COMBINATORICS MT454 / MT5454

MARK WILDON

These notes are intended to give the logical structure of the course; proofs and further remarks will be given in lectures. Further installments will be issued as they are ready. All handouts and problem sheets will be put on Moodle.

I would very much appreciate being told of any corrections or possible improvements to these notes.

You are warmly encouraged to ask questions in lectures, and to talk to me after lectures and in my office hours. I am also happy to answer questions about the lectures or problem sheets by email. My email address is `mark.wildon@rhul.ac.uk`.

Lectures: Monday 12 noon in MFLEC, Tuesday 12 noon in HTL1 and Friday 1pm in ABLT2.

Office hours in McCrea 240: Tuesday 11am, Wednesday 2pm and Friday 3pm.

Date: First term 2012/13.

1. Introduction

Combinatorial arguments may be found lurking in all branches of mathematics. Many people first become interested in mathematics by a combinatorial problem. But, strangely enough, at first many mathematicians tended to sneer at combinatorics. Thus one finds:

> *"Combinatorics is the slums of topology."*
>
> J. H. C. Whitehead (early 1900s, attr.)

Fortunately attitudes have changed, and the importance of combinatorial arguments is now widely recognised:

> *"The older I get, the more I believe that at the bottom of most deep mathematical problems there is a combinatorial problem."*
>
> I. M. Gelfand (1990)

Combinatorics is a very broad subject. Often it will be useful to prove the same result in different ways, in order to see different combinatorial techniques at work. There is no shortage of interesting and easily understood motivating problems.

Overview. This course will give a straightforward introduction to four related areas of combinatorics. Each is the subject of current research, and taken together, they give a good idea of what the subject is about.

- **(A) Enumeration:** Binomial coefficients and their properties. Principle of Inclusion and Exclusion and applications. Rook polynomials.
- **(B) Generating Functions:** Ordinary generating functions and recurrence relations. Partitions and compositions. Catalan Numbers. Derangements.
- **(C) Ramsey Theory:** "Complete disorder is impossible". Pigeonhole Principle. Graph colouring.
- **(D) Probabilistic Methods:** Linearity of expectation. First moment method. Applications to counting permutations. Lovász Local Lemma.

Recommended Reading.

[1] *A First Course in Combinatorial Mathematics.* Ian Anderson, OUP 1989, second edition.
[2] *Discrete Mathematics.* N. L. Biggs, OUP 1989.
[3] *Combinatorics: Topics, Techniques, Algorithms.* Peter J. Cameron, CUP 1994.
[4] *Concrete Mathematics.* Ron Graham, Donald Knuth and Oren Patashnik, Addison-Wesley 1994.
[5] *Invitation to Discrete Mathematics.* Jiri Matoušek and Jaroslav Nešetřil, OUP 2009, second edition.

[6] *Probability and Computing: Randomized Algorithms and Probabilistic Analysis.* Michael Mitzenmacher and Eli Upfal, CUP 2005.

[7] *generatingfunctionology.* Herbert S. Wilf, A K Peters 1994, second edition. Available from `http://www.math.upenn.edu/~wilf/DownldGF.html`.

In parallel with the first few weeks of lectures, you will be asked to do some reading from *generatingfunctionology*: the problem sheets will make clear what is expected.

Prerequisites.

- Permutations and their decomposition into disjoint cycles. (Required for derangements and for some applications in Part D.)
- Basic definitions of graph theory: vertices, edges, complete graphs. (Required for Part C on Ramsey Theory.)
- Basic knowledge of discrete probability. This will be reviewed in lectures when we get to part D of the course. A handout with all the background results needed from probability theory will be issued later in term.

Problem sheets and exercises. There will be weekly problem sheets; the first will be due in on Monday 15th October. Exercises set in these notes are intended to be simple tests that you are following the material. Some will be done in lectures. Please may sure you can do all of them.

Note on optional questions. Optional questions on problem sheets are included for interest and to give extra practice. Harder optional questions are marked ($\star$). **If you can do the compulsory questions and know the bookwork, i.e. the definitions, main theorems, and their proofs, as set out in the handouts and lectures, you should do very well in the exam.**

2. Derangements

In the first two lectures we will see the Derangements Problem and one way to solve it by *ad-hoc* methods. Later in the course we will develop techniques that can be used to solve this problem more easily.

Definition 2.1. A *permutation* of a set X is a bijective function

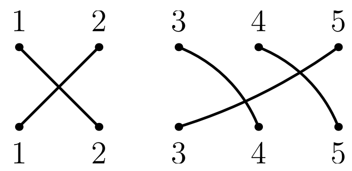
$$\sigma : X \to X.$$

A *fixed point* of a permutation σ of X is an element $x \in X$ such that $\sigma(x) = x$. A permutation is a *derangement* if it has no fixed points.

Usually we will consider permutations of $\{1, 2, \dots, n\}$ for some natural number $n \in \mathbf{N}$. It is often useful to represent permutations by diagrams. For example, the diagram below shows the permutation $\sigma : \{1, 2, 3, 4, 5\} \to \{1, 2, 3, 4, 5\}$ defined by

$$\sigma(1) = 2,\ \sigma(2) = 1,\ \sigma(3) = 4,\ \sigma(4) = 5,\ \sigma(5) = 3.$$

Note that σ is a derangement.



Exercise: For $n \in \mathbf{N}$, how many permutations are there of $\{1, 2, \dots, n\}$? How many of these permutations have 1 as a fixed point?

The principle used to solve this exercise, that when one choice is made after another, the number of choices should be multiplied, will be used many times in this course. In the case where one choice does not affect the next, so we first choose an element of a set A, then an element of a set B, the principle simply says that $|A \times B| = |A||B|$.

More generally, if an object can be specified uniquely by a sequence of n choices so that, when making the ith choice, we always have exactly c_i possibilities to choose from, then there are exactly $c_1 c_2 \dots c_n$ objects.

Problem 2.2 (Derangements)**.** *Let X be a set of size n. How many permutations of X are derangements?*

Let d_n be the number of permutations of $\{1, 2, \dots, n\}$ that are derangements. By definition, although you may regard this as a convention if you prefer, $d_0 = 1$.

Exercise: Check, by listing permutations, that $d_1 = 0$, $d_2 = 1$, $d_3 = 2$ and $d_4 = 9$.

To solve the derangements problem we shall find a recurrence for the numbers d_n.

Lemma 2.3. *If $n \geq 2$ then there are $d_{n-2} + d_{n-1}$ derangements σ of $\{1, 2, \ldots, n\}$ such that $\sigma(1) = 2$.*

Notice the use of another basic counting principle in Lemma 2.3: if we can partition the objects we are counting into two disjoint sets A and B, then the total number of objects is $|A| + |B|$.

Theorem 2.4. *If $n \geq 2$ then $d_n = (n-1)(d_{n-2} + d_{n-1})$.*

Using this recurrence relation it is easy to find values of d_n for much larger n. At this point N. J. A. Sloane's Online Encyclopedia of Integer Sequences is a reliable guide to whether a sequence has already been studied: see `www.research.att.com/~njas/sequences/`.

Corollary 2.5. *For all $n \in \mathbf{N}_0$,*

$$d_n = n!\Big(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!}\Big).$$

Exercise: Check directly that the right-hand side is an integer.

A more systematic way to derive Corollary 2.5 from Theorem 2.4 will be seen in Part B of the course. Question 9 on Sheet 1 gives an alternative proof that does not require knowing the answer in advance.

The proof of Corollary 2.5 and Question 9 show that it is helpful to consider the probability $d_n/n!$ that a permutation of $\{1, 2, \ldots, n\}$, chosen uniformly at random, is a derangement. Here 'uniformly at random' means that each of the $n!$ permutations of $\{1, 2, \ldots, n\}$ is equally likely to be chosen.

Theorem 2.6. *Two probabilistic results on derangements.*

(i) *The probability that a permutation of $\{1, 2, \ldots, n\}$, chosen uniformly at random, is a derangement tends to $1/\mathrm{e}$ as $n \to \infty$.*

(ii) *The average number of fixed points of a permutation of $\{1, 2, \ldots, n\}$ is 1.*

We shall prove more results like this in Part D of the course.

Part A: Enumeration

3. Binomial coefficients and counting problems

The following notation is probably already familiar to you.

Notation 3.1. If Y is a set of size k then we say that Y is a *k-set*, and write $|Y| = k$. To emphasise that Y is a subset of some other set X then we may say that Y is a *k-subset of X*.

We shall define binomial coefficients combinatorially.

Definition 3.2. Let n, $k \in \mathbf{N}_0$. Let $X = \{1, 2, \ldots, n\}$. The *binomial coefficient* $\binom{n}{k}$ is the number of k-subsets of X.

By this definition, if $k \notin \mathbf{N}_0$ then $\binom{n}{k} = 0$. Similarly if $k > n$ then $\binom{n}{k} = 0$. It should be clear that we could replace X with any other set of size n and we would define the same numbers $\binom{n}{k}$.

We should check that the combinatorial definition agrees with the usual definition.

Lemma 3.3. *If n, $k \in \mathbf{N}_0$ and $k \leq n$ then*

$$\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

Many of the basic properties of binomial coefficients can be given combinatorial proofs involving explicit bijections. We shall say that such proofs are *bijective*.

Lemma 3.4. *If n, $k \in \mathbf{N}_0$ then*

$$\binom{n}{k} = \binom{n}{n-k}.$$

Lemma 3.5 (Fundamental Recurrence)**.** *If n, $k \in \mathbf{N}$ then*

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Binomial coefficients are so-named because of the famous binomial theorem. (A binomial is a term of the form $x^r y^s$.)

Theorem 3.6 (Binomial Theorem)**.** *Let x, $y \in \mathbf{C}$. If $n \in \mathbf{N}_0$ then*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Exercise: give inductive or algebraic proofs of the previous three results.

Exercise: in New York, how many ways can one start at a junction and walk to another junction 4 blocks away to the east and 3 blocks away to the north?

We can now answer a basic combinatorial question: *How many ways are there to put k balls into n numbered urns?* The answer depends on whether the balls are distinguishable. We may consider urns of unlimited capacity, or urns that can only contain one ball.

	Numbered balls	Indistinguishable balls
≤ 1 ball per urn		
unlimited capacity		

Three of the entries can be found fairly easily. The entry in the bottom-right can be found in many different ways: two will be demonstrated in this lecture.

Theorem 3.7. *Let $n \in \mathbf{N}$ and let $k \in \mathbf{N}_0$. The number of ways to place k indistinguishable balls into n numbered urns of unlimited capacity is $\binom{n+k-1}{k}$.*

The following reinterpretation of Theorem 3.7 can be useful.

