COMBINATORICS MT454

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.

These notes are based on earlier notes by Dr Yiftach Barnea and Dr Stefanie Gerke. Of course I take full responsibility for any errors. I would very much appreciate being told of any corrections or possible improvements.

My email address is `mark.wildon@rhul.ac.uk`. You are warmly encouraged to talk to me after lectures or in my office hours.

Lecture times: Monday 3pm (C219), Tuesday 10am (C219), Thursday 10am (WIN105).

Office hours in McCrea 240: Monday 4pm, Tuesday 1pm, Thursday noon.

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.

Aim. This course will give a straightforward introduction to four related areas of combinatorics:

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

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 Matoûšek and Jaroslav Neŝetřil, 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. (Useful for derangements and other examples.)
- Basic definitions of graph theory: vertices, edges, complete graphs. (Needed for Part C on Ramsey Theory.)
- Basic knowledge of discrete probability. I will review this 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. There will be weekly problem sheets; the first will be due in on Monday 11th October. According to audience demand, some of the time on Tuesdays will be used to discuss the problems. Please make a serious attempt at the problem sheets.

Exercises set in these notes are intended to be simple tests that you are following the material. They need not be handed in, but please do all of them.

2. Derangements

In the first lecture I will present the Derangements Problem and solve it by *ad-hoc* methods. Later in the course we will see techniques that can be used to solve this problem more easily.

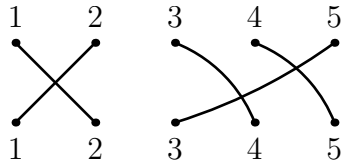
Definition 2.1. A *permutation* of the set $\{1, 2, \ldots, n\}$ is a bijective function

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

A *fixed point* of a permutation σ is an element $k \in \{1, 2, \ldots, n\}$ such that $\sigma(k) = k$. A permutation is a *derangement* if it has no fixed points.

It is often useful to represent permutations by diagrams; the diagram below shows the permutation $\sigma : \{1,2,3,4,5\} \to \{1,2,3,4,5\}$ defined by

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



Problem 2.2 (Derangements)**.** How many of the $n!$ permutations of $\{1,2,\ldots,n\}$ are derangements?

Let d_n be the number of permutations of $\{1,2,\ldots,n\}$ that are derangements. By definition (or convention if you prefer) $d_0 = 1$.

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

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

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

Using this recurrence relation it is easy to find values of d_n for larger n. At this point, N. J. A. Sloane's Online Encyclopedia of Integer Sequences: see `www.research.att.com/~njas/sequences/` can be used to see if a sequence is already known.

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

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

Exercise: (a) check directly that the right-hand side is an integer; (b) use the formula to prove the alternative recurrence relation $d_n = nd_{n-1} + (-1)^n$.

A more systematic way to derive Corollary 2.5 from Theorem 2.4 will be seen in Part B of the course.

Theorem 2.6. *Two probabilistic results:*
(i) *The probability that a randomly chosen permutation of $\{1,2,\ldots,n\}$ 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 will prove more results like this in Part D of the course.

Part A: Enumeration

3. Binomial coefficients and counting problems

We shall define binomial coefficients combinatorially.

Definition 3.1. Let $n, k \in \mathbf{N}_0$. The *binomial coefficient* $\binom{n}{k}$ is the number of k-elements subsets of an n-element set.

By this definition, if $k \notin \mathbf{N}_0$ then $\binom{n}{k} = 0$. Similarly, if $k > n$ then $\binom{n}{k} = 0$. We should check that the combinatorial definition agrees with the usual definition when $k \leq n$.

Lemma 3.2. *If $n, k \in \mathbf{N}_0$ and $k \leq n$ then*

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

Many of the basic properties of binomial coefficients can be given combinatorial proofs involving explicit bijections.

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

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

Lemma 3.4 (Fundamental recurrence)**.** *If $n, k \in \mathbf{N}$ then*

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

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

Theorem 3.5 (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 three results above.

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? What is the connection with Pascal's Triangle?

We can now solve 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 very easily. The entry in the bottom-right can be found in several different ways: two will be demonstrated in this lecture.

Theorem 3.6. *Let $n \in \mathbf{N}$, let $k \in \mathbf{N_0}$. The number of ways to place k indistinguishable balls into n urns of unlimited capacity is $\binom{n+k-1}{k}$.*

The following reinterpretation of this result can be useful.

Corollary 3.7. *Let $n \in \mathbf{N}$, let $k \in \mathbf{N_0}$. The number of solutions of the equation*

$$x_1 + x_2 + \cdots + x_n = k$$

with $x_1, x_2, \ldots, x_n \in \mathbf{N}_0$ is $\binom{n+k-1}{k}$.

4. Further binomial identities

This is a vast subject and we will only cover some aspects. Particularly recommended for further reading is Chapter 5 of *Concrete Mathematics*, reference [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 $k \leq r \leq n$ 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 $n \in \mathbf{N}_0$ 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.

$$\binom{n}{0} + 2\binom{n}{1} + 2^2\binom{n}{2} + \cdots + 2^{n-1}\binom{n}{n-1} + 2^n\binom{n}{n} = 3^n.$$

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

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

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

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

Perhaps surprisingly, there is no simple expression for unsigned row sums $\sum_{k=0}^{r}\binom{n}{k}$. (Except when $r = n$ of course.)

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

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

For the column sums on Pascal's triangle, see Sheet 1, Question 3. For the other diagonal sum see Sheet 1, Question 6.

5. Principle of Inclusion and Exclusion

The *Principle of Inclusion and Exclusion* (PIE) is an elementary way to find the sizes of unions or intersections of finite sets.

If A is a subset of a *universe* set X, we denote by $\bar{A}$ the complement of A in X; i.e.,

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

We start with the two smallest non-trivial examples of the principle.

