Probabilistic and Extremal Combinatorics

MAT3415: Course Notes

Luke Collins*

v1.0† (February 10, 2017)

Table of Contents

I Probabilistic Combinatorics 4

1 An Introduction to Discrete Probability Theory 4

2 Applications in Combinatorics 7
- 2.1 Ramsey Numbers . . . 7
- 2.2 Van der Waerden Numbers . . . 10
- 2.3 Two-Colourable Uniform Families . . . 12
- 2.4 (r, s)-Systems . . . 13
- 2.5 Random Graphs . . . 14

3 Conditional Probability and Independence 15
- 3.1 A Combinatorial Application: Tournaments . . . 20

4 The Lovász Local Lemma 22

5 Random Variables and Expectation 26
- 5.1 Random Variables . . . 26
- 5.2 The Expected Value of a Random Variable . . . 29
- 5.3 Hamiltonian Paths in Tournaments . . . 33
- 5.4 Large Bipartite Subgraphs . . . 34
- 5.5 Independent Sets and Turán's Theorem . . . 34

II Extremal Combinatorics 37

6 Introduction 37

*Based on lectures given by Prof Peter Borg and Dr John B. Gauci at the Mathematics Department, University of Malta (Winter term 2017–18).

†This document is interactive. Whenever you see phrases such as "by theorem x" or "by definition of y", you can click on them to be taken to where they are first stated. If you find any errors or typos whilst reading these notes, please contact the author on luke@maths.com.mt.

7 Exact Intersections **38**

8 Double Counting **40**
8.1 Some Simple Applications 40

9 Intersecting Families **41**
9.1 The Erdős-Ko-Rado Theorem 42

10 Levels of Hereditary Families **45**

11 Antichains **48**

12 Shadows **49**

This set of notes requires some preliminary knowledge of elementary discrete mathematics and graph theory. Notes on these topics can be found at https://maths.com.mt/notes.

PART I

PROBABILISTIC COMBINATORICS

1 An Introduction to Discrete Probability Theory

Definition 1.1 (Finite Probability Space)**.** A *finite probability space* is a pair (Ω, P) where Ω is a finite set, and $\mathsf{P}\colon \wp\,\Omega \to [0,1]$ is a function which assigns a probability to each $A \subseteq \Omega$, such that:

(i) $\mathsf{P}(\Omega) = 1$, and

(ii) $\mathsf{P}(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mathsf{P}(A_i)$ for any finite family of pairwise disjoint subsets $\{A_i\}_{i\in I}$ of Ω.

Proposition 1.2. *Given any finite probability space* (Ω, P)*, then:*

(i) $\mathsf{P}(\emptyset) = 0$*,*

(ii) $\mathsf{P}(A) = \sum_{\omega\in A} \mathsf{P}(\{\omega\})$ *for any* $A \subseteq \Omega$*, and*

(iii) $\mathsf{P}(\Omega \setminus A) = 1 - \mathsf{P}(A)$*.*

Proof. We prove each part separately.

(i) $1 = \mathsf{P}(\Omega) = \mathsf{P}(\Omega \cup \emptyset) = \mathsf{P}(\Omega) + \mathsf{P}(\emptyset) = 1 + \mathsf{P}(\emptyset)$, hence $\mathsf{P}(\emptyset) = 0$.

(ii) $\mathsf{P}(A) = \mathsf{P}\big(\bigcup_{\omega\in A}\{\omega\}\big) = \sum\limits_{\omega\in A} \mathsf{P}(\{\omega\})$, as required.

(iii) $1 = \mathsf{P}(\Omega) = \mathsf{P}(A \cup (\Omega \setminus A)) = \mathsf{P}(A) + \mathsf{P}(\Omega \setminus A)$. □

Definition 1.3 (Event)**.** Let (Ω, P) be a finite probability space. An element $\omega \in \Omega$ is called an *elementary event*, whereas a subset $A \subseteq \Omega$ is called an *event*.

Remark 1.4. A finite probability space provides us with a mathematical formulation for a random experiment, by which we mean:

(i) A set of all possible outcomes of a random experiment, that is, the sample space Ω,

(ii) a set of events $\mathcal{F} \subseteq \wp\,\Omega$ where each event is a set of zero or more elements which represent outcomes, and

(iii) a probability function $\mathsf{P}\colon \mathcal{F} \to [0,1]$ assigning probabilities to the events.

Proposition 1.5. *Let* Ω *be a finite set, and let* $\mathsf{P}\colon \wp\,\Omega \to [0,1]$ *such that* $\mathsf{P}(\Omega) = 1$*. Then* (Ω, P) *is a finite probability space if and only if*

$$\mathsf{P}(A) = \sum_{\omega\in A} \mathsf{P}(\{\omega\})$$

for all $A \subseteq \Omega$*.*

Proof. The 'only if' part follows from proposition 1.2. We prove the 'if' part. Let $\{A_i\}_{i\in I}$ be a finite collection of pairwise disjoint subsets of Ω. Define $J = \{i \in I : A_i \neq \emptyset\}$, and for any $j \in J$, let $a_{j1}, a_{j2}, \ldots, a_{j|A_j|}$ denote the elements of A_j, and let $j_1, j_2, \ldots, j_{|J|}$ denote the elements of J. Since $A_{j_1}, A_{j_2}, \ldots, A_{j_{|J|}}$ are pairwise disjoint and nonempty, then

$$a_{j_1 1}, a_{j_1 2}, \ldots, a_{j_1|A_{j_1}|}, a_{j_2 1}, a_{j_2 2}, \ldots, a_{j_2|A_{j_2}|}, \ldots, a_{j_{|J|}1}, a_{j_{|J|}2}, \ldots, a_{j_{|J|}|A_{j_{|J|}}|}$$

are the distinct elements of $\bigcup_{j\in J} A_j$. Thus

$$\begin{aligned}
\mathsf{P}\Big(\bigcup_{i\in I} A_i\Big) = \mathsf{P}\Big(\bigcup_{j\in J} A_j\Big) &= \sum_{\omega\in \bigcup_{j\in J} A_j} \mathsf{P}(\{\omega\}) && \text{(by the hypothesis)} \\
&= \mathsf{P}(\{a_{j_1 1}\}) + \mathsf{P}(\{a_{j_1 2}\}) + \cdots + \mathsf{P}(\{a_{j_{|J|}|A_{j_{|J|}}|}\}) \\
&= \sum_{k=1}^{|A_{j_1}|} \mathsf{P}(\{a_{j_1 k}\}) + \sum_{k=1}^{|A_{j_2}|} \mathsf{P}(\{a_{j_2 k}\}) + \cdots + \sum_{k=1}^{|A_{j_{|J|}}|} \mathsf{P}(\{a_{j_{|J|}k}\}) \\
&= \mathsf{P}(A_{j_1}) + \mathsf{P}(A_{j_2}) + \cdots + \mathsf{P}(A_{j_{|J|}}) \quad \text{(by the hypothesis)} \\
&= \sum_{j\in J} \mathsf{P}(A_j) = \sum_{i\in I} \mathsf{P}(A_i),
\end{aligned}$$

since the latter may contain some terms $\mathsf{P}(\emptyset)$, which are zero. □

Remark 1.6. Given a set Ω and a function $\mathsf{P}\colon \wp\Omega \to [0,1]$, if we have

$$\sum_{\omega\in\Omega} \mathsf{P}(\{\omega\}) = 1 \qquad \text{and} \qquad \mathsf{P}(A) = \sum_{\omega\in A} \mathsf{P}(\{\omega\})$$

for all $A \subseteq \Omega$, it follows that $\mathsf{P}(\Omega) = 1$. Thus these two properties are sufficient for (Ω, P) to be a (finite[1]) probability space.

Remark 1.7. Consider a probability space (Ω, P) where all elementary events $\omega \in \Omega$ have the same probability $\mathsf{P}(\{\omega\}) = p$. Now $1 = \mathsf{P}(\Omega) = \sum_{\omega\in\Omega} \mathsf{P}(\{\omega\}) = |\Omega|p$, so we must have

$$p = \frac{1}{|\Omega|}.$$

Also, $\mathsf{P}(A) = \sum_{\omega\in A} \mathsf{P}(\{\omega\}) = |A|p$, so

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|}.$$

Example 1.8 (Birthday Problem). We give an example to illustrate what remark 1.7 is telling us. Given a group of n people, we determine the probability that at least two people celebrate their birthday on the same day. To keep things simple, we assume that no one has their birthday on the 29$^{\text{th}}$ of February, and that no birthday is more common than any other.

[1] Henceforth, whenever we write 'probability space', we are still referring to a *finite* probability space.

Let S denote the required event. Clearly if $n > 365$, then $\mathsf{P}(S) = 1$. So suppose $n \leqslant 365$, let the set of people be $[n]$, and for each person $i \in [n]$, define the day $d_i \in [365]$ on which person i has their birthday.

Let Ω be the set of n-tuples $(d_1, d_2, \ldots, d_n)$ with each entry in $[365]$, so we have $|\Omega| = 365^n$. Each n-tuple represents the possibility that person i has their birthday on day d_i, and since we have assumed that each one of these is equally likely, then $\mathsf{P}(A) = |A|/|\Omega|$ for all $A \subseteq \Omega$.

Let D be the event that each person has a different birthday. Clearly the elements of D are n-tuples where $d_i \neq d_j$ for $i \neq j$. There are $365 \cdot 364 \cdot (365 - n + 1)$ tuples of this form. But $S = \Omega \setminus D$, so by proposition 1.2 we have $\mathsf{P}(S) = 1 - \mathsf{P}(D) = 1 - |D|/|\Omega|$, and thus

$$\mathsf{P}(S) = 1 - \frac{1}{365^n} \prod_{i=1}^{n} (366 - i).$$

Rather counter-intuitively, we observe that for a group of $n = 23$ people, the probability that at least two people share the same birthday is over 50%.

Theorem 1.9. *Let (Ω, P) be a probability space, and let $\{A_i\}_{i=1}^n$ be a family of subsets of Ω. Then*

$$\mathsf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) \leqslant \sum_{i=1}^{n} \mathsf{P}(A_i).$$

Proof. By induction on n. For $n = 1$, the result is trivial. Let us show that it holds for $n = 2$.

$$\begin{aligned}
\text{Consider } \mathsf{P}(A_1 \cup A_2) &= \mathsf{P}(A_1 \cup (A_2 \setminus A_1)) \\
&= \mathsf{P}(A_1) + \mathsf{P}(A_2 \setminus A_1) && \text{(since they are disjoint)} \\
&= \mathsf{P}(A_1) + \mathsf{P}(A_2 \setminus (A_1 \cap A_2)) \\
\text{But } \mathsf{P}(A_2) &= \mathsf{P}[(A_1 \cap A_2) \cup (A_2 \setminus (A_1 \cap A_2))] \\
&= \mathsf{P}(A_1 \cap A_2) + \mathsf{P}(A_2 \setminus (A_1 \cap A_2)), && \text{(disjoint)} \\
\text{Therefore } \mathsf{P}(A_1 \cup A_2) &= \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1 \cap A_2),
\end{aligned}$$

which is $\leqslant \mathsf{P}(A_1) + \mathsf{P}(A_2)$, as required. Now consider

$$\begin{aligned}
\mathsf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \mathsf{P}\Big(\Big(\bigcup_{i=1}^{n-1} A_i\Big) \cup A_n\Big) \\
&\leqslant \mathsf{P}\Big(\bigcup_{i=1}^{n-1} A_i\Big) + \mathsf{P}(A_n) && \text{(by the case with } n = 2) \\
&\leqslant \sum_{i=1}^{n-1} \mathsf{P}(A_i) + \mathsf{P}(A_n) && \text{(by the hypothesis)} \\
&= \sum_{i=1}^{n} \mathsf{P}(A_i),
\end{aligned}$$

as required. □

Observe that the base case for $n = 2$ provides us with a useful formula for finding the probability of the union of two events:

Corollary 1.10. *Let (Ω, P) be a probability space, and let $A_1, A_2 \subseteq \Omega$. Then*

$$\mathsf{P}(A_1 \cup A_2) = \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1 \cap A_2).$$

More generally, we have the following.

Theorem 1.11 (Inclusion-Exclusion Principle)**.** *If (Ω, P) is a probability space and $A_1, \ldots, A_n \subseteq \Omega$, then*

$$\begin{aligned}\mathsf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{i=1}^{n} \sum_{\substack{\mathcal{J}\subseteq\{1,2,\ldots,n\}\\ |\mathcal{J}|=i}} (-1)^{i-1}\mathsf{P}\Big(\bigcap_{k\in\mathcal{J}} A_k\Big) \\ &= \sum_{i_1=1}^{n} \mathsf{P}(A_{i_1}) - \sum_{1\leqslant i_1<i_2\leqslant n} \mathsf{P}(A_{i_1}\cap A_{i_2}) + \sum_{1\leqslant i_1<i_2<i_3\leqslant n} \mathsf{P}(A_{i_1}\cap A_{i_2}\cap A_{i_3}) \\ &\quad - \cdots + (-1)^{k+1} \sum_{1\leqslant i_1<\cdots<i_k\leqslant n} \mathsf{P}\Big(\bigcap_{j=1}^{k} A_{i_j}\Big) + \cdots + (-1)^{n+1}\mathsf{P}\Big(\bigcap_{i=1}^{n} A_i\Big).\end{aligned}$$

We will omit the proof of this result, because it goes similarly to the proof of the inclusion-exclusion principle of set-magnitude covered in the notes on discrete mathematics.[2] Aside from that, it does not involve probabilistic techniques, which are the subject of these notes.

2 Applications in Combinatorics

2.1 Ramsey Numbers

We introduce the concept of Ramsey numbers. Consider a group of six people. We claim that at least three people either all know each other (i.e. are friends), or do not know each other (i.e. are strangers), and that this is the case for any group of six people.

Let us use a graph to represent the group of people, where the vertices of the graph represent the people, and edges represent friendships. For example, in figure 1, people 1, 2 and 3 are all friends, whereas 4, 5 and 6 are all strangers.

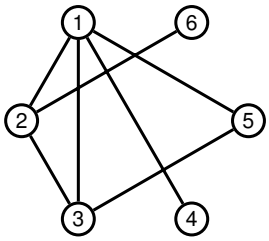

Figure 1

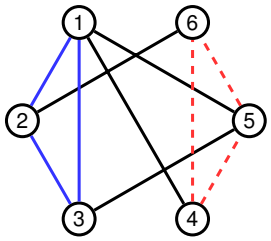

Figure 2

[2] Available here.

The same can be said of 1, 3, and 5 (friends) and 3, 4 and 6 (strangers). Our claim is that at least one of these is true for any graph on six vertices. In other words, we claim that at least three vertices form a triangle (shown blue in figure 2) or a non-triangle (dotted red in figure 2).

Let us show that our claim is true.

Proof. For contradiction, suppose $G = (V, E)$ is a graph on six vertices such that our claim is untrue. Let $v \in V$ be one of the vertices of G, and let $W \subseteq V$ be the set of all its neighbours. Suppose $|W| \geqslant 3$. If G has an edge $e \in E$ on any any two of the vertices in W, then we end up with a triangle on v and the two endpoints of e. Thus G cannot have any edges on the set W, and hence the vertices of W form a non-triangle.

Therefore we cannot have $|W| \geqslant 3$, so suppose $|W| < 3$. Then there are at least three vertices v_1, v_2, v_3 of G outside the set W. These three cannot form a triangle (otherwise our claim would be true), so two of v_1, v_2, v_3 are not on an edge, say, v_1 and v_2. Thus $\{v, v_1, v_2\}$ are a set of non-adjacent vertices which form a non-triangle. Hence G cannot exist; so our claim is always true. □

Definition 2.1 (n-colouring)**.** Let X be a set. If there exists a function $f\colon X \to \{1, 2, \ldots, n\}$, we call f an n-colouring of X.

Note. A set X can be n-coloured in $n^{|X|}$ ways.

Remark 2.2. In our example with the group of six people, we can use blue edges to represent friendships, and red edges to represent non-friendships. This way, our problem becomes about colouring the edges of the complete graph on six vertices, K_6. Thus our claim becomes that for any 2-colouring of the edges of K_6, there must be a monochromatic triangle (that is, a triangle of one colour).

Also, recall that a 'triangle' is simply the complete graph on three vertices, K_3.

Definition 2.3 (Ramsey Number)**.** For any two integers $k, \ell \geqslant 2$, the *Ramsey number* $R(k, \ell)$ is the least integer n such that any 2-colouring of the edges of the complete graph K_n contains either a blue K_k or a red K_ℓ as a subgraph.

In our example, we showed that every graph on 6 vertices must contain either a blue K_3 or a red K_3, Therefore we must have $R(3, 3) \leqslant 6$. Now, is it possible that K_n for $n < 6$ *always* contains either a red/blue K_3? In other words, does our earlier claim hold for any group with less than six people? The answer is no:

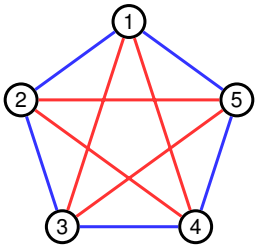

Figure 3: A colouring of the complete graph on $n = 5$ vertices without a monochromatic induced K_3

Here we see that it is possible to colour the edges of the graph K_5 without inducing either a red or a blue triangle, so $R(3,3) > 5$. Combining this with the fact that $R(k,\ell) \leqslant 6$, we get that $R(3,3) = 6$. Therefore we need at least six people for our claim to be true.

Remark 2.4. Let us give some facts about Ramsey numbers.

(i) If $m \geqslant n$, where $n = R(k,\ell)$, then K_n is a subgraph of K_m, so we can still always find a red K_k or a blue K_ℓ. In other words, if we add more people to the group of six, our claim would still hold.

(ii) $R(k,\ell) = R(\ell,k)$. This is obvious, since we are just switching from wanting a blue K_k or a red K_ℓ to the other way around, a red K_k or a blue K_ℓ.

(iii) $R(2,\ell) = \ell$.

For example, take $R(2,5)$. Following a similar exhaustive logic to our example on K_6, that is, attempting to construct a counterexample and finding that it is impossible to do so, one can show that $R(2,5) \leqslant 5$. This is not hard to see – if we start with all the edges coloured red, we would have our red K_5, but then if we colour any one of the edges blue, we immediately get our blue K_2. We also get that $R(2,5) > 4$ similarly, because if we start with an all-red K_4, we have neither a red K_5 nor a blue K_2. Therefore $R(2,5) = 5$.

Theorem 2.5 (Ramsey's Theorem)**.** *$R(k,\ell)$ exists for any $k,\ell \in \mathbb{N}$. Moreover,*

$$R(k,\ell) \leqslant R(k-1,\ell) + R(k,\ell-1).$$

The proof of Ramsey's theorem is by induction on $k+\ell$. It does not utilise any probabilistic techniques, so we omit it here. From this result, we get the following corollary.

Corollary 2.6 (Erdős & Szekeres)**.** $R(k,\ell) \leqslant \binom{k+\ell-2}{k-1}$.

Remark 2.7. It is not easy to evaluate Ramsey numbers. The smallest unknown Ramsey number is $R(5,5)$ (it is known however that $43 \leqslant R(5,5) \leqslant 48$).

The concept of Ramsey numbers can be extended from 2 to n colours, as can the result of Ramsey's theorem. Given a set $C = \{1,2,\ldots,r\}$ of colours and $k_1,k_2,\ldots,k_r \in \mathbb{N}$, there exists $n = R(k_1,k_2,\ldots,k_r)$ such that the edges of K_n can be coloured with the r colours in C and for some $i \in C$ there exists a complete subgraph K_{k_i} whose edges are all coloured i.

So far we have seen upper-bounds for $R(k,\ell)$. The following theorem provides us with a lower-bound, and the proof of the theorem utilises a probabilistic technique.

