Combinatorics: The Art of Counting

Bruce E. Sagan

To Sally, for her love and support

Contents

Preface

Enumerative combinatorics has seen an explosive growth over the last 50 years. The purpose of this text is to give a gentle introduction to this exciting area of research. So, rather than trying to cover many different topics, I have chosen to give a more leisurely treatment of some of the highlights of the field. My goal has been to write the exposition so it could be read by a student at the advanced undergraduate or beginning graduate level, either as part of a course or for independent study. The reader will find it similar in tone to my book on the symmetric group. I have tried to keep the prerequisites to a minimum, assuming only basic courses in linear and abstract algebra as background. Certain recurring themes are emphasized, for example, the existence of sum and product rules first for sets, then for ordinary generating functions, and finally in the case of exponential generating functions. I have also included some recent material from the research literature which, to my knowledge, has not appeared in book form previously, such as the theory of quotient posets and the connection between pattern avoidance and quasisymmetric functions.

Most of the exercises should be doable with a reasonable amount of effort. A few unsolved conjectures have been included among the problems in the hope that an interested student might wish to tackle one of them. They are, of course, marked as such.

A few words about the title are in order. It is in part meant to be a tip of the hat to Donald Knuth's influential series of books *The art of computer programing*, Volumes 1–3 [**51–53**], which, among many other things, helped give birth to the study of pattern avoidance through its connection with stack sorting; see Exercise 36 in Chapter 1. I hope that the title also conveys some of the beauty found in this area of mathematics, for example, the elegance of the Hook Formula (equation (7.10)) for the number of standard Young tableaux. In addition I should mention that, due to my own preferences, this book concentrates on the enumerative side of combinatorics and mostly ignores the important extremal and existential parts of the field. The reader interested in these areas can consult the books of Flajolet and Sedgewick [**25**] and of van Lint [**95**].

This book grew out of the lecture notes which I have compiled over years of teaching the graduate combinatorics course at Michigan State University. I would like to thank the students in these classes for all the feedback they have given me about the various topics and their presentation. I am also indebted to the following colleagues, some of whom taught from a preliminary version of this book, who provided me with suggestions as well as catching numerous typographical errors: Matthias Beck, Moussa Benoumhani, Andreas Blass, Seth Chaiken, Sylvie Corteel, Georges Grekos, Richard Hensh, Nadia Lafrenière, Duncan Levear, and Tom Zaslavsky. Darij Grinberg deserves special mention for providing copious comments and corrections as well as providing a number of interesting exercises. I also received valuable feedback from four anonymous referees. Finally, I wish to express my appreciation of Ina Mette, my editor at the American Mathematical Society. Without her gentle support and persistence, this text would never have seen the light of day. Because I typeset this document myself, all errors can be blamed on my computer.

East Lansing, Michigan, 2020

List of Notation

Symbol	Definition	Page
$A(D)$	arc set of digraph D	21
$A(G)$	adjacency matrix of graph G	60
$\mathcal{A}(G)$	set of acyclic orientations of G	103
$a(G)$	number of acyclic orientations of G	103
$A([n], k)$	set of permutations π in $\mathfrak{S}_n$ having k descents	121
$A(n, k)$	Eulerian number, cardinality of $A([n], k)$	121
$A_n(q)$	Eulerian polynomial	122
$\mathcal{A}(P)$	atom set of poset P	169
$\operatorname{Asc} c$	ascent set of a proper coloring c	279
$\operatorname{asc} c$	ascent number of a proper coloring c	279
$\operatorname{Asc} \pi$	ascent set of permutation π	76
$\operatorname{asc} \pi$	ascent number of permutation π	76
$\operatorname{Av}_n(\pi)$	the set of permutations in $\mathfrak{S}_n$ avoiding π	29
α^r	reversal of composition α	32
$\bar{\alpha}$	expansion of composition α	274
$\alpha(C)$	rank composition of chain C	275
$B(G)$	incidence matrix of graph G	61
$B(T)$	set of partitions of the set T	10
B_n	Boolean algebra on $[n]$	140
B_∞	poset of subsets of $\mathbb{P}$	178
$B(n)$	nth Bell number	10

Symbol	Definition	Page
$\mathbb{C}$	complex numbers	1
$c_i(g)$	number of cycles of length i in group element g	197
CL_n	claw poset with n atoms	169
$\operatorname{co} T$	content of tableau T	225
C_n	cycle with n vertices	19
C_n	chain poset of length n	139
$c_x(P)$	column insertion of element x into tableau P	245
C_∞	chain poset on $\mathbb{N}$	178
$C(n)$	Catalan number	26
$c([n],k)$	set of permutations in $\mathfrak{S}_n$ with k cycles	12
$c(n,k)$	signless Stirling number of the first kind	12
$c_o(L,k)$	ordered k cycle decompositions of permutations of L	127
$\mathbb{C}X$	vector space generated by set X over $\mathbb{C}$	248
$\mathbb{C}[x]$	polynomial algebra in x over $\mathbb{C}$	71
$\mathbb{C}[[x]]$	formal power series algebra in x over $\mathbb{C}$	81
$\mathcal{C}(\pi)$	set of functions compatible with π	236
$\mathcal{C}_m(\pi)$	set of functions compatible with π bounded by m	236
$\operatorname{Des} P$	descent set of tableau P	271
$\operatorname{Des} \pi$	descent set of permutation π	75
$\operatorname{des} \pi$	descent number of permutation π	76
D_n	lattice of divisors of n	140
D_∞	divisibility poset on $\mathbb{P}$	181
$D(n)$	derangement number	43
$\mathcal{D}(n)$	set of Dyck paths of semilength n	26
$\mathcal{D}(V)$	set of all digraphs on vertex set V	21
$\mathcal{D}(V,k)$	set of all digraphs on vertex set V with k edges	21
$\deg m$	degree of a monomial	219
$\deg v$	degree of vertex v in a graph	20
$\Delta f(n)$	forward difference operator of $f(n)$	162
$\delta_{x,y}$	Kronecker delta	7
$\delta(x,z)$	delta function of poset incidence algebra	159
$E(G)$	edge set of graph G	18
$E(L)$	set structure on label set L	125
$\overline{E}(L)$	nonempty set structure on label set L	125
E_n	Euler number	120
e_n	nth elementary symmetric function	221
$E(t)$	generating function for elementary symmetric functions	221
$\operatorname{Exc} \pi$	set of excedances of permutation π	122
$\operatorname{exc} \pi$	number of excedances of permutation π	122

Symbol	Definition	Page
$\operatorname{Fix} f$	fix point set of a function f	44
f_n	Fibonacci number	3
F_n	Fibonacci number	2
$\mathbb{F}_q$	Galois field with q elements	79
$f(x)$	ordinary generating function	81
$f_S(x)$	weight-generating function for weighted set S	86
$F(n)$	binomial poset n-interval factorial function	178
$F(x)$	exponential generating function	117
$F_{\mathcal{S}}(x)$	exponential generating function for structure $\mathcal{S}$	125
F_S	fundamental quasisymmetric for set S	269
F_α	fundamental quasisymmetric for composition α	269
f^λ	number of standard Young tableaux of shape λ	225
Φ	fundamental map on permutations	122
ϕ	bijection between subsets and compositions	16
$G \setminus e$	graph G with edge e deleted	100
G/e	graph G with edge e contracted	101
$\mathrm{GL}(V)$	general linear group over vector space V	287
$\mathcal{G}(V)$	set of all graphs on vertex set V	20
$\mathcal{G}(V,k)$	set of all graphs on vertex set V with k edges	20
G_x	stabilizer of element x under the action of group G	191
$H_c = H_{i,j}$	hook of cell $c = (i,j)$	230
$h_c = h_{i,j}$	hooklength of cell $c = (i,j)$	230
$\mathcal{H}_n$	set of hook diagrams with n cells	278
h_n	nth complete homogeneous symmetric function	221
$H(t)$	complete homogeneous generating function	221
$\operatorname{ideg} v$	in-degree of vertex v in a digraph	21
$\operatorname{Inv} \pi$	inversion set of permutation π	74
$\operatorname{inv} \pi$	inversion number of permutation π	74
$\mathcal{I}(P)$	incidence algebra of poset P	158
$I(S)$	lower-order ideal generated by S in a poset	143
$\mathrm{ISF}(G;t)$	increasing spanning forest generating function of G	105
$\mathrm{ISF}_m(G)$	set of m-edge increasing spanning forests of G	105
$\mathrm{isf}_m(G)$	number of m-edge increasing spanning forests of G	105
$i_\lambda(G)$	number of independent type λ partitions in graph G	254
$\mathcal{J}(P)$	distributive lattice of lower-order ideals of poset P	151
K_n	complete graph with n vertices	19
K_n	lattice of compositions of n	140
$K_{\lambda,\mu}$	number of tableaux of shape λ and content μ	225

Symbol	Definition	Page
$L(G)$	Laplacian of graph G	62
$\mathcal{L}(G)$	bond lattice of graph G	167
$\mathcal{L}(P)$	set of linear extensions of P	238
$\ell(C)$	length of chain C in a poset	147
$\ell(\lambda)$	length of an integer partition λ	15
$\ell(\pi)$	length of a permutation π	4
$\lim_{k\to\infty} f_k(x)$	limit of a sequence of formal power series	84
$\operatorname{lds}\pi$	length of a longest decreasing subsequence of π	245
$\operatorname{lis}\pi$	length of a longest increasing subsequence of π	244
$L_n(q)$	lattice of subspaces of $\mathbb{F}_q^n$	140
$L_\infty(q)$	poset of subspaces of vector space V_∞ over $\mathbb{F}_q$	178
$L(V)$	lattice of subspaces of V	140
$\lambda(F)$	type of partition induced by edge set F	255
$\lambda^!$	multiplicity factorial of partition λ	254
$\operatorname{maj}\pi$	major index of permutation π	76
$M(n)$	Mertens function	183
$M(P)$	monomial quasisymmeric function for poset P	275
M_α	monomial quasisymmetric function	268
m_λ	monomial symmetric function	220
$\mu(P)$	Möbius function value on a poset P	154
$\mu(x)$	one-variable Möbius function evaluated at x	154
$\mu(x,z)$	two-variable Möbius function on the interval $[x,z]$	157
$\mathbb{N}$	nonnegative integers	1
$\mathrm{NBC}_k(G)$	set of no broken circuit sets of k edges of G	102
$\mathrm{nbc}_k(G)$	number of no broken circuit sets of k edges of G	102
$\mathcal{NE}(m,n)$	set of N-E lattice paths from $(0,0)$ to (m,n)	26
$\operatorname{odeg} v$	out-degree of vertex v in a digraph	21
$\mathcal{O}_x$	orbit of an element x under action of a group	190
$O(g)$	big oh notation applied to function g	182
$o(g)$	order of a group element g	210
$\mathbb{P}$	positive integers	1
P^*	dual of poset P	142
$\mathcal{PC}(G)$	set of proper colorings of G with the positive integers	279
$P(G;t)$	chromatic polynomial of graph G	100
$\operatorname{Par} P$	set of P-partitions	238
$\operatorname{Par}_m P$	set of P-partitions bounded by m	238
P_n	path with n vertices	19
$P(n)$	set of partitions of the integer n	13
$p(n)$	number of partitions of the integer n	13
p_n	nth power sum symmetric function	221
$P(t)$	power sum symmetric generating function	221

Symbol	Definition	Page
$P(n,k)$	set of partitions of n into at most k parts	15
$p(n,k)$	number of partitions of n into at most k parts	15
$P(S)$	permutations of a set S	4
$P(S,k)$	permutations of length k of a set S	4
$P((S,k))$	words of length k over a set S	5
$P(\pi)$	insertion tableau of π	242
$\mathcal{P}(u;v)$	set of directed paths from u to v in a digraph	56
Π_n	partition lattice on $[n]$	140
$\Pi(\mathcal{S})$	partition structure on structure $\mathcal{S}$	131
$\Pi_e(\mathcal{S})$	even partition structure on structure $\mathcal{S}$	133
$\Pi_o(\mathcal{S})$	odd partition structure on structure $\mathcal{S}$	133
$\mathbb{Q}$	rational numbers	1
$Q(n)$	set of compositions of the integer n	16
$q(n)$	number of compositions of the integer n	16
$Q(n,k)$	set of compositions of n into k parts	16
$q(n,k)$	number of partitions of n into k parts	16
QSym	algebra of quasisymmetric functions	268
QSym_n	quasisymmetric functions of degree n	268
$Q(\pi)$	recording tableau of π	242
$Q_n(\Pi)$	quasisymmetric function for patterns Π	277
$\mathbb{R}$	real numbers	1
$\mathcal{R}C(\pi)$	set of functions reverse compatible with π	270
$\operatorname{rk} P$	rank of a ranked poset P	147
$\mathrm{Rk}_k\, P$	kth rank set of a ranked poset P	147
$\operatorname{rk} x$	rank of an element x in a ranked poset	147
$\mathcal{R}(k,l)$	set of partitions contained in a $k \times l$ rectangle	79
$\mathrm{RPar}\, P$	set of reverse P-partitions	271
$\mathcal{R}(P)$	reduced incidence algebra of a binomial poset	179
$\mathrm{rpp}_n(\lambda)$	number of shape λ reverse plane partitions of n	233
$\mathrm{rpar}(P;\mathbf{x})$	generating function for reverse P-partitions	271
$r_x(P)$	row insertion of element x into tableau P	241
$\rho(F)$	vertex partition induced by edge set F	255
$\rho:\ G \to \mathrm{GL}(V)$	representation of group G	287
$\mathcal{S}(L)$	labeled structure on label set L	124
$\mathfrak{S}$	pattern poset	140
$\mathfrak{S}_n$	symmetric group on $[n]$	11
$Sf(n)$	summation operator applied to function $f(n)$	162
sgn	sign function on a signed set	44
$\operatorname{sh} T$	shape of tableau T	225
$s(n,k)$	signed Stirling number of the first kind	13

Symbol	Definition	Page
$S(T,k)$	set of partitions of the set T into k blocks	10
$S(n,k)$	Stirling number of the second kind	10
$S_o(L,k)$	set of ordered partitions of the set L into k blocks	127
$\mathcal{ST}(G)$	set of spanning trees of graph G	59
st	statistic on a set	74
$\operatorname{std}\sigma$	standardization of the permutation σ	28
$\operatorname{Supp} x$	support set of x in a product of claws	173
$\operatorname{supp} x$	size of support set of x in a product of claws	173
Sym	algebra of symmetric functions	220
Sym_n	symmetric functions of degree n	220
$\mathrm{SYT}(\lambda)$	set of standard Young tableaux of shape λ	224
$\mathrm{SSYT}(\lambda)$	set of semistandard Young tableaux of shape λ	225
s_λ	Schur function	225
$T_{i,j}$	element in cell (i,j) of tableau T	225
$\mathcal{T}_n$	set of monomino-domino tilings of a row of n squares	3
$U(S)$	upper-order ideal generated by S in a poset	143
$V(D)$	vertex set of digraph D	21
$V(G)$	vertex set of graph G	18
V_∞	vector space with a countably infinite basis over $\mathbb{F}_q$	178
$w_k(P)$	Whitney number of the first kind for a poset P	156
$W_k(P)$	Whitney number of the second kind for a poset P	156
W_n	walk with n vertices	19
wt	weight function on a set	86
$\mathbf{x}$	a countably infinite set of variables	219
$\mathbf{x}^c$	monomial for a coloring c of a graph	253
$\mathbf{x}^f$	monomial for a function f	270
$\mathbf{x}^T$	monomial for a tableau T	225
X^g	fixed points of group element g acting on set X	192
$X(G;\mathbf{x})$	chromatic symmetric function of graph G	253
$X(G;\mathbf{x},q)$	chromatic quasisymmetric function of graph G	280
Y	Young's lattice	140
$\mathbb{Z}$	set of integers	1
$\zeta(x,z)$	zeta function in the incidence algebra of a poset	159
$\zeta(s)$	Riemann zeta function	182
$z(g)$	cycle index of group element g	197
$Z(G)$	cycle index of group G	197
$\#S$	cardinality of the set S	1
$\lvert f\rvert$	size (sum of values) of a function	236
$\lvert S\rvert$	cardinality of the set S	1
$\lvert T\rvert$	sum of entries of tableau T	233
$S \uplus T$	disjoint union of sets S and T	1

Symbol	Definition	Page
$\|\lambda\|$	sum of the parts of partition λ	13
$\lambda \vdash n$	λ is a partition of n	13
$S \times T$	(Cartesian) product of sets S and T	1
$P \uplus Q$	disjoint union of posets P and Q	145
$P \oplus Q$	ordinal sum of posets P and Q	146
$P \times Q$	(Cartesian) product of posets P and Q	146
$[g]$	linear transformation for group element g	287
$[g]_B$	matrix in basis B for group element g	287
$[n]$	set of integers $\{1, 2, \ldots, n\}$	4
$[n]_q$	q-analogue of nonnegative integer n	75
$[n]_q!$	q-analogue of $n!$	75
$[x^n]f(x)$	coefficient of x^n in $f(x)$	83
$n{\downarrow_k}$	n falling factorial with k factors	4
2^S	set of subsets of S	5
$\binom{S}{k}$	set of k-element subsets of S	6
$\binom{n}{k}$	binomial coefficient	7
$\begin{bmatrix} n \\ k \end{bmatrix}_q$	q-binomial coefficient	77
$\begin{bmatrix} V \\ k \end{bmatrix}$	k-dimensional subspaces of vector space V	79
$\{\{a, a, \ldots\}\}$	multiset individual element notation	8
$\{\{a^2, \ldots\}\}$	multiset multiplicity notation	8
$\left(\binom{S}{k}\right)$	set of k-element multisubsets of S	9
$\chi(G)$	chromatic number of G	99
$\chi(g)$	character of group element g	291
$x \lessdot y$	x is covered by y in a poset	140
$y \gtrdot x$	y covers x in a poset	140
$\hat{0}$	the minimum element of a poset	142
$\hat{1}$	the maximum element of a poset	142
$[x, y]$	closed interval from x to y in a poset	143
$x \wedge y$	meet of x and y in a poset	148
$\bigwedge X$	meet of the subset X in a poset	149
$x \vee y$	join of x and y in a poset	149
$U + V$	sum of subspaces U and V	149
$f * g$	convolution of f and g in the incidence algebra	158
$\chi(P; t)$	characteristic polynomial of a ranked poset P	164
$P/\sim$	quotient of poset P by equivalence relation $\sim$	169
ω_n	primitive nth root of unity	210
$\pi \overset{\mathrm{RS}}{\mapsto} (P, Q)$	Robinson–Schensted map	242
$M \overset{\mathrm{RSK}}{\mapsto} (T, U)$	Robinson–Schensted–Knuth map	244

Chapter 1

Basic Counting

In this chapter we will develop the most elementary techniques for enumerating sets. Even though these methods are relatively basic, they will presage more complicated things to come. We denote the integers by $\mathbb{Z}$ and parameters such as n and k are always assumed to be integral unless otherwise indicated. We also use the notation $\mathbb{N}$ and $\mathbb{P}$ for the nonnegative and positive integers, respectively. As usual, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ stand for the rational numbers, real numbers, and complex numbers, respectively. Finally, whenever taking the cardinality of a set we will assume it is finite.

1.1. The Sum and Product Rules for sets

The Sum and Product Rules for sets are the basis for much of enumeration. And we will see various extensions of them later to ordinary and exponential generating functions. Although the rules are very easy to prove, we will include the demonstrations because the results are so useful. Given a finite set S, we will use either of the notations $\#S$ or $|S|$ for its cardinality. We will also write $S \uplus T$ for the disjoint union of S and T, and usage of this symbol implies disjointness even if it has not been previously explicitly stated. Finally, our notation for the (Cartesian) product of sets is

$$S \times T = \{(s,t) \mid s \in S, t \in T\}.$$

Lemma 1.1.1. *Let S, T be finite sets.*

(a) (*Sum Rule*) *If $S \cap T = \emptyset$, then*

$$|S \uplus T| = |S| + |T|.$$

(b) (*Product Rule*) *For any finite sets*

$$|S \times T| = |S| \cdot |T|.$$

Proof. Let $S = \{s_1, \ldots, s_m\}$ and $T = \{t_1, \ldots, t_n\}$. For part (a), if S and T are disjoint, then we have $S \uplus T = \{s_1, \ldots, s_m, t_1, \ldots, t_n\}$ so that $|S \uplus T| = m + n = |S| + |T|$.

For part (b), we induct on $n = |T|$. If $T = \emptyset$, then $S \times T = \emptyset$ so that $|S \times T| = 0$ as desired. If $|T| \geq 1$, then let $T' = T - \{t_n\}$. We can write $S \times T = (S \times T') \uplus (S \times \{t_n\})$. Also $S \times \{t_n\} = \{(s_1, t_n), \ldots, (s_m, t_n)\}$, which has $|S| = m$ elements since the second component is constant. Now by part (a) and induction

$$|S \times T| = |S \times T'| + |S \times \{t_n\}| = m(n-1) + m = mn,$$

which finishes the proof. □

In combinatorial choice problems, one is often given either the option to do one operation or another, or to do both. Suppose there are m ways of doing the first operation and n ways of doing the second. If there is no common operation, then the Sum Rule tells us that the number of ways to do one or the other is $m + n$. And if doing the first operation has no effect on doing the second, then the Product Rule gives a count of mn for doing the first and then the second. More generally if there are m ways of doing the first operation and, no matter which of the m is chosen, the number of ways to continue with the second operation is n, then again there are mn ways to do both. (The actual n second operations available may depend on the choice of the first, but not their number.) So in practice one translates from English to mathematics by replacing "or" with addition and "and" with multiplication.

Another important concept related to cardinalities is that of a bijection. A *bijection* between sets S, T is a function $f : S \to T$ which is both injective (one-to-one) and surjective (onto). If S, T are finite, then the existence of a bijection between them implies that $|S| = |T|$. (One can extend this notion to infinite sets, but we will have no cause to do so here.) In combinatorics, one often uses bijections to prove that two sets have the same cardinality. See, for just one of many examples, the proof of Theorem 1.1.2 below.

We will illustrate these ideas with one of the most famous sequences in all of combinatorics: the Fibonacci numbers. As is sometimes the case, there is an amusing (if somewhat improbable) story attached to the sequence. One starts at the beginning of time with a pair of immature rabbits, one male and one female. It takes one month for rabbits to mature. In every subsequent month a pair gives birth to another pair of immature rabbits, one male and one female. If rabbits only breed with their birth partner and live forever (as I said, the story is somewhat improbable), how many pairs of rabbits are there at the beginning of month n? Let us call this number F_n. It will be convenient to let $F_0 = 0$. Since we begin with only one pair, $F_1 = 1$. And at the beginning of the second month, the pair has matured but produced no offspring, so $F_2 = 1$. In subsequent months, one has all the rabbits from the previous month, counted by F_{n-1}, together with the newborn pairs. The number of newborn pairs equals the number of mature pairs from the previous month, which equals the total number of pairs from the month before which is F_{n-2}. Thus, applying the Sum Rule,

$$F_n = F_{n-1} + F_{n-2} \text{ for } n \geq 2 \text{ with } F_0 = 0 \text{ and } F_1 = 1 \tag{1.1}$$

where we can start the recursion at $n = 2$ rather than $n = 3$ due to letting $F_0 = 0$. The F_n are called the *Fibonacci numbers*. It is also important to note that some authors

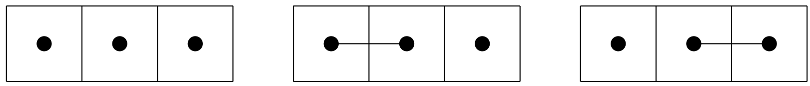


Figure 1.1. $\mathcal{T}_3$

define this sequence by letting

$$f_0 = f_1 = 1 \text{ and } f_n = f_{n-1} + f_{n-2} \text{ for } n \geq 2. \tag{1.2}$$

So it is important to make sure which flavor of Fibonacci is being discussed in a given context.

One might wonder if there is an explicit formula for F_n in addition to the recursive one above. We will see that such an expression exists, although it is far from obvious how to derive it from what we have done so far. Indeed, we will need the theory of ordinary generating functions discussed in Chapter 3 to derive it.

Another thing which might be desired is a combinatorial interpretation for F_n. A *combinatorial interpretation* for a sequence of nonnegative integers $a_0, a_1, a_2, \ldots$ is a sequence of sets $S_0, S_1, S_2, \ldots$ such that $\#S_n = a_n$ for all n. Such interpretations often give rise to very pretty and intuitive proofs about the original sequence and so are highly desirable. One could argue that the story of the rabbits already gives such an interpretation. But we would like something more amenable to mathematical manipulation.

Suppose we are given a row of squares. We are also given two types of tiles: dominos which can cover two squares and monominos which can cover one. A *tiling* of the row is a set of tiles which covers each square exactly once. Let $\mathcal{T}_n$ be the set of tilings of a row of n squares. See Figure 1.1 for a list of the elements of $\mathcal{T}_3$. There is a simple relationship between tilings and Fibonacci numbers.

Theorem 1.1.2. *For $n \geq 1$ we have*

$$F_n = \#\mathcal{T}_{n-1}.$$

Proof. It suffices to prove that both sides of this equation satisfy the same initial conditions and recurrence relation. When the row contains no squares, it only has the empty tiling so $\mathcal{T}_0 = 1 = F_1$. And when there is one square, it can only be tiled by a monomino so $\mathcal{T}_1 = 1 = F_2$. For the recursion, the tilings in $\mathcal{T}_n$ can be divided into two types: those which end with a monomino and those which end with a domino. Removing the last tile shows that these tilings are in bijection with those in $\mathcal{T}_{n-1}$ and those in $\mathcal{T}_{n-2}$, respectively. Thus $\#\mathcal{T}_n = \#\mathcal{T}_{n-1} + \#\mathcal{T}_{n-2}$ as desired. □

To see the power of a good combinatorial interpretation, we will now give a simple proof of an identity for the F_n. Such identities are legion. See, for example, the book of Benjamin and Quinn [**10**].

Corollary 1.1.3. *For $m \geq 1$ and $n \geq 0$ we have*

$$F_{m+n} = F_{m-1}F_n + F_m F_{n+1}.$$

Proof. By the previous theorem, the left-hand side counts the number of tilings of a row of $m+n-1$ squares. So it suffices to show that the same is true of the right. Label the squares $1, \ldots, m+n-1$ from left to right. We can write $\mathcal{T}_{m+n-1} = \mathcal{S} \uplus \mathcal{T}$ where $\mathcal{S}$ contains those tilings with a domino covering squares $m-1$ and m, and $\mathcal{T}$ has the tilings with $m-1$ and m in different tiles. The tilings in $\mathcal{T}$ are essentially pairs of tilings, the first covering the first $m-1$ square and second covering the last n squares. So the Product Rule gives $|\mathcal{T}| = |\mathcal{T}_{m-1}| \cdot |\mathcal{T}_n| = F_m F_{n+1}$. Removing the given domino from the tilings in $\mathcal{S}$ again splits each tiling into a pair with the first covering $m-2$ squares and the second $n-1$. Taking cardinalities results in $|\mathcal{S}| = F_{m-1}F_n$. Finally, applying the Sum Rule finishes the proof. □

The demonstration just given is called a *combinatorial proof* since it involves counting discrete objects. We will meet other useful proof techniques as we go along. But combinatorial proofs are often considered to be the most pleasant, in part because they can be more illuminating than demonstrations just involving formal manipulations.

1.2. Permutations and words

It is always important when considering an enumeration problem to determine whether the objects being considered are ordered or not. In this section we will consider the most basic ordered structures, namely permutations and words.

If S is a set with $\#S = n$, then a *permutation of* S is a sequence $\pi = \pi_1 \ldots \pi_n$ obtained by listing the elements of S in some order. If π is a permutation, we will always use π_i to denote the ith element of π and similarly for other ordered structures. We let $P(S)$ denote the set of all permutations of S. For example,

$$P(\{a, b, c\}) = \{abc,\ acb,\ bac,\ bca,\ cab,\ cba\}.$$

Clearly $\#P(S)$ only depends on $\#S$. So often we choose the canonical n-element set

$$[n] = \{1, 2, \ldots, n\}.$$

We can also consider *k-permutations of S* which are sequences $\pi = \pi_1 \ldots \pi_k$ obtained by linearly ordering k distinct elements of S. Here, k is called the *length* of the permutation and we write $\ell(\pi) = k$. Again, we use the same terminology and notation for other ordered structures. The set of all k-permutations of S is denoted $P(S, k)$. By way of illustration,

$$P(\{a, b, c, d\}, 2) = \{ab,\ ba,\ ac,\ ca,\ ad,\ da,\ bc,\ cb,\ bd,\ db,\ cd,\ dc\}.$$

In particular, if $\#S = n$, then $P(S, n) = P(S)$. Also $P(S, k) = \emptyset$ for $k > n$ since in this case it is impossible to pick k distinct elements from a set with only n. And $P(S, 0) = \{\epsilon\}$ where ϵ is the empty sequence.

To count permutations it will be convenient to introduce the following notation. Given nonnegative integers n, k, we can form the *falling factorial*

$$n\downarrow_k = n(n-1)\ldots(n-k+1).$$

Note that k equals the number of factors in the product.

Theorem 1.2.1. *For $n, k \geq 0$ we have*

$$\#P([n], k) = n\downarrow_k .$$

In particular

$$\#P([n]) = n!\,.$$

Proof. Since $P([n]) = P([n], n)$, it suffices to prove the first formula. Given $\pi = \pi_1 \ldots \pi_k \in P([n], k)$, there are n ways to pick π_1. Since $\pi_2 \neq \pi_1$, there remains $n-1$ choices for π_2. Since the number of choices for π_2 does not depend on the actual element chosen for π, one can continue in this way and apply a modified version of the Product Rule to obtain the result. □

Note that when $0 \leq k \leq n$ we can write

$$n\downarrow_k = \frac{n!}{(n-k)!}. \tag{1.3}$$

But for $k > n$ the product $n\downarrow_k$ still makes sense, even though the product cannot be expressed as a quotient of factorials. Indeed, if $k > n$, then zero is a factor and so $n\downarrow_k = 0$, which agrees with the fact that $P([n], k) = \emptyset$. In the special case $k = 0$ we have $n\downarrow_k = 1$ because it is an empty product. Again, this reflects the combinatorics in that $P([n], 0) = \{\epsilon\}$.

One of the other things to keep track of in a combinatorial problem is whether elements are allowed to be repeated or not. In permutations we have no repetitions. But the case when they are allowed is interesting as well. A *k-word* over a set S is a sequence $w = w_1 \ldots w_k$ where $w_i \in S$ for all i. Note that there is no assumption that the w_i are distinct. We denote the set of k-words over S by $P((S, k))$. Note the use of the double parentheses to denote the fact that repetitions are allowed. Note also that $P(S, k) \subseteq P((S, k))$, but usually the inclusion is strict. To illustrate

$$P((\{a, b, c, d\}, 2)) = P(\{a, b, c, d\}, 2) \uplus \{aa, bb, cc, dd\}.$$

The proof of the next result is almost identical to that of Theorem 1.2.1 and so is left to the reader. When a result is given without proof, this is indicated by a box at the end of its statement.

Theorem 1.2.2. *For $n, k \geq 0$ we have*

$$\#P(([n], k)) = n^k.$$ □

1.3. Combinations and subsets

We will now consider unordered versions of the combinatorial objects studied in the last section. These are sometimes called combinations, although the reader may know them by their more familiar name: subsets.

Given a set S, we let 2^S denote the set of all subsets of S. Notice that 2^S is a set, not a number. For example,

$$2^{\{a,b,c\}} = \{\emptyset,\ \{a\},\ \{b\},\ \{c\},\ \{a, b\},\ \{a, c\},\ \{b, c\},\ \{a, b, c\}\}.$$

The reason for this notation should be made clear by the following result.

Theorem 1.3.1. *For $n \geq 0$ we have*

$$\#2^{[n]} = 2^n.$$

Proof. By Theorem 1.2.2 we have $2^n = \#P((\{0,1\}, n))$. So it suffices to find a bijection

$$f: 2^{[n]} \to P((\{0,1\}, n)),$$

and there is a canonical one. In particular, if $S \subseteq [n]$, then we let $f(S) = w_1 \dots w_n$ where, for all i,

$$w_i = \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{if } i \notin S. \end{cases}$$

To show that f is bijective, it suffices to find its inverse. If $w = w_1 \dots w_n \in P((\{0,1\}, n))$, then we let $f^{-1}(w) = S$ where $i \in S$ if $w_i = 1$ and $i \notin S$ if $w_i = 0$ where $1 \leq i \leq n$. It is easy to check that the compositions $f \circ f^{-1}$ and $f^{-1} \circ f$ are the identity maps on their respective domains. This completes the proof. □

The proof just given is called a *bijective proof* and it is a particularly nice kind of combinatorial proof. This is because bijective proofs can relate different types of combinatorial objects, sometimes revealing unexpected connections. Also note that we proved f bijective by finding its inverse rather than showing directly that it was one-to-one and onto. This is the preferred method as having a concrete description of f^{-1} can be useful later. Finally, when dealing with functions we will always compose them right-to-left so that

$$(f \circ g)(x) = f(g(x)).$$

We now want to count subsets by their cardinality. For a set S we will use the notation

$$\binom{S}{k} = \{T \subseteq S \mid \#T = k\}.$$

As an example,

$$\binom{\{a,b,c\}}{2} = \{\{a,b\},\ \{a,c\},\ \{b,c\}\}.$$

As expected, we now find the cardinality of this set.

Theorem 1.3.2. *For $n, k \geq 0$ we have*

$$\#\binom{[n]}{k} = \frac{n\downarrow_k}{k!}.$$

Proof. Cross-multiplying and using Theorem 1.2.1 we see that it suffices to prove

$$\#P([n], k) = k! \cdot \#\binom{[n]}{k}.$$

To see this, note that we can get each $\pi_1 \dots \pi_k \in P([n], k)$ exactly once by running through the subsets $S = \{s_1, \dots, s_k\} \subseteq [n]$ and then ordering each S in all possible ways. The number of choices for S is $\#\binom{[n]}{k}$ and, by Theorem 1.2.1 again, the number of ways of permuting the elements of S is $k!$. So we are done by the Product Rule. □

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}$$

Figure 1.2. Rows 0 through 4 of Pascal's triangle

Given $n, k \geq 0$, we define the *binomial coefficient*

$$\binom{n}{k} = \#\binom{[n]}{k} = \frac{n\downarrow_k}{k!}. \tag{1.4}$$

The reason for this name is that these numbers appear in the binomial expansion which will be studied in Chapter 3. Often you will see the binomial coefficients displayed in a triangular array called *Pascal's triangle* which has $\binom{n}{k}$ as the entry in the nth row and kth diagonal. When $k > n$ it is traditional to omit the zeros. See Figure 1.2 for rows 0 through 4. (We apologize to the reader for not writing out the whole triangle, but this page is not big enough.) For $0 \leq k \leq n$ we can use (1.3) to write

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}, \tag{1.5}$$

which is pleasing because of its symmetry. We can also extend the binomial coefficients to $k < 0$ by letting $\binom{n}{k} = 0$. This is in keeping with the fact that $\binom{[n]}{k} = \emptyset$ in this case.

In the next theorem, we collect various basic results about binomial coefficients which will be useful in the sequel. In it, we will use the *Kronecker delta function* defined by

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{if } x \neq y. \end{cases}$$

Also note that we do not specify the range of the summation variable k in (c) and (d) because it can be taken as either $0 \leq k \leq n$ or $k \in \mathbb{Z}$ since the extra terms in the larger sum are all zero. Both viewpoints will be useful on occasion.

Theorem 1.3.3. *Suppose $n \geq 0$.*

(a) *The binomial coefficients satisfy the initial condition*

$$\binom{0}{k} = \delta_{k,0}$$

and recurrence relation

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

for $n \geq 1$.

(b) *The binomial coefficients are symmetric, meaning that*

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

(c) *We have*

$$\sum_k \binom{n}{k} = 2^n.$$

(d) *We have*

$$\sum_k (-1)^k \binom{n}{k} = \delta_{n,0}.$$

Proof. (a) The initial condition is clear. For the recursion let $\mathcal{S}_1$ be the set of $S \in \binom{[n]}{k}$ with $n \in S$, and let $\mathcal{S}_2$ be the set of $S \in \binom{[n]}{k}$ with $n \notin S$. Then $\binom{[n]}{k} = \mathcal{S}_1 \uplus \mathcal{S}_2$. But if $n \in S$, then $S - \{n\} \in \binom{[n-1]}{k-1}$. This gives a bijection between $\mathcal{S}_1$ and $\binom{[n-1]}{k-1}$ so that $\#\mathcal{S}_1 = \binom{n-1}{k-1}$. On the other hand, if $n \notin S$, then $S \in \binom{[n-1]}{k}$ and this implies $\#\mathcal{S}_2 = \binom{n-1}{k}$. Applying the Sum Rule completes the proof.

(b) It suffices to find a bijection $f : \binom{[n]}{k} \to \binom{[n]}{n-k}$. Consider the map $f : 2^{[n]} \to 2^{[n]}$ by $f(S) = [n] - S$ where the minus sign indicates difference of sets. Note that the composition f^2 is the identity map so that f is a bijection. Furthermore $S \in \binom{[n]}{k}$ if and only if $f(S) \in \binom{[n]}{n-k}$. So f restricts to a bijection between these two sets.

(c) This follows by applying the Sum Rule to the equation $2^{[n]} = \biguplus_k \binom{[n]}{k}$.

(d) The case $n = 0$ is easy, so we assume $n > 0$. We will learn general techniques for dealing with equations involving signs in the next chapter. But for now, we try to prove the equivalent equality

$$\sum_{k \text{ odd}} \binom{n}{k} = \sum_{k \text{ even}} \binom{n}{k}.$$

Let $\mathcal{T}_1$ be the set of $T \in 2^{[n]}$ with $\#T$ odd and let $\mathcal{T}_2$ be the set of $T \in 2^{[n]}$ with $\#T$ even. We wish to find a bijection $g : \mathcal{T}_1 \to \mathcal{T}_2$. Consider the operation of *symmetric difference*

$$S \,\Delta\, T = (S - T) \uplus (T - S).$$

It is not hard to see that $(S \,\Delta\, T) \,\Delta\, T = S$. Now define $g : 2^{[n]} \to 2^{[n]}$ by $g(T) = T \,\Delta\, \{n\}$ so that, by the previous sentence, g^2 is the identity. Furthermore, g reverses parity and so restricts to the desired bijection. □

As with the case of permutations and words, we want to enumerate "sets" where repetitions are allowed. A *multiset M* is an unordered collection of elements which may be repeated. For example

$$M = \{\{a, a, a, b, c, c\}\} = \{\{c, a, b, a, c, a\}\}.$$

Note the use of double curly brackets to denote a multiset. We will also use multiplicity notation where a^m denotes m copies of the element a. Continuing our example

$$M = \{\{a^3, b, c^2\}\}.$$

As with powers, an exponent of one is optional and an exponent of zero indicates that there are no copies of that element in the multiset. The *cardinality* of a multiset is its number of elements counted with multiplicity. So in our example $\#M = 2 + 1 + 3 = 6$.

If S is a set, then M is a *multiset on S* if every element of M is an element of S. We let $\left(\!\binom{S}{k}\!\right)$ be the set of all multisets on S of cardinality k and

$$\left(\!\!\binom{n}{k}\!\!\right) = \#\left(\!\!\binom{[n]}{k}\!\!\right).$$

To illustrate

$$\left(\!\!\binom{\{a,b,c\}}{2}\!\!\right) = \{\,\{\{a,a\}\},\ \{\{a,b\}\},\ \{\{a,c\}\},\ \{\{b,b\}\},\ \{\{b,c\}\},\ \{\{c,c\}\}\,\}$$

and so $\left(\!\binom{3}{2}\!\right) = 6$.

Theorem 1.3.4. *For $n, k \geq 0$ we have*

$$\left(\!\!\binom{n}{k}\!\!\right) = \binom{n+k-1}{k}.$$

Proof. We wish to find a bijection

$$f : \left(\!\!\binom{[n]}{k}\!\!\right) \to \binom{[n+k-1]}{k}.$$

Given a multiset $M = \{\{m_1 \leq m_2 \leq m_3 \leq \cdots \leq m_k\}\}$ on $[n]$, let

$$f(M) = \{m_1 < m_2 + 1 < m_3 + 2 < \cdots < m_k + k - 1\}.$$

Now the $m_i + i - 1$ are distinct, and the fact that $m_k \leq n$ implies $m_k + k - 1 \leq n + k - 1$. It follows that $f(M) \in \binom{[n+k-1]}{k}$ and so the map is well-defined. It should now be easy for the reader to construct an inverse, proving that f is bijective. □

As with the binomial coefficients, we extend $\left(\!\binom{n}{k}\!\right)$ to negative k by letting it equal zero. In the future we will do the same for other constants whose natural domain of definition is $n, k \geq 0$ without comment.

We do wish to comment on an interesting relationship between counting sets and multisets. Note that definition (1.4) is well-defined for any complex number n since the falling factorial is just a product, and in particular it makes sense for negative integers. In fact, if $n \in \mathbb{N}$, then

$$\begin{aligned} \binom{-n}{k} &= \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!} \\ &= (-1)^k \frac{n(n+1)\cdots(n+k-1)}{k!} \\ &= (-1)^k \left(\!\!\binom{n}{k}\!\!\right) \end{aligned} \tag{1.6}$$

by Theorem 1.3.4. This kind of situation where evaluation of an enumerative formula at negative arguments yields, up to sign, another enumerative function is called *combinatorial reciprocity* and will be studied in Section 3.9.

1.4. Set partitions

We have already seen that disjoint unions are nice combinatorially. So it should come as no surprise that set partitions also play an important role.

A *partition* of a set T is a set ρ of nonempty subsets $B_1, \ldots, B_k$ such that $T = \biguplus_i B_i$, written $\rho \vdash T$. The B_i are called *blocks* and we use the notation $\rho = B_1/\ldots/B_k$ leaving out all curly brackets and commas, even though the elements of the blocks, as well as the blocks themselves, are unordered. For example, one set partition of $T = \{a, b, c, d, e, f, g\}$ is

$$\rho = acf/be/d/g = d/eb/g/cfa.$$

We let $B(T)$ be the set of all $\rho \vdash T$. To illustrate,

$$B(\{a, b, c\}) = \{a/b/c,\ ab/c,\ ac/b,\ a/bc,\ abc\}.$$

The nth *Bell number* is $B(n) = \#B([n])$. Although there is no known expression for $B(n)$ as a simple product, there is a recursion.

Theorem 1.4.1. *The Bell numbers satisfy the initial condition $B(0) = 1$ and the recurrence relation*

$$B(n) = \sum_k \binom{n-1}{k-1} B(n-k)$$

for $n \geq 1$.

Proof. The initial condition counts the empty partition of $\emptyset$. For the recursion, given $\rho \in B([n])$, let k be the number of elements in the block B containing n. Then there are $\binom{n-1}{k-1}$ ways to pick the remaining $k-1$ elements of $[n-1]$ to be in B. And the number of ways to partition $[n] - B$ is $B(n-k)$. Summing over all possible k finishes the proof. □

We may sometimes want to keep track of the number of blocks in our partitions. So define $S(T, k)$ to be the set of all $\rho \vdash T$ with k blocks. The *Stirling numbers of the second kind* are $S(n, k) = \#S([n], k)$. We will introduce Stirling numbers of the first kind in the next section. For example

$$S(\{a, b, c\}, 2) = \{ab/c,\ ac/b,\ a/bc\}$$

so $S(3, 2) = 3$. Just as with the binomial coefficients, the $S(n, k)$ for $1 \leq k \leq n$ can be displayed in a triangle as in Figure 1.3. And like the binomial coefficients, these Stirling numbers satisfy a simple recurrence relation.

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 3 & & 1 & & \\
 & 1 & & 7 & & 6 & & 1 & \\
1 & & 15 & & 25 & & 10 & & 1
\end{array}$$

Figure 1.3. Rows 1 through 5 of Stirling's second triangle

Theorem 1.4.2. *The Stirling numbers of the second kind satisfy the initial condition*

$$S(0,k) = \delta_{k,0}$$

and recurrence relation

$$S(n,k) = S(n-1,k-1) + kS(n-1,k)$$

for $n \geq 1$.

Proof. By now, the reader should be able to explain the initial condition without difficulty. For the recursion, the elements $\rho \in S([n],k)$ are of two flavors: those where n is in a block by itself and those where n is in a block with other elements. Removing n in the first case leaves a partition in $S([n-1],k-1)$ and this is a bijection. This accounts for the summand $S(n-1,k-1)$. Removing n in the second case leaves $\sigma \in S([n-1],k)$, but this map is not a bijection. In particular, given σ, one can insert n into any one of its k blocks to recover an element of $S([n],k)$. So the total count is $kS(n-1,k)$ for this case. □

1.5. Permutations by cycle structure

The ordered analogue of a decomposition of a set into a partition is the decomposition of a permutation of $[n]$ into cycles. These are counted by the Stirling numbers of the first kind.

The *symmetric group* is $\mathfrak{S}_n = P([n])$. As the name implies, $\mathfrak{S}_n$ has a group structure defined as follows. If $\pi = \pi_1 \dots \pi_n \in \mathfrak{S}_n$, then we can view this permutation as a bijection $\pi : [n] \to [n]$ where $\pi(i) = \pi_i$. From this it follows that $\mathfrak{S}_n$ is a group where the operation is composition of functions.

Given $\pi \in \mathfrak{S}_n$ and $i \in [n]$, there is a smallest exponent $\ell \geq 1$ such that $\pi^\ell(i) = i$. This and various other claims below will be proved using digraphs in Section 1.9. In this case, the elements $i, \pi(i), \pi^2(i), \dots, \pi^{\ell-1}(i)$ are all distinct and we write

$$c = (i, \pi(i), \pi^2(i), \dots, \pi^{\ell-1}(i))$$

and call this a *cycle of length* ℓ or simply an ℓ*-cycle* of π. Cycles of length one are called *fixed points*. As an example, if $\pi = 6514237$ and $i = 1$, then we have $\pi(1) = 6, \pi^2(1) = 3, \pi^3(1) = 1$ so that $c = (1,6,3)$ is a cycle of π. We now iterate this process: if there is some $j \in [n]$ which is not in any of the cycles computed so far, we find the cycle containing j and continue until every element is in a cycle. The *cycle decomposition* of π is $\pi = c_1 \dots c_k$ where the c_j are the cycles found in this process. Continuing our example, we could get

$$\pi = (1,6,3)(2,5)(4)(7).$$

To distinguish the cycle decomposition of π from its description as $\pi = \pi_1 \dots \pi_n$ we will call the latter the *one-line notation* for π. This is also distinct from *two-line notation*, which is where one writes

$$\pi = \begin{matrix} 1 & 2 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \end{matrix}. \tag{1.7}$$

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 2 & & 3 & & 1 & & \\ & 6 & & 11 & & 6 & & 1 & \\ 24 & & 50 & & 35 & & 10 & & 1 \end{array}$$

Figure 1.4. Rows 1 through 5 of Stirling's first triangle

Note that an ℓ-cycle can be written in ℓ different ways depending on which of its elements one starts with; for example

$$(1, 6, 3) = (6, 3, 1) = (3, 1, 6).$$

Furthermore, the distinct cycles of π are disjoint. So if we think of the cycle c as the permutation of $[n]$ which agrees with π on the elements of c and has all other elements as fixed points, then the cycles of $\pi = c_1 \dots c_k$ commute where we consider the product as a composition of permutations. Returning to our running example, we could write

$$\pi = (1, 6, 3)(2, 5)(4)(7) = (4)(1, 6, 3)(7)(2, 5) = (5, 2)(3, 1, 6)(7)(4).$$

As mentioned above, we defer the proof of the following result until Section 1.9.