Corollary 3.8. *Let $n \in \mathbf{N}$ and let $k \in \mathbf{N}_0$. The number of k-tuples $(t_1, \ldots, t_n)$ such that $t_1, t_2, \ldots, t_n \in \mathbf{N}_0$ and*

$$t_1 + t_2 + \cdots + t_n = k$$

is $\binom{n+k-1}{k}$.

4. Further binomial identities

This is a vast subject and we shall only cover a few aspects. Particularly recommended for further reading is Chapter 5 of *Concrete Mathematics*, [4] in the list on page 2.

Arguments with subsets. The two identities below are among the most useful in practice.

Lemma 4.1 (Subset of a subset). *If k, r, $n \in \mathbf{N}_0$ and $k \le r \le n$ then*

$$\binom{n}{r}\binom{r}{k} = \binom{n}{k}\binom{n-k}{r-k}.$$

Lemma 4.2 (Vandermonde's convolution). *If a, $b \in \mathbf{N}_0$ and $m \in \mathbf{N}_0$ then*

$$\sum_{k=0}^{m} \binom{a}{k}\binom{b}{m-k} = \binom{a+b}{m}.$$

Corollaries of the Binomial Theorem. The following results can be obtained by making a strategic choice of x and y in the Binomial Theorem.

Corollary 4.3. *If $n \in \mathbf{N}$ then*

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n-1} + \binom{n}{n} = 2^n,$$

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^{n-1}\binom{n}{n-1} + (-1)^n\binom{n}{n} = 0.$$

Corollary 4.4. *For all $n \in \mathbf{N}$ there are equally many subsets of $\{1, 2, \ldots, n\}$ of even size as there are of odd size.*

Corollary 4.5. *If $n \in \mathbf{N}_0$ and $b \in \mathbf{N}$ then*

$$\binom{n}{0}b^n + \binom{n}{1}b^{n-1} + \cdots + \binom{n}{n-1}b + \binom{n}{n} = (1+b)^n.$$

There is a nice bijective proof of Corollary 4.5; it will appear as a question with hints on Sheet 2.

Some identities visible in Pascal's Triangle. There are a number of nice identities that express row, column or diagonal sums in Pascal's Triangle.

Lemma 4.6 (Alternating row sums). *If $n \in \mathbf{N}$, $r \in \mathbf{N}_0$ and $r \le n$ then*

$$\sum_{k=0}^{r} (-1)^k \binom{n}{k} = (-1)^r \binom{n-1}{r}.$$

Perhaps surprisingly, there is no simple formula for the unsigned row sums $\sum_{k=0}^{r} \binom{n}{k}$.

Lemma 4.7 (Diagonal sums, a.k.a. parallel summation)**.** *If* $n \in \mathbf{N}$, $r \in \mathbf{N}_0$ *then*

$$\sum_{k=0}^{r} \binom{n+k}{k} = \binom{n+r+1}{r}.$$

For the column sums on Pascal's Triangle, see Sheet 1, Question 3. For the other diagonal sum, see Sheet 1, Question 7.

5. Principle of Inclusion and Exclusion

The Principle of Inclusion and Exclusion (PIE) is way to find the size of a union of a finite collection of subsets of a finite *universe set* X. The universe set we take will depend on the problem we are solving. If A is a subset of X, we denote by $\bar{A}$ the complement of A in X; i.e.,

$$\bar{A} = X \backslash A = \{x \in X : x \notin A\}.$$

We start with the two smallest non-trivial examples of the Principle of Inclusion and Exclusion.

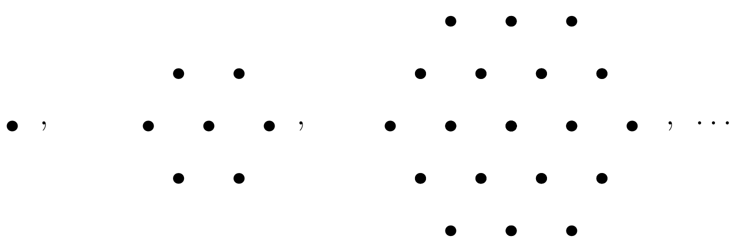
Example 5.1. If A, B, C are subsets of a finite set X then

$$\begin{aligned} |A \cup B| &= |A| + |B| - |A \cap B| \\ |\overline{A \cup B}| &= |X| - |A| - |B| + |A \cap B| \end{aligned}$$

and

$$\begin{aligned} |A \cup B \cup C| &= |A| + |B| + |C| \\ &\qquad - |A \cap B| - |B \cap C| - |C \cap A| + |A \cap B \cap C| \\ |\overline{A \cup B \cup C}| &= |X| - |A| - |B| - |C| \\ &\qquad + |A \cap B| + |B \cap C| + |C \cap A| - |A \cap B \cap C| \end{aligned}$$

Example 5.2. The n-th (centred) hexagonal number is the number of dots in the n-th digure below. The formula for $|A \cup B \cup C|$ gives a nice way a formula for these numbers.



It is easier to find the sizes of the intersections of the three rhombi making up each hexagon than it is to find the sizes of their unions. Whenever this situation occurs, the PIE is likely to work well.

In the general setting we have a finite universe set X and subsets $A_1, A_2, \ldots, A_n \subseteq X$. For each non-empty subset $I \subseteq \{1, 2, \ldots, n\}$ we define

$$A_I = \bigcap_{i \in I} A_i.$$

Thus A_I is the set of elements which belong to all the sets A_i for $i \in I$. For example, if $i, j \in \{1, 2, \ldots, n\}$ then $A_{\{i\}} = A_i$ and $A_{\{i,j\}} = A_i \cap A_j$. By convention we set

$$A_\varnothing = X.$$

Theorem 5.3 (Principle of Inclusion and Exclusion)**.** *If $A_1, A_2, \ldots, A_n$ are subsets of a finite set X then*

$$|\overline{A_1 \cup A_2 \cup \cdots \cup A_n}| = \sum_{I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|} |A_I|.$$

Exercise: Check that Theorem 5.3 holds when $n = 1$ and check that it agrees with Example 5.1 when $n = 2$ and $n = 3$.

Exercise: Deduce from Theorem 5.3 that

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{\substack{I \subseteq \{1,2,\ldots,n\} \\ I \neq \varnothing}} (-1)^{|I|-1} |A_I|.$$

Application to derangements. The Principle of Inclusion and Exclusion gives a particularly elegant proof of the formula for the derangement numbers d_n first proved in Corollary 2.5:

$$d_n = n!\Big(1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}\Big).$$

Recall from Definition 2.1 that a permutation

$$\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$$

is a derangement if and only if it has no fixed points. Let X be the set of all permutations of $\{1, 2, \ldots, n\}$ and let

$$A_i = \{\sigma \in X : \sigma(i) = i\}$$

be the set of permutations which have i as a fixed point. To apply the PIE we need the results in the following lemma.

Lemma 5.4. (i) *A permutation $\sigma \in X$ is a derangement if and only if*

$$\sigma \in \overline{A_1 \cup A_2 \cup \cdots \cup A_n}.$$

(ii) *If $I \subseteq \{1, 2, \ldots, n\}$ then A_I consists of all permutations of $\{1, 2, \ldots, n\}$ which fix the elements of I. If $|I| = k$ then*

$$|A_I| = (n - k)!.$$

It is often helpful to think of each A_i as the set of all objects in X satisfying a property P_i. Then the Principle of Inclusion and Exclusion counts all the objects in X that satisfy *none* of the properties $P_1, \ldots, P_n$. In the derangements example

$$P_i(\sigma) = \text{`}\sigma \text{ has } i \text{ as a fixed point}'$$

and we count the permutations σ such that $P_i(\sigma)$ is false for all $i \in \{1, 2, \ldots, n\}$.

Prime numbers and Euler's ϕ function. Suppose we want to find the number of primes less than some number M. One approach, which is related to the Sieve of Eratosthenes, uses the Principle of Inclusion and Exclusion.

Example 5.5. Let $X = \{1, 2, \ldots, 48\}$. We define three subsets of X:

$$\begin{aligned} B(2) &= \{m \in X : \ m \text{ is divisible by } 2\} \\ B(3) &= \{m \in X : \ m \text{ is divisible by } 3\} \\ B(5) &= \{m \in X : \ m \text{ is divisible by } 5\}. \end{aligned}$$

Any composite number ≤ 48 is divisible by either 2, 3 or 5. So

$$\overline{B(2) \cup B(3) \cup B(5)} = \{1\} \cup \{p : 5 < p \leq 48, \ \ p \text{ is prime}\}.$$

We will find the size of the left-hand side using the PIE, and hence count the number of primes ≤ 48.

The example can be generalized to count numbers not divisible by any of a specified set of primes. Recall that if $x \in \mathbf{R}$ then $\lfloor x \rfloor$ denotes the largest natural number $\leq x$.

Lemma 5.6. *Let r, $M \in \mathbf{N}$. There are exactly $\lfloor M/r \rfloor$ numbers in $\{1, 2, \ldots, M\}$ that are divisible by r.*

Theorem 5.7. *Let $p_1, \ldots, p_n$ be distinct prime numbers and let $M \in \mathbf{N}$. The number of natural numbers $\leq M$ that are not divisible by any of primes $p_1, \ldots, p_n$ is*

$$\sum_{I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|} \left\lfloor \frac{M}{\prod_{i \in I} p_i} \right\rfloor.$$

For $M \in \mathbf{N}$, let $\pi(M)$ be the number of prime numbers $\leq M$. It is possible to use Theorem 5.7 to show that there is a constant C such that

$$\pi(M) \leq \frac{CM}{\log \log M}$$

for all $M \in \mathbf{N}$. This is a bit off-track for this course, but I would be happy to go through the proof in an office-hour or supply a reference.

The next example will be helpful for the questions on Sheet 2. In it, we say that numbers n, M are *coprime* if n and M have no common prime divisors. For example, 12 and 35 are coprime, but 7 and 14 are not.

Example 5.8. Let $M = pqr$ where p, q, r are distinct prime numbers. The numbers of natural numbers less than or equal to pqr that are coprime to M is

$$M\Big(1-\frac{1}{p}\Big)\Big(1-\frac{1}{q}\Big)\Big(1-\frac{1}{r}\Big).$$

6. Rook polynomials

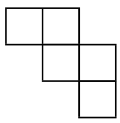
Many enumerative problems can be expressed as problems about counting permutations with some restriction on their structure. The derangements problem is a typical example. In this section we shall see a unified way to solve this sort of problem.

Recommended reading: Ian Anderson, *A First Course in Combinatorial Mathematics*, §5.2 ([1] on the list on page 2) and Victor Bryant, *Aspects of Combinatorics*, Chapter 12 (Cambridge University Press). Examples 6.3 and 6.4 below are based on those in Bryant's book.