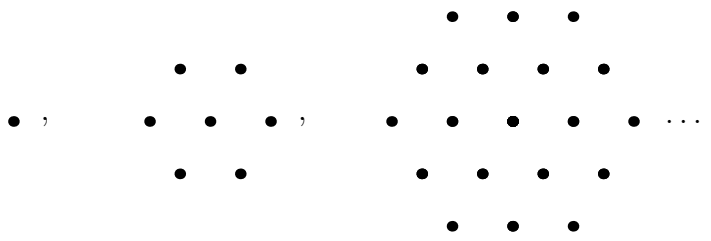
Example 5.1. If A, B, C are subsets of a set X then $|A \cup B| = |A| + |B| - |A \cap B|$ and so

$$\left|\overline{A \cup B}\right| = |X| - |A| - |B| + |A \cap B|.$$

Similarly, $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A| + |A \cap B \cap C|$, so

$$\begin{aligned}\left|\overline{A \cup B \cup C}\right| = |X| - |A| - |B| - |C| \\ +|A \cap B| + |B \cap C| + |C \cap A| - |A \cap B \cap C|.\end{aligned}$$

Example 5.2. The formula for $|A \cup B \cup C|$ gives one of the easiest ways to find the hexagonal numbers.



In the general setting we have a set X and subsets $A_1, A_2, \ldots, A_n$ of X. Let $I \subseteq \{1, 2, \ldots, n\}$ be a non-empty index set. We define

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

Thus A_I is the set of elements of X which belong to all of the sets A_i for $i \in I$. By convention we set

$$A_\varnothing = X.$$

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

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

Exercise: Check Theorem 5.3 when $n = 1$. Check that Theorem 5.3 agrees with Example 5.1 when $n = 2, 3$.

Exercise: Deduce from Theorem 5.3 that

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

6. Applications of the PIE

Derangements. Recall that in Definition 2.1 we defined a derangement of $\{1, 2, \ldots, n\}$ to be a permutation

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

not having any fixed points. Let X be the set of all permutations of $\{1, 2, \ldots, n\}$ and let

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

The set of derangements of $\{1, 2, \ldots, n\}$ is $\overline{A_1 \cup A_2 \cup \cdots \cup A_n}$, and

$$d_n = \left|\overline{A_1 \cup A_2 \cup \cdots \cup A_n}\right|.$$

Lemma 6.1. *Let $I \subseteq \{1, 2, \ldots, n\}$. The set $A_I = \bigcap_{i \in I} A_i$ consists of all permutations of $\{1, 2, \ldots, n\}$ which fix the elements of I. If $|I| = k$ then*

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

Using Lemma 6.1 and the PIE one can give a quick proof of Corollary 2.5, that

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

The PIE is often applicable when we have a set of objects, and we want to count those objects having *none* of a list of properties.

Corollary 6.2. *Let X be a set. Suppose that each $x \in X$ may have some of the properties P_1, P_2, ..., P_n. For $I \subseteq \{1, 2, \ldots, n\}$, let N_I be the number of elements of X which have all the properties P_i for $i \in I$. The number of elements of X having none of the properties is*

$$\sum_{I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|} N_I$$

For example, when $n = 3$, the number of objects with none of the properties is

$$N_\varnothing - N_{\{1\}} - N_{\{2\}} - N_{\{3\}} + N_{\{1,2\}} + N_{\{1,3\}} + N_{\{2,3\}} - N_{\{1,2,3\}}.$$

Note that $N_\varnothing = X$.

Surjective functions. Let $k, n \in \mathbf{N}$. The PIE can be used to count the number of surjective functions from $\{1, 2, \ldots, k\}$ to $\{1, 2, \ldots, n\}$. In this situation Corollary 6.2 is certainly the most useful formulation.

Sieving for primes. 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 6.3. Take $M = 30$. Let $X = \{1, 2, \ldots, 30\}$. We define three subsets of X:

$$B(2) = \{m : 1 \leq n \leq 30,\ m \text{ is divisible by } 2\},$$
$$B(3) = \{m : 1 \leq n \leq 30,\ m \text{ is divisible by } 3\},$$
$$B(5) = \{m : 1 \leq n \leq 30,\ m \text{ is divisible by } 5\}.$$

Any composite number ≤ 30 is divisible by either 2, 3 or 5. Hence

$$\overline{B(2) \cup B(3) \cup B(5)} = \{1\} \cup \{p : 5 < p \leq 30,\ p \text{ is prime}\}.$$

We will find the size of the left-hand side using the PIE, and hence count the number of primes ≤ 30.

The example can be generalized to count numbers not divisible by any of a specified set of primes.

Lemma 6.4. *Let $r, M \in \mathbf{N}$. The number of numbers $\leq M$ that are divisible by r is $\lfloor N/r \rfloor$.*

Theorem 6.5. *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*

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

It is possible to use Theorem 6.5 to show that if $\pi(M)$ is the number of prime numbers $\leq M$ then

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

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

7. Rook polynomials

Many enumerative problems can be expressed in terms of counting permutations with some restriction on their structure. The derangements problem is a typical example. We shall see that rook polynomials give 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). The examples below are taken from Bryant's book.

Example 7.1. 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 lectures at most one course, how many ways are there to assign professors to courses?

Example 7.2. How many derangements σ of $\{1, 2, 3, 4, 5\}$ have the property that $\sigma(i) \neq i + 1$ for $1 \leq i \leq 4$?

Definition 7.3. A *board* is a subset of the squares of an $n \times n$ chessboard. Given a board B, 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

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

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

Lemma 7.4. *The rook polynomial of the $n \times n$ board is*

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

The following lemmas are very useful for finding rook polynomials.

Lemma 7.5. *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*

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

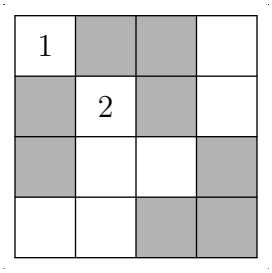
Rook polynomials are, in particular, generating functions. This is the first of many times that multiplying generating functions will help us to solve problems.

Lemma 7.6. *Let B be a board and let s be a square in B. Let*

- *C be the board obtained from B by deleting s;*
- *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 7.7. The rook polynomial for the board in Example 7.1 can be found by applying Lemma 7.6 to the two squares indicated below.



Our final technique for finding rook polynomials is often the most useful in practice. We need the lemma below.

Lemma 7.8. *Let $I \subseteq \{1, 2, \ldots, n\}$ be a subset of size k. If $g : I \to I$ is a permutation then there are $(n-k)!$ permutations $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ such that $f(x) = g(x)$ if $x \in I$.*

We used a special case of this lemma to prove the part of Lemma 6.1 which says that there are $(n-k)!$ permutations of $\{1, 2, \ldots, k\}$ which fix a subset of size k.

