_SX_T) - \mathbb{E}(X_S)\mathbb{E}(X_T) = \begin{cases} 0 & |S \cap T| \leq 1 \\ p^{11} - p^{12} & |S \cap T| = 2 \\ p^9 - p^{12} & |S \cap T| = 3 \\ p^6 - p^{12} & S = T. \end{cases}$$

Thus,

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{S,T} \mathrm{Cov}(X_S, X_T) \\ &= \sum_S \sum_{T:|S\cap T|=2} \mathrm{Cov}(X_S, X_T) + \sum_S \sum_{T:|S\cap T|=3} \mathrm{Cov}(X_S, X_T) + \sum_S \sum_{T:|S\cap T|=4} \mathrm{Cov}(X_S, X_T) \\ &\leq n^6(p^{11} - p^{12}) + n^5(p^9 - p^{12}) + n^4(p^6 - p^{12}) \leq n^6p^{11} + n^5p^9 + n^4p^6. \end{aligned}$$

Indeed, the number of ways to choose S, T of size 4 that intersect on i vertices is at most n^{8-i}, as we have at most n^i ways to choose the vertices in the intersection, and at most n^{4-i} to choose the remaining vertices in S, and similarly for the remaining vertices in T.

Now we prove (b), so let us assume that $n^2p^3 \to \infty$. By Corollary 4.4,

$$\begin{aligned} \mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2} &\leq \frac{1}{2500} \cdot \frac{n^6p^{11} + n^5p^9 + n^4p^6}{n^8p^{12}} \\ &= \frac{1}{2500} \cdot (n^{-2}p^{-1} + n^{-3}p^{-3} + n^{-4}p^{-6}) \to 0. \end{aligned}$$

For the limit, we have $n^2p^3 \to \infty$, implying that $(np)^2 \to \infty$, and thus also $(np)^3 \to \infty$. Moreover, again by assumption we also have $n^2p \to \infty$. Altogether, all three terms tend to 0 as n tends to infinity. This shows that, in this range, $X \geq 1$ with high probability, as required. $\square$

A *graph property* $\mathcal{P}$ is a collection of graphs. For example: being connected, not having a copy of a certain graph H as a subgraph.

A graph property $\mathcal{P}$ is *monotone* if the property is maintained by adding an edge between existing vertices. For example, the property of being connected is monotone, the property of having an even number of edges is not.

Definition 4.10 (Threshold functions)**.** For a graph property $\mathcal{P}$, a function $p_0 : \mathbb{N} \to [0,1]$ is a *threshold function* for $\mathcal{P}$ if the following two properties hold.

- If $\frac{p}{p_0} \to 0$ as $n \to \infty$ then $G(n,p)$ does not satisfy $\mathcal{P}$, with high probability.
- If $\frac{p}{p_0} \to 1$ as $n \to \infty$ then $G(n,p)$ satisfies $\mathcal{P}$, with high probability.

So, what we have shown above is that $p = n^{-2/3}$ is a threshold function for the property of containing a copy of K_4.

4.3 Distinct sums

Definition 4.11 (Distinct sums)**.** A set S of positive integers is said to have *distinct sums* if the sums $\sum_{t \in T} t$, with $T \subseteq S$, are distinct (we think of the sum of the empty set as 0).

Let $f(n)$ be the largest k such that $[n]$ contains a subset of size k with distinct sums.

Proposition 4.12. $\lfloor \log_2 n \rfloor \leq f(n) \leq \log_2 n + \log_2 \log_2 n + 2.$

Proof. For the lower bound, consider the set of powers of 2 in $[n]$.

For the upper bound, notice that if $S \subseteq [n]$ is a set with distinct sums such that $|S| = k$, then its subsets define 2^k distinct sums, each of which is at most kn, showing: $2^k \leq kn$. If $k \geq \log_2 n + \log_2 \log_2 n + 2$, then

$$\frac{2^k}{k} \geq \frac{2^{\log_2 n + \log_2 \log_2 n + 2}}{\log_2 n + \log_2 \log_2 n + 2} = \frac{4n \log_2 n}{\log_2 n + \log_2 \log_2 n + 2} > n,$$

using that the function $\frac{2^k}{k}$ is increasing, and that $\log_2 \log_2 n + 2 \leq 2 \log_2 n$ (which holds for $n \geq 2$; for $n = 1$ the statement of the proposition clearly holds). $\square$

In the next theorem we improve the error term in the upper bound, using the second moment method.

Theorem 4.13. $f(n) \leq \log_2 n + \frac{1}{2}\log_2\log_2 n + 4$.

Proof. Let $x_1, \ldots, x_k \in [n]$ be distinct elements such that $\{x_1, \ldots, x_k\}$ has distinct sums. Let $\varepsilon_1, \ldots, \varepsilon_k$ be independent random variables, with $\mathbb{P}(\varepsilon_i = 0) = \mathbb{P}(\varepsilon_i = 1) = \frac{1}{2}$. Define $X = \sum_{i\in[k]} \varepsilon_i x_i$. Write $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathrm{Var}(X)$ (with $\sigma \geq 0$). Then

$$\begin{aligned}\sigma^2 &= \sum_{i\in[k]} \mathrm{Var}(\varepsilon_i x_i) = \sum_{i\in[k]} x_i^2 \,\mathrm{Var}(\varepsilon_i) = \sum_{i\in[k]} x_i^2(\mathbb{E}(\varepsilon_i^2) - (\mathbb{E}(\varepsilon_i))^2) = \sum_{i\in[k]} x_i^2(\mathbb{E}(\varepsilon_i) - (\mathbb{E}(\varepsilon_i))^2)\\ &= \frac{1}{4}\sum_{i\in[k]} x_i^2 \leq \frac{n^2 k}{4}.\end{aligned}$$

Let $\lambda > 0$ to be determined later. By Chebyshev's inequality (Proposition 4.3),

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

Equivalently,

$$\mathbb{P}\big(|X-\mu| < \lambda\sigma\big) \geq 1 - \frac{1}{\lambda^2}. \tag{11}$$

Crucially, for every positive integer a, we have $\mathbb{P}(X = a) \leq 2^{-k}$, by the distinct sums property. Thus,

$$\mathbb{P}\big(|X-\mu| < \lambda\sigma\big) \leq 2^{-k}\cdot(2\lambda\sigma+1) \leq 2^{-k}\cdot\big(\lambda n\sqrt{k}+1\big). \tag{12}$$

Combining (11) and (12),

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

Rearranging,

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

Assuming that $k \geq \log_2 n + \frac{1}{2}\log_2\log_2 n + 4$, and plugging in $\lambda = 2$ (the precise value of λ does not change much, as long as it is larger than 1), we get

$$n \geq \frac{\frac{3}{4}\cdot 2^k}{3\sqrt{k}} \geq \frac{2^k}{4\sqrt{k}} \geq \frac{16n\sqrt{\log_2 n}}{4\sqrt{\log_2 n + \frac{1}{2}\log_2\log_2 n + 4}} > n,$$

using $\frac{1}{2}\log_2\log_2 n + 4 \leq 8\log_2 n$. □

4.4 Prime divisors

For a positive integer n, let $\nu(n)$ be the number of prime divisors of n. Clearly, $\nu(n)$ can vary substantially: we have $\nu(n) = 1$ if n is prime, and for certain values of n we have $\nu(n) \approx \frac{\log n}{\log\log n}$. Nevertheless, it turns out that 'almost all' value of n satisfy $\nu(n) \approx \log\log n$. This is what we show in the following theorem. Recall that $\log x$ refers to the natural logarithm (namely, the base e logarithm).

Theorem 4.14 (Hardy–Ramanujan, 1920 (this proof is due to Turán, 1934))**.** *For every $\varepsilon > 0$ there is a constant $c > 0$ such that*

$$\left|\left\{x \in [n] : |\nu(x) - \log\log n| > c\sqrt{\log\log n}\right\}\right| \leq \varepsilon n.$$

Proof. Let x be chosen randomly from $[n]$. For a prime p, let Y_p be the indicator random variable for the event $\{p \text{ divides } x\}$, $M = n^{1/2}$ and $Y = \sum_{p \leq M} Y_p$. Notice that every $x \in [n]$ has at most one prime factor that is larger than M, and so $|Y - \nu(x)| \leq 1$. It thus suffices to show that $|Y - \log\log n| \leq c\sqrt{\log\log n}$, with high probability.

Now,

$$\mathbb{E}(Y_p) = \frac{\lfloor n/p \rfloor}{n} = \frac{1}{p} + O(n^{-1}),$$

using $a - 1 < \lfloor a \rfloor \leq a$. By linearity of expectation,

$$\mathbb{E}(Y) = \sum_{p \,\leq\, M \text{ prime}} \mathbb{E}(Y_p) = \sum_{p \,\leq\, M \text{ prime}} \frac{1}{p} + O(1) = \log\log n + O(1).$$

For the last equality, we used a result from number theory which we will not prove here.

Next, we estimate the variance of Y, using the following formula

$$\operatorname{Var} Y = \sum_{p} \operatorname{Var}(Y_p) + \sum_{p \neq q} \operatorname{Cov}(Y_p, Y_q). \tag{13}$$

(Here the sum is over p and q which are primes in $[M]$.)

First, note that $\operatorname{Var}(Y_p) = \mathbb{E}(Y_p^2) - (\mathbb{E}(Y_p)) \leq \mathbb{E}(Y_p)(1 - \mathbb{E}(Y_p)) \leq \mathbb{E}(Y_p) \leq \frac{1}{p}$. Thus,

$$\sum_{p} \operatorname{Var}(Y_p) \leq \sum_{p} \frac{1}{p} = \log\log n + O(1), \tag{14}$$

using the number theory result mentioned above.

Second, note that if p, q are distinct primes in $[M]$ then

$$\begin{aligned}\operatorname{Cov}(Y_p, Y_q) &= \mathbb{E}(Y_p Y_q) - \mathbb{E}(Y_p)\mathbb{E}(Y_q)\\ &\leq \frac{n/pq}{n} - \frac{n/p - 1}{n} \cdot \frac{n/q - 1}{n}\\ &= \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{q} - \frac{1}{n}\right) \leq \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right).\end{aligned}$$

Thus,

$$\sum_{p \neq q} \operatorname{Cov}(Y_p, Y_q) \leq \frac{1}{n} \sum_{p \neq q} \left(\frac{1}{p} + \frac{1}{q}\right) \leq \frac{M}{n} \sum_{p} \frac{1}{p} = O(n^{-1/2} \log\log n) = O(1), \tag{15}$$

using that, trivially, there are at most M primes in $[M]$ for the second inequality.