Theorem 2.8 (Erdős, 1947)**.** *For any $k \geqslant 3$, $R(k,k) > 2^{k/2}$.*

Proof. The theorem is telling us that if $n \leqslant 2^{k/2}$, then there exists a two-colouring of K_n without a monochromatic subgraph K_k. Consider 2-colouring K_n randomly, i.e. assigning colours to the edges in such a way that each colour is equally likely. Define the probability space (Ω,P), where Ω is the set of all 2-colourings of K_n, and $\mathsf{P}\colon \wp\Omega \to [0,1]$ is given by $\mathsf{P}(A) = |A|/|\Omega|$ for all $A \subseteq \Omega$, since each 2-colouring is equally likely (by remark 1.7).

Let B be the set of colourings which induce a blue K_k, and similarly let R be the set of colourings which induce a red K_k. Let M be the set of colourings which induce a monochromatic K_k, i.e. $M = B \cup R$. We show that the theorem is true by showing that for $n \leqslant 2^{k/2}$, $\mathsf{P}(\Omega \setminus M) > 0$, that is, it is not impossible for a colouring *not* to induce a monochromatic subgraph K_k (and therefore $\Omega \setminus M$ is nonempty and a counter-example exists).

Now $|\Omega| = 2^{\binom{n}{2}}$, and $|B| = |R| \leqslant \binom{n}{k} 2^{\binom{n}{2}-\binom{k}{2}}$, since for each of the $\binom{n}{k}$ different ways of fixing a red/blue K_k, there exist $2^{\binom{n}{2}-\binom{k}{2}}$ ways of 2-colouring the remaining edges.[3] Thus:

$$\frac{|B|}{|\Omega|} = \frac{|R|}{|\Omega|} \leqslant \frac{\binom{n}{k} 2^{\binom{n}{2}-\binom{k}{2}}}{2^{\binom{n}{2}}} = \binom{n}{k} 2^{-\binom{k}{2}} = \frac{n(n-1)\cdots(n-k+1)}{k!} 2^{-\binom{k}{2}} < \frac{n^k}{k!} 2^{-\binom{k}{2}}.$$

Now for $n \leqslant 2^{k/2}$, we get

$$\frac{|B|}{|\Omega|} = \frac{|R|}{|\Omega|} < \frac{\left(2^{k/2}\right)^k}{k!} 2^{-\binom{k}{2}} = \frac{2^{1/2}}{k!} < \frac{1}{2} \quad \text{for } k \geqslant 3,$$

and therefore

$$\begin{aligned} \mathsf{P}(\Omega \setminus M) &= 1 - \mathsf{P}(M) = 1 - \mathsf{P}(B \cup R) \geqslant 1 - \mathsf{P}(B) - \mathsf{P}(R) \qquad \text{(by theorem 1.9)} \\ &= 1 - \frac{|B|}{|\Omega|} - \frac{|A|}{|\Omega|} > 1 - \frac{1}{2} - \frac{1}{2} = 0, \end{aligned}$$

as required. □

Remark 2.9. The general idea behind what we are calling the *probabilistic technique* is the following. To prove that a structure with certain desired properties exists, we define an appropriate probability space, and show that the desired properties hold with positive probability within the probability space, as we have just done in the proof of theorem 2.8.

2.2 Van der Waerden Numbers

Another application of the probabilistic technique gives us a lower-bound on Van der Waerden numbers.

Definition 2.10 (Van der Waerden Number)**.** For any positive integer k, the *Van der Waerden number* $W(k)$ is the least integer n such that any two-colouring of the set $[n] = \{1, 2, \ldots, n\}$ yields a monochromatic arithmetic progression of length k.

Example 2.11. Let us try and find $W(2)$ using red and blue to colour the set $[n]$. Clearly n must be at least 2 to induce an arithmetic progression of length 2. Now $W(2) \neq 2$,

[3] We have $\leqslant$ not $=$ here because different choices may still result in the same colouring. For example, if we have $n = 10$ and $k = 5$, then we choose 5 vertices to make up our monochromatic K_5, and then colour the remaining edges randomly. Suppose the remaining edges happened to all be given the same colour as the edges of the K_5. Notice that the resulting colouring may have been obtained if we started with a different choice of 5 vertices.

since $\{1, 2\}$ is a colouring without a monochromatic arithmetic progression of length 2. Can $W(2) = 3$? Consider every possible colouring of $[3]$:

$$\begin{array}{cccc} \{1,2,3\} & \{1,2,3\} & \{1,2,3\} & \{1,2,3\} \\ \{1,2,3\} & \{1,2,3\} & \{1,2,3\} & \{1,2,3\} \end{array}$$

In every case, we observe that a monochromatic arithmetic progression of length $\geqslant 2$ is induced. Therefore $W(2) = 3$.

What about $W(3)$? Well, observe that $W(3) \geqslant 9$, since we can colour $[8]$ in such a way that no arithmetic progression of length 3 is induced: $\{1, 2, 3, 4, 5, 6, 7, 8\}$. If we then add 9 to this set, coloured red/blue, in each case a monochromatic arithmetic progression of length 3 is induced. In fact, $W(3) = 9$.

Remark 2.12. It is not simple to determine Van der Waerden numbers. In fact, the number $W(k)$ is only known for $k = 1, 2, \ldots, 6$, and $W(6) = 1132$.

Theorem 2.13 (Trivial Lower-bound)**.** $W(k) > 2^{k/2}$.

Proof. Suppose $W(k) = n \leqslant 2^{k/2}$. Colour the elements of $[n]$ randomly red and blue, and consider the probability space (Ω, P) where Ω is the set of all such colourings, and $\mathsf{P}(A) = |A|/|\Omega|$ for all $A \subseteq \Omega$. Clearly we have $|\Omega| = 2^n$. Now let $\mathcal{S}$ be the set of all arithmetic progressions in $[n]$ consisting of k terms. For each $S \in \mathcal{S}$, let B_S be the event that S is blue, and similarly let R_S be the event that S is red. Any valid blue progression may be coloured red to give rise to a distinct valid red progression, so clearly $|B_S| = |R_S|$.

Define $A_S := B_S \cup R_S$, so we have $\mathsf{P}(A_S) = \mathsf{P}(B_S) + \mathsf{P}(R_S)$ (B_S and R_S are disjoint). Given an arithmetic progression $S \in \mathcal{S}$ coloured red or blue, the remaining terms of $[n]$ can be coloured red or blue in 2^{n-k} ways, where each distinct colouring gives rise to a member of A_S. In other words, $|A_S| = 2 \cdot 2^{n-k} = 2^{n-k+1}$, and thus

$$\mathsf{P}(A_S) = \frac{|A_S|}{|\Omega|} = \frac{2^{n-k+1}}{2^n} = 2^{1-k}.$$

Now let M be the event that a random two-colouring yields a monochromatic arithmetic progression (in $\mathcal{S}$). Thus $M = \bigcup_{S \in \mathcal{S}} A_S$. Observe that $|\mathcal{S}| \leqslant \binom{n}{2}$, since the first two terms of an arithmetic progression determine its terms uniquely. Thus

$$\begin{aligned} \mathsf{P}(M) = \mathsf{P}\Big(\bigcup_{S \in \mathcal{S}} A_S\Big) &\leqslant \sum_{S \in \mathcal{S}} \mathsf{P}(A_S) && \text{(by theorem 1.9)} \\ &= |\mathcal{S}| 2^{1-k} \leqslant \binom{n}{2} 2^{1-k} = \frac{n^2 - n}{2} 2^{1-k} = \frac{n^2 - n}{2^k}. \end{aligned}$$

We assumed that n is at most $2^{k/2}$, and hence $\mathsf{P}(M)$ is at most $\frac{n^2-n}{2^k} < \frac{n^2}{2^k} = \frac{2^k}{2^k} = 1$, when we put $n = 2^{k/2}$. Thus $\mathsf{P}(M) < 1$, and $\mathsf{P}(\Omega \setminus M) = 1 - \mathsf{P}(M) > 0$, and therefore a colouring of $[n]$ without an induced monochromatic arithmetic progression of length k exists. □

2.3 Two-Colourable Uniform Families

Here we introduce a generalisation of a graph, where edges may join any number of vertices (as opposed to just two). We then prove a theorem on vertex colouring of hypergraphs using the probabilistic technique.

Definition 2.14 (Hypergraph)**.** A *hypergraph* is a pair $H = (V, E)$ where V is a finite set and $E \subseteq \wp V$ is a family of subsets of V. The elements of V are called *vertices*, and the elements of E are called *hyper-edges* (or simply *edges*).

Example 2.15. A hypergraph on the vertices $V = \{v_1, \dots, v_7\}$ having (hyper) edge set $E = \{\{v_1, v_2, v_3\}, \{v_1, v_4\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}\}$ is depicted in figure 4.

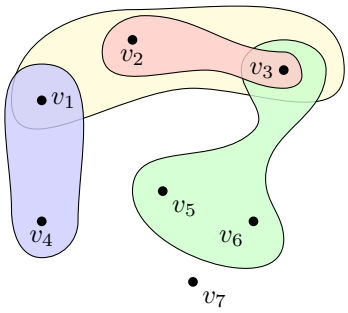

Figure 4: Example of a hypergraph

Definition 2.16 (Uniform Hypergraph)**.** A hypergraph $H = (V, E)$ is said to be *r-uniform* if $|e| = r$ for all $e \in E$. In other words, each of its edges consist of precisely r vertices.

Remark 2.17. Any graph can be considered a 2-uniform hypergraph.

Definition 2.18 (2-Colourable)**.** A hypergraph $H = (V, E)$ is said to be *2-colourable* if there exists a colouring of the set V such that no edge $e \in E$ is monochromatic.

Note. This definition does not prevent vertices within the same edge to have the same colour. For example, colouring v_1, v_2, v_5, v_6 and v_7 all red, then v_3 and v_4 blue, we get an admissible 2-colouring of the graph in example 2.15.

Theorem 2.19 (Erdős, 1963)**.** *Every r-uniform hypergraph with less than 2^{r-1} edges is 2-colourable.*

Proof. Let $H = (V, E)$ be an r-uniform hypergraph, and let $|E| < 2^{r-1}$. Randomly colour the vertices of V red and blue, and consider the probability space (Ω, P), where Ω is the set of all possible 2-colourings ($|\Omega| = 2^{|V|}$), and $\mathsf{P}(A) = |A|/|\Omega| = |A|/2^{|V|}$ for all $A \subseteq \Omega$.

For every $e \in E$, let A_e be the event that e is a monochromatic edge, i.e. all the vertices in e have the same colour. Thus for every e, $|A_e| = 2^{|V|-r+1}$ (since H is r-uniform).

Now let M be the event that some edge is monochromatic, i.e. $M = \bigcup_{e \in E} A_e$. So

$$\begin{aligned} \mathsf{P}(M) = \mathsf{P}\Big(\bigcup_{e\in E} A_e\Big) &\leqslant \sum_{e\in E} \mathsf{P}(A_e) && \text{(by theorem 1.9)} \\ &= \sum_{e\in E} \frac{|A_e|}{|\Omega|} = \sum_{e\in E} \frac{2^{|V|-r+1}}{2^{|V|}} = \sum_{e\in E} \frac{1}{2^{r-1}} = \frac{|E|}{2^{r-1}}, \end{aligned}$$

but $|E| < 2^{r-1}$, so $\mathsf{P}(M) < 1$. Thus $\mathsf{P}(\Omega \setminus M) = 1 - \mathsf{P}(M) > 0$, implying $\Omega \setminus M$ is nonempty and therefore a 2-colouring without a monochromatic edge exists. □

2.4 (r, s)-Systems

Here we introduce yet another combinatorial structure, and prove a result using the probabilistic technique.

Definition 2.20 ((r, s)-system)**.** Let $\mathcal{F} = \{(A_i, B_i)\}_{i=1}^h$ be a family of pairs of sets. Then $\mathcal{F}$ is an (r, s)-*system* if for all $i, j \in [h]$,

i) $|A_i| = r$ and $|B_i| = s$, and

ii) $A_i \cap B_i = \emptyset$, and $A_i \cap B_j \neq \emptyset$ for $i \neq j$.

Example 2.21. We give an example of an (r, s)-system with $r = 2$ and $s = 3$.

$$\begin{aligned} (A_1, B_1) &= (\{1, 2\}, \{3, 4, 5\}) & (A_6, B_6) &= (\{2, 4\}, \{1, 3, 5\}) \\ (A_2, B_2) &= (\{1, 3\}, \{2, 4, 5\}) & (A_7, B_7) &= (\{2, 5\}, \{1, 3, 4\}) \\ (A_3, B_3) &= (\{1, 4\}, \{2, 3, 5\}) & (A_8, B_8) &= (\{3, 4\}, \{1, 2, 5\}) \\ (A_4, B_4) &= (\{1, 5\}, \{2, 3, 4\}) & (A_9, B_9) &= (\{3, 5\}, \{1, 2, 4\}) \\ (A_5, B_5) &= (\{2, 3\}, \{1, 4, 5\}) & (A_{10}, B_{10}) &= (\{4, 5\}, \{1, 2, 3\}) \end{aligned}$$

As we can see, the sets in each pair are disjoint, yet the intersection of each A_i with each B_j where $i \neq j$ is nonempty.

Theorem 2.22 (Bollobás, 1965)**.** *If $\{(A_i, B_i)\}_{i=1}^h$ is an (r, s)-system, then*

$$h \leqslant \binom{r+s}{r}.$$

Proof. Let $X = \bigcup_{i=1}^h (A_i \cup B_i)$, and let $n = |X|$.

Consider the probability space (Ω, P), where Ω is the set of all permutations of X, and $\mathsf{P}(A) = |A|/|\Omega| = |A|/n!$ for all $A \subseteq \Omega$. For each $i \in [h]$, let E_i denote the event that each element of A_i precedes each element of B_i in the set of permutations of X.[4] Then $|E_i| = \binom{n}{r+s} r!s!(n-r-s)! = n!r!s!/(r+s)! = n!/\binom{r+s}{r}$, and thus $\mathsf{P}(E_i) = 1/\binom{r+s}{r}$.

[4]For example, if we have $r = s = 2$, $X = \{1, 2, 3, 4, 5, 6\}$ and $(A_1, B_1) = (\{1, 2\}, \{3, 4\})$, then E_1 contains permutations such as 123456, 561234, 152364, and so on. A construction of a valid permutation in E_i involves first choosing where the $r + s$ elements are to be placed throughout the permutation of length n (which can be done in $\binom{n}{r+s}$ ways), then permuting them in such a way that the elements of A_i appear before those of B_i ($r!s!$ ways), and finally choosing what order to permute the remaining $n - r - s$ elements of X ($(n - r - s)!$ ways).

Now, we claim that all the E_i's are pairwise disjoint. Suppose, for contradiction, that this is not the case. Then for some $i, j \in [h]$ with $i \neq j$, there exists $\pi \in E_i \cap E_j$, i.e. $\pi \in E_i$ and $\pi \in E_j$, and so the last element a of A_i appears before the first element b of B_i, and likewise the last element a' of A_j appears before the first element b' of B_j. Now if a precedes a', then $A_i \cap B_j = \emptyset$, since a' precedes the first element b' of B_j and therefore there is no common intersection. If, on the other hand, a' precedes a, then we would similarly have $A_j \cap B_i = \emptyset$. Either way, we contradict the fact that we have an (r, s)-system ※. Hence the E_i's are pairwise disjoint.

Therefore,

$$1 \geqslant \mathsf{P}\Big(\bigcup_{i=1}^{h} E_i\Big) = \sum_{i=1}^{h} \mathsf{P}(E_i) = \frac{h}{\binom{r+s}{r}}, \qquad \text{(the } E_i\text{'s are disjoint)}$$

and the result follows. □

Remark 2.23. This result provides us with a sharp bound on h. Indeed, for the (r, s)-system

$$\{(A_i, X \setminus A_i) : A \subseteq X \wedge |A_i| = r\}$$

which is given in example 2.21 with $X = [5]$, $r = 2$ and $s = 3$, we have $\binom{r+s}{r} = \binom{2+3}{2} = 10 = h$.

2.5 Random Graphs

We prove that bipartite graphs on n vertices are increasingly sparse for large n using the probabilistic technique.

Definition 2.24 (Random Graph Probability Space)**.** Let $\mathcal{G}_n$ denote the set of all graphs on the vertex set $V = \{1, 2, \ldots, n\}$, and let $(\mathcal{G}_n, \mathsf{P})$ be the probability space in which each graph in $\mathcal{G}_n$ has equal probability, that is, $\mathsf{P}(A) = |A|/|\mathcal{G}_n| = |A|/2^{\binom{n}{2}}$ for all $A \subseteq \mathcal{G}_n$.

Proposition 2.25. *A random graph is almost surely not bipartite.*[5]

Proof. Let $V_n = \{1, 2, \ldots, n\}$. Recall that a graph G is bipartite if and only if V_n can be partitioned in 2 sets U and V $(= V_n \setminus U)$ such that all edges of G join a vertex in U and a vertex in V. Let B_n be the event that G is bipartite. Fix a subset of vertices $U \subseteq V_n$, and for each U, let $B_{n,U}$ be the event that all the edges of G are from U to $V_n \setminus U$.

Let $|U| = r$, and let $E_U = \{\{u, w\} : u \in U \wedge w \in V_n \setminus U\}$. Then we have $|E_U| = r(n-r)$, and hence $|B_{n,U}| = 2^{|E_U|} = 2^{r(n-r)}$, so

$$\begin{aligned}\mathsf{P}(B_{n,U}) &= \frac{2^{r(n-r)}}{2^{\binom{n}{2}}} = 2^{r(n-r)-\binom{n}{2}} \\ &\leqslant 2^{\frac{n}{2}\left(n-\frac{n}{2}\right)-\binom{n}{2}} = 2^{-\frac{n(n-2)}{4}}, \qquad \left(\text{since } r(n-r) \leqslant \frac{n}{2}\left(n-\frac{n}{2}\right)\right)\end{aligned}$$

[5] Formally, if B_n is the event that a random graph $G \in \mathcal{G}_n$ is bipartite, then $\lim_{n\to\infty} \mathsf{P}(B_n) = 0$.

and hence

$$\begin{aligned}\mathsf{P}(B_n) = \mathsf{P}\Big(\bigcup_{U\subseteq V_n} B_{n,U}\Big) &\leqslant \sum_{U\subseteq V_n} \mathsf{P}(B_{n,U}) && \text{(by theorem 1.9)}\\ &\leqslant \sum_{U\subseteq V_n} 2^{-\frac{n(n-2)}{4}} = 2^n \cdot 2^{-\frac{n(n-2)}{4}} = 2^{-\frac{n^2-6n}{4}},\end{aligned}$$

therefore $0 \leqslant \lim_{n\to\infty} \mathsf{P}(B_n) \leqslant \lim_{n\to\infty} 2^{-\frac{n^2-6n}{4}} = 0$, and the result follows. □

3 Conditional Probability and Independence

Definition 3.1 (Conditional Probability)**.** Let (Ω, P) be a probability space, and consider the events $A, B \subseteq \Omega$, where $\mathsf{P}(B) > 0$. The *probability of A given B*, denoted $\mathsf{P}(A|B)$, is defined

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

Example 3.2. Consider the random experiment of throwing a fair die. We have $\Omega = \{\boxdot, \boxdot, \boxdot, \boxdot, \boxdot, \boxdot\}$, and since the experiment is random we have $\mathsf{P}(A) = |A|/|\Omega|$ for any $A \subseteq \Omega$, by remark 1.7. In particular, we have $\mathsf{P}(\{\omega\}) = \frac{1}{6}$ for each $\omega \in \Omega$.

Now let A be the event that the number thrown in less than or equal to 3 ($A = \{\boxdot, \boxdot, \boxdot\}$), and let B be the event that the number is even ($B = \{\boxdot, \boxdot, \boxdot\}$). Then $\mathsf{P}(A) = \mathsf{P}(B) = \frac{1}{2}$, and

$$\mathsf{P}(A|B) = \frac{|\{\boxdot\}|}{|\{\boxdot, \boxdot, \boxdot\}|} = \frac{1}{3}.$$

We can interpret this probability thus: *Given that* we know that B has occurred, that is, the outcome of the die roll is an even number, what is the probability of A?