Theorem 7.9. *Let B be part of the $n \times n$ chessboard with rook polynomial*

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

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

$$n! - (n-1)!\, r_1(B) + (n-2)!\, r_2(B) - \cdots + (-1)^n\, r_0(B).$$

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

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

Part B: Generating functions

8. Introduction to generating functions

Definition 8.1. The *ordinary generating function* associated to a sequence $a_0, a_1, a_2, \ldots$ is the power series

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

Sometimes we shall drop the word 'ordinary' and just write *generating function*. The coefficients a_n will usually be integers, but this is not essential.

If there exists N such that $a_n = 0$ if $n > N$ then the ordinary generating function of the sequence is a polynomial. Rook polynomials (see Definition 7.3) are therefore generating functions.

Analytic and formal interpretations. We can think of a generating function in two ways. Either:

- As a formal power series with x acting as a place-holder. This is the 'clothes-line' interpretation.
- As a function of a real (or complex) variable x convergent for x such that $|x| < r$, where r is a positive real number.

The formal point of view is often convenient, because it allows us to define and manipulate (by adding, multiplying etc.) power series without worrying about convergence. For example,

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

is a perfectly respectable formal power series, even though it only converges when $x = 0$. That said, all the generating functions one normally encounters have positive radius of convergence. So (except when we are proving asymptotic results) we can take either point of view.

In §2.1 of *generatingfunctionology*, Wilf discusses these issues and gives definitions of the sum and product of two formal power series, which agree with the analytic definitions when the two series converge. (He also defines reciprocals and compositions, where this is possible.) For instance, the *product* of formal power series $\sum_{n=0}^{\infty} a_n x^n$ and $\sum_{n=0}^{\infty} b_n x^n$ is defined to be $\sum_{n=0}^{\infty} c_n x^n$ where

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}$$

The exercise below gives his definition of the derivative.

Exercise: Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ be a formal power series. The *derivative* of f is defined by

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

Show that if $f'(x) = f(x)$ then $f(x) = a_0 \exp x$.

Examples of generating functions. We shall look at three typical problems involving ordinary generating functions. The first two are interesting to do using (an extreme version of) the formal point of view.

Example 8.2. How many ways are there to tile a $2 \times n$ path with bricks that are either 1×2 or 2×1?

Example 8.3. How many ways are there to pay for a newspaper costing 30p using only 5p and 2p coins?

The second example suggests that it would be useful to know the power series for $1/(1-x)^n$.

Theorem 8.4. *If $n \in \mathbf{N}$ then*

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

There are (at least) three ways to prove this formula. There is a nice combinatorial proof which uses the urn problem considered in Theorem 3.6. Another proof uses the analytic version of the Binomial Theorem stated below.

Theorem 8.5 (Binomial Theorem for general exponents)**.** *Let $\alpha \in \mathbf{R}$. If $|x| < 1$ then*

$$(1+x)^{\alpha} = \sum_{n=0}^{\infty} \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} x^n.$$

When we deal with Catalan numbers we shall need this formula in the case $\alpha = 1/2$.

Exercise: Give a third proof of Theorem 8.4 by induction on n.

9. Recurrence relations and asymptotics

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 $\sum_{n=0}^{\infty} a_n x^n$;

(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. Terms such as na_n suggest differentiation, and powers of x will usually be needed to get everything to match up correctly. Take care to avoid minor slips! In Step (c) it is often necessary to use partial fractions.

Example 9.1. Will solve the recurrence $a_n = 5a_{n-1} - 6a_{n-2}$ for $n \geq 2$.

The calculations at the end of this example could be simplified by assuming that *some* partial fraction expression exists, and then determining the unknown constants from the first few terms of the recurrence. This procedure is justified by the following theorem. (**Corrected from previous version in which the exponents d_i in the denominators $(1 - x/\beta_i)^{d_i}$ were missing.**)

Theorem 9.2. *Let $f(x)$ and $g(x)$ be polynomials, with $\deg f < \deg g$. Suppose that g has roots $\beta_1, \beta_2 \ldots, \beta_k \in \mathbf{C}$ where β_i has multiplicity d_i. Then there exist polynomials $P_1(x), \ldots, P_k(x) \in \mathbf{C}$ with $\deg P_i < d_i$ such that*

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

The most important case of the theorem is when g has no repeated roots.

Corollary 9.3. *Let $a_0, a_1, a_2, \ldots$ be a sequence with generating function*

$$\frac{f(x)}{(x - \beta_1) \ldots (x - \beta_k)}$$

where $\deg f < d$ *and the* β_i *are distinct. Then there are constants* $P_i \in \mathbf{C}$ *such that*

$$a_n = P_1/\beta_1^n + \cdots + P_k/\beta_k^n$$

for all $n \in \mathbf{N}_0$.

The theorem can be proved quite quickly from the following lemma, which is certainly non-examinable. To avoid getting bogged down in the technicalities, I will not prove the lemma in lectures. For an alternative exposition see Chapter 25 of Biggs *Discrete mathematics*. (Or Chapter 18 in the first edition.)

Lemma 9.4. *Let* $f(x)$ *and* $g(x)$ *be polynomials. If* $g(x) = (x-\beta)^d h(x)$ *where* $h(\beta) \neq 0$, *then there exist polynomials* $A(x), B(x), K(x)$ *with* $\deg A < d$, $\deg B < \deg h$ *such that*

$$\frac{f(x)}{(x-\beta)^d h(x)} = \frac{A(x)}{(x-\beta)^d} + \frac{B(x)}{h(x)} + K(x).$$

Moreover, if $\deg f < \deg g$ *then* $K(x) = 0$.