Theorem 1.5.1. *Every $\pi \in \mathfrak{S}_n$ has a cycle decomposition $\pi = c_1 \dots c_k$ which is unique up to the order of the factors and cyclic reordering of the elements within each c_i.*

We are now in a position to proceed parallel to the development of set partitions with a given number of blocks in the previous section. For $n \geq 0$ we denote by $c([n], k)$ the set of all permutations in $\mathfrak{S}_n$ which have k cycles in their decomposition. Note the difference between "k cycles" referring to the number of cycles and "k-cycles" referring to the length of the cycles. The *signless Stirling numbers of the first kind* are $c(n, k) = \#c([n], k)$. So, analogous to what we have seen before, $c(n, k) = 0$ for $k < 0$ or $k > n$. To illustrate the notation,

$$c([4], 1) = \{(1, 2, 3, 4), (1, 2, 4, 3), (1, 3, 2, 4), (1, 3, 4, 2), (1, 4, 2, 3), (1, 4, 3, 2)\}$$

so $c(4, 1) = 6$. In general, as you will be asked to prove in an exercise, $c([n], 1) = (n-1)!$. Part of Stirling's first triangle is displayed in Figure 1.4. We also have a recursion.

Theorem 1.5.2. *The signless Stirling numbers of the first kind satisfy the initial condition*

$$c(0, k) = \delta_{k,0}$$

and recurrence relation

$$c(n, k) = c(n - 1, k - 1) + (n - 1)c(n - 1, k)$$

for $n \geq 1$.

Proof. As usual, we concentrate on the recurrence. Given $\pi \in c([n], k)$, we can remove n from its cycle. If n was a fixed point, then the resulting permutations are counted by $c(n - 1, k - 1)$. If n was in a cycle of length at least two, then the permutations obtained upon removal are in $c([n - 1], k)$. So one must find the number of ways to insert n into a cycle of some $\sigma \in c([n-1], k)$. There are ℓ places to insert n in a cycle of length ℓ. So the total number of insertion spots is the sum of the cycle lengths of σ, which is $n-1$. □

The reader may have guessed that there are also (*signed*) *Stirling numbers of the first kind* defined by

$$s(n,k) = (-1)^{n-k} c(n,k).$$

It is not immediately apparent why one would want to attach signs to these constants. We will see one reason in Chapter 5 where it will be shown that the $s(n,k)$ are the Whitney numbers of the first kind for the lattice of set partitions ordered by refinement. Here we will content ourselves with proving an analogue of part (d) of Theorem 1.3.3.

Corollary 1.5.3. *For $n \geq 0$ we have*

$$\sum_k s(n,k) = \begin{cases} 1 & \text{if } n = 0 \text{ or } 1, \\ 0 & \text{if } n \geq 2. \end{cases}$$

Proof. The cases when $n = 0$ or 1 are easy to verify, so assume $n \geq 2$. Since $s(n,k) = (-1)^{n-k}c(n,k)$ and $(-1)^n$ is constant throughout the summation, it suffices to show that $\sum_k (-1)^k c(n,k) = 0$. Using Theorem 1.5.2 and induction on n we obtain

$$\begin{aligned}
\sum_k (-1)^k c(n,k) &= \sum_k (-1)^k c(n-1,k-1) + \sum_k (-1)^k (n-1) c(n-1,k) \\
&= -\sum_k (-1)^{k-1} c(n-1,k-1) + (n-1) \sum_k (-1)^k c(n-1,k) \\
&= -0 + (n-1)0 \\
&= 0
\end{aligned}$$

as desired. □

Note the usefulness of considering the sums in the preceding proof as over $k \in \mathbb{Z}$ rather than $0 \leq k \leq n$. This does away with having to consider any special cases at the values $k = 0$ or $k = n$.

1.6. Integer partitions

Just as one can partition a set into blocks, one can partition a nonnegative integer as a sum. Integer partitions play an important role not just in combinatorics but also in number theory and the representation theory of the symmetric group. See the appendix at the end of the book for more information on the latter.

An *integer partition* of $n \geq 0$ is a multiset λ of positive integers such that the sum of the elements of λ, denoted $|\lambda|$, is n. We also write $\lambda \vdash n$. These elements are called the *parts*. Since the parts of λ are unordered, we will always list them in a canonical order $\lambda = (\lambda_1, \ldots, \lambda_k)$ which is weakly decreasing. We let $P(n)$ denote the set of all partitions of n and $p(n) = \#P(n)$. For example,

$$P(4) = \{(1, 1, 1, 1),\ (2, 1, 1),\ (2, 2),\ (3, 1),\ (4)\}$$

so that $p(4) = 5$. Note the distinction between $P([n])$, which is a set of set partitions, and $P(n)$, which is a set of integer partitions. Sometimes we will just say "partition" if the context makes it clear whether we are partitioning sets or integers. We will use multiplicity notation for integer partitions just as we would for any multiset, writing

$$\lambda = (1^{m_1}, 2^{m_2}, \ldots, n^{m_n})$$

where m_i is the multiplicity of i in λ.

There is no known product formula for $p(n)$. In fact, there is not even a simple recurrence relation. One can use generating functions to derive results about these numbers, but that must wait until Chapter 3. Here we will just introduce a useful geometric device for studying $p(n)$. The *Ferrers* or *Young diagram* of $\lambda = (\lambda_1, \ldots, \lambda_k) \vdash n$ is an array of n boxes into left-justified rows such that row i contains λ_i boxes. Dots are also sometimes used in place of boxes and in this case some authors use "Ferrers diagram" for the dot variant and "Young diagram" for the corresponding array of boxes. We often make no distinction between a partition and its Young diagram. The Young diagram of $\lambda = (5, 5, 2, 1)$ is shown in Figure 1.5. We should warn the reader that we are writing our Young diagrams in *English notation* where the rows are numbered from 1 to k from the top down as in a matrix. Some authors prefer *French notation* where the rows are numbered from bottom to top as in a Cartesian coordinate system. The *conjugate* or *transpose* of λ is the partition λ^t whose Young diagram is obtained by reflecting the diagram of λ about its main diagonal. This is done in Figure 1.5, showing that $(5, 5, 2, 1)^t = (4, 3, 2, 2, 2)$. There is also another way to express the parts of the conjugate.

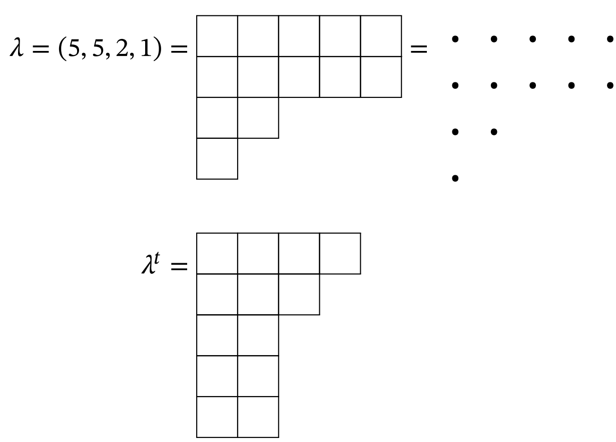


Figure 1.5. A partition, its Young diagram, and its conjugate

Proposition 1.6.1. *If $\lambda = (\lambda_1, \dots, \lambda_k)$ is a partition and $\lambda^t = (\lambda_1^t, \dots, \lambda_l^t)$, then, for $1 \le j \le l$,*

$$\lambda_j^t = \#\{i \mid \lambda_i \ge j\}.$$

Proof. By definition, λ_j^t is the length of the jth column of λ. But that column contains a box in row i if and only if $\lambda_i \ge j$. □

The number of parts of a partition λ is called its *length* and is denoted $\ell(\lambda)$. At this point the reader is probably expecting a discussion of those partitions of n with $\ell(\lambda) = k$. As it turns out, it is a bit simpler to consider $P(n,k)$, the set of all partitions λ of n with $\ell(\lambda) \le k$, and $p(n,k) = \#P(n,k)$. Note that the number of $\lambda \vdash n$ with $\ell(\lambda) = k$ is just $p(n,k) - p(n,k-1)$. So in some sense the two viewpoints are equivalent. But it will be easier to state our results in terms of $p(n,k)$. Note also that

$$p(n,0) \le p(n,1) \le \cdots \le p(n,n) = p(n,n+1) = \cdots = p(n).$$

Because of this behavior, it is best to display the $p(n,k)$ in a matrix, rather than a triangle, keeping in mind that the entries in the nth row eventually stabilize to an infinite repetition of the constant $p(n)$. Part of this array will be found in Figure 1.6. We also assume that $p(n,k) = 0$ if $n < 0$ or $k < 0$. Unlike $p(n)$, one can write down a simple recurrence relation for $p(n,k)$.

Theorem 1.6.2. *The $p(n,k)$ satisfy*

$$p(0,k) = \begin{cases} 0 & \text{if } k < 0, \\ 1 & \text{if } k \ge 0 \end{cases}$$

and

$$p(n,k) = p(n-k,k) + p(n,k-1)$$

for $n \ge 1$

Proof. We skip directly to the recursion. Note that since conjugation is a bijection, $p(n,k)$ also counts the partitions $\lambda = (\lambda_1, \dots, \lambda_l) \vdash n$ such that $\lambda_1 \le k$. It will be convenient to use this interpretation of $p(n,k)$ for the proof. We have two possible cases. If $\lambda_1 = k$, then $\mu = (\lambda_2, \dots, \lambda_l) \vdash n-k$ and $\lambda_2 \le \lambda_1 = k$. So these partitions are counted by $p(n-k,k)$. The other possibility is that $\lambda_1 \le k-1$. And these λ are taken care of by the $p(n,k-1)$ term. □

	0	1	2	3	4	5
0	1	1	1	1	1	1
1	0	1	1	1	1	1
2	0	1	2	2	2	2
3	0	1	2	3	3	3
4	0	1	3	4	5	5

Figure 1.6. The values $p(n,k)$ for $0 \le n \le 4$ and $0 \le k \le 5$

1.7. Compositions

Recall that integer partitions are really unordered even though we usually list them in weakly decreasing fashion. This raises the question about what happens if we considered ways to write n as a sum when the summands are ordered. This is the notion of a composition.

A *composition* of n is a sequence $\alpha = [\alpha_1, \ldots, \alpha_k]$ of positive integers called *parts* such that $\sum_i \alpha_i = n$. We write $\alpha \models n$ and use square brackets to distinguish compositions from integer partitions. This causes a notational conflict between $[n]$ as a composition of n and as the integers from 1 to n, but the context should make it clear which interpretation is meant. Let $Q(n)$ be the set of compositions of n and $q(n) = \#Q(n)$. So the compositions of 4 are

$$Q(4) = \{[1,1,1,1],\ [2,1,1],\ [1,2,1],\ [1,1,2],\ [2,2],\ [3,1],\ [1,3],\ [4]\}.$$

So $q(4) = 8$, which is a power of 2. This, as your author is fond of saying, is not a coincidence.

Theorem 1.7.1. *For $n \geq 1$ we have*

$$q(n) = 2^{n-1}.$$

Proof. There is a famous bijection $\phi : 2^{[n-1]} \to Q(n)$, which we will use to prove this result. This map will be useful when working with quasisymmetric functions in Chapter 8. Given $S = \{s_1, \ldots, s_k\} \subseteq [n-1]$ written in increasing order, we define

$$\phi(S) = [s_1 - s_0,\ s_2 - s_1,\ \ldots,\ s_k - s_{k-1},\ s_{k+1} - s_k] \tag{1.8}$$

where, by definition, $s_0 = 0$ and $s_{k+1} = n$. To show that ϕ is well-defined, suppose $\phi(S) = [\alpha_1, \ldots, \alpha_{k+1}]$. Since S is increasing, $\alpha_i = s_i - s_{i-1}$ is a positive integer. Furthermore

$$\sum_{i=1}^{k+1} \alpha_i = \sum_{i=1}^{k+1} (s_i - s_{i-1}) = s_{k+1} - s_0 = n.$$

Thus $\phi(S) \in Q(n)$ as desired.

To show that ϕ is bijective, we construct its inverse $\phi^{-1} : Q(n) \to 2^{[n-1]}$. Given $\alpha = [\alpha_1, \ldots, \alpha_{k+1}] \in Q(n)$, we let

$$\phi^{-1}(\alpha) = \{\alpha_1,\ \alpha_1 + \alpha_2,\ \alpha_1 + \alpha_2 + \alpha_3,\ \ldots, \alpha_1 + \alpha_2 + \cdots + \alpha_k\}.$$

It should not be hard for the reader to prove that ϕ^{-1} is well-defined and the inverse of ϕ. □

As usual, we wish to make a more refined count by restricting the number of constituents of the object under consideration. Let $Q(n,k)$ be the set of all compositions of n with exactly k parts and let $q(n,k) = \#Q(n,k)$. Since the $q(n,k)$ will turn out to be previously studied constants, we will forgo the usual triangle. The result below follows easily by restricting the function ϕ from the previous proof, so the demonstration is omitted.

Theorem 1.7.2. *The composition numbers satisfy*

$$q(0, k) = \delta_{k,0}$$

and

$$q(n, k) = \binom{n-1}{k-1}$$

for $n \geq 1$. □

1.8. The twelvefold way

We now have all the tools in place to count certain functions. There are 12 types of such functions and so this scheme is called the twelvefold way, an idea which was introduced in a series of lectures by Gian-Carlo Rota. The name was suggested by Joel Spencer and should not be confused with the twelvefold path of Buddhism!

We will consider three types of functions $f : D \to R$, namely, arbitrary functions, injections, and surjections. We will also permit the domain D and range R to be of two types each: either *distinguishable*, which means it is a set, or *indistinguishable*, which means it is a multiset consisting of a single element repeated some number of times. Thus the total number of types of functions under consideration is the product of the number of choices for f, D, and R or $3 \cdot 2 \cdot 2 = 12$. Of course, a function where the domain or range is a multiset is not really well-defined, even though the intuitive notion should be clear. To be precise, when D is a multiset and R is a set, suppose D' is a set with $|D'| = |D|$. Then a *function* $f : D \to R$ is an equivalence class of functions $f : D' \to R$ where f and g are equivalent if $\#f^{-1}(r) = \#g^{-1}(r)$ for all $r \in R$. The reader can come up with the corresponding notions for the other cases if desired. We will assume throughout that $|D| = n$ and $|R| = k$ are both nonnegative integers. We will collect the results in the chart in Table 1.1.

We first deal with the case where both D and R are distinguishable. Without loss of generality, we can assume that $D = [n]$. So a function $f : D \to R$ can be considered as a word $w = f(1)f(2)\dots f(n)$. Since there are k choices for each $f(i)$, we have, by Theorem 1.2.2, that the number of such f is $\#P(([k], n)) = k^n$. If f is injective, then w becomes a permutation, giving the count $\#P([k], n) = k\downarrow_n$ from Theorem 1.2.1. For surjective functions, we need a new concept. If D is a set, then the *kernel* of a function $f : D \to R$ is the partition $\ker f$ of D whose blocks are the nonempty subsets of the form $f^{-1}(r)$ for $r \in R$. For example, if $f : \{a, b, c, d\} \to \{1, 2, 3\}$ is given by $f(a) = f(c) = 2$,

Table 1.1. The twelvefold way

D	R	arbitrary f	injective f	surjective f
dist.	dist.	k^n	$k\downarrow_n$	$k!\,S(n, k)$
indist.	dist.	$\binom{n+k-1}{n}$	$\binom{k}{n}$	$\binom{n-1}{k-1}$
dist.	indist.	$\sum_{j=0}^{k} S(n, j)$	$\delta(n \leq k)$	$S(n, k)$
indist.	indist.	$p(n, k)$	$\delta(n \leq k)$	$p(n, k) - p(n, k-1)$

$f(b) = 3$, and $f(d) = 1$, then $\ker f = ac/b/d$. If f is to be surjective, then the function can be specified by picking a partition of D for $\ker f$ and then picking a bijection g from the blocks of $\ker f$ into R. Continuing our example, f is completely determined by its kernel and the bijection $g(ac) = 2$, $g(b) = 3$, and $g(d) = 1$. The number of ways to choose $\ker f = B_1/\ldots/B_k$ is $S(n,k)$ by definition. And, using the injective case with $n = k$, the number of bijections $g : \{B_1,\ldots,B_k\} \to R$ is $k\downarrow_k = k!$. So the total count is $k!\,S(n,k)$.

Now suppose D is indistinguishable and R is distinguishable where we assume $R = [k]$. Then one can think of $f : D \to R$ as a multiset $M = \{\{1^{m_1},\ldots,k^{m_k}\}\}$ on R where $m_i = \#f^{-1}(i)$. It follows that $\sum_i m_i = \#D = n$. So, by Theorem 1.3.4, the number of all such f is

$$\left(\!\!\binom{k}{n}\!\!\right) = \binom{n+k-1}{n}.$$

If f is to be injective, then we are picking an n-element subset of $R = [k]$ giving a count of $\binom{k}{n}$. If f is to be surjective, then $m_i \geq 1$ for all i so that $[m_1,\ldots,m_k]$ is a composition of n. It follows from Theorem 1.7.2 that the number of functions is $q(n,k) = \binom{n-1}{k-1}$.

To deal with the case when $D = [n]$ is distinguishable and R is indistinguishable, we introduce a useful extension of the Kronecker delta. If S is any statement, we let

$$\delta(S) = \begin{cases} 1 & \text{if } S \text{ is true,} \\ 0 & \text{if } S \text{ is false.} \end{cases} \tag{1.9}$$

Returning to our counting, f is completely determined by its kernel, which is a partition of $[n]$. If we are considering all f, then the kernel can have any number of blocks up to and including k. Summing the corresponding Stirling numbers gives the corresponding entry in Table 1.1. If f is injective, then for such a function to exist we must have $n \leq k$. And in that case there is only one possible kernel, namely the partition into singleton blocks. This count can be summarized as $\delta(n \leq k)$. For surjective f we are partitioning $[n]$ into exactly k blocks, giving $S(n,k)$ possibilities.

If D and R are both indistinguishable, then the nonzero numbers of the form $m_i = \#f^{-1}(r)$ for $r \in R$ completely determine f. And these numbers form a partition of $n = \#D$ into at most $k = \#R$ parts. Recalling the notation of Section 1.6, the total number of such f is $p(n,k)$. The line of reasoning for injective functions follows that of the previous paragraph with the same resulting answer. Finally, for surjectivity we need exactly k parts, which is counted by $p(n,k) - p(n,k-1)$.

1.9. Graphs and digraphs

Graph theory is a substantial part of combinatorics. We will use directed graphs to give the postponed proof of the existence and uniqueness of the cycle decomposition of permutations in $\mathfrak{S}_n$.

A *labeled graph* $G = (V,E)$ consists of a set V of elements called *vertices* and a set E of elements called *edges* where an edge consists of an unordered pair of vertices. We will write $V(G)$ and $E(G)$ for the vertex and edge set of G, respectively, if we wish to emphasize the graph involved. Geometrically, we think of the vertices as nodes and the edges as line segments or curves joining them. Conventionally, in graph theory an

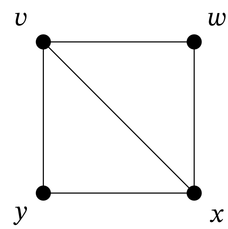


Figure 1.7. A graph G

edge connecting vertices v and w is written $e = vw$ rather than $e = \{v, w\}$. In this case we say that e *contains* v and w, or that e has *endpoints* v and w. We also say that v and w are *neighbors*. For example, a drawing of the graph G with vertices $V = \{v, w, x, y\}$ and edges $E = \{vw, vx, vy, wx, xy\}$ is displayed in Figure 1.7. If $\#V = 1$, then there is only one graph with vertex set V and such a graph is called *trivial*.

Call graph H a *subgraph* of G, written $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. In this case we also say that G *contains* H. There are several types of subgraphs which will play an important role in what follows. A *walk of length* ℓ in G is a sequence of vertices $W : v_0, v_1, \ldots, v_\ell$ such that $v_{i-1}v_i \in E$ for $1 \le i \le \ell$. We say that the walk is *from* v_0 *to* v_ℓ, or is a v_0–v_ℓ *walk*, or that v_0, v_ℓ are the *endpoints* of W. We call W a *path* if all the vertices are distinct and we usually use letters like P for paths. In particular, we will use W_n or P_n to denote a walk or a path having n vertices, respectively. In our example graph, $P : y, v, x, w$ is a path of length 3 from y to w. Notice that length refers to the number of edges in the path, which is one less than the number of vertices. A *cycle of length* ℓ in G is a sequence of distinct vertices $C : v_1, v_2, \ldots, v_\ell$ such that we have distinct edges $v_{i-1}v_i$ for $1 \le i \le \ell$, and subscripts are taken modulo ℓ so that $v_0 = v_\ell$. Returning to our running example, $C : v, x, y$ is a cycle in G of length 3. In a cycle the length is both the number of vertices and the number of edges. The notation C_n will be used for a cycle with n vertices and we will call this an *n-cycle*. We also denote by K_n the *complete graph* which consists of n vertices and all possible $\binom{n}{2}$ edges between them. A copy of a complete graph in a graph G is often called a *clique*. There is a close relationship between some of the parts of a graph which we have just defined.

Lemma 1.9.1. *Let G be a graph and let $u, v \in V$.*

(a) *Any walk from u to v contains a path from u to v.*

(b) *The union of any two different paths from u to v contains a cycle.*

Proof. We will prove (a) and leave (b) as an exercise. Let $W : v_0, \ldots, v_\ell$ be the walk. We will induct on ℓ, the length of W. If $\ell = 0$, then W is a path. So assume $\ell \ge 1$. If W is a path, then we are done. If not, then some vertex of W is repeated, say $v_i = v_j$ for $i < j$. Then we have a u–v walk $W' : v_0, v_1, \ldots, v_i, v_{j+1}, v_{j+2}, \ldots, v_\ell$ which is shorter than W. By induction, W' contains a path P and so W contains P as well. □

To state our first graphical enumeration result, let $\mathcal{G}(V)$ be the set of all graphs on the vertex set V. We will also use $\mathcal{G}(V,k)$ to denote the set of all graphs in $\mathcal{G}(V)$ with k edges.

Theorem 1.9.2. *For $n \ge 1$ and $k \ge 0$ we have*

$$\#\mathcal{G}([n]) = 2^{\binom{n}{2}}$$

and

$$\#\mathcal{G}([n],k) = \binom{\binom{n}{2}}{k}.$$

Proof. If $V = [n]$ is given, then a graph G with vertex set V is completely determined by its edge set. Since there are n vertices, there are $\binom{n}{2}$ possible edges to choose from. So the number of G in $\mathcal{G}([n])$ is the number of subsets of these edges, which, by Theorem 1.3.1, is the given power of 2. The proof for $\mathcal{G}([n],k)$ is similar, just using the definition (1.4). $\square$

A graph is *unlabeled* if the vertices in V are indistinguishable. If the type of graph is clear from the context or does not matter for the particular application at hand, we will omit the adjectives "labeled" and "unlabeled". The enumeration of unlabeled graphs is much more complicated than for labeled ones. So this discussion is postponed until Section 6.4 where we will develop the necessary tools.

If G is a graph and $v \in V$, then the *degree of v* is

$$\deg v = \text{the number of } e \in E \text{ containing } v.$$

In our running example $\deg v = \deg x = 3$ and $\deg w = \deg y = 2$. There is a nice relationship between vertex degrees and the cardinality of the edge set. The demonstration of the next result illustrates an important method of proof in combinatorics, counting in pairs.

Theorem 1.9.3. *For any graph G we have*

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

Proof. Consider

$$P = \{(v,e) \mid v \text{ is contained in } e\}.$$

Then

$$\#P = \sum_{v\in V} (\text{number of } e \text{ containing } v) = \sum_{v\in V} \deg v.$$

On the other hand

$$\#P = \sum_{e\in E} (\text{number of } v \text{ contained in } e) = \sum_{e\in E} 2 = 2|E|.$$

Equating the two counts finishes the proof. $\square$

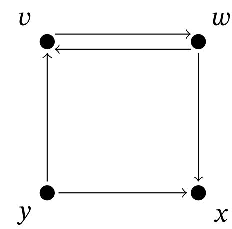


Figure 1.8. A digraph D

Theorem 1.9.3 is often called the Handshaking Lemma because of the following interpretation. Suppose V is the set of people at a party and we draw an edge between person v and person w if they shake hands during the festivities. Then adding up the number of handshakes given by each person gives twice the total number of handshakes.

It is often useful to have specified directions along the edges. A *labeled directed graph*, also called a *digraph*, is $D = (V, A)$ where V is a set of vertices and A is a set of *arcs* which are ordered pairs of vertices. We use the notation $a = \overrightarrow{vw}$ for arcs and say that a goes *from* v *to* w. To illustrate, the digraph with $V = \{v, w, x, y\}$ and $A = \{\overrightarrow{vw}, \overrightarrow{wv}, \overrightarrow{wx}, \overrightarrow{yv}, \overrightarrow{yx}\}$ is drawn in Figure 1.8. We use $V(D)$ and $A(D)$ to denote the vertex set and arc set, respectively, of a digraph D when we wish to be more precise. *Directed walks, paths*, and *cycles* are defined for digraphs similarly to their undirected cousins in graphs, just insisting the $\overrightarrow{v_{i-1}v_i} \in A$ for i in the appropriate range. So, in our example digraph, $P : y, v, w, x$ is a directed path and $C : v, w$ is a directed cycle. Note that w, x, y, v is not a directed path because the arc between x and y goes the wrong way.

Let $\mathcal{D}(V)$ and $\mathcal{D}(V, k)$ be the set of digraphs and the set of digraphs with k arcs, respectively, having vertex set V. The next result is proved in much the same manner as Theorem 1.9.2 so the demonstration is omitted.

Theorem 1.9.4. *For $n \geq 1$ and $k \geq 0$ we have*

$$\#\mathcal{D}([n]) = 2^{n(n-1)}$$

and

$$\#\mathcal{D}([n], k) = \binom{n(n-1)}{k}. \qquad \square$$

In a digraph D there are two types of degrees. Vertex $v \in V$ has *out-degree* and *in-degree*

$$\operatorname{odeg} v = \text{the number of } a \in A \text{ of the form } a = \overrightarrow{vw},$$
$$\operatorname{ideg} v = \text{the number of } a \in A \text{ of the form } a = \overrightarrow{wv},$$

respectively. In Figure 1.8, for example, $\operatorname{odeg} v = 1$ and $\operatorname{ideg} v = 2$. The next result will permit us to finish our leftover business from Section 1.5. The union of digraphs $D \cup E$ is the digraph with vertices $V(D \cup E) = V(D) \cup V(E)$ and arcs $A(D \cup E) = A(D) \cup A(E)$.

Lemma 1.9.5. *Let $D = (V, A)$ be a digraph. We have* $\operatorname{odeg} v = \operatorname{ideg} v = 1$ *for all $v \in V$ if and only if D is a disjoint union of directed cycles.*

Proof. The reverse implication is easy to see since the out-degree and in-degree of any vertex v of D would be the same as those degrees in the directed cycle containing v. But in such a cycle $\operatorname{odeg} v = \operatorname{ideg} v = 1$.

For the forward direction, pick any $v = v_1 \in V$. Since $\operatorname{odeg} v_1 = 1$ there must exist a vertex v_2 with $\overrightarrow{v_1v_2} \in A$. By the same token, there must be a v_3 with $\overrightarrow{v_2v_3} \in A$. Continue to generate a sequence $v_1, v_2, \ldots$ in this manner. Since V is finite, there must be two indices $i < j$ such that $v_i = v_j$. Let i be the smallest such index and let j be the first index after i where repetition occurs. Thus $i = 1$, for if not, then we have $\overrightarrow{v_{i-1}v_i}, \overrightarrow{v_{j-1}v_i} \in A$, contradicting the fact that $\operatorname{ideg} v_i = 1$. By definition of j, we have a directed cycle $C : v_1, v_2, \ldots, v_{j-1}$. Furthermore, no vertex of C can be involved in another arc since that would make its out-degree or in-degree too large. Continuing in this manner, we can decompose D into disjoint directed cycles. □

Sometimes it is useful to allow *loops* in a graph which are edges of the form $e = vv$. Similarly, we can permit loops as arcs $a = \overrightarrow{vv}$ in a digraph. Another possibility is that we would want *multiple edges*, meaning that one could have more than one edge between a given pair of vertices, making E into a multiset. *Multiple arcs* are defined similarly. If we make no specification for our (di)graph, then we are assuming that it has neither loops nor multiple edges. We will now prove Theorem 1.5.1.

Proof (*of Theorem* 1.5.1). To any $\pi \in \mathfrak{S}_n$ we associate its *functional digraph* D_π which has $V = [n]$ and an arc $\vec{ij} \in A$ if and only if $\pi(i) = j$. Now D_π is a digraph with loops. Because π is a function we have $\operatorname{odeg} i = 1$ for all $i \in [n]$. And because π is a bijection we also have $\operatorname{ideg} i = 1$ for all i. The proof of the previous lemma works equally well if one allows loops. So D_π is a disjoint union of cycles. But cycles of the digraph D_π correspond to cycles of the permutation π. Thus the cycle decomposition of π exists. It is also easy to check that the cycles of D_π produced by the algorithm in the demonstration of necessity in Lemma 1.9.5 are unique. This implies the uniqueness statement about the cycles of π and so we are done. □

1.10. Trees

Trees are a type of graph which often occurs in practice, even in domains outside of mathematics. For example, trees are used as data structures in computer science, or to model evolution in genetics. A graph G is *connected* if, for every pair of vertices $v, w \in V$, there is a walk in G from v to w. By Lemma 1.9.1(a), this is equivalent to there being a path from v to w in G. The *connected components* of G are the maximal connected subgraphs. If G is connected, there is only one component. Call G *acyclic* if it contains no cycles. A *forest* is another name for an acyclic graph. The connected components of a forest are called *trees*. So a graph T is a tree if it is both connected and acyclic. Figure 1.9 contains five trees $T_1, \ldots, T_5$.

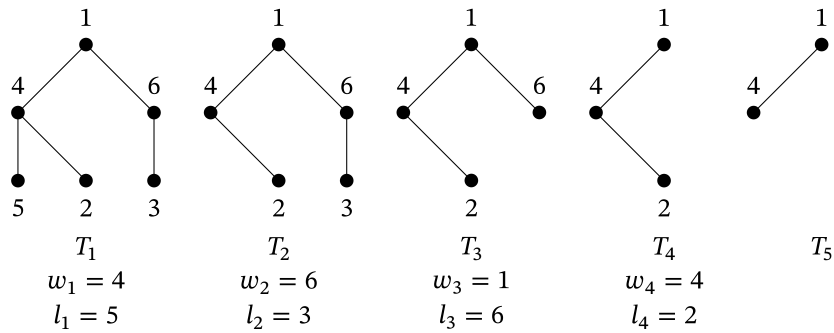


Figure 1.9. The Prüfer algorithm

A *leaf* in a graph G is a vertex v having $\deg v = 1$. The next result will show that nontrivial trees have leaves (regardless of the time of year). Further, it should be clear from this lemma why leaves are a useful tool for induction in trees. In order to state it, we need the following notation. If G is a graph and $W \subseteq V$, then $G - W$ is the graph on the vertex set $V - W$ whose edge set consists of all edges in E with both endpoints in $V - W$. If $W = \{v\}$ for some v, then we write $G - v$ for $G - \{v\}$. In Figure 1.9, $T_2 = T_1 - 5$. Similarly, if $F \subseteq E$, then $G - F$ is the graph with $V(G - F) = V(G)$ and $E(G - F) = E(G) - E(F)$. If F consists of a single edge, then we use a similar abbreviation as for subtracting vertices.

Lemma 1.10.1. *Let T be a tree with $\#V \geq 2$.*

(a) *T has (at least) 2 leaves.*

(b) *If v is a leaf of T, then $T' = T - v$ is also a tree.*

Proof. (a) Let $P : v_0, \ldots, v_\ell$ be a path of maximum length in T. Since T is nontrivial, $v_0 \neq v_\ell$. We claim that v_0, v_ℓ are leaves and we will prove this for v_0 as the same proof works for v_ℓ. Suppose, towards a contradiction, that $\deg v_0 \geq 2$. Then there must be a vertex $w \neq v_1$ such that $v_0 w \in E$. We now have two possibilities. If w is not a vertex of P, then the path $P' : w, v_0, \ldots, v_\ell$ is longer than P, a contradiction to the definition of P. If $w = v_i$ for some $2 \leq i \leq \ell$, then the portion of P from v_0 to v_i together with the edge $v_0 v_i$ forms a cycle in T, again a contradiction.

(b) It is clear that T' is still acyclic since removing vertices cannot create a cycle. To show it is connected, take $x, y \in V(T')$. So x, y are also vertices of T. Since T is connected, Lemma 1.9.1(a) implies that there is a path P from x to y in T. If this path is also in T', then we will be done. But if P goes through v, then, since there is a unique vertex v' adjacent to v, P would have to pass through v' just before and just after v. This contradicts the fact that the vertices of P are distinct. □

There are a number of characterizations of trees. We collect some of them here as they will be useful in the sequel.

Theorem 1.10.2. *Let T be a graph with $\#V = n$ and $\#E = m$. The following are equivalent conditions for T to be a tree:*

(a) *T is connected and acyclic.*

(b) *T is acyclic and $n = m + 1$.*

(c) *T is connected and $n = m + 1$.*

(d) *For every pair of vertices u, v there is a unique path from u to v.*

Proof. We will prove the equivalence of (a), (b), and (c). The equivalence of (a) and (d) is left as an exercise. To prove that (a) implies (b), it suffices to show by induction on n that $n = m + 1$. This is trivial if $n = 1$. If $n \geq 2$, then, by Lemma 1.10.1, T has a leaf v. Induction applies to $T' = T - v$ so that its vertex and edge cardinalities are related by $n' = m' + 1$. But $n = n' + 1$ and $m = m' + 1$ so that $n = m + 1$.

To see why (b) implies (c), consider the connected components $T_1, \ldots, T_k$ of T. Since T is acyclic, each of these components is a tree. Also, from the implication (a) $\Longrightarrow$ (b), we have that $n_i = m_i + 1$ for $1 \leq i \leq k$ where $n_i = \#V(T_i)$ and $m_i = \#E(T_i)$. Adding these equations together and using the fact that $\sum_i n_i = n$ and $\sum_i m_i = m$ we obtain $n = m + k$. But we are given that $n = m + 1$. So we must have $k = 1$. This means that T only has one component and so is connected.

We prove that (c) implies (a) by contradiction. So suppose that T contains a cycle C and let $e = uv \in E(C)$. We claim that $T - e$ is still connected. For if x, y are any two vertices of $T - e$, then there is a walk W from x to y in T. If W does not contain e, then W is still in $T - e$. If W does contain e, then replace e in W with the path $C - e$ to form a new walk W' from x to y in $T - e$. We can keep removing edges in this way until the resulting graph T' is acyclic. Since T' is still connected, it is a tree. And by the first implication we have $n' = m' + 1$. But $n' = n$ and $m' < m$ so that $n < m + 1$, the desired contradiction. □

Let $\mathcal{T}(V)$ be the set of all trees on the vertex set V. There are quite a number of different proofs of the beautiful formula below for $\#\mathcal{T}(V)$, many of which are in Moon's book on the subject [**64**].

Theorem 1.10.3. *For $n \geq 1$ we have*

$$\#\mathcal{T}([n]) = n^{n-2}.$$

Proof. The result is trivial if $n = 1$, so assume $n \geq 2$. By Theorem 1.2.2 it suffices to find a bijection $f : \mathcal{T}([n]) \to P(([n], n - 2))$. There is a famous algorithm for constructing f which is called the *Prüfer algorithm*. An example will be found in Figure 1.9. Given $T \in \mathcal{T}([n])$, to determine $f(T) = w_1 \ldots w_{n-2}$ we will build a sequence of trees $T = T_1, T_2, \ldots, T_{n-1}$ by removing vertices from T as follows. Since the vertices of T are labeled $1, \ldots, n$ it makes sense to talk about, e.g., a maximum vertex because of the ordering on the integers. Given T_i, we find the leaf $l_i \in V(T_i)$ such that l_i is maximum and let $T_{i+1} = T_i - l_i$. By the previous lemma, T_{i+1} will also be a tree. Since l_i is a leaf, it is adjacent to a unique vertex w_i in T_i and we let w_i be the ith element of $f(T)$. Now each $w_i \in [n]$ and $f(T)$ has length $n - 2$ by definition. So $f(T) \in P(([n], n - 2))$.

To show that f is a bijection, we find its inverse. Given $w \in P(([n], n-2))$, we will first construct a permutation $l = l_1 \dots l_{n-2} \in P([n], n-2)$ where l_i will turn out to be the leaf removed from T_i to form T_{i+1}. We construct the l_i inductively by letting

$$l_i = \max([n] - \{l_1, \dots, l_{i-1}, w_i, \dots, w_{n-2}\}). \tag{1.10}$$

Finally we construct $f^{-1}(w) = T$ by letting T have edges $e_i = l_i w_i$ for $1 \le i \le n-2$ as well as the edge $e_{n-1} = l_{n-1} l_n$ where $[n] - \{l_1, \dots, l_{n-2}\} = \{l_{n-1}, l_n\}$. To show that $f^{-1}(w) = T$ is a tree, note first that l_1 is a leaf of T because l_1 is attached to w_1 but to none of the other vertices of T by (1.10) and the definition of e_{n-1}. Consider $w' = w_2 \dots w_{n-2}$ and apply the algorithm for f^{-1} to w' using the ground set $[n] - \{l_1\}$ instead of $[n]$. By induction, the result is a tree T'. And T is formed by adding l_1 as a leaf to T', which makes T a tree as well.

To show that f and f^{-1} are inverses we will show that $f^{-1} \circ f$ is the identity map, leaving the proof for $f \circ f^{-1}$ to the reader. Suppose $f(T) = w_1 \dots w_{n-2}$. Also let the sequence of leaves removed during the construction of $f(T)$ be $l'_1 \dots l'_{n-2}$. Then by definition of the algorithm, the edges of T are exactly $l'_i w_i$ for $1 \le i \le n-2$ and $l'_{n-1} l'_n$ where $[n] - \{l'_1, \dots, l'_{n-2}\} = \{l'_{n-1}, l'_n\}$. Comparing this with the definition of f^{-1} we see that it suffices to show that $l_i = l'_i$ for all i and that this will follow if one can prove the equality holds for $1 \le i \le n-2$. Since l'_i is a leaf in T_i, it cannot be any of the previously removed leaves $l'_1, \dots, l'_{i-1}$. Of the remaining vertices, those which are among $w_i, \dots, w_{n-2}$ are not currently leaves since they are attached to future leaves which are to be removed. And conversely those not among the $w_i, \dots, w_{n-2}$ must be leaves; otherwise, they would be listed as some w_j for $j \ge i$ once all their adjacent leaves were removed. Hence the leaves of T_i are precisely the elements of $[n] - \{l'_1, \dots, l'_{i-1}, w_i, \dots, w_{n-2}\}$. Since we always remove the leaf of maximum value, we see that the rule for choosing l'_i is exactly the same as the one in (1.10). So $l_i = l'_i$ as desired. □

1.11. Lattice paths

Lattice paths lead to many interesting counting problems in combinatorics. They are also important in probability and statistics; see the book of Mohanty [**63**] for examples.

Consider the *integer lattice* in the plane

$$\mathbb{Z}^2 = \{(x, y) \mid x, y \in \mathbb{Z}\}.$$

A *lattice path* is a sequence of elements of $\mathbb{Z}^2$ written

$$P : (x_0, y_0), (x_1, y_1), \dots, (x_\ell, y_\ell).$$

Just as in graph theory, we say the path has *length* ℓ and goes from (x_0, y_0) to (x_ℓ, y_ℓ), which are called its *endpoints*. Unlike graph-theoretic paths, we do not assume the (x_i, y_i) are distinct. To illustrate the notation, if we assume that the left-hand path in Figure 1.10 starts at the origin, then it would be written

$$P : (0,0),\ (0,1),\ (0,2),\ (1,2),\ (1,3),\ (2,3),\ (3,3),\ (3,4),\ (4,4).$$

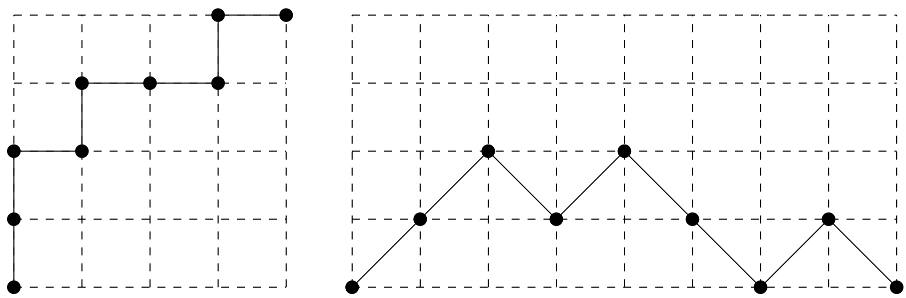


Figure 1.10. Dyck paths

The *step* between (x_{i-1}, y_{i-1}) and (x_i, y_i) on P is the vector $s_i = [x_i - x_{i-1}, y_i - y_{i-1}]$. Note the use of square versus round brackets to distinguish steps from vertices of the path. Note that P is determined up to translation by its steps and that it is determined completely by its steps and initial vertex. If no initial vertex is specified, it is assumed to be the origin. We let $E = [1, 0]$ and $N = [0, 1]$, calling these *east* and *north* steps, respectively. The path on the left in Figure 1.10 could also be represented $P : NNENEENE$.

For our first enumerative result, we use the notation $\mathcal{NE}(m, n)$ for the set of lattice paths from $(0, 0)$ to (m, n) only using steps north and east. We call lattice paths using only N and E steps *northeast paths.*

Theorem 1.11.1. *For $m, n \geq 0$ we have*

$$\#\mathcal{NE}(m, n) = \binom{m+n}{m}.$$

Proof. Let P be a northeast lattice path from $(0, 0)$ to (m, n). Then P has $m + n$ total steps. And once m of them are chosen to be E, the rest must be N. The result follows. □

We will be particularly concerned with a special type of northeast path. A *Dyck path of semilength n* is a northeast lattice path which begins at $(0, 0)$, ends at (n, n), and never goes below the line $y = x$. The first path in Figure 1.10 is of this type. Note that n is called the semilength because the Dyck path itself has $2n$ steps. We let $\mathcal{D}(n)$ denote the set of Dyck paths of semilength n. This should cause no confusion with the notation $\mathcal{D}(V)$ for the set of digraphs on the vertex set V because in the former notation n is a nonnegative integer while in the latter it is a set. We now define that *Catalan numbers* to be

$$C(n) = \#\mathcal{D}(n).$$

The Catalan numbers are ubiquitous in combinatorics. In fact, Stanley has written a book [**92**] containing 214 different combinatorial interpretations of $C(n)$. A few of these are listed in the exercises. The Catalan numbers satisfy a nice recursion.

Theorem 1.11.2. *We have the initial condition*

$$C(0) = 1$$

and recurrence relation

$$C(n) = C(0)C(n-1) + C(1)C(n-2) + C(2)C(n-3) + \cdots + C(n-1)C(0)$$

for $n \geq 1$.

Proof. The initial condition counts the trivial path of a single vertex. For the recursion, take $P : v_0, \ldots, v_{2n} \in \mathcal{D}(n)$ where $v_i = (x_i, y_i)$ for all i. Let $j > 0$ be the smallest index such that v_{2j} is on the line $y = x$. Such an index exists since $v_{2n} = (n, n)$ satisfies this condition. Also note that no vertex of odd subscript is on $y = x$ since the number of north steps and the number of east steps preceding that vertex cannot be equal. It follows that P_1, the portion of P from v_1 to v_{2j-1}, stays above $y = x + 1$. So the number of choices for P_1 is $C(j-1)$. Furthermore, if P_2 is the portion of P from v_{2j} to v_{2n}, then P_2 is (a translation of) a Dyck path of semilength $n - j$. So the number of choices for P_2 is $C(n-j)$. Thus the total number of such P is $C(j-1)C(n-j)$. Summing over $1 \leq j \leq n$ finishes the proof. □

There is an explicit expression for the Catalan numbers. But to derive this formula it will be convenient to use a second set of paths counted by $C(n)$. Call the steps $U = [1, 1]$ and $D = [1, -1]$ *up* and *down*, respectively. An *updown* path is one using only such steps. It should be clear that if we let $\tilde{\mathcal{D}}(n)$ be the set of updown lattice paths from $(0, 0)$ to $(2n, 0)$ never going below the x-axis, then $\#\tilde{\mathcal{D}}(n) = \#\mathcal{D}(n) = C(n)$. In fact one can get from the paths in one set to those in the other by rotation and dilation of the plane. The two paths in Figure 1.10 correspond under this map and the second one would be represented as $P : UUDUDDUD$.

Theorem 1.11.3. *For* $n \geq 0$ *we have*

$$C(n) = \frac{1}{n+1}\binom{2n}{n}.$$

Proof. We rewrite the right-hand side as

$$\frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!\,(n+1)!} = \frac{1}{2n+1}\binom{2n+1}{n}.$$

Let $\mathcal{P}$ be the set of all updown paths starting at $(0, 0)$ and ending at $(2n + 1, -1)$. Such paths have $2n + 1$ steps of which n are up (forcing the other $n + 1$ to be down) so that $\#\mathcal{P} = \binom{2n+1}{n}$. Our strategy will be to find a partition ρ of $\mathcal{P}$ such that

(1) $\#B = 2n + 1$ for every block B of ρ and

(2) there is a bijection between the blocks of ρ and the paths in $\tilde{\mathcal{D}}(n)$.

It will then follow that $\#\tilde{\mathcal{D}}(n)$ is equal to the number of blocks of ρ, which is $\#\mathcal{P}/(2n + 1)$, giving the desired equality.

To determine ρ, we will take any $P \in \mathcal{P}$ and describe the block B containing P. We will refer to the y-coordinate of a vertex v of P as its *height*, written $\operatorname{ht} v$. Suppose P has step representation $P : s_1 s_2 \ldots s_{2n+1}$. Define the rth *rotation* of P to be the path

$$P_r : s_{r+1} s_{r+2} \ldots s_{2n+1} s_1 s_2 \ldots s_r$$

where all paths start at the origin. Let $B = \{P_0, \ldots, P_{2n}\}$. So to show that B has the correct cardinality, we must prove that the P_i are all distinct. Suppose to the contrary that two are equal. By renumbering if necessary, we can assume that $P_0 = P_j$ for some $1 \leq j \leq 2n$. Take j be be minimum. Iterating this equality, we get $P_0 = P_j = P_{2j} = \ldots$. These equalities and the fact that j is as small as possible imply that $P = P_0$ is the concatenation of $P' : s_1 \ldots s_j$ with itself, say k times for some $k \geq 2$. Suppose P' ends at height h. Then P must end at height kh and so $kh = -1$. This forces $k = 1$, which is a contradiction.