Theorem 3.3. *Let (Ω, P) be a finite probability space, let $B \subseteq \Omega$ where $\mathsf{P}(B) > 0$, and let $\mathsf{Q}\colon \wp\Omega \to [0,1]$ such that $\mathsf{Q}(A) = \mathsf{P}(A|B)$ for any $A \subseteq \Omega$. Then (Ω, Q) is a probability space.*

Proof. We show that (Ω, Q) is conformal with definition 1.1. First, we show that $\mathsf{Q}(\Omega) = 1$.

$$\mathsf{Q}(\Omega) = \mathsf{P}(\Omega|B) = \frac{\mathsf{P}(\Omega \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(B)}{\mathsf{P}(B)} = 1,$$

as required. Now let $A_1, A_2, \ldots, A_n$ be pairwise disjoint events in Ω. We show that $\mathsf{Q}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathsf{Q}(A_i)$.

$$\begin{aligned}\mathsf{Q}\Big(\bigcup_{i=1}^{n} A_i\Big) = \mathsf{P}\Big(\bigcup_{i=1}^{n} A_i \Big| B\Big) &= \frac{\mathsf{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap B\right)}{\mathsf{P}(B)}\\ &= \frac{\mathsf{P}\left(\bigcup_{i=1}^{n} (A_i \cap B)\right)}{\mathsf{P}(B)} \qquad \text{(elementary set theory)}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sum_{i=1}^{n} \mathsf{P}(A_i \cap B)}{\mathsf{P}(B)} && \text{(since the } A_i \cap B\text{'s are pairwise disjoint in } \Omega, \text{ and } (\Omega, \mathsf{P}) \text{ is a probability space)} \\
&= \sum_{i=1}^{n} \frac{\mathsf{P}(A_i \cap B)}{\mathsf{P}(B)} \\
&= \sum_{i=1}^{n} \mathsf{P}(A_i|B) \\
&= \sum_{i=1}^{n} \mathsf{Q}(A_i),
\end{aligned}
$$

as required. □

Definition 3.4 (Independence)**.** Let (Ω, P) be a probability space. Then the events $A_1, A_2, \ldots, A_n \subseteq \Omega$ are said to be *independent* if, for any $I \subseteq [n]$,

$$\mathsf{P}\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} \mathsf{P}(A_i).$$

We say that $A_1, A_2, \ldots, A_n$ are *pairwise independent* if the above holds for $|I| = 2$.

Remark 3.5. Independence $\implies$ pairwise independence, but the converse is not true. For example, suppose $\Omega = \{1, 2, 3, 4\}$, and $\mathsf{P}(\{\omega\}) = \frac{1}{4}$ for all $\omega \in \Omega$. Define $A = \{1, 2\}$, $B = \{1, 3\}$ and $C = \{2, 3\}$. We have $\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = \frac{1}{2}$, and

$$
\begin{aligned}
\mathsf{P}(A \cap B) = \mathsf{P}(\{1\}) &= \frac{1}{4} = \mathsf{P}(A) \cdot \mathsf{P}(B) \\
\mathsf{P}(A \cap C) = \mathsf{P}(\{1\}) &= \frac{1}{4} = \mathsf{P}(A) \cdot \mathsf{P}(C) \\
\mathsf{P}(B \cap C) = \mathsf{P}(\{1\}) &= \frac{1}{4} = \mathsf{P}(B) \cdot \mathsf{P}(C),
\end{aligned}
$$

so A, B and C are pairwise disjoint. But $\mathsf{P}(A \cap B \cap C) = \mathsf{P}(\emptyset) = 0 \neq \frac{1}{8} = \mathsf{P}(A) \cdot \mathsf{P}(B) \cdot \mathsf{P}(C)$.

Independent Random Experiments

Example 3.6. Consider two independent random experiments: experiment 1 is a fair coin toss, experiment 2 is throwing a die. So we have $\Omega_1 = \{H, T\}$, $\mathsf{P}_1(\{\omega\}) = \frac{1}{2}$, and $\Omega_2 = \{\boxdot, \boxdot, \boxdot, \boxdot, \boxdot, \boxdot\}$, $\mathsf{P}_2(\{\omega\}) = \frac{1}{6}$. Now suppose A is the event that we toss a heads, and B is the event that we get a 1 or a 2 on the dice. Then the probability of both these occurring is simply the product of the probabilities, $\mathsf{P}(\{H\} \text{ and } \{\boxdot, \boxdot\}) = \frac{1}{2} \times \frac{2}{6} = \frac{1}{6}$.

Observe that when considering this probability, we were working in the probability space $\Omega_1 \times \Omega_2$. In the following proposition 3.7, we see that indeed forms a valid probability space conformal with definition 1.1.

Proposition 3.7. *For $i = 1, 2, \ldots, n$, let (Ω_i, P_i) be a probability space. Let $\Omega := \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$, and let $\mathsf{P} : \wp\Omega \to \mathbb{R}$ such that for $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega$,*

$\mathsf{P}(\{\omega\}) := \prod_{i=1}^{n} \mathsf{P}_i(\{\omega\})$ *and* $\mathsf{P}(A) := \sum_{\omega\in A} \mathsf{P}(\{\omega\})$ *for* $A \subseteq \Omega$. *Then* (Ω, P) *is a probability space and*

$$\mathsf{P}(A_1 \times A_2 \times \cdots \times A_n) = \prod_{i=1}^{n} \mathsf{P}_i(A_i)$$

for any $A_1 \subseteq \Omega_1, A_2 \subseteq \Omega_2, \ldots, A_n \subseteq \Omega_n$.

This result is an immediate consequence of the following fact, which can be proved by induction on n.

Fact 3.8. Let $I_1, I_2, \ldots, I_n$ be finite sets. For each $j \in \{1, 2, \ldots, n\}$ and $i_j \in I_j$, let x_{ji_j} be a real number. Then

$$\sum_{(i_1,i_2,\ldots,i_n)\in I_1\times I_2\times\cdots\times I_n} x_{1i_1}x_{2i_2}\cdots x_{ni_n} = \Bigg(\sum_{i_1\in I_1} x_{1i_1}\Bigg)\Bigg(\sum_{i_2\in I_2} x_{2i_2}\Bigg)\cdots\Bigg(\sum_{i_n\in I_n} x_{ni_n}\Bigg).$$

Proof of proposition 3.7. Note that for each $i \in [n]$ we have $\mathsf{P}(\Omega_i) = 1$, since (Ω_i, P_i) is a probability space, and by proposition 1.2(ii), $\sum_{\omega_i\in A_i} \mathsf{P}_i(\{\omega_i\}) = \mathsf{P}_i(A_i)$. Now suppose we have $A = A_1 \times A_2 \times \cdots \times A_n \subseteq \Omega$. Then

$$\begin{aligned}
\mathsf{P}(A) &= \sum_{\omega\in A} \mathsf{P}(\{\omega\}) = \sum_{\substack{(\omega_1,\omega_2,\ldots,\omega_n)\\ \in A_1\times A_2\times\cdots\times A_n}} \mathsf{P}(\{(\omega_1, \omega_2, \ldots, \omega_n)\}) \\
&= \sum_{\substack{(\omega_1,\omega_2,\ldots,\omega_n)\\ \in A_1\times A_2\times\cdots\times A_n}} \mathsf{P}_1(\{\omega_1\})\mathsf{P}_2(\{\omega_2\})\cdots\mathsf{P}_n(\{\omega_n\}) \\
&= \Bigg(\sum_{\omega_1\in A_1} \mathsf{P}_1(\{\omega_1\})\Bigg)\Bigg(\sum_{\omega_2\in A_2} \mathsf{P}_2(\{\omega_2\})\Bigg)\cdots\Bigg(\sum_{\omega_n\in A_n} \mathsf{P}_n(\{\omega_n\})\Bigg) \quad \text{(by fact 3.8)} \\
&= \mathsf{P}_1(A_1)\mathsf{P}_2(A_2)\cdots\mathsf{P}_n(A_n) = \prod_{i=1}^{n} \mathsf{P}_i(A_i).
\end{aligned}$$

Now we show that (Ω, P) is a probability space. We have $\mathsf{P}(\Omega) = \prod_{i=1}^{n} \mathsf{P}_i(\Omega_i) = \prod_{i=1}^{n} 1 = 1$. Also, for any $A \subseteq \Omega$, we have

$$0 \leqslant \mathsf{P}(A) = \sum_{\omega\in A} \mathsf{P}(\{\omega\}) \leqslant \sum_{\omega\in\Omega} \mathsf{P}(\{\omega\}) = \mathsf{P}(\Omega) = 1.$$

Thus $\mathsf{P}(\wp\,\Omega) \subseteq [0, 1]$, that is, the range of P is some subset of $[0, 1]$. Hence by proposition 1.5, the result follows. □

We now extend the result of proposition 3.7 as follows.

Proposition 3.9. *Consider the probability spaces*

$$\begin{array}{cccccc}
(\Omega_{11}, \mathsf{P}_{11}), & (\Omega_{12}, \mathsf{P}_{12}), & (\Omega_{13}, \mathsf{P}_{13}), & \cdots & (\Omega_{1m_1}, \mathsf{P}_{1m_1}), \\
(\Omega_{21}, \mathsf{P}_{21}), & (\Omega_{22}, \mathsf{P}_{22}), & (\Omega_{23}, \mathsf{P}_{23}), & \cdots & (\Omega_{2m_2}, \mathsf{P}_{2m_2}), \\
(\Omega_{31}, \mathsf{P}_{31}), & (\Omega_{32}, \mathsf{P}_{32}), & (\Omega_{33}, \mathsf{P}_{33}), & \cdots & (\Omega_{3m_3}, \mathsf{P}_{3m_3}), \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
(\Omega_{n1}, \mathsf{P}_{n1}), & (\Omega_{n2}, \mathsf{P}_{n2}), & (\Omega_{n3}, \mathsf{P}_{n3}), & \cdots & (\Omega_{nm_n}, \mathsf{P}_{nm_n}).
\end{array}$$

Let $\Omega := \Omega_{11} \times \cdots \Omega_{1m_1} \times \cdots \times \Omega_{n1} \times \cdots \times \Omega_{nm_n}$*, and define* $\mathsf{P}\colon \wp\,\Omega \to [0,1]$ *similarly to proposition 3.7 (hence* (Ω, P) *is a probability space). For* $i = 1, 2, \ldots, n$*, let* $A'_i \subseteq \Omega_{i1} \times \cdots \times \Omega_{im_1}$*, and define the set*

$$A_i := \{(\omega_{11}, \ldots, \omega_{1m_1}, \ldots, \omega_{n1}, \ldots, \omega_{nm_n}) \in \Omega : (\omega_{i1}, \ldots, \omega_{im_i}) \in A'_i\}.$$

Then the events $A_1, \ldots, A_n$ *in* (Ω, P) *are independent, and moreover, for any* $i = 1, 2, \ldots, n$ *we have*

$$\mathsf{P}(A_i) = \sum_{(\omega_{i1}, \ldots, \omega_{im_i}) \in A'_i} \left(\prod_{j=1}^{m_i} \mathsf{P}_{ij}(\{\omega_{ij}\}) \right).$$

Proof. For $i = 1, 2, \ldots, n$, define $\omega_i := (\omega_{i1}, \ldots, \omega_{im_i})$ and $p_{i\omega_i} := \prod_{j=1}^{m_i} \mathsf{P}_{ij}(\{\omega_{ij}\})$. Let $A := \bigcap_{i=1}^{n} A_i$. Clearly A is the set $\{(\omega_{11}, \ldots, \omega_{1m_1}, \ldots, \omega_{n1}, \ldots, \omega_{nm_n}) \in \Omega : (\omega_{i1}, \ldots, \omega_{im_i}) \in A'_i \text{ for } i = 1, 2, \ldots, n\}$. Thus by the definition of P in the proposition, A is the sum of probabilities $\mathsf{P}(\{(\omega_{11}, \ldots, \omega_{1m_1}, \ldots, \omega_{n1}, \ldots, \omega_{nm_n})\})$ which satisfy $(\omega_{i1}, \ldots, \omega_{im_i}) \in A'_i$ for each $i = 1, 2, \ldots, n$. Also by definition of P, we have

$$\mathsf{P}(\{\omega_{11}, \ldots, \omega_{1m1}, \ldots, \omega_{n1}, \ldots, \omega_{nm_n}\}) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \mathsf{P}(\{\omega_{ij}\}),$$

so

$$\begin{aligned}
\mathsf{P}\left(\bigcap_{i=1}^{n} A_i \right) &= \sum_{(\omega_{11}, \ldots, \omega_{1m_1}) \in A'_1} \cdots \sum_{(\omega_{n1}, \ldots, \omega_{nm_n}) \in A'_n} \mathsf{P}(\{(\omega_{11}, \ldots, \omega_{1m_1}, \ldots, \omega_{n1}, \ldots, \omega_{nm_n})\}) \\
&= \sum_{(\omega_{11}, \ldots, \omega_{1m_1}) \in A'_1} \cdots \sum_{(\omega_{n1}, \ldots, \omega_{nm_n}) \in A'_n} \prod_{i=1}^{n} \prod_{j=1}^{m_i} \mathsf{P}(\{\omega_{ij}\}) \\
&= \sum_{\omega_1 \in A'_1} \cdots \sum_{\omega_n \in A'_n} \prod_{i=1} p_{iw_i} = \sum_{\substack{(\omega_1, \omega_2, \ldots, \omega_n) \in \\ A'_1 \times A'_2 \times \cdots \times A'_n}} \prod_{j=1}^{m_i} \mathsf{P}(\{\omega_{ij}\}) \\
&= \left(\sum_{\omega_1 \in A'_1} p_{1\omega_1} \right) \left(\sum_{\omega_2 \in A'_2} p_{2\omega_2} \right) \cdots \left(\sum_{\omega_n \in A'_n} p_{n\omega_n} \right) \qquad \text{(by fact 3.8)} \\
&= \mathsf{P}(A_1)\mathsf{P}(A_2) \cdots \mathsf{P}(A_n) = \prod_{i=1}^{n} \mathsf{P}(A_i),
\end{aligned}$$

because

$$\begin{aligned}
\sum_{\omega_1 \in A'_1} p_{1\omega_1} &= \sum_{\omega_1 \in A'_1} p_{1\omega_1} \mathsf{P}_{21}(\Omega_{21}) \cdots \mathsf{P}_{nm_n}(\Omega_{mn_n}) \qquad (\text{since } \mathsf{P}_{ij}(\Omega_{ij}) = 1 \text{ for all } i, j) \\
&= \sum_{\omega_1 \in A'_1} p_{1\omega_1} \left(\sum_{\omega_{21} \in \Omega_{21}} \mathsf{P}_{21}(\{\omega_{21}\}) \right) \cdots \left(\sum_{\omega_{nm_n} \in \Omega_{nm_n}} \mathsf{P}_{nm_n}(\{\omega_{nm_n}\}) \right) \\
&= \sum_{\omega_1 \in A'_1} p_{1\omega_1} \sum_{\substack{(\omega_{21}, \ldots, \omega_{nm_n}) \\ \in \Omega_{21} \times \cdots \times \Omega_{nm_n}}} \mathsf{P}_{21}(\{\omega_{21}\}) \cdots \mathsf{P}_{nm_n}(\{\omega_{nm_n}\}) \qquad \text{(by fact 3.8)}
\end{aligned}$$

$$= \sum_{\substack{(\omega_{11},\dots,\omega_{1m_1})\\ \in A_1'}} \sum_{\substack{(\omega_{21},\dots,\omega_{nm_n})\\ \in \Omega_{21}\times\cdots\times\Omega_{nm_n}}} p_{1\omega_1}\mathsf{P}_{21}(\{\omega_{21}\})\cdots\mathsf{P}_{nm_n}(\{\omega_{nm_n}\})$$

$$= \sum_{\substack{(\omega_{11},\dots,\omega_{nm_n})\\ \in A_1}} \mathsf{P}_{11}(\{\omega_{11}\})\cdots\mathsf{P}_{1m_1}(\{\omega_{1m_1}\})\mathsf{P}_{21}(\{\omega_{21}\})\cdots\mathsf{P}_{nm_n}(\{\omega_{nm_n}\})$$

$$= \sum_{\substack{(\omega_{11},\dots,\omega_{nm_n})\\ \in A_1}} \mathsf{P}(\{(\omega_{11},\dots,\omega_{nm_n})\}) = \sum_{\omega\in A_1}\mathsf{P}(\{\omega\}) = \mathsf{P}(A_1),$$

and similarly $\sum_{\omega_i\in A_i'} p_{i\omega_i} = \mathsf{P}(A_i)$ for $i = 1, 2, \dots, n$, which proves the second part of the result. Note that we have also proved the first part of the result for $I = \{1, 2, \dots, n\}$, that is, $\mathsf{P}\big(\bigcap_{i=1}^{n} A_i\big) = \prod_{i=1}^{n}\mathsf{P}(A_i)$ for any $A_i' \subseteq \Omega_{i1}\times\cdots\Omega_{im_i}$ where $i = 1, \dots, n$. We now prove the result for any $I \subseteq [n]$ as in definition 3.4. Indeed, let $I \subseteq [n]$ and define $J := [n] \setminus I$. For all $j \in J$, let us take $A_j' := \Omega_{j1}\times\cdots\times\Omega_{jm_j}$, so that we get $A_j = \Omega$. Then

$$\mathsf{P}\Big(\bigcup_{i\in I} A_i\Big) = \mathsf{P}\Big(\Big(\bigcup_{i\in I} A_i\Big)\cap\Omega\Big) = \mathsf{P}\Big(\Big(\bigcup_{i\in I} A_i\Big)\cap\Big(\bigcup_{j\in J} A_j\Big)\Big) = \mathsf{P}\Big(\bigcup_{i=1}^{n} A_i\Big)$$

$$= \prod_{i=1}^{n}\mathsf{P}(A_i) = \prod_{i\in I}\mathsf{P}(A_i)\prod_{j\in J}\mathsf{P}(A_j) = \prod_{i\in I}\mathsf{P}(A_i),$$

since $\mathsf{P}(A_j) = \mathsf{P}(\Omega) = 1$ for all $j \in J$. □

Remark 3.10. Due to the heavy use of notation, proposition 3.9 may seem difficult to grasp, but this is not the case. Loosely speaking, what this is telling us is that if we carry out a set of independent random experiments and we partition this set, then the probability that something happens in a part (subset) is independent of what happens in another part (a disjoint subset). Moreover, the probability that something happens in a subset of the set of experiments (and 'anything' happens in the other experiments) can be quantified by 'ignoring' the other experiments.

Example 3.11. Consider 10 fair coin tosses. Then $(\Omega_1, \mathsf{P}_1), \dots, (\Omega_{10}, \mathsf{P}_{10})$ each have $\Omega_i = \{H, T\}$ and $\mathsf{P}_i(\{H\}) = \mathsf{P}_i(\{T\}) = \frac{1}{2}$. The product $\Omega = \Omega_1\times\cdots\times\Omega_{10}$ is the set $\{H, T\}^{10}$, and the probability function $\mathsf{P}\colon \wp\omega \to [0, 1]$ is defined as in proposition 3.7.

Consider the events $A_1, A_2, A_3, A_4 \subseteq \Omega$ where:

- A_1 is the event that the 1st, 3rd and 9th coins turn out H.
- A_2 is the event that the 5th and 10th coins turn out H.
- A_3 is the event that the 4th and 6th coins turn out H.
- A_4 is the event that the 2nd, 7th and 8th coins turn out H.

Then $A_1 = \{(\omega_1, \omega_2, \dots, \omega_{10}) : (\omega_1, \omega_3, \omega_9) \in A_1'\}$ where $A_1' = \{(H, H, H)\} \subseteq \Omega_{11}\times\Omega_{12}\times\Omega_{13}$, where $\Omega_{11} = \Omega_1, \Omega_{12} = \Omega_3$ and $\Omega_{13} = \Omega_9$. Therefore

$$\mathsf{P}(A_1) = \sum_{(\omega_1,\omega_3,\omega_9)\in A_1'}\Big(\prod_{j=1}^{3}\mathsf{P}_{ij}(\{\omega_{ij}\})\Big)$$

$$= \sum_{(\omega_1,\omega_3,\omega_9)\in A_1'} \mathsf{P}_{11}(\{\omega_{i1}\})\,\mathsf{P}_{12}(\{\omega_{i2}\})\,\mathsf{P}_{13}(\{\omega_{i3}\})$$
$$= \mathsf{P}_1(\{H\})\,\mathsf{P}_3(\{H\})\,\mathsf{P}_9(\{H\}) = \frac{1}{8}.$$

Similarly we get $\mathsf{P}(A_2) = \mathsf{P}(A_3) = \frac{1}{4}$ and $\mathsf{P}(A_4) = \frac{1}{8}$.