Proof. The polynomials $(x-\beta)^d$ and $h(d)$ are coprime. Hence, by the Euclidean Algorithm, there exist polynomials $C(x)$, $D(x)$ such that

$$C(x)h(x) + D(x)(x-\beta)^d = 1.$$

Multiplying by $f(x)$ we get

$$(\star) \qquad f(x)C(x)h(x) + f(x)D(x)(x-\beta)^d = f(x).$$

By polynomial division we may write

$$f(x)C(x) = A(x) + q_C(x)(x-\beta)^d$$
$$f(x)D(x) = B(x) + q_D(x)h(x)$$

where $\deg A < d$, $\deg B < \deg h$. Substituting into $(\star)$ and rearranging gives

$$A(x)h(x) + B(x)(x-\beta)^d = f(x) - (q_C(x) + q_D(x))(x-\beta)^d h(x).$$

Now divide through by $(x-\beta)^d h(x)$ to get

$$\frac{f(x)}{(x-\beta)^d h(x)} = \frac{A(x)}{(x-\beta)^d} + \frac{B(x)}{h(x)} + K(x)$$

where $K(x) = -(q_C(x)+q_D(x))$. If $\deg f < d + \deg h$ then the left-hand side tends to 0 as $x \to \infty$, so we must have $K(x) = 0$. (Alternatively compare degrees in the previous equation.) $\square$

Example 9.5. Suppose that the generating function of the sequence $a_0, a_1, a_2, \ldots$ is

$$\frac{4x^2 - 13x + 12}{(x-2)^2(x-1)}$$

Theorem 9.2 can be used to find a formula for the a_n.

It is worth noting the MATHEMATICA command `Apart` for finding partial fraction expansions.

Example 9.6. The recurrence $a_n = 4(a_{n-1} - a_{n-2})$ for $n \geq 2$ with $a_0 = 1$, $a_1 = 4$ has the unique solution $a_n = 2^n(n+1)$.

In Theorem 2.4 we derived the recurrence $d_n = (n-1)(d_{n-2} + d_{n-1})$ for the number of derangements of the set $\{1, 2, \ldots, n\}$. Generating functions give a systematic way to obtain the formula first stated in Corollary 2.5.

Theorem 9.7. *Let $p_n = d_n/n!$ be the probability that a randomly chosen permutation of $\{1, 2, \ldots, n\}$ is a derangement. Then*

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

The remainder of this section is non-examinable, and will be omitted if time is pressing. We shall need a standard piece of notation.

Definition 9.8. Given a sequence a_0, a_1, a_2 and a function $t : \mathbf{R} \to \mathbf{R}$, we write $a_n = O(t(n))$ if there exists a constant $c \in \mathbf{R}$ such that $|a_n| < ct(n)$ for all $n \in \mathbf{N}_0$.

Theorem 9.9. *Let $F(x) = \sum_{n=0}^{\infty} a_n x^n$ be the ordinary generating function for the sequence $a_0, a_1, a_2, \ldots$. Suppose that $F(x) = f(x)/g(x)$ where $f(x), g(x)$ are polynomials and $\deg g \geq 1$. Let $\beta \in \mathbf{C}$ be the root of g of minimum modulus. Given any $\varepsilon > 0$,*

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

More generally, if $F(x) = \sum_{n=0}^{\infty} a_n x^n$ is the ordinary generating function for the sequence $a_0, a_1, a_2, \ldots$, and $F(x)$ has no singularities with modulus $< \beta$, then the conclusion of the theorem still holds. See Theorem 2.25 and the discussion in §5.2 of Wilf *generatingfunctionology*.

10. Convolutions and the Catalan numbers

Definition 10.1. 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

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

Keeping the notation from the definition, let $F(x) = \sum_{n=0}^{\infty} a_n x^n$, let $G(x) = \sum_{n=0}^{\infty} b_n x^n$ and let $H(z) = \sum_{n=0}^{\infty} c_n x^n$. By definition of the product of formal power series, we have $F(x)G(x) = H(x)$. This makes generating functions ideal for finding sequences defined by convolutions.

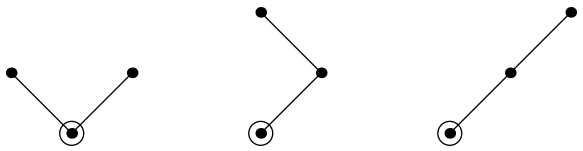
Convolutions frequently arise in combinatorial problems. See Problem Sheet 4 for some more examples.

Example 10.2. Given a pile of indistinguishable building blocks, how many ways are there to use n blocks to make an equilateral triangle and a square?

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

Definition 10.3. 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 $C_3 = 5$. **Corrected from the wrong** $C_4 = 5$. Three of them are shown below, with the root vertex circled. The other two can be obtained by reflection.



Lemma 10.4. *If* $n \in \mathbf{N}$ *then*

$$C_n = C_0 C_{n-1} + C_1 C_{n-2} + \cdots + C_{n-2} C_1 + C_{n-1} C_0.$$

Theorem 10.5. *If* $n \in \mathbf{N}_0$ *then* $C_n = \frac{1}{n+1}\binom{2n}{n}$.

We shall prove Theorem 10.5 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. The steps (given in outline at the end of lecture 15) are:

(a) Use the recurrence in Lemma 10.4 to show that $F(x)$ satisfies the equation
$$xF(x)^2 = F(x) - 1.$$

(b) Solve this quadratic equation to get the closed form
$$xF(x) = \frac{1 - \sqrt{1-4x}}{2}.$$
We choose the negative root because when $x = 0$ the left-hand side is 0. So the right-hand side must also be 0 when $x = 0$. This is the case if we take the negative root, but not if we take the positive root.

(c) Use the general version of the Binomial Theorem stated in Theorem 8.5 to get a formula for the coefficients.

The resulting formula for the Catalan numbers is surprisingly simple, but not easy to prove without using generating functions. I hope you will agree that while the generating function proof takes some work, each step is essentially routine.

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

Lemma 10.6. *If $n \in \mathbf{N}_0$ then*
$$\sum_{r=0}^{n} \binom{n}{r} d_{n-r} = n!.$$

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

Theorem 10.7. *If $G(x) = \sum_{m=0}^{\infty} d_m x^m/m!$ 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.

Remark: The *exponential generating function* associated to the sequence $a_0, a_1, a_2, \ldots$ is
$$\sum_{n=0}^{\infty} \frac{a_n}{n!} x^n.$$

The argument used above is a typical example of how convolutions of exponential generating functions are used in practice. Question 10 on Sheet 5 on the Bell numbers gives another application. See Wilf *generatingfunctionology* Chapter 3 for a full account.

11. Partitions

Definition 11.1. 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) $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq 1$
(ii) $\lambda_1 + \lambda_2 + \ldots + \lambda_k = n$.

Let $p(n)$ be the number of partitions of n. The entries in a partition are called *parts*.

By this definition, $\varnothing$ is the unique partition of 0. The sequence of partition numbers begins 1, 1, 2, 3, 5, 7, 11, 15, ...