Definition 6.1. A *board* is a subset of the squares of an $n \times n$ grid. Given a board B, we let $r_k(B)$ denote the number of ways to place k rooks on B, so that no two rooks are in the same row or column. Such rooks are said to be *non-attacking*. The *rook polynomial* of B is defined to be

$$f_B(x) = r_0(B) + r_1(B)x + r_2(B)x^2 + \cdots + r_n(B)x^n.$$

Example 6.2. Let B be the board shown below.



The rook polynomial of B is $1 + 5x + 6x^2 + x^3$.

Exercise: Let B be a board. Convince yourself that $r_0(B) = 1$ and that $r_1(B)$ is the number of squares in B.

Example 6.3. After the recent spate of cutbacks, only four professors remain at the University of Erewhon. Prof. W can lecture courses 1 or 4; Prof. X is an all-rounder and can lecture 2, 3 or 4; Prof. Y refuses to lecture anything except 3; Prof. Z can lecture 1 or 2. If each professor must lecture exactly one course, how many ways are there to assign professors to courses?

Example 6.4. How many derangements σ of $\{1,2,3,4,5\}$ have the property that $\sigma(i) \neq i + 1$ for $1 \leq i \leq 4$?

Lemma 6.5. *The rook polynomial of the $n \times n$ board is*

$$\sum_{k=0}^{n} k! \binom{n}{k}^2 x^k.$$

The two following lemmas are very useful when calculating rook polynomials.

Lemma 6.6. *Let B be a board. Suppose that the squares in B can be partitioned into sets C and D so that no square in C lies in the same row or column as a square of D. Then*

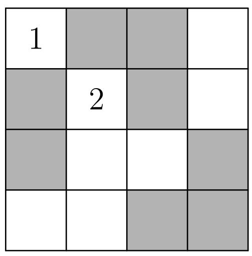
$$f_B(x) = f_C(x) f_D(x).$$

Rook polynomials are, in particular, generating functions. This is the first of many times that multiplying generating functions will help us to solve combinatorial problems.

Lemma 6.7. *Let B be a board and let s be a square in B. Let C be the board obtained from B by deleting s and let D be the board obtained from B by deleting the entire row and column containing s. Then*

$$f_B(x) = f_C(x) + x f_D(x).$$

Example 6.8. The rook-polynomial of the boards in Examples 6.3 and 6.4 can be found using Lemma 6.7. For the board in Example 6.3 it works well to apply the lemma first to the square marked 1, then to the square marked 2 (in the new boards).



Our final result on rook polynomials is often the most useful in practice. The proof uses the Principle of Inclusion and Exclusion. The following lemma isolates the key idea. Its proof needs the same idea we used in Lemma 5.4(ii) to count permutations with a specified set of fixed points.

Lemma 6.9. *Let B be a board contained in an $n \times n$ grid and let $0 \le k \le n$. The number of ways to place k red rooks on B and $n-k$ blue rooks anywhere on the grid, so that the n rooks are non-attacking, is $r_k(B)(n-k)!$.*

Theorem 6.10. *Let B be a board contained in an $n \times n$ grid. Let $\bar{B}$ denote the board formed by all the squares in the grid that are not in B. The number of ways to place n non-attacking rooks on $\bar{B}$ is*

$$n! - (n-1)!r_1(B) + (n-2)!r_2(B) - \cdots + (-1)^n r_n(B).$$

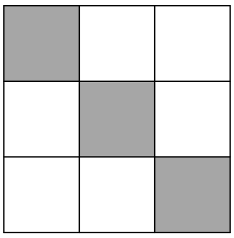
As an easy corollary we get our third proof of the derangements formula (Corollary 2.5), that

$$d_n = n!\Big(1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}\Big).$$

See Problem Sheet 3 for some other applications of Theorem 6.10.

Theorem 6.10 is one of the harder results in the course. If you find the proof difficult, you may find the following exercise helpful.

Exercise: Let $n = 3$ and let B be the board formed by the *shaded* squares below.



Draw the rook placements lying in each of the sets $A_{\varnothing}$, $A_{\{1\}}$, $A_{\{2\}}$, $A_{\{3\}}$, $A_{\{1,2\}}$, $A_{\{1,3\}}$, $A_{\{2,3\}}$, $A_{\{1,2,3\}}$ defined in the proof of Theorem 6.10, and check the main claim in the proof for $k = 0, 1, 2, 3$. For instance, for $k = 1$, you should find that $|A_{\{1\}}| + |A_{\{2\}}| + |A_{\{3\}}|$ is the number of non-attacking placements with one red rook on B and two blue rooks anywhere on the grid; according to Lemma 6.9 there are $r_1(B)(3-1)!$ such placements.

Part B: Generating Functions

7. Introduction to generating functions

Generating functions can be used to solve the sort of recurrence relations that often arise in combinatorial problems. But better still, they can help us to think about combinatorial problems in new ways and suggest new results.

Definition 7.1. The *ordinary generating function* associated to the sequence $a_0, a_1, a_2, \ldots$ is the power series

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots .$$

To indicate that $F(x)$ is the ordinary generating function of the sequence $a_0, a_1, a_2, \ldots$ we may use the notation in §2.2 of Wilf *generatingfunctionology* and write

$$(a_n) \stackrel{ogf}{\longleftrightarrow} F(x).$$

Usually we shall drop the word 'ordinary' and just write 'generating function'.

If there exists $N \in \mathbf{N}$ such that $a_n = 0$ if $n > N$, then the generating function of the sequence $a_0, a_1, a_2, \ldots$ is a polynomial. Rook polynomials (see Definition 6.1) are therefore generating functions.

Operations on generating functions. Let $F(x) = \sum_{n=0}^{\infty} a_n x^n$ and $G(x) = \sum_{n=0}^{\infty} b_n x^n$ be generating functions. From

$$F(x) + G(x) = \sum_{n=0}^{\infty} (a_n + b_n) x^n$$

and

$$F(x)G(x) = \sum_{n=0}^{\infty} c_n x^n$$

where $c_n = \sum_{m=0}^{n} a_m b_{n-m}$. The derivative of $F(x)$ is

$$F'(x) = \sum_{n=0}^{\infty} n x^{n-1}.$$

Note that if $(a_n) \stackrel{ogf}{\longleftrightarrow} F(x)$ and $(b_n) \stackrel{ogf}{\longleftrightarrow} G(x)$ then

$$(a_n + b_n) \stackrel{ogf}{\longleftrightarrow} F(x) + G(x).$$

The sequence (c_n) such that $(c_n) \stackrel{ogf}{\longleftrightarrow} F(x)G(x)$ often arises in combinatorial problems. This was seen for rook polynomials in Lemma 6.6, and will be studied in §9.

It is also possible to define $1/F(x)$ whenever $a_0 \neq 0$. By far the most important case is the case $F(x) = 1 - x$, when

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

is the usual formula for the sum of a geometric progression.

Analytic and formal interpretations. There are at least two ways to think of a generating function $\sum_{n=0}^{\infty} a_n x^n$. Either:

- As a formal power series with x acting as a place-holder. This is the 'clothes-line' interpretation (see Wilf *generatingfunctionology*, page 4), in which we regard the power-series merely as a convenient way to display the terms in our sequence.
- As a function of a real or complex variable x convergent when $|x| < r$, where r is the radius of convergence of $\sum_{n=0}^{\infty} a_n x^n$.

The formal point of view is often the most convenient because it allows us to define and manipulate power series by the operations on the previous page without worrying about convergence. From this point of view,

$$0! + 1!x + 2!x^2 + 3!x^3 + \cdots$$

is a perfectly respectable formal power series, even thought it only converges when $x = 0$. The analytic point of view is useful for proving asymptotic results.[1]

All the generating functions one normally encounters have positive radius of convergence, so in practice, the two approaches are equivalent. For a more careful discussion of these issues and the general definition of $1/F(x)$, see §2.1 of Wilf *generatingfunctionology*.

Examples of generating functions. We shall look at two typical problems that can be solved using ordinary generating functions. In each case, the generating function for the sequence is at least as useful as an explicit formula for the terms in the sequence.

Example 7.2. How many ways are there to tile a $2 \times n$ path with bricks that are either 1×2 or 2×1?

The result in the next example has already been proved as Corollary 3.8. We shall use generating functions to give an independent approach.

---

[1]From the analytic perspective, the formula for the derivative $F'(x)$ on the previous page expresses a non-trivial theorem, namely that power series are differentiable functions, with derivatives given by term-by-term differentiation. A similar remark applies to the formulae for the sum $F(x) + G(x)$ and product $F(x)G(x)$.

Example 7.3. Let $k \in \mathbf{N}_0$ How many n-tuples $(x_1, \ldots, x_n)$ are there such that $x_i \in \mathbf{N}_0$ for each i and $x_1 + \cdots + x_n = k$? Such an n-tuple is said to be a *composition* of k with *n parts*.

To complete Example 7.3 we used the power series

$$\frac{1}{(1-x)^n} = \sum_{k=0}^{\infty} \binom{k+n-1}{k} x^k$$

found on Question 5 of Sheet 3. A more general result is stated below.

Theorem 7.4 (Binomial Theorem for general exponent)**.** *If $\alpha \in \mathbf{R}$ then*

$$(1+y)^\alpha = \sum_{n=0}^{\infty} \frac{\alpha(\alpha-1)\ldots(\alpha-(n-1))}{n!} y^n$$

for all y such that $|y| < 1$.

Exercise: Let $\alpha \in \mathbf{Z}$.

(i) Show that if $\alpha \geq 0$ then Theorem 7.4 agrees with the Binomial Theorem for integer exponents, proved in Theorem 3.6.

(ii) Show that if $\alpha < 0$ then Theorem 7.4 agrees with Question 5 on Sheet 3. (Substitute $-x$ for y.)

We shall need the case $\alpha = 1/2$ of the general Binomial Theorem to find the Catalan Numbers in §9.

8. Recurrence relations and asymptotics

We have seen that combinatorial problems often lead to recurrence relations. For example, in §2 we found the derangement numbers d_n by solving the recurrence relation in Theorem 2.4. See also Questions 5 and 7 on Sheet 1 for other examples.

Generating functions are very useful for solving recurrence relations. The method is clearly explained at the end of §1.2 of Wilf *generatingfunctionology*. Given a recurrence satisfied by the sequence a_0, a_1, $a_2, \ldots$ proceed as follows:

(a) Use the recurrence to write down an equation satisfied by the generating function $F(x) = \sum_{n=0}^{\infty} a_n x^n$;

(b) Solve the equation to get a closed form for the generating function;

(c) Use the closed form for the generating function to find a formula for the coefficients.