3.1 A Combinatorial Application: Tournaments

Definition 3.12 (Digraph)**.** A *digraph* is a pair (V, D) where V is a set of vertices, and $D \subseteq V \times V$ is a set of ordered pairs of vertices, called arcs.

Example 3.13. In figure 5 we have an ordinary graph on four vertices, joined by edges. In figure 6 we see an example of a digraph. Observe that since arcs are ordered pairs, we interpret $(v_1, v_2) \in D$ to be pointing from v_1 to v_2, and consider it different from the arc (v_2, v_1). In fact, as we see in the figure, both (v_1, v_2) and (v_2, v_1) may appear as distinct arcs in the same digraph. Note that digraphs may also have loops, that is, arcs of the form (v_1, v_1), as indicated in figure 7.

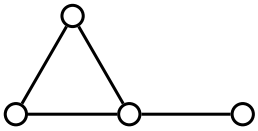

Figure 5: Graph

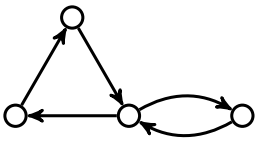

Figure 6: Digraph

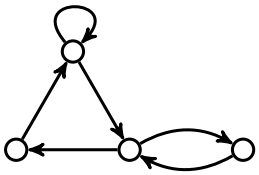

Figure 7: Digraph with loop

Definition 3.14 (Tournament)**.** A *tournament* $T = (V, D)$ is a digraph without loops such that for any distinct $u, v \in V$, exactly one of (u, v) and (v, u) is in D.

Example 3.15. In figure 8, the tournament is invalid because both $(1, 2)$ and $(2, 1)$ are in D. Furthermore, we have the loop $(3, 3) \in D$, which is not allowed. In figure 9, the tournament is invalid because neither $(1, 3)$ nor $(3, 1)$ are in D (the same can be said about $(2, 4)$ and $(4, 2)$). Figure 10 gives us an example of a valid tournament on four vertices.

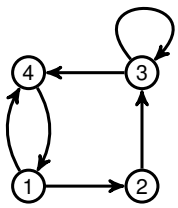

Figure 8: Invalid Tournament

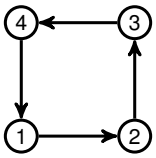

Figure 9: Invalid Tournament

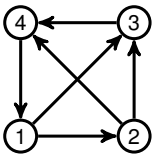

Figure 10: Valid Tournament

Remark 3.16. A tournament can be considered a complete graph with oriented edges.

We use the term 'tournament' here because we can use this type of digraph to represent a series of competitions, where the vertices represent the competitors, and an edge (A, B) represents the fact that competitor A has beaten competitor B.

Definition 3.17 (Property $\mathcal{S}_k$)**.** A tournament $T = (V, D)$ is said to *have property* $\mathcal{S}_k$ if for every k-subset $K \subseteq V$, there exists a vertex $v \in V \setminus K$ such that $(v, w) \in D$ for all $w \in K$.

Example 3.18. The tournament in figure 11 has property $\mathcal{S}_1$, since every 1-subset of V has a vertex which 'beats' all the vertices in the (singleton) set.

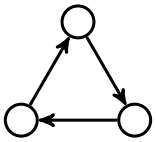

Figure 11: Tournament with property $\mathcal{S}_1$

Is it possible to find a tournament on, say, four vertices, having property $\mathcal{S}_2$? Let us try and construct such an example by attempting to orient the edges of K_4. Let us start with the non-oriented K_4 shown in figure 12. Without loss of generality, suppose 1 beats 2 and 3, shown in figure 13. Now, someone must beat 3 and 4, and it cannot be 1 (otherwise no one beats 1), so it must be 2.

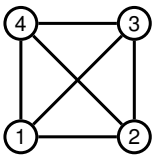

Figure 12: Non-oriented K_4

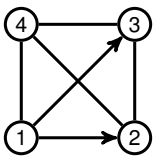

Figure 13: 1 beats 2 and 3

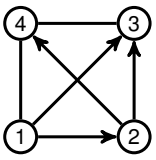

Figure 14: 2 must beats 3 and 4

At this stage, it is impossible for a vertex to beat 1 and 4. Thus it is impossible to construct a tournament with property $\mathcal{S}_2$ on 4 vertices. In fact, the smallest number of vertices required for a tournament to have property $\mathcal{S}_2$ is $n = 7$. Indeed, any 2-subset of vertices of the tournament in figure 15 has a vertex beating both vertices.

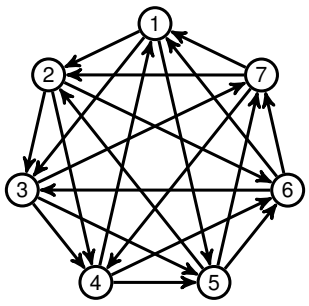

Figure 15: Tournament on $n = 7$ vertices with property $\mathcal{S}_2$

Theorem 3.19 (Erdős, 1963)**.** *If* $\binom{n}{k}(1-2^{-k})^{n-k} < 1$*, then there exists a tournament on* n *vertices with property* $\mathcal{S}_k$*.*

Proof. Consider a tournament T on n vertices such that $\binom{n}{k}(1-2^{-k})^{n-k} < 1$ where k is some fixed positive integer, and for every pair of vertices v and w, let the toss of a coin decide whether we put (v,w) or (w,v) in the arc set D. Let $K \subseteq V$ with $|K| = k$. For each such subset K, let A_K be the event that there is no vertex in $V \setminus K$ that beats all members of K, and for each $v \in V \setminus K$, let $A_{K,v}$ be the event that v does not beat all the members of K. Then $\Omega \setminus A_{K,v}$ is the event that $(v,w) \in D$ for all $w \in K$, and we have $\mathsf{P}(\Omega \setminus A_{K,v}) = \left(\frac{1}{2}\right)^k = 2^{-k}$, and $\mathsf{P}(A_{K,v}) = 1 - 2^{-k}$ by proposition 1.2(iii).

But A_K is the event that $A_{K,v}$ holds for all $v \in V \setminus K$, and note that $A_{K,v}$ and $A_{K,v'}$ are two events which concern restrictions on a subset of experiments which are independent. Thus by proposition 3.9, we get that $\mathsf{P}(A_K) = \prod_{v \in V \setminus K} \mathsf{P}(A_{K,v}) = (1-2^{-k})^{n-k}$, and hence

$$\mathsf{P}\Big(\bigcup_{K \in \binom{V}{k}} A_K\Big) \leqslant \sum_{K \in \binom{V}{k}} \mathsf{P}(A_K) = \binom{n}{k}\mathsf{P}(A_K) = \binom{n}{k}(1-2^{-k})^{n-k} < 1$$

by the hypothesis. Therefore $\mathsf{P}(\Omega \setminus \bigcup_{K \in \binom{V}{k}} A_K) > 0$, so there exists a random tournament on n-vertices having property $\mathcal{S}_k$. □

Remark 3.20. In theorem 3.19, we have that if n is sufficiently large, then the property $\mathcal{S}_k$ occurs with positive probability. In fact, we have that for $n \geqslant 2^k k^2 (\ln 2)(1 + o(1))$,[6] a tournament with property $\mathcal{S}_k$ certainly exists.

4 The Lovász Local Lemma

Notation. Let (Ω, P) be a probability space, and let $A \subseteq \Omega$ be an event. We denote the complement of A in Ω by $\overline{A}$, that is, $\overline{A} := \Omega \setminus A$.

The following proposition is a simple consequence of De Morgan's laws.[7]

Proposition 4.1. *Let* $A_1, \ldots, A_n$ *be events in the probability space* (Ω, P)*. Then* $A_1, \ldots, A_n$ *are independent if and only if* $\overline{A}_1, \ldots, \overline{A}_n$ *are independent.*

Thus if $A_1, A_2, \ldots, A_n$ are independent such that $\mathsf{P}(A_i) < 1$ for all $i = 1, 2, \ldots, n$, then

$$\Omega \setminus \bigcup_{i=1}^n A_i = \overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \overline{A}_i \qquad \text{(by De Morgan's laws)}$$

$$\implies \mathsf{P}\Big(\Omega \setminus \bigcup_{i=1}^n A_i\Big) = \mathsf{P}\Big(\bigcap_{i=1}^n \overline{A}_i\Big) = \prod_{i=1}^n \mathsf{P}(\overline{A}_i) = \prod_{i=1}^n (1 - \mathsf{P}(A_i)) > 0,$$

since each $\mathsf{P}(A_i) < 1$, and therefore the probability that no A_i occurs is positive.

[6] The notation $o(1)$ here denotes *little-o*, a Landau-notation which signifies a quantity which gets asymptotically smaller for larger values of k.

[7] De Morgan's laws for sets: For arbitrary index set I and family of sets $\{A_i\}_{i \in I}$, $\overline{\bigcap_{i \in I} A_i} = \bigcup_{i \in I} \overline{A}_i$, and $\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A}_i$.

Definition 4.2 (Set Independence)**.** Let $A_1, A_2, \dots, A_n \subseteq \Omega$ be events in a probability space (Ω, P). The event $A \subseteq \Omega$ is said to be *independent of* the set of events $\{A_1, A_2, \dots, A_n\}$ if for any $I \subseteq [n]$,

$$\mathsf{P}\big(A \cap \bigcap_{i \in I} A_i\big) = \mathsf{P}(A)\,\mathsf{P}\big(\bigcap_{i \in I} A_i\big).$$

Remark 4.3. This is a weakening of the notion of *independence* as in definition 3.4. In fact, given a set of independent events $\{A_1, A_2, \dots, A_n\}$, each A_i is independent of the set $\{A_1, A_2, \dots, A_n\} \setminus \{A_i\}$, since for any $I \subseteq [n] \setminus \{i\}$,

$$\mathsf{P}\big(A_i \cap \bigcap_{j \in I} A_j\big) = \mathsf{P}\big(\bigcap_{j \in I \cup \{i\}} A_j\big) = \prod_{j \in I \cup \{i\}} \mathsf{P}(A_j) = \mathsf{P}(A_i) \prod_{j \in I} \mathsf{P}(A_j) = \mathsf{P}(A_i)\,\mathsf{P}\big(\bigcap_{j \in I} A_j\big).$$

Theorem 4.4 (Asymmetric Lovász Local Lemma)**.** *Let $A_1, A_2, \dots, A_n \subseteq \Omega$ be events in a probability space (Ω, P). If $S_1, S_2, \dots, S_n \subseteq [n]$ and $x_1, x_2, \dots, x_n \in [0, 1)$ such that for each $i \in [n]$,*

(i) A_i is independent of $\{A_j : j \in [n] \setminus (S_i \cup \{i\})\}$, and

(ii) $\mathsf{P}(A_i) \leqslant x_i \prod_{j \in S_i} (1 - x_j)$,

then

$$\mathsf{P}\big(\bigcap_{i=1}^{n} \overline{A_i}\big) \geqslant \prod_{i=1}^{n} (1 - x_i).$$

We will not give a proof of the Lovász local lemma in this form. We will instead reformulate the result in terms of what are known as *dependency digraphs*, and then prove a special case known as the *symmetric* Lovász Local Lemma.

Definition 4.5 (Dependency Digraph)**.** Let (Ω, P) be a probability space. A *dependency digraph* of the events $A_1, A_2, \dots, A_n \subseteq \Omega$ is a digraph (V, D) where $V = \{A_1, A_2, \dots, A_n\}$ and for $i = 1, \dots, n$, A_i is independent of $\{A_j : j \in [n] \setminus \{i\} \ \wedge \ (A_i, A_j) \notin D\}$.

Remark 4.6. In other words, an event $A \subseteq \Omega$ in a dependency digraph (V, D) is independent of the set of events $B \subseteq \Omega$ where there is no arc from A to B in D. Note that a dependency digraph for a set of given events is not unique. For example, consider the random experiment of tossing two coins simultaneously. Then

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}.$$

Suppose A_1 is the event that the first coin is heads, A_2 is the event that second coin is heads, and A_3 is the event that the first and second coin are the same. Then

$$\mathsf{P}(A_1) = \frac{|\{(H,H),(H,T)\}|}{|\Omega|} = \frac{1}{2}, \qquad \mathsf{P}(A_2) = \frac{|\{(H,H),(T,H)\}|}{|\Omega|} = \frac{1}{2},$$
$$\mathsf{P}(A_3) = \frac{|\{(H,H),(T,T)\}|}{|\Omega|} = \frac{1}{2}.$$

First of all, notice that

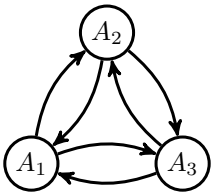

Figure 16

is a valid dependency digraph for the events A_1, A_2 and A_3. The reason being that by definition, it is the *absence* of an arc which claims that events are independent. The existence of an arc does *not* imply anything about the events in question (in particular, an arc $(A, B) \in D$ does not mean that the event A is 'dependent', i.e. not independent, of B). So in fact, the above dependency digraph will work for any 3-set of events in a probability space.

Now let us give some less trivial dependency digraphs. Consider the figures below.

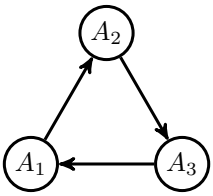

Figure 17

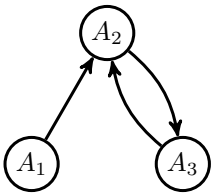

Figure 18

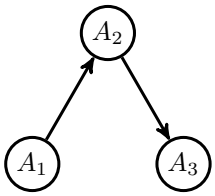

Figure 19

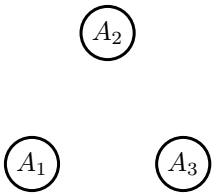

Figure 20

In figure 17, we have that A_1 is independent of $\{A_3\}$, A_2 is independent of $\{A_1\}$, and A_3 is independent of $\{A_2\}$. Let us check whether these claims are true.

$$\mathsf{P}\big(A_1 \cap \bigcap_{i\in\{3\}} A_i\big) = \mathsf{P}\big(A_1 \cap A_3\big) = \frac{|\{(H,H)\}|}{|\Omega|} = \frac{1}{4} = \mathsf{P}(A_1)\mathsf{P}(A_3) = \mathsf{P}(A_1)\,\mathsf{P}\big(\bigcap_{i\in\{3\}} A_i\big),$$

$$\mathsf{P}(A_2 \cap A_1) = \frac{|\{(H,H)\}|}{|\Omega|} = \frac{1}{4} = \mathsf{P}(A_2)\mathsf{P}(A_1),$$

$$\mathsf{P}(A_3 \cap A_2) = \frac{|\{(H,H)\}|}{|\Omega|} = \frac{1}{4} = \mathsf{P}(A_3)\mathsf{P}(A_2).$$

Since all the claims are true, then the dependency digraph in figure 17 is valid for the events A_1, A_2, and A_3.

In figure 18, the claims are that A_1 is independent of $\{A_3\}$, A_2 is independent of $\{A_1\}$, and A_3 is independent of $\{A_1\}$. One can similarly check that these claims are true, and that the dependency digraph in figure 18 is valid.

In figure 19 the claims are that A_1 is independent of $\{A_3\}$, A_2 is independent of $\{A_1\}$, and A_3 is independent of $\{A_1, A_2\}$. The last claim is false:

$$\mathsf{P}(A_3 \cap (A_1 \cap A_2)) = \mathsf{P}(A_1 \cap A_2 \cap A_3) = \frac{|\emptyset|}{|\Omega|} = 0 \neq \frac{1}{8} = \mathsf{P}(A_3)\,\mathsf{P}(A_1 \cap A_2),$$

and therefore the dependency digraph in figure 19 is not valid.

In figure 20, the claim is that each event is independent of all the other two, which by remark 4.3, is equivalent to the events A_1, A_2, A_3 being independent. This is false:

$$\mathsf{P}(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = \mathsf{P}(A_1)\mathsf{P}(A_2)\mathsf{P}(A_3).$$

In terms of dependency digraphs, the Lovász local lemma becomes

Theorem 4.7 (Lovász Local Lemma Revisited)**.** *Let $A_1, A_2, \ldots, A_n \subseteq \Omega$ be events in a probability space (Ω, P). If $A_1, A_2, \ldots A_n$ have dependency digraph (V, D) and $x_1, x_2, \ldots, x_n \in [0, 1)$ such that for each $i \in [n]$, $\mathsf{P}(A_i) \leqslant x_i \prod_{(i,j)\in D} (1 - x_j)$, then*

$$\mathsf{P}\Big(\bigcap_{i=1}^{n} \overline{A_i}\Big) \geqslant \prod_{i=1}^{n} (1 - x_i).$$

Let us now state and prove the special case known as the *symmetric* Lovász Local Lemma. In the proof, we assume the asymmetric Lovász local lemma as it is given above.

Theorem 4.8 (Symmetric Lovász Local Lemma)**.** *Let $A_1, A_2, \ldots, A_n \subseteq \Omega$ be events in a probability space (Ω, P). If $A_1, A_2, \ldots A_n$ have dependency digraph (V, D) with maximum out-degree less than or equal to d, and $\mathsf{P}(A_i) \leqslant \frac{1}{e(d+1)}$ for all $i = 1, \ldots, n$, then*

$$\mathsf{P}\Big(\bigcap_{i=1}^{n} \overline{A}_i\Big) > 0.$$

Proof. Let (V, D) be a dependency digraph of the events $A_1, A_2, \ldots, A_n$ with maximum out-degree at most d. Note that if d is zero, then the events are all independent and the result is trivial. So suppose $d \geqslant 1$. We show that $\mathsf{P}(A_i) \leqslant x_i \prod_{(i,j)\in D} (1 - x_j)$, then the result follows from the Lovász local lemma (theorem 4.7).

For all $i \in [n]$, let $x_i = 1/(1 + d)$. Then

$$\mathsf{P}(A_i) \leqslant \frac{1}{e(d+1)} = \frac{x_i}{e} < \frac{x_i}{\left(1 + \frac{1}{d}\right)^d} = x_i \left(1 - \frac{1}{d+1}\right)^d =\leqslant x_i \prod_{(i,j)\in D} (1 - x_j),$$

since $e = \lim(1 + \frac{1}{n})^n$ and for fixed $i \in V$, the number of arcs $(i, j) \in D$ is at most d. □

The Lovász local lemma leads to the following theorem.

Theorem 4.9 (Erdős & Lovász, 1975)**.** *For any $k \geqslant 9$, every k-uniform, k-regular hypergraph is 2-colourable.*

Note that k-regular here means that every vertex $v \in V$ lies in precisely k edges.

Proof. Let $H = (V, E)$ be a k-regular, k-uniform hypergraph with edges $e_1, \ldots, e_n$. For every $v \in V$, let the toss of a coin determine the colouring (say, red/blue) of v, and

for every $i \in [n]$, let A_i be the event that the edge e_i is monochromatic. Since H is k-uniform, $\mathsf{P}(A_i) = |A_i|/|\Omega| = 1/2^{k-1} = 2^{1-k}$.