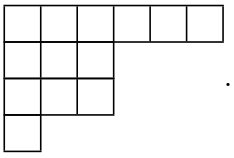
Example 11.2. Example 8.3 can be re-interpreted in terms of partitions: the number of ways to pay for something costing n pence with 2p and 5p coins is the number of partitions of n into parts of sizes 2 and 5. We saw that the associated generating function is

$$\frac{1}{(1-x^2)(1-x^5)}.$$

Theorem 11.3. *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



.

Sometimes it is more convenient to use dots rather than boxes.

The next theorem has a very simple proof using Young diagrams.

Theorem 11.4. *Let $n, k \in \mathbf{N}$ 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.*

Definition 11.5. We say that a partition has *distinct parts* if it has at most one part of any given size. Let $d(n)$ be the number of partitions of n with distinct parts.

Exercise: Show that the sequence $d(n)$ for $n \in \mathbf{N}_0$ starts 1, 1, 1, 2, 2, 3, 4, 5, 6, 8, 10, 12.

The following theorem is easily proved using generating functions. There are also bijective proofs using Young diagrams, but none are completely straightforward. Note how we adapt the proof of Theorem 11.4 to get the generating functions for the special types of partition.

Theorem 11.6. *Let $n \in \mathbf{N}$. The number of partition of n into parts of odd size is equal to the number of partition of n with distinct parts.*

12. Euler's Pentagonal Number Theorem

Definition 12.1. A *pentagonal number* is any number of the form

$$\frac{m(3m \pm 1)}{2}$$

for $m \in \mathbf{N}$.

The sequence of pentagonal numbers starts $1, 2, 5, 7, 12, 15, 22, 26, \ldots$. See Question 8 on Sheet 5 for why they are so-named.

Considering only the partitions of n with distinct parts, let

- $d_o(n)$ be the number of partitions with an odd number of parts;
- $d_e(n)$ be the number of partitions with an even number of parts.

Our aim in this section is to prove the following theorem.

Theorem 12.2 (Euler's Pentagonal Number Theorem)**.** *Let $n \in \mathbf{N}$.*

(i) *If n is not a pentagonal number then $d_o(n) = d_e(n)$;*
(ii) *If $n = m(3m \pm 1)/2$ then $d_o(n) = d_e(n) + (-1)^{m+1}$.*

The sign in (ii) can be reconstructed from the cases $n = 1$ and $n = 2$, so all one has to remember is that usually $d_o(n) = d_e(n)$, with an error of ± 1 in the exceptional cases when n is a pentagonal number.

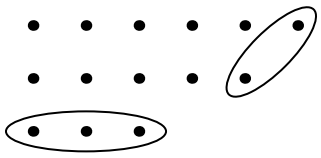
The bijective proof given below is due to F. Franklin (1881). It was said by H. Rademacher (who proved an exact asymptotic formula for $p(n)$ by building on the work of Hardy and Ramanujan) to be 'the first major achievement of American mathematics'.

Definition 12.3. Let λ be a partition of n with distinct parts. The

- *base* of λ is all dots in the bottom row of its Young diagram;
- *slope* of λ consists of the dot at the end of the largest row and the dots diagonally below it to the south-west.

Let $b(\lambda)$ be the number of dots in the base and let $s(\lambda)$ be the number of dots in the slope.

For example, the base and slope of the partition $(6, 5, 3)$ are ringed in its Young diagram below; $b(6, 5, 3) = 3$ and $s(6, 5, 3) = 2$.



The terminology in the next definition is not standard, but seems convenient.

Definition 12.4. Let λ be a partition. To perform a

- *base move* on λ, remove the base and add it as a new slope;
- *slope move* on λ, remove the slope and add it as a new base.

We only allow these moves to be applied when they lead to a new partition. Say that a partition is

- *thin* if a base move can be applied to it;
- *thick* if a slope move can be applied to it.

At most one type of move can be applied to any partition.

Lemma 12.5. *If λ is a partition with m distinct parts then*

(i) *λ is thin if and only if $b(\lambda) \leq s(\lambda)$ and*

$$\lambda \neq (2m-1, 2m-2, \ldots, m+1, m);$$

(ii) *λ is thick if and only if $b(\lambda) > s(\lambda)$ and*

$$\lambda \neq (2m, 2m-1, \ldots, m+1).$$

Theorem 12.6. *The base and slope moves are mutually inverse bijections between the thin and thick partitions of n.*

We have now done almost all the work needed to prove Theorem 12.2.

Proof of Euler's Pentagonal Number Theorem. If n is not a pentagonal number then any partition of n is either thin or thick. By Theorem 12.6, the two classes are in bijection by maps that changes the number of parts by 1. Hence $d_o(n) = d_e(n)$. If $n = m(3m \pm 1)/2$ then the relevant partition with m parts from Lemma 12.5 is left over. So if m is odd then $d_o(n) = d_e(n) + 1$ and if m is even then $d_e(n) = d_o(n) + 1$. $\square$

We now use generating functions to turn Euler's Pentagonal Theorem into a recurrence relation for the $p(n)$.

Corollary 12.7.

$$\prod_{n\geq 1}(1-x^n) = 1 + \sum_{k=1}^{\infty}(-1)^k\left(x^{k(3k-1)/2} + x^{k(3k+1)/2}\right)$$

Exercise: check by expanding the product by hand that the coefficients of x^n on either side agree for small n.

Corollary 12.8. *If $n \in \mathbf{N}$ then*

$$\begin{aligned} p(n) &= \sum_{k=1}^{\infty}(-1)^{k-1}\Big((p(n-\tfrac{1}{2}k(3k-1)) + p(n-\tfrac{1}{2}k(3k+1)\Big) \\ &= p(n-1) + p(n-2) - p(n-5) - p(n-7) + \cdots \end{aligned}$$

where we set $p(m) = 0$ if $m < 0$.

Previously we have usually started with a recurrence relation, and then used it to find a generating function. For instance, in Example 8.2 we started with the Fibonacci recurrence $a_n = a_{n-1} + a_{n-2}$ and showed that the associated generating function $F(x) = \sum_{n=0}^{\infty} a_n x^n$ satisfied $(1 - x - x^2)F(x) = 1$. This time we have started with the generating function and used its reciprocal to find a (highly non-obvious!) recurrence relation.