Altogether, combining (13), (14) and (15), we get

$$\operatorname{Var}(Y) = \sum_p \operatorname{Var}(Y_p) + \sum_{p \neq q} \operatorname{Cov}(Y_p, Y_q) = \log\log n + O(1).$$

By Chebyshev's inequality (Proposition 4.3),

$$\mathbb{P}\left(|Y - \mathbb{E}(Y)| \geq \lambda\sqrt{\operatorname{Var}(Y)}\right) \leq \frac{1}{\lambda^2}.$$

That is, the number of elements $x \in [n]$ that satisfy $|Y(x) - \mathbb{E}(Y)| \geq \lambda\sqrt{\operatorname{Var}(Y)}$ is at most $\frac{n}{\lambda^2}$. For every $x \in [n]$ *not* satisfying this we have

$$\begin{aligned} |\nu(x) - \log\log n| &\leq |\nu(x) - Y(x)| + |Y(x) - \mathbb{E}(Y)| + |\mathbb{E}(Y) - \log\log n| \\ &\leq \lambda\sqrt{\operatorname{Var}(Y)} + O(1) = \lambda\sqrt{\log\log n} + O(1) \leq 2\lambda\sqrt{\log\log n}. \end{aligned}$$

using $\nu(x) \leq Y(x) \leq \nu(x) + 1$, the assumption on x, and $\mathbb{E}(Y) = \log\log n + O(1)$ and $\mathbb{E}(Y) = \log\log n + O(1)$ for the equality. In other words

$$\left|\left\{x \in [n] : |\nu(x) - \log\log n| > 2\lambda\sqrt{\log\log n}\right\}\right| \leq \frac{n}{\lambda^2}.$$

This proves the theorem (for $\varepsilon > 0$, can take $c = \frac{2}{\sqrt{\varepsilon}}$). □

4.5 Clique number of random graphs

Definition 4.15. A *clique* in a graph G is a set of vertices in G whose every two vertices are joined by an edge. The *clique number* of G, denoted $\omega(G)$, is the size of a largest clique in G (see Figure 12).

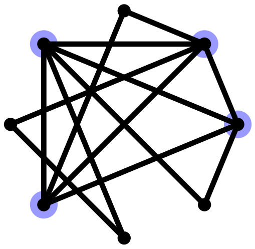


Figure 12: A graph with clique number 4

In this section we calculate (with high probability) the clique number of $G(n, 1/2)$. Surprisingly, this number is concentrated on one or two values, depending on n. Define $f : \mathbb{N} \to \mathbb{N}$ as follows.

$$f(k) = \binom{n}{k} 2^{-\binom{k}{2}}.$$

Notice that $f(k)$ is the expected number of cliques of size k in $G(n, 1/2)$. Denote by $k_0 = k_0(n)$ the largest k such that $f(k) \geq n^{1/4}$; such a k_0 exists because $f(1) = n \geq n^{1/4}$ and $f(n) < n^{1/4}$.

Theorem 4.16. *With high probability, the clique number of $G(n, 1/2)$ is either k_0 or $k_0 + 1$.*

Before the proof of the theorem, we prove two numerical claims. The first estimates k_0.

Claim 4.17. $2\log_2 n - 4\log_2\log_2 n \leq k_0 \leq 2\log_2 n$ *for large enough n.*

Proof. If $k \geq 2\log_2 n$ then

$$f(k) \leq \left(\frac{en}{k}\right)^k 2^{-k(k-1)/2} = \left(\frac{\sqrt{2}en}{k2^{k/2}}\right)^k \leq \left(\frac{4n}{kn}\right)^k \leq 1,$$

where the last inequality holds when $k \geq 4$, say, which is the case when $n \geq 2$. This shows $k_0 \leq 2\log_2 n$.

Now, if $\log_2 n \leq k \leq 2\log_2 n - 4\log_2\log_2 n$, then

$$f(k) \geq \left(\frac{n}{k}\right)^k 2^{-k^2/2} = \left(\frac{n}{k2^{k/2}}\right)^k \geq \left(\frac{n}{2\log_2 n \cdot \frac{n}{(\log_2 n)^2}}\right)^k = \left(\frac{\log_2 n}{2}\right)^k \geq 2^{\log_2 n} = n \geq n^{1/4},$$

where the penultimate inequality holds for $n \geq 16$. It follows that $k_0 \geq 2\log_2 n - 4\log_2\log_2 n$ for large enough n. □

Notice that this claim, along with Theorem 4.16 which we will prove shortly, implies the following corollary.

Corollary 4.18. *With high probability, $\omega(G(n, 1/2)) = 2\log_2 n + O(\log\log n)$.*

The next claim lower bounds the ration $\frac{f(k)}{f(k+1)}$ (for large enough k).

Claim 4.19. *Let $k \geq 1.99\log_2 n$. Then $\frac{f(k)}{f(k+1)} \geq \sqrt{n}$.*

Proof.

$$\begin{aligned} \frac{f(k)}{f(k+1)} = \frac{\binom{n}{k}2^{-\binom{k}{2}}}{\binom{n}{k+1}2^{-\binom{k+1}{2}}} &= \frac{(k+1)2^k}{n-k} \\ &\geq \frac{1}{n} \cdot 2^{1.99\log_2 n} = n^{0.99n} \geq n^{1/2}. \end{aligned} \tag{16}$$

Here we used that $\frac{k+1}{n-k} \geq \frac{1}{n}$ for $k \geq 0$ (as the fraction increases with k) and the lower bound on k for evaluating 2^k. □

Finally, we prove the theorem.

Proof of Theorem 4.16. In order to prove the theorem, we need to prove that the following two things holds with high probability: the clique number of $G = G(n, 1/2)$ is at most $k_0 + 1$; and that it is at least k_0. The first task is easier, so we do it first.

Let k satisfy $k \geq k_0$. Then, for large enough n, by Claims 4.17 and 4.19, we have $f(k_0 + 2) \leq n^{-1/2} \cdot f(k_0 + 1) \leq n^{-1/4}$. By Markov's inequality, recalling that $f(k_0 + 2)$ is the expected number of cliques of size $k_0 + 2$, we get

$$\mathbb{P}(\text{there is at least one cliques of size } k_0 + 2) \leq f(k_0 + 2) \leq n^{-1/4}.$$

So, with high probability, there are no cliques of size $k_0 + 2$, i.e. the clique number is at most $k_0 + 1$.

Now, we turn to the second task, of showing that, with high probability, there is a clique of size k_0. From now on, we set $k := k_0$ to avoid notational clutter. For a set of k vertices S, denote by X_S the indicator random variable for the event $\{S \text{ is a clique}\}$, and let $X = \sum_S X_S$. For two sets S, T of k vertices, writing $i := |S \cap T|$, we have

$$\operatorname{Cov}(X_S, X_T) \begin{cases} = 0 & i \in \{0, 1\} \\ \leq \mathbb{E}(X_S X_T) = 2^{-2\binom{k}{2} + \binom{i}{2}} & i \geq 2. \end{cases}$$

Thus,

$$\begin{aligned} \operatorname{Var}(X) = \sum_{S,T} \operatorname{Cov}(X_S, X_T) &\leq \sum_{2 \leq i \leq k} \sum_{S,T : |S \cap T| = i} 2^{-2\binom{k}{2} + \binom{i}{2}} \\ &= \sum_{2 \leq i \leq k} \binom{n}{k} \binom{k}{i} \binom{n-k}{k-i} 2^{-2\binom{k}{2} + \binom{i}{2}}, \end{aligned}$$

and, writing $g(i) := \frac{\binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}}}{\binom{n}{k}}$,

$$\frac{\operatorname{Var}(X)}{(\mathbb{E}(X))^2} = \frac{\operatorname{Var}(X)}{\left(\binom{n}{k} 2^{-\binom{k}{2}}\right)^2} \leq \sum_{2 \leq i \leq k} \frac{\binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}}}{\binom{n}{k}} = \sum_{2 \leq i \leq k} g(i).$$

To complete the proof, it suffices to show that $g(i) \leq n^{-1/4}$ for every $i \in [2, k]$. Indeed, if that is the case we get

$$\frac{\operatorname{Var}(X)}{(\mathbb{E}(X))^2} \leq k n^{-1/4} \leq 2 \log_2 n \cdot n^{-1/4} \to 0.$$

By Corollary 4.4, this would show that $X \geq 1$, with high probability. Equivalently, with high probability, there is a clique of size k.

It is quite easy to show that $g(i) \le n^{-1/4}$ for $i \in \{2, k\}$. Indeed,

$$\begin{aligned} g(2) &= \frac{\binom{k}{2}\binom{n-k}{k-2}2^{\binom{2}{2}}}{\binom{n}{k}} \le \frac{k^2 \frac{(n-k)^{k-2}}{(k-2)!}}{\frac{(n-k)^k}{k!}} \le \frac{k^4}{(n-k)^2} \le n^{-1/4}. \\ g(k) &= \frac{2^{\binom{k}{2}}}{\binom{n}{k}} = \frac{1}{f(k)} \le n^{-1/4}, \end{aligned}$$

where we used that n is large for the first line, and the choice of k for the second.

Claim 4.20. $g(i) \le \max\{g(2), g(k)\}$ *for* $i \in [2, k]$.

The proof of Theorem 4.16 follows from the claim, as explained above. □

End of lecture 11

Proof of Claim 4.20. We will show that there is an i_0 such that $g(2) \ge g(3) \ge \ldots \ge g(i_0)$ and $g(i_0) \le g(i_0+1) \le \ldots \le g(k)$. This would show $g(i) \le g(2)$ if $i \in [2, i_0]$, and $g(i) \le g(k)$ for $i \in [i_0, k]$, which implies the claim. Write $h(i) := \frac{g(i+1)}{g(i)}$. Then

$$h(i) = \frac{g(i+1)}{g(i)} = \frac{\binom{k}{i+1}\binom{n-k}{k-i-1}2^{\binom{i+1}{2}}}{\binom{k}{i}\binom{n-k}{k-i}2^{\binom{i}{2}}} = \frac{(k-i)^2 \cdot 2^i}{(i+1)(n-2k+i+1)}.$$

So our task is to show that there is i_0 such that $h(i) \le 1$ for $i \in [2, i_0 - 1]$ and $h(i) \ge 1$ for $i \in [i_0, k-1]$. It is easy to check that $h(1) < 1$ and $h(k-1), h(k-2), h(k-3) > 1$. Now, for $i \in [1, k-4]$,

$$\begin{aligned} \frac{h(i+1)}{h(i)} &= \frac{(k-i-1)^2 \cdot 2^{i+1}}{(i+2)(n-2k+i+2)} \cdot \frac{(i+1)(n-2k+i+1)}{(k-i)^2 \cdot 2^i} \\ &= 2\left(1 - \frac{1}{k-i}\right)^2 \left(1 - \frac{1}{i+2}\right)\left(1 - \frac{1}{n-2k+i+2}\right) > 1. \end{aligned}$$

It follows that $h(1) < \ldots < h(k-3)$. Since $h(1) < 1$ and $h(k-1), h(k-2), h(k-3) > 1$, the existence of a suitable i_0 follows. □

5 Concentration inequalities: Chernoff's bound

5.1 Chernoff bounds

The second moment method allowed us to show that a random variable X is, with high probability, close to its mean. In this section we show that for X which is the sum of independent identically-distributed random variables, X is close to its mean, with very high probability.

Here is the simplest version of this statement.