For each $i \in [n]$, each vertex of the edge e_i is in precisely $k-1$ other edges of H, and every e_i contains precisely k vertices. Thus there are at most $d = k(k-1)$ edges in $E \setminus \{e_i\}$ which intersect e_i (in other words, $e_i \cap e_j \neq \emptyset$). Indeed, define $S_i := \{j \in [n] \setminus \{i\} : e_i \cap e_j = \emptyset\}$. Then $|S_i| \geqslant n - d - 1$, and for any $s \in S_i$, the set of k experiments under consideration for A_s is disjoint from that of A_i. So for any $I \subseteq S_i$, the set of experiments under consideration for $\bigcap_{s\in I} A_s$ is disjoint from that of A_i. Thus by proposition 3.9,

$$\mathsf{P}\big(A_i \cap \bigcap_{s\in I} A_s\big) = \mathsf{P}(A_i)\,\mathsf{P}\big(\bigcap_{s\in I}\big)A_s,$$

that is, A_i is independent of the events $\{A_j : j \in S_i\}$. Now let (W, D) be a digraph such that $W = \{A_1, \dots, A_n\}$ and for any $i, j \in [n]$, $(A_i, A_j) \notin D \iff j \in \{i\} \cup S_i$. Thus (W, D) is a dependency digraph of $A_1, \dots, A_n$, and the maximum out-degree is at most d, since for all $i \in [n]$, $|\{i\} \cup S_i| \geqslant n - d$. So there are at most d arcs from A_i to A_j for each i. Also,

$$\begin{aligned}\mathsf{P}(A_i) = 2^{1-k} = \frac{1}{2^{k-1}} &< \frac{1}{e(k(k-1)+1)} \quad (\text{since } 2^{k-1} > e(k(k-1)+1) \text{ for } k \geqslant 0)\\ &= \frac{1}{e(d+1)}.\end{aligned}$$

Applying the symmetric Lovász local lemma, we get that

$$\mathsf{P}\big(\bigcap_{i=1}^{n} \bar{A}_i\big) > 0,$$

and therefore the probability that no A_i occurs is positive. Thus there exists $\omega \in \Omega \setminus \bigcap_{i=1}^{n} A_i$ where $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ whose entries are outcomes of $|V|$ experiments not in any A_i. □

5 Random Variables and Expectation

5.1 Random Variables

Definition 5.1 (Random Variable)**.** Let (Ω, P) be a probability space. A *random variable* X on (Ω, P) is a function $X \colon \Omega \to \mathbb{R}$.

For $x \in \mathbb{R}$, we denote the probability $\mathsf{P}(\{\omega \in \Omega : X(\omega) = x\})$ by $\mathsf{P}_X(x)$ or $\mathsf{P}(X = x)$. More generally, if $\sim \subseteq \mathbb{R} \times S$ denotes a relation between $\mathbb{R}$ and some set S, then for $s \in S$, we denote $\mathsf{P}(\{\omega \in \Omega : X(\omega) \sim s\})$ by $\mathsf{P}_X(\sim s)$ or $\mathsf{P}(X \sim s)$.

Example 5.2. Consider the experiment of tossing two dice. We have

$$\Omega = \{(⚀, ⚀), (⚀, ⚁), \dots, (⚅, ⚅)\}.$$

We can define a random variable $X \colon \Omega \to \mathbb{R}$ to be the sum of the outcomes, $X(a, b) = a + b$. It is not difficult to check that the range $X(\Omega) = \{2, 3, 4, \dots, 12\}$, and these now have unequal probability. For example,

$$\mathsf{P}(X = 2) = \frac{|\{(⚀, ⚀)\}|}{|\Omega|} = \frac{1}{36},$$

$$\mathsf{P}(X=7)=\frac{|\{(\boxdot,\boxplus),(\boxdot,\boxplus),(\boxdot,\boxplus),(\boxdot,\boxplus),(\boxdot,\boxplus),(\boxplus,\boxdot)\}|}{|\Omega|}=\frac{1}{6}.$$

Another example,

$$\mathsf{P}(X\leqslant 11)=1-\mathsf{P}(X=12)=1-\frac{|\{(\boxplus,\boxplus)\}|}{|\Omega|}=\frac{35}{36}.$$

It is not hard to show that for all $x\in X(\Omega)$, we have

$$\mathsf{P}(X=x)=\frac{6-|x-7|}{36},$$

plotted in figure 21.

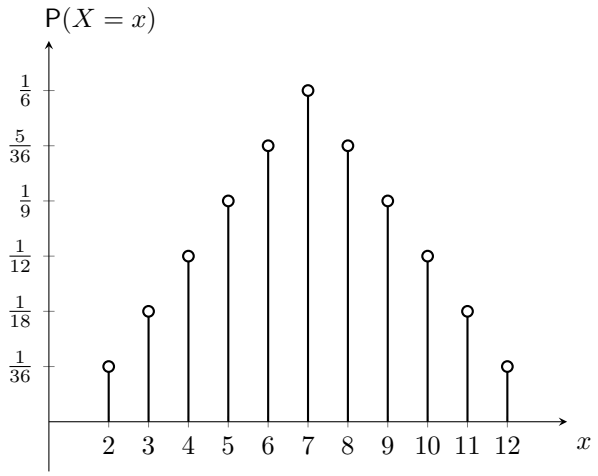

Figure 21: Plot of $\mathsf{P}(X=x)$ for $x\in X(\Omega)$

Proposition 5.3. *Let X be a random variable on the probability space (Ω,P), and let $X(\Omega)=\{x_1,\dots,x_n\}$. For $i=1,2,\dots,n$, let $A_i:=\{\omega\in\Omega:X(\omega)=x_i\}$. Then $A_1,A_2,\dots A_n$ partition Ω.*

Proof. Since X is a function, then for any $\omega\in\Omega$ there exists $i\in[n]$ such that $X(\omega)=x_i$. It follows that $\Omega=A_1\cup A_2\cup\dots\cup A_n$, and that $A_1,A_2,\dots,A_n$ are pairwise disjoint. □

Proposition 5.4. *Let X be a random variable on the probability space (Ω,P), and let $Y\subseteq\mathbb{R}$ be countable. Then*

$$\mathsf{P}(X\in Y)=\sum_{y\in Y}\mathsf{P}(X=y).$$

Proof. We have

$$\mathsf{P}(X\in Y)=\mathsf{P}(\{\omega\in\Omega:X(\omega)\in Y\})=\mathsf{P}\Big(\bigcup_{y\in Y}\{\omega\in\Omega:X(\omega)=y\}\Big)$$

$$= \sum_{y \in Y} \mathsf{P}(\{\omega \in \Omega : X(\omega) = y\}) \qquad \text{(disjoint)}$$
$$= \sum_{y \in Y} \mathsf{P}(X = y),$$

as required. □

Proposition 5.5. *Let X be a random variable on the probability space (Ω, P). Then*

$$\mathsf{P}(X \in X(\Omega)) = \sum_{x \in X(\Omega)} \mathsf{P}(X = x) = 1.$$

Proof. We have

$$\begin{aligned}
\mathsf{P}(X \in X(\Omega)) &= \mathsf{P}(\{\omega \in \Omega : X(\omega) \in X(\Omega)\}) \\
&= \sum_{x \in X(\Omega)} \mathsf{P}(X = x) && \text{(by proposition 5.4)} \\
&= \sum_{i=1}^{n} \mathsf{P}(A_i) && \text{(by proposition 5.3)} \\
&= \mathsf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) && \text{(since the } A_i \text{ are pairwise disjoint)} \\
&= \mathsf{P}(\Omega) = 1,
\end{aligned}$$

as required. □

Example 5.6. Let Ω be the set of all eight digit binary strings, that is, $\Omega = \{0,1\}^8$. Suppose each string occurs with equal probability, so that for any $\omega \in \Omega$,

$$\mathsf{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{2^8}.$$

Now define the random variable X to be the number of 1's in the string, that is, for $\omega = (\omega_1, \omega_2, \ldots, \omega_8) \in \Omega$,

$$X(\omega) = \sum_{i=1}^{8} \omega_i.$$

Clearly $X(\Omega) = \{0, 1, \ldots, 8\}$ and $\mathsf{P}(X = x) = \frac{1}{2^8}\binom{8}{x}$. Also, we get that

$$\mathsf{P}(X \leqslant x) = \mathsf{P}(X \in \{0, 1, \ldots, x\}) = \sum_{i=0}^{x} \binom{8}{i} \frac{1}{2^8},$$

by proposition 5.4.

Definition 5.7 (Compound Random Variable)**.** Let $X_1, X_2, \ldots, X_n$ be random variables defined on the probability space (Ω, P), that is, $X_i \colon \Omega \to \mathbb{R}$ for $i = 1, 2, \ldots, n$. Let $g \colon \mathbb{R}^n \to \mathbb{R}$ be a function. By $X = g(X_1, X_2, \ldots, X_n)$, it is meant that the random variable $X \colon \Omega \to \mathbb{R}$ is defined by

$$X(\omega) = g(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$$

for all $\omega \in \Omega$.

Example 5.8. A simple example: given n random variables $X_1, X_2, \dots X_n$, we can define $X : \Omega \to \mathbb{R}$ to be their sum:

$$X = X_1 + X_2 + \cdots + X_n.$$

In other words, $X(\omega) = X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)$ for all $\omega \in \Omega$. Indeed, suppose $n = 2$, let $\Omega = \{1, 2, 3\}$ and $\mathsf{P}(\{1\}) = \mathsf{P}(\{2\}) = \frac{1}{4}$ and $\mathsf{P}(\{3\}) = \frac{1}{2}$. Also, let:

$$\begin{aligned} X_1(1) &= 0 & X_1(2) &= 0 & X_1(3) &= 2 \\ X_2(1) &= 1 & X_2(2) &= 0 & X_2(3) &= -1 \end{aligned}$$

We determine the probabilities $P(X = 0)$, $P(X = 1)$, $P(X = 2)$ and $P(X = 3)$. First of all,

$$\begin{aligned} X(1) &= X_1(1) + X_2(1) = 1 \\ X(2) &= X_1(2) + X_2(2) = 0 \\ X(3) &= X_1(3) + X_2(3) = 1 \end{aligned}$$

Therefore $\mathsf{P}(X = 0) = \mathsf{P}(\{2\}) = \frac{1}{4}$, $\mathsf{P}(X = 1) = \mathsf{P}(\{1, 3\}) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$, and $\mathsf{P}(X = 2) = \mathsf{P}(X = 3) = 0$.

5.2 The Expected Value of a Random Variable

Definition 5.9 (Expected Value)**.** Let $X : \Omega \to \mathbb{R}$ be a random variable on (Ω, P). The *expected value of* X, denoted $\mathsf{E}[X]$, is defined by

$$\mathsf{E}[X] = \sum_{x \in X(\Omega)} x \, \mathsf{P}(X = x).$$

Example 5.10. Consider the experiment of tossing two dice, and let the random variable X be the sum of the result. We have $X(\Omega) = \{2, 3, \dots, 12\}$, and $\mathsf{P}(X = x) = \frac{6 - |x-7|}{36}$ by example 5.2. Then

$$\mathsf{E}[X] = \sum_{x \in X(\Omega)} x \, \mathsf{P}(X = x) = \sum_{i=2}^{12} x \, \frac{6 - |x - 7|}{36} = 7.$$

Indeed, as *expected*, the expected value of X is indeed 7, since it occurs with the highest probability.

Proposition 5.11. *If X is a random variable on (Ω, P), then there exist $\omega, \omega' \in \Omega$ such that*

$$X(\omega) \leqslant \mathsf{E}[X] \leqslant X(\omega').$$

Proof. Suppose for contradiction that $X(\omega) > \mathsf{E}[X]$ for all $\omega \in \Omega$. Then $x > \mathsf{E}[X]$ for all $x \in X(\Omega)$, and therefore

$$\mathsf{E}[X] = \sum_{x \in X(\Omega)} x \, \mathsf{P}(X = x) > \sum_{x \in X(\Omega)} \mathsf{E}[X] \, \mathsf{P}(X = x) = \mathsf{E}[X] \sum_{x \in X(\Omega)} P(X = x) = \mathsf{E}[X]$$

by proposition 5.5, a contradiction. Similarly if we assume $X(\omega) < \mathsf{E}[X]$ for all $\omega \in \Omega$, we obtain another contradiction, so the result follows. □

Proposition 5.12 (Law of Total Probability)**.** *Let* (Ω, P) *be a probability space and let* $A_1, A_2, \ldots, A_n$ *partition* Ω*. then for any* $A \subseteq \Omega$*,*

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(A \cap A_i).$$

Proof. $\mathsf{P}(A) = \mathsf{P}(A \cap \Omega) = \mathsf{P}\big(A \cap \bigcup_{i=1}^{n} A_i\big) = \mathsf{P}\big(\bigcup_{i=1}^{n} (A \cap A_i)\big) = \sum_{i=1}^{n} \mathsf{P}(A \cap A_i)$, since the A_i are disjoint. □

Remark 5.13. If $\mathsf{P}(A_i) > 0$ for all $i = 1, 2, \ldots, n$, then

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(A \cap A_i) = \sum_{i=1}^{n} \mathsf{P}(A_i)\mathsf{P}(A|A_i).$$

Notation. If $X_1, X_2, \ldots, X_n$ are random variables on (Ω, P), then we abbreviate

$$\mathsf{P}(\{\omega \in \Omega : X_1(\omega) = x_1 \ \wedge \ X_2(\omega) = x_2 \ \wedge \cdots \wedge \ X_n(\omega) = x_n\})$$

to $\mathsf{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$.

Corollary 5.14. *If* X *and* Y *are random variables on* (Ω, P)*, then*

$$\mathsf{P}(X = x) = \sum_{y \in Y(\Omega)} \mathsf{P}(X = x, Y = y).$$

Proof. Let the range $Y(\Omega) = \{y_1, y_2 \ldots, y_n\}$, let $A = \{\omega \in \Omega : X(\omega) = x\}$, and for $i = 1, 2, \ldots, n$, let $A_i = \{\omega \in \Omega : Y(\omega) = y_i\}$. Then

$$\begin{aligned} P(X = x) = \mathsf{P}(A) &= \sum_{i=1}^{n} \mathsf{P}(A \cap A_i) \qquad \text{(by proposition 5.3 and proposition 5.12)} \\ &= \sum_{i=1}^{n} \mathsf{P}(\{\omega \in \Omega : X(\omega) = x \ \wedge \ Y(\Omega) = y_i\}) \\ &= \sum_{y \in Y(\Omega)} \mathsf{P}(X = x, Y = y), \end{aligned}$$

as required. □

Theorem 5.15 (Linearity of Expectation)**.** *If* $X_1, X_2, \ldots, X_n$ *are random variables on a probability space* (Ω, P)*, then*

$$\mathsf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathsf{E}[X_i].$$

Proof. By induction on n. For the base case, consider the case $n = 2$. Let $X := X_1 + X_2$, that is, $X : \Omega \to \mathbb{R}$ such that $X(\omega) = X_1(\omega) + X_2(\omega)$ for all $\omega \in \Omega$. We need to show

that $\mathsf{E}[X] = \mathsf{E}[X_1] + \mathsf{E}[X_2]$. Clearly the range $X(\Omega)$ is given by $\{X_1(\omega) + X_2(\omega) : \omega \in \Omega\}$ and

$$\begin{aligned}\mathsf{P}(X = x) &= \mathsf{P}(\{\omega \in \Omega : X_1(\omega) + X_2(\omega) = x\}) \\ &= \mathsf{P}\Big(\bigcup_{i \in X_1(\Omega)} \{\omega \in \Omega_1 : X_1(\omega) = i, X_2(\Omega) = x - i\}\Big) \\ &= \sum_{i \in X_1(\Omega)} \mathsf{P}(\{\omega \in \Omega_1 : X_1(\omega) = i, X_2(\Omega) = x - i\}),\end{aligned}$$

since the sets $\{\omega \in \Omega_i : X_1(\omega) = i, X_2(\omega) = x - i\}$ are pairwise disjoint for distinct $i \in X_1(\Omega)$ (since X_1 is a function). Therefore $\mathsf{P}(X = x) = \sum_{i \in X_1(\Omega)} \mathsf{P}(X_1 = i, X_2 = x - i)$. Hence

$$\begin{aligned}\mathsf{E}[X] &= \sum_{x \in X(\Omega)} x\, \mathsf{P}(X = x) \\ &= \sum_{x \in X(\Omega)} \Big(x \sum_{i \in X_1(\Omega)} \mathsf{P}(\{\omega \in \Omega_1 : X_1(\omega) = i, X_2(\Omega) = x - i\})\Big) \\ &= \sum_{i \in X_1(\Omega)} \sum_{x \in X(\Omega) \,:\, x - i \in X_2(\Omega)} x\, \mathsf{P}(X_1 = i, X_2 = x - i) \\ &\qquad\qquad \text{(since } \mathsf{P}(X_1 = i, X_2 = x - i) = 0 \text{ if } x - i \notin X_2(\Omega)\text{)} \\ &= \sum_{i \in X_1(\Omega)} \sum_{\substack{j = x - i \\ \text{for some } x \in X(\Omega)}} (i + j)\, \mathsf{P}(X_1 = i, X_2 = j)\end{aligned}$$

(Justification: suppose $j \in X_2(\Omega)$ is such that there does not exists $x \in X(\Omega)$ with $j = x - i$. Then there does not exist $\omega \in \Omega$ such that $j = X(\omega) - i$, that is, $X_1(\omega) + X_2(\Omega) = i + j$. Therefore $\{\omega \in \Omega : X_1(\omega) = i \ \wedge \ X_2(\omega) = j\} = \emptyset$, so $\mathsf{P}(X_1 = i, X_2 = j) = 0$.)

$$\begin{aligned}\implies \mathsf{E}[X] &= \sum_{i \in X_1(\Omega)} \sum_{j \in X_2(\Omega)} i\, \mathsf{P}(X_1 = i, X_2 = j) + \sum_{i \in X_1(\Omega)} \sum_{j \in X_2(\Omega)} j\, \mathsf{P}(X_1 = i, X_2 = j) \\ &= \sum_{i \in X_1(\Omega)} \sum_{j \in X_2(\Omega)} i\, \mathsf{P}(X_1 = i, X_2 = j) + \sum_{j \in X_2(\Omega)} \sum_{i \in X_1(\Omega)} j\, \mathsf{P}(X_2 = j, X_1 = i) \\ &= \sum_{i \in X_1(\Omega)} \Big(i \sum_{j \in X_2(\Omega)} \mathsf{P}(X_1 = i, X_2 = j)\Big) + \sum_{j \in X_2(\Omega)} \Big(j \sum_{i \in X_1(\Omega)} \mathsf{P}(X_2 = j, X_1 = i)\Big) \\ &= \sum_{i \in X_1(\Omega)} i\, \mathsf{P}(X_1 = i) + \sum_{j \in X_2(\Omega)} j\, \mathsf{P}(X_2 = j) \qquad \text{(by corollary 5.14)} \\ &= \mathsf{E}[X_1] + \mathsf{E}[X_2].\end{aligned}$$

So the theorem is true for when $n = 2$. Now for $n > 2$, define $Y := X_1 + \cdots + X_{n-1}$. Then $\mathsf{E}[Y + X_n] = \mathsf{E}[Y] + \mathsf{E}[X_n]$ by the above, and by the inductive hypothesis, $\mathsf{E}[Y] = \sum_{i=1}^{n-1} \mathsf{E}[X_i]$. Therefore

$$\mathsf{E}\Big[\sum_{i=1}^{n} X_i\Big] = \mathsf{E}[Y + X_n] = \mathsf{E}[Y] + \mathsf{E}[X_n] = \sum_{i=1}^{n-1} \mathsf{E}[X_i] = \sum_{i=1}^{n} \mathsf{E}[X_i].$$

as required. □

Proposition 5.16. *If X is a random variable on (Ω, P) and $g\colon \mathbb{R} \to \mathbb{R}$ is an injective function, then*

$$\mathsf{E}[g(X)] = \sum_{x \in X(\Omega)} g(x)\, \mathsf{P}(X = x).$$

Proof. Let $Y = g(X)$. So Y is a random variable on (Ω, P) such that $Y(\omega) = g(X(\omega))$ for all $\omega \in \Omega$. Since g is injective, it follows that if $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$, then the image $Y(\Omega) = \{g(x_1), g(x_2), \ldots, g(x_n)\}$, and so

$$\begin{aligned}
\mathsf{P}(Y = y_i) &= \mathsf{P}(\{\omega \in \Omega : Y(\omega) = y_i\}) \\
&= \mathsf{P}(\{\omega \in \Omega : g(X(\omega)) = g(x_i)\}) \\
&= \mathsf{P}(\{\omega \in \Omega : X(\omega) = x_i\}) \qquad \text{(since } g \text{ is injective)} \\
&= \mathsf{P}(X = x_i),
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathsf{E}[Y] = \sum_{y \in Y(\Omega)} y\, \mathsf{P}(Y = y) &= \sum_{i=1}^{n} y_i\, \mathsf{P}(Y = y_i) \\
&= \sum_{i=1}^{n} g(x_i)\, \mathsf{P}(X = x_i) = \sum_{x \in X(\Omega)} g(x) \mathsf{P}(X = x),
\end{aligned}$$

as required. □

Proposition 5.17. *If X is a random variable on (Ω, P) and $a, b \in \mathbb{R}$, then*

$$\mathsf{E}[aX + b] = a\mathsf{E}[X] + b.$$

Proof. This follows from the previous proposition 5.16 with $g(x) = ax + b$. Indeed,

$$\mathsf{E}[aX+b] = \sum_{x \in X(\Omega)} (ax+b)\mathsf{P}(X = x) = a \sum_{x \in X(\Omega)} x\, \mathsf{P}(X = x) + b \sum_{x \in X(\Omega)} \mathsf{P}(X = x) = a\, \mathsf{E}[X] + b,$$

by proposition 5.5. □

Theorem 5.18. *If $X_1, X_2, \ldots, X_n$ are random variables on (Ω, P) and $a_1, a_2, \ldots, a_n \in \mathbb{R}$, then*

$$\mathsf{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i\, \mathsf{E}[X_i].$$

Proof. This follows from theorem 5.15 and the previous proposition 5.17 with $a = a_i$ and $b = 0$. Indeed,

$$\mathsf{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} \mathsf{E}[a_i X_i] = \sum_{i=1}^{n} a_i\, \mathsf{E}[X_i],$$

as required. □

5.3 Hamiltonian Paths in Tournaments

Recall that in a digraph (V, D), a *Hamiltonian path* is a directed path traversing all the vertices of the digraph without repetition. For example, in figure 22 no path traversing each vertex without repetition exists. In figure 23 on the other hand, we may traverse all the vertices without repetition in the order $1 \to 3 \to 4 \to 2$.