Asymptotics of $p(n)$. It is possible to end with an open problem. In 1918 Hardy and Ramanujan proved that $p(n)$ is asymptotic to

$$\frac{1}{4n\sqrt{3}}\mathrm{e}^{a\sqrt{n}}$$

where $a = 2\sqrt{\pi^2/6}$. Their paper introduced a number of ideas, including the circle-method, which have been highly influential in later work. Since their paper, several easier proofs of weaker results, for example, that $\log p(n) \leq a\sqrt{n}$, have been found. (See Question 10 on Sheet 6.)

However, the problem of finding good bounds for $p(n)$ by entirely combinatorial arguments is still largely open. For instance, is there a combinatorial proof that there is a constant $A \in \mathbf{R}$ such that

$$p(n) \leq A^{\sqrt{n}} \quad \text{for all } n \in \mathbf{N}?$$

Part C: Ramsey Theory

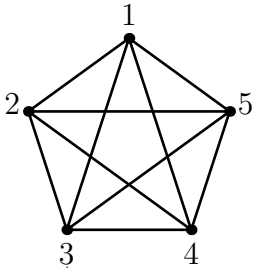
13. Introduction to Ramsey Theory

The idea behind Ramsey theory is that any sufficiently large structure should 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).

Most of the results in this area concern graphs: we shall concentrate on the finite case.

Definition 13.1. A *graph* is a set X of *vertices* together with a set E of 2-subsets of X called *edges*. The *complete graph* on X is the graph whose edge set is all 2-subsets of X.

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



We denote the complete graph with n vertices by K_n. The graph K_3 is often called a *triangle*.

Exercise: Find the number of edges in K_n.

Definition 13.2. Let $c \in \mathbf{N}$ and let G be a complete graph, with edge set E. A *c-colouring* of G is a function from E to $\{1,2,\ldots,c\}$. If Y is an r-set of vertices of G such that all edges between vertices in Y have the same colour, then we say that Y is a *monochromatic K_r*.

Note that it is the edges that are coloured, *not the vertices*.

In practice we shall specify graphs and colourings rather less formally. It seems to be a standard convention that colour 1 is red, colour 2 is blue and colour 3 (which we won't need for a while) is green.

Example 13.3. In any two-colouring of the edges of K_6, there is either a red triangle, or a blue triangle.

Definition 13.4. Given $s, t \in \mathbf{N}$, 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 13.2 that $R(3,3) \leq 6$. We will prove in Theorem 15.2 that all Ramsey numbers exist; please assume this in the two exercises below.

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

Exercise: Let $s, t \in \mathbf{N}$. Show that $R(s,t) = R(t,s)$. Show that $R(s,t) \leq R(s',t')$ whenever $s \leq s'$ and $t \leq t'$.

Two families of Ramsey numbers are easily found.

Lemma 13.5. *If $s \in \mathbf{N}$ then $R(1,s) = 1$ and $R(2,s) = s$.*

The main idea need to prove Theorem 15.2 appears in the next example.

Example 13.6. 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 13.7 (Hand-Shaking Lemma)**.** *Let G be a graph with vertex set $\{1, 2, \ldots, n\}$ 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 13.8. $R(3,4) = 9$.

The proof of the final theorem is left to you: see Questions 2 and 3 on Sheet 6.

Theorem 13.9. $R(4,4) = 18$.

For a survey of other known results on $R(s,t)$ for small s and t, see Stanisław Radziszowski, *Small Ramsey Numbers*, Electronic Journal of Combinatorics, available from `www.combinatorics.org/Surveys`.

14. The Pigeonhole Principle

The Pigeonhole Principle can be stated as follows.

Theorem 14.1 (Pigeonhole Principle)**.** *If m balls are coloured with $m-1$ colours, then there are two balls of the same colour.*

We used a variant form in Examples 13.3 and 13.5. Let $a+b=n$. If $n-1$ edges are coloured red or blue then either there are a red edges, or there are b blue edges.

The remaining material in this section is non-examinable, and included for interest only. Dedekind's original application of the Pigeonhole Principle was as follows.

Theorem 14.2. *Let $\alpha \in \mathbf{R}$ and let $N \in \mathbf{N}$. There exists $p, q \in \mathbf{N}$ such that $q \leq N$ and*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2}.$$

Like many results proved using the Pigeonhole Principle or results from Ramsey Theory, the proof does not give us an efficient algorithm for finding p and q. (This can be done using continued fractions.)

Another typical application of the Pigeonhole Principle:

Theorem 14.3. *Let $n \in \mathbf{N}$ and let $A \subseteq \{1, 2, \ldots, 2n\}$ with $|A| = n+1$.*
(i) *There exists $x, y \in A$ such that x and y are coprime.*
(ii) *There exist $x, y \in A$ such that x divides y.*

If $|A| = n$ then both (i) and (ii) can fail to hold.

For three further applications see Question 7 on Sheet 6. It is also possible to Question 8 by a clever application of the Pigeonhole Principle; the proof suggested in the hint is much easier!

15. Ramsey's Theorem

We shall prove an upper bound for the Ramsey numbers $R(s,t)$ by induction on $s+t$. The following lemma gives the inductive step.

Lemma 15.1. *Let $s, t \in \mathbf{N}$ with $s, t \geq 2$. 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 15.2. *For any $s, t \in \mathbf{N}$ the Ramsey number $R(s,t)$ exists and*

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

Corollary 15.3. *If $s \in \mathbf{N}$ then*

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

and there exists a constant $C \in \mathbf{R}$ such that

$$R(s,s) \leq 4^s C$$

for all $s \in \mathbf{R}$.

Using Stirling's Formula one can show that $\binom{2s-2}{s-1} \leq 4^{s-1}/\sqrt{s}$, and so get the stronger bound

$$R(s,s) \leq 4^{s-1}/\sqrt{s}.$$

This result was due to 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) $\sqrt{s}$ can be replaced with s.

In 1947 Erdös proved the lower bound

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

His argument becomes clearest when stated using the language of probability: we will see it in part D of the course.

To end this introduction to Ramsey Theory we shall give two interesting applications of Theorem 15.2.

Many colours.