To finish the proof, we must show that the blocks of ρ are in bijection with the paths in $\tilde{\mathcal{D}}(n)$. Let $\tilde{\mathcal{D}}'(n)$ denote the set of paths obtained by appending a down step to each path in $\tilde{\mathcal{D}}(n)$. So ρ partitions $\mathcal{P} \supseteq \tilde{\mathcal{D}}'(n)$. Thus it suffices to show that there is a unique path from $\tilde{\mathcal{D}}'(n)$ in each block B of ρ. Let B be generated by rotating a path P as in the previous paragraph and let $P : v_0 \ldots v_{2n+1}$ be the lattice point representation of P. Let h be the minimum height of a vertex of P, and among all vertices of P of height h let v_r be the left most. We claim that $P_r \in \tilde{\mathcal{D}}'(n)$ and no other P_s is in this set for $s \in \{0, 1, \ldots, n\} - \{r\}$. We will prove the first of these two claims and leave the second, whose demonstration is similar, as an exercise. Since v_r is translated to the origin and has smallest height in P, the translations of all v_i for $i \geq r$ will lie weakly above the x-axis. As for the v_i with $i < r$, they must be translated so the v_r becomes the last vertex of P_r which is of height -1. But since v_r was the first vertex of minimum height in P, the vertices before it must be translated to have height greater than -1 and so must also lie weakly above the x-axis. It follows that only the last vertex of P_r is below the x-axis, which is what we wished to prove. $\square$

1.12. Pattern avoidance

Pattern avoidance is a relatively recent area of study in combinatorics. It has seen strong growth in part because of its connections to algebraic geometry and computer science. For more information about this topic, see the books of Bóna [**18**] or Kitaev [**48**].

Let S be a set of integers with $\#S = k$ and consider a permutation $\sigma \in P(S)$. The *standardization* of σ is the permutation $\operatorname{std} \sigma \in P([k])$ obtained by replacing the smallest element of σ by 1, the next smallest by 2, and so on. For example, if $\sigma = 263$, then $\operatorname{std} \sigma = 132$. Given $\sigma \in \mathfrak{S}_n$ and $\pi \in \mathfrak{S}_k$ in one-line notation, we say that σ *contains a copy* of π if there is a subsequence σ' of σ such that $\operatorname{std} \sigma' = \pi$. Note that a subsequence need not consist of consecutive elements of π. In this case, π is called the *pattern*. To illustrate, $\sigma = 425613$ contains the pattern $\pi = 132$ since $\sigma' = 263$ standardizes to π. On the other hand, we say that σ *avoids* π if it has no subsequence σ' with $\operatorname{std} \sigma = \pi$. Continuing our example, one can check that σ avoids 4321 since σ does not contain a decreasing subsequence of length four. There is an equivalent

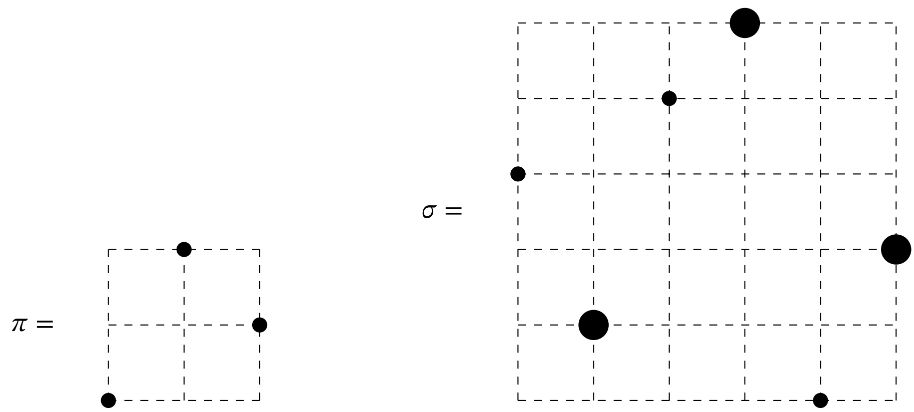


Figure 1.11. The diagrams for $\pi = 132$ and $\sigma = 425613$

definition of pattern containment which the reader will see in the literature. If S, T are sets with $\#S = \#T = k$, then call $\sigma = \sigma_1 \ldots \sigma_k \in P(S)$ and $\tau = \tau_1 \ldots \tau_k \in P(T)$ *order isomorphic* if $\sigma_i < \sigma_j$ is equivalent to $\tau_i < \tau_j$ for all i, j. It is easy to see that σ contains a copy of π if and only if σ contains a subsequence order isomorphic to π.

To study patterns, it will be useful to have a geometric model of a permutation analogous to its permutation matrix. Again, the integer lattice will come into play. Given $\sigma = \sigma_1 \ldots \sigma_n \in \mathfrak{S}_n$, its *diagram* is the set of points $(i, \sigma_i) \in \mathbb{Z}^2$ for $1 \le i \le n$. In displaying the diagram, the lower-left corner is always assumed to have coordinates $(1, 1)$. Using our running example, the diagrams for $\pi = 132$ and $\sigma = 425613$ are shown in Figure 1.11. The points corresponding to the copy 263 of π in σ have been enlarged to emphasize how easily one can see pattern containment using diagrams.

From an enumerative point of view, avoidance often turns out to be easier to work with than containment. So given $\pi \in \mathfrak{S}_k$, we consider

$$\mathrm{Av}_n(\pi) = \{\sigma \in \mathfrak{S}_n \mid \sigma \text{ avoids } \pi\}.$$

Note that many authors use $\mathfrak{S}_n(\pi)$ instead of $\mathrm{Av}_n(\pi)$ for this set. Call π and π' *Wilf equivalent*, written $\pi \equiv \pi'$, if $\# \mathrm{Av}_n(\pi) = \# \mathrm{Av}_n(\pi')$ for all $n \ge 0$. It is easy to see that this is an equivalence relation on $\mathfrak{S}_n$. We will prove that any two permutations in $\mathfrak{S}_3$ are Wilf equivalent, although this is not as startling as it might first sound.

Certain Wilf equivalences follow easily from manipulation of diagrams. Consider the *dihedral group of the square*

$$D = \{\rho_0, \rho_{90}, \rho_{180}, \rho_{270}, r_0, r_1, r_{-1}, r_\infty\} \tag{1.11}$$

where ρ_θ is rotation by θ degrees counterclockwise and r_m is reflection across a line of slope m. If σ contains a copy σ' of π and $f \in D$, then $f(\sigma)$ contains a copy $f(\sigma')$ of $f(\pi)$. Using f^{-1}, we see that the converse of the previous assertion is also true. It follows that σ avoids π if and only if $f(\sigma)$ avoids $f(\pi)$. We have proven the following result.

Lemma 1.12.1. *For any $\pi \in \mathfrak{S}_k$ and any $f \in D$ we have $\pi \equiv f(\pi)$.* □

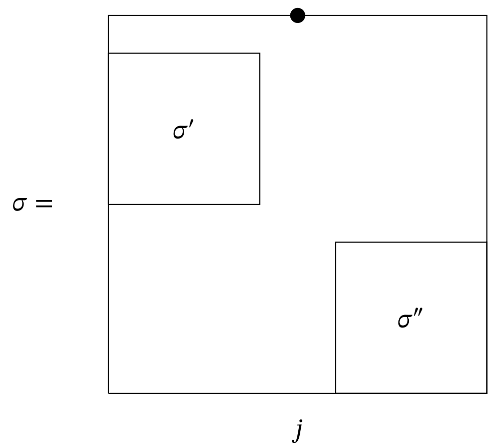


Figure 1.12. Decomposing $\sigma \in \mathrm{Av}_n(132)$

The equivalences in this lemma are called *trivial Wilf equivalences*. In particular, in $\mathfrak{S}_3$ one sees by repeatedly applying ρ_{90} that $132 \equiv 231 \equiv 213 \equiv 312$ and $123 \equiv 321$. In fact, all six permutations are Wilf equivalent and their avoidance sets are counted by the Catalan numbers. We start with 132 from the first set of equivalent permutations.

Theorem 1.12.2. *For $n \geq 0$ we have*

$$\#\,\mathrm{Av}_n(132) = C(n).$$

Proof. We will induct on n, using the initial condition and recurrence relation for $C(n)$ given in Theorem 1.11.2. As usual, we concentrate on the latter. Pick $\sigma = \sigma_1 \ldots \sigma_n \in \mathrm{Av}_n(132)$ and suppose $\sigma_j = n$. So we can write $\sigma = \sigma' n \sigma''$ where $\sigma' = \sigma_1 \ldots \sigma_{j-1}$ and $\sigma'' = \sigma_{j+1} \ldots \sigma_n$. Clearly σ' and σ'' must avoid 132 since they are subsequences of σ. We also claim that $\min \sigma' > \max \sigma''$ so that we can think of the diagram of σ decomposing as in Figure 1.12. Indeed, if there is $s \in \sigma'$ and $t \in \sigma''$ with $s < t$, then σ contains snt, which is a copy of 132, a contradiction. Thus σ' and σ'' are permutations of $\{n-1, n-2, \ldots, n-j+1\}$ and $[n-j]$, respectively, both of which avoid 132. Conversely, if the diagram of σ has the form given in Figure 1.12 with σ', σ'' avoiding 132, then σ must avoid 132. This is a case-by-case proof by contradiction, considering where the elements of a copy of 132 could lie in the diagram if one existed. We leave the details to the reader. To finish the count, from what we have shown and induction there are $C(j-1)$ choices for σ' and $C(n-j)$ for σ''. Taking their product and summing over $j \in [n]$ shows that there are $C(n)$ choices for σ. □

Next we will tackle 123, but to do so we will need some new concepts. The *left-right minima* of $\sigma = \sigma_1 \ldots \sigma_n \in \mathfrak{S}_n$ are the σ_i satisfying $\sigma_i < \min\{\sigma_1, \sigma_2, \ldots, \sigma_{i-1}\}$. For example $\sigma = 698371542$ has left-right minima $\sigma_1 = 6$, $\sigma_4 = 3$, and $\sigma_6 = 1$. The indices i such that σ_i is a left-right minimum are called the left-right minimum *positions*. If necessary to distinguish from the positions, the σ_j themselves are called the left-right minimum *values*. Reading the left-right minima in order from left to right, the positions and values always satisfy

$$1 = i_1 < i_2 < \cdots < i_l \quad \text{and} \quad m_1 > m_2 > \cdots > m_l = 1 \tag{1.12}$$

for some $l \geq 1$.

We will need to determine, given a set of values and positions, whether a permutation exists with left-right minima having these values and positions. To do this, we introduce the dominance order on compositions, which is also useful in other areas of combinatorics and representation theory. A *weak composition* of n is a sequence $\alpha = [\alpha_1, \ldots, \alpha_l]$ of nonnegative integers with $\sum_i \alpha_i = n$. So in weak compositions zero is permitted as a part and we will use 0 subscripts on notation for compositions when used for weak compositions. If $\alpha, \beta \vDash_0 n$, then α is *dominated* by β, written $\alpha \trianglelefteq \beta$, if we have

$$\alpha_1 + \alpha_2 + \cdots + \alpha_j \leq \beta_1 + \beta_2 + \cdots + \beta_j$$

for all $j \geq 1$, where $\alpha_j = 0$ if $j > \ell(\alpha)$ and similarly for β. To illustrate, $[2,2,1,1] \trianglelefteq [3,1,2]$ because $2 \leq 3$, $2+2 \leq 3+1$, $2+2+1 \leq 3+1+2$, and $2+2+1+1 = 3+1+2+0$. Since $\alpha, \beta \vDash_0 n$ the last inequality always becomes an equality. In the next result, the reader will notice a similarity between the construction of ι and μ and the map ϕ defined by (1.8).

Lemma 1.12.3. *Let $\sigma \in \mathfrak{S}_n$.*

(a) *We have $\sigma \in \mathrm{Av}_n(123)$ if and only if its subsequence of non-left-right minima is decreasing.*

(b) *There exists $\sigma \in \mathrm{Av}_n(123)$ with left-right minima positions and values given by* (1.12) *if and only if $\iota \trianglelefteq \mu$ where*

$$\iota = (i_2 - i_1 - 1,\ i_3 - i_2 - 1,\ \ldots,\ i_{l+1} - i_l - 1),$$
$$\mu = (m_0 - m_1 - 1,\ m_1 - m_2 - 1,\ \ldots,\ m_{l-1} - m_l - 1),$$

and $i_{l+1} = m_0 = n + 1$. In this case, σ is unique.

Proof. (a) We will prove this statement in its contrapositive form. Suppose first that σ contains a copy $\sigma_i\sigma_j\sigma_k$ of 123. Then σ_j, σ_k cannot be left-right minima since σ_i is smaller than both and to their left in σ. Since $\sigma_j < \sigma_k$, the non-left-right minima subsequence contains an increase. Conversely, suppose $\sigma_j < \sigma_k$ with $j < k$ and both non-left-right minima. Let σ_i be the left-right minimum closest to σ_j on its left. We have that σ_i exists since σ begins with a left-right minimum. Then $\sigma_i < \sigma_j < \sigma_k$, giving a copy of 123.

(b) Clearly if σ exists, then it must be unique since the positions and values of its left-right minima are given by (1.12) and the rest of the elements can only be arranged in one way by (a). We can attempt to build σ satisfying the given conditions as follows. An example will be found following the proof. Start with a row of n blank positions. Now fill in the values $m_1 > \cdots > m_l$ at the positions $i_1 < \cdots < i_l$. Filling in the rest of the positions with the elements of $S = [n] - \{m_1, \ldots, m_l\}$ (the set of non-left-right minima) in decreasing order gives a σ avoiding 123 since σ is a union of two decreasing subsequences. So the only question is whether doing this will result in a permutation having the m_j as its left-right minima. We have that m_1 is always a left-right minimum regardless of the other entries. Now m_{j+1} will be the next left-right minimum after m_j if and only if the blanks before position i_{j+1} are filled with elements larger than m_j. Note that $\iota_j = i_{j+1} - i_j - 1$ is the number of spaces between positions i_j and i_{j+1}. Also $\mu_j = m_{j-1} - m_j - 1$ is the number of $s \in S$ with $m_j < s < m_{j-1}$. It follows that $\iota_1 + \cdots + \iota_j$ is the number of blanks before position i_{j+1} and $\mu_1 + \cdots + \mu_j$ is the number of elements of S greater than m_j. So filling in the spaces preserves the left-right minima if and only if the inequalities for $\iota \trianglelefteq \mu$ are satisfied. This completes the proof. $\square$

Suppose we want to see if there is $\sigma \in \mathrm{Av}_9(123)$ with left-right minima $6 > 3 > 1$ in positions $1 < 4 < 6$. We start off with the diagram

$$\sigma = 6\,\_\,\_\,3\,\_\,1\,\_\,\_\,\_. \tag{1.13}$$

We wish to check whether filling the blanks with the remaining elements of $[9]$ in decreasing order will result in a permutation which has the initial elements as left-right minima. One way to do this is just to fill the blanks and verify that the desired elements become left-right minima: $\sigma = 6\,9\,8\,3\,7\,1\,5\,4\,2$. Another way is to use the ι and μ compositions. Note that $\iota_1 = 4 - 1 - 1 = 2$ is the number of blanks between $m_1 = 6$ and $m_2 = 3$ in the original diagram. Similarly $\mu_1 = 10 - 6 - 1 = 3$ is the number of elements of $S = [9] - \{6, 3, 1\}$ greater than $m_1 = 6$. In order to fill the blanks between 6 and 3 so that 6 is a left-right minimum, the numbers used must all be greater than 6. This is possible exactly when $\iota_1 \le \mu_1$. Similarly $\iota_1 + \iota_2 \le \mu_1 + \mu_2$ ensures that one can fill the blanks to the left of $m_3 = 1$ with numbers greater than $m_2 = 3$, and so forth. So checking whether $\iota \trianglelefteq \mu$ also determines whether σ has the correct left-right minima.

We will need an analogue of Lemma 1.12.3 for elements of $\mathrm{Av}_n(132)$. To state it, we define the *reversal* of a weak composition $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_l]$ to be

$$\alpha^r = [\alpha_l, \alpha_{l-1}, \ldots, \alpha_1].$$

Lemma 1.12.4. *Let $\sigma \in \mathfrak{S}_n$.*

(a) *We have $\sigma \in \mathrm{Av}_n(132)$ if and only if, for every left-right minimum m, the elements of σ to the right of and greater than m form an increasing subsequence.*

(b) *There exists $\sigma \in \mathrm{Av}_n(132)$ with left-right minima positions and values given by* (1.12) *if and only if $\mu^r \trianglelefteq \iota^r$ where ι, μ are as given in Lemma* 1.12.3. *In this case, σ is unique*

Proof. Much of the proof of this result is similar to the demonstration of Lemma 1.12.3 and so will be left as an exercise. Here we will only present the construction of $\sigma \in \mathrm{Av}_n(132)$ from its diagram of left-right minima and blanks. Again, an example follows the explanation. We keep the notation of the proof of the previous lemma. We start by filling the blanks to the right of $m_l = 1$ with the elements $s \in S$ such that $m_l < s < m_{l-1}$ in increasing order and as far left as possible (so they will be consecutive). Next we fill in the remaining blanks to the right of m_{l-1} with those $s \in S$ such that $m_{l-1} < s < m_{l-2}$ so that they form an increasing subsequence which is as far left as possible given the spaces already filled. Continue in this manner until all blanks are occupied. $\square$

Suppose we wish to fill in the diagram (1.13) so that σ avoids 132. If $m_3 < s < m_2$, then $s = 2$, so we put 2 just to the right of $m_3 = 1$ to get $\sigma = 6\ \_\ \_\ 3\ \_\ 1\ 2\ \_\ \_$. Similarly, $m_1 < s < m_2$ is satisfied by $s = 4, 5$ so we put these elements following $m_2 = 3$ in increasing order using the left-most blanks available to obtain $\sigma = 6\ \_\ \_\ 3\ 4\ 1\ 2\ 5\ \_$. Finally, we do the same for the elements greater than $m_1 = 6$ to get the end result $\sigma = 6\,7\,8\,3\,4\,1\,2\,5\,9$.

We need one last observation before we achieve our goal of showing all elements of $\mathfrak{S}_3$ are Wilf equivalent. Suppose $\alpha = [\alpha_1, \ldots, \alpha_l]$ and $\beta = [\beta_1, \ldots, \beta_l]$ are weak compositions of n. We claim $\alpha \trianglelefteq \beta$ if and only if $\beta^r \trianglelefteq \alpha^r$. To see this, note that the inequality $\alpha_1 + \cdots + \alpha_j \le \beta_1 + \cdots + \beta_j$ is equivalent to $n - (\beta_1 + \cdots + \beta_j) \le n - (\alpha_1 + \cdots + \alpha_j)$. But $n - (\alpha_1 + \cdots + \alpha_j) = \alpha_r + \alpha_{r-1} + \cdots + \alpha_{j+1}$ and similarly for β. Making this substitution we get the necessary inequalities for $\beta^r \trianglelefteq \alpha^r$ and all steps are reversible. Finally, we say that a bijection $f : S \to T$ *preserves property* P if $s \in S$ having property P is equivalent to $f(s)$ having property P for all $s \in S$.

Theorem 1.12.5. *For $n \ge 0$ and any $\pi \in \mathfrak{S}_3$ we have*

$$\#\operatorname{Av}_n(\pi) = C(n).$$

Proof. By Theorem 1.12.2 and the discussion just before it, it suffices to show that we have $\#\operatorname{Av}_n(123) = C(n)$. This will be true if we can find a bijection $f : \operatorname{Av}_n(123) \to \operatorname{Av}_n(132)$. In fact, f will preserve the values and positions of left-right minima. Suppose $\sigma \in \operatorname{Av}_n(123)$ has its positions and values given by (1.12). By Lemma 1.12.3 there is a unique such σ and we must also have $\iota \trianglelefteq \mu$. But, as noted just before this theorem, this is equivalent to $\mu^r \trianglelefteq \iota^r$. So, using Lemma 1.12.4, there is a unique $\sigma' \in \operatorname{Av}_n(132)$ having the given positions and values of its left-right minima and we let $f(\sigma) = \sigma'$. Because of the existence and uniqueness of σ and σ', this is a bijection. $\square$

Note that the description of f in the previous proof can be made constructive. Given $\sigma \in \operatorname{Av}_n(123)$, we remove its non-left-right minima and rearrange them using the algorithm in the proof of Lemma 1.12.4. So, using our running example, $f(698371542) = 678341259$.

Exercises

(1) Prove each of the following identities for $n \ge 1$ in two ways: one inductive and one combinatorial.

(a) $\displaystyle\sum_{i=1}^{n} F_i = F_{n+2} - 1.$

(b) $\displaystyle\sum_{i=1}^{n} F_{2i} = F_{2n+1} - 1.$

(c) $\displaystyle\sum_{i=1}^{n} F_{2i-1} = F_{2n}.$

(2) Prove that if $k, n \in \mathbb{P}$ with $k \mid n$ (meaning k divides evenly into n), then $F_k \mid F_n$.

(3) Given $m \in \mathbb{P}$, show that the sequence of Fibonacci numbers is periodic modulo m; that is, there exists $p \in \mathbb{P}$ such that

$$F_{n+p} \equiv F_n \pmod{m}$$

for all $n \geq 0$. The *period modulo m* is the smallest p such that this congruence holds. Note that it is an open problem to find the period of the Fibonacci sequence for an arbitrary m.

(4) The *Lucas numbers* are defined by $L_0 = 2$, $L_1 = 1$, and

$$L_n = L_{n-1} + L_{n-2} \text{ for } n \geq 2.$$

Prove the following identities for $m, n \geq 1$.

(a) $L_n = F_{n-1} + F_{n+1}$.

(b) Let $\mathcal{C}_n$ be the set of tilings of n boxes arranged in a circle with dominos and monominos. Show that $\#\mathcal{C}_n = L_n$.

(c) $L_{m+n} = F_{m-1}L_n + F_mL_{n+1}$.

(d) $F_{2n} = F_nL_n$.

(5) Prove Theorem 1.2.2.

(6) Check that the two maps defined in the proof of Theorem 1.3.1 are inverses.

(7) (a) Prove Theorem 1.3.3(b) using equation (1.5).

(b) Give an inductive proof of Theorem 1.3.3(c).

(c) Give an inductive proof of Theorem 1.3.3(d).

(8) Let S, T be sets.

(a) Show that $S \,\Delta\, T = (S \cup T) - (S \cap T)$.

(b) Show that $(S \,\Delta\, T) \,\Delta\, T = S$.

(9) Given nonnegative integers satisfying $n_1 + n_2 + \cdots + n_m = n$. the corresponding *multinomial coefficient* is

$$\binom{n}{n_1, n_2, \ldots, n_m} = \frac{n!}{n_1!\; n_2!\; \ldots\; n_m!}. \tag{1.14}$$

We extend this definition to negative n_i by letting the multinomial coefficient be zero if any $n_i < 0$. Note that when $m = 2$ we recover the binomial coefficients as

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

(a) Find and prove analogues of Theorem 1.3.3(a), (b), and (c) for multinomial coefficients.

(b) A *permutation* of a multiset $M = \{\{1^{n_1}, 2^{n_2}, \ldots, m^{n_m}\}\}$ is a linear arrangement of the elements of M. Let $P(M)$ denote the set of permutations of M. For example

$$P(\{\{1^2, 2^2\}\}) = \{1122,\ 1212,\ 1221,\ 2112,\ 2121,\ 2211\}.$$

$$
\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 0 & & 1 & & & & & \\
 & & & & 1 & & 1 & & 1 & & 1 & & & & \\
 & & & 1 & & 0 & & 0 & & 0 & & 1 & & & \\
 & & 1 & & 1 & & 0 & & 0 & & 1 & & 1 & & \\
 & 1 & & 0 & & 1 & & 0 & & 1 & & 0 & & 1 & \\
1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1
\end{array}
$$

Figure 1.13. Pascal's triangle modulo 2

Prove that

$$\#P(\{\{1^{n_1}, 2^{n_2}, \ldots, m^{n_m}\}\}) = \binom{n}{n_1, n_2, \ldots, n_m}$$

in three ways:

(i) combinatorially,

(ii) by induction on n,

(iii) by proving that

$$\binom{n}{n_1, n_2, \ldots, n_m} = \binom{n}{n_1}\binom{n - n_1}{n_2, \ldots, n_m}$$

and then inducting on m.

(10) (a) Prove the Pascal triangle is fractal modulo 2. Specifically, if one replaces each binomial coefficient by its remainder on division by 2, then, for any $k \geq 0$, the triangle consisting of rows 0 through $2^k - 1$ is repeated on the left and on the right in rows 2^k through $2^{k+1} - 1$ with an inverted triangle of zeros in between. See Figure 1.13 for the first eight rows. Hint: Induct on k.

(b) Formulate and prove an analogous result modulo p for any prime p.

(11) Find the inverse for the map in the proof of Theorem 1.3.4, proving that it is well-defined and the inverse to the given function.

(12) For $n \geq 0$ define the nth *Fibotorial* to be the product $F_n^! = F_1 F_2 \ldots F_n$. Also, for $0 \leq k \leq n$ define a *Fibonomial coefficient* by

$$\binom{n}{k}_F = \frac{F_n^!}{F_k^! F_{n-k}^!}.$$

Note that from this definition it is not clear that this is an integer.

(a) Show that the Fibonomial coefficients satisfy the initial conditions $\binom{n}{0}_F = \binom{n}{n}_F = 1$ and recurrence

$$\binom{n}{k}_F = F_{n-k+1}\binom{n-1}{k-1}_F + F_{k-1}\binom{n-1}{k}_F$$

for $0 < k < n$.

(b) Show that $\binom{n}{k}_F$ is an integer for all $0 \le k \le n$.
(c) Find a combinatorial interpretation of $\binom{n}{k}_F$.

(13) For $n \ge 1$ show that the Stirling numbers of the second kind have the following values.
(a) $S(n, 1) = 1$.
(b) $S(n, 2) = 2^{n-1} - 1$.
(c) $S(n, n) = 1$.
(d) $S(n, n-1) = \binom{n}{2}$.
(e) $S(n, n-2) = \binom{n}{3} + 3\binom{n}{4}$.

(14) For $n \ge 1$ show that the signless Stirling numbers of the first kind have the following values.
(a) $c(n, 1) = (n-1)!$.
(b) $c(n+1, 2) = n! \sum_{i=1}^{n} \frac{1}{i}$.
(c) $c(n, n) = 1$.
(d) $c(n, n-1) = \binom{n}{2}$.
(e) $c(n, n-2) = 2\binom{n}{3} + 3\binom{n}{4}$.

(15) Call an integer partition λ *self-conjugate* if $\lambda^t = \lambda$. Show that the number of self-conjugate $\lambda \vdash n$ equals the number of $\mu \vdash n$ having parts which are distinct (no part can be repeated) and odd. Hint: Use Young diagrams and try to guess a bijection inductively by first seeing what it has to be for small n. Then try to construct a bijection for $n+1$ which will be consistent in some way with the one for previous values. Finally try to describe your bijection in a noninductive manner.

(16) The *main diagonal* of a Young diagram is the set of squares starting with the one at the top left and moving diagonally right and down. So in Figure 1.5, the main diagonal of λ consists of two squares. Prove the following.
(a) If λ is self-conjugate as defined in the previous exercise, then $|\lambda| \equiv d \pmod 2$ where d is the length (number of squares) of the main diagonal.
(b) Let $p_d(n)$ be the number of partitions of n whose diagonal has length d. Then

$$p_d(n) = \sum_{m \ge 0} p(m, d)p(n - m - d^2, d).$$

(17) Define $p_e(n, k)$ to be the number of $\lambda \vdash n$ having exactly k parts. Prove the following under the assumption that $n \ge 4$, where $\lfloor \cdot \rfloor$ is the round-down function.
(a) $p_e(n, k) = p(n - k, k)$.
(b) $p_e(n, 1) = 1$.
(c) $p_e(n, 2) = \lfloor n/2 \rfloor$.
(d) $p_e(n, n-2) = 2$.
(e) $p_e(n, n-1) = 1$.
(f) $p_e(n, n) = 1$.

(18) Finish the proof of Theorem 1.7.1.

(19) Prove Theorem 1.7.2.

(20) Consider a line of n copies of the integer 1. One can now put slashes in the spaces between the 1's and count the number of 1's between each pair of adjacent slashes to form a composition of n. For example, if $n = 6$, then we start with 1 1 1 1 1 1. One way of inserting slashes is 1 1/1/1 1 1, which corresponds to the composition $2 + 1 + 3 = 6$. Give alternate proofs of Theorems 1.7.1 and 1.7.2 using this idea.

(21) A *weak composition* of n into k parts is a sequence of k nonnegative integers summing to n. Find a formula for the number of weak compositions of n into k parts and then prove it in three different ways:
  (a) by using a variant of the map ϕ defined in (1.8),
  (b) by finding a relation between weak compositions and compositions and then using the statement of Theorem 1.7.2 (as opposed to their proofs as in part (a)),
  (c) by modifying the construction in the previous exercise.

(22) Show that the last two columns in Table 1.1 agree when f is bijective, that is, when $n = k$.

(23) Prove Lemma 1.9.1(b).

(24) A graph $G = (V, E)$ is *regular* if all of its vertices have the same degree. If $\deg v = r$ for all vertices v, then G is *regular of degree* r.
  (a) Show that if G is regular of degree r, then
$$|E| = \frac{r|V|}{2},$$
  (b) Call G *bipartite* if there is a set partition of $V = V_1 \uplus V_2$ such for all $uv \in E$ we have $u \in V_1$ and $v \in V_2$ or vice versa. Show that a bipartite graph regular of degree $r \geq 1$ has $|V_1| = |V_2|$.

(25) A graph G is *planar* if it can be drawn in the plane $\mathbb{R}^2$ without edge crossings. In this case the *regions* of G are the topologically connected components of the set-theoretic difference $\mathbb{R}^2 - G$. Let R be the set of regions of G. If $r \in R$, then let $\deg r$ be the number of edges on the boundary of r. Show that
$$\sum_{r \in R} \deg r \leq 2|E|.$$
Find, with proof, a condition on the cycles of G which is equivalent to having equality.

(26) Two graphs G, H are *isomorphic*, written $G \cong H$, if they are equal as unlabeled graphs. The *complement* of a graph $G = (V, E)$ is the graph $\bar{G}$ with vertices V and with uv an edge of $\bar{G}$ if and only if $uv \notin E$. Call G *self-complementary* if $G \cong \bar{G}$.
  (a) Show that there exists a self-complementary graph with n vertices if and only if $n \equiv 0 \pmod 4$ or $n \equiv 1 \pmod 4$.
  (b) Show that in a self-complementary graph with n vertices where $n \equiv 1 \pmod 4$ there must be at least one vertex of degree $(n-1)/2$. Hint: Show that the number of vertices of degree $(n-1)/2$ must be odd.

(27) Prove Theorem 1.9.4.

(28) Prove that in any digraph $D = (V, A)$ we have
$$\sum_{v \in V} \operatorname{ideg} v = \sum_{v \in V} \operatorname{odeg} v = |A|.$$

(29) Prove the equivalence of (a) and (d) in Theorem 1.10.2. Hint: Use Lemma 1.9.1(b).

(30) Consider a sequence of nonnegative integers $d : d_1, \ldots, d_n$. Call d a *degree sequence* if there is a graph with vertices $v_1, \ldots, v_n$ such that $\deg v_i = d_i$ for all i.
  (a) Let T be a tree with n vertices and arrange its degree sequence in weakly decreasing order. Prove that for $1 \le i \le n$ we have
$$d_i \le \left\lceil \frac{n-1}{i} \right\rceil.$$
  (b) Let T be a tree with n vertices and let $k \ge 2$ be an integer. Suppose the degree sequence of T satisfies $d_i = 1$ or k for all i. Prove that $d_i = k$ for exactly $(n-2)/(k-1)$ indices i.

(31) Finish the proof of Theorem 1.10.3.

(32) Consider n cars $C_1, \ldots, C_n$ passing in this order by a line of n parking spaces numbered $1, \ldots, n$. Each car C_i has a preferred space number c_i in which to park. If C_i gets to space c_i and it is free, then it parks. Otherwise it proceeds to the next empty space (which will have a number greater than c_i) and parks there if such a space exists. If no such space exists, it does not park. Call $c = (c_1, \ldots, c_n)$ a *parking function of length n* if all the cars end up in a parking space.
  (a) Show that c is a parking function if and only if its unique weakly increasing rearrangement $d = (d_1, \ldots, d_n)$ satsifies $d_i \le i$ for all $i \in [n]$.
  (b) Use a counting argument to show that the number of parking functions of length n is $(n+1)^{n-1}$. Hint: Consider parking where there are $n+1$ spaces arranged in a circular manner and $n+1$ is an allowed preference for cars.
  (c) Reprove (b) by finding a bijection between parking functions of length n and trees on $n+1$ vertices. Hint: Let T be a tree on $n+1$ vertices labeled $0, \ldots, n$ and call vertex 0 the *root* of the tree. Draw T in the plane so that the vertices connected to the root, called the root's *children*, are in increasing order read left to right. Continue to do the same thing for the children of each child of the root, and so forth. Create a permutation π by reading the children of the root left to right, then the grandchildren of the root left to right, etc. Finally, orient each edge of T so that it points from a vertex to its parent and call this set of arcs A. Map T to $c = (c_1, \ldots, c_n)$ where
$$c_i = \begin{cases} 1 & \text{if } \vec{\imath 0} \in A, \\ 1 + j & \text{if } \vec{\imath \pi_j} \in A. \end{cases}$$

(33) Consider *EW-lattice paths* along the x-axis which are paths starting at the origin and using steps $E = [1, 0]$ and $W = [-1, 0]$.
  (a) Show that if an EW-lattice path has length n and ends at $(k, 0)$, then n and k have the same parity and $|k| \le n$.
  (b) Show that the number of EW-lattice paths of length n ending at $(k, 0)$ is
$$\binom{n}{\frac{n+k}{2}}.$$

(c) Show that the number of EW-lattice paths of length $2n$ ending at the origin and always staying on the nonnegative side of the axis is $C(n)$.

(34) Show that the Catalan numbers $C(n)$ also count the following objects:

(a) ballot sequences which are words $w = w_1 \ldots w_{2n}$ containing n ones and n twos such that in any prefix $w_1 \ldots w_i$ the number of ones is always at least as great as the number of twos,

(b) sequences of positive integers

$$1 \leq a_1 \leq a_2 \leq \ldots \leq a_n$$

with $a_i \leq i$ for $1 \leq i \leq n$,

(c) triangulations of a convex $(n+2)$-gon using nonintersecting diagonals,

(d) noncrossing partitions $\rho = B_1/\ldots/B_k \vdash [n]$ where a crossing is $a < b < c < d$ such that $a, c \in B_i$ and $b, d \in B_j$ for $i \neq j$.

(35) Fill in the details of the proof of Theorem 1.11.3.

(36) A *stack* is a first-in first-out (FIFO) data structure with two operations. One can put something on the top of a stack, called *pushing*, or take something from the top of the stack, called *popping*. A permutation $\sigma = \sigma_1 \ldots \sigma_n \in \mathfrak{S}_n$ is considered *sorted* if its elements have been rearranged to form the permutation $\tau = 12 \ldots n$. Consider the following algorithm for sorting σ. Start with an empty stack and an empty output permutation τ. At each stage there are two options. If the stack is empty or the current first element s of σ is smaller than the top element of the stack, then one pushes s onto the stack. If σ has become empty or the top element t of the stack is smaller than the first element of σ, then one pops t from the stack and appends it to the end of τ. An example showing the sorting of $\sigma = 3124$ will

τ	stack	σ
ϵ	ϵ	3124
ϵ	3	124
	1	
ϵ	3	24
1	3	24
	2	
1	3	4
12	3	4
123	ϵ	4
123	4	ϵ
1234	ϵ	ϵ

Figure 1.14. A stack-sorting algorithm

be found in Figure 1.14. Note that the input permutation σ is on the right and the output permutation τ is on the left so that the head of σ and the tail of τ are nearest the stack.

(a) Show that this algorithm sorts σ if and only if $\sigma \in \mathrm{Av}_n(231)$.
(b) Show that if there is a sequence of pushes and pops which sorts σ, then it must be the sequence given by the algorithm.

(37) Suppose $\pi = \pi_1 \ldots \pi_k \in \mathfrak{S}_k$. Prove the following descriptions of actions of elements of D in terms of one-line notation.

(a) $r_\infty(\pi) = \pi_k \ldots \pi_1 := \pi^r$, the *reversal* of π.
(b) $r_0(\pi) = (k+1-\pi_1)\ldots(k+1-\pi_k) := \pi^c$, the *complement* of π.
(c) $r_1(\pi) = \pi^{-1}$, the group-theoretic inverse of π.

(38) Finish the proof of Theorem 1.12.2.

(39) Given any set of permutations Π we let

$$\mathrm{Av}_n(\Pi) = \{\sigma \in \mathfrak{S}_n \mid \sigma \text{ avoids all } \pi \in \Pi\}.$$

If $\pi = \pi_1\pi_2\ldots\pi_m \in \mathfrak{S}_m$ is a permutation and $n \in \mathbb{N}$, then we can construct a new permutation

$$\pi + n = \pi_1 + n, \pi_2 + n, \ldots, \pi_m + n.$$

Given permutations π, σ of disjoint sets, we denote by $\pi\sigma$ the permutation obtained by concatenating them. Define two other concatenations on $\pi \in \mathfrak{S}_m$ and $\sigma \in \mathfrak{S}_n$, the *direct sum*

$$\pi \oplus \sigma = \pi(\sigma + m)$$

and *skew sum*

$$\pi \ominus \sigma = (\pi + n)\sigma.$$

Finally for $n \geq 0$ we use the notation

$$\iota_n = 12\ldots n$$

for the increasing permutation of length n, and

$$\delta_n = n \ldots 21$$

for the decreasing one. Prove the following.

(a) $\mathrm{Av}_n(213, 321) = \{\iota_{k_1} \oplus (\iota_{k_2} \ominus \iota_{k_3}) \mid k_1 + k_2 + k_3 = n\}$.
(b) $\mathrm{Av}_n(132, 213) = \{\iota_{k_1} \ominus \iota_{k_2} \ominus \cdots \mid \sum_i k_i = n\}$.
(c) $\mathrm{Av}_n(132, 213, 321) = \{\iota_{k_1} \ominus \iota_{k_2} \mid k_1 + k_2 = n\}$.
(d) $\mathrm{Av}_n(132, 231, 312) = \{\delta_{k_1} \oplus \iota_{k_2} \mid k_1 + k_2 = n\}$.
(e) $\mathrm{Av}_n(132, 231, 321) = \{(1 \ominus \iota_{k_1}) \oplus \iota_{k_2} \mid k_1 + k_2 = n - 1\}$.
(f) $\mathrm{Av}_n(123, 132, 213) = \{\iota_{k_1} \ominus \iota_{k_2} \ominus \cdots \mid \sum_i k_i = n \text{ and } k_i \leq 2 \text{ for all } i\}$.

(40) Finish the proof of Lemma 1.12.4.

Chapter 2

Counting with Signs

In the previous chapter, we concentrated on counting formulae where all of the terms were positive. But there are interesting things to say when one permits negative terms as well. This chapter is devoted to some of the principal techniques which one can use in such a situation.

2.1. The Principle of Inclusion and Exclusion

The Principle of Inclusion and Exclusion, or PIE, is one of the classical methods for counting using signs. After presenting the Principle itself, we will give an application to derangements which are permutations having no fixed points.

In the Sum Rule, Lemma 1.1.1(a), we assumed that the sets S, T are disjoint. Of course, it is easy to see that for any finite sets S, T we have

$$|S \cup T| = |S| + |T| - |S \cap T|. \tag{2.1}$$

Indeed, $|S| + |T|$ counts $S \cap T$ twice and so to count it only once we must subtract the cardinality of the intersection. But one could ask if there is a similar formula for the union of any number of sets. It turns out that it is often more useful to consider these sets as subsets of some universal set S and count the number of elements in S which are not in any of the subsets, similar to the viewpoint used in pattern avoidance. To set up notation, let S be a set and let $S_1, \ldots, S_n \subseteq S$. We wish to find a formula for $|S - \bigcup_i S_i|$. When $n = 1$ we clearly have

$$|S - S_1| = |S| - |S_1|.$$

And for $n = 2$ equation (2.1) yields

$$|S - (S_1 \cup S_2)| = |S| - |S_1| - |S_2| + |S_1 \cap S_2|.$$

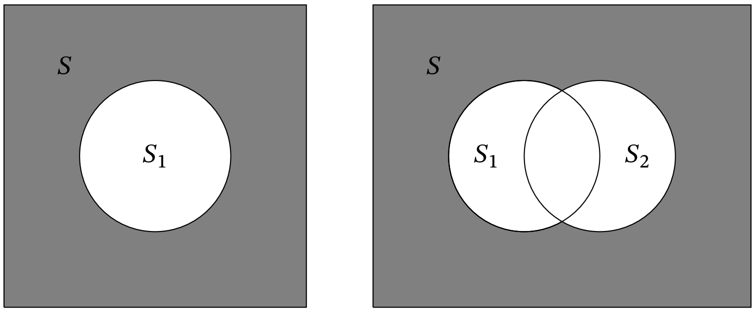


Figure 2.1. The PIE for $n = 1, 2$

Venn diagrams showing the shaded region counted for these two cases are given in Figure 2.1. The reader may have already guessed the generalization for arbitrary n. This type of enumeration where one alternately adds and subtracts cardinalities is sometimes called a *sieve*.

Theorem 2.1.1 (Principle of Inclusion and Exclusion, PIE). *If S is a finite set with subsets $S_1, \ldots, S_n$, then*

$$\left| S - \bigcup_{i=1}^{n} S_i \right| = |S| - \sum_{1 \le i \le n} |S_i| + \sum_{1 \le i < j \le n} |S_i \cap S_j| - \cdots + (-1)^n \left| \bigcap_{i=1}^{n} S_i \right|. \tag{2.2}$$

Proof. For any set S we have $|S| = \sum_{s \in S} 1$. We will use the notation $|S| = \sum_{s \in S} 1_s$ so that 1_s will keep track of the contribution of s to the sum. So it suffices to show that the coefficient of 1_s in the alternating sum is one if $s \notin \bigcup_i S_i$ and zero otherwise. In the first case, 1_s only occurs in $|S|$, giving the desired coefficient. In the second case, suppose $s \in S_i$ for exactly $m \ge 1$ indices i. Now $s \in S_{i_1} \cap \cdots \cap S_{i_k}$ precisely when $S_{i_1}, \ldots, S_{i_k}$ are k of the m subsets containing s. It follows that the number of summands 1_s in the sum for k-fold intersections is $\binom{m}{k}$. So the coefficient of 1_s to the right-hand side of (2.2) is

$$\binom{m}{0} - \binom{m}{1} + \binom{m}{2} - \cdots = 0$$

by Theorem 1.3.3(d). This completes the proof. $\square$

To simplify notation we will usually write just $\bigcup S_i$ for $\bigcup_{i=1}^{n} S_i$. We will also write S_I in place of $\bigcap_{i \in I} S_i$.

As an application of the PIE, we will count permutations without fixed points. This problem is sometimes accompanied by the following story. Suppose that n jolly revelers (and it is important that they be jolly) put their n identical bowler hats on a hat stand before dinner at a restaurant. During the meal, the hat stand gets overturned (I told you they were jolly) so that the hats, having no identifying markings, are returned at random when the revelers leave. What is the probability that no man gets his own hat back?

If one numbers the men $1, \ldots, n$ and similarly number the hats where hat i belongs to man i, then a way of returning the hats is just a permutation $\pi = \pi_1 \ldots \pi_n \in \mathfrak{S}_n$ where $\pi_i = j$ means that man i gets back hat j. So the condition that no man gets his own hat means that $\pi_i \neq i$ for all i; that is, π has no fixed points. Such a permutation is called a *derangement* and the number of derangements in $\mathfrak{S}_n$ is denoted $D(n)$ and is called the nth *derangement number*.

We now wish to set this problem up so that we can use the PIE. In particular, we want to define S and subsets $S_1, \ldots, S_n$ so that $D(n) = |S - \bigcup S_i|$. To do this, we think of the problem as counting a set of elements subject to certain restrictions and then let

(i) S be the set of objects with no restrictions and

(ii) $S_1, \ldots, S_n$ be subsets so that removing S_i from S imposes the ith restriction.

We will have chosen S and the S_i correctly if the cardinalities on the right-hand side of (2.2) can be computed. In the case under consideration, we want to count permutations with no fixed points. So we should let $S = \mathfrak{S}_n$, the set of all permutations without any restriction on their fixed points. We will also let S_i be the set of $\pi \in \mathfrak{S}_n$ with $\pi_i = i$ so that we will remove those permutations having i as a fixed point. Note that we do *not* choose subsets S_i' defined as the set of $\pi \in \mathfrak{S}_n$ with i fixed points, for if we did so, then the S_i' would be disjoint so that $|S - \bigcup S_i'| = |S| - |S_1'| - \cdots - |S_n'|$. Because of this, computing the cardinalities of the S_i' is about as hard as computing the cardinality of the set difference directly and so one does not gain anything. However, our original choice of subsets will turn out to be very nice.

We now compute the necessary cardinalities. Of course, $|S| = |\mathfrak{S}_n| = n!$. Next, if $\pi \in S_1$, then $\pi = 1\pi_2 \ldots \pi_n$ where $\pi_2 \ldots \pi_n$ form a permutation of $2, \ldots, n$. So $|S_1| = (n-1)!$. Clearly the same argument could be applied to any S_i, so

$$\sum_i |S_i| = n \cdot (n-1)! = n!\,.$$

Similarly, $S_1 \cap S_2 \cap \cdots \cap S_k$ is the set of all permutations of the form $\pi = 12 \ldots k\pi_{k+1} \ldots \pi_n$ and there are $(n-k)!$ ways to choose $\pi_{k+1}, \ldots, \pi_n$. In fact, all the terms in the k-fold sum have this value and there are $\binom{n}{k}$ such terms giving a total of

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

Summing up, so to speak, we have proved the following.

Theorem 2.1.2. *The nth derangement number is given by*

$$D(n) = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^n \frac{1}{n!}\right)$$

for $n \geq 0$. □

The reader should recognize the series in the previous result as a truncation of the series for $1/e$. Since the probability that no man gets his hat back is the number of ways this could happen over the total number of permutations for returning the hats, or $D(n)/n!$, we get a very pretty answer to the question originally posed.

Corollary 2.1.3. *In the limit as $n \to \infty$, the probability that no man gets his hat back is $1/e$.* □

It is striking that e, one of the quintessential transcendental numbers, should occur in the solution to a combinatorial problem which, at the outset, involves only integers.

2.2. Sign-reversing involutions

Sign-reversing involutions are a powerful way of proving identities involving signs, and even identities which do not explicitly have signs in them. As we will see, these maps can be used to prove the PIE itself and play an important role in the Garsia–Milne Involution Principle, which we will study in the next section.

Let S be a (not necessarily finite) set. A function $\iota : S \to S$ is an *involution* if ι^2 is the identity map on S. Equivalently, ι is a bijection such that $\iota^{-1} = \iota$. There is another nice characterization of involutions which will be crucial once we introduce signs. For any $f : S \to S$, its *fixed point set* is

$$\operatorname{Fix} f = \{s \in S \mid f(s) = s\}.$$

We also say that distinct elements $s, t \in S$ form a 2-*cycle* of f if $f(s) = t$ and $f(t) = s$. In this case we write (s, t) or $s \leftrightarrow t$ to denote the 2-cycle.

Lemma 2.2.1. *Consider $\iota : S \to S$. The function ι is an involution if and only if S is the disjoint union of the fixed points and 2-cycles of ι.*

Proof. For the forward direction, it suffices to show that if $s \in S$ is not a fixed point, then it is in a 2-cycle. So suppose $\iota(s) = t$. Then $\iota(t) = \iota^2(s) = s$ as desired.