Theorem 5.1 (Chernoff)**.** *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = 1/2$, and write $X = X_1 + \ldots + X_n$. Then, for every $a \geq 0$,*

$$\mathbb{P}(X > a) \leq e^{-a^2/2n}, \qquad \mathbb{P}(X < -a) \leq e^{-a^2/2n}.$$

Proof. Notice that $\mathbb{P}(X \geq a) = \mathbb{P}(-X \geq a) = \mathbb{P}(X \leq -a)$, by symmetry. It thus suffices to prove the first inequality. Let $\lambda \geq 0$ be a parameter to be determined later.

$$\mathbb{P}(X \geq a) = \mathbb{P}\left(e^{\lambda X} \geq e^{\lambda a}\right) \leq \frac{\mathbb{E}\left(e^{\lambda X}\right)}{e^{\lambda a}}. \tag{17}$$

We now estimate $\mathbb{E}\left(e^{\lambda X}\right)$.

$$\mathbb{E}\left(e^{\lambda X}\right) = \mathbb{E}\left(e^{\sum_i \lambda X_i}\right) = \prod_i \mathbb{E}\left(e^{\lambda X_i}\right) = \prod_i \frac{e^{-\lambda} + e^{\lambda}}{2} \leq \prod_i e^{\lambda^2/2} = e^{n\lambda^2/2}. \tag{18}$$

Here the second equality follows from the independence of the X_i's and the inequality can be easily verified using the Taylor expansions of e^x. Indeed, recall that $e^x = \sum_{i \geq 0} \frac{x^i}{i!}$. Thus

$$\frac{e^{-\lambda} + e^{\lambda}}{2} - e^{\lambda^2/2} = \sum_{i \geq 0} \frac{1}{2}\left(\frac{(-\lambda)^i}{i!} + \frac{\lambda^i}{i!}\right) - \sum_{i \geq 0} \frac{(\lambda^2/2)^i}{i!} = \sum_{i \geq 0}\left(\frac{\lambda^{2i}}{(2i)!} - \frac{\lambda^{2i}}{2^i i!}\right) \geq 0,$$

using $(2i)! \geq 2^i i!$ which holds for all $i \geq 0$.

Combining (17) and (18), we get

$$\mathbb{P}(X \geq a) \leq \exp\left(\frac{n\lambda^2}{2} - \lambda a\right).$$

Plugging in $\lambda = a/n$ (obtained from optimising), we get $\mathbb{P}(X \geq a) \leq e^{-a^2/2n}$, as required. □

Remark 5.2. Let us compare the performance of the Chebyshev and Chernoff bounds, for X as in Theorem 5.1. Taking $a = \lambda\sqrt{n}$, Chebyshev's inequality give

$$\mathbb{P}\left(|X| \geq \lambda\sqrt{n}\right) \leq \frac{\mathrm{Var}(X)}{\lambda^2 n} = \frac{1}{\lambda^2},$$

using that $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) = n$. Chernoff gives

$$\mathbb{P}\left(|X| \geq \lambda\sqrt{n}\right) \leq 2e^{-\lambda^2}.$$

This is much better than what Chebyshev gives (at least for somewhat large λ).

Corollary 5.3. *Let $X_1, \ldots, X_n$ be independent random variables satisfying $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$, and write $X = X_1 + \ldots + X_n$. Then, for every $a \geq 0$,*

$$\mathbb{P}(X > \mathbb{E}(X) + a) \leq e^{-2a^2/n}, \qquad \mathbb{P}(X < \mathbb{E}(X) - a) \leq e^{-2a^2/n}.$$

Proof. Write $Y_i = 2X_i - 1$ for $i \in [n]$ and $Y = Y_1 + \ldots + Y_n$. Then $X_i = (Y_i + 1)/2$ and $X = Y/2 + n/2 = Y/2 + \mathbb{E}(X)$. So $\mathbb{P}(Y_i = -1) = \mathbb{P}(Y_i = 1) = 1/2$, and the random variables $Y_1, \ldots, Y_n$ are independent. Then, by Theorem 5.1,

$$\mathbb{P}(X \geq \mathbb{E}(X) + a) = \mathbb{P}\big(Y/2 + \mathbb{E}(X) \geq \mathbb{E}(X) + a\big) = \mathbb{P}(Y \geq 2a) \leq e^{-2a^2/n}.$$

This proves the first inequality; the second can be proved similarly. □

Theorem 5.1 has many extensions, allowing for more flexibility in the distributions of the X_i's. Here is a pretty general version (which we will not prove, but which can be proved similarly).

Theorem 5.4. *Let $X_1, \ldots, X_n$ be independent random variables with values in $\{0, 1\}$, and write $X = X_1 + \ldots + X_n$. Then, for every $a \geq 0$,*

$$\mathbb{P}\left(X - \mathbb{E}(X) > a\right) \leq e^{-2a^2/n}, \qquad \mathbb{P}\left(X - \mathbb{E}(X) < -a\right) \leq e^{-2a^2/n}.$$

Here is another useful version (which we will not prove).

Theorem 5.5. *Let $X_1, \ldots, X_n$ be independent random variables taking values in $\{0, 1\}$. Then, for $\delta > 0$,*

$$\mathbb{P}(X < (1-\delta)\mathbb{E}(X)) \leq e^{-\delta^2 \mathbb{E}(X)/2}, \qquad \mathbb{P}(X > (1+\delta)\mathbb{E}(X)) \leq e^{(-\delta^2 + \delta^3)\mathbb{E}(X)/2}.$$

In particular, if $0 \leq \delta \leq 1/3$,

$$\mathbb{P}(X > (1+\delta)\mathbb{E}(X)) \leq e^{-\delta^2 \mathbb{E}(X)/3}.$$

Remark 5.6. Let us compare the last two theorems. Suppose that $\mathbb{E}(X) = pn$ (for some $p \in [0, 1]$). The Theorem 5.4 gives

$$\mathbb{P}\left(X \leq (1-\delta)\mathbb{E}(X)\right) \leq e^{-\frac{2(\delta \mathbb{E}(X))^2}{n}} = e^{-2\delta^2 p^2 n},$$

and Theorem 5.5 gives

$$\mathbb{P}\left(X \leq (1-\delta)\mathbb{E}(X)\right) \leq e^{-\delta^2 pn}.$$

If p is small, then the latter estimate is stronger.

5.2 Hajós's conjecture

Definition 5.7 (Subdivision)**.** A *subdivision* of a graph H, is a graph obtained from H by replacing each edge e in H by a path P_e, whose interior vertices are new, and the interiors of P_e for $e \in E(H)$ are pairwise disjoint. Visually, a subdivision of H is obtained by adding some new vertices on edges of H (see Section 5.2).

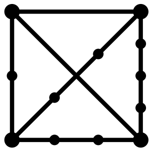


Figure 13: A subdivision of K_4

Definition 5.8. A proper colouring of G is a colouring of its vertices so that no two adjacent vertices share a colour (see Figure 14). The *chromatic number* of a graph G, denoted $\chi(G)$, is the minimum number of colours in a proper colouring of G.

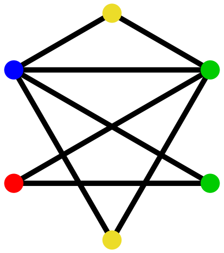


Figure 14: A proper colouring

A conjecture of Hajós (1961) asserts that if G has chromatic number k, then it has a subdivision of K_k. We will show that $G(n, 1/2)$ provides a counter example to this conjecture, with high probability.

Theorem 5.9. *With high probability $G(n, 1/2)$ has chromatic number at least $\frac{n}{3 \log_2 n}$ and has no subdivision of K_k for $k \geq 10\sqrt{n}$.*

Proof. Write $G = G(n, 1/2)$. First, recall that we showed that, with high probability, the clique number of G is $2 \log_2 n + O(\log \log n)$ (see Corollary 4.18). By symmetry, this show that, with high probability, the independence number of G, denoted $\alpha(G)$, is $2 \log_2 n + O(\log \log n)$. In particular, it is at most $3 \log_2 n$ for large enough n.

Note that $\chi(H) \geq \frac{|V(H)|}{\alpha(H)}$ for every graph H. Indeed, otherwise there is a proper colouring with fewer than $\frac{|V(H)|}{\alpha(H)}$ colours, showing that there is a colour class with more than $\alpha(H)$ edges. But each colour class is an independent set, so we found an independent set of size larger than $\alpha(H)$, a contradiction. It follows that, with high probability, $\chi(G) \geq \frac{n}{3 \log_2 n}$.

Claim 5.10. *With high probability, for every $m \geq 1000 \log n$, every set of m vertices in G has at least $\frac{1}{3}\binom{m}{2}$ non-edges.*

Proof. Consider a set S of m vertices, with $m \geq 1000 \log n$, and let X be the number of non-edges in S. Notice that X is a sum of $\binom{m}{2}$ independent random variable, each being 0 or 1 with probability

1/2. It follows that $\mathbb{E}(X) = \frac{1}{2}\binom{m}{2}$. Thus, by Theorem 5.5,

$$\mathbb{P}\left(X \leq \frac{1}{3}\binom{m}{2}\right) = \mathbb{P}\left(X \leq \left(1 - \frac{1}{6}\right) \cdot \mathbb{E}(X)\right)$$
$$\leq \exp\left(-\frac{1}{72} \cdot \mathbb{E}(X)\right) = \exp\left(-\frac{1}{144} \cdot \binom{m}{2}\right) \leq \exp\left(-2m \log n\right) = n^{-2m}.$$

using $m \geq 1000 \log n$.

Taking a union bound over all sets of size at least $1000 \log n$, we get that the probability that there is a set of m vertices, with $m \geq 1000 \log n$, that has fewer than $\frac{1}{3}\binom{m}{2}$ non-edges, is at most

$$\sum_{1000 \log n \leq m \leq n} \binom{n}{m} n^{-2m} \leq \sum_{m} n^{-m} \leq n \cdot n^{-1000 \log n} \to 0.$$

Here we used the inequality $\binom{n}{m} \leq n^m$ for the first inequality. This proves the claim. □

Suppose that the conclusion of the above claim holds for G, namely that every set of m vertices, with $m \geq 1000 \log n$, has at least $\frac{1}{3}\binom{m}{2}$ non-edges. We will show that there is no subdivision of K_k for $k = 10\sqrt{n}$. Then there is a set S of k vertices in G, and paths P_{xy} for every (unordered) pair xy of vertices of S, such that P_{xy} has ends x and y, and the interiors of these paths are in $V(G) - S$ and are pairwise vertex disjoint. In particular, at most $n - k \leq n$ of these paths have length more than 1 (otherwise, two such paths would share an interior vertex, a contradiction). So, at least $\binom{k}{2} - n$ of the paths P_{xy} are, in fact, the edge xy. In particular, there are at least $\binom{k}{2} - n$ edges in S, i.e. there are at most n non-edges in S. But

$$\frac{1}{3}\binom{k}{2} \geq \frac{k^2}{12} > n,$$

a contradiction to the assumption that S has at least $\frac{1}{3}\binom{k}{2}$ non-edges. So, indeed, with high probability, G has no subdivision of K_k for $k \geq 10\sqrt{n}$.

Altogether, we showed that each of the following holds with high probability: $\chi(G) \geq \frac{n}{3\log_2 n}$; and G has no subdivision of K_k for $k \geq 10\sqrt{n}$. By the union bound, both assertions hold simultaneously, as required for the theorem. □

5.3 Consistent edges in tournaments

Recall that a tournament is an oriented graph where for every two vertices x and y, exactly one of the pairs xy and yx is an edge. Given a tournament T and a permutation π of its vertices, an edge xy is *consistent* with π if x appears before y in π.

A *random tournament* on n vertices is the tournament on vertex set $[n]$, where for any two vertices x, y, one of xy and yx is chosen to be a directed edge, randomly and independently.

Lemma 5.11. *Let T be a random tournament on n vertices. Then, with high probability, for every permutation π of $V(T)$, at most $\frac{1}{2}\binom{n}{2} + O(n^{3/2}\sqrt{\log n})$ edges of T are consistent with π.*

Proof. For a permutation π of $V(T)$, let A_π be the event

$$\left\{\text{there are more than } \frac{1}{2}\binom{n}{2} + n^{3/2}\sqrt{\log n} \text{ edges consistent with } \pi\right\}.$$

We show that $\mathbb{P}(A_\pi)$ is small for every permutation π.