Theorem 15.4. *There exists $n \in \mathbf{N}$ such that if the edges of the complete graph on $\{1, 2, \ldots, n\}$ are coloured red, blue and green, then there exists a monochromatic triangle.*

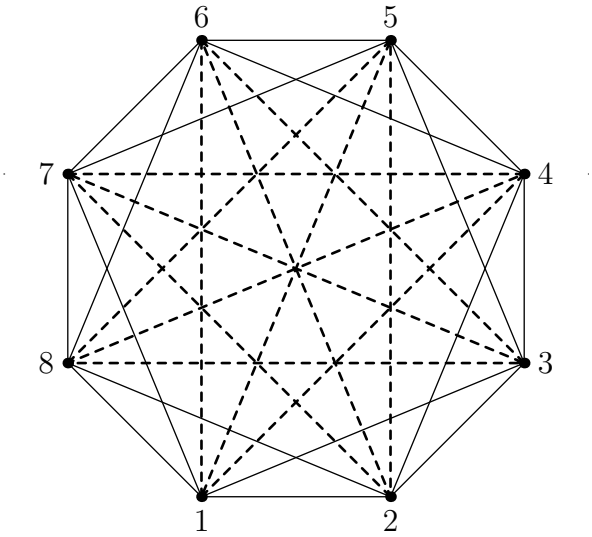
There are (at least) two ways to prove Theorem 15.4. The first adapts our usual argument, looking at the edges coming out of vertex 1 and concentrating on those vertices joined by edges of the majority colour. The second uses a neat trick to reduce to the two-colour case.

The following more general theorem can be proved by either of these arguments.

Theorem 15.5. *Let $c \in \mathbf{N}$. There exists $n \in \mathbf{N}$ such that if the edges of the complete graph on $\{1, 2, \ldots, n\}$ are coloured with c different colours, then there exists a monochromatic triangle.*

Schur's Theorem. Recall that we used the graph below to show that $R(3, 4) > 8$. (Dashed edges are red and solid lines are blue.) Observe that it has a strong regularity property: the colour of the edge $\{x, y\}$ depends only on $|x - y|$.

Exercise: Check that the dashed edges have differences $3, 4, 5$ and the blue edges differences $1, 2, 6, 7$.



In our application, we shall construct such colourings of K_n when n is big enough to guarantee there will be a monochromatic triangle. The smallest interesting example is given in the following lemma.

Lemma 15.6. *If $\{1, 2, 3, 4, 5\}$ is partitioned into two subsets so that $\{1, 2, 3, 4, 5\} = Y \cup Z$, then either there exist $y, y', y'' \in Y$ such that $y + y' = y''$, or there exist $z, z', z'' \in Z$ such that $z + z' = z''$.*

The general theorem is due to Schur (1916).

Theorem 15.7. *Let $c \in \mathbf{N}$. There exists n such that if $\{1, 2, \ldots, n\}$ is partitioned into c subsets $Y_1, Y_2, \ldots, Y_c$ then there exists a subset Y_k and $y, y', y'' \in Y_k$ such that $y + y' = y''$.*

Schur's Theorem was the first in a long line of deep theorems combining arithmetic with combinatorics. A descendant is the 2004 result of Ben Green and Terence Tao that the primes contain arbitrarilyy long arithmetic progressions.

Part D: Probabilistic Methods

17. 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 setup of probability spaces and random variables recalled in §16. It will be particularly important for you to ask questions if the use of anything from this section seems unclear.

Fix $n \in \mathbf{N}$. Let Ω be the set of all permutations of the set $\{1, 2, \ldots, n\}$. For each $\sigma \in \Omega$, let $p_\sigma = 1/n!$; 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 $A = \{\sigma \in \Omega : \sigma(x) = x\}$. Then A is the event that a permutation fixes x. What is the probability of A?

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

Theorem 17.1. *Define a random variable $X : \Omega \to \mathbf{N}$ by letting $X(\sigma)$ be the number of fixed-points of the permutation σ. Then $\mathbf{E}[X] = 1$.*

To proceed further we need cycles and the cycle decomposition of permutations.

Definition 17.2. A permutation τ of $\{1, 2, \ldots, n\}$ is a *k-cycle* if there is a k-subset

$$\{x_1, x_2, \ldots, x_k\} \subseteq \{1, 2, \ldots, n\}$$

such that

$$\tau(x_1) = x_2, \tau(x_2) = x_3, \ldots, \tau(x_k) = x_1$$

and $\tau(y) = y$ if $y \notin \{x_1, \ldots, x_k\}$. We shall write $\tau = (x_1, x_2, \ldots, x_k)$. We say that cycles $\tau = (x_1, \ldots, x_k)$ and $\rho = (y_1, \ldots, y_\ell)$ are *disjoint* if

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

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

Lemma 17.3. *Any permutation can be written as a composition of disjoint cycles. The cycles in this composition are uniquely determined by the permutation.*

Given a permutation σ of $\{1, 2, \ldots, n\}$ and $x \in \{1, 2, \ldots, n\}$ we can ask: what is the probability that x lies in a k-cycle of σ, for some given k? We have already seen 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 17.4. *Let $1 \leq k \leq n$ and let $x \in \{1, 2, \ldots, n\}$. The probability that x lies in an k-cycle of a permutation of $\{1, 2, \ldots, n\}$ chosen uniformly at random is $1/n$.*

Theorem 17.5. *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 10.6: there we got a recurrence by considering fixed points; here we get a recurrence by considering cycles.

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

Corollary 17.6. *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 17.1.

Theorem 17.7. *Let C_k 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 randomly chosen permutation has an k-cycle.

18. Ramsey Numbers and the First Moment Method

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

Lemma 18.1 (First Moment Method)**.** *Let Ω be a probability space and let $X : \Omega \to \mathbf{N}_0$ be a random variable. If $\mathbf{E}[X] = x$ then*

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

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

Exercise: check that the lemma holds in the case where

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

models the throw of two fair dice and $X(x, y) = x + y$.

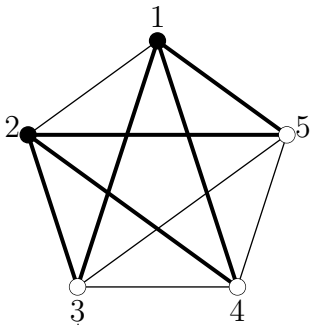
More generally, the k-th *moment* of X is $\mathbf{E}[X^k]$. Sometimes stronger results can be obtained by considering these higher moments. We shall concentrate on first moments, where the power is the method is closely related to the linearity property of expectation (see Lemma 16.8).

Our applications will come from graph theory.