Step (a) may seem the most mysterious, but it will become routine with practice. To obtain terms like na_{n-1}, try differentiating $F(x)$. Powers of x will usually be needed to get everything to match up correctly. In Step (c) it is often necessary to use partial fractions.

Example 8.1. Will solve (i) using generating functions, and perform step (a) of the three-step programme on (ii).

(i) $a_{n+2} = 5a_{n+1} - 6a_n$ for $n \in \mathbf{N}_0$, $a_0 = A$, $a_1 = B$;

(ii) $b_r = 3b_{r-1} - 4b_{r-3}$ for $r \geq 3$.

Another way to deal with (i) is to first rewrite it as $a_n = 5a_{n-1} - 6a_{n-2}$ for $n \geq 2$; then the shifts are done by multiplication by x and x^2 rather than division.

The next theorem gives a general form for the partial fraction expressions needed to solve these recurrences. Recall that if $f(x) = b_d x^d + b_{d-1}x^{d-1} + \cdots + b_0$ and $b_d \neq 0$ then f is said to have *degree* d; we write this as $\deg f = d$.

Theorem 8.2. *Let $f(x)$ and $g(x)$ be polynomials with $\deg f < \deg g$.*

(i) *If $g(x) = \alpha(x - \beta_1)(x - \beta_2)\ldots(x - \beta_k)$ where $\alpha, \beta_1, \beta_2, \ldots, \beta_k$ are non-zero complex numbers, then there exist $C_1, \ldots, C_k \in \mathbf{C}$ such that*

$$\frac{f(x)}{g(x)} = \frac{C_1}{1 - x/\beta_1} + \cdots + \frac{C_k}{1 - x/\beta_k}.$$

(ii) *If $g(x) = \alpha(x-\beta_1)^{d_1}(x-\beta_2)^{d_2}\ldots(x-\beta_k)^{d_k}$ where $\alpha, \beta_1, \beta_2, \ldots, \beta_k$ are non-zero complex numbers and $d_1, d_2, \ldots, d_k \in \mathbf{N}$, then there exist polynomials $P_1, \ldots, P_k$ such that $\deg P_i < d_i$ and*

$$\frac{f(x)}{g(x)} = \frac{P_1(1 - x/\beta_1)}{(1 - x/\beta_1)^{d_1}} + \cdots + \frac{P_k(1 - x/\beta_k)}{(1 - x/\beta_k)^{d_k}}$$

where $P_i(1 - x/\beta_i)$ is P_i evaluated at $1 - x/\beta_i$.

Example 8.3. As an example of Theorem 8.2(ii), will finish steps (b) and (c) of the three-step programme on the recurrence in Example 8.1, $b_r = 3b_{r-1} - 4b_{r-3}$ for $r \geq 3$, with initial values $b_0 = 1$, $b_1 = 1$, $b_2 = 0$.

The next exercise gives a similar example of Theorem 8.2.

Exercise: As in Example 7.2, let a_n be the number of ways to tile a a $2 \times n$ path with bricks that are either 1×2 ($\square\!\square$) or 2×1 ($\begin{smallmatrix}\square\\\square\end{smallmatrix}$). We saw that $a_n = a_{n-1} + a_{n-2}$, and that the generating function

$$F(x) = \sum_{n=0}^{\infty} a_n x^n$$

satisfies $(1-x-x^2)F(x) = 1$. Show that

$$x^2 + x - 1 = (x-\beta_1)(x-\beta_2)$$

where $\beta_1 = -1/2 + \sqrt{5}/2$ and $\beta_2 = -1/2 - \sqrt{5}/2$. Show that $1/\beta_1 = 1/2 + \sqrt{5}/2$ and $1/\beta_2 = 1/2 - \sqrt{5}/2$ and deduce from Theorem 8.2(i) that

$$a_n = C_1\Big(\frac{1+\sqrt{5}}{2}\Big)^n + C_2\Big(\frac{1-\sqrt{5}}{2}\Big)^n$$

for some $C_1, C_2 \in \mathbf{C}$. Use $a_0 = a_1 = 1$ to show that

$$C_1 = \frac{1}{2} + \frac{1}{2\sqrt{5}} \quad \text{and} \quad C_2 = \frac{1}{2} - \frac{1}{2\sqrt{5}}.$$

In §2 we used the recurrence $d_n = (n-1)(d_{n-1} + d_{n-2})$ for the derangement numbers to prove Theorem 2.5 by induction on n. This required us to already know the formula. Generating functions give a more systematic approach. (You are asked to fill in the details in this proof in Question 2 on Sheet 4.)

Theorem 8.4. *Let $p_n = d_n/n!$ be the probability that a permutation of $\{1, 2, \ldots, n\}$, chosen uniformly at random, is a derangement. Then*

$$np_n = (n-1)p_{n-1} + p_{n-2}$$

for all $n \geq 2$ and

$$p_n = 1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}.$$

The steps needed in this proof can readily be performed using computer algebra packages. Indeed, MATHEMATICA implements a more refined version of our three step programme for solving recurrences in its `RSolve` command. (See the discussion in Appendix A of Wilf *generatingfunctionology*.)

Often it is possible to get some information about the asymptotic growth of a sequence from its generating function. We shall need a standard piece of notation.

Definition 8.5. Given a sequence a_0, a_1, a_2, … of real numbers and a function $t : \mathbf{R} \to \mathbf{R}$, we write $a_n = \mathrm{O}\big(t(n)\big)$ if there exists a constant $B \in \mathbf{R}$ such that $|a_n| < Bt(n)$ for all $n \in \mathbf{N}_0$.

Theorem 8.6. *Let $F(x) = \sum_{n=0}^{\infty} a_n x^n$ be the generating function for the sequence $a_0, a_1, a_2, \ldots$. Suppose that $F(x) = f(x)/g(x)$ where $f(x)$ and $g(x)$ are polynomials and $\deg f < \deg g$. If β is the root of $g(x)$ of minimum modulus then*

$$a_n = \mathrm{O}\Big(\big(\frac{1}{|\beta|} + \varepsilon\big)^n\Big)$$

for all $\varepsilon > 0$.

More generally if F has no singularities in the complex plane with modulus $< |\beta|$ then $F(z)$ converges for all $z \in \mathbf{C}$ such that $|z| < |\beta|$, and the conclusion of Theorem 8.6 still holds. See §2.4 in Wilf *generatingfunctionology* for a proof.

9. Convolutions and the Catalan numbers

Definition 9.1. The convolution of the sequences $a_0,\ a_1,\ a_2,\ \ldots$ and $b_0,\ b_1,\ b_2,\ \ldots$ is the sequence $c_0,\ c_1,\ c_2, \ldots$ defined by $c_n = \sum_{m=0}^{n} a_m b_{n-m}$.

Convolutions are closely related to generating functions. The next lemma states that convolution of sequences correspond to products of power series. It can be proved in one line using the definition of the product of formal power series (see bottom of page 15).

Lemma 9.2. *Let $a_0, a_1, a_2, \ldots$ and $b_0, b_1, b_2, \ldots$ be sequences and let $c_0, c_1, c_2, \ldots$ be their convolution. Let $F(x) = \sum_{n=0}^{\infty} a_n x^n$, $G(x) = \sum_{n=0}^{\infty} b_n x^n$ and $H(x) = \sum_{n=0}^{\infty} c_n x^n$. Then*

$$F(x)G(x) = H(x).$$

Typically convolutions arise in combinatorial problems when we have a combinatorial object that can be decomposed into two different subobjects of the same total size in different ways.

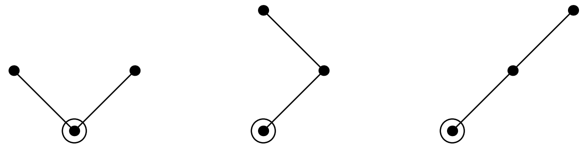
For example, in Lemma 6.6, we decomposed a placement of k rooks on the board B into placements of j rooks on D and $k-j$ rooks on the board D.

Example 9.3. A resident of Flatland is given an enormous number of indistinguishable 1×1 unit squares for his birthday. How many ways can he make a 'T' shape, using at least one brick for the vertical section and at least two bricks for the horizontal section?

The canonical application of convolutions is to the Catalan numbers. These numbers have many different combinatorial interpretations; we shall define them using rooted binary trees drawn in the plane.

Definition 9.4. A rooted binary tree is either empty, or consists of a *root vertex* together with a pair of rooted binary trees: a *left subtree* and a *right subtree*. The *Catalan number C_n* is the number of rooted binary trees on n vertices.

For example, there are five rooted binary trees with three vertices, so $C_3 = 5$. Three of them are shown below, with the root vertex circled. The other two can be obtained by reflecting the two asymmetric diagrams.



Lemma 9.5. *For each $n \in \mathbf{N}_0$ we have*

$$C_{n+1} = C_0C_n + C_1C_{n-1} + \cdots + C_{n-1}C_1 + C_nC_0.$$

Theorem 9.6. *If $n \in \mathbf{N}_0$ then $C_n = \frac{1}{n+1}\binom{2n}{n}$.*

We shall prove Theorem 9.6 using our usual three-step programme. Let $F(x) = \sum_{n=0}^{\infty} C_n x^n$ be the generating function for the Catalan numbers. In outline the steps are:

(a) Use the recurrence relation in Lemma 9.5 to show that $F(x)$ satisfies the quadratic equation

$$xF(x)^2 = F(x) - 1.$$

(b) Solve the quadratic equation to get the closed form

$$xF(x) = \frac{1 - \sqrt{1-4x}}{2}.$$

(c) Use the general version of the Binomial Theorem in Theorem 7.5 to deduce the formula for C_n.

Our final application of convolutions will give yet another proof (the shortest yet!) of the formula for the derangement numbers d_n.

Lemma 9.7. *If $n \in \mathbf{N}_0$ then*

$$\sum_{k=0}^{n} \binom{n}{k} d_{n-k} = n!.$$

The sum in the lemma becomes a convolution after a small amount of rearranging.

Theorem 9.8. *If $G(x) = \sum_{n=0}^{\infty} d_m x^m/m!$ then*

$$G(x)\exp(x) = \frac{1}{1-x}.$$

It is now easy to deduce the formula for d_n; the argument needed is the same as the final step in the proof of Theorem 9.7. The generating function G used above is an example of an *exponential generating function*.

Exercise: Explain the unusual structure of the decimal expansion of

$$\tfrac{1}{2} - \sqrt{\tfrac{1}{4} - \tfrac{1}{1000}} = 0.001\,001\,002\,005\,014\,042\ldots.$$

The Catalan Numbers have a vast number of combinatorial interpretations. See Question 4 on Sheet 6 for one more. A further 64 (and counting) are given in Exercise 6.19 in Stanley *Enumerative Combinatorics 2*, CUP 2001.