Conversely, suppose that S is such a disjoint union and pick $s \in S$. If $s \in \operatorname{Fix}\iota$, then $\iota^2(s) = \iota(s) = s$. Otherwise, s is in a 2-cycle (s, t) so that $\iota^2(s) = \iota(t) = s$. So ι^2 is the identity map and we are done. □

A *signed set* is a set S together with a function $\operatorname{sgn} : S \to \{+1, -1\}$. In this case we let

$$S^+ = \{s \in S \mid \operatorname{sgn} s = +1\}$$

and similarly for S^-. If $\iota : S \to S$ is an involution on S, then we say that ι is *sign reversing* if $\operatorname{sgn}\iota(s) = -\operatorname{sgn} s$ for every s which is in a 2-cycle of ι. A pictorial representation of this situation will be found in Figure 2.2. Now suppose that S is finite. It follows that

$$\sum_{s \in S} \operatorname{sgn} s = \sum_{s \in \operatorname{Fix}\iota} \operatorname{sgn} s. \tag{2.3}$$

Indeed, if s is in a 2-cycle $(s, \iota(s))$, then on the left-hand side we have $\operatorname{sgn} s + \operatorname{sgn}\iota(s) = 0$. So all elements in 2-cycles cancel from the sum, which leaves only terms from $\operatorname{Fix}\iota$. This formula can be very useful if the sum on the right has far fewer terms than the one on the left. And if all the fixed points of ι have the same sign so that the right-hand side of (2.3) is $\pm|\operatorname{Fix}\iota|$, then we may be able to glean even more information. The

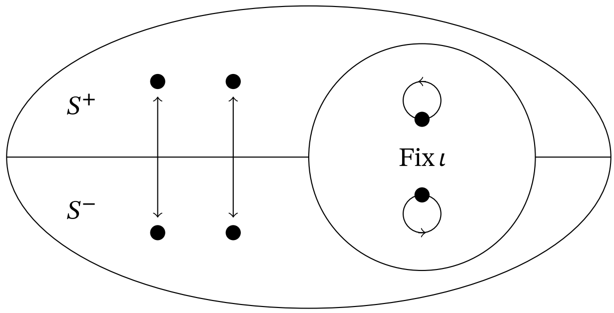


Figure 2.2. A sign-reversing involution on a set S

general method for trying to prove facts about a signed sum $\sum_{k\geq 0}(-1)^k a_k$ for positive integers a_k is as follows:

(i) Find a set S enumerated by the positive sum $\sum_k a_k$.

(ii) Sign S so that the left-hand side of (2.3) equals $\sum_k (-1)^k a_k$.

(iii) Devise a sign-reversing involution ι on S with many 2-cycles.

As our first application of sign-reversing involutions, we will reprove the formula for the alternating sum of the binomial coefficients in Theorem (1.3.3)(d). In fact, the original demonstration was a closet version of this technique. But now we can present the involution proof in its full glory. We restate the identity here for ease of reference:

$$\sum_k (-1)^k \binom{n}{k} = \delta_{n,0}. \tag{2.4}$$

Proof. As usual, we assume $n \geq 1$ since $n = 0$ is trivial. From the sum with the signs removed, it is clear that we should let $S = 2^{[n]}$. And from the way k is being used in the original sum, one would be inclined to let $\operatorname{sgn} s = (-1)^{\#s}$ for $s \subseteq [n]$. We now need to check that the left-hand sides of (2.3) and (2.4) agree. The technique we will use, of turning a single sum into a double sum and then grouping terms, is a common one in enumerative combinatorics. In this case

$$\begin{aligned}
\sum_{s\in S} \operatorname{sgn} s &= \sum_{s\subseteq[n]} (-1)^{\#s} \\
&= \sum_k \sum_{s\in\binom{[n]}{k}} (-1)^k \\
&= \sum_k (-1)^k \binom{n}{k}
\end{aligned}$$

as desired.

As for the sign-reversing involution, we already saw it in the original demonstration of this result. Define $\iota\colon 2^{[n]} \to 2^{[n]}$ by $\iota(s) = s \,\Delta\, \{n\}$. As noted previously, this is an involution. To see that it is sign reversing, we have that $|s \,\Delta\, \{n\}| = |s| \pm 1$. So

$\operatorname{sgn}\iota(s) = (-1)^{|s|\pm 1} = -\operatorname{sgn} s$. Finally, we just need to determine $\operatorname{Fix}\iota$. But $s \,\Delta\, \{n\} \neq s$ for all $s \subseteq [n]$. Thus the right-hand side of (2.3) is the empty sum. Since this equals zero the proof is complete. □

Given that (2.4) was a crucial tool in proving the PIE, it may not come as a surprise that the principle itself can be proved using a sign-reversion involution. We restate the PIE here, in part so as not to conflict with the notation we have set up for sign-reversing involutions. So given a finite set A and subsets $A_1, \ldots, A_n$ we wish to prove

$$\left| A - \bigcup_{i=1}^{n} A_i \right| = |A| - \sum_{1\le i\le n} |A_i| + \sum_{1\le i<j\le n} |A_i \cap A_j| - \cdots + (-1)^n \left| \bigcap_{i=1}^{n} A_i \right|. \tag{2.5}$$

Proof. An example illustrating the proof will be found after the demonstration. We cannot take $S = A$ since the same element of A is counted in many of the terms on the right side of (2.5). To take care of these multiplicities, let

$$S = \{(a, I) \in A \times 2^{[n]} \mid a \in A_I\}, \tag{2.6}$$

recalling the notation

$$A_I = \bigcap_{i\in I} A_i. \tag{2.7}$$

Notice how pairs come into play here even though they are not apparent from the original statement of the result to be proved, just as in the case of the demonstration of Theorem 1.9.3. Note that $A_\emptyset = A$. So $(a, \emptyset)$ is a pair for all $a \in A$, and if $a \notin \bigcup A_i$, then this is the only pair in which a appears. Since the signs in (2.5) come from the number of subsets in an intersection, we define

$$\operatorname{sgn}(a, I) = (-1)^{\#I}.$$

It follows that

$$\begin{aligned}
\sum_{s\in S} \operatorname{sgn} s &= \sum_{(a,I)\in S} (-1)^{\#I} \\
&= \sum_{I\in 2^{[n]}} \sum_{a\in A_I} (-1)^{\#I} \\
&= \sum_{k=0}^{n} \sum_{I\in\binom{[n]}{k}} \sum_{a\in A_I} (-1)^k \\
&= \sum_{k=0}^{n} (-1)^k \sum_{I\in\binom{[n]}{k}} |A_I|
\end{aligned}$$

as we wished.

To construct an involution, define for each $a \in \bigcup A_i$ the index

$$m(a) = \max\{i \mid a \in A_i\}.$$

Finally, we let

$$\iota(a, I) = \begin{cases} (a, I \,\Delta\, \{m(a)\}) & \text{if } a \in \bigcup A_i, \\ (a, I) & \text{otherwise.} \end{cases}$$

It is clear from the definition that this is an involution whose fixed points are in bijection with the elements of $A - \bigcup A_i$ and whose 2-cycles contain elements of opposite signs. Since elements of $A - \bigcup A_i$ each occur in exactly one pair, it follows that the right side of (2.3) is just the cardinality of this set, as desired. □

To illustrate the proof, suppose $A = \{a, b, c, d\}$, $A_1 = \{a, b\}$, and $A_2 = \{b, c\}$. Then, leaving out curly brackets and commas in the index sets I for readability,

$$S = \{(a, \emptyset),\ (a, 1),\ (b, \emptyset),\ (b, 1),\ (b, 2),\ (b, 12),\ (c, \emptyset),\ (c, 2),\ (d, \emptyset)\}.$$

Also $m(a) = 1$ and $m(b) = m(c) = 2$ so that the involution creates the following 2-cycles:

$$(a, \emptyset) \leftrightarrow (a, 1),\ (b, \emptyset) \leftrightarrow (b, 2),\ \ (b, 1) \leftrightarrow (b, 12),\ (c, \emptyset) \leftrightarrow (c, 2).$$

The only fixed point is $(d, \emptyset)$ and $A - (A_1 \cup A_2) = \{d\}$.

It would be nice to prove something we have not seen before using our new technique. Here is an identity involving Stirling numbers of the second kind.

Theorem 2.2.2. *For $n \ge 0$ we have*

$$\sum_{k \ge 0} (-1)^k k!\, S(n, k) = (-1)^n.$$

Proof. The first order of business will be to give a combinatorial interpretation to the summands. A *composition* of a set T is a sequence of nonempty subsets $\rho = (B_1, \ldots, B_k)$ such that $\biguplus_i B_i = T$. In this case we write $\rho \vDash T$. So the number of $\rho \vDash [n]$ with k blocks is $k!\, S(n, k)$ since we can start with any of the $S(n, k)$ partitions in $S([n], k)$ and order its blocks in $k!$ ways. The reader should have enough experience with signed sets at this point to see that we are going to want to take S to be all $\rho \vDash [n]$ with $\operatorname{sgn} \rho = (-1)^k$ if ρ has k blocks. Verifying that this gives the correct alternating sum andis easy and is left as an exercise.

The involution will be more interesting. We will break it into two cases which will be inverses of each other. As often, an example follows the proof. Given $\rho = (B_1, \ldots, B_k) \vDash [n]$, we say that B_j is *splittable* if $\#B_j \ge 2$. In this case the *splitting map* applied to B_j is defined by

$$\sigma(B_1, \ldots, B_k) = (B_1,\ \ldots,\ B_{j-1},\ \{b\},\ B_j - \{b\},\ B_{j+1},\ \ldots,\ B_k)$$

where $b = \min B_j$. In other words B_j is replaced by a pair of blocks, the first containing its minimum element and the other all the rest of its elements. Although the notation σ does not indicate which block is to be split, this will be made clear from the context. We will now define the part of the involution which will undo splitting. Given ρ, we say that B_j can be *merged* with B_{j+1} if

(1) $B_j = \{b\}$ for some element $b \in [n]$ and
(2) $b < \min B_{j+1}$.

In this case the *merging map* applied to B_j is defined by

$$\mu(B_1, \ldots, B_k) = (B_1, \ldots, B_{j-1}, B_j \cup B_{j+1}, B_{j+2}, \ldots, B_k).$$

It should be clear that if B_i can be split into B'_i and B'_{i+1}, then the primed blocks can be merged back into B_i and vice versa. To define the involution ι, suppose we are given $\rho = (B_1, \ldots, B_k)$. We scan ρ from left to right until we find the first index j, if any, such that B_j can either be split or merged with B_{j+1}. (Clearly one cannot do both since splitting implies that $\#B_j \geq 2$ and merging that $\#B_j = 1$.) Now define

$$\iota(\rho) = \begin{cases} \sigma(\rho) & \text{if } B_j \text{ can be split,} \\ \mu(\rho) & \text{if } B_j \text{ can be merged.} \end{cases}$$

If no such index exists, then ρ will be a fixed point of ι.

We have some work to do to verify that ι is an involution. Specifically, we must show that if $\iota(\rho) = \rho'$ is obtained from ρ by splitting at index j, then $\iota(\rho')$ will be obtained by merging at the same index and vice versa. We will do the first case and leave the second to the reader. First note that since no B_i, $i < j$, could be split in ρ we must have $B_i = \{b_i\}$ for some b_i for each i in this range. Furthermore, since none of these B_i could be merged into B_{i+1}, we must also have $b_1 > b_2 > \cdots > b_{j-1} > b_j = \min B_j$. Now in ρ' we have $B'_i = \{b_i\}$ for $i \leq j$ with $b_1 > \cdots > b_j$. As a consequence, no B'_i can be split or merged for $i < j$ and so $\iota(\rho')$ will merge B'_j into B'_{j+1}. Thus $\iota(\rho') = \rho$ as desired.

It is clear that ι is sign reversing since $\iota(\rho)$ has one more or one fewer block than ρ. So we just need to find the fixed points. But if $\rho \in \operatorname{Fix} \iota$, then all ρ's blocks contain a single element; otherwise one could be split. It follows that $\rho = (\{b_1\}, \ldots, \{b_n\})$. Furthermore, none of the blocks can be merged and so $b_1 > \cdots > b_n$. But this forces our set composition to be $\rho = (\{n\}, \{n-1\}, \ldots, \{1\})$ and $\operatorname{sgn} \rho = (-1)^n$, completing the proof. □

To illustrate, suppose $n = 8$. As we have done previously, we will dispense with brackets and commas in sets. Consider $\rho = (B_1, \ldots, B_5) = (5, 3, 147, 2, 68)$. Then B_3 is splittable and splitting it results in $\sigma(\rho) = (5, 3, 1, 47, 2, 68)$. Also, B_4 can be merged into B_5 in ρ since $B_4 = \{2\}$ and $2 < \min B_5 = 6$. Merging these two blocks gives $\mu(\rho) = (5, 3, 147, 268)$. To decide which operation to use we start with B_1. It cannot be split, having only one element. And it cannot be merged with B_2 since $5 > \min B_2 = 3$. Similarly B_2 cannot be split or merged with B_3. But we have already seen that B_3 can be split so that $\iota(\rho) = (5, 3, 1, 47, 2, 68) = \rho'$. To check that $\iota(\rho') = \rho$ is similar.

Involutions involving merging and splitting often come up when finding formulae for antipodes in Hopf algebras. One can consult the papers of Benedetti-Bergeron [**7**], Benedetti-Hallam-Machacek [**8**], Benedetti-Sagan [**9**], or Bergeron-Ceballos [**12**] for examples.

2.3. The Garsia–Milne Involution Principle

So far we have used sign-reversing involutions to explain cancellation in alternating sums. But can they also furnish a bijection for proving that two given sets have the same cardinality? The answer in certain cases is "yes" and the standard technique for doing this is called the Garsia–Milne Involution Principle. Garsia and Milne [**30**] introduced this method to give the first bijective proof of the Rogers–Remanujan identities, famous formulas which involve certain sets of integer partitions. Since then the Involution Principle has found a number of other applications. See, for example, the articles of Remmel [**73**] or Wilf [**100**].

In order to prove the Garsia–Milne result, we will need a version of Lemma 1.9.5 which applies to a slightly wider class of digraphs. Since the demonstration of the next result is similar to that of the earlier one, we leave the pleasure of proving it to the reader.

Lemma 2.3.1. *Let $D = (V, A)$ be a digraph. We have* $\operatorname{odeg} v, \operatorname{ideg} v \le 1$ *for all $v \in V$ if and only if D is a disjoint union of directed paths and directed cycles.* □

The basic idea of the Involution Principle is that, under suitable conditions, if one has two signed sets each with their own sign-reversing involution, then we can use a bijection between these sets to create a bijection between their fixed-point sets. So let S and T be disjoint signed sets with sign-reversing involutions $\iota : S \to S$ and $\kappa : T \to T$ such that $\operatorname{Fix} \iota \subseteq S^+$ and $\operatorname{Fix} \kappa \subseteq T^+$. Furthermore, suppose we have a bijection $f : S \to T$ which *preserves signs* in that $\operatorname{sgn} f(s) = \operatorname{sgn} s$ for all $s \in S$. A picture of this setup can be found in Figure 2.3. Note that although all arrows are really double-headed, we have only shown them in one direction because of what is to come. And the circular arrows on the fixed points have been ignored. We now construct a map $F : \operatorname{Fix} \iota \to \operatorname{Fix} \kappa$ as follows. To define $F(s)$ for $s \in \operatorname{Fix} \iota$ we first compute $f(s) \in T^+$. If $f(s) \in \operatorname{Fix} \kappa$, then we let $F(s) = f(s)$. If not, we apply the functional composition $\phi = f \circ \iota \circ f^{-1} \circ \kappa$ to $f(s)$. Remembering that we compose from right to left, this takes $f(s)$ to T^-, S^-, S^+, and T^+ in that order. If this brings us to an element of $\operatorname{Fix} \kappa$, then we let $F(s) = \phi(f(s))$. Otherwise we apply ϕ as many times as necessary, say m, to arrive at an element of $\operatorname{Fix} \kappa$ and define

$$F(s) = \phi^m(f(s)). \tag{2.8}$$

Continuing the example in Figure 2.3 we see that $f(s) = u \notin \operatorname{Fix} \kappa$. So we apply ϕ, which takes u to v, r, q, and t in turn. Since $t \in \operatorname{Fix} \kappa$ we let $F(s) = t$. Of course, we have to worry whether this is all well-defined; e.g., does m always exist? And we also need to prove that F is a bijection. This is taken care of by the next theorem.

Theorem 2.3.2 (Garsia–Milne Involution Principle)**.** *With the notation of the previous paragraph, the map $F : \operatorname{Fix} \iota \to \operatorname{Fix} \kappa$ is a well-defined bijection.*

Proof. Recall the notion of a functional digraph as used in the proof from Section 1.9 of Theorem 1.5.1. Define the following functions by restriction of their domains:

$$\bar{f} = f|_{S^+}, \quad \bar{g} = f^{-1}|_{T^-}, \quad \bar{\iota} = \iota|_{S^-}, \quad \bar{\kappa} = \kappa|_{T^+ - \operatorname{Fix} \kappa}.$$

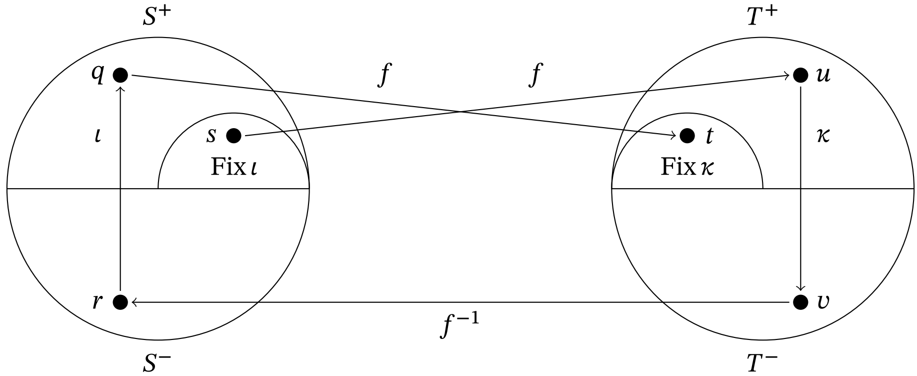


Figure 2.3. An example of the Garsia–Milne construction

Consider D which is the union of the functional digraphs for $\overline{f}, \overline{g}, \bar{\iota}$, and $\bar{\kappa}$. It is easy to verify from the definitions that $x \in V(D)$ has in-degrees and out-degrees given by the following table depending on the subset of $S \cup T$ containing x:

subset	odeg x	ideg x
Fix ι	1	0
Fix κ	0	1
$(S - \operatorname{Fix}\iota) \cup (T - \operatorname{Fix}\kappa)$	1	1

For example, if $x \in \operatorname{Fix}\iota$, then the only arc containing x comes from $\overline{f}$ and so $\operatorname{odeg} x = 1$ and $\operatorname{ideg} x = 0$. On the other hand, if $x \in S^+ - \operatorname{Fix}\iota$, then x has an arc going out from $\overline{f}$ and one coming in from $\bar{\iota}$ giving $\operatorname{odeg} x = \operatorname{ideg} x = 1$.

Now D satisfies the hypothesis of the forward direction of Lemma 2.3.1. It follows that D is a disjoint union of directed paths and directed cycles. Each directed path must start at a vertex with out-degree 1 and in-degree 0 and end at a vertex with these degrees switched. Furthermore, all other vertices have out-degree and in-degree both 1. From these observations and the chart, it follows that these paths define a 1-to-1 correspondence between the vertices of Fix ι and those of Fix κ. Furthermore, from the definition of D we see that each path corresponds exactly to a functional composition $\phi^m f(s)$ for $s \in \operatorname{Fix}\iota$ and some $m \geq 0$. So F is the bijection defined by these paths. □

Before we give an application of the previous theorem, we should mention an approach which can be useful in setting up the necessary sets and bijections. Here is one way to try to find a bijection $F : X \to Y$ between two finite sets X, Y.

(i) As with the PIE, construct a set A with subsets $A_1, \ldots, A_n$ such that $X = A - \bigcup A_i$. Similarly construct B and $B_1, \ldots, B_n$ for Y.

(ii) Use the method of our second proof of the PIE to set up a sign-reversing involution ι on the set S as given by (2.6). Similarly construct κ on a set T.

(iii) Find a bijection $f : S \to T$ of the form

$$f(a, I) = (b, I)$$

which is well-defined in that $a \in A_I$ if and only if $b \in B_I$.

Recall that $\operatorname{Fix}\iota = (a, \emptyset)$ where $a \in A - \bigcup A_i$. Thus $\operatorname{Fix}\iota \subseteq S^+$ as needed to apply the Involution Principle, and there is a natural bijection between $\operatorname{Fix}\iota$ and X. Note also that f is automatically sign preserving since $\operatorname{sgn}(a, I) = (-1)^{\#I} = \operatorname{sgn}(b, I)$. So once these three steps have been accomplished, Theorem 2.3.2 guarantees that we have a bijection $X \to Y$.

As already remarked, the Involution Principle is useful in proving integer partition identities. Say that partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ has *distinct parts* if $\lambda_1 > \lambda_2 > \cdots > \lambda_k$ (as opposed to the usual weakly decreasing condition). On the other hand, say that λ has *odd parts* if all the λ_i are odd. The next result is a famous theorem of Euler. As has become traditional, an example follows the proof.

Theorem 2.3.3 (Euler). *Let $P_d(n)$ be the set of partitions of n with distinct parts and let $P_o(n)$ be the set of partitions of n with odd parts. For $n \geq 0$ we have*

$$\#P_d(n) = \#P_o(n).$$

Proof. It suffices to show that there is a bijection $P_d(n) \to P_o(n)$. To apply the PIE to $P_d(n)$ we can take $A = P(n)$, the set of all partitions of n, with subsets $A_1, \ldots, A_n$ where

$$A_i = \{\lambda \vdash n \mid \lambda \text{ has (at least) two copies of the part } i\}.$$

Note that $A_i = \emptyset$ if $i > n/2$, but this does no harm and keeps the notation simple. It should be clear from the definitions that $P_d(n) = A - \bigcup A_i$. Similarly, for $P_o(n)$ we let $B = P(n)$ with subsets

$$B_i = \{\mu \vdash n \mid \mu \text{ has a part of the form } 2i\}$$

for $1 \leq i \leq n$. Again, it is easy to see that $P_o(n) = B - \bigcup B_i$.

The construction of S, ι, T, and κ are now exactly the same as in the second proof of the PIE. So it suffices to construct an appropriate bijection $f : S \to T$. Given $(\lambda, I) \in S$, we replace, for each $i \in I$, a pair of i's in λ by a part $2i$ to form μ. So if $\lambda \in A_i$, then $\mu \in B_i$ for all $i \in I$ and the map $f(\lambda, I) = (\mu, I)$ is well-defined. It is also easy to construct f^{-1}, taking an even part $2i$ in μ and replacing it with two copies of i to form λ as i runs over I. Appealing to Theorem 2.3.2 finishes the proof. □

To illustrate this demonstration, suppose we start with $(6, 2, 1) \in P_d(9)$. For the pairs in S and T, we will dispense with delimiters and commas as usual. So

$$\begin{aligned}
(621, \emptyset) &\overset{f}{\mapsto} (621, \emptyset) \overset{\kappa}{\mapsto} (621, 3) \overset{f^{-1}}{\mapsto} (3321, 3) \overset{\iota}{\mapsto} (3321, \emptyset) \\
&\overset{f}{\mapsto} (3321, \emptyset) \overset{\kappa}{\mapsto} (3321, 1) \overset{f^{-1}}{\mapsto} (33111, 1) \overset{\iota}{\mapsto} (33111, 13) \\
&\overset{f}{\mapsto} (621, 13) \overset{\kappa}{\mapsto} (621, 1) \overset{f^{-1}}{\mapsto} (6111, 1) \overset{\iota}{\mapsto} (6111, \emptyset) \\
&\overset{f}{\mapsto} (6111, \emptyset) \overset{\kappa}{\mapsto} (6111, 3) \overset{f^{-1}}{\mapsto} (33111, 3) \overset{\iota}{\mapsto} (33111, \emptyset) \\
&\overset{f}{\mapsto} (33111, \emptyset).
\end{aligned}$$

It follows that we should map $(6, 2, 1) \overset{F}{\mapsto} (3, 3, 1, 1, 1)$. Clearly one might like to find a more efficient bijection if one exists. This issue will be further explored in the exercises.

2.4. The Reflection Principle

The Reflection Principle is a geometric method for working with certain combinatorial problems involving lattice paths. In particular, it will permit us to give a very simple proof of the binomial coefficient formula for the Catalan numbers. It is also useful in proving unimodality, an interesting property of real number sequences.

Consider the integer lattice $\mathbb{Z}^2$ and northeast paths in this lattice. Suppose we are given a line in the plane of the form $L : y = x + b$ for some $b \in \mathbb{Z}$. Note that the reflection in L of any northeast path is again a northeast path. If P is a path from u to v, then we write $P : u \to v$ or $u \overset{P}{\to} v$. Suppose $P : u \to v$ intersects L and let x be its

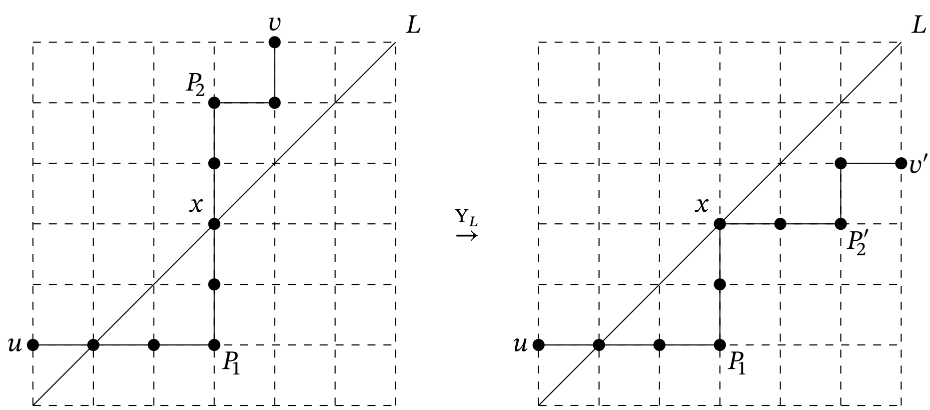


Figure 2.4. The map Y_L

last (northeast-most) point of intersection. Then P can be written as the concatenation

$$P : u \stackrel{P_1}{\to} x \stackrel{P_2}{\to} v.$$

For example, on the left in Figure 2.4 we have the path $P = EEENNNNEN$ with $P_1 = EEENN$ and $P_2 = NNEN$. We now define a new path

$$\Upsilon_L(P) : u \stackrel{P_1}{\to} x \stackrel{P_2'}{\to} v'$$

where P_2' and v' are the reflections of P_2 and v in L, respectively. Returning to our example, $P_2' = EENE$, which is obtained from P_2 by merely interchanging north and east steps. So $\Upsilon_L(P) = EEENNEENE$ as on the right in Figure 2.4. This is the fundamental map for using the Reflection Principle. To state it precisely, let $\mathcal{NE}(u; v)$ denote the set of northeast paths from u to v and let $\mathcal{NE}_L(u; v)$ be the subset of paths which intersect L. If u is omitted, then it is assumed that $u = (0, 0)$. Also, be sure to distinguish the notation $\mathcal{NE}(u; v)$ for the northeast paths from u to v and $\mathcal{NE}(m, n)$ for the northeast paths from $(0, 0)$ to (m, n). The former contains a semicolon where the latter has a comma.

Theorem 2.4.1 (Reflection Principle). *Given $L : y = x + b$ for $b \in \mathbb{Z}$ and $v \in \mathbb{Z}^2$, we let v' be the reflection of v in L. Then the map $\Upsilon_L : \mathcal{NE}_L(u; v) \to \mathcal{NE}_L(u; v')$ is a bijection.*

Proof. In fact, we can show that Υ_L is an involution on $\mathcal{NE}_L(u; v) \cup \mathcal{NE}_L(u; v')$. This follows from the fact that reflection in L is an involution and that the set of intersection points does not change when passing from $P \cap L$ to $\Upsilon_L(P) \cap L$. □

As a first application of Theorem 2.4.1, we will give a simpler, although not as purely combinatorial, proof of Theorem 1.11.3. We restate the formula here for reference:

$$C(n) = \frac{1}{n+1}\binom{2n}{n}.$$

Proof. Recall that $C(n)$ counts the set $\mathcal{D}(n)$ of northeast Dyck paths from $(0, 0)$ to (n, n). From Theorem 1.11.1 we know that the total number of all northeast paths P from the origin to (n, n) is

$$\#\mathcal{NE}(n, n) = \binom{2n}{n}.$$

Note that P does not stay weakly above $y = x$ if and only if P intersects the line $L : y = x - 1$. And by the Reflection Principle, such paths are in bijection with $\mathcal{NE}_L((0, 0); (n + 1, n - 1))$ since $(n + 1, n - 1)$ is the reflection of (n, n) in L. But all paths from $(0, 0)$ to $(n + 1, n - 1)$ cross L since these two points are on opposite sides of the line. Thus, using Theorem 1.11.1 again,

$$\#\mathcal{NE}_L((0, 0); (n + 1, n - 1)) = \#\mathcal{NE}(n + 1, n - 1) = \binom{2n}{n+1}.$$

So subtracting the number of non-Dyck paths from the total number of paths in $\mathcal{NE}(n, n)$ gives

$$\begin{aligned}
C(n) &= \binom{2n}{n} - \binom{2n}{n+1} \\
&= \frac{(2n)!}{n!\ n!} - \frac{(2n)!}{(n+1)!\ (n-1)!} \\
&= \left(1 - \frac{n}{n+1}\right)\binom{2n}{n} \\
&= \frac{1}{n+1}\binom{2n}{n}
\end{aligned}$$

as desired. □

The Reflection Principle can also be used to prove that certain sequences have a property called unimodality. A sequence of real numbers $a_0, a_1, \ldots, a_n$ is said to be *unimodal* if there is an index m such that

$$a_0 \le a_1 \le \cdots \le a_m \ge a_{m+1} \ge \cdots \ge a_n.$$

So this is the next most complicated behavior after being weakly increasing or weakly decreasing. In fact the latter are the special cases of unimodality where $m = n$ or $m = 0$. Many sequences arising in combinatorics, algebra, and geometry are unimodal. See the survey articles of Stanley [**89**], Brenti [**20**], or Brändén [**19**] for more details. The term "unimodal" comes from probability and statistics where one thinks of the a_i as giving you the distribution of a random variable taking values in $\{0, 1, \ldots, n\}$. Then a unimodal distribution has only one hump.

We have already met a number of unimodal sequences, although we have not remarked on the fact. Here is the simplest.

Theorem 2.4.2. *For $n \ge 0$ the sequence*

$$\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$$

is unimodal.

Proof. Because the binomial coefficients are symmetric, Theorem 1.3.3(b), it suffices to prove that this sequence is increasing up to its halfway point. So we want to show

$$\binom{n}{k} \le \binom{n}{k+1}$$

for $k < \lfloor n/2 \rfloor$. From Theorem 1.11.1, we know that

$$\binom{n}{k} = \#\mathcal{NE}(k, n-k) \quad \text{and} \quad \binom{n}{k+1} = \#\mathcal{NE}(k+1, n-k-1).$$

So it suffices to find an injection $i : \mathcal{NE}(k, n-k) \to \mathcal{NE}(k+1, n-k-1)$. Let L be the perpendicular bisector of the line segment from $(k, n-k)$ to $(k+1, n-k-1)$. It is easy to check that L has the form $y = x + b$ for $b \in \mathbb{Z}$. From the Reflection Principle, we have a bijection $\Upsilon_L : \mathcal{NE}_L(k, n-k) \to \mathcal{NE}_L(k+1, n-k-1)$. But since $k < \lfloor n/2 \rfloor$ the lattice points $(0,0)$ and $(k, n-k)$ are on opposite sides of L so that $\mathcal{NE}_L(k, n-k) = \mathcal{NE}(k, n-k)$. Furthermore $\mathcal{NE}_L(k+1, n-k-1) \subseteq \mathcal{NE}(k+1, n-k-1)$. So extending the range of Υ_L provides the desired injection. □

It turns out that the Stirling number sequences

$$c(n,0), c(n,1), \ldots, c(n,n) \quad \text{and} \quad S(n,0), S(n,1), \ldots, S(n,n)$$

are also unimodal. But this is not so easy to prove directly. One reason for this is that these sequences are not symmetric like the one for the binomial coefficients. And there is no known simple expression for the index m where they achieve their maxima. Instead it is better to use another property of real sequences, called log-concavity, which can imply unimodality. This is one of the motivations for the next section.

2.5. The Lindström–Gessel–Viennot Lemma

The lemma in question is a powerful technique for dealing with certain determinantal identities. It was first discovered by Lindström [**57**] and then used to great effect by Gessel and Viennot [**31**] as well as many other authors. Like the Reflection Principle, this method uses directed paths. On the other hand, it uses multiple paths and is not restricted to the integer lattice. In particular, when there are two paths, then log-concavity results can be obtained.

A sequence of real numbers $a_0, a_1, \ldots, a_n$ is called *log-concave* if, for all $0 < k < n$, we have

$$a_k^2 \geq a_{k-1}a_{k+1}. \tag{2.9}$$

As usual, we can extend this to all $k \in \mathbb{Z}$ by letting $a_k = 0$ for $k < 0$ or $k > n$. Log-concave sequences, like unimodal ones, are ubiquitous in combinatorics, algebra, and geometry. See the previously cited survey articles of Stanley, Brenti, and Brändén for details. For example, a row of Pascal's triangle or either of the Stirling triangles is log-concave.

The name "log-concave" comes from the following scenario. Suppose that we have a function $f : \mathbb{R} \to \mathbb{R}$ which is concave down. So if one takes any two points on the graph of f, then the line segment connecting them lies weakly below f. Taking the points to be $(k-1, f(k-1))$ and $(k+1, f(k+1))$ and comparing the y-coordinate of the midpoint of the corresponding line segment with that coordinate on f gives $(f(k-1) + f(k+1))/2 \leq f(k)$. Now if $f(x) > 0$ for all x and the function $\log f(x)$ is concave down, then substituting into the previous inequality and exponentiating gives $f(k-1)f(k+1) \leq f(k)^2$ just like the definition of log-concavity for sequences.

It turns out that log-concavity and unimodality are related.

Proposition 2.5.1. *Suppose that $a_0, a_1, \ldots, a_n$ is a sequence of positive reals. If the sequence is log-concave, then it is unimodal.*

Proof. To show a sequence is unimodal it suffices to show that after its first strict decrease, then it continues to weakly decrease. But $a_{k-1} > a_k$ is equivalent to $a_{k-1}/a_k > 1$ for positive a_k. Rewriting (2.9) as $a_k/a_{k+1} \geq a_{k-1}/a_k$ we see that if $a_{k-1}/a_k > 1$, then $a_{l-1}/a_l > 1$ for all $l \geq k$. So the sequence is unimodal. □

Even though log-concavity implies unimodality for positive sequences, it is paradoxically often easier to prove log-concavity rather than proving unimodality directly. This comes in part from the fact that the log-concave condition is a uniform one for all k, as opposed to unimodality where one must know where the maximum of the sequence occurs.

We can rewrite (2.9) as $a_k^2 - a_{k-1}a_{k+1} \geq 0$, or in terms of determinants as

$$\begin{vmatrix} a_k & a_{k+1} \\ a_{k-1} & a_k \end{vmatrix} \geq 0. \tag{2.10}$$

To prove that the determinant is nonnegative, we could show that it counts something and that is exactly what the Lindström–Gessel–Viennot Lemma is set up to do. We will first consider the case of 2×2 determinants and at the end of the section indicate how to do the general case. As a running example, we will show how to prove log-concavity of the sequence of binomial coefficients considered in Theorem 2.4.2.

Let D be a digraph which is *acyclic* in that it contains no directed cycles. Given two vertices of $u, v \in V(D)$, we let $\mathcal{P}(u; v)$ denote the set of directed paths from u to v. We will assume that u, v are always chosen so that $p(u; v) = \#\mathcal{P}(u; v)$ is finite even if D itself is not. To illustrate, let D be the digraph with vertices $\mathbb{Z}^2$ and arcs from (m, n) to $(m + 1, n)$ and to $(m, n + 1)$ for all $m, n \in \mathbb{Z}$. Then $\mathcal{P}(u; v)$ is just the set of northeast lattice paths from u to v, denoted $\mathcal{NE}(u; v)$ in the previous section. We will continue to use the notation for general paths from that section for any acyclic digraph. We also extend that notation as follows. Given a directed path $P : u \to v$ and vertices x coming before y on P, we let $x \xrightarrow{P} y$ be the portion of P between x and y.

Continuing the general exposition, suppose we are given $u_1, u_2 \in V$ called the *initial vertices* and $v_1, v_2 \in V$ which are the *final vertices*. We wish to consider determinants of the form

$$\begin{vmatrix} p(u_1; v_1) & p(u_1; v_2) \\ p(u_2; v_1) & p(u_2; v_2) \end{vmatrix} = p(u_1; v_1)p(u_2; v_2) - p(u_1; v_2)p(u_2; v_1). \tag{2.11}$$

Note that $p(u_1; v_1)p(u_2; v_2)$ counts pairs of paths

$$(P_1, P_2) \in \mathcal{P}(u_1; v_1) \times \mathcal{P}(u_2; v_2) := \mathcal{P}_{12}$$

and similarly for $p(u_1; v_2)p(u_2; v_1)$ and

$$\mathcal{P}(u_1; v_2) \times \mathcal{P}(u_2; v_1) := \mathcal{P}_{21}.$$

Returning to our example, if we wish to show

$$\binom{n}{k}^2 - \binom{n}{k-1}\binom{n}{k+1} \geq 0,$$

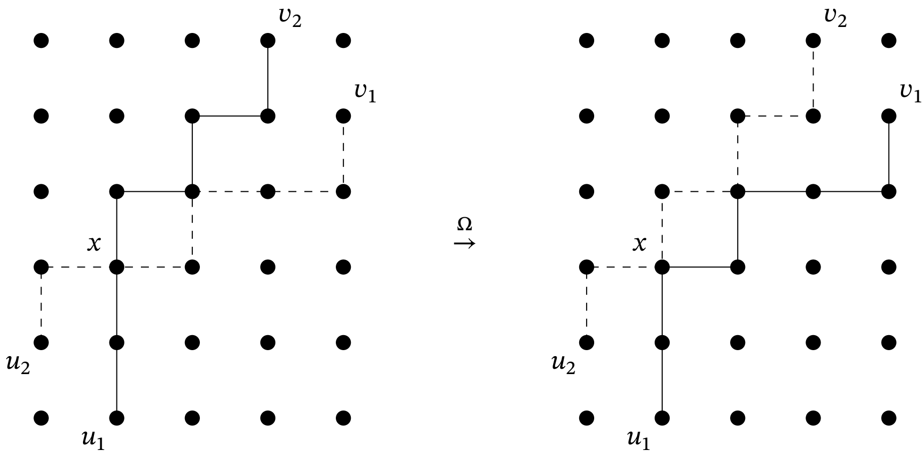


Figure 2.5. The Lindström–Gessel–Viennot Involution

then we could take

$$u_1 = (1,0),\ u_2 = (0,1),\ v_1 = (k+1, n-k),\ v_2 = (k, n-k+1).$$

It follows from Theorem 1.11.1 that $p(u_1;v_1) = p(u_2;v_2) = \binom{n}{k}$, while $p(u_1;v_2) = \binom{n}{k-1}$ and $p(u_2;v_1) = \binom{n}{k+1}$. More specifically, if $n = 7$ and $k = 3$, then in Figure 2.5 we have a pair of paths in $\mathcal{P}_{21}$ counted by $\binom{7}{2}\binom{7}{4}$ on the left and another pair in $\mathcal{P}_{12}$ counted by $\binom{7}{3}^2$ on the right. For readablity, the grid for the integer lattice has been suppressed, leaving only the vertices of $\mathbb{Z}^2$.

To prove that the determinant (2.11) is nonnegative, we will construct a sign-reversing involution Ω on the set $\mathcal{P} := \mathcal{P}_{12} \cup \mathcal{P}_{21}$ where

$$\operatorname{sgn}(P_1,P_2) = \begin{cases} +1 & \text{if } (P_1,P_2) \in \mathcal{P}_{12}, \\ -1 & \text{if } (P_1,P_2) \in \mathcal{P}_{21}. \end{cases}$$

We will construct Ω so that every pair in $\mathcal{P}_{21}$ is in a 2-cycle with a pair in $\mathcal{P}_{12}$. Furthermore, the remaining fixed points in $\mathcal{P}_{12}$ will be exactly the path pairs in $\mathcal{P}$ which do not intersect. It follows that (2.11) is just the number of nonintersecting path pairs in $\mathcal{P}$ and therefore must be nonnegative.

To define Ω, consider a path pair $(P_1,P_2) \in \mathcal{P}$. If $P_1 \cap P_2$ is empty, then this pair is in $\mathcal{P}_{12}$, since every pair in $\mathcal{P}_{21}$ intersects. So in this case we let $\Omega(P_1,P_2) = (P_1,P_2)$, a fixed point. If $P_1 \cap P_2 \neq \emptyset$, then consider the list of intersections $x_1,\ldots,x_t$ in the order in which they are encountered on P_1. We claim they must also be encountered in this order on P_2. For if there were intersections x, y such that x comes before y on P_1 and y comes before x on P_2, then one can show that the directed walk $x \stackrel{P_1}{\to} y \stackrel{P_2}{\to} x$ contains a directed cycle, as the reader will be asked to do in the exercises. This contradicts the assumption that D is acyclic. So there is a well-defined notion of a first intersection

$x = x_1$. We now let $\Omega(P_1, P_2) = (P_1', P_2')$ where

$$P_1' = u_1 \xrightarrow{P_1} x \xrightarrow{P_2} v_2,$$
$$P_2' = u_2 \xrightarrow{P_2} x \xrightarrow{P_1} v_1,$$

if $(P_1, P_2) \in \mathcal{P}_{12}$, and similarly if $(P_1, P_2) \in \mathcal{P}_{21}$ with v_1 and v_2 reversed. An illustration of this map is shown in Figure 2.5.

Because the set of intersections in (P_1, P_2) is the same as in (P_1', P_2'), the first intersection remains the same and this makes Ω an involution. It is also clear from its definition that it changes sign. We have proved the following lemma and corollary.

Lemma 2.5.2. *Let D be an acyclic digraph. Let $u_1, u_2, v_1, v_2 \in V(D)$ be such that each pair of paths $(P_1, P_2) \in \mathcal{P}_{21}$ intersects. Then*

$$\begin{vmatrix} p(u_1; v_1) & p(u_1; v_2) \\ p(u_2; v_1) & p(u_2; v_2) \end{vmatrix} = \text{number of nonintersecting pairs } (P_1, P_2) \in \mathcal{P}_{12}.$$

In particular, the determinant is nonnegative. □

Corollary 2.5.3. *For $n \geq 0$ the sequence*

$$\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$$

is log-concave.

Lemma 2.5.2 can be extended to $n \times n$ determinants as follows. Let $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ be n-tuples of distinct vertices in an acyclic digraph. For $\pi \in \mathfrak{S}_n$, we let

$$\mathcal{P}_\pi = \{(P_1, \ldots, P_n) \mid P_i : u_i \to v_{\pi(i)} \text{ for all } i \in [n]\}$$

and

$$\mathcal{P} = \bigcup_{\pi \in \mathfrak{S}_n} \mathcal{P}_\pi.$$

To make $\mathcal{P}$ into a signed set, recall from abstract algebra that the *sign* of $\pi \in \mathfrak{S}_n$ is

$$\operatorname{sgn} \pi = (-1)^{n-k}$$

if π has k cycles in its disjoint cycle decomposition. There are other ways to define $\operatorname{sgn} \pi$, but they are all equivalent. One crucial property of this sign function is that if $A = [a_{i,j}]$ is a matrix, then

$$\det A = \sum_{\pi \in \mathfrak{S}_n} (\operatorname{sgn} \pi) a_{1,\pi(1)} a_{2,\pi(2)} \ldots a_{n,\pi(n)}.$$

Now if $(P_1, \ldots, P_n) \in \mathcal{P}_\pi$, then we let $\operatorname{sgn}(P_1, \ldots, P_n) = \operatorname{sgn} \pi$.

To extend the involution Ω, call $P = (P_1, \ldots, P_n)$ *intersecting* if there is some pair P_i, P_j which intersects. Given an intersecting P, we find the smallest i such that P_i intersects another path of P and let x be the first intersection of P_i with another path.

Now take the smallest $j > i$ such that P_j goes through x. We now let $\Omega(P) = P'$ where P' is P with P_i, P_j replaced by

$$\begin{aligned} P_i' &= u_i \xrightarrow{P_i} x \xrightarrow{P_j} v_{\pi(j)}, \\ P_j' &= u_j \xrightarrow{P_j} x \xrightarrow{P_i} v_{\pi(i)}, \end{aligned} \tag{2.12}$$

respectively. One now needs to check that Ω is a sign-reversing involution. As before, nonintersecting path families P are fixed points of Ω. Modulo the details about Ω, we have now proved the following.

Lemma 2.5.4 (Lindström–Gessle–Viennot)**.** *Let D be an acyclic digraph. Consider two sequences of vertices $u_1, \ldots, u_n, v_1, \ldots, v_n \in V(D)$ such that every $P \in \mathcal{P}_\pi$ is intersecting for $\pi \neq \mathrm{id}$, the identity permutation. We have*

$$\det[p(u_i; v_j)]_{1\le i,j\le n} = \textit{number of nonintersecting } P \in \mathcal{P}_{\mathrm{id}}.$$

In particular, the determinant is nonnegative. □

This theorem also has something to say about real sequences. Any sequence $a_0, \ldots,$ a_n has a corresponding *Toeplitz matrix* which is the infinite matrix $A = [a_{j-i}]_{i,j\ge 0}$. So

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n & 0 & 0 & 0 & \ldots \\ 0 & a_0 & a_1 & a_2 & \cdots & a_n & 0 & 0 & \ldots \\ 0 & 0 & a_0 & a_1 & a_2 & \cdots & a_n & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

The sequence is called *Pólya frequency*, or *PF* for short, if every square submatrix of A has a nonnegative determinant. Notice that, in particular, we get the determinants in (2.10) so that PF implies log-concave. Lemma 2.5.4 can be used to prove that a sequence is PF in much the same way that Lemma 2.5.2 can be used to prove that it is log-concave. The reader should now have no difficulty in proving the following result.

Theorem 2.5.5. *For $n \ge 0$ the sequence*

$$\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$$

is PF. □

2.6. The Matrix-Tree Theorem

We end this chapter with another application of determinants. There are many places where these animals abide in enumerative combinatorics and a good survey will be found in the articles of Krattenthaler [**54**, **55**]. Here we will be concerned with counting spanning trees using a famous result of Kirchhoff called the Matrix-Tree Theorem.