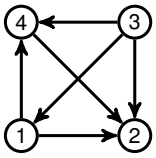

Figure 22: No Hamiltonian path exists

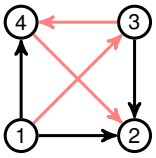

Figure 23: A Hamiltonian path exists

Theorem 5.19 (Szele, 1943)**.** *For any integer n, there exists a tournament on n vertices with at least $2^{1-n}\, n!$ Hamiltonian paths.*

Proof. Let (Ω, P) be the finite probability space where Ω consists of all tournaments T on the vertex set $V = [n]$, and $\mathsf{P}(\{T\}) = 1/|\Omega| = 1/2^{\binom{n}{2}}$. Let $\mathcal{P}$ be the set of all Hamiltonian paths on V. For any $P \in \mathcal{P}$, we will define an *indicator variable*[8] $X_P \colon \Omega \to \{0, 1\}$ by

$$X_P(T) \coloneqq \begin{cases} 1 & \text{if } P \text{ is a valid path is in } T \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $X_P(T) = 1$ if the arcs $(x_1, x_2), (x_2, x_3), \ldots, (x_{n-1}, x_n)$ forming the path P are all present in the arc set D of $T = (V, D)$. How many tournaments T contain P? Being a Hamiltonian path, P must consist of precisely $n - 1$ arcs, leaving $\binom{n}{2} - (n-1)$ edges free to be oriented either way, giving rise to $2^{\binom{n}{2}-(n-1)}$ tournaments containing P, and therefore $\mathsf{P}(X_P = 1) = 2^{\binom{n}{2}-(n-1)}/2^{\binom{n}{2}} = 2^{1-n}$. Now let $X \colon \Omega \to \mathbb{R}$ be the random variable given by $X \coloneqq \sum_{P \in \mathcal{P}} X_P$. By theorem 5.15, we have

$$\mathsf{E}[X] = \sum_{P \in \mathcal{P}} \mathsf{E}[X_P] = \sum_{P \in \mathcal{P}} \mathsf{P}(X_P = 1) = \sum_{P \in \mathcal{P}} 2^{1-n} = |\mathcal{P}| 2^{1-n}.$$

Now each $P = \{(x_1, x_2), (x_2, x_3), \ldots, (x_{n-1}, x_n)\} \in \mathcal{P}$ can be represented uniquely by the permutation $(x_1, x_2, \ldots, x_n)$ of the vertices in V and vice-versa, so $|\mathcal{P}| = n!$ (remember $\mathcal{P}$ contains all possible Hamiltonian paths on n vertices). Hence $\mathsf{E}[X] = 2^{1-n}\, n!$, and by proposition 5.11, there exists a tournament $T' \in \Omega$ such that $X(T') \geqslant 2^{1-n}\, n!$, that is,

$$\sum_{P \in \mathcal{P}} X_P(T') \geqslant 2^{1-n}\, n!$$

So there exist at least $2^{1-n}\, n!$ paths $P \in \mathcal{P}$ such that P is contained in T'. □

[8] All this means is a variable which takes on the value 1 if something is true, and 0 otherwise.

5.4 Large Bipartite Subgraphs

Theorem 5.20. *Let $G = (V, E)$ be a graph on n vertices and m edges. Then G contains a bipartite subgraph with at least $m/2$ edges.*

Note. In the following proof, instead of defining X_{ij} on subsets U of V, we instead define it on what we will call the *characteristic vectors* of these subsets; by which we mean n-vectors which have a 1 in entry i if $v_i \in U$, and 0 otherwise. As we shall see, this enables us to apply proposition 3.9.

Proof. Let $G = (V, E)$ where $V = [n]$. We set up the following n independent random experiments. For each $i \in V$, let (Ω_i, P_i) be given by $\Omega_i = \{0, 1\}$ and $\mathsf{P}_i(\{0\}) = \mathsf{P}_i(\{1\}) = \frac{1}{2}$, and define (Ω, P) as in proposition 3.7. For each edge $\{i, j\} \in E$, let $X_{ij} \colon \Omega \to \{0, 1\}$ be the random variable such that, for any $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega$,

$$X_{ij} := \begin{cases} 1 & \text{if } (\omega_i, \omega_j) \in \{(0,1), (1,0)\} \\ 0 & \text{otherwise.} \end{cases}$$

$\mathsf{P}(X_{ij} = 1)$ is the probability that $(\omega_i, \omega_j) = (1, 0)$ or $(\omega_i, \omega_j) = (0, 1)$, which is therefore $\mathsf{P}_i(\{1\})\mathsf{P}_j(\{0\}) + \mathsf{P}_i(\{0\})\mathsf{P}_j(\{1\}) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$ by proposition 3.9, that is, $\mathsf{P}(X_{ij} = 1) = \frac{1}{2}$. Now define the random variable $X \colon \Omega \to \mathbb{R}$ by $X := \sum_{\{i,j\} \in E} X_{ij}$. By theorem 5.15, we have

$$\mathsf{E}[X] = \sum_{\{i,j\} \in E} \mathsf{E}[X_{ij}] = \sum_{\{i,j\} \in E} \mathsf{P}(X_{ij} = 1) = \sum_{\{i,j\} \in E} \frac{1}{2} = \frac{m}{2}.$$

By proposition 5.11, there exists $\omega' = (\omega'_1, \omega'_2, \ldots, \omega'_n) \in \Omega$ such that $X(\omega') \geqslant \frac{m}{2}$, that is,

$$\sum_{\{i,j\} \in E} X_{ij}(w') \geqslant \frac{m}{2}.$$

This means that there exists a subset $D \subseteq E$ of size $\geqslant \frac{m}{2}$ such that $X_{i,j}(w') = 1$ for all $\{i, j\} \in D$, that is, $(w'_i, w'_j) = (1, 0)$ or $(0, 1)$ for all $\{i, j\} \in D$. Therefore by setting $V_1 = \{i \in [n] : \omega'_i = 1\}$ and $V_2 = V \setminus V_1 = \{i \in [n] : \omega'_i = 0\}$, we observe that all the edges $\{i, j\} \in D$ have one vertex in V_1 and the other vertex in V_2. So the edges in D are on th bipartite subgraph of G with partite sets V_1 and V_2. □

5.5 Independent Sets and Turán's Theorem

The following is an important graph-theoretical result proven independently by Caro (1979) and Wei (1981). Recall that in a graph $G = (V, E)$, an *independent set* is a subset $U \subseteq V$ such that no vertex in U is adjacent to any other vertex in U. For example, the red vertices in the graph of figure 24 form an independent set. In other words, an independence set is an induced subgraph with an empty vertex set. Given a graph G, we denote the size of a *largest* independent set by $\alpha(G)$, called the *independence number* of G.

Theorem 5.21. *For any graph $G = (V, E)$,*

$$\alpha(G) \geqslant \sum_{v \in V} \frac{1}{\deg v + 1}.$$

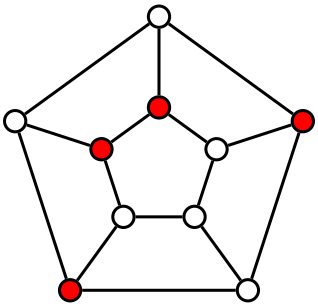

Figure 24

Proof. Let $V = [n]$, and let (Ω, P) be the probability space where $\Omega = S_n$, the set of all permutations of V, and $\mathsf{P}(\{\sigma\}) = 1/|\Omega| = 1/n!$ for all $\sigma \in \Omega$. For each permutation $\sigma \in \Omega$, define

$$I_\sigma := \{v \in V : v \text{ preceeds all its neighbours in } \sigma\}.$$

In mathematical notation, $I_\sigma = \{v \in V : \sigma(v) < \sigma(w) \text{ for all } w \in N(w)\}$ where $N(v)$ denotes the set of neighbours of v. For each $v \in V$, let $X_v \colon \Omega \to \mathbb{R}$ be the random variable defined by

$$X_v(\sigma) := \begin{cases} 1 & \text{if } v \in I_\sigma \\ 0 & \text{otherwise} \end{cases} \qquad \text{for each } \sigma \in \Omega.$$

Let $A_v = \{\sigma \in \Omega : v \text{ preceeds all its neighbours in } \sigma\}$. As we have seen in the proof of theorem 2.22, $|A_v| = \binom{n}{1+\deg v} 1! \deg v! (n - 1 - \deg v)!$. Now

$$\mathsf{E}[X_v] = \mathsf{P}(X_v = 1) = \mathsf{P}(\{\sigma \in \Omega : v \in I_\sigma\}) = \mathsf{P}(A_v) = \frac{|A_v|}{|\Omega|} = \frac{1}{1 + \deg v}.$$

Let $X \colon \Omega \to \mathbb{R}$ be the random variable $X := \sum_{v \in V} X_v$. By theorem 5.15,

$$\mathsf{E}[X] = \sum_{v \in V} \mathsf{E}[X_v] = \sum_{v \in V} \frac{1}{1 + \deg v}.$$

By proposition 5.11, there exists a $\sigma' \in \Omega$ such that $X(\sigma') \geqslant \sum_{v \in V} \frac{1}{1+\deg v}$, that is, $\sum_{v \in V} X_v(\sigma') \geqslant \sum_{v \in V} \frac{1}{1+\deg v}$. Now $|I_{\sigma'}| = \sum_{v \in V} X_v(\sigma')$, so $|I_{\sigma'}| \geqslant \sum_{v \in V} \frac{1}{1+\deg v}$. Suppose $\{x, y\} \in E$ for some distinct $x, y \in I_{\sigma'}$. By definition of $I_{\sigma'}$, x precedes y in σ' and y precedes x in σ', a contradiction. So $I_{\sigma'}$ is an independent set and hence $\alpha(G) \geqslant |I_{\sigma'}|$. □

We now use this result to prove a classical theorem known as Turán's theorem. The following lemma is also needed, which can easily be proved by induction on n.

Lemma 5.22. *Let $x_1, x_2, \ldots, x_n > 0$ be real numbers, and let $\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i$ (the mean). Then*

$$\sum_{i=1}^{n} \frac{1}{x_i} \geqslant \frac{n}{\bar{x}}.$$

Theorem 5.23. *Let $G = (V, E)$ be a graph on n vertices and m edges. Then*

$$\alpha(G) \geqslant \frac{n^2}{n + 2m}.$$

Proof. Let $d \coloneqq \frac{1}{n} \sum_{v \in V} (1 + \deg v) = \frac{1}{n}(n + \sum_{v \in V} \deg v) = 1 + \frac{1}{n} \sum_{v \in V} \deg v$. By the handshaking lemma, $d = 1 + \frac{2m}{n} = \frac{n+2m}{n}$. By theorem 5.21 and lemma 5.22, $\alpha(G) \geqslant \sum_{v \in V} \frac{1}{1+\deg v} \geqslant \frac{n}{d}$. So $\alpha(G) \geqslant \frac{n^2}{n+2m}$. □

Corollary 5.24 (Turán, 1941)**.** *Let $G = (V, E)$ be a graph with n vertices and m edges such that G contains no complete graph K_k as a subgraph. Then*

$$m \leqslant \frac{n^2}{2} \left(\frac{k-2}{k-1} \right).$$

Proof. Let $G' = (V, \binom{V}{2} \setminus E)$ be the complement of G. So G' has $m' = \binom{n}{2} - m$ edges. Then by theorem 5.23, G' has an independent set I such that

$$|I| \geqslant \frac{n^2}{n + 2(\binom{n}{2} - m)} = \frac{n^2}{n + n(n-1) - 2m} = \frac{n^2}{n^2 - 2m}.$$

This means that G has a complete graph of order $\geqslant \frac{n^2}{n^2-2m}$ on the vertex set I. Since G contains no complete subgraph of order k, we get that $k \geqslant \frac{n^2}{n^2-2m} + 1$. Re-arranging the inequality gives the result. □

PART II

EXTREMAL COMBINATORICS

6 Introduction

The result in theorem 5.21 from the last section tell us how *small* an independent set can be given the degrees of the vertices in a Graph. Turán's theorem also told us how *large* the number of edges can be, subject to some other constraints. Results like these are called *extremal*, and belong to the branch of graph theory known as extremal graph theory.

Here we will be dealing mostly with sets — and what we are concerned about is how large/small a finite set can be subject to certain constraints. Some motivating examples:

1. How large can a subfamily $\mathcal{A} \subseteq \wp[n]$ be if no subset in $\mathcal{A}$ is a subset of any other set in $\mathcal{A}$?

   Take $\wp[5] = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1,2\}, \{1,3\}, \ldots, \{1,2,3,4,5\}\}$ as an example. A possible candidate for $\mathcal{A}$ would be the subfamily

   $$\mathcal{A}_1 = \binom{[5]}{1} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}.$$

   No subset in $\mathcal{A}_1$ is a subset of any other subset in $\mathcal{A}_1$ as desired, and $|\mathcal{A}_1| = 5$. Is this the best we can do? If we consider the subfamily

   $$\mathcal{A}_2 = \binom{[5]}{2} = \{\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{2,3\}, \{2,4\}, \{2,5\}, \{3,4\}, \{3,5\}, \{4,5\}\},$$

   this works as well. So now we have a subfamily of size $\binom{5}{2} = 10$. Can we do any better?

   Later on in theorem 11.4 we show that $|\mathcal{A}| \leqslant \binom{n}{\lfloor \frac{n}{2} \rfloor}$, which in this case is 10. This is known as Sperner's theorem.

2. How large can a subfamily $\mathcal{A} \subseteq \wp[n]$ be if no two subsets in $\mathcal{A}$ are disjoint?

   Again, we take the example of $\wp[5]$. We could perhaps pick an arbitrary element, say $1 \in [n]$, and take all the sets which contain 1 as a candidate:

   $$\mathcal{A}_1 = \big\{\{1\}, \{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,3,4\}, \{1,3,5\}, \{1,4,5\}, \{1,2,3,4\}, \{1,2,3,5\}, \{1,2,4,5\}, \{1,3,4,5\}, \{1,2,3,4,5\}\big\}.$$

   Effectively this adds $\cup\{1\}$ to each member of $\wp\{2,3,4,5\}$, so we have $|\wp\{2,3,4,5\}| = 2^4 = 16$ subsets so far. Is this the best we can do? Note that it is not necessary for there to be a common element throughout, another example which works is

   $$\mathcal{A}_2 = \{\{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,3,4\}, \{1,3,5\}, \{2,3,4\}, \{2,3,5\}, \{1,2,3,4\}, \{1,2,3,5\}, \{1,2,4,5\}, \{1,3,4,5\}, \{2,3,4,5\}, \{1,2,3,4,5\}\},$$

which also has $|\mathcal{A}_2| = 16$. Also note that if we try $\mathcal{A}_3 = \binom{[5]}{3} \cup \binom{[5]}{4} \cup \binom{[5]}{5}$, this will definitely work, since any two subsets in this family are at least of size 3, and if they were disjoint, then their union would give us a subset of $[5]$ of size 6, which is impossible. But this again gives us a subfamily of size $\binom{5}{3} + \binom{5}{4} + \binom{5}{5} = 16$. Can we do better?

In theorem 9.4 we show that $|\mathcal{A}| \leqslant 2^{n-1}$, which in this case is 16.

3. How large can a subfamily $\mathcal{A} \subseteq \binom{[n]}{k}$ be if no two subsets in $\mathcal{A}$ are disjoint?

   By a similar reasoning to that in the previous example when describing $\mathcal{A}_3$, if $\frac{n}{2} < k < n$, then trivially $\mathcal{A} = \binom{[n]}{k}$ works, since no two subsets can be disjoint. For an example where $k \leqslant \frac{n}{2}$, suppose $n = 5$ and $k = 2$, so our family is $\mathcal{F} = \binom{[n]}{k} = \{\{1,2\},\{1,3\},\{1,4\},\ldots,\{4,5\}\}$.

   If we always include 1, we obtain $\mathcal{A}_1 = \{\{1,2\},\{1,3\},\{1,4\},\{1,5\}\}$ as a possible candidate of size $|\mathcal{A}_1| = 4$.

   In fact, in theorem 9.5 we show that $|\mathcal{A}| \leqslant \binom{n-1}{r-1}$, which in this case is 4. This is known as the Erdős-Ko-Rado theorem.

7 Exact Intersections

How large can a subfamily $\mathcal{A}$ of a given family $\mathcal{F}$ be if any two sets of $\mathcal{A}$ intersect in exactly t elements? For example, If we have $\mathcal{F} = \wp[5] = \{\emptyset,\{1\},\ldots,\{1,2,3,4,5\}\}$ and $t = 0$, then we may take $\mathcal{A}$ to be the subfamily $\{\emptyset,\{1\},\{2\},\{3\},\{4\},\{5\}\}$, since any two subsets intersect in 0 elements (they are disjoint). Note that $|\mathcal{A}| \leqslant 6$. If instead we have $t = 2$, then a suitable $\mathcal{A}$ would be $\{\{1,2\},\{1,2,3\},\{1,2,4\},\{1,2,5\}\}$. Another observation we make is that unless $|\mathcal{A}| = 1$, each $A \in \mathcal{A}$ must have $|A| \geqslant t$ (otherwise they cannot intersect in at least t elements).

Theorem 7.1 (Fischer's Inequality)**.** *If $\mathcal{A} \subseteq \wp[n]$ such that $|A \cap B| = t$ for all $A, B \in \mathcal{A}$ (where $A \neq B$), then $|\mathcal{A}| \leqslant n+1$, and if furthermore $t \neq 0$, then $|A| \leqslant n$.*

Proof. If $t = 0$, then $\mathcal{A} = \{A_1, A_2, \ldots, A_{|\mathcal{A}|}\}$ is a family of disjoint subsets of $[n]$. Then

$$n = |[n]| \geqslant \left|\bigcup_{i=1}^{|\mathcal{A}|} A_i\right| = |A_1| + |A_2| + \cdots + |A_{|\mathcal{A}|}| \geqslant 0 + (|\mathcal{A}| - 1) \cdot 1,$$

since at most one of the A_i can be the empty set, and therefore the rest of the sets must have size at least 1. It follows that $|\mathcal{A}| \leqslant n+1$.