10. Partitions

Definition 10.1. A *partition* of a number $n \in \mathbf{N}_0$ is a sequence of natural numbers $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ such that
(i) $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 1$.
(ii) $\lambda_1 + \lambda_2 + \cdots + \lambda_k = n$.
The entries in a partition λ are called the *parts* of λ. Let $p(n)$ be the number of partitions of n.

By this definition the unique partition of 0 is the empty partition $\varnothing$, and so $p(0) = 1$. The sequence of partition numbers begins $1, 1, 2, 3, 5, 7, 11, 15, \ldots$.

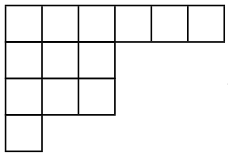
Example 10.2. Let a_n be the number of ways to pay for an item costing n pence using only 2p and 5p coins. Equivalently, a_n is the number of partitions of n into parts of size 2 and size 5. Will find the generating function for a_n.

This example suggests it is useful to encode a partition as a list of multiplicities of its parts. For example, the partition $(6, 3, 3, 2, 1, 1, 1)$ of 17 can be encoded as the list of multiplicities $(3, 1, 2, 0, 0, 1)$.

Theorem 10.3. *The generating function for $p(n)$ is*

$$\sum_{n=0}^{\infty} p(n)x^n = \frac{1}{(1-x)(1-x^2)(1-x^3)\ldots}.$$

It is often useful to represent partitions by *Young diagrams*. The Young diagram of $(\lambda_1, \ldots, \lambda_k)$ has k rows of boxes, with λ_i boxes in row i. For example, the Young diagram of $(6, 3, 3, 1)$ is



.

The next theorem has a very simple proof using Young diagrams. (See also Question 9 on Sheet 5.)

Theorem 10.4. *Let $n \in \mathbf{N}$ and let $k \leq n$. The number of partitions of n into parts of size $\leq k$ is equal to the number of partitions of n with at most k parts.*

While there are bijective proofs of the next theorem using Young diagrams, it is much easier to prove it using generating functions. Note how we adapt the proof of Theorem 10.3 to get the generating functions for two special types of partitions.

Theorem 10.5. *Let $n \in \mathbf{N}$. The number of partitions of n with at most one part of any given size is equal to the number of partitions of n into odd parts.*

For a generalization see Question 9 on Sheet 6.

There are many deep combinatorial and number-theoretic properties of the partition numbers. For example, in 1919 Ramanujan used analytic arguments with generating functions to prove that

$$p(4), p(9), p(14), p(19), \ldots, p(5m + 4), \ldots$$

are all divisible by 5. In 1944 Freeman Dyson found a bijective proof of this result while still an undergraduate. A number of deep generalizations of Ramanujan's congruences have since been proved, most recently by Mahlburg in 2005.

The problem of finding an estimate for the size of the partition number $p(n)$ was solved in 1919 by Hardy and Ramanujan as the original application of the circle method. The crudest version of their result is

$$p(n) \sim \frac{\mathrm{e}^{c\sqrt{n}}}{4n\sqrt{3}}$$

where $c = 2\sqrt{\frac{\pi^2}{6}}$, and $\sim$ means that the ratio of the two sides tends to 1 as $n \to \infty$.

We end with a much more elementary result that helps to explain the constant c in the Hardy–Ramanujan theorem. This is included for interest only, and may be considered non-examinable. It will be skipped if time is pressing.

Theorem 10.6 (Van Lint's upper bound)**.** *If* $n \in \mathbf{N}$ *then* $p(n) \leq \mathrm{e}^{c\sqrt{n}}$ *where* $c = 2\sqrt{\frac{\pi^2}{6}}$.

Outline proof. Let $P(x) = \prod_{m=1}^{\infty} 1/(1-x^m)$ be the generating function for the partition numbers found in Theorem 10.3. Taking logs we get

$$\log P(x) = -\sum_{m=1}^{\infty} \log(1 - x^m).$$

Using the power series $-\log(1-y) = \sum_{r=1}^{\infty} y^r/r$ we get

$$\log P(x) = \sum_{r=1}^{\infty} \frac{x^r}{r(1-x^r)}.$$

Substitute $x = \mathrm{e}^{-y}$ where $y > 0$ gives

$$\log P(\mathrm{e}^{-y}) \leq \sum_{r=1}^{\infty} \frac{\mathrm{e}^{-ry}}{r(1-\mathrm{e}^{-ry})}.$$

Now rewrite each summand as $1/r(\mathrm{e}^{ry} - 1)$ and use the inequality $\mathrm{e}^{ry} - 1 \geq ry$ to get

$$\log P(\mathrm{e}^{-y}) \leq \sum_{r=1}^{\infty} \frac{1}{r^2 y}.$$

Since $\sum_{r=1}^{\infty} 1/r^2 = \pi^2/6$, we have $\log P(\mathrm{e}^{-y}) \leq \pi^2/6y$. The result now follows by making a strategic choice of y: see Question 10 on Sheet 6 for the remaining steps. $\square$

Part C: Ramsey Theory

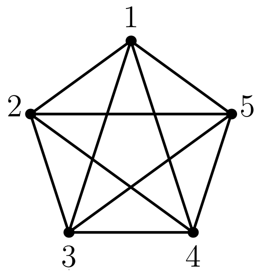
11. Introduction to Ramsey Theory

A typical result in Ramsey Theory says that any sufficiently large combinatorial structure always contain a substructure with some regular pattern. For example, any infinite sequence of real numbers contains either an increasing or a decreasing subsequence (the Bolzano–Weierstrass theorem). The finite version of this result will appear on Problem Sheet 7.

Most of the results in Ramsey Theory are naturally stated in terms of graphs. We concentrate on the finite case.

Definition 11.1. A *graph* consists of a set V of vertices together with a set E of 2-subsets of V called *edges*. The *complete graph* with vertex set V is the graph whose edge set is all 2-subsets of V.

For example, the complete graph on $V = \{1, 2, 3, 4, 5\}$ is drawn below. Its edge set is $\big\{\{1, 2\}, \{1, 3\}, \ldots, \{4, 5\}\big\}$.



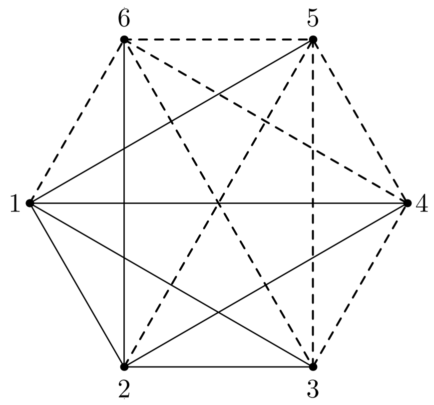
We denote the complete graph on $\{1, 2, \ldots, n\}$ by K_n.

Definition 11.2. Let $c, n \in \mathbf{N}$. A *c-colouring* of the complete graph K_n is a function from the edge set of K_n to $\{1, 2, \ldots, c\}$. If S is an s-subset of the vertices of K_n such that all the edges between vertices in S have the same colour, then we say that S is a *monochromatic* K_s

A monochromatic K_3 is usually said to be a monochromatic triangle. Note that it is the edges of the complete graph K_n that are coloured, *not the vertices.*

In practice we shall specify graphs and colours rather less formally. It seems to be a standard convention that colour 1 is red and colour 2 is blue. In these notes, red will be indicated by solid lines and blue by dashed lines.

Exercise: Show that in the colouring of K_6 below there is a unique blue K_4 and exactly two red triangles. Find all the blue triangles.



Example 11.3. In any red-blue colouring of the edges of K_6 there is either a red triangle or a blue triangle.

Definition 11.4. Given s, $t \in \mathbf{N}$, with $s, t \geq 2$, we define the Ramsey number $R(s,t)$ to be the smallest n (if one exists) such that in any red-blue colouring of the complete graph on n vertices, there is either a red K_s or a blue K_t.

For example, we know from Example 11.3 that $R(3,3) \leq 6$.

Exercise: Show that if $N \geq R(s,t)$ then in any red-blue colouring of K_N there is either a red K_s or a blue K_t.

By Question 2 on Sheet 6 there is a red-blue colouring of K_5 with no monochromatic triangle. Hence, by the exercise $R(3,3) > 5$. It now follows from Example 11.3 that $R(3,3) = 6$.

We will prove in Theorem 12.3 that all the two-colour Ramsey numbers $R(s,t)$ exist, and that $R(s,t) \leq \binom{s+t-2}{s-1}$. (But please do not assume this result when doing Sheet 6.) One family of Ramsey numbers is easily found.

Lemma 11.5. *For any $s \in \mathbf{N}$ we have $R(2,s) = s$.*

The main idea need to prove Theorem 12.3 appears in the next example.

Example 11.6. In any two-colouring of K_{10} there is either a red K_3 or a blue K_4. Hence $R(3,4) \leq 10$.

This bound can be improved; to do this we shall need a result from graph theory. Recall that if v is a vertex of a graph G then the *degree* of v is the number of edges of G that meet v.

Lemma 11.7 (Hand-Shaking Lemma)**.** *Let G be a graph with vertex set $\{1, 2, ..., n\}$ and exactly e edges. If d_i is the degree of vertex i then*

$$2e = d_1 + d_2 + \cdots + d_n.$$

In particular, the number of vertices of odd degree is even.

Theorem 11.8. $R(3, 4) = 9$.

The proof of the final theorem is left to you: see Question 3 on Sheet 6.

Theorem 11.9. $R(4, 4) \leq 18$.

There is a red-blue colouring of K_{17} with no red K_4 or blue K_4 so $R(4, 4) = 18$. The construction will appear on Sheet 7.

It is a very hard problem to find the exact values of Ramsey numbers for larger s and t. For a survey of other known results on $R(s, t)$ for small s and t, see Stanislaw Radziszowski, *Small Ramsey Numbers*, Electronic Journal of Combinatorics, available at `www.combinatorics.org/Surveys`. For example, it was shown in 1965 that $R(4, 5) = 25$, but all that is known about $R(5, 5)$ is that it lies between 43 and 49. It is probable that no-one will ever know the exact value of $R(6, 6)$.

12. Ramsey's Theorem

Since finding the Ramsey numbers $R(s, t)$ exactly is so difficult, we settle for proving that they exist, by proving an upper bound for $R(s, t)$. We work by induction on $s + t$. The following lemma gives the critical inductive step.

Lemma 12.1. *Let $s, t \in \mathbf{N}$ with $s, t \geq 3$. If $R(s - 1, t)$ and $R(s, t - 1)$ exist then $R(s, t)$ exists and*

$$R(s, t) \leq R(s - 1, t) + R(s, t - 1).$$

Theorem 12.2. *For any $s, t \in \mathbf{N}$ with $s, t \geq 2$, the Ramsey number $R(s, t)$ exists and*

$$R(s, t) \leq \binom{s + t - 2}{s - 1}.$$

We now get a bound on the diagonal Ramsey numbers $R(s, s)$. Note that because of the use of induction on $s+t$, we could not have obtained this result without first bounding all the Ramsey numbers $R(s, t)$.