A subgraph $H \subseteq G$ is called *spanning* if $V(H) = V(G)$. So a spanning subgraph is completely determined by its edge set. A *spanning tree* T of G is a spanning subgraph which is a tree. Clearly for a spanning tree to exist, G must be connected. Let $\mathcal{ST}(G)$ be the set of spanning trees of G. If one considers the graph G on the left in Figure 2.6,

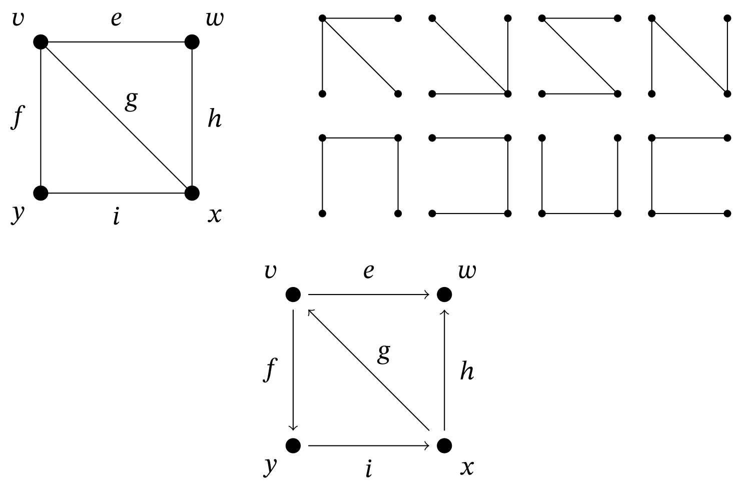


Figure 2.6. A graph G, its spanning trees, and an orientation

then the list of its eight spanning trees is in the middle of the figure (shrunk to half size so they will fit and without the vertex and edge labels). To develop the tools needed to prove our main theorem, we first need to make some remarks about combinatorial matrices.

We will often have occasion to create matrices whose rows and columns are indexed by sets rather than numbers. If S, T are sets, then an $S \times T$ *matrix* M is constructed by giving a linear order to the elements of S and to those of T and using them to index the rows and columns of M, respectively. So if $(s,t) \in S \times T$, then $m_{s,t}$ is the entry in M in the row indexed by s and the column indexed by t. The reader may have noted that such a matrix depends not just on S, T, but also on their linear orderings. However, changing these orderings merely permutes rows and columns in M which will usually have no effect on the information we wish to extract from it.

If $G = (V, E)$ is a graph, then there are several important matrices associated with it. The *adjacency matrix* of G is the $V \times V$ matrix $A = A(G)$ with

$$a_{v,w} = \begin{cases} 1 & \text{if } vw \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Using the ordering v, w, x, y, the graph on the left in Figure 2.6 has adjacency matrix

$$A = \begin{array}{c c} & \begin{array}{cccc} v & w & x & y \end{array} \\ \begin{array}{c} v \\ w \\ x \\ y \end{array} & \left[\begin{array}{cccc} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{array}\right] \end{array}.$$

The adjacency matrix is always symmetric since vw and wv denote the same edge. It also has zeros on the diagonal since our graphs are (usually) loopless.

A second matrix associated with G is its *incidence matrix*, $B = B(G)$, which is the $V \times E$ matrix with entries

$$b_{v,e} = \begin{cases} 1 & \text{if } v \text{ is an endpoint of } e, \\ 0 & \text{otherwise.} \end{cases}$$

Returning to our example, the graph has

$$B = \begin{array}{c} \\ v \\ w \\ x \\ y \end{array} \begin{array}{c} \begin{array}{ccccc} e & f & g & h & i \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

By construction, row v of B contains $\deg v$ ones, and every column contains 2 ones. We will also need the diagonal $V \times V$ matrix $C(G)$ which has diagonal entries $c_{v,v} = \deg v$. These three matrices are nicely related.

Proposition 2.6.1. *For any graph G we have*

$$BB^t = A + C.$$

Proof. The (v, w) entry of BB^t is the inner product of rows v and w of B. If $v = w$, then this is, using the notation (1.9),

$$\sum_e b_{v,e}^2 = \sum_e \delta(v \text{ is an endpoint of } e)^2 = \deg v = c_{v,v}$$

since $0^2 = 0$ and $1^2 = 1$. Similarly, if $v \neq w$, then the entry is

$$\begin{aligned} \sum_e b_{v,e} b_{w,e} &= \sum_e \delta(v \text{ is an endpoint of } e) \cdot \delta(w \text{ is an endpoint of } e) \\ &= \delta(vw \in E) \\ &= a_{v,w}, \end{aligned}$$

which completes the proof. □

Interestingly, to compute the number of spanning trees of G we will have to turn G into a digraph. An *orientation* of G is a digragh D with $V(D) = V(G)$ and, for each edge $vw \in E(G)$, either the arc $\overrightarrow{vw}$ or the arc $\overrightarrow{wv}$ in $A(D)$. In this case G is called the *underlying graph* of D. The digraph on the right in Figure 2.6 is an orientation of our running example graph G. The adjacency matrix of a digraph is defined just as for graphs and will not concern us here. But we will need the *directed incidence matrix*, $B = B(D)$, defined by

$$b_{v,a} = \begin{cases} -1 & \text{if } a = \overrightarrow{vw} \text{ for some } w, \\ 1 & \text{if } a = \overrightarrow{wv} \text{ for some } w, \\ 0 & \text{otherwise.} \end{cases}$$

For the digraph in Figure 2.6 we have

$$B = \begin{array}{c} \\ v \\ w \\ x \\ y \end{array} \begin{array}{c} \begin{array}{ccccc} e & f & g & h & i \end{array} \\ \left[\begin{array}{rrrrr} -1 & -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & -1 & 1 \\ 0 & 1 & 0 & 0 & -1 \end{array}\right] \end{array}.$$

Here are the two properties of $B(D)$ which will be important for us.

Proposition 2.6.2. *Let D be a digraph and let $B = B(D)$.*

(a) *If the rows of B are $\mathbf{b}_1, \ldots, \mathbf{b}_n$, then*

$$\mathbf{b}_1 + \cdots + \mathbf{b}_n = \mathbf{0}$$

where $\mathbf{0}$ is the zero vector.

(b) *If D is an orientation of a graph G, then*

$$BB^t = C(G) - A(G). \tag{2.13}$$

Proof. For (a), just note that every column of B contains a single 1 and a single -1, which will cancel in the sum. The proof of (b) is similar to that for Proposition 2.6.1 and so is left to the reader. □

It is interesting to note that although the matrix B on the left-hand side of (2.13) depends on D, the right-hand side only depends on the underlying graph G. The matrix $L(G) = C(G) - A(G)$ is called the *Laplacian* of G and controls many combinatorial aspects of the graph. Returning to our example, we have

$$L(G) = \left[\begin{array}{rrrr} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 2 \end{array}\right].$$

Note that the sum of the rows of $L = L(G)$ is zero since, for all $v \in V$, column v contains $\deg v$ on the diagonal and then $\deg v$ other nonzero entries which are all -1. So $\det L = 0$. But removing the last row and column of the previous displayed matrix and taking the determinant gives

$$\det \left[\begin{array}{rrr} 3 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 3 \end{array}\right] = 8.$$

The reader may recall that 8 was also the number of spanning trees of G. This is not a coincidence! But before we can prove the implied theorem, we need one more result.

Let M be an $S \times T$ matrix and let $I \subseteq S$ and $J \subseteq T$. Let $M_{I,J}$ denote the submatrix of M whose rows are indexed by I and columns by J. In $B(G)$ for our example graph G with $I = \{v, x\}$ and $J = \{f, g, i\}$ we would have

$$B_{I,J} = \left[\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 1 & 1 \end{array}\right].$$

If $I = S - \{s\}$ for some $s \in S$ and $J = T - \{t\}$ for some $t \in T$, then we use the abbreviation $M_{\hat{s},\hat{t}}$ for $M_{I,J}$. In this case when $S = T = [n]$, the (i, j) *cofactor* of M is

$$m_{\hat{\imath},\hat{\jmath}} = (-1)^{i+j} \det M_{\hat{\imath},\hat{\jmath}}.$$

We will need the following famous result about determinants called the Cauchy–Binet Theorem. Since this is really a statement about linear algebra rather than combinatorics, we will just outline a proof in the exercises.

Theorem 2.6.3 (Cauchy–Binet Theorem). *Let Q be an $[m] \times [n]$ matrix and let R be $[n] \times [m]$. Then*

$$\det QR = \sum_{K \in \binom{[n]}{m}} \det Q_{[m],K} \cdot \det R_{K,[m]}. \qquad \square$$

Note that in the special case $m = n$ this reduces to the well-known statement that $\det QR = \det Q \cdot \det R$.

Theorem 2.6.4 (Matrix-Tree Theorem). *Let G be a graph with $V = [n]$, $E = [m]$, and let $L = L(G)$. We have for any $i, j \in [n]$*

$$\#\mathcal{ST}(G) = \ell_{\hat{\imath},\hat{\jmath}}.$$

Proof. We will do the case when $i = j = n$ as the other cases are similar. So

$$\ell_{\hat{n},\hat{n}} = (-1)^{n+n} \det L_{\hat{n},\hat{n}} = \det L_{\hat{n},\hat{n}}.$$

Let D be any orientation of G and $B = B(D)$. By Proposition 2.6.2(b), we have that $L = C(G) - A(G) = BB^t$. It follows that

$$L_{\hat{n},\hat{n}} = B_{W,E}(B_{W,E})^t$$

where $W = [n-1]$. Applying the Cauchy–Binet Theorem we get

$$\ell_{\hat{n},\hat{n}} = \sum_{F \in \binom{E}{n-1}} \det B_{W,F} \cdot \det(B_{W,F})^t = \sum_{F \in \binom{E}{n-1}} (\det B_{W,F})^2.$$

So the theorem will be proved if we can show that

$$\det B_{W,F} = \begin{cases} \pm 1 & \text{if the edges of } F \text{ are a spanning tree of } G, \\ 0 & \text{otherwise.} \end{cases} \tag{2.14}$$

Note that $B_{W,F}$ is the incidence matrix of the digraph D_F having $V(D_F) = V$ and $A(D_F) = F$ but with the row of vertex n removed. We say that D_F is a tree if its underlying graph is one.

We first consider the case when D_F is not a tree. We know $\#F = n-1$ so, by Theorem 1.10.2, D_F must be disconnected. Thus there is a component of D_F not containing the vertex n. And the sum of the row vectors of $B_{W,F}$ corresponding to that component is $\mathbf{0}$ by Proposition 2.6.2(a). Thus $\det B_{W,F} = 0$ in this case.

Now suppose that D_F is a tree. To prove this case of (2.14), it suffices to permute the rows and columns of $B_{W,F}$ so that the matrix becomes lower triangular with ± 1 on the diagonal. Such a permutation corresponds to a relabeling of the vertices and edges of D_F. If $n = 1$, then $B_{W,F}$ is the empty matrix which has determinant 1. If $n > 1$, then, by Lemma 1.10.1, D_F has at least two leaves. So in particular there is a leaf in $W = [n-1]$. By relabeling D_F we can assume that $v = 1$ is the leaf and $a = 1$ is the sole arc containing v. It follows that the first row of $B_{W,F}$ has ± 1 in the $(1,1)$ position and zeros elsewhere. Now we consider $D_F - v$ and recurse to finish constructing the matrix. □

We can use this theorem to rederive Cayley's result, Theorem 1.10.3, enumerating all trees on a given vertex set. To do so, consider the complete graph K_n with vertex set $V = [n]$. Clearly the number of trees on n vertices is the same as the number of spanning trees of K_n. The Laplacian $L(K_n)$ consists of $n-1$ down the diagonal with -1 everywhere else. So $L_{\hat{n},\hat{n}}$ is the same matrix but with dimensions $(n-1)\times(n-1)$. Add all the rows of this matrix to the first row. The result is a first row which is all ones since every column consists of an $n-1$ as well as $n-2$ minus ones in some order. Next add the first row to each of the other rows. This will cancel all the minus ones in those rows as well as changing each diagonal entry from $n-1$ to n. Now the matrix is upper triangular and, since elementary row operations do not change the determinant, we have that $\ell_{\hat{n},\hat{n}}$ is the product of the diagonal entries which consist of a one and $n-2$ copies of n. Cayley's Theorem follows.

Exercises

(1) Let n be a positive integer and let $p_1, \ldots, p_k$ be distinct primes. Prove that the number of integers between 1 and n not divisible by any of the p_i is

$$n - \sum_{1\le i\le k}\left\lfloor \frac{n}{p_i}\right\rfloor + \sum_{1\le i<j\le k}\left\lfloor \frac{n}{p_i p_j}\right\rfloor - \cdots + (-1)^k\left\lfloor \frac{n}{p_1 p_2 \cdots p_k}\right\rfloor.$$

(2) Let $A(n)$ be the number of $\rho \vdash [n]$ such that i and $i+1$ never occur in the same block of ρ for any $i \in [n-1]$.
   (a) Show that
$$A(n) = \sum_{i=0}^{n-1}(-1)^i\binom{n-1}{i}B(n-i)$$
   where $B(n)$ is the nth Bell number.
   (b) Find and prove a similar identity involving the Stirling numbers of the second kind.
   (c) Show that part (a) follows from part (b).

(3) Fix positive integers $k \le n$. Use the Principle of Inclusion and Exclusion to find a formula for the number of compositions $\alpha = [\alpha_1, \ldots, \alpha_k] \models n$ with the property that $\alpha_i \ge 2$ for all $i \in [k]$.

(4) Prove that for $n \geq 3$ we have

$$D(n) = (n-1)(D(n-1) + D(n-2))$$

in two ways:
(a) by using Theorem 2.1.2,
(b) by a combinatorial argument.

(5) Prove that for $n \geq 1$ we have

$$D(n) = nD(n-1) + (-1)^n.$$

(6) Call two positive integers k, n *relatively prime* if $\gcd(k, n) = 1$ where gcd is the greatest common divisor. The *Euler totient function*, also called the *Euler phi function*, is

$$\phi(n) = \#\{k \in [n] \mid \gcd(k, n) = 1\}.$$

Using the PIE, show that

$$\phi(n) = n \prod_p \left(1 - \frac{1}{p}\right)$$

where the product is over all primes p dividing n.

(7) Given another proof of Lemma 2.2.1 when S is finite by using Theorem 1.5.1.

(8) Fix a set A and subsets $A_1, \ldots, A_n \subseteq A$. Define A_I for $I \subseteq [n]$ by (2.7). Show that

$$A_\emptyset = A.$$

(9) Prove that for the (signed) Stirling numbers of the first kind

$$\sum_k s(n, k) = \begin{cases} 1 & \text{if } n = 0 \text{ or } 1, \\ 0 & \text{if } n \geq 2, \end{cases}$$

using a sign-reversing involution.

(10) Fill in the details of the proof of Theorem 2.2.2.

(11) Consider permutations $\pi \in P(S)$ and $\sigma \in P(T)$ where $S \cap T = \emptyset$. The set of *shuffles* of π and σ is

$$\pi \sqcup\!\sqcup \sigma = \{\tau \in P(S \uplus T) \mid \pi \text{ and } \sigma \text{ are subwords of } \tau\}.$$

For example

$$31 \sqcup\!\sqcup 24 = \{3124, 3214, 3241, 2314, 2341, 2431\}.$$

We take linear combinations of permutations as if they were vectors. For example

$$6(3124) - 7(3241) - 9(3124) + (3241) = -3(3124) - 6(3241).$$

And a set of permutations represents the sum of all the elements in the set with coefficient one. So we would also write

$$31 \sqcup\!\sqcup 24 = 3124 + 3214 + 3241 + 2314 + 2341 + 2431$$

and let the context determine whether $31 \sqcup\!\sqcup 24$ means the set or the sum. Show that

$$\sum_{k\geq 1}(-1)^k \sum_{w_1\cdot w_2\cdot\ldots\cdot w_k=12\ldots n} w_1 \sqcup\!\sqcup\, w_2 \sqcup\!\sqcup \cdots \sqcup\!\sqcup\, w_k = (-1)^n(n\ldots 21),$$

where the sum is over all concatenations $w_1 \cdot w_2 \cdot \ldots \cdot w_k = 12\ldots n$. For example, when $n = 3$, then the concatenations are

$$123 = 1 \cdot 2 \cdot 3 = 1 \cdot 23 = 12 \cdot 3 = 123.$$

Hint: Consider a permutation v contained in a shuffle $w_1 \sqcup\!\sqcup\, w_2 \sqcup\!\sqcup \cdots \sqcup\!\sqcup\, w_k$. Find the largest index $j \geq 0$, if any, such that

(i) $|w_1| = |w_2| = \cdots = |w_j| = 1$ (which implies that $w_i = i$ for $i \in [j]$) and

(ii) $j\ldots 21$ is a subword of v.

Use the relative positions of j and $j+1$ in v together with merging and splitting to find a copy of v in another shuffle of opposite sign.

(12) Prove Lemma 2.3.1.

(13) Here is a way to obtain a direct bijection $g : P_d(n) \to P_o(n)$. Consider $\lambda \in P_d(n)$. Each part p of λ can be uniquely written as $p = q2^r$ for some odd q and integer $r \geq 0$. Replace p by 2^r copies of q to get a partition $\mu = g(\lambda)$. For example, if $\lambda = (6, 4, 1)$, then $6 = 3 \cdot 2^1 = 3 + 3$, $4 = 1 \cdot 2^2 = 1 + 1 + 1 + 1$, and $1 = 1 \cdot 2^0 = 1$. So $g(6, 4, 1) = (3, 3, 1, 1, 1, 1, 1)$.

(a) Prove that g is a bijection.

(b) Prove that g is the same as the bijection obtained using the Involution Principle in the proof of Theorem 2.3.3.

(14) Call a graph G *rooted* if each component has a distinguished vertex called the *root* of that component. Say that two unlabled, rooted graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are *equal* if there is a bijection $f : V_1 \to V_2$ which preserves both the roots (r is a root of G_1 if and only if $f(r)$ is a root of G_2) and edges ($vw \in E_1$ if and only if $f(v)f(w) \in E_2$). Call a rooted tree T *even* if there is some edge rv, where r is the root, such that removing this edge from T and making v the root of its component results in a graph with two equal components. Call a rooted forest *distinct* if all of its component trees are not equal.

(a) Use the Garsia–Milne Involution principle to find a bijection between the rooted forests on n vertices with no component tree being even and the rooted forests on n vertices which are distinct.

(b) Describe a bijection for (a) using the ideas from Exercise 13.

(c) Show that the bijections in (a) and (b) are actually the same.

(15) One can generalize Theorem 2.3.3 in the following way. Fix a positive integer m. Let $P_{<m}(n)$ be the set of $\lambda \vdash n$ where each part is repeated fewer than m times. Let $P_{\not\equiv m}(n)$ be the set of $\lambda \vdash n$ such that none of the parts is divisible by m.

(a) Show that $P_{<2}(n) = P_d(n)$ and $P_{\not\equiv 2}(n) = P_o(n)$.

(b) Prove that $\#P_{<m}(n) = \#P_{\not\equiv m}(n)$ by generalizing the bijection of the previous exercise.

(c) Reprove that $\#P_{<m}(n) = \#P_{\not\equiv m}(n)$ using the Involution Principle.

(d) Show that the bijections in (b) and (c) are the same.

(16) Let $\mathcal{S} = (S; S_1, \ldots, S_n)$ where S is a finite set and $S_1, \ldots, S_n$ are subsets. Similarly define $\mathcal{T} = (T; T_1, \ldots, T_n)$. Call $\mathcal{S}$ and $\mathcal{T}$ *sieve equivalent* if $\#S_I = \#T_I$ for all $I \subseteq [n]$.

(a) Use the PIE to show that if $\mathcal{S}$ and $\mathcal{T}$ are sieve equivalent, then

$$\left| S - \bigcup_{i=1}^{n} S_i \right| = \left| T - \bigcup_{i=1}^{n} T_i \right|.$$

(b) Show that if $\mathcal{S}$ and $\mathcal{T}$ are sieve equivalent, then the Involution Principle can be used to construct a bijection proving (a).

(17) (a) Check that the line L used in the proof of Theorem 2.4.2 has the correct form. Use this equation to verify that $(0, 0)$ and $(k, n - k)$ are on opposite sides of L.

(b) Give a second proof of this theorem using the factorial expression for binomial coefficients.

(c) Give a third proof of this theorem using induction.

(18) Consider lattice paths of length n, starting at the origin and ending at (x, y) and using steps N, E, S, W where $S = [0, -1]$ and $W = [-1, 0]$. Let $r = (n - x - y)/2$ and $s = (n + x - y)/2$.

(a) Show that the number of such paths is given by

$$\binom{n}{r}\binom{n}{s}.$$

Hint: Find a bijection with pairs of EW-lattice paths which are defined in Exercise 33 of Chapter 1.

(b) Show that the number of such paths staying weakly above the x-axis is

$$\binom{n}{r}\binom{n}{s} - \binom{n}{r-1}\binom{n}{s-1}.$$

(c) Show that for integers $n, r \geq 0$ the sequence

$$\binom{n}{r}\binom{n}{0}, \binom{n}{r-1}\binom{n}{1}, \ldots, \binom{n}{0}\binom{n}{r}$$

is unimodal.

(19) Let D be a digraph.

(a) Show that any directed walk from u to v with $u \neq v$ contains a directed path from u to v.

(b) Show that any directed walk of length at least 2 from u to v with $u = v$ contains a directed cycle.

(20) Show that for $n \in \mathbb{N}$ the sequence

$$\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$$

is log concave by using the formula for a binomial coefficient in terms of factorials.

(21) Let $a_0, a_1, \dots, a_n$ be a sequence of positive reals. Show that the sequence is log-concave if and only if for all $0 < k \le l < n$ we have

$$a_k a_l \ge a_{k-1} a_{l+1}.$$

Hint: Use the ideas in the proof of Proposition 2.5.1.

(22) (a) Let $t(n,k)$ be a triangular array of real numbers for $0 \le k \le n$. Call the array *log-concave in* k if the sequence $t(n,0), \dots, t(n,n)$ is log-concave for all n. Suppose that the $t(n,k)$ satisfy the recursion

$$t(n,k) = a(n,k)t(n-1,k-1) + b(n,k)t(n-1,k)$$

for $n \ge 1$ where $a(n,k), b(n,k), t(n,k)$ are nonnegative reals and $a(n,k) = b(n,k) = t(n,k) = 0$ for $k < 0$ or $k > n$. Also assume that
(i) $a(n,k)$ and $b(n,k)$ are log-concave in k and
(ii) $a(n,k-1)b(n,k+1) + a(n,k+1)b(n,k-1) \le 2a(n,k)b(n,k)$ for $n \ge 1$.
Prove that $t(n,k)$ is log-concave in k.
(b) Use part (a) to prove that $\binom{n}{k}, c(n,k)$ (unsigned Stirling numbers of the first kind), and $S(n,k)$ (Stirling numbers of the second kind) are all log-concave in k.

(23) Suppose $0 \le k < n$. Prove in two ways that

$$\binom{n}{k}^2 \ge \binom{n-1}{k}\binom{n+1}{k},$$

by using the expression for binomial coefficients in terms of factorials and by using lattice paths.

(24) Check that Ω as defined for general path families $P = (P_1, \dots, P_n)$ is a sign-reversing involution.

(25) Prove Theorem 2.5.5.

(26) Consider the sequence $c(n,0), \dots, c(n,n)$ of signless Stirling numbers of the first kind.
(a) Use Lemma 2.5.2 to prove that this sequence is log-concave. Hint: Try to construct D with $V = \mathbb{Z}^2$ such that the number of paths from $(0,0)$ to (n,k) is $c(n,k)$. It will be helpful to use multiple, but distinguishable, arcs.
(b) Use Lemma 2.5.4 to show that, in fact, this is a PF sequence.

(27) (a) Find a sequence of positive reals which is unimodal but not log-concave.
(b) Find a sequence of positive reals which is log-concave but not PF.

(28) (a) Show that the (v,w) entry of $A(G)^n$ is the number of walks going from v to w of length n.
(b) Show that a similar result holds for digraphs.

(29) Use the matrix $B(G)$ to prove the Handshaking Lemma, Theorem 1.9.3.

(30) Prove Proposition 2.6.2(b).

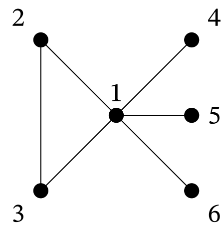


Figure 2.7. The graph G_6

(31) Give two proofs of Theorem 2.6.3 as follows.
  (a) Give one proof using the Lindström–Gessle–Viennot Lemma.
  (b) Give a second demonstration based on the outline below.
    (i) Show that if $m > n$, then both sides are zero.
    (ii) Assume that $m \le n$, write out the entries of QR, and expand about the columns of the product using multilinearity to show that
$$\det QR = \sum_{\pi \in P(([n],m))} (\det Q_{\bullet,\pi}) r_{\pi_1,1} r_{\pi_2,2} \cdots r_{\pi_m,m}$$
    where $Q_{\bullet,\pi}$ is the matrix whose jth column is the π_j column of Q.
    (iii) Show that in the previous sum, $\det Q_{\bullet,\pi} = 0$ if π contains a repeated entry.
    (iv) Show that if $K \in \binom{[n]}{m}$, then $\det Q_{[m],K}$ can be factored out of all the terms in the sum where π is a permutation of K and that what remains sums to $\det R_{K,[m]}$.

(32) Prove the case of Theorem 2.6.4 where $i = 1$ and $j = 2$.

(33) Let G_n be the graph with vertex set $V = [n]$ and edge set
$$E = \{12,\ 13,\ 14,\ \ldots,\ 1n,\ 23\}.$$
Graph G_6 is displayed in Figure 2.7. Find the number of spanning trees of G_n in two ways: by a direct count and by using the Matrix-Tree Theorem.

(34) The *complete bipartite graph*, $K_{m,n}$, has $V = \{v_1, \ldots, v_m, w_1, \ldots, w_n\}$ and edge set consisting of $v_i w_j$ for all i, j (and no other edges). Show that
$$\#ST(K_{m,n}) = m^{n-1} n^{m-1}.$$

Counting with Ordinary Generating Functions

This chapter introduces one of the most powerful techniques in the enumerator's toolkit: generating functions. Wilf [**101**] wrote a whole book devoted to their properties. There are several types of generating functions and we will start with the simplest, which are called ordinary generating functions. In Chapters 4, 7, and 8 we will deal with other types. The basic idea in all cases is to take a sequence of numbers in which we are interested and replace it by an algebraic object, namely a polynomial or power series. The advantage of doing this is that one can then bring a host of algebraic techniques to bear in order to study the original sequence. This makes it possible to give proofs of results about the sequence which have the following advantages:

(1) The proofs can be very short.
(2) Many demonstrations can be done by straightforward manipulations which do not require the cleverness of other approaches.
(3) Sometimes no other method is known for obtaining a given result.

3.1. Generating polynomials

Let x be a variable. A sequence

$$a_0, a_1, a_2, \dots, a_n \tag{3.1}$$

of complex numbers has *ordinary generating polynomial*

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = \sum_{k=0}^{n} a_kx^k.$$

Here, "ordinary" is to distinguish this generating polynomial from other types. Since we will only be dealing with the ordinary case in this chapter, we will usually drop the adjective. Note that $f(x)$ is an element of the algebra $\mathbb{C}[x]$ of polynomials in x

with complex coefficients. We will also often call $f(x)$ the *generating function* for the sequence (3.1) since it is a special case of the generating function for a sequence with a countable, but perhaps not finite, number of terms. This more general setting will be discussed in Section 3.3.

To begin with a simple example, consider the sequence of binomial coefficients found in a row of Pascal's triangle

$$\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \dots, \binom{n}{n}.$$

The corresponding generating function is

$$f(x) = \sum_{k=0}^{n} \binom{n}{k} x^k.$$

In particular, when $n = 4$ we get

$$f(x) = 1 + 4x + 6x^2 + 4x^3 + x^4 = (1+x)^4.$$

The power of this generating function is that it can be expressed as a product which is just the well-known Binomial Theorem. We will give two proofs of this result, one combinatorial and one using algebraic manipulations.

Theorem 3.1.1 (Binomial Theorem). *For $n \in \mathbb{N}$ we have*

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

Proof (Combinatorial). Consider expanding the product

$$(1+x)^n = \overbrace{(1+x)(1+x)\cdots(1+x)}^{n}$$

using the distributive law. One obtains a term x^k in the expansion by picking the x in k of the factors and picking the 1 in the remaining $n-k$. But the number of ways of choosing k objects from n objects is $\binom{n}{k}$. So that is the coefficient of x^k in the product and we are done. □

Proof (Algebraic). We will induct on n. The result is clearly true for $n = 0$ so assume $n \geq 1$. Note that, because of our conventions for binomial coefficients, we can write the generating function as

$$\sum_{k=0}^{n} \binom{n}{k} x^k = \sum_{k=-\infty}^{\infty} \binom{n}{k} x^k.$$

The advantage of doing this is that we will not have to worry about boundary cases when $k = 0$ or $k = n$ and so we will suppress the limits. Now using the binomial recursion in Theorem 1.3.3(a), reindexing, and induction

$$\begin{aligned}\sum_k \binom{n}{k} x^k &= \sum_k \binom{n-1}{k-1} x^k + \sum_k \binom{n-1}{k} x^k \\ &= x\sum_k \binom{n-1}{k-1} x^{k-1} + \sum_k \binom{n-1}{k} x^k \\ &= x\sum_k \binom{n-1}{k} x^k + \sum_k \binom{n-1}{k} x^k \\ &= x(1+x)^{n-1} + (1+x)^{n-1} \\ &= (1+x)^n\end{aligned}$$

as desired. □

The first proof illustrates the use of the Product Rule for weight-generating functions which will be discussed in Section 3.4. The second proof is an example of the point made in the chapter introduction about how proofs involving generating functions can be based on routine manipulations. And the trick of extending the domain of summation is one which we will often use to simplify demonstrations. We now wish to give an illustration of how a generating function, once derived, can be used to give simple proofs of other results. In particular, setting $x = 1$ in the Binomial Theorem we immediately get

$$\sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n,$$

which is part (c) of Theorem 1.3.3. Similarly, letting $x = -1$ in Theorem 3.1.1 gives

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = (1-1)^n = 0^n = \delta_{0,n},$$

which is Theorem 1.3.3(d).

We end this section by stating the generating function for the Stirling numbers of the first kind. This result can be proved similarly to the algebraic proof of the Binomial Theorem so its demonstration will be left as an exercise. Finding a generating function for the Stirling numbers of the second kind will have to wait until after we have discussed formal power series in Section 3.3.

Theorem 3.1.2. *For $n \in \mathbb{N}$ we have*

$$\sum_{k=0}^{n} c(n,k) x^k = x(x+1)(x+2)\dots(x+n-1).$$ □

Note that by setting $x = 1$ in the previous displayed equation we obtain the special case

$$\#P([n]) = \sum_k c(n,k) = n!\,.$$

So this proposition can be considered a generalization of Theorem 1.2.1. Such extensions are called q-analogues and will be discussed in the next section.

3.2. Statistics and q-analogues

One way of constructing generating functions is through the use of statistics and q-analogues. Because of connections with the theory of hypergeometic series, the variable q is usually used for these generating functions. This is a mnemonic choice since sometimes, as we will see below, q stands for the power of a prime p. There is no formal definition of a q-analogue, so we will start with an example which will illustrate the meta-definition we will eventually give.

A *statistic* on a set S is a function $\mathrm{st}: S \to \mathbb{N}$. Because the range of a statistic is $\mathbb{N}$ we can define, for finite S, a corresponding generating polynomial

$$f(q) = \sum_{s \in S} q^{\mathrm{st}\, s}.$$

This generating function is sometimes called the *distribution* of st over S because it can also be written

$$f(q) = \sum_{k \geq 0} a_k q^k$$

where a_k is the number of $s \in S$ satisfying $\mathrm{st}\, s = k$ and this parallels the distribution of a random variable in probability theory. One of the most famous statistics on permutations is the inversion number. A permutation $\pi = \pi_1 \ldots \pi_n \in P([n])$ has *inversion set*

$$\mathrm{Inv}\, \pi = \{(i, j) \mid i < j \text{ and } \pi_i > \pi_j\}.$$

One can think of this as the set of pairs of indices where the corresponding elements of π are out of their natural increasing order. Note that one uses pairs of indices rather than the elements of π because this makes it easier to generalize this concept to words where repetitions are allowed. For example, if $\pi = \pi_1\pi_2\pi_3\pi_4\pi_5 = 41532$, then

$$\mathrm{Inv}\, \pi = \{(1, 2),\ (1, 4),\ (1, 5),\ (3, 4),\ (3, 5),\ (4, 5)\}.$$

The *inversion number* of π is just

$$\mathrm{inv}\, \pi = \#\, \mathrm{Inv}\, \pi.$$

We will often use the convention of beginning functions having to do with sets with uppercase letters and their corresponding cardinalities with lowercase. Continuing our example, $\mathrm{inv}\, 41532 = 6$. Clearly $\mathrm{inv}: P([n]) \to \mathbb{N}$ is a statistic and it has a very interesting generating polynomial.

Theorem 3.2.1. *For $n \geq 0$ we have*

$$\sum_{\pi \in P([n])} q^{\mathrm{inv}\, \pi} = (1)(1 + q)(1 + q + q^2) \cdots (1 + q + q^2 + \cdots + q^{n-1}).$$

Proof. We will induct on n, omitting the trivial base case. Every $\pi \in P([n])$ can be obtained uniquely from a $\sigma \in P([n-1])$ by inserting n into one of the n spaces between the elements of σ (including the space before σ_1 and the space after σ_{n-1}). Let σ^i be the result of placing n in the ith space from the right where the space after σ_{n-1} is considered space zero. Then clearly

$$\operatorname{inv}\sigma^i = i + \operatorname{inv}\sigma.$$

Using this equation and induction we see that

$$\begin{aligned}\sum_{\pi\in P([n])} q^{\operatorname{inv}\pi} &= \sum_{\sigma\in P([n-1])} \sum_{i=0}^{n-1} q^{\operatorname{inv}\sigma^i}\\ &= \sum_{\sigma\in P([n-1])} q^{\operatorname{inv}\sigma}\cdot\sum_{i=0}^{n-1} q^i\\ &= (1+q)(1+q+q^2)\cdots(1+q+q^2+\cdots+q^{n-1})\end{aligned}$$

as we wished to prove. □

Note that by plugging $q = 1$ into this result one obtains

$$\#P([n]) = \sum_{\pi\in P([n])} 1 = n!\,,$$

which is the second statement in Theorem 1.2.1.

Now that we have met some q-analogues (although they have not been named as such), their meta-definition should make more sense. A *q-analogue* of a combinatorial object $\mathcal{O}$ is an object $\mathcal{O}(q)$ such that

$$\lim_{q\to 1}\mathcal{O}(q) = \mathcal{O}.$$

Note that $\mathcal{O}$ could be many things: a number, a definition, or a theorem. For example, one of the standard q-analogues of $n \in \mathbb{N}$ is the polynomial

$$[n]_q = 1 + q + q^2 + \cdots + q^{n-1}. \tag{3.2}$$

Clearly $[n]_1 = n$. Another possible q-analogue of n is the rational function $(1-q^n)/(1-q)$. In this case one cannot just substitute $q = 1$ but must take a limit. Of course, this quotient and $[n]_q$ are equal when $q \neq 1$. Another q-analogue is the *q-factorial*

$$[n]_q! = [1]_q[2]_q\dots[n]_q.$$

So Theorem 3.2.1 can be restated as

$$\sum_{\pi\in P([n])} q^{\operatorname{inv}\pi} = [n]_q!\,.$$

Note that we will sometimes write $[n]_q$ as just $[n]$. This could cause confusion with the use of $[n]$ as a set, so we will only use this simplification if it is clear which of the two possible meanings is meant. Similarly, we will often drop the q subscript from other q-analogues when convenient.

There is another famous statistic which has $[n]_q!$ as its distribution. The *descent set* of $\pi \in P([n])$ is

$$\operatorname{Des}\pi = \{i \mid \pi_i > \pi_{i+1}\} \tag{3.3}$$

with corresponding *descent number* $\operatorname{des}\pi = \#\operatorname{Des}\pi$. Equivalently $i \in \operatorname{Des}\pi$ if and only if $(i, i+1) \in \operatorname{Inv}\pi$. We also define the *ascent set*, $\operatorname{Asc}\pi$, and *ascent number*, $\operatorname{asc}\pi$, analogously by reversing the inequality in definition (3.3). Using our previous example we have $\operatorname{Des} 41532 = \{1, 3, 4\}$ and $\operatorname{des} 41532 = 3$. The *major index* of π is

$$\operatorname{maj}\pi = \sum_{i \in \operatorname{Des}\pi} i.$$

So $\operatorname{maj} 41532 = 1 + 3 + 4 = 8$. The term "major index" was coined by Dominique Foata [**26**] in honor of Percy MacMahon who first studied this statistic [**61**] and was a major in the British army.

Theorem 3.2.2. *For $n \geq 0$ we have*

$$\sum_{\pi \in P([n])} q^{\operatorname{maj}\pi} = [n]_q!\,.$$

Proof. We start as in the proof of Theorem 3.2.1 but now number the spaces of σ differently. First number the spaces between σ_i and σ_{i+1} where i is a descent, as well as the space after σ_{n-1}, from right to left starting with zero. Now number the remaining spaces, including the one before σ_1, from left to right with the numbers $\operatorname{des}\sigma + 1, \operatorname{des}\sigma + 2, \ldots, n - 1$. An example follows this proof.

Let $\sigma^{(j)}$ denote the result of placing n in space j with this maj labeling. We claim that

$$\operatorname{maj}\sigma^{(j)} = j + \operatorname{maj}\sigma. \tag{3.4}$$

Indeed, if space j is in a descent or at the end of σ, then inserting n just moves the j descents to the right of and including the given descent one position to the right. By definition of major index, this adds a total of j to $\operatorname{maj}\sigma$. If space j is in an ascent or at the beginning of σ, then inserting n creates a new descent as well as moving descents to the right of the space one position to the right. It is easy to check for these j that if inserting n in space j caused $\operatorname{maj}\sigma$ to increase by j, then inserting n in place $j + 1$ increases $\operatorname{maj}\sigma$ by $j + 1$. So, by induction, equation (3.4) continues to hold in this range of j. The completion of the proof is now done exactly as in the demonstration of Theorem 3.2.1. $\square$

Continuing on with $\sigma = 41532$ having $\operatorname{maj}\sigma = 8$, the spaces are labeled using subscripts as follows:

$${}_4 4_3 1_5 5_2 3_1 2_0.$$

Inserting 6 into each space in turn gives

j	0	1	2	3	4	5
$\sigma^{(j)}$	415326	415362	415632	461532	641532	416532
$\operatorname{maj}\sigma^{(j)}$	8	9	10	11	12	13

It turns out that there are many permutation statistics whose distribution is $[n]_q!$ and these statistics were dubbed *Mahonian* by Foata. One can consult the article of Babson and Steingrímsson [**3**] for a list of Mahonian statistics.

Having found q-analogues involving permutations, the reader may suspect that they also exist for combinations. For integers $0 \le k \le n$, define the *q-binomial coefficients* or *Gaussian polynomials* to be

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{[n]_q!}{[k]_q!\,[n-k]_q!}.$$

As usual, we let this function be zero if $k < 0$ or $k > n$. For example

$$\begin{aligned} \begin{bmatrix} 4 \\ 2 \end{bmatrix} &= \frac{[4]!}{[2]!\,[2]!} \\ &= \frac{[4][3]}{[2][1]} \\ &= \frac{(1+q+q^2+q^3)(1+q+q^2)}{(1+q)} \\ &= 1+q+2q^2+q^3+q^4. \end{aligned} \tag{3.5}$$

It is not at all clear from the definition just given that this is actually a polynomial in q rather than just a rational function. But this follows easily using induction and our next result. Note that this theorem gives two q-analogues for the ordinary binomial recursion. This illustrates a general principle that q-analogues are not necessarily unique as we have also seen in the inv and maj interpretations of $[n]_q!$.

Theorem 3.2.3. *We have*

$$\begin{bmatrix} 0 \\ k \end{bmatrix}_q = \delta_{0,k}$$

and, for $n \ge 1$,

$$\begin{aligned} \begin{bmatrix} n \\ k \end{bmatrix}_q &= q^k \begin{bmatrix} n-1 \\ k \end{bmatrix}_q + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q \\ &= \begin{bmatrix} n-1 \\ k \end{bmatrix}_q + q^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q. \end{aligned}$$

Proof. The initial condition is trivial. We will prove the first recursion for the q-binomial, leaving the other as an exercise. Using the definition in terms of q-factorials and finding a common denominator gives

$$\begin{aligned} q^k \begin{bmatrix} n-1 \\ k \end{bmatrix} + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} &= \frac{[n-1]!}{[k]!\,[n-k]!}\left(q^k[n-k]+[k]\right) \\ &= \frac{[n-1]!}{[k]!\,[n-k]!} \cdot [n] \\ &= \begin{bmatrix} n \\ k \end{bmatrix} \end{aligned}$$

as desired. □

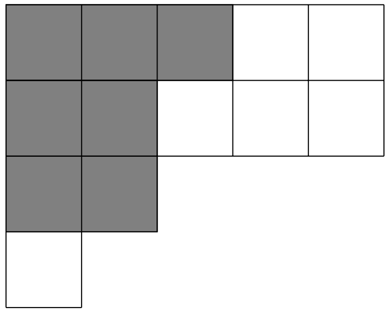


Figure 3.1. The Young diagrams for $(5, 5, 2, 1) \supseteq (3, 2, 2)$

We will now give a q-analogue of the Binomial Theorem (Theorem 3.1.1). Let q, t be two variables.

Theorem 3.2.4. *For $n \geq 0$ we have*

$$(1+t)(1+qt)(1+q^2t)\cdots(1+q^{n-1}t) = \sum_{k=0}^{n} q^{\binom{k}{2}} \begin{bmatrix} n \\ k \end{bmatrix}_q t^k. \tag{3.6}$$

Proof. We will induct on n where the case $n = 0$ is easy to check. For $n > 0$ we can use the second recursion in the previous result and the induction hypothesis to write

$$\begin{aligned}
\sum_k q^{\binom{k}{2}} \begin{bmatrix} n \\ k \end{bmatrix}_q t^k &= \sum_k q^{\binom{k}{2}} \begin{bmatrix} n-1 \\ k \end{bmatrix}_q t^k + \sum_k q^{\binom{k}{2}+n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q t^k \\
&= (1+t)(1+qt)\cdots(1+q^{n-2}t) + q^{n-1}t \sum_k q^{\binom{k-1}{2}} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q t^{k-1} \\
&= (1+t)(1+qt)\cdots(1+q^{n-2}t) + q^{n-1}t(1+t)(1+qt)\cdots(1+q^{n-2}t) \\
&= (1+t)(1+qt)\cdots(1+q^{n-1}t),
\end{aligned}$$

which is what we wished to prove. □

There are many combinatorial interpretations of the q-binomial coefficients. We will content ourselves with presenting two of them here. If $\lambda = (\lambda_1, \ldots, \lambda_k)$ and $\mu = (\mu_1, \ldots, \mu_l)$ are integer partitions, then we say that λ *contains* μ, written $\lambda \supseteq \mu$, if $k \geq l$ and $\lambda_i \geq \mu_i$ for $i \leq l$. Equivalently, the Young diagram of λ contains the Young diagram of μ if they are placed so that their northwest corners align. As an example, $(5, 5, 2, 1) \supseteq (3, 2, 2)$ and Figure 3.1 shows the diagram of λ with the squares of μ shaded inside. The notation $\mu \subseteq \lambda$ should be self-explanatory. Given $\mu \subseteq \lambda$, one also has the corresponding *skew partition*

$$\lambda/\mu = \{(i, j) \in \lambda \mid (i, j) \notin \mu\}. \tag{3.7}$$

The cells of the skew partition in Figure 3.1 are white.

The $k \times l$ *rectangle* is the integer partition whose multiplicity notation is (k^l). Consider the set of partitions contained in this rectangle

$$\mathcal{R}(k,l) = \{\lambda \mid \lambda \subseteq (k^l)\}.$$

Recalling that $|\lambda|$ is the sum of the parts of λ, we consider the generating function $\sum_{\lambda\in\mathcal{R}(k,l)} q^{|\lambda|}$. For example, if $k = l = 2$, then we have

$\lambda \subseteq (2^2)$	$\emptyset$	(1)	(2)	(1^2)	$(2,1)$	(2^2)
$q^{\lvert\lambda\rvert}$	1	q	q^2	q^2	q^3	q^4

which gives

$$\sum_{\lambda\in\mathcal{R}(2,2)} q^{|\lambda|} = 1 + q + 2q^2 + q^3 + q^4.$$

The reader will have noticed the similarity to (3.5), which is not an accident.

Theorem 3.2.5. *For $k, l \geq 0$ we have*

$$\sum_{\lambda\in\mathcal{R}(k,l)} q^{|\lambda|} = \begin{bmatrix} k+l \\ k \end{bmatrix}_q.$$

Proof. We induct on k where the case $k = 0$ is left to the reader. If $k > 0$ and $\lambda \subseteq (k^l)$, then there are two possibilities. Either $\lambda_1 < k$ in which case $\lambda \subseteq ((k-1)^l)$ or $\lambda_1 = k$ so that λ can be written as $\lambda = (k, \lambda')$ where λ' is the partition containing the parts of λ other than λ_1. So $\lambda' \subseteq (k^{l-1})$. Notice that in this case $|\lambda| = |\lambda'| + k$. We now use induction and Theorem 3.2.3 to obtain

$$\begin{aligned}
\sum_{\lambda\in\mathcal{R}(k,l)} q^{|\lambda|} &= \sum_{\lambda\in\mathcal{R}(k-1,l)} q^{|\lambda|} + \sum_{\lambda'\in\mathcal{R}(k,l-1)} q^{|\lambda'|+k} \\
&= \begin{bmatrix} k+l-1 \\ k-1 \end{bmatrix} + q^k \begin{bmatrix} k+l-1 \\ k \end{bmatrix} \\
&= \begin{bmatrix} k+l \\ k \end{bmatrix},
\end{aligned}$$

which finishes the proof. □

For our second combinatorial interpretation of the Gaussian polynomials we will need some linear algebra. Let q be a prime power and let $\mathbb{F}_q$ be the Galois field with q elements. Let V be a vector space of dimension $\dim V = n$ over $\mathbb{F}_q$. We will use $W \leq V$ to indicate that W is a subspace of V. Let

$$\begin{bmatrix} V \\ k \end{bmatrix} = \{W \leq V \mid \dim W = k\}.$$

The subspaces of dimension k are in bijective correspondence with $k \times n$ row-reduced echelon matrices of full rank, that is, with no zero rows. For example, if $n = 4$ and

$k = 2$, then the possible matrices are

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \end{bmatrix},$$

$$\begin{bmatrix} 1 & * & * & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & * & 0 & * \\ 0 & 0 & 1 & * \end{bmatrix}, \begin{bmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \end{bmatrix},$$

where the stars represent arbitrary elements of $\mathbb{F}_q$. So the number of subspaces corresponding to one of these star diagrams is q^s where s is the number of stars. Thus

$$\#\begin{bmatrix} \mathbb{F}_q^4 \\ 2 \end{bmatrix} = 1 + q + 2q^2 + q^3 + q^4,$$

which should look very familiar at this point! Note however that, in contrast to previous cases, this actually represents an integer rather than a polynomial since q is a prime power. Of course, this example generalizes. Because of this result people sometimes talk half-jokingly about sets being vector spaces over the (nonexistent) Galois field with one element.