Definition 18.2. Let G be a graph with vertex set V. A *cut* of G is a partition of V into two disjoint subsets A and B. The *capacity* of the cut is the number of edges of G that meet both A and B.

Note that $B = V \setminus A$ and $A = V \setminus B$, so a cut can be specified by giving either of the sets in the partition.

For example, the diagram below shows the cut in the complete graph on $\{1, 2, 3, 4, 5\}$ where $A = \{1, 2, 3\}$ and $B = \{4, 5\}$. The capacity of this cut is 6, corresponding to the 6 edges $\{x, y\}$ for $x \in A$, $y \in B$ shown with thicker lines.



Theorem 18.3. *Let G be a graph with n vertices and 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 18.4. *Let G be the complete graph on $\{1,2,\ldots,n\}$ and let Ω be the set of all red-blue colourings of G. Let*

$$p_\omega = \frac{1}{2^{\binom{n}{2}}}$$

for each $\omega \in \Omega$. Then

(i) *Ω is a probablity space in which each colouring is equally probable.*

(ii) *For each edge $\{x,y\}$ of G,*

$$\mathbf{P}[\{x,y\} \text{ is red}] = \mathbf{P}[\{x,y\} \text{ is blue}] = 1/2.$$

Theorem 18.5. *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 and no blue K_s.

Corollary 18.6. *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 $\{1,2,\ldots,42\}$ at random, then we will fairly soon get a colouring with no monochromatic K_8. However, to check we have succeeded, we will have to check all $\binom{42}{8} = 118030185 \approx 1.18 \times 10^8$ subsets of $\{1,2,\ldots,42\}$. Unsuccessful colourings may be spotted more quickly, but there is still a big cost.

Therefore Theorem 18.5 does not give an effective construction. It is a major unsolved problem to find, for each $s \in \mathbf{N}$, an explicit colouring of the complete graph on $n = 1.01^s$ vertices with no monochromatic K_s. (Here 1.01 could be replaced with any $\alpha > 1$.)

19. Lovász Local Lemma

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

In the proof of Theorem 18.5, we considered a random colouring of the complete graph on $\{1, 2, \ldots, n\}$ and used Lemma 18.1 to show that, provided

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

there was a positive probability that the colouring had no monochromatic K_s.

As motivation for the Lovász Local Lemma, consider the following alternative argument, which avoids the use of Lemma 18.1.

Alternative proof of Theorem 18.5. 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 of $\{1, 2, \ldots, n\}$ let E_A be the event that A is monochromatic. The event that no K_s is monochromatic is $\bigcap_A \overline{E_A}$, where the intersection is taken over all s-subsets A of $\{1, 2, \ldots, n\}$ and $\overline{E_A} = \Omega \backslash E_A$. So it will suffice to show that $\mathbf{P}[\bigcap_A \overline{E_A}] > 0$, or equivalently, that

$$\mathbf{P}\big[\bigcup_A E_A\big] < 1.$$

It it always the case that the probability of a union of events is at most the sum of their probabilities. So it will be enough to show that

$$\sum_A \mathbf{P}[E_A] < 1.$$

The probability of E_A was found in lectures to be $2^{1-\binom{s}{2}}$ (for any A). Hence

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

which is < 1 by assumption. □

Now, if the events E_A were independent, we would have

$$\mathbf{P}[\bigcap_A \overline{E_A}] = \prod_A \mathbf{P}[\overline{E_A}].$$

Since each event $\overline{E_A}$ has non-zero probability, this would show at once that their intersection has non-zero probability, as required. 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 19.1. An event E is *mutually independent* of a collection T of events, if for all $U \subset T$, $U' \subset T \setminus U$, we have

$$\mathbf{P}\left[E \,\middle|\, \bigcap_{E_u \in U} E_u \cap \bigcap_{E_{u'} \in U'} \overline{E_{u'}}\right] = \mathbf{P}[E].$$

For example, if the events E_A are as defined above, then E_A is independent of the set $\{E_B : |A \cap B| \leq 1\}$. This is because if $A \cap B$ has at most one element, then no edge is common to both A and B. Hence knowing whether or not A is monochromatic gives no information about B.

Lemma 19.2 (Symmetric Lovász Local Lemma)**.** *Let $d \in \mathbf{N}$. Let S be a collection of events such that $\mathbf{P}[E] \leq p$ for all $E \in S$. Suppose that for each event $E \in S$, there is a subset $\mathcal{T}_E$ of $\mathcal{S}$ such that*

(i) $|\mathcal{T}_E| \geq |S| - d$
(ii) *E is independent of $\mathcal{T}_E$.*

If $\mathrm{e}p(d+1) \leq 1$ then

$$\mathbf{P}[\bigcap_{A} \overline{E_A}] > 0.$$

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 slightly weaker result is given in §6.7 of Michael Mitzenmacher and Eli Upfal *Probability and Computing* ([6] in the list of page 2).

The Lovász Local Lemma can be used to prove a slightly stronger version of Theorem 18.5.

Theorem 19.3. *Let $n, s \in \mathbf{N}$. If*

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

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

Proof: Keep the notation from the proof of Theorem 18.5. Let A be an s-subset of $\{1, 2, \ldots, n\}$. We remarked above that E_A is independent of the set of events $\{E_B : |A \cap B| \leq 1\}$. There are at most

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

sets B which meet A in ≥ 2 elements, since we can choose two common elements in $\binom{s}{2}$ ways, and then choose any $s-2$ elements to complete B. (There is some overcounting here, so this is only an upper bound.) Therefore we let $d = \binom{s}{2}\binom{n-2}{s-2}$. Since

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

for all A, we take $p = 2^{1-\binom{n}{s}}$. Then we can apply the Lovász Local Lemma provided that $ep(d+1) \leq 1$, which is exactly the hypothesis of the theorem. Hence

$$\mathbf{P}[\bigcap_A \overline{E_A}] > 0$$

as required. □

Theorem 19.2 is stronger than Theorem 18.5 when s is reasonably large.

Example 19.4. When $s = 15$, the largest n such that

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

is $n = 792$. So Theorem 18.5 tells us that $R(15, 15) > 792$. But

$$\mathrm{e}\left(\binom{15}{2}\binom{n-2}{13} + 1\right) \leq 2^{\binom{15}{2}-1}$$

provided $n \leq 947$. Theorem 19.2 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

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

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