Otherwise if $t \geqslant 1$, two cases arise.

Case I There exists $B \in \mathcal{A}$ such that $|B| \leqslant t$.

If $|B| < t$, then no set in $\wp[n]$ intersects B on t elements, therefore $\mathcal{A} = \{B\}$. Otherwise if $|B| = t$, then $A \cap B = B$ for all $A \in \mathcal{A}$, and furthermore, $A_1 \cap A_2 = B$ for all $A_1, A_2 \in \mathcal{A}$. This implies that $\{A \setminus B : A \in \mathcal{A}\}$ is a family of disjoint subsets of $[n] \setminus B$, and can have at most $n - |B| + 1$ elements (by the case when $t = 0$). Therefore $|\mathcal{A}| \leqslant n - |B| + 1 = n - t + 1 \leqslant n$, since $t \geqslant 1$.

Case II $|A| > t$ for all $A \in \mathcal{A}$.

Now for every $A \in \mathcal{A}$, define the vector $\mathbf{x}_A \in \mathbb{R}^n$ such that $\mathbf{x}_A = (x_1, x_2, \ldots, x_n)$ where

$$x_i = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{otherwise.} \end{cases}$$

Thus for all $A, B \in \mathcal{A}$,

$$\mathbf{x}_A \cdot \mathbf{x}_B = \begin{cases} |A| & \text{if } A = B \\ t & \text{if } A \neq B. \end{cases}$$

We now show that the set of vectors $\{\mathbf{x}_A : A \in \mathcal{A}\}$ forms a set of linearly independent vectors in $\mathbb{R}^n$. Indeed, for all $A \in \mathcal{A}$, let $\alpha_A \in \mathbb{R}$ such that $\sum_{A\in\mathcal{A}} \alpha_A \mathbf{x}_A = \mathbf{0}$. Now

$$\begin{aligned}
0 &= \mathbf{0} \cdot \mathbf{x}_B && \text{(for any } B \in \mathcal{A}) \\
&= \left(\sum_{A\in\mathcal{A}} \alpha_A \mathbf{x}_A \right) \cdot \mathbf{x}_B \\
&= \left(\alpha_B \mathbf{x}_B + \sum_{A\in\mathcal{A}\setminus B} \alpha_A \mathbf{x}_A \right) \cdot \mathbf{x}_B \\
&= \alpha_B (\mathbf{x}_B \cdot \mathbf{x}_B) + \sum_{A\in\mathcal{A}\setminus B} \alpha_A (\mathbf{x}_A \cdot \mathbf{x}_B) \\
&= \alpha_B |B| + \sum_{A\in\mathcal{A}\setminus B} \alpha_A t \\
&= \alpha_B |B| + \sum_{A\in\mathcal{A}} \alpha_A t - \alpha_B t \\
&= \alpha_B (|B| - t) + \sum_{A\in\mathcal{A}} \alpha_A t \\
&= \alpha_B (|B| - t) + ts && \text{(where } s := \sum_{A\in\mathcal{A}} \alpha_A t) \\
\implies \alpha_B &= \frac{-ts}{|B| - t} \qquad \circledast
\end{aligned}$$

Now

$$s = \sum_{B\in\mathcal{A}} \alpha_B t = \sum_{B\in\mathcal{A}} \frac{-ts}{|B| - t}, \qquad \text{so} \qquad s\left(1 + \sum_{B\in\mathcal{A}} \frac{t}{|B| - t}\right) = 0.$$

Notice that $\frac{t}{|B|-t} > 0$ for all $B \subseteq \mathcal{A}$, since for case II we assumed $|B| > t$. Therefore the bracket above cannot be zero, so it follows that $s = 0$, which by $\circledast$ implies that $\alpha_B = 0$ for all $B \in \mathcal{A}$. Therefore we have established that $\{\mathbf{x}_A : A \in \mathcal{A}\}$ is a set of linearly independent vectors in $\mathbb{R}^n$, so there cannot be more than n of them. But each $\mathbf{x}_A$ corresponds to a set $A \in \mathcal{A}$, so $|\mathcal{A}| \leqslant n$. □

Remark 7.2. When is the bound attained?

(i) When $t = 0$, the family $\mathcal{A} = \{\emptyset, \{1\}, \{2\}, \ldots, \{n\}\}$ has $|\mathcal{A}| = n + 1$, attaining the bound.

(ii) When $t = 1$, the family $\mathcal{A} = \{\{1\}, \{1, 2\}, \ldots, \{1, n\}\}$ attains the bound $|\mathcal{A}| = n$.

(iii) When $t = n-2$, the family $\mathcal{A} = \binom{[n]}{n-1}$ attains the bound $|\mathcal{A}| = n$.

For example with $n = 5$, we get

$$\mathcal{A} = \binom{[5]}{4} = \{\{1,2,3,4\},\{1,2,3,5\},\{1,2,4,5\},\{1,3,4,5\},\{2,3,4,5\}\}.$$

(iv) When $t \leqslant n-2$, the family $\mathcal{A} = \binom{[t+2]}{t+1}$ has size $|\mathcal{A}| = t+2$.

For example, with $n = 5$ and $t = 2$, we have $\mathcal{A} = \binom{[4]}{3} = \{\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\}\}$.

(v) By case I in the proof, we can always construct a family of size $n-t+1$.

Not much is known for other values of t. Note however that by combining (iv) and (v), we can always construct an exact intersection of size at least $\frac{n+3}{2}$. Indeed, considering $\max\{t+2, n-t+1\}$, if $t > \frac{n-1}{2}$, then $t+2 > \frac{n+3}{2}$ so we use $\binom{[t+2]}{t+1}$, whereas if $t \leqslant \frac{n-1}{2}$, then $n-t+1 \geqslant n - \frac{n-1}{2} + 1 = \frac{n+3}{2}$, so we construct a family as described in case I of the proof.

8 Double Counting

Let X and Y be two finite sets, and let $\sim \subseteq X \times Y$ be a relation defined on these two sets. Additionally, let each $x \in X$ be related to at most q elements in Y, and let each $y \in Y$ be related to at least p elements of X. The aim of this section is to count the number σ of pairs (x, y) such that $x \sim y$.

For all $(x, y) \in X \times Y$, define $\chi(x, y)$ by

$$\chi(x,y) \coloneqq \begin{cases} 1 & \text{if } x \sim y \\ 0 & \text{otherwise.} \end{cases}$$

Therefore

$$\sigma = \sum_{x\in X}\sum_{y\in Y}\chi(x,y) = \sum_{y\in Y}\sum_{x\in X}\chi(x,y).$$

But $\sum_{x\in X}\sum_{y\in Y}\chi(x,y) \geqslant \sum_{x\in X} q = q|X|$, and similarly $\sum_{y\in Y}\sum_{x\in X}\chi(x,y) \geqslant \sum_{y\in Y} p = p|Y|$. Thus it follows that $|X|q \leqslant \sigma \leqslant |Y|p$, which rearranges to give

$$|X| \leqslant \frac{p}{q}|Y|.$$

This simple result is the principle of *double counting*.

Remark 8.1. If instead of "at most" and "at least", both read "precisely", that is, if we had that every $x \in X$ were related to precisely q elements in Y and every $y \in Y$ were related to precisely p elements in X, then we get equality: $|X| = \frac{p}{q}|Y|$.

8.1 Some Simple Applications

Remark 8.2. If we consider the bipartite graph $B = (X \cup Y, E)$ where $E = \{\{x, y\} : x \sim y\} \subseteq X \times Y$, then double counting essentially relates the number of edges incident to

vertices in X to the number of edges incident to vertices in Y, and we get that

$$|X| \leqslant \frac{\Delta(Y)}{\delta(X)}|Y|,$$

where $\Delta(U)$ and $\delta(U)$ denote the maximum and minimum degrees of all the vertices in $U \subseteq V$ respectively.

The double counting principle provides us with an easy proof of the handshaking lemma.

Proposition 8.3 (Handshaking Lemma for Graphs)**.** *Let $G = (V, E)$ be a graph. Then*

$$\sum_{v \in V} \deg v = 2|E|.$$

Proof. Let $v \in V$ be related to $e \in E$ if $v \in e$, that is, v is one of the end points of the edge e. Every $v \in V$ is related to precisely $\deg v$ elements in E, and every element $e \in E$ is related to precisely 2 elements of V. Thus by double counting,

$$\sum_{v \in V} \deg v = \sum_{e \in E} 2 = 2|E|,$$

as required. □

By a similar reasoning, we obtain the following result as well:

Proposition 8.4 (Handshaking Lemma for Hypergraphs)**.** *Let $H = (V, E)$ be a hypergraph. Then*

$$\sum_{v \in V} = \sum_{e \in E} |e|.$$

If, furthermore, the hypergraph is k-uniform, then

$$\sum_{v \in V} = k|E|.$$

9 Intersecting Families

How large can a subfamily $\mathcal{A}$ of a given family $\mathcal{F}$ be if any two sets of $\mathcal{A}$ intersect in *at least* t elements?

Definition 9.1 (t-Intersecting Family)**.** A family $\mathcal{A}$ of sets is said to be *t-intersecting* if $|A \cap B| \geqslant t$ for any two subsets $A, B \in \mathcal{A}$ (where $A \neq B$). If $t = 1$, we simply say $\mathcal{A}$ is *intersecting*.

Definition 9.2 (t-star)**.** Given a family $\mathcal{F}$ and a set T, the subfamily $\mathcal{F}\langle T\rangle$ defined by

$$\mathcal{F}\langle T\rangle \coloneqq \{F \in \mathcal{F} : T \subseteq F\}$$

is said to be a *t-star* of $\mathcal{F}$ if it is nonempty and $|T| \geqslant t$. A 1-star of $\mathcal{F}$ is simply called a *star* of $\mathcal{F}$.

Example 9.3. Let $\mathcal{F}_1 = \wp[n]$, let $t = 1$, and let $T = \{5\}$. Then

$$\mathcal{F}_1\langle T\rangle = \begin{cases} \emptyset & \text{if } n \leqslant 4 \\ \{\{5\},\{1,5\},\{2,5\},\{3,5\},\dots,[n]\} & \text{otherwise.} \end{cases}$$

In the case when $n \geqslant 5$, the star $\mathcal{F}_1\langle T\rangle$ is the power set of $[n] \setminus \{5\}$ with the operation $\cup\{5\}$ applied to each set inside. Thus $|\mathcal{F}_1\langle T\rangle| = 2^{n-1}$. In fact, this is a largest intersecting subfamily of $\mathcal{F}_1$.

Another example, let $\mathcal{F}_2 = \binom{[5]}{2}$, let $t = 1$, and let $T = \{2\}$. Then

$$\mathcal{F}_2\langle T\rangle = \{\{1,2\},\{2,3\},\{2,4\},\{2,5\}\}.$$

$|\mathcal{F}_2\langle T\rangle| = \binom{5-1}{2-1} = 4$, which is also in fact a largest intersecting subfamily of $\mathcal{F}_2$.

Theorem 9.4. *If $\mathcal{A}$ is an intersecting subfamily of $\wp[n]$, then*

$$|\mathcal{A}| \leqslant 2^{n-1},$$

and equality holds if $\mathcal{A}$ is a star of $\wp[n]$.

Proof. Let $\mathcal{S} = \{A \in \wp[n] : 1 \in A\}$. Clearly $|\mathcal{S}| = 2^{n-1}$. Define $f\colon \mathcal{A} \to \mathcal{S}$ such that

$$f(A) = \begin{cases} A & \text{if } A \in \mathcal{S} \\ [n] \setminus A & \text{otherwise.} \end{cases}$$

We show that f is an injection. Suppose $f(A) = f(B)$ for $A, B \in \mathcal{A}$. First of all, notice that either both $A, B \in \mathcal{S}$ or both $A, B \notin \mathcal{S}$. Indeed, if $A \in \mathcal{S}$ and $B \notin \mathcal{S}$, then $f(A) = A$ and $f(B) = [n] \setminus B$, thus $A = [n] \setminus B$, and in particular, $A \cap B = \emptyset$. But $A, B \in \mathcal{A}$, so $|A \cap B| \geqslant 1$ ※. Now, if both $A, B \in \mathcal{S}$, then $A = f(A) = f(B) = B$. If, on the other hand, both $A, B \notin \mathcal{S}$, then $[n] \setminus A = f(A) = f(B) = [n] \setminus B$, which implies that $A = B$ as well. Therefore f is injective.

Thus, $|\mathrm{Dom}(f)| \leqslant |\mathrm{Codom}(f)|$, that is, $|\mathcal{A}| \leqslant |\mathcal{S}| = 2^{n-1}$. Also, the stars of $\wp[n]$ are families of the form $\{A \in \wp[n] : i \in A\}$ for $i \in [n]$, and each have size 2^{n-1}. □

9.1 The Erdős-Ko-Rado Theorem

Now, consider the family $\mathcal{F} = \binom{[n]}{k}$. If $\frac{n}{2} < k < n$, we have already seen in the third motivating example of the introductory section that $\mathcal{F}$ itself is already an intersecting family. Indeed, let $A, B \in \mathcal{F} = \binom{[n]}{k}$ where $\frac{n}{2} < k < n$. Then $n \geqslant |A \cup B| = |A| + |B| - |A \cap B| = 2k - |A \cap B| > n - |A \cap B|$, that is, $|A \cap B| > 0$.

If $1 \leqslant k \leqslant \frac{n}{2}$, the problem is nontrivial. In fact, we have the following well known result.

Theorem 9.5 (Erdős-Ko-Rado Theorem (EKR), 1961)**.** *If $1 \leqslant k \leqslant \frac{n}{2}$ and $\mathcal{A}$ is an intersecting subfamily of $\binom{[n]}{k}$, then*

$$|\mathcal{A}| \leqslant \binom{n-1}{k-1},$$

and equality holds if $\mathcal{A}$ is a star of $\binom{[n]}{k}$.

We need some lemmata before we can prove the EKR theorem.

Note. One can obtain a complete characterisation of the extremal cases/structures:

(i) If $k < \frac{n}{2}$, the bound is attained if and only if $\mathcal{A}$ is a star.

(ii) If $k = \frac{n}{2}$, the bound is attained if and only if for each $A \in \binom{[n]}{k}$, exactly one of A and $[n] \setminus A$ is in $\mathcal{A}$. For example, if $\mathcal{F} = \binom{[4]}{2}$, then $\mathcal{A} = \{\{1,2\},\{1,3\},\{1,4\}\}$ is an extremal case.

Definition 9.6. Let X be a set of size m.

(i) A bijection $\sigma\colon X \to [m]$ is called a *cyclic ordering* of X.

(ii) If the elements of $A = \{a_1, a_2, \ldots, a_n\} \subseteq X$ where $n \leqslant m$ are consecutively labelled by σ in a cyclic manner, that is, if for some $k \in [m]$ we have $\sigma(a_i) = k + i \bmod m$ for all $i \in [n]$, then we say that A *meets* σ.

Example 9.7. Let $X = \{A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}$ represent the cards in a suite of playing cards. Define the cyclic ordering $\sigma\colon X \to [13]$ by $\sigma(A) = 1$, $\sigma(2) = 2$, …, $\sigma(Q) = 12$, $\sigma(K) = 13$, effectively labelling the cards in the usual order.

The subset $\{4,5,6,7\} \subseteq X$ meets σ, so do $\{10, J, Q, K\}$, $\{J, Q, K, A, 2\}$ and $\{A, J, Q, K\}$ since each of these are labelled consecutively by σ (in a cyclic fashion). The sets $\{2,4,5,6\}$ or $\{A, J, K\}$ however do not meet σ.

Lemma 9.8 (Katona, 1972)**.** *Let $1 \leqslant k \leqslant \frac{m}{2}$, let X be a set of size m, let σ be a cyclic ordering of X, and let*

$$\mathcal{C} = \left\{ C \in \binom{X}{k} : C \text{ meets } \sigma \right\}.$$

For any intersecting subfamily $\mathcal{B}$ of $\mathcal{C}$, $|\mathcal{B}| \leqslant k$.

Proof. Fix a cyclic ordering $\sigma\colon X \to [m]$ of the elements in X which labels them $x_1, x_2, \ldots, x_m$. Let $B^* = \{x_1, x_2, \ldots, x_k\} \in \mathcal{B}$ be the first k elements of X under the labelling σ, and for all $i \in [k]$, let $C_i \in \mathcal{C}$ be the k-set starting with x_i, and let $C_i' \in \mathcal{C}$ be the k-set ending in x_i (illustrated in figure 25).

Note that $B^* = C_1 = C_k'$, and since $k \leqslant \frac{m}{2}$, we have that $C_j \cap C_{j-1}' = \emptyset$ for each $j \in \{2, \ldots, k\}$, and therefore at most one of C_j and C_{j-1}' is in $\mathcal{B}$ since it is an intersecting subfamily. Therefore

$$\mathcal{B} \subseteq \{B^*, \text{one of } C_2 \text{ or } C_1', \text{one of } C_3 \text{ or } C_2', \ldots, \text{one of } C_k \ \&\ C_{k-1}'\},$$

so $|\mathcal{B}| \leqslant k$. □

Now we present a simple proof of the EKR theorem due to Katona.

Proof of EKR theorem (Katona, 1972). Let $X = \mathcal{A}$ and let Y be the set of all cyclic orderings of $[n]$ (thus $|Y| = n!$). Now for each $A \in X$ and $\sigma \in Y$, we say that they are related ($A \sim \sigma$) if and only if A meets σ.

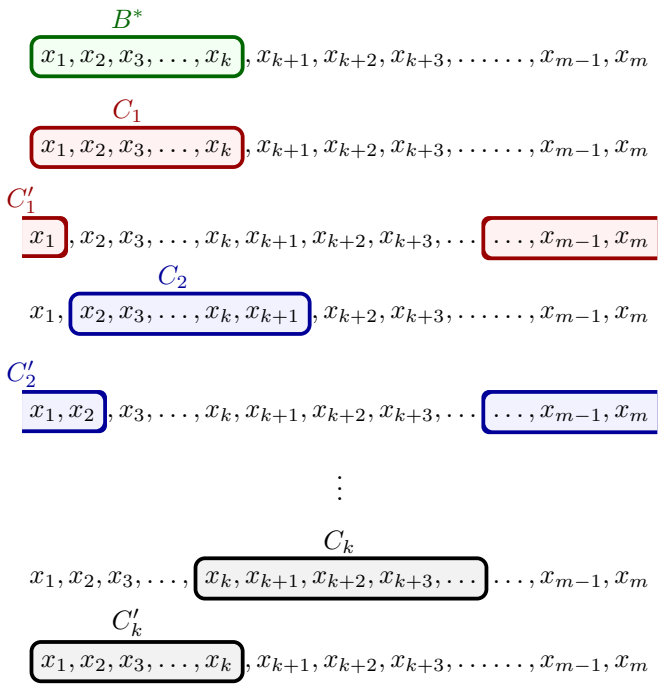

Figure 25: An illustration of the sets B^*, C_i and C_i' in the proof of lemma 9.8

Let q be the number of members of Y to which each member $A \in X$ is related. For every $K \in \binom{[n]}{k}$, there are precisely $q = k!(n-k)!n$ members $\sigma \in Y$ such that K meets σ.[9]

Now let p be the number of members of X to which each member $\sigma \in Y$ is related. The required members of X form a subfamily $\mathcal{B}$ of the family $\{C \in \binom{X}{k} : C \text{ meets } \sigma\}$. Furthermore, $\mathcal{B} \subseteq \mathcal{A}$, an intersecting subfamily; so $\mathcal{B}$ itself is an intersecting subfamily. Thus by lemma 9.8, $|\mathcal{B}| \leqslant k$, and by double counting,

$$|X| \leqslant \frac{p}{q}|Y| = \frac{k}{k!(n-k)!n}n! = \frac{(n-1)!}{(k-1)!(n-k)!} = \binom{n-1}{k-1},$$

as required. □

Note. Stars of $\wp[n]$ and $\binom{[n]}{r}$ are extremal intersecting families, but not necessarily the only ones. For example, consider the family $\mathcal{F} = \wp[n]$. The subfamily $\mathcal{A} = \{A \in \wp[n] : |A \cap [3]| \geqslant 2\}$ is another family with $|\mathcal{A}| = 2^{n-1}$. In particular, when $n = 5$,

$$\mathcal{A} = \{\{1,2\},\{1,3\},\{2,3\},\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\},\{1,2,3,4\},\{1,2,5\},\{1,3,5\},$$
$$\{2,3,5\},\{1,2,3,5\},\{1,2,4,5\},\{1,3,4,5\},\{2,3,4,5\},\{1,2,3,4,5\}\},$$

which is not a star.