Corollary 12.3. *If $s \in \mathbf{N}$ and $s \geq 2$ then*
$$R(s,s) \leq \binom{2s-2}{s-1} \leq 4^{s-1}.$$

One version of Stirling's Formula states that if $m \in \mathbf{N}$ then
$$\sqrt{2\pi m}\left(\frac{m}{\mathrm{e}}\right)^m \leq m! \leq \sqrt{2\pi m}\left(\frac{m}{\mathrm{e}}\right)^m \mathrm{e}^{1/12m}.$$
These bounds lead to the asymptotically stronger result that
$$R(s,s) \leq \frac{4^s}{\sqrt{s}} \quad \text{for all } s \in \mathbf{N}.$$

Corollary 12.3 was proved by Erdös and Szekeres in 1935. We have followed their proof above. The strongest improvement known to date is due to David Conlon, who showed in 2004 that, up to a rather technical error term, $R(s,s) \leq 4^s/s$. In 1947 Erdös proved the lower bound $R(s,s) \geq 2^{(s-1)/2}$. His argument becomes clearest when stated in probabilistic language: we will see it in Part D of the course.

To end this introduction to Ramsey Theory we give two results related to Theorem 12.2.

Pigeonhole Principle. The Pigeonhole Principle states that if n pigeons are put into $n-1$ holes, then some hole must contain two or more pigeons. See Question 8 on Sheet 6 for some applications of the Pigeonhole Principle.

In Examples 11.3 and 11.6, and Lemma 12.1, we used a similar result: if $r+s-1$ objects (in these cases, edges) are coloured red and blue, then either there are r red objects, or s blue objects. This is probably the simplest result that has some of the general flavour of Ramsey theory.

Multiple colours. It is possible to generalize all the results proved so far to three or more colours.

Theorem 12.4. *There exists $n \in \mathbf{N}$ such that if the edges of K_n are coloured red, blue and yellow then there exists a monochromatic triangle.*

There are (at least) two ways to prove Theorem 12.4. The first adapts our usual argument, looking at the edges coming out of vertex 1 and concentrating on those vertices joined by edges of the majority colour. The second uses a neat trick to reduce to the two-colour case.

Part D: Probabilistic Methods

13. Revision of Discrete Probability

This section is intended to remind you of the definitions and language of discrete probability theory, on the assumption that you have seen most of the ideas before. These notes are based on earlier notes by Dr Barnea and Dr Gerke; of course any errors are my responsibility.

For further background see any basic textbook on probability, for example Sheldon Ross, *A First Course in Probability*, Prentice Hall 2001.

Definition 13.1.

- A *probability measure* p on a finite set Ω assigns a real number p_ω to each $\omega \in \Omega$ so that $0 \leq p_\omega \leq 1$ for each ω and
$$\sum_{\omega \in \Omega} p_\omega = 1.$$
We say that p_ω is the *probability of* ω.
- A *probability space* is a finite set Ω equipped with a probability measure. The elements of a probability space are sometimes called *outcomes*.
- An *event* is a subset of Ω.
- The *probability* of an event $A \subseteq \Omega$, denoted $\mathbf{P}[A]$ is the sum of the probability of the outcomes in A; that is
$$\mathbf{P}[A] = \sum_{\omega \in A} p_\omega.$$

It follows at once from this definition that $\mathbf{P}[\{\omega\}] = p_\omega$ for each $\omega \in \Omega$. We also have $\mathbf{P}[\varnothing] = 0$ and $\mathbf{P}[\Omega] = 1$.

Example 13.2

(1) To model a throw of a single unbiased die, we take
$$\Omega = \{1, 2, 3, 4, 5, 6\}$$
and put $p_\omega = 1/6$ for each outcome $\omega \in \Omega$. The event that we throw an even number is $A = \{2, 4, 6\}$ and as expected, $\mathbf{P}[A] = p_2 + p_4 + p_6 = 1/6 + 1/6 + 1/6 = 1/2$.

(2) To model a throw of a pair of dice we could take
$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$
and give each element of Ω probability $1/36$, so $p_{(i,j)} = 1/36$ for all $(i, j) \in \Omega$. Alternatively, if we know we only care about the sum of the two dice, we could take $\Omega = \{2, 3, \ldots, 12\}$ with $p_2 = 1/36$, $p_3 = 2/36$, $\ldots$, $p_6 = 5/36$, $p_7 = 6/36$, $p_8 = 5/36$, $\ldots$, $p_{12} = 1/36$. The former is natural and more flexible.

(3) A suitable probability space for three flips of a coin is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

where H stands for heads and T for tails, and each outcome has probability $1/8$. To allow for a biased coin we fix $0 \le q \le 1$ and instead give an outcome with exactly k heads probability $q^k(1-q)^{3-k}$.

(4) Let $n \in \mathbf{N}$ and let Ω be the set of all permutations of $\{1, 2, \ldots, n\}$. Set $p_\sigma = 1/n!$ for each permutation $\sigma \in \Omega$. This gives a suitable setup for Theorem 2.6. Later we shall use the language of probability theory to give a shorter proof of part (ii) of this theorem.

It will often be helpful to specify events (i.e. subsets of Ω) a little informally. For example, in (3) above we might write $\mathbf{P}$[at least two heads], rather than $\mathbf{P}[\{HHT, HTH, THH, HHH\}]$.

Unions, intersections and complements. Let Ω be a probability space. If A, $B \subseteq \Omega$ then

$$\begin{aligned}\mathbf{P}[A \cup B] = \sum_{\omega \in A \cup B} p_\omega = \sum_{\omega \in A} p_\omega + \sum_{\omega \in B} p_\omega - \sum_{\omega \in A \cap B} p_\omega \\ = \mathbf{P}[A] + \mathbf{P}[B] - \mathbf{P}[A \cap B].\end{aligned}$$

In particular, if A and B are disjoint, i.e. $A \cap B = \varnothing$, then $P[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B]$. The *complement* of an event $A \subseteq \Omega$ is defined to be

$$\bar{A} = \{\omega \in \Omega : \omega \notin A\}.$$

Since

$$1 = \mathbf{P}[\Omega] = \mathbf{P}[A \cup \bar{A}] = \mathbf{P}[A] + \mathbf{P}[\bar{A}]$$

we have $\mathbf{P}[\bar{A}] = 1 - \mathbf{P}[A]$.

Exercise: Show that if $A_1, \ldots, A_n \subseteq \Omega$ then

$$\mathbf{P}[A_1 \cup \cdots \cup A_n] \le \mathbf{P}[A_1] + \cdots + \mathbf{P}[A_n].$$

Condition probability and independence.

Definition 13.3. Let Ω be a probability space, and let A, $B \subseteq \Omega$ be events.

- If $\mathbf{P}[B] \neq 0$ then we define the *conditional probability of A given B* by
$$\mathbf{P}[A|B] = \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[B]}.$$
- The events A, B are said to be *independent* if $\mathbf{P}[A \cap B] = \mathbf{P}[A]\mathbf{P}[B]$.

Suppose that each element of Ω has equal probability p. Then

$$\mathbf{P}[A|B] = \frac{|A \cap B|p}{|B|p} = \frac{|A \cap B|}{|B|}$$

is the proportion of elements of B that also lie in A; informally, if we know that the event B has occurred, then the probability that A has also occurred is $\mathbf{P}[A|B]$.

Exercise: Show that if A and B are events in a probability space such that $\mathbf{P}[A]$, $\mathbf{P}[B] \neq 0$, then $\mathbf{P}[A|B] = \mathbf{P}[A]$ if and only if A and B are independent.

Conditional probability can be quite subtle.

Exercise: Let $\Omega = \{HH, HT, TH, TT\}$ be the probability space for two flips of a fair coin, so each outcome has probability $\frac{1}{4}$. Let A be the event that both flips are heads, and let B be the event that at least one flip is a head. Write A and B as subsets of Ω and show that $\mathbf{P}[A|B] = 1/3$.

Example 13.4 (The Monty Hall Problem)**.** On a game show you are offered the choice of three doors. Behind one door is a car, and behind the other two are goats. You pick a door and then the host, *who knows where the car is*, opens another door to reveal a goat. You may then either open your original door, or change to the remaining unopened door. Assuming you want the car, should you change?

Most people find the answer to the Monty Hall problem a little surprising. The Sleeping Beauty Problem, stated below, is even more controversial.

Example 13.5. Beauty is told that if a coin lands heads she will be woken on Monday and Tuesday mornings, but after being woken on Monday she will be given an amnesia inducing drug, so that she will have no memory of what happened that day. If the coin lands tails she will only be woken on Tuesday morning. At no point in the experiment will Beauty be told what day it is. Imagine that you are Beauty and are awoken as part of the experiment and asked for your credence that the coin landed heads. What is your answer?

The related statistical issue in the next example is also widely misunderstood.

Example 13.6. Suppose that one in every 1000 people has disease X. There is a new test for X that will always identify the disease in anyone who has it. There is, unfortunately, a tiny probability of 1/250 that the test will falsely report that a healthy person has the disease. What is the probability that a person who tests positive for X actually has the disease?

Random variables.

Definition 13.7. Let Ω be a probability space. A *random variable* on Ω is a function $X : \Omega \to \mathbf{R}$.

Definition 13.8. If X, $Y : \Omega \to \mathbf{R}$ are random variables then we say that X and Y are *independent* if for all x, $y \in \mathbf{R}$ the events

$$\begin{aligned} A &= \{\omega \in \Omega : X(\omega) = x\} \quad \text{and} \\ B &= \{\omega \in \Omega : Y(\omega) = y\} \end{aligned}$$

are independent.

The following shorthand notation is very useful. If $X : \Omega \to \mathbf{R}$ is a random variable, then '$X = x$' is the event $\{\omega \in \Omega : X(\omega) = x\}$. Similarly '$X \geq x$' is the event $\{\omega \in \Omega : X(\omega) \geq x\}$. We mainly use this shorthand in probabilities, so for instance

$$\mathbf{P}[X = x] = \mathbf{P}\big[\{\omega \in \Omega : X(\omega) = x\}\big].$$

Exercise: Show that X, $Y : \Omega \to \mathbf{R}$ are independent if and only if

$$\mathbf{P}[(X = x) \cap (Y = y)] = \mathbf{P}[X = x]\mathbf{P}[Y = y]$$

for all $x, y \in \mathbf{R}$. (This is just a trivial restatement of the definition.)