Theorem 3.2.6. *If V is a vector space over $\mathbb{F}_q$ of dimension n, then*

$$\#\begin{bmatrix} V \\ k \end{bmatrix} = \begin{bmatrix} n \\ k \end{bmatrix}_q.$$

Proof. Given $W \le V$ with $\dim W = k$, we first count the number of possible ordered bases $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ for W. Note that since $\dim V = n$ we have $\#V = \#\mathbb{F}_q^n = q^n$. We can pick any nonzero vector for $\mathbf{v}_1$ so the number of choices is $q^n - 1$. For $\mathbf{v}_2$ we can choose any vector in V which is not in the span of $\mathbf{v}_1$, which gives $q^n - q$ possibilities. Continuing in this way, the total count will be

$$(q^n - 1)(q^n - q)(q^n - q^2)\ldots(q^n - q^{k-1}).$$

By a similar argument, the number of different ordered bases which span a given W of dimension k is

$$(q^k - 1)(q^k - q)(q^k - q^2)\ldots(q^k - q^{k-1}).$$

So the number of possible W's is

$$\begin{aligned} \frac{(q^n-1)(q^n-q)\ldots(q^n-q^{k-1})}{(q^k-1)(q^k-q)\ldots(q^k-q^{k-1})} &= \frac{q^{\binom{k}{2}}(q^n-1)(q^{n-1}-1)\ldots(q^{n-k+1}-1)}{q^{\binom{k}{2}}(q^k-1)(q^{k-1}-1)\ldots(q-1)} \\ &= \frac{(q-1)^k[n][n-1]\ldots[n-k+1]}{(q-1)^k[k][k-1]\ldots[1]} \\ &= \begin{bmatrix} n \\ k \end{bmatrix}, \end{aligned}$$

as advertised. □

There is a beautiful proof of this result due to Knuth [**50**] using row-reduced echelon matrices as in the previous example. The reader will be asked to supply the details in the exercises.

3.3. The algebra of formal power series

We now wish to generalize the concept of generating function from finite to countably infinite sequences. To do so, we will have to use power series. But we wish to avoid the questions of convergence which come up when using analytic power series. Instead, we will work in the algebra of formal power series. This will mean that we have to be careful since, in an algebra, one is only permitted to apply an operation like addition or multiplication a finite number of times. But there is another concept of convergence which will take care of this issue. We should note that there is a whole branch of combinatorics which uses analytic techniques to extract useful information about a sequence, such as its rate of growth, from the corresponding power series. For information about this approach, see the book of Flajolet and Sedgewick [**25**].

A *formal power series* is an expression of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots = \sum_{n=0}^{\infty} a_n x^n,$$

where the a_n are complex numbers. We also say that $f(x)$ is the *ordinary generating function* or *ogf* for the sequence a_n, $n \geq 0$. Often we will leave out the adjective "ordinary" in this chapter since we will not have met any other type of generating function yet.

Note that these series are considered formal in the sense that the powers of x are just place holders and we are not permitted to substitute a value for x. Because of this rule, analytic convergence is not an issue and we can happily talk about formal power series such as $\sum_{n\geq 0} n!\, x^n$ which converge nowhere except at $x = 0$. We will use the notation

$$\mathbb{C}[[x]] = \left\{ \sum_{n\geq 0} a_n x^n \mid a_n \in \mathbb{C} \text{ for all } n \geq 0 \right\}.$$

The set is an algebra, the *algebra of formal power series*, under the three operations of addition, scalar multiplication, and multiplication defined by

$$\sum_{n\geq 0} a_n x^n + \sum_{n\geq 0} b_n x^n = \sum_{n\geq 0} (a_n + b_n) x^n,$$

$$c \sum_{n\geq 0} a_n x^n = \sum_{n\geq 0} (c a_n) x^n,$$

$$\sum_{n\geq 0} a_n x^n \cdot \sum_{n\geq 0} b_n x^n = \sum_{n\geq 0} c_n x^n,$$

where $c \in \mathbb{C}$ and

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

The reader may object that, as mentioned earlier, in an algebra one is only permitted a finite number of additions yet the very elements of $\mathbb{C}[[x]]$ seem to involve infinitely many. But this is an illusion. Remember that x is a formal parameter so that the expression $\sum_n a_n x^n$ is only meant to be a mnemonic device which gives intuition to the definitions of the three algebra operations, especially that of multiplication. We could just as easily have defined $\mathbb{C}[[x]]$ to be the set of all complex vectors $(a_0, a_1, a_2, \dots)$ subject to the operation of vector addition

$$(a_0, a_1, a_2, \dots) + (b_0, b_1, b_2, \dots) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots)$$

and similarly for the other two. What *is* true is that one is only permitted to add or multiply a finite number of elements of $\mathbb{C}[[x]]$. So one can only perform operations which will alter the coefficient of a given power of x a finite number of times.

Now given a sequence of complex numbers $a_0, a_1, a_2, \dots$, we associate with it the *ordinary generating function*

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots \in \mathbb{C}[[x]].$$

We will sometimes say that this series *counts* the objects enumerated by the a_n if appropriate. As with generating polynomials, the reason for doing so is to exploit properties of $\mathbb{C}[[x]]$ to obtain information about the original sequence. We will often write this generating function as $\sum_n a_n x^n$, assuming the that range of indices is $n \geq 0$.

Let us start with a simple example. Consider the sequence $1, 1, 1, \dots$ with generating function $\sum_n x^n$. We would like to simplify this as a geometric series to

$$1 + x + x^2 + \cdots = \frac{1}{1-x}. \tag{3.8}$$

But what does the right-hand side even mean since $1/(1-x)$ appears to be a rational function and so not an element of $\mathbb{C}[[x]]$? The way out of this conundrum is to remember that given an element a in an algebra A, it is possible for a to have an inverse, namely an element a^{-1} such that $a \cdot a^{-1} = 1$ where 1 is the identity element of A. So to prove (3.8) in this setting we must show that $\sum_n x^n$ and $1 - x$ are inverses. This is easily done by using the distributive law:

$$\begin{aligned}(1-x)(1 + x + x^2 + \cdots) &= (1 + x + x^2 + \cdots) - x(1 + x + x^2 + \cdots)\\ &= (1 + x + x^2 + \cdots) - (x + x^2 + x^3 + \cdots)\\ &= 1.\end{aligned}$$

This example illustrates a general principle that often well-known results about analytic power series carry over to their formal counterparts, although some work may be required to check that this is true. For the most part, we will assume the truth of a standard formula in this setting without further comment. But it would be wise to also give a couple of examples to show that caution may be needed. One illustration is that the expression $1/x$ has no meaning in $\mathbb{C}[[x]]$ because x does not have an inverse. For suppose we have $xf(x) = 1$ for some formal power series $f(x)$. Then on the left-hand side the constant coefficient is 0 while on the right it is 1, a contradiction.

As another example, consider the sequence $1/n!$ for $n \geq 0$. We would like to write

$$e^x = \sum_{n\geq 0} \frac{x^n}{n!}$$

for the corresponding generating function but, again, run into the problem that e^x is not a priori an element of $\mathbb{C}[[x]]$. The solution this time is to *define* e^x to be a formal symbol which stands for this power series. Then, of course, to be complete we would need to verify formally that all the usual rules of exponents hold such as $e^{2x} = (e^x)^2$. We will not take the time to do this. But we will point out a case where the rules do not hold. In particular, in $\mathbb{C}[[x]]$ one can*not* write

$$e^{1+x} = ee^x.$$

This is because the left-hand side is not well-defined. Indeed, when expanding $\sum_n (1+x)^n/n!$ there are infinitely many additions needed to compute the coefficient of any given power of x which, as we have already noted, is not permitted.

Although we will not verify every specific analytic identity needed for formal power series in this text, it would be good to have some general results about which operations are permitted in $\mathbb{C}[[x]]$. First we deal with the issue of when a formal power series is invertible.

Theorem 3.3.1. *If $f(x) = \sum_n a_n x^n$, then $f(x)^{-1}$ exists in $\mathbb{C}[[x]]$ if and only if $a_0 \neq 0$.*

Proof. For the forward direction, suppose $f(x)g(x) = 1$ where $g(x) = \sum_n b_n x^n$. Taking the constant coefficient on both sides gives $a_0 b_0 = 1$. So $a_0 \neq 0$.

Now assume $a_0 \neq 0$. We will construct an inverse $g(x) = \sum_n b_n x^n$. We want $f(x)g(x) = 1$. Comparing coefficients of x^n on both sides we see that we wish to have $a_0 b_0 = 1$ and, for $n \geq 1$,

$$a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0 = 0.$$

Since $a_0 \neq 0$ we can take $b_0 = 1/a_0$. By the same token, when $n \geq 1$ we can solve for b_n in the previous displayed equation giving a recursive formula for its value. Thus we can construct such a $g(x)$ and are done. □

Our example with e^x shows that we also need to be careful about substitution. We wish to define the *substitution* of $g(x)$ into $f(x) = \sum_n a_n x^n$ to be

$$f(g(x)) = \sum_{n\geq 0} a_n g(x)^n.$$

But now the right-hand side is an infinite sum of formal power series, not just formal variables. To be able to talk about such sums, we need to introduce a notion of convergence in $\mathbb{C}[[x]]$.

It will be convenient to have the notation that for a formal power series $f(x)$

$$[x^n]f(x) = \text{the coefficient of } x^n \text{ in } f(x),$$

which we have usually been calling a_n. Suppose that we have a sequence $f_0(x), f_1(x), f_2(x), \ldots$ of formal power series. We say that this sequence *converges* to $f(x) \in \mathbb{C}[[x]]$

and write

$$\lim_{k\to\infty} f_k(x) = f(x),$$

if, for any n, the coefficient of x^n in the sequence is eventually constant and equals the coefficient of x^n in $f(x)$. Formally, given n, there exists a corresponding K such that $[x^n]f_k(x) = [x^n]f(x)$ for all $k \geq K$. Otherwise we say that the sequence *diverges* or that the limit *does not exist*.

As an illustration, consider the sequence

$$f_0(x) = 1,\ f_1(x) = 1 + x,\ f_2(x) = 1 + x + x^2,\ \ldots$$

so that $f_k(x) = 1 + x + \cdots + x^k$. Then this sequence has a limit; namely

$$\lim_{k\to\infty} f_k(x) = \sum_{n\geq 0} x^n = \frac{1}{1-x}.$$

To prove this, note that given n we can let $K = n$. So for $k \geq n$ we have $[x^n]f_k = [x^n]f_n = 1$. On the other hand, consider the sequence

$$f_0(x) = 1 + x,\ f_1(x) = 1/2 + x/2,\ f_2(x) = 1/4 + x/4,\ \ldots$$

and in general $f_k(x) = 1/2^k + x/2^k$. This sequence does *not* converge in $\mathbb{C}[[x]]$ since for any n we have that $[x]f_k(x)$ is always different for different k. This is in contrast to the analytic situation where this sequence converges to zero.

As in analysis, we now use convergence of sequences to define convergence of series. Given $f_0(x), f_1(x), f_2(x), \ldots$, we say that their *sum exists and converges to* $f(x)$, written $\sum_{k\geq 0} f_k(x) = f(x)$, if

$$\lim_{k\to\infty} s_k(x) = f(x)$$

where

$$s_k(x) = f_0(x) + f_1(x) + \cdots + f_k(x) \tag{3.9}$$

is the kth partial sum. Divergence is defined as expected. Note that this definition is consistent with our notation for formal power series since given a sequence $a_0, a_1, a_2, \ldots$, we can let $f_k(x) = a_k x^k$ and then prove that $\sum_{k\geq 0} f_k(x) = f(x)$ where $f(x) = \sum_{k\geq 0} a_k x^k$.

To state a criterion for convergence of series, it will be useful to define the *minimum degree* of $f(x) = \sum_n a_n x^n$ to be

$$\operatorname{mdeg} f(x) = \text{smallest } n \text{ such that } a_n \neq 0$$

if $f(x) \neq 0$, and let $\operatorname{mdeg} f(x) = \infty$ if $f(x) = 0$. It turns out that to show a sum of power series converges, it suffices to take a limit of integers.

Theorem 3.3.2. *Given $f_0(x), f_1(x), f_2(x), \ldots \in \mathbb{C}[[x]]$, then $\sum_{k\geq 0} f_k(x)$ exists if and only if*

$$\lim_{k\to\infty} (\operatorname{mdeg} f_k(x)) = \infty.$$

Proof. We will prove the forward direction, leaving the other implication as an exercise. We are given that the sequence $s_k(x)$ as defined by (3.9) converges. So given n, there is a K such that

$$[x^n]s_K(x) = [x^n]s_{K+1}(x) = [x^n]s_{K+2}(x) = \cdots .$$

But for $j \geq 0$ we have

$$s_{K+j}(x) = s_K(x) + f_{K+1}(x) + f_{K+2}(x) + \cdots + f_{K+j}(x).$$

It follows that $[x^n]f_k(x) = 0$ for $k > K$. Now given n, take N to be the maximum of all the K-values associated to integers less than or equal to n. From what we have shown, this forces $\operatorname{mdeg} f_k(x) > n$ for $n > N$. But by definition of a limit of real numbers, this means $\lim_{k\to\infty}(\operatorname{mdeg} f_k(x)) = \infty$. □

We are now on a firm footing with our definition of substitution as we know what it means for a sum of power series to converge. We can use the previous result to give a simple criterion for convergence when substituting one generating function into another.

Theorem 3.3.3. *Given $f(x), g(x) \in \mathbb{C}[[x]]$, then the composition $f(g(x))$ exists if and only if*

(1) *$f(x)$ is a polynomial or*

(2) *$g(x)$ has zero constant term.*

Proof. If $f(x)$ is a polynomial, then $f(g(x))$ is a finite sum and so obviously converges. So assume $f(x) = \sum_n a_n x^n$ is not polynomial.

If $g(x)$ has no constant term, we have $\operatorname{mdeg} a_n g(x)^n \geq n$. So the limit in the previous theorem is infinity and $f(g(x))$ is well-defined.

To finish the proof, consider the remaining case where $[x^0]g(x) \neq 0$. Since $f(x)$ is not a polynomial, there are an infinite number of n such that $a_n \neq 0$. But for these n we have $\operatorname{mdeg} a_n g(x)^n = 0$. So the desired limit cannot be infinity and $f(g(x))$ does not exist in $\mathbb{C}[[x]]$. □

We will also find it useful to consider certain infinite products. We approach their convergence just as we did for infinite sums. Given a sequence $f_0(x), f_1(x), f_2(x), \ldots$, we say that their *product exists and converges to* $f(x)$, written $\prod_{k\geq 0} f_k(x) = f(x)$, if

$$\lim_{k\to\infty} p_k(x) = f(x)$$

where

$$p_k(x) = f_0(x)f_1(x)\ldots f_k(x).$$

We have the following result whose proof is similar enough to that of Theorem 3.3.2 that we will leave it to the reader.

Theorem 3.3.4. *Let $f_0(x), f_1(x), f_2(x), \ldots$ be power series with zero constant terms. Then $\prod_{k\geq 0}(1 + f_k(x))$ exists if and only if*

$$\lim_{k\to\infty}(\operatorname{mdeg} f_k(x)) = \infty.$$ □

Let us end this section by showing how the previous result can give simple verifications that a product does or does not exist. Consider $\prod_{k\geq 1}(1+x^k)$. In this case $f_k(x) = x^k$ and $\operatorname{mdeg} x^k = k$. So the desired limit is infinity and this product exists. As we will see in Section 3.5, it counts integer partitions with distinct parts. By contrast, the product $\prod_{k\geq 0}(1+x/2^k)$ does not converge since $\operatorname{mdeg} x/2^k = 1$.

3.4. The Sum and Product Rules for ogfs

Just as for sets there is a Sum Rule and a Product Rule for ordinary generating functions. In order to state these results, we need the idea of a weight-generating function. This approach makes it possible to construct generating functions for various sequences in a very combinatorial manner. As a first application, we make a more deep exploration of the Binomial Theorem.

Let S be a set. Then a *weighting* of S is a function $\operatorname{wt}: S \to \mathbb{C}[[x]]$. Most often if $s \in S$, then $\operatorname{wt} s$ will just be a monomial reflecting some property of s. For example, if st is any statistic on S, then we could take $\operatorname{wt} s = x^{\operatorname{st} s}$. For a more concrete illustration which we will continue to use throughout this section, let $S = 2^{[n]}$ and define for $T \in S$

$$\operatorname{wt} T = x^{|T|}. \tag{3.10}$$

Given a weighted set S, we can form the corresponding *weight-generating function*

$$f(x) = f_S(x) = \sum_{s\in S} \operatorname{wt} s.$$

We must be careful that this sum exists in $\mathbb{C}[[x]]$, and if it does, then we say S is a *summable set*. Of course, when S is finite then it is automatically summable. To illustrate for $S = 2^{[3]}$ we have

T	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$\operatorname{wt} T$	1	x	x	x	x^2	x^2	x^2	x^3

so that

$$f_S(x) = 1 + 3x + 3x^2 + x^3.$$

More generally, for $S = 2^{[n]}$ we have

$$\begin{aligned} f_S(x) &= \sum_{T\in 2^{[n]}} x^{|T|} \\ &= \sum_{k=0}^{n} \sum_{T\in\binom{[n]}{k}} x^k \\ &= \sum_{k=0}^{n} \binom{n}{k} x^k \end{aligned}$$

and we have recovered the generating function for a row of Pascal's triangle.

The following theorem will permit us to manipulate weight-generating functions with ease. For the Sum Rule, if S, T are disjoint weighted sets, then we weight $u \in S \uplus T$

using u's weight in S or in T depending on whether $u \in S$ or $u \in T$, respectively. For arbitrary S, T we weight $S \times T$ by letting

$$\operatorname{wt}(s,t) = \operatorname{wt} s \cdot \operatorname{wt} t.$$

Lemma 3.4.1. *Let S, T be summable sets.*

(a) (*Sum Rule*) *The set $S \cup T$ is summable. If $S \cap T = \emptyset$, then*

$$f_{S \uplus T}(x) = f_S(x) + f_T(x).$$

(b) (*Product Rule*) *The set $S \times T$ is summable and*

$$f_{S \times T}(x) = f_S(x) \cdot f_T(x).$$

Proof. (a) Since S is summable, given any $n \in \mathbb{N}$, there are only a finite number of $s \in S$ such that $\operatorname{wt} s$ has a nonzero coefficient of x^n. And the same is true of T. It follows that only finitely many elements of $S \cup T$ have such a coefficient, which mean this set is summable. To prove the desired equality, we compute as follows:

$$f_{S \uplus T}(x) = \sum_{u \in S \uplus T} \operatorname{wt} u = \sum_{u \in S} \operatorname{wt} u + \sum_{u \in T} \operatorname{wt} u = f_S(x) + f_T(x).$$

(b) The statement about summability of $S \times T$ is safely left as an exercise. Computing the weight-generating function gives

$$f_{S \times T}(x) = \sum_{(s,t) \in S \times T} \operatorname{wt}(s,t) = \sum_{s \in S} \operatorname{wt} s \cdot \sum_{t \in T} \operatorname{wt} t = f_S(x) \cdot f_T(x),$$

so we are done. □

We can now use these rules to derive various generating functions in a straightforward manner. We begin by reproving the Binomial Theorem as stated in Theorem 3.1.1. We have already seen that the summation side is the weight-generating function for $S = 2^{[n]}$. For the product side, it will be useful to reformulate S in terms of multiplicity notation. Specifically, consider

$$S' = \{T' = (1^{m_1}, 2^{m_2}, \ldots, n^{m_n}) \mid m_i = 0 \text{ or } 1 \text{ for all } i\}$$

weighted by

$$\operatorname{wt} T' = x^{\sum_i m_i}.$$

Clearly we have a bijection $f : S \to S'$ given by $f(T) = (1^{m_1}, 2^{m_2}, \ldots, n^{m_n})$ where

$$m_i = \begin{cases} 0 & \text{if } i \notin T, \\ 1 & \text{if } i \in T. \end{cases}$$

(In fact, this is the map used in the proof of Theorem 1.3.1.) Furthermore, this bijection is *weight preserving* in that $\operatorname{wt} f(T) = \operatorname{wt} T$. As a concrete example, if $n = 5$ and $T = \{2, 4, 5\}$, then $f(T) = (1^0, 2^1, 3^0, 4^1, 5^1)$ and $\operatorname{wt} f(T) = x^3 = \operatorname{wt} T$. The advantage of using S' is that it is clearly a weighted product of the sets $\{i^0, i^1\}$ for $i \in [n]$ where $\operatorname{wt} i^0 = 1$ and $\operatorname{wt} i^1 = x$. So we can write, where for distinct elements a and b we use the shorthand $a \uplus b$ for $\{a\} \uplus \{b\}$,

$$S' = \{1^0, 1^1\} \times \{2^0, 2^1\} \times \cdots \times \{n^0, n^1\} = (1^0 \uplus 1^1) \times (2^0 \uplus 2^1) \times \cdots \times (n^0 \uplus n^1).$$

Translating this expression using both parts of Lemma 3.4.1, we obtain

$$\begin{aligned}\sum_{k=0}^{n}\binom{n}{k}x^k &= f_S(x)\\ &= f_{S'}(x)\\ &= (\operatorname{wt}1^0+\operatorname{wt}1^1)(\operatorname{wt}2^0+\operatorname{wt}2^1)\cdots(\operatorname{wt}n^0+\operatorname{wt}n^1)\\ &= (1+x)^n.\end{aligned}$$

Since this was the reader's first example of the use of weight-generating functions, we were careful to write out all the details. However, in practice one is usually more concise, for example, making no distinction between S and S' with the understanding that they yield the same weight-generating function and so can be considered the same set in this situation. We also usually omit checking summability, assuming that such details could be filled in if necessary. We are now ready for a more substantial example, namely the Binomial Theorem for negative exponents.

Theorem 3.4.2. *If $n \in \mathbb{N}$, then*

$$\frac{1}{(1-x)^n} = \sum_{k\ge 0}\left(\!\!\binom{n}{k}\!\!\right)x^k.$$

Proof. The summation side suggests that we should consider

$$S = \{T \mid T \text{ is a multiset on } [n]\}$$

with weight function given by (3.10). We are rewarded for our choice since

$$f_S(x) = \sum_{T\in S}\operatorname{wt}T = \sum_{k\ge 0}\sum_{T\in\left(\!\binom{[n]}{k}\!\right)} x^k = \sum_{k\ge 0}\left(\!\!\binom{n}{k}\!\!\right)x^k.$$

We now write

$$\begin{aligned}S &= \{(1^{m_1},2^{m_2},\ldots,n^{m_n}) \mid m_i \ge 0 \text{ for all } i\}\\ &= (1^0 \uplus 1^1 \uplus 1^2 \uplus \ldots) \times (2^0 \uplus 2^1 \uplus 2^2 \uplus \ldots) \times \cdots \times (n^0 \uplus n^1 \uplus n^2 \uplus \ldots)\end{aligned}$$

with weight function $\operatorname{wt} i^k = x^k$. Using Lemma 3.4.1 yields

$$\begin{aligned}f_S(x) &= (\operatorname{wt}1^0+\operatorname{wt}1^1+\operatorname{wt}1^2+\cdots)\cdots(\operatorname{wt}n^0+\operatorname{wt}n^1+\operatorname{wt}n^2+\cdots)\\ &= (1+x+x^2+\cdots)^n\\ &= \frac{1}{(1-x)^n}\end{aligned}$$

and the theorem is proved. □

There are several remarks which should be made about this result. First of all, contrast it with our first version of the Binomial Theorem. In Theorem 3.1.1 we are counting subsets of $[n]$ where repeated elements are not allowed and the resulting generating function is $(1+x)^n$. In Theorem 3.4.2 we are counting multisets on $[n]$ so that repetitions are allowed and these are counted by $1/(1-x)^n$. We will see another example of this in the next section.

We can also make Theorem 3.4.2 look almost exactly like Theorem 3.1.1. Indeed, if $n \leq 0$, then by Theorem 3.4.2 and equation (1.6) (with $-n$ substituted for n) we have

$$(1+x)^n = \frac{1}{(1-(-x))^{-n}} = \sum_{k\geq 0} \left(\!\!\binom{-n}{k}\!\!\right)(-x)^k = \sum_{k\geq 0} \binom{n}{k} x^k.$$

This is exactly like Theorem 3.1.1 except that we have an infinite series whereas for positive n we have a polynomial.

Analytically, the Binomial Theorem makes sense for any $n \in \mathbb{C}$ as long as $|x| < 1$ so that the series converges. In $\mathbb{C}[[x]]$ one can make sense of $(1+x)^n$ for any rational $n \in \mathbb{Q}$; see Exercise 12 of this chapter, and prove the following.

Theorem 3.4.3. *For any $n \in \mathbb{Q}$ we have*

$$(1+x)^n = \sum_{k\geq 0} \binom{n}{k} x^k. \qquad \square$$

3.5. Revisiting integer partitions

The theory of integer partitions is one place where ordinary generating functions have played a central role. In this context and others it will be necessary to consider infinite products. But then, as we have seen in the previous section, we must take care that these products converge. There is a corresponding restriction on the sets which we can use to construct weight-generating functions. We begin by discussing this matter.

Let S be a weighted set. We say that S is *rooted* if there is an element $r \in S$ called the *root* satisfying

(1) $\operatorname{wt} r = 1$ and

(2) if $s \in S - \{r\}$, then $\operatorname{wt} s$ has zero constant term.

For example, the sets $(n^0, n^1, n^2, \dots)$ used in the proof of Theorem 3.4.2 were rooted with $r = n^0$ since $\operatorname{wt} n^0 = 1$ and $\operatorname{wt} n^k = x^k$ for $k \geq 1$. Given a sequence $S_1, S_2, S_3 \dots$ of rooted sets with S_i having root r_i, their *direct sum* is defined to be

$$\begin{aligned} &S_1 \oplus S_2 \oplus S_3 \oplus \cdots \\ &\quad = \{(s_1, s_2, s_3, \dots) \mid s_i \in S_i \text{ for all } i \text{ and } s_i \neq r_i \text{ for only finitely many } i\}. \end{aligned}$$

Note that when the number of S_i is finite, then their direct sum is the same as their product. But when their number is infinite the root condition kicks in. Note that, because of this condition, we have a well-defined weighting on $\oplus_{i\geq 1} S_i$ given by

$$\operatorname{wt}(s_1, s_2, s_3, \dots) = \prod_{i\geq 1} \operatorname{wt} s_i$$

since the product has only finitely many factors not equal to 1. In addition, the Product Rule in Lemma 3.4.1 has to be modified appropriately to get convergence. But the proof is similar to the former result and so is left as an exercise.

Theorem 3.5.1. *Let $S_1, S_2, S_3, \dots$ be a sequence of summable, rooted sets such that*

$$\lim_{i\to\infty} [\operatorname{mdeg}(f_{S_i}(x) - 1)] = \infty.$$

Then the direct sum $S_1 \oplus S_2 \oplus S_3 \oplus \cdots$ *is summable and*

$$f_{S_1 \oplus S_2 \oplus S_3 \oplus \cdots}(x) = \prod_{i \geq 1} f_{S_i}(x). \qquad \square$$

We will now prove a theorem of Euler giving the generating function for $p(n)$, the number of integer partitions of n. The reader should contrast the proof with that given for counting multisets in Theorem 3.4.2, which has evident parallels.

Theorem 3.5.2. *We have*

$$\sum_{n \geq 0} p(n) x^n = \prod_{i \geq 1} \frac{1}{1 - x^i}.$$

Proof. Motivated by the sum side we consider the set S of all integer partitions λ of all numbers $n \geq 0$ with weight

$$\operatorname{wt} \lambda = x^{|\lambda|}, \tag{3.11}$$

recalling that $|\lambda|$ is the sum of the parts of λ. It follows that

$$f_S(x) = \sum_{\lambda \in S} \operatorname{wt} \lambda = \sum_{n \geq 0} \sum_{|\lambda| = n} x^n = \sum_{n \geq 0} p(n) x^n.$$

We now express S as a direct sum, using multiplicity notation, as

$$\begin{aligned} S &= \{(1^{m_1}, 2^{m_2}, 3^{m_3}, \ldots) \mid m_i \geq 0 \text{ for all } i \text{ and only finitely many } m_i \neq 0\} \\ &= (1^0 \uplus 1^1 \uplus 1^2 \uplus \cdots) \oplus (2^0 \uplus 2^1 \uplus 2^2 \uplus \cdots) \oplus (3^0 \uplus 3^1 \uplus 3^2 \uplus \cdots) \oplus \cdots. \end{aligned}$$

Note that since we want the exponent on $\operatorname{wt} \lambda$ to be the sum of its parts, and i^k represents a part i repeated k times, we must take

$$\operatorname{wt} i^k = x^{ik}$$

in contrast to the weight used in the proof of Theorem 3.4.2. Translating into generating functions by using the previous theorem gives

$$\begin{aligned} f_S(x) &= \prod_{i \geq 1} (\operatorname{wt} i^0 + \operatorname{wt} i^1 + \operatorname{wt} i^2 + \operatorname{wt} i^3 + \cdots) \\ &= \prod_{i \geq 1} (1 + x^i + x^{2i} + x^{3i} + \cdots) \\ &= \prod_{i \geq 1} \frac{1}{1 - x^i}, \end{aligned}$$

which is the desired result. $\square$

The reader will notice that the previous proof actually shows much more. In particular, the factor $1/(1 - x^i)$ is responsible for keeping track of the parts equal to i in λ. We can make this precise as follows.

Proposition 3.5.3. *Given $n \in \mathbb{N}$ and $P \subseteq \mathbb{P}$, let $p_P(n)$ be the number of partitions of n all of whose parts are in P.*

(a) *We have*

$$\sum_{n\geq 0} p_P(n)x^n = \prod_{i\in P} \frac{1}{1-x^i}.$$

(b) *In particular, for $k \in \mathbb{P}$,*

$$\sum_{n\geq 0} p_{[k]}(n)x^n = \frac{1}{(1-x)(1-x^2)\cdots(1-x^k)}.$$

Proof. For (a), one uses the ideas in the proof of Theorem 3.5.2 except that the elements of S only contain components of the form r^{m_r} for $r \in P$. And (b) follows immediately from (a). □

Instead of restricting the set of parts of a partition, we can restrict the number of parts. Recall that $p(n,k)$ is the number of $\lambda \vdash n$ with length $\ell(\lambda) \leq k$.

Corollary 3.5.4. *For $k \geq 0$ we have*

$$\sum_{n\geq 0} p(n,k)x^n = \frac{1}{(1-x)(1-x^2)\cdots(1-x^k)}.$$

Proof. From the previous result, it suffices to show that there is a size-preserving bijection between the partitions counted by $p_{[k]}(n)$ and those counted by $p(n,k)$. The map $\lambda \to \lambda^t$ is such a map. Indeed, λ only uses parts in $[k]$ if and only if $\lambda_1 \leq k$. In terms of Young diagrams, this means that the first row of λ has length at most k. It follows that the first column of λ^t has length at most k, which is equivalent to λ^t having at most k parts. □

In the previous section we pointed out a relationship between the generating functions for sets and for multisets. The same holds for integer partitions. Let $p_d(n)$ be the number of partitions of n into distinct parts as defined in Section 2.3.

Theorem 3.5.5. *We have*

$$\sum_{n\geq 0} p_d(n)x^n = \prod_{i\geq 1}(1+x^i).$$

Proof. Up to now we have been writing out most of the gory details of our proofs by weight-generating function as the reader gets familiar with the method. But by now it should be sufficient just to write out the highlights. We begin by letting S be all partitions of all $n \in \mathbb{N}$ into distinct parts and using the weighting in (3.11). It is routine to show that $f_S(x)$ results in the sum side of the theorem. To get the product side we write

$$\begin{aligned} S &= \{(1^{m_1}, 2^{m_2}, 3^{m_3}, \dots) \mid m_i = 0 \text{ or } 1 \text{ for all } i, \text{ only finitely many } m_i \neq 0\} \\ &= \bigoplus_{i\geq 1}(i^0 \uplus i^1). \end{aligned}$$

The generating function translation is

$$f_S(x) = \prod_{i \geq 1} (1 + x^i)$$

and we are done. □

As mentioned in the introduction to this chapter, one of the reasons for using generating functions is that they can give quick and easy proofs for various results. Here is an example where we reprove Euler's distinct parts-odd parts result, Theorem 2.3.3, which we restate here for convenience. Let $p_o(n)$ be the number of partitions of n into parts all of which are odd.

Theorem 3.5.6 (Euler). *For all $n \geq 0$,*

$$p_o(n) = p_d(n).$$

Proof. It suffices to show that these two sequence have the same generating function. Using Theorems 3.5.3(a) and 3.5.5 as well as multiplying by a strange name for one, we get

$$\begin{aligned}
\sum_{n \geq 0} p_d(n) x^n &= (1+x)(1+x^2)(1+x^3)\cdots \\
&= (1+x)(1+x^2)(1+x^3)\cdots \frac{(1-x)(1-x^2)(1-x^3)\cdots}{(1-x)(1-x^2)(1-x^3)\cdots} \\
&= \frac{(1-x^2)(1-x^4)(1-x^6)\cdots}{(1-x)(1-x^2)(1-x^3)\cdots} \\
&= \frac{1}{(1-x)(1-x^3)(1-x^5)\cdots} \\
&= \sum_{n \geq 0} p_o(n) x^n,
\end{aligned}$$

which completes this short and slick proof. □

3.6. Recurrence relations and generating functions

The reader may have noticed that many of the combinatorial sequences described in Chapter 1 satisfy recurrence relations. If one has a sequence defined by a recursion, then generating functions can often be used to find an explicit expression for the terms of the sequence. It is also possible to glean information from the generating function derived from a recurrence which is hard to extract from the recurrence itself. This section is devoted to exploring these ideas.

We start with a simple algorithm using generating functions to solve a recurrence relation. Given a sequence $a_0, a_1, a_2, \ldots$ defined by a recursion and boundary conditions, we wish to find a self-contained formula for the nth term.

(1) Multiply the recurrence by x^n; usually a good choice for n is the largest index of all the terms in the recurrence. Sum over all $n \geq d$ where d is the smallest index for which the recurrence is valid.

(2) Let
$$f(x) = \sum_{n\geq 0} a_n x^n$$
and express the equation in step (1) in terms of $f(x)$ using the boundary conditions.

(3) Solve for $f(x)$.

(4) Find a_n as the coefficient of x^n in $f(x)$.

We note that partial fraction expansion can be a useful way to accomplish step (4).

For a simple example, suppose that our sequence is defined by $a_0 = 2$ and $a_n = 3a_{n-1}$ for $n \geq 1$. Calculating the first few values we get $a_1 = 2\cdot 3$, $a_2 = 2\cdot 3^2$, $a_3 = 2\cdot 3^3$. So it is easy to guess and then prove by induction that $a_n = 2 \cdot 3^n$. We would now like to obtain this result using generating functions. Step (1) is easy as we just write $a_n x^n = 3a_{n-1}x^n$ and then sum to get
$$\sum_{n\geq 1} a_n x^n = \sum_{n\geq 1} 3a_{n-1}x^n.$$
Letting $f(x)$ be as in step (2) we see that
$$\sum_{n\geq 1} a_n x^n = f(x) - a_0 = f(x) - 2$$
and
$$\sum_{n\geq 1} 3a_{n-1}x^n = 3x \sum_{n\geq 1} a_{n-1}x^{n-1} = 3xf(x)$$
where the last equality is obtained by substituting n for $n-1$ in the sum. For step (3) we have

$$f(x) - 2 = 3xf(x) \implies f(x) - 3xf(x) = 2 \implies f(x) = \frac{2}{1-3x}. \tag{3.12}$$

As far as step (4), we can now expand $1/(1-3x)$ as a geometric series (that is, use (3.8) and substitute $3x$ for x) to obtain
$$f(x) = 2\sum_{n\geq 0} 3^n x^n = \sum_{n\geq 0} 2\cdot 3^n x^n.$$
Extracting the coefficient of x^n we see that $a_n = 2 \cdot 3^n$ as expected.

In the previous example, it was easier to guess the formula for a_n and then prove it by induction rather than use generating functions. However, there are times when it is impossible to guess the solution this way, but generating functions still give a straightforward method for obtaining the answer. An example of this is given by the Fibonacci sequence. Our result will be slightly nicer if we use the definition of this sequence given by (1.1). Following the algorithm, we write
$$\sum_{n\geq 2} F_n x^n = \sum_{n\geq 2} (F_{n-1} + F_{n-2})x^n.$$

Writing $f(x) = \sum_{n\geq 0} F_n x^n$ we obtain

$$\sum_{n\geq 2} F_n x^n = f(x) - F_0 - F_1 x = f(x) - x$$

and

$$\sum_{n\geq 2} (F_{n-1} + F_{n-2})x^n = x(f(x) - F_0) + x^2 f(x) = (x + x^2)f(x).$$

Setting the expressions for the left and right sides equal and solving for $f(x)$ yields

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

For the last step we wish to use partial fractions and so must factor $1 - x - x^2$. Using the quadratic formula, we see that the denominator has roots

$$r_1 = \frac{-1 + \sqrt{5}}{2} \quad \text{and} \quad r_2 = \frac{-1 - \sqrt{5}}{2}.$$

It follows that

$$1 - x - x^2 = \left(1 - \frac{x}{r_1}\right)\left(1 - \frac{x}{r_2}\right)$$

since both sides vanish at $x = r_1, r_2$ and both sides have constant term 1. So we have the partial fraction decomposition

$$f(x) = \frac{x}{\left(1 - \frac{x}{r_1}\right)\left(1 - \frac{x}{r_2}\right)} = \frac{A}{\left(1 - \frac{x}{r_1}\right)} + \frac{B}{\left(1 - \frac{x}{r_2}\right)} \tag{3.13}$$

for constants A, B. Clearing denominators gives

$$x = A\left(1 - \frac{x}{r_2}\right) + B\left(1 - \frac{x}{r_1}\right).$$

Setting $x = r_1$ reduces this equation to $r_1 = A(1 - r_1/r_2)$ and solving for A shows that $A = 1/\sqrt{5}$. Similarly letting $x = r_2$ yields $B = -1/\sqrt{5}$. Plugging these values back into (3.13) and expanding the series,

$$f(x) = \frac{1}{\sqrt{5}} \cdot \sum_{n\geq 0} \frac{x^n}{r_1^n} - \frac{1}{\sqrt{5}} \cdot \sum_{n\geq 0} \frac{x^n}{r_2^n}.$$

By rationalizing denominators one can check that $1/r_1 = (1 + \sqrt{5})/2$ and $1/r_2 = (1 - \sqrt{5})/2$. So taking the coefficient of x^n on both sides of the previous displayed equation gives

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1 + \sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1 - \sqrt{5}}{2}\right)^n. \tag{3.14}$$

This example shows the true power of the generating function method. It would be impossible to guess the formula in (3.14) from just computing values of F_n. In fact, it is not even obvious that the right-hand side is an integer!

Our algorithm can be used to derive generating functions in the case where one has a triangle of numbers rather than just a sequence. Here we illustrate this using the

Stirling numbers. Recall that the signless Stirling numbers of the first kind satisfy the recurrence relation and boundary conditions in Theorem 1.5.2. Translating these to the signed version gives $s(0,k) = \delta_{0,k}$ and

$$s(n,k) = s(n-1,k-1) - (n-1)s(n-1,k)$$

for $n \geq 1$. We wish to find the generating function $f_n(x) = \sum_k s(n,k)x^k$ where we are using the fact that $s(n,k) = 0$ for $k < 0$ or $k > n$ to sum over all integers n. Applying our algorithm we have

$$\begin{aligned} f_n(x) &= \sum_k s(n,k)x^k \\ &= \sum_k [s(n-1,k-1) - (n-1)s(n-1,k)]x^k \\ &= xf_{n-1}(x) - (n-1)f_{n-1}(x) \\ &= (x-n+1)f_{n-1}(x), \end{aligned}$$

giving us a recursion for the sequence of generating functions $f_n(x)$. From the boundary condition for $s(0,k)$ we have $f_0(x) = 1$. It is now easy to guess a formula for $f_n(x)$ by writing out the first few values and proving that pattern holds by induction to obtain the theorem below, which also follows easily from Theorem 3.1.2.

Theorem 3.6.1. *For $n \geq 0$ we have*

$$\sum_k s(n,k)x^k = x(x-1)\cdots(x-n+1). \qquad \square$$

In an entirely analogous manner, one can obtain a generating function for the Stirling numbers of the second kind. Because of the similarity, the proof is left to the reader.

Theorem 3.6.2. *For $k \geq 0$ we have*

$$\sum_n S(n,k)x^n = \frac{x^k}{(1-x)(1-2x)\cdots(1-kx)}. \qquad \square$$

Comparing the previous two results, the reader will note a similar relationship as between generating functions for objects without repetitions (sets, distinct partitions) and those where repetitions are allowed (multisets, ordinary partitions). As already mentioned, this will be explained in Section 3.9.

So far, all the generating functions we have derived from recurrences have been rational functions. This is because the recursions are linear and we will prove a general result to this effect in the next section. We will end this section by illustrating that more complicated generating functions, for example algebraic ones, do arise in practice. Let us consider the Catalan numbers $C(n)$ and the generating function $c(x) = \sum_{n\geq 0} C(n)x^n$. Using the recursion and boundary condition in Theorem 1.11.2 and computing in the way we have become accustomed to, we obtain

$$c(x) = 1 + \sum_{n\geq 1} C(n)x^n = 1 + \sum_{n\geq 1}\left(\sum_{i+j=n-1} C(i)C(j)\right)x^n = 1 + xc(x)^2.$$

Writing $xc(x)^2 - c(x) + 1 = 0$ and solving for $c(x)$ using the quadratic formula yields

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

Two things seem to be wrong with this formula for $c(x)$. First of all, we don't know whether the plus or minus solution is the correct one. And second, we seem to have left the ring of formal power series because we are dividing by x which has no inverse. Both of these can be solved simultaneously by choosing the sign so that the numerator has no constant term. Then one can divide by x simply by reducing the power of each term in the top by one. By Theorem 3.4.3 we see that the generating function for $\sqrt{1-4x} = (1-4x)^{1/2}$ has constant term $\binom{1/2}{0} = 1$. So the correct sign is negative and we have proved the following.

Theorem 3.6.3. *We have*

$$\sum_{n\geq 0} C(n)x^n = \frac{1-\sqrt{1-4x}}{2x}. \qquad \square$$

One can use this generating function to rederive the explicit expression for $C(n)$ in Theorem 1.11.3, and the reader will be asked to carry out the details in the exercises.

3.7. Rational generating functions and linear recursions

The reader may have noticed in the previous section that, both in the initial example and for the Fibonacci sequence, the solution of the recursion for a_n was a linear combination of functions of the form r^n where r varied over the reciprocals of the roots of the denominator of the corresponding generating function. This happens for a wide variety of recursions which we will study in this section. Before giving a theorem which characterizes this situation, we will study one more example to illustrate what can happen.

Consider the sequence defined by $a_0 = 1, a_1 = -4$, and

$$a_n = 4a_{n-1} - 4a_{n-2} \quad \text{for } n \geq 2. \tag{3.15}$$

Following the usual four-step program, we have, for $f(x) = \sum_{n\geq 0} a_n x^n$,

$$\begin{aligned} f(x) - 1 + 4x &= \sum_{n\geq 2} a_n x^n \\ &= 4x \sum_{n\geq 2} a_{n-1}x^{n-1} - 4x^2 \sum_{n\geq 2} a_{n-2}x^{n-2} \\ &= 4x(f(x)-1) - 4x^2 f(x). \end{aligned}$$

Solving for $f(x)$ and evaluating the constants in the partial fraction expansion yields

$$f(x) = \frac{1-8x}{1-4x+4x^2} = \frac{1-8x}{(1-2x)^2} = \frac{4}{1-2x} - \frac{3}{(1-2x)^2}.$$

Taking the coefficient of x^n on both sides using the Theorem 3.4.2 (interchanging the roles of n and k) together with the fact that

$$\left(\!\!\binom{k}{n}\!\!\right) = \binom{n+k-1}{n} = \binom{n+k-1}{k-1} \tag{3.16}$$

gives a final answer of

$$a_n = 4 \cdot 2^n - 3\binom{n+1}{1}2^n = (1-3n)2^n. \tag{3.17}$$

So now, instead of a constant times r^n we have a polynomial in n as the coefficient. And that polynomial has degree less than the multiplicity of $1/r$ as a root of the denominator. These observations generalize.

Consider a sequence of complex numbers a_n for $n \geq 0$. We say that the sequence satisfies a (*homogeneous*) *linear recursion of degree d with constant coefficients* if there is a $d \in \mathbb{P}$ and constants $c_1, \ldots, c_d \in \mathbb{C}$ with $c_d \neq 0$ such that

$$a_{n+d} + c_1 a_{n+d-1} + c_2 a_{n+d-2} + \cdots + c_d a_n = 0. \tag{3.18}$$

To simplify things later, we have put all the terms of the recursion on the left-hand side of the equation and made a_{n+d} the term of highest index rather than a_n. One can also consider the nonhomogeneous case where one has a summand c_{d+1} which does not multiply any term of the sequence, but we will have no cause to do so here. It turns out that the sequences satisfying a recursion (3.18) are exactly the ones having rational generating functions.

Theorem 3.7.1. *Given a sequence a_n for $n \geq 0$ and $d \in \mathbb{P}$, the following are equivalent.*

(a) *The sequence satisfies* (3.18).

(b) *The generating function $f(x) = \sum_{n\geq 0} a_n x^n$ has the form*

$$f(x) = \frac{p(x)}{q(x)} \tag{3.19}$$

where

$$q(x) = 1 + c_1 x + c_2 x^2 + \cdots + c_d x^d \tag{3.20}$$

and deg $p(x) < d$.

(c) *We can write*

$$a_n = \sum_{i=1}^{k} p_i(n) r_i^n$$

where the r_i are distinct, nonzero complex numbers satisfying

$$1 + c_1 x + c_2 x^2 + \cdots + c_d x^d = \prod_{i=1}^{k} (1 - r_i x)^{d_i} \tag{3.21}$$

and $p_i(n)$ is a polynomial with deg $p_i(n) < d_i$ *for all i.*

Proof. We first prove the equivalence of (a) and (b). Showing that (a) implies (b) is essentially an application of our algorithm. Multiplying (3.18) by x^{n+d} and summing over $n \geq 0$ gives

$$\begin{aligned}
0 &= \sum_{n\geq 0} a_{n+d}x^{n+d} + c_1 x \sum_{n\geq 0} a_{n+d-1}x^{n+d-1} + \cdots + c_d x^d \sum_{n\geq 0} a_n x^n \\
&= \left[f(x) - \sum_{n=0}^{d-1} a_n x^n \right] + c_1 x \left[f(x) - \sum_{n=0}^{d-2} a_n x^n \right] + \cdots + c_d x^d f(x) \\
&= q(x)f(x) - p(x)
\end{aligned}$$

where $q(x)$ is given by (3.20) and $p(x)$ is the sum of the remaining terms, which implies $\deg p(x) < d$. Solving for $f(x)$ completes this direction.

To prove (b) implies (a), cross multiply (3.19) and use (3.20) to write

$$p(x) = q(x)f(x) = (1 + c_1 x + c_2 x^2 + \cdots + c_d x^d)f(x).$$

Since $\deg p(x) < d$ we have that $[x^{n+d}]p(x) = 0$ for all $n \geq 0$. So taking the coefficient of x^{n+d} on both sides of the previous displayed equation gives the recursion (3.18).

We now show that (b) and (c) are equivalent. The fact that (b) implies (c) again follows from the algorithm. Specifically, using equations (3.19), (3.20), and (3.21), as well as partial fraction expansion, we have

$$f(x) = \frac{p(x)}{\prod_{i=1}^{k}(1 - r_i x)^{d_i}} = \sum_{i=1}^{k}\sum_{j=1}^{d_i} \frac{A_{i,j}}{(1 - r_i x)^j} \tag{3.22}$$

for certain constants $A_{i,j}$. But by Theorem 3.4.2 and equation (3.16) we have that

$$[x^n]\frac{1}{(1 - r_i x)^j} = \left(\!\!\binom{j}{n}\!\!\right) r_i^n = \binom{n+j-1}{j-1} r_i^n$$

where

$$\binom{n+j-1}{j-1} = \frac{(n+j-1)(n+j-2)\cdots(n+1)}{(j-1)!}$$

is a polynomial in n of degree $j-1$ for any given j. Now taking the coefficient of x^n on both sides of (3.22) gives

$$a_n = \sum_{i=1}^{k} \left[\sum_{j=1}^{d_i} A_{i,j} \binom{n+j-1}{j-1} \right] r_i^n.$$

Calling the polynomial inside the brackets $p_i(n)$, we have derived the desired expansion.