Fix π, define X_{xy} to be the indicator random variable for the event

$$\{\text{the edge with ends } x \text{ and } y \text{ is consistent with } \pi\},$$

and write $X = \sum_{x,y} X_{xy}$, where x, y run over all (unordered) pairs of vertices. Then X counts the number of edges consistent with π. Notice that $A_\pi = \{X > \frac{1}{2}\binom{n}{2} + n^{3/2}\sqrt{\log n}\}$. Thus, by independence of the X_{xy}'s, and by Corollary 5.3,

$$\mathbb{P}(A_\pi) = \mathbb{P}\left(X > \frac{1}{2}\binom{n}{2} + n^{3/2}\sqrt{\log n}\right) \leq \exp\left(-\frac{2n^3\log n}{\binom{n}{2}}\right) \leq \exp\left(-n\log n\right) = n^{-n}.$$

(Here we used $\binom{n}{2} \leq n^2$ for the second inequality.) Taking a union bound, we get

$$\begin{aligned}&\mathbb{P}\left(\text{there is a permutation with more than } \frac{1}{2}\binom{n}{2} + n^{3/2}\sqrt{\log n} \text{ consistent edges}\right)\\ &\quad = \mathbb{P}\left(\bigcup_\pi A_\pi\right) \leq \sum_\pi \mathbb{P}(A_\pi) \leq n! \cdot n^{-n} \leq n^{n-1}\cdot n^{-n} = \frac{1}{n} \to 0.\end{aligned}$$

Here for the penultimate inequality we used the crude bound $n! \leq n^{n-1}$. This shows that, with high probability, there is no permutation which is consistent with more than $\frac{1}{2}\binom{n}{2} + n^{3/2}\sqrt{\log n}$ edges. □

In fact, Lemma 5.11 can be improved a bit.

Theorem 5.12 (de la Vega, 1983)**.** *Let T be a random tournament on n vertices. Then, with high probability, for every permutation π of $V(T)$, at most $\frac{1}{2}\binom{n}{2} + O(n^{3/2})$ edges are consistent with π.*

Proof. For simplicity, we assume that $n = 2^k$. Write $V = V(T)$. For $i \in [k]$, define

$$s_i = 2^{2(k-i)+i-1} \qquad\qquad t_i = n^{3/2}2^{-i/2}\sqrt{i}. \tag{19}$$

For $i \in [k]$, let A_i be the event that there is an equipartition $\{V_1, \ldots, V_{2^i}\}$ of V such that there are more than $\frac{1}{2}s_i + t_i$ edges xy with $x \in V_{2j-1}$ and $y \in V_{2j}$ for some $j \in [2^{i-1}]$, where an *equipartition* is a partition into equal parts. Notice that s_i is the number of unordered pairs xy such that $x \in V_{2j-1}$ and $y \in V_{2j}$ for some j.

For example, A_1 is the event that there is an equipartition $\{V_1, V_2\}$ of V, with more than $\frac{1}{2}\left(\frac{n}{2}\right)^2 + \frac{1}{2}n^{3/2}$ edges directed from V_1 to V_2.

Claim 5.13. *With high probability, none of the events $A_1, \ldots, A_k$ hold.*

Proof of Claim 5.13. For $i \in [k]$, let P_i be the collection of all equi-partitions $\{V_1, \ldots, V_{2^i}\}$ of V into 2^i parts. Then $|\mathsf{P}_i| \le 2^{in}$, because each of n vertices has at most 2^i choices of a part. Fix $\mathcal{P} = (V_1, \ldots, V_{2^i}) \in \mathsf{P}_i$, and let $A_{\mathcal{P}}$ be the event: there are more than $\frac{1}{2}s_i + t_i$ edges xy with $x \in V_j$ and $y \in V_{j+1}$, for some j. For a pair (x, y) with $x \in V_{i,j}$ and $y \in V_{i,j+1}$, for some j, let X_{xy} be the indicator random variable for the event $\{xy \text{ is an edge}\}$, and set $X = \sum_{xy} X_{xy}$ for all such pairs (x, y). Then, as X is a sum of s_i independent random variables,

$$\mathbb{P}(A_{\mathcal{P}}) = \mathbb{P}\left(X \ge \frac{s_i}{2} + t_i\right) \le \exp\left(-\frac{2t_i^2}{s_i}\right)$$
$$= \exp\left(-\frac{2n^3 2^{-i} i}{2^{2(k-i)+i-1}}\right) \le \exp\left(-\frac{2n^3 2^{-i} i}{n^2 2^{-i}}\right) = e^{-2in}.$$

It follows from a union bound that

$$\mathbb{P}(A_i) = \mathbb{P}\left(\bigcup_{\mathcal{P} \in \mathsf{P}_i} A_{\mathcal{P}}\right) \le \sum_{\mathcal{P} \in \mathsf{P}} \mathbb{P}(A_{\mathcal{P}}) \le 2^{in} e^{-2in} \le e^{-in}.$$

Another union bound gives

$$\mathbb{P}\left(\bigcup_{i \in [k]} A_i\right) \le \sum_{i \in [k]} \mathbb{P}(A_i) \le \sum_{i \in [k]} e^{-in} \le \sum_{i \ge 1} e^{-in} = \frac{e^{-n}}{1 - e^{-n}} \le \frac{1}{2}e^{-n}.$$

So, with high probability, all of $A_1, \ldots, A_k$ hold. □

Fix an outcome of T where none of the events $A_1, \ldots, A_k$ hold, and fix a permutation π. We will show that the number of edges consistent with π is $\frac{1}{2}\binom{n}{2} + O\left(n^{3/2}\right)$.

For $i \in [0, k]$ we define an equipartition $\mathcal{P}_i = \{V_{i,1}, \ldots, V_{i,2^i}\}$ of V as follows: write $\pi = (v_1, \ldots, v_n)$, and take $V_{i,j} = \{v_{(j-1)2^{k-i}+1}, \ldots, v_{j2^{k-i}}\}$ for $j \in [2^i]$. For example, $\mathcal{P}_0$ partitions V into just one set, and $\mathcal{P}_k$ partitions V into singletons.

For $i \in [k]$, let E_i be the set of ordered pairs xy such that there exists $j \in [2^{i-1}]$ such that $x \in V_{i,2j-1}$ and $y \in V_{i,2j}$.

Claim 5.14. *For every ordered pair xy where x appears before y in π, there is a unique index i such that $xy \in E_i$.*

Proof. Write $x = v_a$ and $y = v_b$, so $a < b$. Let i be maximal such that x and y are in the same set in $\mathcal{P}_i$ (notice that such i exists, because x and y are in same set in $\mathcal{P}_0$), and let j be such that $x, y \in V_{i,j}$. Notice that $V_{i+1,2j-1}, V_{i+1,2j}$ partition $V_{i,j}$, with the elements in $V_{i+1,2j-1}$ preceding those of $V_{i+1,2j}$ in π. Thus, by choice of i and because $a < b$, we have $x \in V_{i+1,2j-1}$ and $y \in V_{i+1,2j}$, showing that $xy \in E_{i+1}$, as claimed.

For uniqueness, notice that

$$\sum_{i \in [k]} |E_i| = \sum_{i \in [k]} s_i = \sum_{i \in [k]} 2^{2(k-i)+i-1} = 2^{2k-2} \sum_{i \in [k]} 2^{-(i-1)} = \frac{n^2}{4} \frac{1-2^{-k}}{1-1/2} = \frac{n^2}{4} \cdot \frac{1-1/n}{1/2} = \binom{n}{2},$$

So, by the first paragraph, the union of the E_i's is the set $\{v_a v_b : a < b\}$, a set of size $\binom{n}{2}$. By the equality $\sum_i |E_i| = \binom{n}{2}$, this is a disjoint union, proving uniqueness. □

Let E_i' be the set of ordered pairs $xy \in E_i$ which are edges in T. By Claim 5.14, $\{E_1', \ldots, E_k'\}$ is a partition of the edges in T that are consistent with π. By the assumption on T, we have $|E_i'| \leq \frac{1}{2}|E_i| + t_i$. Thus, the number of edges consistent with π is at most

$$\begin{aligned}\sum_{i \in [k]} |E_i'| \leq \sum_{i \in [k]} \left(\frac{1}{2}|E_i| + t_i\right) &= \frac{1}{2}\binom{n}{2} + \sum_{i \in [k]} n^{3/2} 2^{-i/2} \sqrt{i} \\ &\leq \frac{1}{2}\binom{n}{2} + n^{3/2} \sum_{i \geq 1} 2^{-i/2}\sqrt{i} \leq \frac{1}{2}\binom{n}{2} + cn^{3/2},\end{aligned}$$

where $c = \sum_{i \geq 1} 2^{-i/2}\sqrt{i}$ (notice that the series converges). we showed that at most $\frac{1}{2}\binom{n}{2} + cn^{3/2}$ edges are consistent with π, for every permutation π, under the assumption that the conclusion of Claim 5.13 holds. This proves the theorem, since the conclusion of the claim holds with high probability. □

6 The local lemma

In the last few sections we used concentrations inequalities (Chebyshev and Chernoff) to prove that an outcome holds with positive probability. In fact, it was often the case that the same methods actually allowed us to prove that it holds with high probability.

In contrast, here is another way of showing that a certain outcome holds with positive probability: suppose that $A_1, \ldots, A_n$ are independent events, each holding with positive probability. Then, by independence, $\mathbb{P}(A_1 \cap \ldots \cap A_n) = \prod_{i \in [n]} \mathbb{P}(A_i) > 0$. In particular, the event $A_1 \cap \ldots \cap A_n$ holds with positive, but possibly small, probability.

This is, however, a very specific situation. In this section we will see a useful generalisation of this situation, namely when we have many events, with 'few dependencies'.

Definition 6.1 (Mutual independence)**.** We say that an event A is *mutually independent* of a collection of events $B_1, \ldots, B_n$, if for every choice of events $C_i \in \{B_i, B_i^{\complement}\}$, the events A and $\bigcap_{i \in [n]} C_i$ are independent.

Theorem 6.2 (Lovász, 1975). *Let $A_1, \ldots, A_n$ be events in a probability space. Suppose that $p \in [0,1)$ and d is a non-negative integer, such that: $\mathbb{P}(A_i) \leq p$ and A_i is mutually independent of a collection of all but at most d other events A_j, for $i \in [n]$; and $ep(d+1) \leq 1$. Then*

$$\mathbb{P}\left(\bigcap_{i\in[n]} A_i^C\right) > 0.$$

Remark 6.3. When $d = 0$ the independence condition implies that the events $A_1, \ldots, A_n$ are independent. Indeed, if say A, B, C are three events any one of which is mutually independent of the other two, then we get: $\mathbb{P}(A\cap B) = \mathbb{P}(A\cap B\cap C) + \mathbb{P}(A\cap B\cap C^{\mathsf{C}}) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) + \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C^{\mathsf{C}}) = \mathbb{P}(A)\mathbb{P}(B)$, showing that A and B are independent. Similarly, we can get that if A_i is mutually independent of $\{A_j : j \in [n] - \{i\}\}$ then the probability of the intersection of any subfamily of A_j's is the product of the probabilities in the family, showing independence of the family. We thus get that $\mathbb{P}\left(\cap_{i\in[n]}(A_i)^{\mathsf{C}}\right) = \prod_{i\in[n]} \mathbb{P}(A_i^{\mathsf{C}}) > 0$, assuming that $\mathbb{P}(A_i) < 1$.