Example 13.9. Let $\Omega = \{HH, HT, TH, TT\}$ be the probability space for two flips of a fair coin. Define $X : \Omega \to \mathbf{R}$ to be 1 if the first coin is heads, and zero otherwise. So

$$X(HH) = X(HT) = 1 \quad \text{and} \quad X(TH) = X(TT) = 0.$$

Define $Y : \Omega \to \mathbf{R}$ similarly for the second coin.

(i) The random variables X and Y are independent.
(ii) Let Z be 1 if exactly one flip is heads, and zero otherwise. Then X and Z are independent, and Y and Z are independent.
(iii) There exist $x, y, z \in \{0, 1\}$ such that
$$\mathbf{P}[X = x, Y = y, Z = z] \neq \mathbf{P}[X = x]\mathbf{P}[Y = y]\mathbf{P}[Z = z].$$

This shows that one has to be quite careful when defining independence for a family of random variables. (Except in the Lovász Local Lemma, we will be able to manage with the pairwise independence defined above.)

Given random variables X, $Y : \Omega \to \mathbf{R}$ we can define new random variables by taking functions such as $X + Y$, aX for $a \in \mathbf{R}$ and XY. For instance $(X + Y)(\omega) = X(\omega) + Y(\omega)$, and so on. Notice that if $z \in \mathbf{R}$ then

$$\{\omega \in \Omega : (X + Y)(\omega) = z\} = \bigcup_{x+y=z} \{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}.$$

The events above are disjoint for different x, y, so we get

$$\mathbf{P}[X+Y=z] = \sum_{x+y=z} \mathbf{P}[(X=x)\cap(Y=y)].$$

If X and Y are independent then

$$\mathbf{P}[(X=x)\cap(Y=y)] = \mathbf{P}[X=x]\mathbf{P}[Y=y]$$

and so

$$\mathbf{P}[X+Y=z] = \sum_{x+y=z} \mathbf{P}[X=x]\mathbf{P}[Y=y].$$

(Note that all of these sums have only finitely many non-zero summands, so they are well-defined.)

Exercise: Show similarly that if X, $Y : \Omega \to \mathbf{R}$ are independent random variables then

$$\mathbf{P}[XY=z] = \sum_{xy=z} \mathbf{P}[X=x]\mathbf{P}[Y=y].$$

Expectation and linearity.

Definition 13.10. Let Ω be a probability space with probability measure p. The *expectation* $\mathbf{E}[X]$ of a random variable $X : \Omega \to \mathbf{R}$ is defined to be

$$\mathbf{E}[X] = \sum_{\omega\in\Omega} X(\omega)p_w.$$

Intuitively, the expectation of X is the average value of X on elements of Ω, if we choose $\omega \in \Omega$ with probability p_ω. We have

$$\mathbf{E}[X] = \sum_{\omega\in\Omega} X(\omega)p_\omega = \sum_{x\in R} \sum_{\substack{\omega \\ X(\omega)=x}} xp_\omega = \sum_{x\in\mathbf{R}} x\mathbf{P}[X=x].$$

A critical property of expectation is that it is linear. Note that we *do not need to assume independence* in this lemma.

Lemma 13.11. *Let Ω be a probability space. If $X_1, X_2, \ldots, X_k : \Omega \to \mathbf{R}$ are random variables then*

$$\mathbf{E}[a_1X_1 + a_2X_2 + \cdots + a_kX_k] = a_1\mathbf{E}[X_1] + a_2\mathbf{E}[X_2] + \cdots + a_k\mathbf{E}[X_k]$$

for any $a_1, a_2, \ldots, a_k \in \mathbf{R}$.

Proof. By definition the left-hand side is

$$\begin{aligned}\sum_{\omega\in\Omega} p_\omega\big(a_1X_1 + \cdots + a_kX_k\big)(\omega) &= \sum_{\omega\in\Omega} p_\omega\big(a_1X_1(\omega) + \cdots + a_kX_K(\omega)\big) \\ &= a_1\sum_{\omega\in\Omega} p_\omega X_1(\omega) + \cdots + a_k\sum_{\omega\in\Omega} X_k(\omega)\end{aligned}$$

which is the right-hand side. $\square$

When X, $Y : \Omega \to \mathbf{R}$ are independent random variables, there is a very useful formula for $\mathbf{E}[XY]$.

Lemma 13.12. *If X, $Y : \Omega \to \mathbf{R}$ are independent random variables then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.*

Exercise: Prove Lemma 13.11 by arguing that

$$\mathbf{E}[XY] = \sum_{z \in \mathbf{R}} z\mathbf{P}[XY = z] = \sum_{z \in R} z \sum_{xy=z} \mathbf{P}[(X = x) \cap (Y = y)]$$

and using independence.

VARIANCE.

Definition 13.13. Let Ω be a probability space. The *variance* $\mathbf{Var}[X]$ of a random variable $X : \Omega \to \mathbf{R}$ is defined to be

$$\mathbf{Var}[X] = \mathbf{E}\big[(X - \mathbf{E}[X])^2\big].$$

The variance measures how much X can be expected to depart from its mean value $\mathbf{E}[X]$. So it is a measure of the 'spread' of X.

It is tempting to define the variance as $\mathbf{E}\big[X - \mathbf{E}[X]\big]$, but by linearity this expectation is $\mathbf{E}[X] - \mathbf{E}[X] = 0$. One might also consider the quantity $\mathbf{E}\big[\,\big|X - \mathbf{E}[X]\big|\,\big]$, but the absolute value turns out to be hard to work with. The definition above works well in practice.

Lemma 13.14. *Let Ω be a probability space.*

(i) *If $X : \Omega \to \mathbf{R}$ is a random variable then*

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

(ii) *If X, $Y : \Omega \to \mathbf{R}$ are independent random variables then*

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y].$$

Exercise: Show that (ii) can fail if X and Y are not independent. [*Hint: usually a random variable is not independent of itself.*]

14. Introduction to probabilistic methods

In this section we shall solve some problems involving permutations (including, yet again, the derangements problem) using probabilistic arguments. We shall use the language of probability spaces and random variables recalled in §13. It will be particularly important for you to ask questions if the use of anything from this section is unclear.

Throughout this section we fix $n \in \mathbf{N}$ and let Ω be the set of all permutations of the set $\{1, 2, \ldots, n\}$. We define a probability measure $q : \Omega \to \mathbf{R}$ by $q_\sigma = 1/n!$ for each permutation σ of $\{1, 2, \ldots, n\}$. This makes Ω into a probability space in which all the permutations have equal probability. We say that the permutations are *chosen uniformly at random.*

Recall that, in probabilistic language, *events* are subsets of Ω.

Exercise: Let $x \in \{1, 2, \ldots, n\}$ and let $A_x = \{\sigma \in \Omega : \sigma(x) = x\}$. Then A_x is the event that a permutation fixes x. What is the probability of A_x?

Building on this we can give a better proof of Theorem 2.6(ii).

Theorem 14.1. *Let $F : \Omega \to \mathbf{N}_0$ be defined so that $F(\sigma)$ is the number of fixed points of the permutation $\sigma \in \Omega$. Then $\mathbf{E}[F] = 1$.*

To give a more general result we need cycles and the cycle decomposition of a permutation.

Definition 14.2. A permutation σ of $\{1, 2, \ldots, n\}$ *acts as a k-cycle* on a k-subset $S \subseteq \{1, 2, \ldots, n\}$ if S has distinct elements $x_1, x_2, \ldots, x_k$ such that

$$\sigma(x_1) = x_2,\ \sigma(x_2) = x_3,\ \ldots,\ \sigma(x_k) = x_1.$$

If $\sigma(y) = y$ for all $y \in \{1, 2, \ldots, n\}$ such that $y \notin S$ then we say that σ is a *k-cycle*, and write

$$\sigma = (x_1, x_2, \ldots, x_k).$$

Note that there are k different ways to write a k-cycle. For example, the 3-cycle $(1, 2, 3)$ can also be written as $(2, 3, 1)$ and $(3, 1, 2)$.

Definition 14.3. We say that cycles $(x_1, x_2, \ldots, x_k)$ and $(y_1, y_2, \ldots, y_\ell)$ are *disjoint* if

$$\{x_1, x_2, \ldots, x_k\} \cap \{y_1, y_2, \ldots, y_\ell\} = \varnothing.$$

Lemma 14.4. *A permutation σ of $\{1, 2, \ldots, n\}$ can be written as a composition of disjoint cycles. The cycles in this composition are uniquely determined by σ.*

The proof of Lemma 14.4 is non-examinable and will not be given in full in lectures. What is more important is that you can apply the result. We shall use it below in Theorem 14.5

Exercise: Write the permutation of $\{1,2,3,4,5,6\}$ defined by $\sigma(1) = 3$, $\sigma(2) = 4$, $\sigma(3) = 1$, $\sigma(4) = 6$, $\sigma(5) = 5$, $\sigma(6) = 2$ as a composition of disjoint cycles.

Given a permutation σ of $\{1, 2, \ldots, n\}$ and $k \in \mathbf{N}$, we can ask: what is the probability that a given $x \in \{1, 2, \ldots, n\}$ lies in a k-cycle of σ? The first exercise in this section shows that the probability that x lies in a 1-cycle is $1/n$.

Exercise: Check directly that the probability that 1 lies in a 2-cycle of a permutation of $\{1, 2, 3, 4\}$ selected uniformly at random is $1/4$.

Theorem 14.5. *Let $1 \leq k \leq n$ and let $x \in \{1, 2, \ldots, n\}$. The probability that x lies in a k-cycle of a permutation of $\{1, 2, \ldots, n\}$ chosen uniformly at random is $1/n$.*

Theorem 14.6. *Let p_n be the probability that a permutation of $\{1, 2, \ldots, n\}$ chosen uniformly at random is a derangement. Then*

$$p_n = \frac{p_{n-2}}{n} + \frac{p_{n-3}}{n} + \cdots + \frac{p_1}{n} + \frac{p_0}{n}.$$

It may be helpful to compare this result with Lemma 9.7: there we get a recurrence by considering fixed points; here we get a recurrence by considering cycles.

We now use generating functions to recover the usual formula for p_n.

Corollary 14.7. *For all $n \in \mathbf{N}$,*

$$p_n = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!}.$$

We can also generalize Theorem 14.1.

Theorem 14.8. *Let $C_k : \Omega \to \mathbf{R}$ be the random variable defined so that $C_k(\sigma)$ is the number of k-cycles in the permutation σ of $\{1, 2, \ldots, n\}$. Then $\mathbf{E}[C_k] = 1/k$ for all k such that $1 \leq k \leq n$.*

Note that if $k > n/2$ then a permutation can have at most one k-cycle. So in these cases, $\mathbf{E}[C_k]$ is the probability that a permutation of $\{1, 2, \ldots, n\}$, chosen uniformly at random, has a k-cycle.