[9] For example, if $n = 4$, $k = 2$ and $K = \{1,3\}$, then σ can be any of the orderings 1324, 1342, 2134, 4132, 2413, 4213, 3241, 3421, 3124, 3142, 2314, 4312, 2431, 4231, 1243, 1423. The important thing is that 1 and 3 are (cyclically) adjacent.

Proposition 9.9. *Every intersecting subfamily of $\wp[n]$ can be extended to one of maximum size, 2^{n-1}.*

Proof. Let $\mathcal{A} = \{A_1, \ldots, A_k\}$. We construct $\mathcal{B} \supseteq \mathcal{A}$ with $|\mathcal{B}| = 2^{n-1}$. Consider $A \in \wp[n]$. If $A \cap A_i \neq \emptyset$ for all $i = 1, \ldots, k$, then put $A \in \mathcal{B}$. Otherwise if $A \cap A_i = \emptyset$ for some $i = \alpha$, then $A \cap A_\alpha = \emptyset \implies ([n] \setminus A) \cap A_\alpha = A_\alpha \implies ([n] \setminus A) \cap A_i \neq \emptyset$ for all $i = 1, \ldots, k$, so put $[n] \setminus A$ in $\mathcal{B}$. So for each set $A \in \wp[n]$, we are either putting A or $[n] \setminus A$ in $\mathcal{B}$, accounting for half of the 2^n sets in $\wp[n]$, so $|\mathcal{B}| = \frac{1}{2}2^n = 2^{n-1}$. □

10 Levels of Hereditary Families

Definition 10.1 (Hereditary Family)**.** A family $\mathcal{H}$ of sets is said to be *hereditary* (or a downset, an ideal, a simlicial complex) if for all $H \in \mathcal{H}$, all the subsets of H are also in $\mathcal{H}$ (i.e. $\wp H \subseteq \mathcal{H}$).

Example 10.2. We give a few examples of hereditary families.

(i) The family $\mathcal{F} = \{\{1\}, \{2\}, \{3\}, \{1, 2, 3\}\}$ is not hereditary, because $\{1, 2, 3\} \supseteq \{1, 2\} \notin \mathcal{F}$. An even simpler reason for which $\mathcal{F}$ fails to be hereditary is that $\emptyset \notin \mathcal{F}$.

$\mathcal{H} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}\}$ is hereditary.

(ii) Consider the following graph.

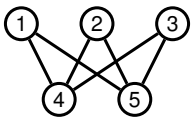

Figure 26

The family $\mathcal{H}$ containing all the independent sets of vertices is hereditary:

$$\mathcal{H} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}, \{4, 5\}\}.$$

(iii) Consider the following graph.

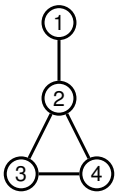

Figure 27

The family $\mathcal{H}$ containing all the cliques (induced complete subgraphs) is hereditary:

$$\mathcal{H} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{2, 3, 4\}\}.$$

(iv) Given any finte set X, the power set $\wp X$ is trivially hereditary.

(v) Consider the following graph:

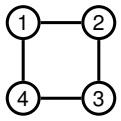

Figure 28

The set of cycles in a graph is not hereditary in general.

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}, \{3,4\}, \{1,4\}, \{1,2,3,4\}\}.$$

Definition 10.3 (Base)**.** A member B of a family $\mathcal{F}$ is said to be a *base* of $\mathcal{F}$ if it is not a subset of any other member of $\mathcal{F}$.

Proposition 10.4. *A family is hereditary if and only if it is a union of power sets.*

Proof. We start with the 'only if' part. Let $\mathcal{H}$ be hereditary, and let $B_1, \ldots, B_k$ be its bases. Let $H \in \bigcup_{i=1}^{k} \wp B_i$. Then $H \in \wp B_j$ for some $j \in [k]$, so $H \in \mathcal{H}$. Therefore $\bigcup_{i=1}^{k} \wp B_i \subseteq \mathcal{H}$.

Now let $H \in \mathcal{H}$, and let B be the largest set in $\mathcal{H}$ such that $H \subseteq B$. Clearly B is a base, since otherwise $B \subsetneq C \in \mathcal{H}$ and this contradicts that B is the largest set containing H. Thus for some $j \in [k]$, $B = B_j$, and thus $H \subseteq B_j$, that is, $H \in \wp B_j$. It follows that $\mathcal{H} \subseteq \bigcup_{i=1}^{k} \wp B_i$, and therefore that $\mathcal{H} = \bigcup_{i=1}^{k} \wp B_i$.

Conversely, let $\mathcal{F}$ be a union of power sets. Then $\mathcal{F} = \bigcup_{i=1}^{k} \wp X_i$ for some sets $X_1, \ldots, X_k$. We show that given $F \in \mathcal{F}$, $E \in \mathcal{F}$ for all $E \subseteq F$. Indeed,

$$\begin{aligned}
F \in \mathcal{F} &\implies F \in \bigcup_{i=1}^{k} \wp X_i \\
&\implies F \in \wp X_j \qquad \text{for some } j \in [k] \\
&\implies F \subseteq X_j \\
&\implies E \subseteq X_j \\
&\implies E \in \wp X_j \\
&\implies E \in \bigcup_{i=1}^{k} \wp X_i,
\end{aligned}$$

that is, $E \in \mathcal{F}$. □

Definition 10.5 (kth Level)**.** For any integer $k \geqslant 0$, the family of all sets in $\mathcal{F}$ of size k, denoted $\mathcal{F}^{(k)}$, is called the *kth level* of $\mathcal{F}$, therefore

$$\mathcal{F}^{(k)} = \{F \in \mathcal{F} : |F| = k\}.$$

Example 10.6. If $\mathcal{F} = \wp[n]$, then $\mathcal{F}^{(k)} = \binom{[n]}{k}$, and if $\mathcal{F} = \binom{[n]}{k}$, then $\mathcal{F}^{(k)} = \binom{[n]}{k} = \mathcal{F}$.

Notation. The size of a smallest base of a family $\mathcal{F}$ is denoted by $\mu(\mathcal{F})$.

Theorem 10.7 (Peter Borg, 2009)**.** *If $\mathcal{H}$ is a hereditary family and $r \leqslant s \leqslant \mu(\mathcal{H})$, then*

$$|\mathcal{H}^{(s)}| \geqslant \frac{\binom{\mu(\mathcal{H})-r}{s-r}}{\binom{s}{s-r}}|\mathcal{H}^{(r)}|.$$

Proof. Let $X = \mathcal{H}^{(r)}$ and $Y = \mathcal{H}^{(s)}$. For each $A \in X$ and each $B \in Y$, we say that A and B are related ($A \sim B$) if and only if $A \subseteq B$.

Suppose M is a base of $\mathcal{H}$ such that $A \subseteq M$. Then $|M| \geqslant \mu(\mathcal{H}) \geqslant s$. But since $\mathcal{H}$ is hereditary, every s-subset of M is a member of $\mathcal{H}$ (and therefore of $\mathcal{H}^{(s)}$), i.e. we have $\binom{M}{s} \subseteq \mathcal{H}^{(s)}$. Since $|A| = r$ and $A \subseteq M$, then there exist $\binom{|M|-r}{s-r}$ sets in $\binom{M}{s}$ that contain A, and since $|M| \geqslant \mu(\mathcal{H})$, then $\binom{|M|-r}{s-r} \geqslant \binom{\mu(\mathcal{H})-r}{s-r} =: q$.

Now suppose each $B \in Y$ is related to p members of X. Since B is an s-subset of $\mathcal{H}$ where $s \geqslant r$, then all the r-subsets of B are members of $\mathcal{H}^{(r)}$. Thus $p = \binom{s}{r} = \binom{s}{s-r}$. The result follows by double counting. □

Remark 10.8. What is an extremal case of this theorem? Consider the family $\mathcal{H} = \wp[n]$, which is hereditary. Its only base is the set $[n]$, so $\mu(\wp[n]) = n$. Then

$$\frac{\binom{\mu(\mathcal{H})-r}{s-r}}{\binom{s}{s-r}}|\mathcal{H}^{(r)}| = \frac{\binom{n-r}{s-r}}{\binom{s}{s-r}}\binom{n}{r} = \binom{n}{s} = |\mathcal{H}^{s}|, \text{ so the bound is attained.}$$

Corollary 10.9. *If $\mathcal{H}$ is a hereditary family and $r < s < \mu(\mathcal{H}) - r$, then*

$$|\mathcal{H}^{(s)}| > |\mathcal{H}^{(r)}|.$$

Proof. Since $\binom{\mu(H)-r}{s-r} > \binom{s}{s-r}$, the result follows from theorem 10.7. □

Corollary 10.10. *If $\mathcal{H}$ is a hereditary family and $r < s < \mu(\mathcal{H})/2$, then*

$$|\mathcal{H}^{(s)}| > |\mathcal{H}^{(r)}|.$$

Proof. $r < \mu(\mathcal{H})/2 \iff 2r < \mu(\mathcal{H}) \iff r < s < \mu(\mathcal{H}) - r$. The result follows from corollary 10.9. □

Remark 10.11. Let $\mathcal{H} = \wp[n]$. Then $|\mathcal{H}^{(s)}| = \binom{n}{s}$ and $|\mathcal{H}^{(r)}| = \binom{n}{r}$. Since $\mu(\mathcal{H}) = n$, then by corollary 10.10 it follows that the values of $\binom{n}{t}$ are unimodal; that is, they increase, then they decrease, since they are symmetric about $t = \lfloor \frac{n}{2} \rfloor$ (because $\binom{n}{t} = \binom{n}{n-t}$). Indeed,

$$\binom{n}{0} < \binom{n}{1} < \cdots < \binom{n}{\lfloor \frac{n}{2} \rfloor} = \binom{n}{\lceil \frac{n}{2} \rceil} > \cdots > \binom{n}{n-1} > \binom{n}{n}.$$

11 Antichains

Definition 11.1 (Antichain)**.** A family $\mathcal{A}$ is said to be an *antichain* (or Sperner family) if no member of $\mathcal{A}$ is a subset of another member of $\mathcal{A}$, that is, $A \nsubseteq B$ for all $A, B \in \mathcal{A}$ (where $A \neq B$).

Definition 11.2 (r-uniform)**.** A family $\mathcal{F}$ is said to be r-uniform if every member of $\mathcal{F}$ contains precisely r elements.

Remark 11.3. Every r-uniform family is an antichain. In particular, $\binom{[n]}{r}$ is an antichain. If we consider the family $\mathcal{F} = \wp[n]$, then $\mathcal{F}^{(\lfloor n/2 \rfloor)} = \binom{[n]}{\lfloor n/2 \rfloor}$ is a level of $\mathcal{F}$ of maximum size. This follows by remark 10.11, as well as by Sperner's theorem:

Theorem 11.4 (Sperner, 1928)**.** *If $\mathcal{A} \subseteq \wp[n]$ is an antichain, then*

$$|A| \leqslant \binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

This is a consequence of the well known LYM inequality, also known as the (B)LYM inequality:

Theorem 11.5 ((Bollobász)-Lubell-Yamamoto-Meshalkin, (B)LYM)**.** *If $\mathcal{A} \subseteq \wp[n]$ is an antichain, then*

$$\sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{|\mathcal{A}|}} \leqslant 1.$$

Clearly Sperner's theorem follows, since $1 \geqslant \sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{|\mathcal{A}|}} \geqslant \sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{\lfloor n/2 \rfloor}} = \frac{|A|}{\binom{n}{\lfloor n/2 \rfloor}}$ by remark 10.11.

Proof of (B)LYM inequality. Consider $A \in \wp[n]$, and let $\sigma = x_1 x_2 \cdots x_n$ be a permutation of $[n]$. We will say that A meets σ if the first $|A|$ entries of σ form A, that is, $A = \{x_1, x_2, \ldots, x_{|A|}\}$. A subset $\mathcal{A}$ of $[n]$ meets precisely $|A|!(n - |A|)!$ permutations of $[n]$. Indeed, let P denote the set S_n of permutations of $[n]$, and for all $A \in \mathcal{A}$, let P_A be the set of permutations of $[n]$ met by A. Then we have

$$|P_A| = |A|!(n - |A|)!$$

Now observe that $P_A \cap P_B = \emptyset$ if $A \neq B$. Otherwise, if there exist two sets $A, B \in \mathcal{A}$ with $P_A \cap P_B \neq \emptyset$ (where without loss of generality we assume that $|A| \leqslant |B|$), then A and B both meet $\sigma \in P_A \cap P_B$, and therefore $A \subseteq B$, contradicting the fact that $\mathcal{A}$ is an antichain. Thus the P_A are pairwise disjoint subsets, and

$$\sum_{A \in \mathcal{A}} |P_A| = \left| \bigcup_{A \in \mathcal{A}} P_A \right| \leqslant |P| = n!$$
$$\implies \sum_{A \in \mathcal{A}} \frac{|P_A|}{n!} = \sum_{A \in \mathcal{A}} \frac{|A|!(n - |A|)!}{n!} \leqslant 1,$$

as required. □

12 Shadows

Definition 12.1 (sth shadow)**.** The *sth shadow* of a family $\mathcal{A}$, denoted $\partial_s(\mathcal{A})$, is the family of all s-subsets of sets in A, i.e.

$$\partial_s(\mathcal{A}) = \bigcup_{A\in\mathcal{A}} \binom{A}{s}.$$

Remark 12.2. Let $\mathcal{H}$ be the hereditary family $\bigcup_{A\in\mathcal{A}} \wp A$. Then $\partial_s(A) = \mathcal{H}^{(s)}$.

Given an r-uniform family of size m, what is the smallest the sth shadow can be? Or, to rephrase, given $m \geqslant 1$, how can one construct an r-uniform family of size m that is guaranteed to have an sth shadow of minimal size? We answer this question in the Kruskal-Katona theorem, however first we define the notion of colexicographical order.

Definition 12.3 (Colexicographical order)**.** Let $A = \{a_1, \dots, a_k\}$ and $B = \{b_1, \dots, b_k\}$ be k-subsets of $\mathbb{N}$, where the elements of A and B are labelled in increasing order (i.e. $a_1, < \cdots < a_k$ and $b_1 < \cdots < b_k$). We define the order relation $<_c$ on the family $\binom{\mathbb{N}}{k}$ of k-subsets of $\mathbb{N}$ by

$$A <_c B \iff \text{there exists } p \in [k] \text{ such that } a_p < b_p \text{ and } a_i = b_i \text{ for all } i > p.$$

Note. It is not difficult to show that the relation $<_c$ defines a *well-ordering* of the family $\binom{\mathbb{N}}{k}$, by which we mean that the following are satisfied for every $A, B, C \in \binom{\mathbb{N}}{k}$.

(i) If $A <_c B$ and $B <_c C$ then $A <_c C$, (TRANSITIVITY)

(ii) Precisely one of $A <_c B$ and $B <_c A$ is true, (TOTALITY)

(iii) Every non-empty subfamily $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ has a least element under $<_c$. (MINIMUM)

Example 12.4. For example, consider the set $\binom{\mathbb{N}}{4}$. We abbreviate $\{a_1, a_2, a_3, a_4\} \in \mathbb{N}$ to $a_1a_2a_3a_4$. The first few elements in colexicographical order are:

$$\begin{aligned}
&1234 <_c 1235 <_c 1245 <_c 1345 <_c 2345 <_c 1236 <_c 1246 <_c 1346 <_c 2346 <_c 1256 \\
&<_c 1356 <_c 2356 <_c 1456 <_c 2456 <_c 3456 <_c 1237 <_c 1247 <_c 1347 <_c 2347 <_c 1257 \\
&<_c 1357 <_c 2357 <_c 1457 <_c 2457 <_c 3457 <_c 1267 <_c 1367 <_c 2367 <_c 1467 <_c 2467 \\
&<_c 3467 <_c 1567 <_c 2567 <_c 3567 <_c 4567 <_c 1238 <_c 1248 <_c 1348 <_c 2348 <_c 1258 \\
&<_c 1358 <_c 2358 <_c 1458 <_c 2458 <_c 3458 <_c 1268 <_c 1368 <_c 2368 <_c 1468 <_c 2468 \\
&<_c 3468 <_c 1568 <_c 2568 <_c 3568 <_c 4568 <_c 1278 <_c 1378 <_c 2378 <_c 1478 <_c 2478 \\
&<_c 3478 <_c 1578 <_c 2578 <_c 3578 <_c 4578 <_c 1678 <_c 2678 <_c 3678 <_c 4678 <_c 5678.
\end{aligned}$$

Notation. The family of the first m sets in $\binom{\mathbb{N}}{k}$ is denoted by $\mathcal{C}_{k,m}$. Therefore in the previous example, we listed the elements of $\mathcal{C}_{4,8}$.

Remark 12.5. The first $\binom{t}{k}$ k-sets in colexicographical order of $\binom{\mathbb{N}}{k}$ are the sets in the family $\binom{[t]}{k}$, so

$$\mathcal{C}_{k,\binom{t}{k}} = \binom{[t]}{k}.$$

Also for $r \leqslant k$, the rth shadow of $\binom{[t]}{k}$ is $\binom{[t]}{r}$, so we have

$$|\partial_r(\mathcal{C}_{k,\binom{t}{k}})| = \binom{t}{r}.$$

Theorem 12.6 (Kruskal-Katona Theorem)**.** *Let $\mathcal{A}$ be an r-uniform family of size m. Then*

$$|\partial_s(\mathcal{A})| \geqslant |\partial_s(\mathcal{C}_{r,m})|.$$

In other words, the optimal family is one whose elements are in colexicographical order. We do not give a proof of the theorem 12.6 here, however we show that it gives us a nice way to prove the EKR theorem as a corollary.

Proof of EKR theorem (Daykin, 1974). Let $\mathcal{B} = \{[n] \setminus A : A \in \mathcal{A}\}$, and let $\mathcal{C} = \partial_k(\mathcal{B})$. Every $C \in \mathcal{C}$ is a k-subset of some $B \in \mathcal{B}$, therefore $C \subseteq [n] \setminus A$ for some $A \in \mathcal{A}$, and thus $C \cap A = \emptyset$. We can therefore conclude that $C \notin \mathcal{A}$, since $\mathcal{A}$ is an intersecting family.

Thus $\mathcal{A}$ and $\mathcal{C}$ are disjoint subfamilies of $\binom{[n]}{k}$, and so $|\mathcal{A}| + |\mathcal{C}| = |\mathcal{A} \cup \mathcal{C}| \leqslant |\binom{[n]}{k}| = \binom{n}{k}$. ⊛

For contradiction, suppose that the EKR theorem is false, i.e. that $|A| > \binom{n-1}{k-1}$. Since $\mathcal{B}$ is an $(n-k)$-uniform family and $|\mathcal{B}| = |\mathcal{A}| > \binom{n-1}{k-1} = \binom{n-1}{n-k}$, then by the Kruskal-Katona theorem,

$$|\mathcal{C}| = |\partial_k(\mathcal{B})| \geqslant |\partial_k(\mathcal{C}_{n-k,|\mathcal{B}|})| \geqslant |\partial_k(\mathcal{C}_{n-k,\binom{n-1}{n-k}})| = \binom{n-1}{k}$$

by remark 12.5. But now $|\mathcal{A}| + |\mathcal{C}| > \binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}$, contradicting ⊛. Therefore the Erdős-Ko-Rado theorem must be true. □

The End