We will prove the theorem later in the section. Before that, let us see some applications.

6.1 Ramsey numbers

Here is an application of the local lemma to Ramsey numbers, giving a small improvement over Theorem 3.1.

Theorem 6.4. *If $e \cdot \binom{k}{2}\binom{n-2}{k-2} \cdot 2^{1-\binom{k}{2}} < 1$, then $r(k,k) > n$.*

Proof. Colour the edges of K_n red and blue, randomly and independently. For a set S of k vertices, let A_S be the event that S is monochromatic. Then $\mathbb{P}(A_s) = 2^{1-\binom{k}{2}}$. Notice that A_S is mutually independent of the set $\{A_T : T \text{ is a set of } k \text{ vertices with } |S \cap T| \leq 1\}$, a set which contains all but at most $\binom{k}{2}\binom{n}{k-2} - 1$ sets of k vertices different from S. The local lemma, applied with $p = 2^{1-\binom{k}{2}}$ and $d = \binom{k}{2}\binom{n-2}{k-2} - 1$, along with the assumption, proves the proposition. □

Corollary 6.5. *For every $\varepsilon > 0$ there exists k_0 such that if $k \geq k_0$ then*

$$r(k,k) \geq (1-\varepsilon) \cdot \frac{\sqrt{2}}{e} \cdot k \cdot 2^{k/2}.$$

Proof. Fix $\varepsilon > 0$ and suppose that $n \leq (1-\varepsilon) \cdot \frac{\sqrt{2}}{e} \cdot k \cdot 2^{k/2}$. Then

$$\begin{aligned} e \cdot \binom{k}{2}\binom{n-2}{k-2} \cdot 2^{1-\binom{k}{2}} &\leq ek^2 \cdot 2^{-\frac{1}{2}(k+1)(k-2)} \left(\frac{e(n-2)}{k-2}\right)^{k-2} \\ &\leq ek^2 \cdot \left(\frac{(1+\varepsilon)\cdot en}{k \cdot 2^{(k+1)/2}}\right)^{k-2} \\ &\leq ek^2(1-\varepsilon^2)^{k-2}. \end{aligned}$$

(For the penultimate inequality we used $(1+\varepsilon)(k-2) \geq k$, which holds for large enough k.) This expression tends to 0 as k tends to infinity, so for large enough k it is at most 1. Thus, by Theorem 6.4, we have $r(k,k) \geq n$ for large enough k. By choice of n, this gives $r(k,k) \geq (1-\varepsilon)\frac{\sqrt{2}}{e}k2^{k/2}$, as claimed. □

Remark 6.6. This improve the previous lower bound for $r(k,k)$ that we obtained in Theorem 3.1 by about a factor of 2.

6.2 Colouring hypergraphs

Theorem 6.7. *Let $r \geq 2$ and suppose that H is an r-uniform hypergraph each of its edges intersects at most d other edges. If $e(d+1) \leq 2^{r-1}$ then H is 2-colourable.*

Proof. We wish to show that the vertices of H can be red-blue coloured so that no edge is monochromatic. As usual, consider a random red-blue colouring of H (namely, each vertex is coloured randomly and independently). For an edge e, let A_e be the event that e is monochromatic. We would like to show that $\mathbb{P}\left(\bigcap_{e\in E(H)} A_e^{\mathsf{c}}\right) > 0$, i.e. there is a colouring where no edge is monochromatic. To do so, we apply the local lemma (Theorem 6.2). Note that $\mathbb{P}(A_e) = 2 \cdot 2^{-r} = 2^{-(r-1)}$, so we set $p = 2^{-(r-1)}$. By assumption, each edge e touches at most d other edges, so A_e is mutually independent of a collection of all but at most d other events A_f, and $ep(d+1) \leq 1$. Thus, by the local lemma, we have $\mathbb{P}\left(\bigcap_e A_e^{\mathsf{c}}\right) > 0$, as desired. □

Remark 6.8. Notice that Theorem 6.4 is a corollary of Theorem 6.7. Indeed, take H to be the hypergraph with $V(H)$ being the collection of edges of K_n, and $E(H)$ the collection of edge sets of copies of K_k in K_n.

Corollary 6.9. *Let $r \geq 10$ and suppose that H is an r-uniform hypergraph which is r-regular, namely each of its vertices is in exactly r edges. Then H is 2-colourable.*

Proof. Notice that each edge in H touches at most r^2 other edges. Putting $d = r^2$, by Theorem 6.7, it suffices to show that $e(r^2+1) \leq 2^{r-1}$, which indeed holds for $r \geq 10$. □

6.3 Colouring real numbers

Theorem 6.10. *Let n and k be two positive integers satisfying*

$$e(n(n-1)+1) \cdot k\left(1-\frac{1}{k}\right)^n \leq 1.$$

Then for every set S of n real numbers, there is a colouring of $\mathbb{R}$ with k colours such that every translation of S (namely a set of form $x+S$) has elements of all colours.

Proof. Fix $S \subseteq \mathbb{R}$ of size n. We first prove: for every *finite* $R \subseteq \mathbb{R}$, there is a k-colouring of R with no monochromatic translations of S. To see this, fix a finite $R \subseteq \mathbb{R}$, and colour each vertex with a colour from $[k]$, chosen randomly and independently. For a translation $T = x + S$ of S which is contained in R, let A_T be the event that T is monochromatic. Then $\mathbb{P}(A_T) \le k\left(1 - \frac{1}{k}\right)^n$. Notice that A_T is mutually independent of the set of translations of S contained in R which are disjoint of T. As there are at most $n(n-1)$ translations of S distinct from T that intersect T (each such translation is determined by a choice of elements $s \in S$ and $t \in T$, such that $t \neq x + s$, and taking the unique translation of S where s is mapped to t). In particular, A_T is mutually independent of all but at most $n(n-1)$ other events $A_{T'}$. Apply the local lemma (Theorem 6.2) with $p = k\left(1 - \frac{1}{k}\right)^n$ and $d = n(n-1)$ to deduce that there is a colouring where all events $A_T^{\complement}$ holds, i.e. where there are no monochromatic translations of S.

We now prove the same for all countable sets $R \subseteq \mathbb{R}$. To do this, fix $R \subseteq \mathbb{R}$, which is countable and infinite. Enumerate it as $(r_i)_{i \ge 1}$, and define $R_i = \{r_1, \ldots, r_i\}$. By the above, for every i there is a colouring $f_i : R_i \to [k]$ where no translation of S is monochromatic. For $i < j$, we say that a colouring f of R_j *extends* a colouring g of R_i if the restriction of f to R_i agrees with g.

Claim 6.11. *There is a sequence $(g_i)_{i \ge 1}$ satisfying: g_i is a colouring $g_i : R_i \to [k]$; g_i extends g_{i-1} for every $i \ge 2$; and for infinitely many $j \ge i$, the colouring f_j extends g_i.*

Proof. We prove by induction that there is a sequence $g_1, \ldots, g_i$ as in the claim, for $i \ge 0$, noting that there is nothing to prove for $i = 0$. Suppose that we have defined a sequence $g_1, \ldots, g_i$ satisfying the above properties. Let X be an infinite set of indices $j > i$ such that f_j extends g_i; such a set exists by the properties of the sequence. By the pigeon hole principle, there exists $c \in [k]$ such that $f_j(s_{i+1}) = c$ for infinitely many $j \in X$. Let $g_{i+1} : S_{i+1} \to [k]$ defined by

$$g_{i+1}(r_j) = \begin{cases} g_i(r_j) & j \in [i] \\ c & j = i + 1. \end{cases}$$

It is easy to check that g_{i+1} satisfies the required properties. □

Let $(g_i)_{i \ge 1}$ be a sequence as guaranteed by the above claim. Now consider the colouring $g : R \to [k]$, obtained by taking $g(t_i) = g_i(t_i)$. We claim that this is a k-colouring of R with no monochromatic translations of S. Indeed, consider a translation T of S which is contained in R, and let i be such that $T \subseteq R_i$. Notice that the restriction of g to R_i is g_i, and recall that f_j extends g_i for infinitely many $j > i$, showing that g_i has no monochromatic translations of S. In particular, T is not monochromatic in g_i and thus in g.

Write $S = \{s_1, \ldots, s_{|S|}\}$ and $\mathrm{Span}_{\mathbb{Z}}(S) = \{a_1 s_1 + \ldots + a_{|S|} s_{|S|} : a_1, \ldots, a_{|S|} \in \mathbb{Z}\}$. Finally, we conclude that $\mathbb{R}$ can be k-coloured with no monochromatic translations of S. Define a binary relation $\sim$ as follows: for $x, y \in \mathbb{R}$, we write $x \sim y$ if $x - y \in \mathrm{Span}_{\mathbb{Z}}(S)$. It is easy to check that $\sim$ is an equivalence relations, and that every translation of S is contained in an equivalence class. Thus it suffices to show that each equivalence class can be k-coloured with no monochromatic translation of S. This easily follows by the previous paragraph, by observing that each equivalence class is a translation of $\mathrm{Span}_{\mathbb{Z}}(S)$. □

Remark 6.12. For the last part of the proof, we implicitly used the axiom of choice, which asserts that any product of non-empty sets is non-empty. This allows us to simultaneously choose an appropriate colouring for each equivalence class.

Remark 6.13. For k large, it suffices to take $n = 4k \log k$. Indeed, then

$$e(n(n-1)+1)k\left(1-\frac{1}{k}\right)^n \le en^2 k \exp\left(-\frac{n}{k}\right) = 16ek^3(\log k)^3 \exp(-4\log k) = \frac{16e(\log k)^2}{k} \le 1,$$

where the latter holds for large k (as the expression tends to 0).

6.4 Proof of the local lemma

Proof of Theorem 6.2. We will prove the following claim.

Claim 6.14. *For every subset $S \subseteq [n]$ and every $i \in [n] \setminus S$,*

$$\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j\in S} A_j^{\mathsf{C}}\right) \le \frac{1}{d+1}. \tag{20}$$

Before proving Claim 6.14, we prove that it implies the theorem. To see this, note that for every $\ell \in [n-1]$,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i\in[\ell]} A_i^{\mathsf{C}} \,\middle|\, \bigcap_{i\in[\ell+1,n]} A_i^{\mathsf{C}}\right) &= \mathbb{P}\left(\bigcap_{i\in[\ell-1]} A_i^{\mathsf{C}} \,\middle|\, \bigcap_{i\in[\ell,n]} A_i^{\mathsf{C}}\right)\cdot \mathbb{P}\left(A_\ell^{\mathsf{C}} \,\middle|\, \bigcap_{i\in[\ell+1,n]} A_i^{\mathsf{C}}\right)\\ &= \mathbb{P}\left(\bigcap_{i\in[\ell-1]} A_i^{\mathsf{C}} \,\middle|\, \bigcap_{i\in[\ell,n]} A_i^{\mathsf{C}}\right)\cdot \left(1-\mathbb{P}\left(A_\ell \,\middle|\, \bigcap_{i\in[\ell+1,n]} A_i^{\mathsf{C}}\right)\right)\\ &\ge \mathbb{P}\left(\bigcap_{i\in[\ell-1]} A_i^{\mathsf{C}} \,\middle|\, \bigcap_{i\in[\ell,n]} A_i^{\mathsf{C}}\right)\cdot \left(1-\frac{1}{d+1}\right).\end{aligned}$$

using $\mathbb{P}(A\cap B \,|\, C) = \mathbb{P}(A \,|\, B\cap C)\cdot \mathbb{P}(B \,|\, C)$, which holds for any events A, B, C. Iterating this,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i\in[n]} A_i^{\mathsf{C}}\right) &= \mathbb{P}\left(\bigcap_{i\in[n-1]} A_i^{\mathsf{C}} \,\middle|\, A_n^{\mathsf{C}}\right)\cdot\left(1-\frac{1}{d+1}\right)\\ &\ge \mathbb{P}\left(\bigcap_{i\in[n-2]} A_i^{\mathsf{C}} \,\middle|\, A_{n-1}^{\mathsf{C}}\cap A_n^{\mathsf{C}}\right)\cdot\left(1-\frac{1}{d+1}\right)^2 \ge \ldots \ge \left(1-\frac{1}{d+1}\right)^n > 0,\end{aligned}$$

as required.