15. Ramsey Numbers and the First Moment Method

The grandly named 'First Moment Method' is nothing more than the following simple observation.

Lemma 15.1 (First Moment Method)**.** *Let Ω be a probability space and let $M : \Omega \to \mathbf{N}_0$ be a random variable taking values in $\mathbf{N}_0$. If $\mathbf{E}[M] = x$ then*

(i) $\mathbf{P}[M \geq x] > 0$, *so there exists $\omega \in \Omega$ such that $M(\omega) \geq x$.*
(ii) $\mathbf{P}[M \leq x] > 0$, *so there exists $\omega' \in \Omega$ such that $M(\omega') \leq x$.*

Exercise: Check that the lemma holds in the case when

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$

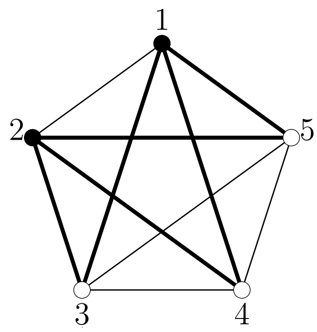
models the throw of two fair dice (see Example 13.2(2)) and if $(\alpha, \beta) \in \Omega$ then $M(\alpha, \beta) = \alpha + \beta$.

The kth *moment* of a random variable X is defined to be $\mathbf{E}[X^k]$. Sometimes stronger results can be obtained by considering higher moments. We shall concentrate on first moments, where the power of the method is closely related to the linearity property of expectation (see Lemma 13.11).

Our applications will come from graph theory.

Definition 15.2. Let G be a graph with vertex set V. A *cut* (S, T) of G is a partition of V into subsets A and B. The *capacity* of a cut (S, T) is the number of edges of G that meet both S and T.

Note that $T = V \backslash S$ and $S = V \backslash T$, so a cut can be specified by giving either of the sets making up the partition. The diagram below shows the cut in the complete graph on $\{1, 2, 3, 4, 5\}$ where $S = \{1, 2, 3\}$ and $T = \{4, 5\}$. The capacity of the cut is 6, corresponding to the 6 edges $\{x, y\}$ with $x \in S$ and $y \in T$ shown with thicker lines.



Theorem 15.3. *Let G be a graph with vertex set $\{1, 2, \ldots, n\}$ and exactly m edges. There is a cut of G with capacity $\geq m/2$.*

In 1947 Erdös proved a lower bound on the Ramsey Numbers $R(s,s)$ that is still almost the best known result in this direction. Our version of his proof will use the First Moment Method in the following probability space.

Lemma 15.4. *Let $n \in \mathbf{N}$ and let Ω be the set of all red-blue colourings of the complete graph K_n. Let $p_\omega = 1/|\Omega|$ for each $\omega \in \Omega$. Then*

(i) *each colouring in Ω has probability $1/2^{\binom{n}{2}}$;*
(ii) *given any m edges in G, the probability that all m of these edges have the same colour is 2^{1-m}.*

Theorem 15.5. *Let $n, s \in \mathbf{N}$. If*

$$\binom{n}{s} 2^{1-\binom{s}{2}} < 1$$

then there is a red-blue colouring of the complete graph on $\{1, 2, \ldots, n\}$ with no red K_s or blue K_s.

Corollary 15.6. *For any $s \in \mathbf{N}$ we have*

$$R(s,s) \geq 2^{(s-1)/2}.$$

For example, since

$$\binom{42}{8} 2^{1-\binom{8}{2}} \approx 0.879 < 1,$$

if we repeatedly colour the complete graph on $\{1, 2, ..., 42\}$ at random, then we will fairly soon get a colouring with no monochromatic K_8. However, to check that we have found such a colouring, we will have to look at all $\binom{42}{8} \approx 1.18 \times 10^8$ subsets of $\{1, 2, \ldots, 42\}$. Thus Theorem 15.5 does not give an effective construction.

It is a major open problem to find, for each $s \geq 2$, an explicit colouring of the complete graph on 1.01^s vertices with no monochromatic K_s. (Here 1.01 could be replaced with $1 + \varepsilon$ for any $\varepsilon > 0$.

The bound in Corollary 15.6 can be slightly improved by the Lovász Local Lemma: see the final section.

16. Lovász Local Lemma

The section is non-examinable, and is included for interest only.

In the proof of Theorem 15.5, we considered a random colouring of the complete graph on $\{1, 2, \ldots, n\}$ and used Lemma 15.1 to show that, provided $\binom{n}{s} 2^{1-\binom{s}{2}} < 1$ there was a positive probability that this colouring had no monochromatic K_s. As motivation for the Lovász Local Lemma, consider the following alternative argument, which avoids the use of Lemma 15.1.

Alternative proof of Theorem 15.5. As before, let Ω be the probability space of all colourings of the complete graph on $\{1, 2, \ldots, n\}$, where each colouring gets the same probability. For each s-subset

$$S \subseteq \{1, 2, \ldots, n\},$$

let E_S be the event that S is a monochromatic K_s. The event that no K_s is monochromatic is then $\bigcap_S \bar{E}_S$, where the intersection is taken over all s-subsets $S \subseteq \{1, 2, \ldots, n\}$ and $\bar{E}_S = \Omega \backslash E_S$. So it will suffice to show that $\mathbf{P}[\bigcap \bar{E}_S] > 0$, or equivalently, that $\mathbf{P}[\bigcup E_S] < 1$.

In lectures we used Lemma 15.4 to show that if S is any s-subset of $\{1, 2, \ldots, n\}$ then

$$\mathbf{P}[E_S] = 2^{1-\binom{n}{s}}.$$

By the exercise on page 30, the probability of a union of events is at most the sum of their probabilities, so

$$\mathbf{P}[\bigcup_S E_S] \leq \binom{n}{s} 2^{1-\binom{n}{s}}.$$

Hence the hypothesis implies that $\mathbf{P}\left[\bigcup_S E_S\right] < 1$, as required. $\square$

If the events E_S were independent, we would have

$$\mathbf{P}\left[\bigcap_S \bar{E}_S\right] = \prod_S \mathbf{P}[\bar{E}_S].$$

Since each event E_S has non-zero probability, it would follow that their intersection has non-zero probability, giving another way to finish the proof. However, the events are not independent, so this is not an admissible strategy. The Lovász Local Lemma gives a way to get around this obstacle.

We shall need the following definition.

Definition 16.1. An event E is *mutually independent* of a collection $\mathcal{A}$ of events, if for all $U \subseteq \mathcal{A}$ and $U' \subseteq \mathcal{A} \backslash U$ we have

$$\mathbf{P}\left[E \,\middle|\, \Big(\bigcap_{C \in U} C\Big) \cap \Big(\bigcap_{D \in U'} \bar{D}\Big)\right] = \mathbf{P}[E]$$

whenever $\left(\bigcap_{C \in U} C\right) \cap \left(\bigcap_{D \in U'} \bar{D}\right)$ is non-empty.

For example, if the events E_S are as defined above, then E_S is independent of the events $\{E_T : |S \cap T| \leq 1\}$. This can be checked quite easily: informally the reason is that since each $S \cap T$ has at most one vertex, no edge is common to both S and T, and so knowing whether or not T is monochromatic gives no information about S.

Lemma 16.2 (Symmetric Lovász Local Lemma). *Let $d \in \mathbf{N}$. Let $\mathcal{A}$ be a collection of events such that $\mathbf{P}[E] \le p$ for all $E \in \mathcal{A}$. Suppose that for each event $E \in \mathcal{A}$, there is a subset $\mathcal{A}_E$ of $\mathcal{A}$ such that*
(i) $|\mathcal{A}_E| \ge |\mathcal{A}| - d$;
(ii) *E is independent of $\mathcal{A}_E$.*
If $\mathrm{e}p(d+1) \le 1$ *then*

$$\mathbf{P}\Big[\bigcap_{E \in \mathcal{A}} \overline{E}\Big] > 0$$

For a proof of the lemma, see Chapter 5 of Noga Alon and Joel H. Spencer *The Probabilistic Method*, 3rd edition. A simpler proof of a very similar result, where $\mathrm{e}p(d+1)$ is replaced with $4pd$, is given in §6.7 of Michael Mitzenmacher and Eli Upfal Probability and Computing (see [6] in the list of page 2).

The Lovász Local Lemma can be used to prove a slightly stronger version of Theorem 15.5.

Theorem 16.3. *Let $n, s \in \mathbf{N}$. If*

$$\mathrm{e}\Big(\binom{s}{2}\binom{n-2}{s-2} + 1\Big)2^{1-\binom{s}{2}} < 1$$

then there is a red-blue colouring of the complete graph K_n with no red K_s or blue K_s.

Proof. Define the events E_S as at the start of this section. We remarked that if S is an s-subset of $\{1, 2, \ldots, n\}$ then the event E_S is independent of the events E_T for those s-subsets T such that $S \cap T \le 1$. There are at most

$$\binom{s}{2}\binom{n-2}{s-2}$$

s-subsets T such that $S \cap T \ge 2$, since we can choose two common elements in $\binom{s}{2}$ ways, and then choose any $s-2$ of the remaining $n-2$ elements of $\{1, 2, \ldots, n\}$ to complete T. (There is some over-counting here, so this is only an upper bound.)

Therefore we let $d = \binom{s}{2}\binom{n-2}{s-2}$. Since

$$\mathbf{P}[E_S] = 2^{1-\binom{n}{s}}.$$

for all S, we take $p = 2^{1-\binom{n}{s}}$. Then we can apply the Lovász Local Lemma, provided that $\mathrm{e}p(d+1) \le 1$, which is one of the hypotheses of the theorem. Hence

$$\mathbf{P}\Big[\bigcap_{S} \overline{E_S}\Big] > 0$$

and so there is a red-blue colouring with no monochromatic K_s, as required. $\square$

Theorem 16.2 is stronger than Theorem 15.5 when s is reasonably large.

Example 16.4. When $s = 15$, the largest n such that

$$\binom{n}{15} 2^{1-\binom{15}{2}} < 1$$

is $n = 792$. So Theorem 15.5 tells us that $R(15, 15) > 792$. But

$$\mathrm{e}\left(\binom{15}{2}\binom{n-2}{15-2} + 1\right) 2^{1-\binom{15}{2}} < 1$$

provided $n \leq 947$. Theorem 16.2 therefore gives the stronger result that $R(15, 15) > 947$.

A more general version of the Lovász Local Lemma can be used to get the bound

$$R(s, 3) \geq \frac{Cs^2}{(\log s)^2}$$

for some constant C. For an outline of the proof and references to further results, see Alon and Spencer, Chapter 5.