The proof that (c) implies (b) essentially reverses the steps of the forward direction. So it is left as an exercise. □

We note that the preceding theorem is not just of theoretical significance but is also very useful computationally. In particular, because of the equivalence of (a) and (c), one can solve a linear, constant coefficient recursion in a more direct manner without having to deal with generating functions. To illustrate, suppose we have a sequence satisfying (3.18). Then we know that the solution in (c) is in terms of the r_i which are the reciprocals of the roots of $q(x)$ as given by (3.20). To simplify things, we consider the polynomial

$$r(x) = x^d q(1/x) = x^d + c_1 x^{d-1} + c_2 x^{d-2} + \cdots + c_d.$$

Comparison with (3.21) shows that the r_i are the roots of $r(x)$. We now find the $p_i(n)$ by solving for the coefficients of these polynomials using the initial conditions. To be quite concrete, consider again the example (3.15) with which we began this section. Since $a_n - 4a_{n-1} + 4a_{n-2} = 0$ for $n \geq 2$ we factor $r(x) = x^2 - 4x + 4 = (x-2)^2$. So $a_n = p(n)2^n$ where $\deg p(n) < 2$. It follows that $p(n) = A + Bn$ for constants A, B. Plugging in $n = 0$ we get $1 = a_0 = A2^0 = A$. Now letting $n = 1$ gives $-4 = a_1 = (1 + B)2^1$ or $B = -3$. Thus a_n is again given as in (3.17). But this solution is clearly simpler than the first one given. This is called the *method of undetermined coefficients*. Of course, the advantage of using generating functions is that they can be used to solve recursions even when they are not linear and constant coefficient.

There is a striking resemblance between the theory we have developed in this section and the method of undetermined coefficients for solving linear differential equations with constant coefficients. This is not an accident and the material in this section may be considered as part of the theory of finite differences which is a discrete analogue of the theory of differential equations. We will have more to say about finite differences when we study Möbius inversion in Section 5.5.

3.8. Chromatic polynomials

Sometimes generating functions or polynomials appear in unexpected ways. We now illustrate this phenomenon using the chromatic polynomial of a graph.

Let $G = (V, E)$ be a graph. A (*vertex*) *coloring* of G from a set S is a function $c : V \to S$. We refer to S as the *color set*. Figure 3.2 contains a graph which we will be using as our running example together with two colorings using the set $S = \{\text{white, gray, black}\}$. We say that c is *proper* if, for all edges $uv \in E$ we have $c(u) \neq c(v)$. The first coloring in Figure 3.2 is proper while the second is not since the edge vx has both endpoints colored gray. The *chromatic number* of G, denoted $\chi(G)$, is the minimum cardinality of a set S such that there is a proper coloring $c : V \to S$. In our example, $\chi(G) = 3$ because we have displayed a proper coloring with three colors in Figure 3.2 (black, white, and gray), and one cannot use fewer colors because of the triangle uvx.

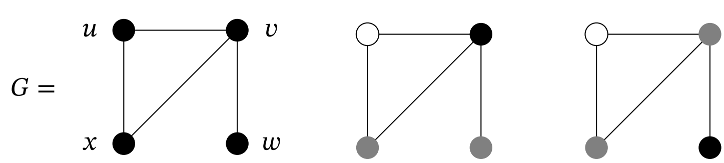


Figure 3.2. A graph and two colorings

The chromatic number is an important invariant in graph theory. But by its definition, it belongs more to extremal combinatorics (which studies structures which minimize or maximize a constraint) than the enumerative side of the subject. Although we will not have much more to say about $\chi(G)$ here, we would be remiss if we did not state one of the most famous mathematical theorems in which it plays a part. Call a graph *planar* if it can be drawn in the plane without any pair of edges crossing.

Theorem 3.8.1 (The Four Color Theorem). *If G is planar, then*

$$\chi(G) \le 4. \qquad \square$$

Note that this result is in stark contrast to ordinary graphs which can have arbitrarily large chromatic number. Complete graphs, for example, have $\chi(K_n) = n$. The Four Color Theorem caused quite a stir when it was proved in 1977 by Appel and Haken (with the help of Koch) [**1**,**2**]. For one thing, it had been the Four Color Conjecture for over 100 years. Also their proof was the first to make heavy use of computers to do the calculations for all the various cases and the demonstration could not be completely checked by a human.

We now turn to the enumerating graph colorings. Let $t \in \mathbb{N}$. The *chromatic polynomial* of G is defined to be

$$P(G;t) = \text{the number of proper colorings } c\colon\ V \to [t].$$

This concept was introduced by George Birkhoff [**13**]. It is not clear at this point why $P(G;t)$ should be called a polynomial, but let us compute it for the graph in Figure 3.2. Consider coloring the vertices of G in the order u, v, w, x. There are t choices for the color of u. This leaves $t-1$ possibilities for v since it cannot be the same color as u. By the same token, the number of choices for w is $t-1$. Finally, x can be colored in $t-2$ ways since it cannot have the colors of u or v and these are different. So the final count is

$$P(G) = P(G;t) = t(t-1)(t-1)(t-2) = t^4 - 4t^3 + 5t^2 - 2t. \tag{3.23}$$

This is a polynomial in t, the number of colors! Before proving that this is always the case, we have a couple of remarks. First of all, there is a close relationship between $P(G;t)$ and $\chi(G)$; namely $P(G;t) = 0$ if $0 \le t < \chi(G)$ but $P(G;\chi(G)) > 0$. This follows from the definitions of P and χ since the latter is the smallest nonnegative integer for which proper colorings of G exist and the former counts such colorings. Secondly, it is not always possible to compute $P(G;t)$ in the manner above and express it as a product of factors $t-k$ for integers k. For example, consider the cycle C_4 with vertices labeled clockwise as u, v, w, x. If we try to use this method to compute $P(C_4;t)$, then everything is fine until we get to coloring x, for x is adjacent to both u and w. But we cannot be sure whether u and w have the same color or not since they themselves are not adjacent.

It turns out that the same ideas can be used both for proving that $P(G;t)$ is always a polynomial in t and to rectify the difficulty in computing $P(C_4;t)$. Consider a graph $G = (V, E)$ and an edge $e \in E$. The graph obtained by *deleting* e from G is denoted $G \setminus e$ and has vertices V and edges $E - \{e\}$. The middle graph in Figure 3.3 is obtained from our running example by deleting $e = vx$. The graph obtained by *contracting* e in G is

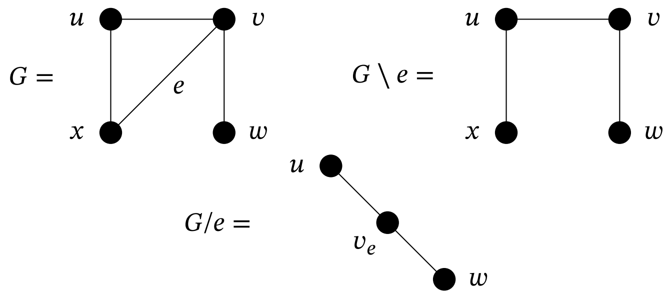


Figure 3.3. Deletion and contraction

denoted G/e and is obtained by shrinking e to a new vertex v_e, making v_e adjacent to the vertices which were adjacent to either endpoint of e and leaving all other vertices and edges of G the same. Contracting vx in our example graph results in the graph on the right in Figure 3.3. The next lemma is crucial in the study of $P(G;t)$. It is ideally set up for induction on $\#E$ since both $G \setminus e$ and G/e have fewer edges than G.

Lemma 3.8.2 (Deletion-Contraction Lemma)**.** *If G is a graph, then for any $e \in E$ we have*

$$P(G;t) = P(G \setminus e;t) - P(G/e;t).$$

Proof. We will prove this in the form $P(G \setminus e) = P(G) + P(G/e)$. Suppose $e = uv$. Since e is no longer present in $G \setminus e$, its proper colorings are of two types: those where $c(u) \neq c(v)$ and those where $c(u) = c(v)$. If $c(u) \neq c(v)$, then properly coloring $G \setminus e$ is the same as properly coloring G. So there are $P(G)$ colorings of the first type. There is also a bijection between the proper colorings of $G \setminus e$ where $c(u) = c(v)$ and those of G/e; namely color v_e with the common color of u and v and leave all the other colors the same. It follows that there are $P(G/e)$ colorings of the second type and the lemma is proved. □

We can now easily show that $P(G;t)$ lives up to its name.

Theorem 3.8.3. *We have that $P(G;t)$ is a polynomial in t for any graph G.*

Proof. We induct on $\#E$. If G has no edges, then clearly $P(G;t) = t^{\#V}$ which is a polynomial in t. If $\#E \geq 1$, then pick $e \in E$. By deletion-contraction $P(G;t) = P(G \setminus e;t) - P(G/e;t)$. And by induction we have that both $P(G \setminus e;t)$ and $P(G/e;t)$ are polynomials in t. Thus the same is true of their difference. □

We can also use Lemma 3.8.2 to compute the chromatic polynomial of C_4. Recall our notation of P_n and K_n for paths and complete graphs on n vertices, respectively. Now picking any edge $e \in E(C_4)$ we can use deletion-contraction and then determine the polynomials of the resulting graphs via coloring vertex by vertex to obtain

$$P(C_4) = P(P_4) - P(K_3) = t(t-1)^3 - t(t-1)(t-2) = t(t-1)(t^2 - 3t + 3).$$

Note that the quadratic factor has complex roots, thus substantiating our claim that $P(G)$ does not always have roots which are integers.

One can use induction and Lemma 3.8.2 to prove a host of results about $P(G;t)$. Since these demonstrations are all similar, we will leave them to the reader. We will use a nonstandard way of writing down the coefficients of this polynomial, which will turn out to be convenient later.

Theorem 3.8.4. *Let $G = (V, E)$ and write*

$$P(G;t) = a_0 t^n - a_1 t^{n-1} + a_2 t^{n-2} - \cdots + (-1)^n a_n.$$

(a) $n = \#V$.

(b) $\operatorname{mdeg} P(G;t) =$ *the number of components of G.*

(c) $a_i \geq 0$ *for all i and* $a_i > 0$ *for* $0 \leq i \leq n - \operatorname{mdeg} P(G;t)$.

(d) $a_0 = 1$ *and* $a_1 = \#E$. □

Now that we know $P(G;t)$ is a polynomial we can ask if there is any combinatorial interpretation for its coefficients, the reverse of our approach up to now, which has been to start with a sequence and then find its generating function. Put a total order on the edge set E, writing $e < f$ if e is less than f in this order and similarly for other notation. If C is a cycle in G, then the corresponding *broken circuit B* is the set of edges obtained from $E(C)$ by removing the smallest edge in the total order. Returning to the graph in Figure 3.2, let $b = uv$, $c = ux$, $d = vw$, $e = vx$ and impose the order $b < c < d < e$. The only cycle has edges b, c, e and the corresponding broken circuit is $B = \{c, e\}$ which are the edges of a path. Say that a set of edges $A \subseteq E$ *contains no broken circuit* or is an *NBC set* if $A \not\supseteq B$ for any broken circuit B. Let

$$\mathrm{NBC}_k = \mathrm{NBC}_k(G) = \{A \subseteq E \mid \#A = k \text{ and } A \text{ is an NBC set}\}$$

and $\mathrm{nbc}_k = \mathrm{nbc}_k(G) = \#\,\mathrm{NBC}_k(G)$. In our example graph

k	$\mathrm{NBC}_k(G)$	$\mathrm{nbc}_k(G)$
0	$\{\emptyset\}$	1
1	$\{\{b\}, \{c\}, \{d\}, \{e\}\}$	4
2	$\{\{b,c\}, \{b,d\}, \{b,e\}, \{c,d\}, \{d,e\}\}$	5
3	$\{\{b,c,d\}, \{b,d,e\}\}$	2
4	$\emptyset$	0

Comparison of the last column of this table with the coefficients of $P(G;t)$ presages our next result, which is due to Whitney [**99**]. It is surprising that the conclusion does not depend on the total order given to the edges. The proof we give is based on the demonstration of Blass and Sagan [**16**].

Theorem 3.8.5. *If $\#V = n$, then, given any ordering of E,*

$$P(G;t) = \sum_{k=0}^{n} (-1)^k \,\mathrm{nbc}_k(G)\, t^{n-k}.$$

Proof. Identify each $A \in \mathrm{NBC}_k(G)$ with the associated spanning subgraph. Then A is acyclic since any cycle contains a broken circuit. It follows from Theorem 1.10.2 that

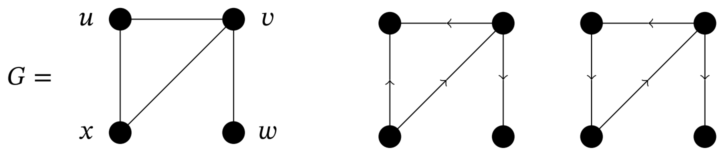


Figure 3.4. Two orientations of a graph

A is a forest with $n-k$ component trees. Hence $\mathrm{nbc}_k(G)\, t^{n-k}$ is the number of pairs (A, c) where $A \in \mathrm{NBC}_k(G)$ and $c : V \to [t]$ is a coloring constant on each component of A. We call such a coloring *A-improper.* Make the set of such pairs into a signed set by letting $\mathrm{sgn}(A, c) = (-1)^{\#A}$. So the theorem will be proved if we exhibit a sign-reversion involution ι on these pairs whose fixed points have positive sign and are in bijection with the proper coloring of G.

Define the fixed points of ι to be the (A, c) such that $A = \emptyset$ and c is proper. These pairs clearly have the desired characteristics. For any other pair, c is not a proper coloring so there must be an edge $e = uv$ with $c(u) = c(v)$. Let e be the smallest such edge in the total order. Now define $\iota(A, c) = (A \,\Delta\, \{e\}, c) \coloneqq (A', c)$. It is clear that ι reverses signs. And it is an involution because c does not change, and so the smallest monochromatic edge is the same in a pair and its image. We just need to check that ι is well-defined. If $A' = A - \{e\}$, then obviously A is still an NBC set and c is A'-improper. If $A' = A \cup \{e\}$, then, since e joined two vertices of the same color, c is still A'-improper. But assume, towards a contradiction, that A' is no longer NBC. Then $A' \supseteq B$ where B is a broken circuit, and $e \in B$ since A is NBC. Since c is A'-improper, all edges in B have vertices colored $c(u)$. But e is the smallest edge having that property, and so the smaller edge removed from a cycle to get B cannot exist. Thus A' is NBC, ι is well-defined, and we are done with the proof. $\square$

One of the amazing things about the chromatic polynomial is that it often appears in places where a priori it has no business being because no graph coloring is involved. We now give two illustrations of this. Recall from Section 2.6 that an *orientation O* of a graph G is a digraph with the same vertex set obtained by replacing each edge uv of G by one of the possible arcs $\overrightarrow{uv}$ or $\overrightarrow{vu}$. See Figure 3.4 for two orientations of our usual graph. Call O *acyclic* if it does not contain any directed cycles and let $\mathcal{A}(G)$ and $a(G)$ denote the set and number of acyclic orientations of G, respectively. The first of the orientations just given is acyclic while the second is not. The total number of orientations of the cycle u, v, x, u is 2^3 and the number of those which produce a cycle is 2 (clockwise and counterclockwise). Since neither orientation of vw can produce a cycle, we see that $a(G) = 2(2^3 - 2) = 12$. We now do something very strange and plug $t = -1$ into the chromatic polynomial (3.23) and get $P(G; -1) = (-1)(-2)(-2)(-3) = 12$. Although it is *not* at all clear what it means to color a graph with -1 colors, we have just seen an example of the following celebrated theorem of Stanley [**85**].

Theorem 3.8.6. *For any graph G with $\#V = n$ we have*

$$P(G; -1) = (-1)^n\, a(G).$$

Proof. We induct on $\#E$ where the base case is an easy check. It suffices to show that $(-1)^n P(G;-1)$ and $a(G)$ satisfy the same recursion. Using the Deletion-Contraction Lemma for the former, we see that we need to show $a(G) = a(G \setminus e) + a(G/e)$ for a fixed $e = uv \in E$. Consider the map

$$\phi\colon A(G) \to A(G \setminus e)$$

which sends $O \mapsto O'$ where O' is obtained from O by removing the arc corresponding to e. Clearly O' is still acyclic so the function is well-defined.

We claim ϕ is onto. Suppose to the contrary that there is some $O' \in A(G \setminus e)$ such that adding back $\overrightarrow{uv}$ creates a directed cycle C, and similarly with $\overrightarrow{vu}$ creating a cycle C'. Then $(C - \overrightarrow{uv}) \cup (C' - \overrightarrow{vu})$ is a closed, directed walk which must contain a directed cycle by Exercise 14(b) of Chapter 1. But this third directed cycle is contained in O', which is the desired contradiction.

If $O' \in A(G \setminus e)$, then by definition of the map $\#\phi^{-1}(O') \le 2$. And from the previous paragraph $\#\phi^{-1}(O') \ge 1$. So $a(G) = x + 2y$ where $x = \#\{O' \mid \phi^{-1}(O') = 1\}$ and $y = \#\{O' \mid \phi^{-1}(O') = 2\}$. Since $a(G \setminus e) = x + y$ it suffices to show that $a(G/e) = y$. We will do this by constructing a bijection

$$\psi\colon \{O' \in A(G \setminus e) \mid \phi^{-1}(O') = 2\} \to A(G/e).$$

Let Y be the domain of ψ. If there are a pair of edges $wu, wv \in E(G)$, then any $O' \in Y$ contains either both $\overrightarrow{wu}$ and $\overrightarrow{wv}$ or both $\overrightarrow{uw}$ and $\overrightarrow{vw}$. This is because in all other cases adding back one of the orientations of e would create an orientation of G with a cycle, contradicting the fact that $\phi^{-1}(O') = 2$. So define $O'' = \psi(O')$ to be the orientation of G/e which agrees with O' on all arcs not containing the new vertex v_e, and on any edge of the form wv_e uses the same orientation as either wu or wv. (As just shown, these two orientations are either both towards or both away from e). Proving that ψ is a well-defined bijection is left as an exercise. $\square$

We should mention that Stanley actually proved a stronger result giving a combinatorial interpretation to $P(G;-t)$ for all negative integers $-t$. See Exercise 28(b) for details. So, as we saw with the binomial coefficients in (1.6), we have another instance of combinatorial reciprocity. We will study this phenomenon more generally in the next section.

For our second example of the protean nature of the chromatic polynomial, we will need to assume that our graphs G have vertex set $[n]$ so that one has a total order on the vertices. Let F be a spanning forest of G and root each component tree T of F at its smallest vertex r. Say that F is *increasing* if the integers on any path starting at r form an increasing sequence for all roots r. In Figure 3.5 the reader will find the usual

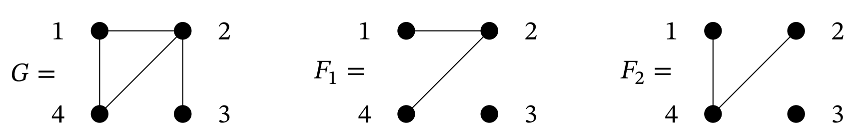


Figure 3.5. A graph and two spanning forests

graph, now labeled with [4], and two spanning forests. We have that F_1 is increasing as any singleton node is increasing, and in the nontrivial tree the only maximal path from the root is $1, 2, 4$, which is an increasing sequence. On the other hand F_2 is not increasing because of the path $1, 4, 2$.

For a graph $G = (V, E)$ we define

$$\mathrm{ISF}_m(G) = \{F \mid F \text{ is an increasing spanning forest of } G \text{ with } m \text{ edges}\}$$

and $\mathrm{isf}_m(G) = \#\,\mathrm{ISF}_m(G)$. If $\#V = n$, then consider the corresponding generating polynomial

$$\mathrm{isf}(G) = \mathrm{isf}(G;t) = \sum_{m=0}^{n} (-1)^m \,\mathrm{isf}_m(G) t^{n-m}.$$

Let us compute this for our example graph. Any tree with zero or one edge is increasing so that $\mathrm{isf}_0(G) = 1$ and $\mathrm{isf}_1(G) = \#E = 4$. Any of the pairs of edges of G form an increasing forest except for the pair giving F_2 in Figure 3.5. So $\mathrm{isf}_2(G) = \binom{4}{2} - 1 = 5$. Similarly one checks that $\mathrm{isf}_3(G) = 2$. And $\mathrm{isf}_4(G) = 0$ since G itself is not a forest. So

$$\mathrm{isf}(G;t) = t^4 - 4t^3 + 5t^2 - 2t = t(t-1)^2(t-2) = P(G;t).$$

We cannot always have $\mathrm{isf}(G) = P(G)$ because the former depends on the labeling of the vertices (even though our notation conceals that fact) while the latter does not. So we will put aside deciding when they are equal for now and concentrate on the factorization over $\mathbb{Z}$ which we have just seen and which, as we will see, is not a coincidence. In fact, the roots will be the cardinalities of the edge sets defined by

$$E_k = \{ik \in E \mid i < k\} \tag{3.24}$$

for $1 \le k \le n$. In our example $E_1 = \emptyset$ (since there is no vertex smaller than 1), $E_2 = \{12\}$, $E_3 = \{23\}$, and $E_4 = \{14, 24\}$.

Lemma 3.8.7. *If G has $V = [n]$, then a spanning subgraph F is an increasing forest if and only if it is obtained by picking at most one edge from each E_k for $k \in [n]$.*

Proof. For the forward direction assume, towards a contradiction, that F contains both ik and jk with $i, j < k$. So if r is the root of the tree containing i, j, k, then, by the increasing condition, i must be the vertex preceding k on the unique r–k path. But the same must be true of j, which is a contradiction.

For the reverse implication, we must first verify that F is acyclic. But if F contains a cycle C, then let k be its maximum vertex. It follows that there are $ik, jk \in E(C)$ and, by the maximum requirement, $i, j < k$. This contradicts the assumption in this direction. Similarly one can show that if F is not increasing, then one can produce two edges from the same E_k and so we are done. □

It is now a simple matter to prove the following result of Hallam and Sagan [**41**]. The proof given here was obtained by Hallam, Martin, and Sagan [**40**].

Theorem 3.8.8. *If G has $V = [n]$, then*

$$\mathrm{isf}(G;t) = \prod_{k=1}^{n} (t - |E_k|).$$

Proof. The coefficient of t^{n-m} in the product is, up to sign, the sum of all terms of the form $|E_{i_1}||E_{i_2}|\cdots|E_{i_m}|$ where the i_j are distinct indices. But this product is the number of ways to pick one edge out of each of the sets $E_{i_1}, E_{i_2}, \ldots, E_{i_m}$. So, by the previous lemma, the sum is the number of increasing forests with m edges, finishing the proof. □

Returning to the question of when the chromatic and increasing spanning forest polynomials are equal, we need the following definition. Graph G has a *perfect elimination order* (peo) if there is a total ordering of V as $v_1, v_2, \ldots, v_n$ such that, for all k, the set of vertices coming before v_k in this order and adjacent to v_k form the vertices of a clique (complete subgraph of G). This definition may seem strange at first glance, but it has been useful in various graph-theoretic contexts. Returning to our running example graph we see that the order $1, 2, 3, 4$ is a peo since 1 is adjacent to no earlier vertex, 2 and 3 are both adjacent to a single previous vertex which is a K_1, and 4 is adjacent to 1 and 2 which form an edge also known as a K_2. We can now prove another result from [**40**].

Lemma 3.8.9. *Let G have $V = [n]$. Write the edges of G as ij with $i < j$ and order them lexicographically. For all $m \geq 0$ we have*

$$\mathrm{ISF}_m(G) \subseteq \mathrm{NBC}_m(G).$$

Furthermore, we have equality for all m if and only if the natural order on $[n]$ is a peo.

Proof. To prove the inclusion we suppose that F is an increasing spanning forest which contains a broken circuit B and derive a contradiction. By the lexicographic ordering of the edges, B must be a path of the form $v_1, v_2, \ldots, v_l$ where $v_1 = \min\{v_1, \ldots, v_l\}$ and $v_2 > v_l$. So there must be a smallest index $i \geq 2$ such that $v_i > v_{i+1}$. It follows that $v_{i-1}, v_{i+1} < v_i$ so that the two corresponding edges of B contradict Lemma 3.8.7.

For the forward direction of the second statement we must show that if $i, j < k$ and $ik, jk \in E(G)$, then $ij \in E(G)$. By Lemma 3.8.7 again, $\{ik, jk\}$ is not the edge set of an increasing spanning forest. So, by the assumed equality, this set must contain a broken circuit. Since there are only two edges, this set must actually be a broken circuit, and $ij \in E(G)$ must be the edge used to complete the cycle. The reverse implication is left as an exercise. □

From this result we immediately conclude the following.

Theorem 3.8.10. *Let G have $V = [n]$. Then* $\mathrm{isf}(G; t) = P(G; t)$ *if and only if the natural order on $[n]$ is a peo.* □

3.9. Combinatorial reciprocity

When plugging a negative parameter into a counting function results in a sign times another enumerative function, then this is called *combinatorial reciprocity*. This concept was introduced and studied by Stanley [**86**]. We have already seen two examples of this in equation (1.6) and Theorem 3.8.6 (and, more generally, Exercise 28 of this chapter). Here we will make a connection with recurrences and rational generating functions. See the text of Beck and Sanyal [**5**] for a whole book devoted to this subject.

Before stating a general theorem, we return to the example with which we began Section 3.6. This was the sequence defined by $a_0 = 2$ and $a_n = 3a_{n-1}$ for $n \geq 1$. One can extend the domain of this recursion to all integral n, in which case one gets $2 = a_0 = 3a_{-1}$ so that $a_{-1} = 2/3$. Then $2/3 = a_{-1} = 3a_{-2}$ yielding $a_{-2} = 2/3^2$, and so forth. An easy induction shows that for $n \leq 0$ we have $a_n = 2 \cdot 3^n$ just as for $n \geq 0$. We can also compute the generating function for the negatively indexed part of the sequence, where it is convenient to start with a_{-1}, which is the geometric series

$$\sum_{n\geq 1} a_{-n}x^n = \sum_{n\geq 1} \frac{2x^n}{3^n} = \frac{2x/3}{1-x/3} = \frac{2x}{3-x}.$$

Comparing this to $f(x) = \sum_{n\geq 0} a_n x^n$ as found in (3.12) we see that

$$-f(1/x) = \frac{-2}{1-3/x} = \frac{2x}{3-x} = \sum_{n\geq 1} a_{-n}x^n.$$

The reader should have some qualms about writing $f(1/x)$ since we have gone to great pains to point out that x has no inverse in $\mathbb{C}[[x]]$. Indeed, if we use the definition that $f(x) = \sum_{n\geq 0} a_n x^n$, then $f(1/x) = \sum_{n\geq 0} a_n/x^n$, which is not a formal power series! But if $f(x)$ can be expressed as a rational function $f(x) = p(x)/q(x)$ where $\deg p(x) \leq \deg q(x) \coloneqq d$, then we can make sense of this substitution as follows. Since $q(x)$ has degree d we have that $x^d q(1/x)$ is also a polynomial and is invertible since its constant coefficient is nonzero (Theorem 3.3.1). Furthermore, $x^d p(1/x)$ is also a polynomial since $d \geq \deg p(x)$. So we can *define*

$$f(1/x) = \frac{x^d p(1/x)}{x^d q(1/x)} \tag{3.25}$$

and stay inside the formal power series ring. With this convention, the following result makes sense.

Theorem 3.9.1. *Suppose that a_n is a sequence defined for all $n \in \mathbb{Z}$ and satisfying the linear recurrence relation with constant coefficients* (3.18) *for all such n. Letting $f(x) = \sum_{n\geq 0} a_n x^n$, we have*

$$\sum_{n\geq 1} a_{-n}x^n = -f(1/x).$$

Proof. By (3.25), to prove the theorem it suffices to show that

$$x^d q(1/x) \sum_{n\geq 1} a_{-n}x^n = -x^d p(1/x).$$

Note that by (3.20) we have

$$x^d q(1/x) = x^d + c_1 x^{d-1} + c_2 x^{d-2} + \cdots + c_d.$$

So if $m \geq 1$, then, using (3.18) and the fact that $x^d p(1/x)$ has degree at most d,

$$\begin{aligned}[x^{m+d}]x^d q(1/x) \sum_{n\geq 1} a_{-n}x^n &= a_{-m} + c_1 a_{-m-1} + c_2 a_{-m-2} + \cdots + c_d a_{-m-d} \\ &= 0 \\ &= [x^{m+d}](-x^d p(1/x)).\end{aligned}$$

Similarly we can prove that this equality of coefficients continues to hold for $-d \leq m \leq 0$. This completes the proof. □

To illustrate this theorem, we consider the negative binomial expansion. So if one fixes $n \geq 1$, then using Theorem 3.4.2 and equation (3.16)

$$f(x) := \frac{1}{(1-x)^n} = \sum_{k\geq 0} \left(\!\!\binom{n}{k}\!\!\right) x^k = \sum_{k\geq 0} \binom{n+k-1}{n-1} x^k.$$

Note that since n is fixed we are thinking of $\binom{n+k-1}{n-1}$ as a function of k. Substituting $-k$ for k in the binomial coefficient, we wish to consider the corresponding generating function

$$g(x) = \sum_{k\geq 1} \binom{n-k-1}{n-1} x^k.$$

We note that $\binom{n-k-1}{n-1} = 0$ for $1 \leq k < n$ since then $0 \leq n-k-1 < n-1$. So x^n can be factored out from $g(x)$ and, using (1.6) and the above expression for the negative binomial expansion,

$$\begin{aligned}g(x) &= x^n \sum_{k\geq n} \binom{n-k-1}{n-1} x^{k-n} \\ &= x^n \sum_{j\geq 0} \binom{-j-1}{n-1} x^j \\ &= (-1)^{n-1} x^n \sum_{j\geq 0} \left(\!\!\binom{j+1}{n-1}\!\!\right) x^j \\ &= (-1)^{n-1} x^n \sum_{j\geq 0} \binom{n+j-1}{n-1} x^j \\ &= \frac{(-1)^{n-1}x^n}{(1-x)^n}.\end{aligned}$$

On the other hand, we could apply Theorem 3.9.1 and write

$$g(x) = \frac{-1}{(1-1/x)^n} = \frac{-x^n}{(x-1)^n} = \frac{(-1)^{n-1}x^n}{(1-x)^n}$$

giving the same result but with less computation.

Exercises

(1) Prove that for $n \in \mathbb{N}$
$$\sum_{k=0}^{n} 2^k \binom{n}{k} = 3^n$$
in two ways.
 (a) Use the Binomial Theorem.
 (b) Use a combinatorial argument.
 (c) Generalize this identity by replacing 2^k by c^k for any $c \in \mathbb{N}$, giving both a proof using the Binomial Theorem and one which is combinatorial.

(2) For $m, n, k \in \mathbb{N}$ show that
$$\binom{m+n}{k} = \sum_{l \geq 0} \binom{m}{l}\binom{n}{k-l}$$
in three ways: by induction, using the Binomial Theorem, and using a combinatorial argument.

(3) Let $x_1, \ldots, x_m$ be variables. Prove the multinomial coefficient identity
$$\sum_{n_1+\cdots+n_m=n} \binom{n}{n_1, \ldots, n_m} x_1^{n_1} \cdots x_m^{n_m} = (x_1 + \cdots + x_m)^n,$$
in three ways:
 (a) by inducting on n,
 (b) by using the Binomial Theorem and inducting on m,
 (c) by a combinatorial argument.

(4) (a) Prove Theorem 3.1.2.
 (b) Use this generating function to rederive Corollary 1.5.3.

(5) (a) Recall that an inversion of $\pi \in P([n])$ is a pair (i, j) with $i < j$ and $\pi_i > \pi_j$ and we call π_i the *maximum* of the inversion. The *inversion table* of π is $I(\pi) = (a_1, a_2, \ldots, a_n)$ where a_k is the number of inversions with maximum k. Show that $0 \leq a_k < k$ for all k and that
$$\operatorname{inv} \pi = \sum_{k=1}^{n} a_k.$$
 (b) Let
$$\mathcal{I}_n = \{(a_1, a_2, \ldots, a_n) \mid 0 \leq a_k < k \text{ for all } k\}.$$
 Show that the map $\pi \mapsto I(\pi)$ is a bijection $P([n]) \to \mathcal{I}_n$.
 (c) Use part (b) and weight-generating functions to rederive Theorem 3.2.1.

(6) Let $\operatorname{st} : \mathfrak{S}_n \to \mathbb{N}$ be a statistic on permutations. Recall that for a permutation π we let $\operatorname{Av}_n(\pi)$ denote the set of permutations in $\mathfrak{S}_n$ avoiding π. Say that permutations π, σ are st-*Wilf equivalent*, written $\pi \overset{\text{st}}{\equiv} \sigma$, if for all $n \geq 0$ we have equality of the

generating functions

$$\sum_{\tau \in \mathrm{Av}_n(\pi)} q^{\mathrm{st}\,\tau} = \sum_{\tau \in \mathrm{Av}_n(\sigma)} q^{\mathrm{st}\,\tau}.$$

(a) Show that if π and σ are st-Wilf equivalent, then they are Wilf equivalent.

(b) Show that

$$132 \stackrel{\mathrm{inv}}{\equiv} 213$$

and

$$231 \stackrel{\mathrm{inv}}{\equiv} 312$$

and that there are no other inv-Wilf equivalences between two permutations in $\mathfrak{S}_3$.

(c) Show that

$$132 \stackrel{\mathrm{maj}}{\equiv} 231$$

and

$$213 \stackrel{\mathrm{maj}}{\equiv} 312$$

and that there are no other maj-Wilf equivalences between two permutations in $\mathfrak{S}_3$.

(7) (a) Prove that

$$\left[\begin{matrix} n \\ k \end{matrix}\right] = \left[\begin{matrix} n \\ n-k \end{matrix}\right]$$

in three ways: using the q-factorial definition, using the interpretation in terms of integer partitions, and using the interpretation in terms of subspaces.

(b) Prove the second recursion in Theorem 3.2.3 in two ways: by mimicking the proof of the first recursion and by using the first recursion in combination with part (a).

(c) If $S \subseteq [n]$, then let ΣS be the sum of the elements of S. Give two proofs of the following q-analogue of the fact that $\#\binom{[n]}{k} = \binom{n}{k}$:

$$\sum_{S \in \binom{[n]}{k}} q^{\Sigma S} = q^{\binom{k+1}{2}} \left[\begin{matrix} n \\ k \end{matrix}\right]_q,$$

one proof by induction and the other using Theorem 3.2.5.

(d) Give two other proofs of Theorem 3.2.6: one by inducting on n and one by using Theorem 3.2.5.

(8) (a) Reprove Theorem 3.2.4 in two way: using integer partitions and using subspaces.

(b) The Negative q-Binomial Theorem states that

$$\frac{1}{(1-t)(1-qt)(1-q^2t)\dots(1-q^{n-1}t)} = \sum_{k \geq 0} \left[\begin{matrix} n+k-1 \\ k \end{matrix}\right] t^k.$$

Give three proofs of this result: inductive, using integer partitions, and using subspaces.

(9) (a) Given $n_1+n_2+\cdots+n_m = n$, define the corresponding *q-multinomial coefficient* to be

$$\left[\begin{array}{c} n \\ n_1, n_2, \ldots, n_m \end{array}\right]_q = \frac{[n]_q!}{[n_1]_q!\,[n_2]_q!\,\ldots\,[n_m]_q!}$$

if all $n_i \geq 0$ or zero otherwise. Prove that

$$\left[\begin{array}{c} n \\ n_1, n_2, \ldots, n_m \end{array}\right]_q = \sum_{i=1}^{m} q^{n_1+n_2+\cdots+n_{i-1}} \left[\begin{array}{c} n-1 \\ n_1, \ldots, n_{i-1}, n_i-1, n_{i+1}, \ldots, n_m \end{array}\right]_q.$$

(b) Define inversions, descents, and the major index for permutations (linear orderings) of the multiset $M = \{\{1^{n_1}, 2^{n_2}, \ldots, m^{n_m}\}\}$ exactly the way as was done for permutations without repetition. Let $P(M)$ be the set of permutations of M. Prove that

$$\sum_{\pi \in P(M)} q^{\operatorname{inv} \pi} = \sum_{\pi \in P(M)} q^{\operatorname{maj} \pi} = \left[\begin{array}{c} n \\ n_1, n_2, \ldots, n_m \end{array}\right]_q.$$

(c) Let V be a vector space over $\mathbb{F}_q$ of dimension n. Assume $S = \{s_1 < \cdots < s_m\} \subseteq \{0, 1, \ldots, n\}$. Then a *flag of type S* is a chain of subspaces

$$F : W_1 < W_2 < \cdots < W_m \leq V$$

such that $\dim W_i = s_i$ for all i. The reason for this terminology is that when $n = 2$ and $S = \{0, 1, 2\}$, then F consists of a point (the origin) contained in a line contained in a plane which could be viewed as a drawing of a physical flag with the point being the hole in the ground, the line being the flag pole, and the plane being the cloth flag itself. Give two proofs that

$$\#\{F \text{ of type } S\} = \left[\begin{array}{c} n \\ s_1,\ s_2-s_1,\ s_3-s_2,\ \ldots,\ s_m-s_{m-1},\ n-s_m \end{array}\right]_q,$$

one by mimicking the proof of Theorem 3.2.6 and one by induction on m.

(10) (a) Prove that in $\mathbb{C}[[x]]$ we have $e^{kx} = (e^x)^k$ for any $k \in \mathbb{N}$.

(b) Define formal power series for the trigonometric functions using their usual Taylor expansions. Prove that in $\mathbb{C}[[x]]$ we have $\sin^2 x + \cos^2 x = 1$ and $\sin 2x = 2 \sin x \cos x$.

(c) If one is given a sequence $a_0, a_1, a_2, \ldots$ and defines

$$f_k(x) = a_k x^k,$$

then show that $\sum_{k\geq 0} f_k(x) = f(x)$ where $f(x) = \sum_{k\geq 0} a_k x^k$.

(d) Prove the backwards direction of Theorem 3.3.2.

(e) Use Theorems 3.3.1 and 3.3.3 to reprove that $1/x$ and e^{1+x} are not well-defined in $\mathbb{C}[[x]]$.

(f) Prove Theorem 3.3.4

(11) Prove that if S, T are summable sets, then so is $S \times T$.

(12) Say that $f(x) \in \mathbb{C}[[x]]$ has a *square root* if there is $g(x) \in \mathbb{C}[[x]]$ such that $f(x) = g(x)^2$.

(a) Prove that $f(x)$ has a square root if and only if $\operatorname{mdeg} f(x)$ is even.

(b) Show that as formal power series

$$(1+x)^{1/2} = \sum_{k\geq 0} \binom{1/2}{k} x^k.$$

(c) Show that as formal power series

$$(e^x)^{1/2} = \sum_{k\geq 0} \frac{x^k}{2^k\, k!}.$$

(d) Generalize the previous parts of this exercise to mth roots for $m \in \mathbb{P}$.

(13) Give a second proof of Theorem 3.4.2 by using induction.

(14) Prove Theorem 3.4.3.

(15) Prove Theorem 3.5.1.

(16) (a) Finish the proof of Theorem 3.5.3(a).

(b) Give a second proof of Theorem 3.5.3(b) by using Theorem 3.2.5.

(c) Show that the generating function for the number of partitions of n with largest part k equals the generating function for the number of partitions of n with exactly k parts and that both are equal to the product

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

(d) The *Durfee square* of a Young diagram λ is the largest square partition (d^d) such that $(d^d) \subseteq \lambda$. Use this concept to prove that

$$\sum_{n\geq 0} p(n)x^n = \sum_{d\geq 0} \frac{x^{d^2}}{(1-x)^2(1-x^2)^2\cdots(1-x^d)^2}.$$

(17) Let a_n be the number of integer partitions of n such that any part i is repeated fewer than i times and let b_n be the number of integer partitions of n such that no part is a square. Use generating functions to show that $a_n = b_n$ for all n.

(18) Given $m \geq 2$, use generating functions to show that the number of partitions of n where each part is repeated fewer than m times equals the number of partitions of n into parts not divisible by m. Note that bijective proofs of this result were given in Exercise 15 of Chapter 2.

(19) (a) Show that F_n is the closest integer to

$$\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$$

for $n \geq 1$.

(b) Prove that

$$F_n^2 = \begin{cases} F_{n-1}F_{n+1} + 1 & \text{if } n \text{ is odd,} \\ F_{n-1}F_{n+1} - 1 & \text{if } n \text{ is even} \end{cases}$$

in two ways: using equation (3.14) and by a combinatorial argument.

(20) (a) Use the algorithm in Section 3.6 to rederive Theorem 3.1.1.
(b) Complete the proof of Theorem 3.6.1.
(c) Give a second proof of Theorem 3.6.1 using Theorem 3.1.2
(d) Prove Theorem 3.6.2.
(e) Let s be the infinite matrix with rows and columns indexed by $\mathbb{N}$ and with $s(n,k)$ being the entry in row n and column k. Similarly define S with entries $S(n,k)$. Show that $Ss = sS = I$ where I is the $\mathbb{N} \times \mathbb{N}$ identity matrix. Hint: Use Theorems 3.6.1 and 3.6.2.

(21) Given $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, define the corresponding *q-Stirling number of the second kind* by $S[0,k] = \delta_{0,k}$ and, for $n \geq 1$,

$$S[n,k] = S[n-1,k-1] + [k]_q S[n-1,k].$$

(a) Show that

$$\sum_{n \geq 0} S[n,k]x^n = \frac{x^k}{(1-[1]_q x)(1-[2]_q x)\cdots(1-[k]_q x)}.$$

(b) All set partitions $\rho = B_1/B_2/\ldots/B_k$ in the rest of this problem will be written in *standard form*, which means that

$$1 = \min B_1 < \min B_2 < \cdots < \min B_k.$$

An *inversion* of ρ is a pair (b, B_j) where $b \in B_i$ for some $i < j$ and $b > \min B_j$. We let $\operatorname{inv}\rho$ be the number of inversions of ρ. For example, $\rho = B_1/B_2/B_3 = 136/25/4$ has inversions $(3,B_2)$, $(6,B_2)$, $(6,B_3)$, and $(5,B_3)$ so that $\operatorname{inv}\rho = 4$. Show that

$$S[n,k] = \sum_{\rho \in S([n],k)} q^{\operatorname{inv}\rho}.$$

(c) A *descent* of a set partition ρ is a pair (b, B_{i+1}) where $b \in B_i$ and $b > \min B_j$. We let $\operatorname{des}\rho$ be the number of descents of ρ. In the previous example, ρ has descents $(3,B_2)$, $(6,B_2)$, and $(5,B_3)$ so that $\operatorname{des}\rho = 3$. The *descent multiset* of ρ is denoted $\operatorname{Des}\rho$ and is the multiset

$$\{\{1^{d_1}, 2^{d_2}, \ldots, k^{d_k} \mid \text{for all } i,\ d_i = \text{number of descents } (b, B_{i+1})\}\}.$$

The *major index* of ρ is

$$\operatorname{maj}\rho = \sum_{i \in \operatorname{Des}\rho} i = d_1 + 2d_2 + \cdots + kd_k.$$

In our running example $\operatorname{Des}\rho = \{\{1^2, 2^1\}\}$ so that $\operatorname{maj}\rho = 1+1+2 = 4$. Show that

$$S[n,k] = \sum_{\rho \in S([n],k)} q^{\operatorname{maj}\rho}.$$

(22) Given $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, define the corresponding *signless q-Stirling number of the first kind* by $c[0,k] = \delta_{0,k}$ and, for $n \geq 1$,

$$c[n,k] = c[n-1,k-1] + [n-1]_q c[n-1,k]$$

(a) Show that

$$\sum_{k \geq 0} c[n,k]x^k = x(x+[1]_q)(x+[2]_q)\cdots(x+[n-1]_q).$$

(b) The *standard form* of $\pi \in \mathfrak{S}_n$ is $\pi = \kappa_1\kappa_2\cdots\kappa_k$ where the κ_i are the cycles of π,

$$\min \kappa_1 < \min \kappa_2 < \cdots < \min \kappa_k,$$

and each κ_i is written beginning with $\min \kappa_i$. Define the *cycle major index* of π to be $\operatorname{maj}_c \pi = \operatorname{maj} \pi'$ where π' is the permutation in one-line form obtained by removing the cycle parentheses in the standard form of π. If, for example, $\pi = (1,7,2)(3,6,8)(4,5)$, then $\pi' = 17236845$ so that $\operatorname{maj}_c \pi = 2+6 = 8$. Show that

$$c[n,k] = \sum_{\pi \in c([n],k)} q^{\operatorname{maj}_c \pi}.$$

(23) Reprove the formula

$$C(n) = \frac{1}{n+1}\binom{2n}{n}$$

by using Theorem 3.6.3.

(24) (a) Show that if $k, l \in \mathbb{N}$ are constants, then $\binom{n+l}{k}$ is a polynomial in n of degree k.

(b) Show that the polynomials

$$\binom{n}{0}, \binom{n+1}{1}, \binom{n+2}{2}, \ldots$$

form a basis for the algebra of polynomials in n.

(c) Use part (a) to complete the proof of Theorem 3.7.1.

(25) Redo the solution for the first recursion in Section 3.6 as well as the one for F_n using the method of undetermined coefficients.

(26) Prove for the n-cycle that

$$P(C_n; t) = (t-1)^n + (-1)^n(t-1)$$

in two ways: using deletion-contraction and using NBC sets.

(27) Prove Theorem 3.8.4 using induction and give a second proof of parts (b)–(d) using NBC sets.

(28) (a) Complete the proof of Theorem 3.8.6.

(b) Let $G = (V, E)$ be a graph and $t \in \mathbb{P}$. Call an acyclic orientation O and a (not necessarily proper) coloring $c : V \to [t]$ *compatible* if, for all arcs $\vec{uv}$ of O, we have $c(u) \le c(v)$. Show that if $\#V = n$, then

$$P(G; -t) = (-1)^n(\text{number of compatible pairs } (O, c)).$$

(c) Show that Theorem 3.8.6 is a special case of part (b).

(29) Finish the proofs of Lemma 3.8.7 and Lemma 3.8.9.

(30) (a) Call a permutation σ which avoids $\Pi = \{231, 312, 321\}$ *tight*. Show that σ is tight if and only if σ is an involution having only 2-cycles of the form $(i, i+1)$ for some i.

(b) Let G be a graph with $V = [n]$. Call a spanning forest F of G *tight* if the sequence of labels on any path starting at a root of F avoids Π as in part (a). Let

$$\operatorname{TSF}_m(G) = \{F \mid F \text{ is a tight spanning forest of } G \text{ with } m \text{ edges}\}.$$

Show that if G has no 3-cycles, then for all $m \geq 0$

$$\mathrm{TSF}_m(G) \subseteq \mathrm{NBC}_m(G).$$

(c) A *candidate path* in a graph G with no 3-cycles is a path of the form

$$a, c, b, v_1, v_2, \ldots, v_m = d$$

such that $a < b < c$, $m \geq 1$, and v_m is the only v_i smaller than c. A total order on $V(G)$ is called a *quasiperfect ordering* (qpo) if every candidate path satisfies the following condition: either $ad \in E(G)$, or $d < b$ and $cd \in E(G)$. Consider the generating function

$$\mathrm{tsf}(G;t) = \sum_{m \geq 0} (-1)^m \, \mathrm{tsf}_m(G) t^{n-m}.$$

Show that $\mathrm{tsf}(G;t) = P(G;t)$ if and only if the natural order on $[n]$ is a qpo.