Proof of Claim 6.14. The proof is by induction on $|S|$. Notice that when $|S| = 0$ (i.e. $S = \emptyset$), this amounts to showing that $\mathbb{P}(A_i) \leq \frac{1}{d+1}$ for every $i \in [n]$, which holds by assumption due to $p \leq \frac{1}{e(d+1)} < \frac{1}{d+1}$.

Now let $S \subseteq [n]$ and $i \in [n] \setminus S$, and suppose that (20) holds for all subsets $S' \subseteq [n]$ with $|S'| < |S|$ (and all $i \in [n] \setminus S'$). Let $T \subseteq S$ be a set of size at most d such that A_i is mutually independent of $\{A_j : j \in S \setminus T\}$, and write $R = S \setminus T$. Then

$$\mathbb{P}\left(A_i \,\middle|\, \bigcap_{j \in S} A_j^{\mathsf{c}}\right) = \frac{\mathbb{P}\left(\bigcap_{j\in T} A_j^{\mathsf{c}} \cap A_i \,\middle|\, \bigcap_{j \in R} A_j^{\mathsf{c}}\right)}{\mathbb{P}\left(\bigcap_{j\in T} A_j^{\mathsf{c}} \,\middle|\, \bigcap_{j \in R} A_j^{\mathsf{c}}\right)}. \tag{21}$$

We now estimate the numerator in (21).

$$\mathbb{P}\left(\bigcap_{j\in T} A_j^{\mathsf{c}} \cap A_i \,\middle|\, \bigcap_{j \in R} A_j^{\mathsf{c}}\right) \leq \mathbb{P}\left(A_i \,\middle|\, \bigcap_{j \in R} A_j^{\mathsf{c}}\right) = \mathbb{P}(A_i) \leq p. \tag{22}$$

Here the equality follows from the independence of A_i with $\bigcap_{j\in R} A_j^{\mathsf{c}}$.

Next, we estimate the denominator. Write $T = \{i_1, \ldots, i_t\}$. For every $\ell \in [t]$ we have

$$\begin{aligned}
&\mathbb{P}\left(\bigcap_{j\in[\ell]} A_{i_j}^{\mathsf{c}} \,\middle|\, \bigcap_{j\in[\ell+1,t]} A_{i_j}^{\mathsf{c}} \cap \bigcap_{j\in R} A_j^{\mathsf{c}}\right)\\
&= \mathbb{P}\left(\bigcap_{j\in[\ell-1]} A_{i_j}^{\mathsf{c}} \,\middle|\, \bigcap_{j\in[\ell,t]} A_{i_j}^{\mathsf{c}} \cap \bigcap_{j\in R} A_j^{\mathsf{c}}\right) \cdot \mathbb{P}\left(A_{j_\ell}^{\mathsf{c}} \,\middle|\, \bigcap_{j\in[\ell+1,t]} A_{i_j}^{\mathsf{c}} \cap \bigcap_{j\in R} A_j^{\mathsf{c}}\right)\\
&= \mathbb{P}\left(\bigcap_{j\in[\ell-1]} A_{i_j}^{\mathsf{c}} \,\middle|\, \bigcap_{j\in[\ell,t]} A_{i_j}^{\mathsf{c}} \cap \bigcap_{j\in R} A_j^{\mathsf{c}}\right) \cdot \left(1 - \mathbb{P}\left(A_{j_\ell} \,\middle|\, \bigcap_{j\in[\ell+1,t]} A_{i_j}^{\mathsf{c}} \cap \bigcap_{j\in R} A_j^{\mathsf{c}}\right)\right)\\
&\geq \mathbb{P}\left(\bigcap_{j\in[\ell-1]} A_{i_j}^{\mathsf{c}} \,\middle|\, \bigcap_{j\in[\ell,t]} A_{i_j}^{\mathsf{c}} \cap \bigcap_{j\in R} A_j^{\mathsf{c}}\right) \cdot \left(1 - \frac{1}{d+1}\right),
\end{aligned}$$

where for the first equality we used $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid B \cap C) \cdot \mathbb{P}(B \cap C)$, which holds for any events A, B, C, and for the inequality we used the induction hypothesis, noting that the set $R \cup \{i_{\ell+1}, \ldots, i_t\}$

does not contain i_ℓ and thus has size less than $|S|$. Applying this iteratively, we get

$$\begin{aligned}\mathbb{P}\left(\bigcap_{j\in T} A_j^{\complement} \;\middle|\; \bigcap_{j\in R} A_j^{\complement}\right) &= \mathbb{P}\left(\bigcap_{j\in[t]} A_{i_j}^{\complement} \;\middle|\; \bigcap_{j\in R} A_j^{\complement}\right)\\ &\geq \mathbb{P}\left(\bigcap_{j\in[t-1]} A_{i_j}^{\complement} \;\middle|\; (A_{i_t})^{\complement}\cap\bigcap_{j\in R} A_j^{\complement}\right)\cdot\left(1-\frac{1}{d+1}\right)\\ &\geq \mathbb{P}\left(\bigcap_{j\in[t-2]} A_{i_j}^{\complement} \;\middle|\; (A_{i_{t-1}})^{\complement}\cap(A_{i_t})^{\complement}\cap\bigcap_{j\in R} A_j^{\complement}\right)\cdot\left(1-\frac{1}{d+1}\right)^2\\ \ldots &\geq \left(1-\frac{1}{d+1}\right)^t \geq \left(1-\frac{1}{d+1}\right)^d \geq e^{-1}.\end{aligned} \tag{23}$$

For the last inequality, we used that $\left(1+\frac{1}{d}\right)^d$ is an increasing sequences whose limit is e and $\left(1-\frac{1}{d+1}\right)^d = \left(1+\frac{1}{d}\right)^{-d}$.

Plugging in (22) and (23) into (21), we get

$$\mathbb{P}\left(A_i \;\middle|\; \bigcap_{j\in S} A_j^{\complement}\right) \leq ep \leq \frac{1}{d+1},$$

using $ep(d+1) < 1$, as required for (20), proving the claim. □

As explained above, the claim implies the theorem. □

6.5 Cycles in directed graphs

The following is the final application of the local lemma that we will see. The statement consider directed cycles in digraphs, which does not involve any colouring, unlike previous applications. Neverthelesss, as we shall see, the proof does introduce colours.

Theorem 6.15 (Alon–Linial, 1989)**.** *Let d, k be positive integers satisfying*

$$e(d(d+1)+1)\left(1-\frac{1}{k}\right)^d \leq 1.$$

Suppose that D is a d-regular digraph (meaning that each vertex has both in- and out-degree equal to d). Then D contains a directed cycle of length divisible by k.

Proof. We colour each vertex with a colour from $[k]$, chosen randomly and independently, denoting the resulting colouring by c. We will show that, with positive probability, every vertex c has an out-neighbour u such that $c(u) \equiv c(v) + 1 \pmod{k}$.

Before proving this, let us see why this would imply the result. Fix c satisfying the above property, and consider the subdigraph $D' \subseteq D$ obtained by keeping edges uv such that $c(u) \equiv c(v) + 1 \pmod{k}$. So D' has minimum out-degree at least 1, and thus it has a directed cycle $C = (v_1 \dots v_\ell)$. Indeed, let $u_1 \dots u_t$ be a longest directed path in D'. By maximality of the path and the minimum degree assumption on D', the vertex u_t has an out-neighbour among $\{u_1, \dots, u_{t-1}\}$; denote it by u_i. Then $(u_i \dots u_\ell)$ is a directed cycle in D'. By choice of D' we have $c(v_{i+1}) \equiv c(v_i) + 1 \pmod{k}$ for $i \in [\ell - 1]$ and $c(v_1) \equiv c(v_\ell) + 1 \pmod{k}$. Thus $c(v_1) \equiv c(v_\ell) + 1 \equiv \dots \equiv c(v_1) + \ell \pmod{k}$. This implies $\ell \equiv 0 \pmod{k}$; namely, C is a directed cycle whose length is divisible by k.

Define $N^+(v)$ to be the out-neighbourhood of the vertex v, and let $N^+[v]$ be the *closed* out-neighbourhood of v, namely the union $N^+(v) \cup \{v\}$. We now prove that c has the desired property with positive probability. To see this, define A_v to be the event that v does not have an out-neighbour u with $c(u) \equiv c(v) + 1 \pmod{k}$.

First notice that $\mathbb{P}(A_v) = \left(1 - \frac{1}{k}\right)^d$. Indeed, A_v fails to hold whenever none of the out-neighbours of v have colour $c(v) + 1 \pmod{k}$.

Next, notice that A_v only depends on the outcome of v in the closed out-neighbourhood $N^+[v]$. As such, A_v is mutually independent of the events A_u where $N^+[u] \cap N^+[v] = \emptyset$. Suppose that $N^+[u] \cap N^+[v] \neq \emptyset$ and $u \neq v$. Then one of the following holds: $u \in N^+(v)$; $v \in N^+(u)$; $N^+(u) \cap N^+(v) \neq \emptyset$. There are exactly d vertices u satisfying each of the first two properties (by regularity, and because the second is equivalent to $u \in N^-(v)$), and there are at most $d(d-1)$ vertices u satisfying the third property (it is equivalent to u being an in-neighbour of an out-neighbour of v which is not v itself; there are d ways of choosing an out-neighbour w of v and $d - 1$ ways of choosing an in-neighbour of w which is not u). This leaves at most $(d-1)d+d+d = d(d+1)$ options for u (each vertex in $N^+(v)$ contributes at most $d - 1$ in-neighbours other than v, v contributes d in-neighbours, and u could also be in $N^+(v)$).

By Theorem 6.2, applied with $p = \left(1 - \frac{1}{k}\right)^d$ and $d_{6.2} = d(d + 1)$, and the assumption regarding the relation between p and d, there is an outcome of c where none of the events A_v hold, as claimed. □

Remark 6.16. For large k, it suffices to take $d = 3k \log k$. Indeed, then

$$e(d(d+1)+1)\left(1 - \frac{1}{k}\right)^d \leq 2ed^2 \exp\left(-\frac{d}{k}\right) = 18ek^3(\log k)^2 \exp(-3\log k) = \frac{18e(\log k)^2}{k} \leq 1.$$

7 Concentration inequalities: McDiarmid's inequality

7.1 McDiarmid's inequality

Recall that Chernoff's bounds allowed us to upper bound the probability that a random variable X, which is the sum of independent random variables $X_1, \dots, X_n$, is far from its expectation. While this is a very useful tool, the assumption on X is quite specific. The following inequality, due to McDiarmid[4] proves a bound similar to Chernoff's, which applies to functions of independent random variables satisfying a 'Lipschitz property'. We will not proves this inequality.

Theorem 7.1 (McDiarmid's inequality, 1989)**.** *Let $X_1, \ldots, X_n$ be independent random variables, where X_i takes values in the set S_i. Let $c > 0$ and let $f : S_1 \times \ldots \times S_n \to \mathbb{R}$ be a function satisfying $|f(x) - f(x')| \leq c$ for every $x, x' \in S_1 \times \ldots \times S_n$ differing on at most one coordinate. Then*

$$\mathbb{P}\big(f(X_1, \ldots, X_n) \leq \mathbb{E}\left(f(X_1, \ldots, X_n) - t\big) \leq \exp\left(-\frac{2t^2}{c^2 n}\right)\right.$$
$$\mathbb{P}\big(f(X_1, \ldots, X_n) \geq \mathbb{E}\left(f(X_1, \ldots, X_n) + t\big) \leq \exp\left(-\frac{2t^2}{c^2 n}\right)\right..$$

Remark 7.2. Notice that by taking $S_i = \{0, 1\}$ and taking $f(s_1, \ldots, s_n) = s_1 + \ldots + s_n$ we recover a version of Chernoff's bound (Theorem 5.4).

Nevertheless, like in Remark 5.6, stronger variants of Chernoff's bound such as Theorem 5.5 yield significantly stronger bounds than Theorem 7.1 when $\mathbb{E}(X)$ is much smaller than n.

7.2 Random functions

To illustrate the power of Theorem 7.1, we start with an easy consequence of it.

Theorem 7.3. *Let f be a function from $[n]$ to $[n]$, chosen randomly among all such functions, and let X be the random variable counting the number of elements in $[n]$ that are not in the image of f. Then*

$$\mathbb{P}\left(\left|X - n\left(1 - \frac{1}{n}\right)^n\right| \geq \lambda\sqrt{n}\right) \leq 2e^{-2\lambda^2}. \tag{24}$$

In particular, with high probability, $X \approx \frac{n}{e}$.

Proof. First, we note that $\mathbb{E}(X) = n\left(1 - \frac{1}{n}\right)^n$. Indeed, this follows from linearity of expectation, and noting that the probability that i is not in the image of f is $\left(1 - \frac{1}{n}\right)^n$. Next, observe that f can be thought of as constructed from n independent random variables $X_1, \ldots, X_n$, each of which is uniformly distributed on $[n]$. Indeed, given $X_1, \ldots, X_n$ we assign $f(i) = X_i$. Notice also that X satisfies the Lipschitz property with constant 1, namely changing one coordinate in the sequence of outcomes of the X_i's changes X by at most 1 (the size of the set $\{X_1, \ldots, X_n\}$ can change by at most 1 by changing X_i: at most one element is removed, and at most one element is added). Thus, by Theorem 7.1, we get (24). The 'In particular' part easily follows from noting that $\left(1 - \frac{1}{n}\right)^n$ tends to e^{-1} as n tends to infinity, and taking, say, $\lambda = \log n$. □

7.3 Isoperimetric inequality

Recall that the *hypercube* Q_n of dimension n is the graph on vertices $\{0, 1\}^n$ whose edges are pairs of vertices that differ on exactly one coordinate (see Definition 3.14).

[4] You may have heard or may hear of 'Azuma–Hoeffding's inequality'; McDiarmid's inequality is a consequence of it.

The *distance* between two vertices $v, u \in V(Q_n)$, denoted $\mathrm{dist}(v, u)$, is the number of coordinates on which v and u differ (in fact, this is just the usual notion of distance in a graph, namely the minimum number of edges in a path from v to u). The distance between a set of vertices $U \subseteq V(Q_n)$ and a vertex $v \in V(Q_n)$, denoted $\mathrm{dist}(v, U)$, is the minimum of $\mathrm{dist}(v, u)$ over all $u \in U$.

The *ball* of radius r around a vertex v, denoted $B_r(v)$, is the set of vertices at distance at most r from v. For a set of vertices U we denote by $B_r(U)$ the set of vertices that are at distance at most r from U.

Theorem 7.4. *For every $\varepsilon > 0$ there is a constant λ such that for large enough n the following holds. If U is a set of at least $\varepsilon 2^n$ vertices in the hypercube Q_n then $|B_{\lambda\sqrt{n}}(U)| \geq (1-\varepsilon)2^n$.*

Remark 7.5. Theorem 7.4 can be thought of as a form of isoperimetric result. An classic isoperimetric problem is the following: among all shapes in $\mathbb{R}^2$ with a given area, which one minimises the boundary? It is well known that the answer is a disk.

One can define a notion of boundary in graphs. Given a graph G, the *vertex boundary* of a set of vertices U, denoted $\partial_v(U)$, is the set of vertices in G that have a neighbour in U but are not in U (so, for example, $\partial_v(V(G)) = \emptyset$). A theorem of Harper (1966) asserts that, in the hypercube, the Hamming ball of radius r, namely the set $B_r(v)$ for any vertex v, minimises the vertex boundary among all sets of vertices of the same size (standard proofs of this are combinatorial). One could use this theorem to prove Theorem 7.4, but we will see a proof using McDiarmid's inequality.

Proof of Theorem 7.4. Let $X_1, \ldots, X_n$ be independent random variables, each chosen uniformly from $\{0, 1\}$. Write $X = (X_1, \ldots, X_n)$, and let $D = \mathrm{dist}(X, U)$. Notice that D is 1-Lipschitz with respect to $(X_1, \ldots, X_n)$. Indeed, changing one coordinate of X would change the distance of X from any point in Q_n by at most 1, thus changing D by at most 1. Thus, by McDiarmid's inequality,

$$\mathbb{P}\left(|D - \mathbb{E}(D)| \geq \lambda\sqrt{n}\right) \leq 2\exp\left(-2\lambda^2\right) < \varepsilon,$$

where the last inequality holds for sufficiently large λ.

Notice that $\mathbb{P}(D = 0) = \frac{|U|}{2^n} \geq \varepsilon$, as X a uniformly random element of $V(Q_n)$. Thus, $\mathbb{E}(D) < \lambda\sqrt{n}$. Indeed, otherwise,

$$\varepsilon \leq \mathbb{P}(D = 0) \leq \mathbb{P}\big(\,|D - \mathbb{E}(D)| \geq \lambda\sqrt{n}\big) < \varepsilon,$$

a contradiction.

Hence,

$$\mathbb{P}\big(X \notin B_{2\lambda\sqrt{n}}(U)\big) = \mathbb{P}\big(D \geq 2\lambda\sqrt{n}\big) \leq \mathbb{P}\left(|D - \mathbb{E}(D)| \geq \lambda\sqrt{n}\right) \leq \varepsilon.$$

In particular, $\mathbb{P}(X \in B_{2\lambda\sqrt{n}}(U)) \geq 1 - \varepsilon$. Equivalently, $|B_{2\lambda\sqrt{n}}(U)| \geq (1-\varepsilon)2^n$, as required for the theorem (taking 2λ instead of λ). □

Remark 7.6. Notice that we did not need to know $\mathbb{E}(X)$ for the above proof. This is a common occurence when using McDiarmid's inequality.

7.4 The chromatic number of random graphs

In the remainder of these notes, we will be interested in the chromatic number of the random graph $G(n, p)$. First, we give a straightforward consequence of McDiarmid's inequality to prove that, with high probability, the the chromatic number of $G(n, p)$ is quite close to its expectation.

Theorem 7.7 (Shamir–Spencer, 1987)**.** *For every $\varepsilon > 0$ there exists $\lambda > 0$ such that the following holds. For every n and $p = p(n) \in (0, 1)$ there is an interval I of at most $\lambda\sqrt{n}$ integers such that $\chi(G(n, p)) \in I$ with probability at least $1 - \varepsilon$.*

Proof. Write $G = G(n, p)$ and $Y = \chi(G)$. We think of Y as a function of n independent random variables $X_1, \dots, X_n$, where X_i encodes the outcomes of the edges ij with $j \in [i-1]$. We claim that Y is 1-Lipschitz. Indeed, changing X_i amounts to adding and/or removing some edges incident to the vertex i, which we claim can change the chromatic number by at most 1. To see this, suppose that H_1, H_2 are two graphs that differ only on the edges touching a vertex v. Then

$$\chi(H_1) \leq \chi(H_1 - \{v\}) + 1 = \chi(H_2 - \{v\}) + 1 \leq \chi(H_2) + 1.$$

(for the second inequality, given a proper colouring of $H_1 - \{v\}$ with $\chi(H_1 - \{v\})$ colours, colour v with a new colour.) By symmetry, we also have $\chi(H_2) \leq \chi(H_1) + 1$, and thus $\chi(H_2) - 1 \leq \chi(H_1) \leq \chi(H_2) + 1$, showing $|\chi(H_1) - \chi(H_2)| \leq 1$.

Thus, by McDiarmid's inequality (Theorem 7.1),

$$\mathbb{P}\left(|Y - \mathbb{E}(Y)| \geq \lambda\sqrt{n}\right) \leq 2\exp(-2\lambda^2).$$

Taking λ to be large enough so that $2\exp(-2\lambda^2) \leq \varepsilon$ and taking the interval $I = [\mathbb{E}(Y) - \lambda\sqrt{n}, \mathbb{E}(Y) + \lambda\sqrt{n}]$, we get that I is an interval of length $2\lambda\sqrt{n}$ such that $\chi(G(n, p)) \in I$ with probability at least $1 - \varepsilon$, as required (again taking 2λ instead of λ). □

Remark 7.8. Again, this proof works without knowing (even approximately) the expected value of $\chi(G(n, p))$.

Remark 7.9. A more natural way of thinking of $G(n, p)$ as a function of indepedent random variables would be to have a separate random variable for pair ij with $i, j \in [n]$, encoding whether or not ij is an edge. This is known as the *edge exposure martingale* (though we will not define martingales in this module). This point of view is not helpful here because this would mean having $\binom{n}{2}$ random variables instead of just n, leading to a much worse bound on the probability of deviating from the mean. Instead, the above proof used the *vertex exposure martingale.* This name makes sense, as $X_1, \dots, X_i$ determine the subgraph of $G(n, p)$ induced on $[i]$.

7.5 The chromatic number of dense random graphs

The following results determines, asymptotically and with high probability, the chromatic number of $G(n, 1/2)$.

Theorem 7.10 (Bollobás, 1988). *With high probability, the chromatic number of $G(n, 1/2)$ is* $(1 + o(1))\frac{n}{2\log_2 n}$.

As in Section 4.5, we define $f : \mathbb{N} \to \mathbb{N}$ as follows

$$f(k) = \binom{n}{k} 2^{-\binom{k}{2}}.$$

Let $k_0 = k_0(n)$ be the largest k such that $f(k) \geq 1$ (this is a slightly different choice) and write $k_1 = k_1(n) = k_0 - 4$. Claim 4.17 and Claim 4.19 yield that $k_0 = (2 + o(1))\log_2 n$ and $\frac{f(k)}{f(k+1)} \geq n^{1/2}$ for $k \geq 1.99 \log_2 n$, implying $k_1 = (2 + o(1))\log_2 n$ and $f(k_1) \geq n^2$.

Lemma 7.11. *With probability at least* $1 - e^{n^2/(\log_2 n)^{13}}$*, the clique number of* $G(n, 1/2)$ *is at least* k_1.

Remark 7.12. By symmetry, the results are the same. This results is stronger than Section 4.5 because here we give a much stronger bound on the probability of failure. Section 4.5 is stronger in the sense that the bound on the clique number there is a bit stronger.

First, we show how the lemma implies Theorem 7.10.

Proof of Theorem 7.10 using Lemma 7.11. Write $m = m(n) = \frac{n}{(\log_2 n)^2}$ and let $k_2 = k_1(m)$. Then

$$k_2 = (2 + o(1))\log_2 m = (2 + o(1))\log_2 n,$$

and by Lemma 7.11, with probability at least $1 - \exp(-m^2/(\log_2 m)^{13}) \geq 1 - \exp(-n^2/(\log_2 n)^{18})$, the independence number of $G(m, 1/2)$ is at least k_2 (using the symmetry between $G(n, 1/2)$ and its complement, which allows us to deduce the analogue of the lemma for independent sets of size k_2). Thus, the probability that there is a set of m vertices in $G = G(n, 1/2)$ with independence number less than k_2 is at most

$$\binom{n}{m} \exp\left(-\frac{n^2}{(\log_2 n)^{18}}\right) \leq 2^{n - n^{3/2}} = o(1).$$

So, with high probability, every set of m vertices contains an independent set of size k_2. It follows that, with high probability, the chromatic number of $G(n, 1/2)$ is at most

$$\frac{n}{k_2} + m = (1 + o(1))\frac{n}{2\log_2 n}.$$

Indeed, a maximal collection of pairwise vertex disjoint independent sets of size at least k_2 covers all but at most m vertices and consists of at most $\frac{n}{k_2}$ sets; we can colour each such independent set with its own colour and colour each vertex not covered by the collection with a new unique colour. □

Proof of Lemma 7.11. Write $k = k_1(n)$ for convenience. Define Y to be the maximum number of pairwise edge-disjoint cliques of size k that can be found in $G = G(n, 1/2)$.

Claim 7.13. $\mathbb{E}(Y) \geq \frac{n^2}{k^6}$.

We will prove the claim next time. We now prove the lemma using the claim.

We think of Y as a function of $\binom{n}{2}$ independent random variables $(X_{ij} : 1 \leq i < j \leq n)$, where X_{ij} encodes whether or not ij is an edge. Notice that Y is 1-Lipschitz. Indeed, changing one variable X_{ij}, which amounts to adding or removing one edge, can add or remove at most one clique from a largest collection of pairwise edge-disjoint k-cliques. We may thus apply McDiarmid's inequality (Theorem 7.1), to deduce

$$\begin{aligned}\mathbb{P}(Y=0) \leq \mathbb{P}\big(|Y - \mathbb{E}(Y)| \geq \mathbb{E}(Y)\big) &\leq 2\exp\left(-\frac{2(\mathbb{E}(Y))^2}{\binom{n}{2}}\right)\\ &\leq 2\exp\left(-\frac{\left(\frac{n^2}{k^6}\right)^2}{n^2}\right) \leq 2\exp\left(-\frac{n^2}{k^{12}}\right) \leq 2\exp\left(-\frac{n^2}{(\log_2 n)^{13}}\right).\end{aligned}$$

This completes the proof of the lemma, as $Y=0$ exactly when G has no cliques of size k. □

Proof of Claim 7.13. Let X be the number of cliques of size k in G, and let Z be the number of ordered pairs of cliques of size k that intersect in at least two vertices. Write $\mu = \mathbb{E}(X)$ and $\nu = \mathbb{E}(Z)$. Then $\mu = f(k) \geq n^2$. We first show that

$$\frac{\nu}{\mu^2} \leq \frac{2k^5}{n^2}. \tag{25}$$

To see this, write $g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}2^{\binom{i}{2}}}{\binom{n}{k}}$. Then

$$\nu = \mathbb{E}(Z) = \sum_{i\in[2,k]} \binom{n}{k}\binom{k}{i}\binom{n-k}{k-i}2^{-2\binom{k}{2}+\binom{i}{2}} = \left(\binom{n}{k}2^{-\binom{k}{2}}\right)^2 \sum_{i\in[2,k]} g(i) = \mu^2 \sum_{i\in[2,k]} g(i).$$

By Claim 4.20 (albeit with slightly different value of k, which does not change the estimates), we have $g(i) \leq \max\{g(2), g(k)\}$. Hence, it suffices to show that $g(i) \leq \frac{2k^4}{n^2}$ for $i \in \{2, k\}$.

$$\begin{aligned} g(2) &= \frac{\binom{k}{2}\cdot\binom{n-k}{k-2}2^{\binom{2}{2}}}{\binom{n}{k}} \leq \frac{k^2\,\frac{(n-k)^{k-2}}{(k-2)!}}{\frac{(n-k)^k}{k!}} \leq \frac{k^4}{(n-k)^2} \leq \frac{2k^4}{n^2}.\\ g(k) &= \frac{1}{\binom{n}{k}2^{-\binom{k}{2}}} = \frac{1}{\mu} \leq \frac{1}{n^2} \leq \frac{2k^4}{n^2}.\end{aligned}$$

The penultimate inequality in the second line follows from the choice of k which implies $f(k) \geq n^2$. This completes the proof of (25).

Now, let q be a probability to be determined. Let $\mathcal{K}$ be a family of k-cliques (i.e. cliques of size k) in G, obtained by including each k-clique in G with probability q, indepedently. Let $\mathcal{F}$ be the family of ordered pairs of k-cliques in $\mathcal{K}$ that intersect in at least two vertices. Finally, let $\mathcal{K}'$ be the subfamily of $\mathcal{K}$, obtained by removing a k-clique from each pair in $\mathcal{F}$. Then

$$\mathbb{E}(|\mathcal{K}'|) \ge \mathbb{E}(|\mathcal{K}|) - \mathbb{E}(|\mathcal{F}|) = q\mathbb{E}(|\mathcal{K}|) - q^2\mathbb{E}(Z) = q\mu - q^2\nu.$$

Take $q = \frac{\mu}{2\nu}$. (Notice that this is less than 1, as $\nu \ge \mu$.) This gives $\mathbb{E}(|\mathcal{K}'|) \ge \frac{\mu^2}{4\nu} \ge \frac{n^2}{8k^5} \ge \frac{n^2}{k^6}$. Since $\mathcal{K}'$ is a family of pairwise edge-disjoint k-cliques in G, we have $\mathbb{E}(Y) \ge \mathbb{E}(|\mathcal{K}'|) \ge \frac{n^2}{k^6}$. □

7.6 The chromatic number of sparse random graphs

Theorem 7.14 (Łuczak, 1991)**.** *Let $p = p(n) \in (0,1)$ satisfy $p \cdot n^{5/6} \to 0$. Then there exists u such that, with high probability,*

$$u \le \chi(G(n,p)) \le u + 3.$$

Lemma 7.15. *Let $c > 0$ be a constant and let $p = p(n) \in (0,1)$ satisfy $p \cdot n^{5/6} \to 0$. Then, with high probability, every subgraph of $G(n,p)$ on at most $c\sqrt{n}$ vertices is 3-colourable.*

Proof. We first show that, with high probability, for every $t \le c\sqrt{n}$, every t vertices induce fewer than $3t/2$ edges. Indeed, the probability of this failing for some $t \le c\sqrt{n}$ is at most

$$\begin{aligned}
\sum_{4\le t\le c\sqrt{n}} \binom{n}{t}\binom{\binom{t}{2}}{\frac{3t}{2}} p^{3t/2} &\le \sum_{4\le t\le c\sqrt{n}} \left(\frac{en}{t}\right)^t \left(\frac{2et^2}{3t}\right)^{3t/2} p^{3t/2} \\
&\le \sum_{4\le t\le c\sqrt{n}} \left(10n^{2/3}pt^{1/3}\right)^{3t/2} \\
&\le \sum_{4\le t\le c\sqrt{n}} \left(10c^{1/3}n^{5/6}p\right)^{3t/2} \\
&\le \sum_{t=4}^{\infty} \left(10c^{1/3}n^{5/6}p\right)^t \le 2\left(10c^{1/3}n^{5/6}p\right)^4 \to 0,
\end{aligned}$$

using that $p \cdot n^{5/6} \to 0$.

Now suppose that there is a subgraph G' of $G = G(n,p)$ on at most $c\sqrt{n}$ vertices which is not 3-colourable, and take G_0 to be a minimal such graph. By the above, we may assume that every subgraph of G on t vertices, with $t \le c\sqrt{n}$, has fewer than $3t/2$ edges. In particular, the average degree of G_0 is less than 3, showing that there is a vertex $v \in V(G_0)$ with degree at most 2 in G_0. By minimality of G_0, the graph $G_0 - v$ is 3-colourable. By this means that G_0 is also 3-colourable (colour v by a colour not present in its neighbourhood), a contradiction. □

Proof of Theorem 7.14. Write $G = G(n,p)$. Let $\varepsilon > 0$ be an arbitrary constant, and let $u = u(n,p,\varepsilon)$ be the least integer satisfying

$$\mathbb{P}(\chi(G) \le u) \ge \varepsilon.$$

(It is easy to see that such u exists.)

Let Y be the minimal size of a set S such that $G - S$ is u-colourable. As in Theorem 7.7, we think of Y as a function of n independent random variables $X_1, \ldots, X_n$, where X_i encodes the edges from the vertex i to the vertices $[i-1]$. Notice that Y is 1-Lipschitz, namely changing the value of X_i changes Y by at most 1 (if G and G' differ only on edges touching a vertex v, then if S is a set for which $G - S$ is u-colourable, then $G' - (S \cup \{v\})$ is also u-colourable, and vice versa). Hence, by McDiarmid's inequality (Theorem 7.1), letting $\mu = \mathbb{E}(Y)$,

$$\mathbb{P}\left(|Y - \mathbb{E}(Y)| \geq \lambda\sqrt{n}\right) \leq 2e^{-2\lambda^2} < \varepsilon, \tag{26}$$

where λ is a constant satisfying $e^{-2\lambda^2} < \varepsilon$.

We claim that $\mathbb{E}(Y) \leq \lambda\sqrt{n}$. Indeed, otherwise $\mathbb{P}(Y = 0) \leq \mathbb{P}(|Y - \mathbb{E}(Y)| \geq \lambda\sqrt{n}) < \varepsilon$. Since $Y = 0$ exactly when G is u-colourable, this is a contradiction to the choice of u. By $\mathbb{E}(Y) \leq \lambda\sqrt{n}$, with probability at least $1 - \varepsilon$, we have $Y \leq 2\lambda\sqrt{n}$. Assume this holds, and take S to be a set of minimal size such that $G - S$ is u-colourable, so $|S| \leq 2\lambda\sqrt{n}$. By Lemma 7.15, the subgraph of G induced by S is 3-chromatic. Altogether, we find that G is $(u+3)$-chromatic (we can colour $G - S$ by u colours, and colour the rest by three different colours).

In summary, we have shown that $\mathbb{P}(u \leq \chi(G) \leq u + 3) \geq 1 - 2\varepsilon$. Since ε was arbitrary, this proves the theorem. □

Remark 7.16. With a little more effort, Łuczak's proof actually gives a 2-point concentration, namely it shows the existence of u such that, with high probability, $\chi(G(n,p)) \in \{u, u+1\}$, for p (roughly) in the same range. This was improved by Alon–Krivelevich (1997) to a 2-point concentration result for $p \geq n^{-1/2+\varepsilon}$, for any constant $\varepsilon > 0$.