(31) Fill in the details of the case $-d \leq m \leq 0$ in the proof of Theorem 3.9.1.

(32) (a) Extend the Fibonacci numbers F_n to all $n \in \mathbb{Z}$ by insisting that their recursion continue to hold for $n < 0$. Show that if $n \geq 0$, then

$$F_{-n} = (-1)^{n-1} F_n.$$

(b) Find $\sum_{n \geq 1} F_{-n} x^n$ in two ways: by using part (a) and by using Theorem 3.9.1.

Counting with Exponential Generating Functions

Given a sequence $a_0, a_1, a_n, \dots$ of complex numbers, one can associate with it an exponential generating function where a_n is the coefficient of $x^n/n!$. In certain cases it turns out that the exponential generating function is easier to deal with than the ordinary one. This is particularly true if the a_n count combinatorial objects obtained from some set of labels. We give a method for dealing with such structures which again give rise to Sum and Product Rules as well as an Exponential Formula unique to this setting.

4.1. First examples

Given $a_0, a_1, a_n, \dots$ where $a_n \in \mathbb{C}$ for all n, the corresponding *exponential generating function* (egf) is

$$F(x) = a_0 + a_1\frac{x}{1!} + a_2\frac{x^2}{2!} + \cdots = \sum_{n\geq 0} a_n\frac{x^n}{n!}.$$

In order to distinguish these from the ordinary generating functions (ogfs) in the previous chapter, we will often use capital letters for egfs and lowercase ones for ogfs. The use of the adjective "exponential" is because in the simple case when $a_n = 1$ for all n, the corresponding egf is $F(x) = \sum_{n\geq 0} x^n/n! = e^x$.

To illustrate why egfs may be useful in studying a sequence, consider $a_n = n!$ for $n \geq 0$. The ogf is $f(x) = \sum_{n\geq 0} n!\, x^n$ and this power series cannot be simplified. On the other hand, the egf is

$$F(x) = \sum_{n\geq 0} n!\, \frac{x^n}{n!} = \frac{1}{1-x}$$

which can now be manipulated if necessary.

To get some practice in using egfs, we will now compute some examples. One technique which occurs often in determining generating functions is the interchange of summations. As an example, consider the derangement numbers $D(n)$ and the formula which was given for them in Theorem 2.1.2. So

$$\begin{aligned}\sum_{n\geq 0} D(n)\frac{x^n}{n!} &= \sum_{n\geq 0} n!\left(\sum_{k=0}^{n}(-1)^k\frac{1}{k!}\right)\frac{x^n}{n!}\\ &= \sum_{k\geq 0}\frac{(-1)^k}{k!}\sum_{n\geq k} x^n\\ &= \sum_{k\geq 0}\frac{(-1)^k}{k!}\frac{x^k}{1-x}\\ &= \frac{1}{1-x}\sum_{k\geq 0}\frac{(-x)^k}{k!}\\ &= \frac{e^{-x}}{1-x}.\end{aligned}$$

We will be able to give a much more combinatorial derivation of this formula once we have introduced the theory of labeled structures in Section 4.3. For now, we just record the result for future reference.

Theorem 4.1.1. *We have*

$$\sum_{n\geq 0} D(n)\frac{x^n}{n!} = \frac{e^{-x}}{1-x}. \qquad \square$$

If the given sequence is defined by a recurrence relation, then one can use a slight modification of the algorithm in Section 3.6 to compute its egf. One just multiplies by $x^n/n!$, rather than x^n, and sums. Note that the largest index in the recursion may not be the best choice to use as the power on x because of the following considerations. Given $f(x) = \sum_{n\geq 0} a_n x^n$, then its *formal derivative* is the formal power series

$$f'(x) = \sum_{n\geq 0} n a_n x^{n-1}.$$

This derivative enjoys most of the usual properties of the ordinary analytic derivative such as linearity and the product rule. One similarly defines formal integrals. Note that if one starts with an egf $F(x) = \sum_{n\geq 0} a_n x^n/n!$, then

$$F'(x) = \sum_{n\geq 0} n a_n \frac{x^{n-1}}{n!} = \sum_{n\geq 1} a_n \frac{x^{n-1}}{(n-1)!} = \sum_{n\geq 0} a_{n+1}\frac{x^n}{n!} \tag{4.1}$$

which is just the egf for the same sequence shifted up by one. So it can simplify things if the subscript on an element of the sequence is greater than the exponent of the corresponding power of x.

Let us consider the Bell numbers and their recurrence relation given in Theorem 1.4.1. Let $B(x) = \sum_{n\geq 0} B_n x^n/n!$. It will be convenient to replace n by $n+1$ and k by $k+1$ in the recursion before multiplying by $x^n/n!$ and summing. So using (4.1) and the summation interchange trick we obtain

$$\begin{aligned}
B'(x) &= \sum_{n\geq 0} B(n+1)\frac{x^n}{n!} \\
&= \sum_{n\geq 0}\left(\sum_{k=0}^{n}\binom{n}{k}B(n-k)\right)\frac{x^n}{n!} \\
&= \sum_{n\geq 0}\sum_{k=0}^{n}\frac{1}{k!\,(n-k)!}B(n-k)x^n \\
&= \sum_{k\geq 0}\frac{x^k}{k!}\sum_{n\geq k}B(n-k)\frac{x^{n-k}}{(n-k)!} \\
&= \sum_{k\geq 0}\frac{x^k}{k!}B(x) \\
&= e^x B(x).
\end{aligned}$$

We now have a differential equation to solve, but we must take some care to make sure this can be done formally. To this end we define the *formal natural logarithm* by

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots = \sum_{n\geq 1}\frac{(-1)^{n-1}x^n}{n} \tag{4.2}$$

which can be thought of as formally integrating the geometric series for $1/(1+x)$. Note that since $\ln(1+x)$ has an infinite number of terms, to have $\ln(1+f(x))$ well-defined it must be that $f(x)$ has constant term 0 by Theorem 3.3.3. In other words, for an infinite series $g(x)$, we have $\ln g(x)$ is only defined if the constant term of $g(x)$ is 1. Luckily this is true of $B(x)$ so we can separate variables above to get $B'(x)/B(x) = e^x$ and then formally integrate to get $\ln B(x) = e^x + c$ for some constant c. By definition (4.2) a natural log has no constant term so we must take $c = -1$. Solving for $B(x)$ we obtain the following result. Again, a more combinatorial proof will be given later.

Theorem 4.1.2. *We have*

$$\sum_{n\geq 0} B(n)\frac{x^n}{n!} = e^{e^x - 1}. \qquad \square$$

We end this section by discussing certain permutations whose descent sets have a nice structure. To compute their egf we will need to define the *formal sine power series* by

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots = \sum_{n\geq 0}(-1)^n\frac{x^{2n+1}}{(2n+1)!}$$

and

$$\cos x = (\sin x)', \quad \sec x = \frac{1}{\cos x}, \quad \tan x = \frac{\sin x}{\cos x}.$$

Note that $\sec x$ and $\tan x$ are well-defined by Theorem 3.3.1.

Call a permutation $\pi \in P([n])$ *alternating* if

$$\pi_1 > \pi_2 < \pi_3 > \pi_4 < \cdots \tag{4.3}$$

or equivalently if $\operatorname{Des}\pi$ consists of the odd number in $[n]$. The nth *Euler number* is

$$E_n = \text{ the number of alternating } \pi \in P([n]).$$

For example, when $n = 4$ then the alternating permutations are

$$2143, 3142, 3241, 4132, 4231$$

so $E_4 = 5$. A permutation is *complement alternating* if $\pi_1 < \pi_2 > \pi_3 < \pi_4 > \cdots$ or equivalently π^c is alternating where π^c is the complement of π as defined in Exercise (37)(b) of Chapter 1. Clearly E_n also counts the number of complement alternating $\pi \in P([n])$. More generally, define any sequence of integers to be alternating using (4.3) and similarly for complement alternating. We have the following result for the Euler numbers.

Theorem 4.1.3. *We have $E_0 = E_1 = 1$ and, for $n \geq 1$,*

$$2E_{n+1} = \sum_{k=0}^{n} \binom{n}{k} E_k E_{n-k}.$$

Also

$$\sum_{n\geq 0} E_n \frac{x^n}{n!} = \sec x + \tan x.$$

Proof. To prove the recurrence it will be convenient to consider the set S which is the union of all permutations which are either alternating or complement alternating in $P([n+1])$. So $\#S = 2E_{n+1}$. Pick $\pi \in S$ and suppose $\pi_k = n+1$. Then π factors as a word $\pi = \pi'(n+1)\pi''$. Suppose first that π is alternating. Then k is odd, π' is alternating, and π'' is complement alternating. The number of ways of choosing the elements of π' is $\binom{n}{k}$ and the remaining ones are used for π''. The number of ways of arranging the elements for π' in alternating order is E_k, and for π'' it is E_{n-k}. So the total number of such π is $\binom{n}{k}E_k E_{n-k}$ where k is odd. Similar considerations show that the same formula holds for even k when π is complement alternating. The summation side of the recursion follows.

Now let $E(x)$ be the egf for the E_n. Multiplying the recurrence by $x^n/n!$ and summing over $n \geq 1$ one obtains the differential equation and boundary condition

$$2E'(x) = E(x)^2 + 1 \quad \text{and} \quad E(0) = 1$$

where, to make things well-defined for formal power series, $E(0)$ is an abbreviation for the constant term of $E(x)$. One now obtains the unique solution $E(x) = \sec x + \tan x$, either by separation of variables or by verifying that this function satisfies the differential equation and initial condition. $\square$

4.2. Generating functions for Eulerian polynomials

In Section 3.2 we saw that the inv and maj statistics have the same distribution. It turns out that there are statistics that have the same distribution as des and these are called *Eulerian*. The polynomials having this distribution have nice corresponding generating functions of both the ordinary and exponential type. We will discuss them in this section. A whole book devoted to this topic has been written by Petersen [**69**].

Given $n, k \in \mathbb{N}$ with $0 \le k < n$, the corresponding *Eulerian number* is

$$A(n,k) = \#\{\pi \in \mathfrak{S}_n \mid \operatorname{des}\pi = k\}.$$

As usual, we let $A(n,k) = 0$ if $k < 0$ or $k \ge n$ with the exception $A(0,0) = 1$. For example, if $n = 3$, then we have

k	0	1	2
$\operatorname{des}\pi = k$	123	132, 213, 231, 312	321

so that

$$A(3,0) = 1, \quad A(3,1) = 4, \quad A(3,2) = 1.$$

Be sure not to confuse these Eulerian numbers with the Euler numbers introduced in the previous section. Also, some authors use $A(n,k)$ to denote the number of permutations in $\mathfrak{S}_n$ having $k-1$ descents. Some elementary properties of the $A(n,k)$ are given in the next result. It will be convenient to let

$$A([n],k) = \{\pi \in \mathfrak{S}_n \mid \operatorname{des}\pi = k\}.$$

Theorem 4.2.1. *Suppose $n \ge 0$.*

(a) *The Eulerian numbers satisfy the initial condition*

$$A(0,k) = \delta_{k,0}$$

and recurrence relation

$$A(n,k) = (k+1)A(n-1,k) + (n-k)A(n-1,k-1)$$

for $n \ge 1$.

(b) *The Eulerian numbers are symmetric in that*

$$A(n,k) = A(n, n-k-1).$$

(c) *We have*

$$\sum_k A(n,k) = n!\,.$$

Proof. We leave all except the recursion as an exercise. Suppose $\pi \in A([n],k)$. Then removing n from π results in $\pi' \in A([n-1],k)$ or $\pi'' \in A([n-1],k-1)$ depending on the relative size of the elements to either side of n in π. A permutation π' will result if either the n in π is in the space corresponding to a descent of π', or at the end of π'. So a given π' will result $k+1$ times by this method, which accounts for the first term in the sum. Similarly, one obtains a π'' from π if n is either in the space of an ascent or at the beginning. So the total number of repetitions in this case is $n-k$ and the recurrence is proved. □

The *nth Eulerian polynomial* is

$$A_n(q) = \sum_{k\geq 0} A(n,k)q^k = \sum_{\pi\in\mathfrak{S}_n} q^{\operatorname{des}\pi}.$$

Any statistic having distribution $A_n(q)$ is said to be an *Eulerian statistic*. One of the other famous Eulerian statistics counts excedances. An *excedance* of a permutation $\pi \in \mathfrak{S}_n$ is an integer i such that $\pi(i) > i$. This gives rise to the *excedance set*

$$\operatorname{Exc}\pi = \{i \mid \pi(i) > i\}$$

and *excedance statistic*

$$\operatorname{exc}\pi = \#\operatorname{Exc}\pi.$$

To illustrate, if $\pi = 3167542$, then $\pi(1) = 3$, $\pi(3) = 6$, and $\pi(4) = 7$ while $\pi(i) \leq i$ for other i. So $\operatorname{Exc}\pi = \{1, 3, 4\}$ and $\operatorname{exc}\pi = 3$. Making a chart for $n = 3$ as we did for des gives

k	0	1	2
$\operatorname{exc}\pi = k$	123	132, 213, 312, 321	231

so that the number of permutations in each column is given by the $A(3, k)$, even though the sets of permutations in the two tables are not necessarily equal.

In order to prove that $A(n, k)$ also counts permutations by excedances, we will need a map that is so important in enumerative combinatorics that it is sometimes called the fundamental bijection. Before we can define this function, we will need some more concepts. Similar to what was done in Section 1.12, call an element π_i of $\pi \in \mathfrak{S}_n$ a *left-right maximum* if

$$\pi_i > \max\{\pi_1, \pi_2, \ldots, \pi_{i-1}\}.$$

Note that π_1 and n are always left-right maxima and that the left-right maxima increase left-to-right. To illustrate, the left-right maxima of $\pi = 51327846$ are 5, 7, and 8. The left-right maxima of π determine the *left-right factorization* of π into factors $\pi_i\pi_{i+1}\ldots\pi_{j-1}$ where π_i is a left-right maximum and π_j is the next such maximum. In our example π, the factorization is 5132, 7, and 846.

Recall that since disjoint cycles commute, there are many ways of writing the disjoint cycle decomposition $\pi = c_1c_2\cdots c_k$. We wish to distinguish one which is analogous to the left-right factorization. The *canonical cycle decomposition* of π is obtained by writing each c_i so that it starts with $\max c_i$ and then ordering the cycles so that

$$\max c_1 < \max c_2 < \cdots < \max c_k.$$

To illustrate, the permutation $\pi = (7, 1, 8)(2, 4, 5, 3)(6)$ written canonically becomes $\pi = (5, 3, 2, 4)(6)(8, 7, 1)$.

The *fundamental map* is $\Phi : \mathfrak{S}_n \to \mathfrak{S}_n$ where $\Phi(\pi)$ is obtained by replacing each left-right factor $\pi_i\pi_{i+1}\ldots\pi_{j-1}$ by the cycle $(\pi_i, \pi_{i+1}, \ldots, \pi_{j-1})$. For example

$$\Phi(51327846) = (5, 1, 3, 2)(7)(8, 4, 6) = 35261874.$$

Note that from the definitions it follows that the cycle decomposition of $\Phi(\pi)$ obtained will be the canonical one. It is also easy to construct an inverse for Φ: given $\sigma \in \mathfrak{S}_n$, we construct its canonical cycle decomposition and then just remove the parentheses and commas to get π. These maps are inverses since the inequalities defining the left-right factorization and canonical cycle decomposition are the same. We have proved the following.

Theorem 4.2.2. *The fundamental map* $\Phi : \mathfrak{S}_n \to \mathfrak{S}_n$ *is a bijection.* □

Corollary 4.2.3. *For* $n, k \geq 0$ *we have*

$$A(n,k) = \text{number of } \pi \in \mathfrak{S}_n \text{ with } k \text{ excedances.}$$

Proof. A *coexcedance* of $\pi \in \mathfrak{S}_n$ is $i \in [n]$ such that $\pi(i) < i$. Note that the distribution of coexcedances over $\mathfrak{S}_n$ is the same as for excedances. Indeed, we have a bijection on $\mathfrak{S}_n$ defined by $\pi \mapsto \pi^{-1}$. And this bijection has the property that the number of excedances of π is the number of coexcedances of π^{-1} because one obtains the two-line notation for π^{-1} (as defined in Section 1.5) by taking the two-line notation for π, interchanging the top and bottom lines, and then permuting the columns until the first row is $12\ldots n$.

We now claim that if $\Phi(\pi) = \sigma$ where Φ is the fundamental bijection, then $\operatorname{des}\pi$ is the number of coexcedances of σ which will finish the proof. But if we we have a descent $\pi_i > \pi_{i+1}$ in π, then π_i, π_{i+1} must be in the same factor of the left-right factorization. So in $\Phi(\pi)$ we have a cycle mapping π_i to π_{i+1}. This makes π_i a coexcedance of σ. Similar ideas show that no ascent of σ gives rise to a coexcedance of σ and so we are done. □

We will now derive two generating functions involving the Eulerian polynomials, one ordinary and one exponential.

Theorem 4.2.4. *For* $n \geq 0$ *we have*

$$\frac{A_n(q)}{(1-q)^{n+1}} = \sum_{m\geq 0} (m+1)^n q^m. \tag{4.4}$$

Proof. We will count *descent partitioned permutations* $\overline{\pi}$ which consist of a permutation $\pi \in \mathfrak{S}_n$ which has bars inserted in some of its spaces either between elements or before π_1 or after π_n, subject to the restriction that the space between each descent $\pi_i > \pi_{i+1}$ must have a bar. For example, if $\pi = 2451376$, then we could have $\overline{\pi} = 24|5||137|6|$. Let $b(\overline{\pi})$ be the number of bars in $\overline{\pi}$. We will show that both sides of (4.4) are the generating function $f(q) = \sum_{\overline{\pi}} q^{b(\overline{\pi})}$.

First of all, given π, what is its contribution to $f(q)$? First we must put bars in the descents of π which results in a factor of $q^{\operatorname{des}\pi}$. Now we can choose the rest of the bars by putting them in any of the $n+1$ spaces of π with repetition allowed which, by Theorem 3.4.2, gives a factor of $1/(1-q)^{n+1}$. So

$$f(q) = \sum_{\pi\in\mathfrak{S}_n} \frac{q^{\operatorname{des}\pi}}{(1-q)^{n+1}} = \frac{A_n(q)}{(1-q)^{n+1}}$$

which is the first half of what we wished to prove.

On the other hand, the coefficient of q^m in $f(q)$ is the number of $\overline{\pi}$ which have exactly m bars. One can construct these permutations as follows. Start with m bars which create $m+1$ spaces between them. Now place the numbers $1,\dots,n$ between the bars, making sure that the numbers between two consecutive bars form an increasing sequence. So we are essentially placing n distinguishable balls into $m+1$ distinguishable boxes since the ordering in each box is fixed. By the twelvefold way, there are $(m+1)^n$ ways of doing this, which completes the proof. □

We can now use the ogf just derived to find the egf for the polynomials $A_n(x)$.

Theorem 4.2.5. *We have*

$$\sum_{n\ge 0} A_n(q)\frac{x^n}{n!} = \frac{q-1}{q-e^{(q-1)x}}. \tag{4.5}$$

Proof. Multiply both sides of the equality in the previous theorem by the quantity $(1-q)^n x^n/n!$ and sum over n. The left-hand side becomes

$$\sum_{n\ge 0} \frac{(1-q)^n A_n(q)x^n}{(1-q)^{n+1}n!} = \frac{1}{1-q}\sum_{n\ge 0} A_n(q)\frac{x^n}{n!}.$$

And the right side is now

$$\begin{aligned}
\sum_{n\ge 0}\sum_{m\ge 0} q^m \frac{(1-q)^n(m+1)^n x^n}{n!} &= \sum_{m\ge 0} q^m \sum_{n\ge 0}\frac{[(1-q)x(m+1)]^n}{n!}\\
&= \sum_{m\ge 0} q^m e^{(1-q)x(m+1)}\\
&= \frac{e^{(1-q)x}}{1-qe^{(1-q)x}}\\
&= \frac{1}{e^{(q-1)x}-q}.
\end{aligned}$$

Setting the two sides equal and solving for the desired generating function completes the proof. □

4.3. Labeled structures

There is a method for working combinatorially with exponential generating functions which we will present in the following sections. It is based on Joyal's theory of species [**47**]. His original method used the machinery of categories and functors. But for the type of enumeration we will be doing, it is not necessary to use this level of generality. An exposition of the full theory can be found in the textbook of Bergeron, Labelle, and Leroux [**11**].

A *labeled structure* is a function $\mathcal{S}$ which assigns to each finite set L a finite set $\mathcal{S}(L)$ such that

$$\#L = \#M \implies \#\mathcal{S}(L) = \#\mathcal{S}(M). \tag{4.6}$$

We call L the *label set* and $\mathcal{S}(L)$ the set of *structures on L*. We let

$$s_n = \#\mathcal{S}(L)$$

for any L of cardinality n, and this is well-defined because of (4.6). We sometimes use $\mathcal{S}(\cdot)$ as an alternative notation for the structure $\mathcal{S}$. We also have the corresponding egf

$$F_{\mathcal{S}} = F_{\mathcal{S}(\cdot)}(x) = \sum_{n\geq 0} s_n \frac{x^n}{n!}.$$

Although these definitions may seem very abstract, we have already seen many examples of labeled structures. It is just that we have not identified them as such. The rest of this section will be devoted to putting these examples in context. A summary can be found in Table 4.1.

To start, consider the labeled structure defined by $\mathcal{S}(L) = 2^L$. So $\mathcal{S}$ assigns to each label set L the set of subsets of L. To illustrate

$$\mathcal{S}(\{a,b\}) = \{\emptyset,\ \{a\},\ \{b\},\ \{a,b\}\}.$$

Clearly $\mathcal{S}$ satisfies (4.6) with $s_n = \#2^{[n]} = 2^n$. So the associated generating function is

$$F_{2^\cdot}(x) = \sum_{n\geq 0} 2^n \frac{x^n}{n!} = e^{2x}. \tag{4.7}$$

We will also want to specify the size of the subsets under consideration by using the structure $\mathcal{S}(L) = \binom{L}{k}$ for some fixed $k \geq 0$. Now we have $s_n = \binom{n}{k}$ and, using the fact that this binomial coefficient is zero for $n < k$,

$$F_{\binom{\cdot}{k}}(x) = \sum_{n\geq 0} \binom{n}{k} \frac{x^n}{n!} = \sum_{n\geq k} \frac{n!}{k!\,(n-k)!} \cdot \frac{x^n}{n!} = \frac{x^k}{k!} \sum_{n\geq k} \frac{x^{n-k}}{(n-k)!} = \frac{x^k}{k!} e^x. \tag{4.8}$$

It will be convenient to have a map which adds no extra structure to the label set. So define the labeled structure $E(L) = \{L\}$. Note that E returns the set consisting of L itself, not the set consisting of the elements of L. Consequently $s_n = 1$ for all n and $F_E = e^x$. The use of E for this labeled structure reflects both the fact that its egf is the exponential function and also that the French word for “set” is “ensemble”. (Joyal is a francophone.)

We will also need to specify that a set be nonempty by defining

$$\overline{E}(L) = \begin{cases} \{L\} & \text{if } L \neq \emptyset, \\ \emptyset & \text{if } L = \emptyset. \end{cases}$$

Note that $E(\emptyset) = \{\emptyset\}$ while $\overline{E}(\emptyset) = \emptyset$. For $\overline{E}$ we clearly have $s_n = 1$ for $n \geq 1$ and $s_0 = 0$. It is also obvious that

$$F_{\overline{E}} = e^x - 1. \tag{4.9}$$

For partitions of sets we will use the structure $L \mapsto B(L)$ where $B(L)$ is defined as in Section 1.4. So $s_n = B(n)$ and, by Theorem 4.1.2, the egf is

$$F_B = \sum_{n\geq 0} B(n) \frac{x^n}{n!} = e^{e^x - 1}.$$

We will be able to give a combinatorial derivation of this fact once we derive the Exponential Formula in Section 4.5, rather than using the recursion and formal manipulations as we did before.

Table 4.1. Labeled structures

$\mathcal{S}(L)$	Counts	s_n	egf
2^L	subsets	2^n	$\sum_{n\ge0} 2^n \frac{x^n}{n!} = e^{2x}$
$\binom{L}{k}$	k-subsets	$\binom{n}{k}$	$\sum_{n\ge0} \binom{n}{k}\frac{x^n}{n!} = \frac{x^k e^x}{k!}$
$E(L)$	sets	1	$\sum_{n\ge0} \frac{x^n}{n!} = e^x$
$\overline{E}(L)$	nonempty sets	$1-\delta_{n,0}$	$\sum_{n\ge1} \frac{x^n}{n!} = e^x - 1$
$B(L)$	set partitions	B_n	$\sum_{n\ge0} B_n \frac{x^n}{n!} = e^{e^x-1}$
$S(L,k)$	set partitions with k blocks	$S(n,k)$	$\sum_{n\ge0} S(n,k)\frac{x^n}{n!} = \frac{1}{k!}(e^x-1)^k$
$S_o(L,k)$	ordered version of $S(L,k)$	$k!\,S(n,k)$	$k!\sum_{n\ge0} S(n,k)\frac{x^n}{n!} = (e^x-1)^k$
$\mathfrak{S}(L)$	permutations	$n!$	$\sum_{n\ge0} n!\,\frac{x^n}{n!} = \frac{1}{1-x}$
$c(L,k)$	permutations with k cycles	$c(n,k)$	$\sum_{n\ge0} c(n,k)\frac{x^n}{n!} = \frac{1}{k!}\left(\ln\frac{1}{1-x}\right)^k$
$c_o(L,k)$	ordered version of $c(L,k)$	$k!\;c(n,k)$	$k!\sum_{n\ge0} c(n,k)\frac{x^n}{n!} = \left(\ln\frac{1}{1-x}\right)^k$
$c(L)$	permutations with 1 cycle	$(n-1)!$	$\sum_{n\ge1}(n-1)!\,\frac{x^n}{n!} = \ln\frac{1}{1-x}$

Just as with subsets, we will restrict our attention to partitions with a given number of blocks by using $L \mapsto S(L,k)$. Now we have $s_n = S(n,k)$, a Stirling number of the second kind. But we have yet to find a closed form for the egf $\sum_{n\ge0} S(n,k)x^n/n!$ to verify that entry in Table 4.1. We will be able to do this easily once we have the sum and product rules for egfs presented in the next section.

We will sometimes work with set partitions where there is a specified ordering on the blocks and use the notation

$$S_o(L,k) = \{(B_1, B_2, \ldots, B_k) \mid B_1/B_2/\ldots/B_k \vdash L\}.$$

We call these *ordered set partitions* or *set compositions*. Note that the ordering is of the blocks themselves, not of the elements in each block, so that $(\{1,3\},\{2\}) \neq (\{2\},\{1,3\})$ but $(\{1,3\},\{2\}) = (\{3,1\},\{2\})$. Clearly the labeled structure $L \mapsto S_o(L,k)$ has $s_n = k!\,S(n,k)$ and a similar statement can be made for the egf. We will also need *weak set compositions* where we will allow empty blocks.

One can look at labeled structures on permutations analogously to what we have just seen for set partitions. In this context, consider a permutation of L to be a bijection $\pi : L \to L$ decomposed into cycles as we did when $L = [n]$ in Section 1.5. Let $\mathcal{S}(L) = \mathfrak{S}(L)$ be the labeled structure of all permutations of L so that $s_n = n!$ and $F_{\mathfrak{S}} = \sum_n n!\,x^n/n! = 1/(1-x)$. We have the associated structures

$$c(L,k) = \{\pi = c_1 c_2 \cdots c_k \mid \pi \text{ is a permutation of } L \text{ with } k \text{ cycles } c_i\}$$

with ordered variant

$$c_o(L,k) = \{(c_1, c_2, \ldots, c_k) \mid \text{the } c_i \text{ are the cycles of a permutation of } L\}.$$

Using the signless Stirling numbers of the first kind we see that the sequences enumerating these two structures are $c(n,k)$ and $k!\,c(n,k)$, respectively. Again, we will wait to evaluate the corresponding egfs.

Finally, we will find the special case $c(L) \coloneqq c(L,1)$ of having only one cycle particularly useful. In this case, the enumerator is easy to compute.

Proposition 4.3.1. *We have*

$$\#c([n]) = \begin{cases} (n-1)! & \text{if } n \geq 1, \\ 0 & \text{if } n = 0. \end{cases}$$

Proof. The empty permutation has no cycles so that $c(\emptyset) = 0$. Suppose $n \geq 1$ and consider $(a_1, a_2, \ldots, a_n) \in c([n])$. Then the number of such cycles is the number of ways to order the a_i divided by the number of orderings which give the same cycle, namely $n!\,/n = (n-1)!$. □

It follows from the previous proposition that

$$F_c = \sum_{n\geq 1} (n-1)!\,\frac{x^n}{n!} = \sum_{n\geq 1} \frac{x^n}{n} = \ln \frac{1}{1-x}$$

by definition (4.2).

4.4. The Sum and Product Rules for egfs

Just as with sets and ogfs, there is a Sum Rule and a Product Rule for egfs. To derive these results, we will first need corresponding rules for labeled structures.

Suppose $\mathcal{S}$ and $\mathcal{T}$ are labeled structures. If $\mathcal{S}(L) \cap \mathcal{T}(L) = \emptyset$ for any finite set L, then we say $\mathcal{S}$ and $\mathcal{T}$ are *disjoint*. In this case we define their *disjoint union structure*, $\mathcal{S} \uplus \mathcal{T}$, by

$$(\mathcal{S} \uplus \mathcal{T})(L) = \mathcal{S}(L) \uplus \mathcal{T}(L).$$

It is easy to see, and we will prove in Proposition 4.4.1 below, that $\mathcal{S} \uplus \mathcal{T}$ satisfies the definition of a labeled structure. As examples of this concept, suppose $\#L = n$. Then 2^L can be partitioned into the subsets of L having size k for $0 \leq k \leq n$. In other words

$$2^L = \binom{L}{0} \uplus \binom{L}{1} \uplus \cdots \uplus \binom{L}{n}. \tag{4.10}$$

Note that to make a statement about all L regardless of cardinality we can write $2^L = \biguplus_{k\geq 0} \binom{L}{k}$ since $\binom{L}{k} = \emptyset$ for $k > \#L$. Similarly, we have

$$B(L) = S(L,0) \uplus S(L,1) \uplus \cdots \uplus S(L,n) \tag{4.11}$$

and

$$\mathfrak{S}(L) = c(L,0) \uplus c(L,1) \uplus \cdots \uplus c(L,n). \tag{4.12}$$

To define products, let $\mathcal{S}$ and $\mathcal{T}$ be arbitrary labeled structures. Their *product*, $\mathcal{S}\times\mathcal{T}$, is defined by

$$\begin{aligned}&(\mathcal{S} \times \mathcal{T})(L)\\ &\quad = \{(S,T) \mid S \in \mathcal{S}(L_1),\ T \in \mathcal{T}(L_2) \text{ with } (L_1, L_2) \text{ a weak composition of } L\}.\end{aligned}$$

Intuitively, we carve L up into two subsets in all possible ways and put an $\mathcal{S}$-structure on the first subset and a $\mathcal{T}$-structure on the second. Again, we will show that this is indeed a labeled structure in Proposition 4.4.1. Strictly speaking, $\mathcal{S} \times \mathcal{T}$ should be a multiset since it is possible that the same pair (S, T) could arise from two different ordered partitions. However, in the examples we will use, this will never be the case. And the theorems we will prove about the product will still be true in the more general context if we count with multiplicity.

In order to give some examples using products, we will need a notion of equivalence of structures. Say that labeled structures $\mathcal{S}$ and $\mathcal{T}$ are *equivalent*, and write $\mathcal{S} \equiv \mathcal{T}$ if

$$\#\mathcal{S}(L) = \#\mathcal{T}(L)$$

for all finite L. Sometimes we will write $\mathcal{S}(L) \equiv \mathcal{T}(L)$ for this concept if the context makes inclusion of a generic label set L convenient. Clearly if $\mathcal{S} \equiv \mathcal{T}$, then $F_{\mathcal{S}}(x) = F_{\mathcal{T}}(x)$.

As a first illustration of these concepts, we claim that

$$2^{\cdot} \equiv (E \times E)(\cdot). \tag{4.13}$$

To see this, note that subsets $S \in 2^L$ are in bijection with weak compositions via the map
$$S \leftrightarrow (S, L - S).$$
So $\#2^L = \#(E \times E)(L)$ as we wished to show. These same ideas demonstrate that we have $S_o(\cdot, 2) = (\overline{E} \times \overline{E})(\cdot)$ and more generally, for any $k \geq 0$,

$$S_o(\cdot, k) = \overline{E}^k(\cdot). \tag{4.14}$$

In a similar manner, we obtain

$$c_o(\cdot, k) = c^k(\cdot). \tag{4.15}$$

It is time to prove the Sum and Product Rules for labeled structures. In so doing, we will also be showing that they satisfy the definition for a labeled structure (4.6).

Proposition 4.4.1. *Let $\mathcal{S}, \mathcal{T}$ be labeled structures and let*
$$s_n = \#\mathcal{S}(L), \quad t_n = \#\mathcal{T}(L)$$
where $\#L = n$.

(a) (*Sum Rule*) *If $\mathcal{S}$ and $\mathcal{T}$ are disjoint, then*
$$\#(\mathcal{S} \uplus \mathcal{T})(L) = s_n + t_n.$$

(b) (*Product Rule*) *For any $\mathcal{S}, \mathcal{T}$*
$$\#(\mathcal{S} \times \mathcal{T})(L) = \sum_{k=0}^{n} \binom{n}{k} s_k t_{n-k}.$$

Proof. For part (a) we have
$$\#(\mathcal{S} \uplus \mathcal{T})(L) = \#(\mathcal{S}(L) \uplus \mathcal{T}(L)) = \#\mathcal{S}(L) + \#\mathcal{T}(L) = s_n + t_n.$$

Now consider part (b). In order to construct $(S, T) \in (\mathcal{S} \times \mathcal{T})(L)$ we must first pick a weak composition $L = L_1 \uplus L_2$. This is equivalent to just picking L_1 as then $L_2 = L - L_1$. So if $\#L_1 = k$, then there are $\binom{n}{k}$ ways to perform this step. Next we must put an $\mathcal{S}$-structure on L_1 and a $\mathcal{T}$-structure on L_2 which can be done in $s_k t_{n-k}$ ways. Multiplying together the two counts and summing over all possible k yields the desired formula. □

As application of this result, note that applying the Sum Rule to (4.10) just gives $2^n = \sum_k \binom{n}{k}$ which is Theorem 1.3.3(c). And if we apply the Product Rule to (4.13), we get
$$2^n = \sum_{k=0}^{n} \binom{n}{k} \cdot 1 \cdot 1 = \sum_{k=0}^{n} \binom{n}{k}$$
again. Somewhat more interesting formulas are derived in Exercise 8(b) of this chapter.

We can now translate Proposition 4.4.1 into the corresponding rules for exponential generating functions. This will permit us to fill in the entries in Table 4.1 which were postponed in the previous section.

Theorem 4.4.2. *Let $\mathcal{S}, \mathcal{T}$ be labeled structures.*

(a) (*Sum Rule*) *If $\mathcal{S}$ and $\mathcal{T}$ are disjoint, then*

$$F_{\mathcal{S} \uplus \mathcal{T}}(x) = F_{\mathcal{S}}(x) + F_{\mathcal{T}}(x).$$

(b) (*Product Rule*) *For any $\mathcal{S}, \mathcal{T}$*

$$F_{\mathcal{S} \times \mathcal{T}}(x) = F_{\mathcal{S}}(x) \cdot F_{\mathcal{T}}(x).$$

Proof. Let $s_n = \mathcal{S}([n])$ and $t_n = \mathcal{T}([n])$. Using the Sum Rule in Proposition 4.4.1 gives

$$\begin{aligned} F_{\mathcal{S}}(x) + F_{\mathcal{T}}(x) &= \sum_{n\geq 0} s_n \frac{x^n}{n!} + \sum_{n\geq 0} t_n \frac{x^n}{n!} \\ &= \sum_{n\geq 0} (s_n + t_n) \frac{x^n}{n!} \\ &= \sum_{n\geq 0} \#(\mathcal{S} \uplus \mathcal{T})([n]) \frac{x^n}{n!} \\ &= F_{\mathcal{S} \uplus \mathcal{T}}(x). \end{aligned}$$

Now using the Product Rule of the same proposition yields

$$\begin{aligned} F_{\mathcal{S}}(x) F_{\mathcal{T}}(x) &= \left(\sum_{n\geq 0} s_n \frac{x^n}{n!}\right)\left(\sum_{n\geq 0} t_n \frac{x^n}{n!}\right) \\ &= \sum_{n\geq 0} \left(\sum_{k=0}^{n} \frac{s_k}{k!} \cdot \frac{t_{n-k}}{(n-k)!}\right) x^n \\ &= \sum_{n\geq 0} \left(\sum_{k=0}^{n} \binom{n}{k} s_k t_{n-k}\right) \frac{x^n}{n!} \\ &= \sum_{n\geq 0} \#(\mathcal{S} \times \mathcal{T})([n]) \frac{x^n}{n!} \\ &= F_{\mathcal{S} \times \mathcal{T}}(x), \end{aligned}$$

which completes the proof. □

As an illustration of how this result can be used, we can apply the Sum Rule to (4.10), keeping in mind the comment following the equation, to write

$$F_{2^{\cdot}}(x) = \sum_{k\geq 0} F_{\binom{\cdot}{k}}(x).$$

We can check this using (4.7) and (4.8):

$$\sum_{k\geq 0} F_{\binom{\cdot}{k}}(x) = \sum_{k\geq 0} \frac{x^k}{k!} e^x = e^x \sum_{k\geq 0} \frac{x^k}{k!} = e^x \cdot e^x = e^{2x} = F_{2^{\cdot}}(x).$$

One can also apply the Product Rule to (4.13) and obtain

$$F_{2^\cdot}(x) = F_E(x)F_E(x).$$

Again, this yields a simple identity; namely $e^{2x} = e^x \cdot e^x$.

The true power of Theorem 4.4.2 is that it can be used to derive egfs which are more complicated to prove by other means. For example, applying the Product Rule to equation (4.14) along with (4.9) yields

$$F_{S_o(\cdot,k)}(x) = F_{\overline{E}}(x)^k = (e^x - 1)^k,$$

a new entry for Table 4.1. Furthermore, since $F_{S_o(\cdot,k)}(x) = k!\, F_{S(\cdot,k)}(x)$ we obtain

$$F_{S(\cdot,k)}(x) = \frac{(e^x - 1)^k}{k!}.$$

This permits us to give another derivation of the egf for $B(n)$. Using the Sum Rule and (4.11) gives

$$F_B(x) = \sum_{k\geq 0} F_{S(\cdot,k)}(x) = \sum_{k\geq 0} \frac{(e^x-1)^k}{k!} = e^{e^x-1}.$$

These same ideas can be used to derive the egfs for permutations with a given number of cycles, as the reader is asked to do in the exercises.

4.5. The Exponential Formula

Often in combinatorics and other areas of mathematics there are objects which can be broken down into components. For example, the components of set partitions are blocks and the components of permutations are cycles. The exponential formula determines the egf of a labeled structure in terms of the egf for its components. It can also be considered as an analogue of the Product Rule for egfs where one carves L into an arbitrary number of subsets (rather than just 2) and the subsets are unordered (rather than ordered).

To make these ideas precise, let $\mathcal{S}$ be a labeled structure satisfying

$$\mathcal{S}(L) \cap \mathcal{S}(M) = \emptyset \text{ if } L \neq M. \tag{4.16}$$

The corresponding *partition structure*, $\Pi(\mathcal{S})$, is defined by

$$(\Pi(\mathcal{S}))(L) = \{\{S_1, S_2, \dots\} \mid \text{for all } L_1/L_2/\dots \vdash L \text{ with } S_i \in \mathcal{S}(L_i) \text{ for all } i\}.$$

Intuitively, to form $(\Pi(\mathcal{S}))(L)$ we partition the label set L in all possible ways and then put a structure from $\mathcal{S}$ on each block of the partition, again in all possible ways. Condition (4.16) is imposed so that each element of $(\Pi(\mathcal{S}))(L)$ can only arise in one way from this process. To illustrate,

$$B(L) = \{L_1/L_2/\dots \vdash L\} \equiv (\Pi(E))(L) \tag{4.17}$$

since L_i is the only element of $E(L_i)$ for any i. In much the same way, we see that

$$\begin{aligned}\mathfrak{S}(L) &= \{c_1c_2\cdots \mid c_i \text{ a cycle on } L_i \text{ for all } i \text{ for all } L_1/L_2/\dots \vdash L\}\\ &\equiv (\Pi(c))(L).\end{aligned} \tag{4.18}$$

There is a simple relationship between the egf for $\Pi(\mathcal{S})$ and the egf for $\overline{\mathcal{S}}$ which is the labeled structure defined by

$$\overline{\mathcal{S}}(L) = \begin{cases} \mathcal{S}(L) & \text{if } L \neq \emptyset, \\ \emptyset & \text{if } L = \emptyset. \end{cases}$$

So if $s_n = \#\mathcal{S}([n])$, then

$$F_{\overline{\mathcal{S}}}(x) = \sum_{n \geq 1} s_n \frac{x^n}{n!}.$$

We need $F_{\overline{\mathcal{S}}}(x)$ to have a zero constant term so that the composition in the next result will be well-defined, see Theorem 3.3.3.

Theorem 4.5.1 (Exponential Formula). *If $\mathcal{S}$ is a labeled structure satisfying* (4.16)*, then*

$$F_{\Pi(\mathcal{S})}(x) = e^{F_{\overline{\mathcal{S}}}(x)}.$$

Proof. We have

$$e^{F_{\overline{\mathcal{S}}}(x)} = \sum_{k \geq 0} \frac{F_{\overline{\mathcal{S}}}(x)^k}{k!}.$$

From the Product Rule for egfs in Theorem 4.4.2 we see that $F_{\overline{\mathcal{S}}}(x)^k$ is the egf for putting $\mathcal{S}$-structures on partitions of the label set into k ordered, nonempty blocks. So, by (4.16), $F_{\overline{\mathcal{S}}}(x)^k/k!$ is the egf for putting $\mathcal{S}$-structures on partitions of the label set into k unordered, nonempty blocks. Now using the Sum Rule for egfs, again from Theorem 4.4.2, it follows that $\sum_{k \geq 0} F_{\overline{\mathcal{S}}}(x)^k/k!$ is the egf for putting $\mathcal{S}$-structures on partitions of the label set into any number of unordered, nonempty blocks. But this is exactly the structure $\Pi(\mathcal{S})$ and so we are done. $\square$

As a first application of the Exponential Formula, consider (4.17). In this case $\mathcal{S} = E$ and $F_{\overline{E}}(x) = e^x - 1$. So, applying the previous theorem,

$$F_B(x) = F_{\Pi(\overline{E})}(x) = e^{F_{\overline{E}}(x)} = e^{e^x - 1}.$$

Even though we already knew this generating function, this proof is definitely the simplest both computationally and conceptually.

We can use (4.18) in a similar manner. Now $\mathcal{S} = c$ and

$$F_c(x) = \ln(1/(1-x)) = F_{\overline{c}}(x)$$

since the original egf already has no constant term. Applying the Exponential Formula gives

$$F_{\mathfrak{S}}(x) = F_{\Pi(\overline{c})}(x) = e^{F_{\overline{c}}(x)} = e^{\ln(1/(1-x))} = \frac{1}{1-x}$$

which at least agrees with what we computed previously for this egf, even though this in now a more roundabout way of getting it. But with Theorem 4.5.1 in hand it is easy to get more refined information about permutations or other labeled structures. For example, suppose we wish to give a simpler and more combinatorial derivation for the egf of the derangement numbers $D(n)$ found in Theorem 4.1.1. The corresponding structure is defined by

$$\mathcal{D}(L) = \text{ derangements on } L.$$

In order to express $\mathcal{D}(L)$ as a partition structure we need to permit only cycles of length two or greater. So let

$$\mathcal{S}(L) = \begin{cases} c(L) & \text{if } \#L \geq 2, \\ \emptyset & \text{otherwise.} \end{cases}$$

It follows that $\mathcal{D} \equiv \Pi(\mathcal{S})$. Furthermore,

$$s_n = \begin{cases} (n-1)! & \text{if } n \geq 2, \\ 0 & \text{otherwise,} \end{cases}$$

so that

$$F_{\overline{\mathcal{S}}}(x) = \sum_{n\geq 2}(n-1)!\,\frac{x^n}{n!} = \sum_{n\geq 2}\frac{x^n}{n} = \ln\left(\frac{1}{1-x}\right) - x.$$

Applying the Exponential Formula gives

$$\sum_{n\geq 0} D(n)\frac{x^n}{n!} = F_{\mathcal{D}}(x) = F_{\Pi(\mathcal{S})}(x) = \exp\left(\ln\left(\frac{1}{1-x}\right) - x\right) = \frac{e^{-x}}{1-x}. \tag{4.19}$$

One can mine even more information from Theorem 4.5.1 since the proof shows that each of the summands $F_{\overline{\mathcal{S}}}(x)^k/k!$ has a combinatorial meaning. Define the *hyperbolic sine* and *cosine* functions to be the formal power series

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots = \sum_{n\geq 0}\frac{x^{2n+1}}{(2n+1)!}$$

and

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots = \sum_{n\geq 0}\frac{x^{2n}}{(2n)!}.$$

It is easy to see that for any formal power series $f(x) = \sum_{n\geq 0} a_n x^n$ we can extract the series of odd or even terms by

$$\sum_{n\geq 0} a_{2n+1}x^{2n+1} = \frac{f(x)-f(-x)}{2} \quad \text{and} \quad \sum_{n\geq 0} a_{2n}x^{2n} = \frac{f(x)+f(-x)}{2}. \tag{4.20}$$

It follows that

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad \text{and} \quad \cosh x = \frac{e^x + e^{-x}}{2}. \tag{4.21}$$

Define the *odd partition structure* $\Pi_o(\mathcal{S})$ by

$$\begin{aligned}(\Pi_o(\mathcal{S}))(L) \\ &= \{\{S_1, S_2, \ldots\} \in (\Pi(\mathcal{S}))(L) \mid L \text{ partitioned into an odd number of blocks}\}\end{aligned}$$

and similarly define the *even partition structure* $\Pi_e(\mathcal{S})$. A proof like that of the Exponential Formula can be used to demonstrate the following.

Theorem 4.5.2. *If $\mathcal{S}$ is a labeled structure satisfying* (4.16)*, then*

$$F_{\Pi_o(\mathcal{S})}(x) = \sinh F_{\overline{\mathcal{S}}}(x)$$

and

$$F_{\Pi_e(\mathcal{S})}(x) = \cosh F_{\overline{\mathcal{S}}}(x).$$

□

Now suppose we wish to find the egf for a_n which is the number of permutations of $[n]$ that have an odd number of cycles. As before $\mathcal{S} = c$ with $F_c(x) = F_{\bar{c}}(x) = \ln(1/(1-x))$. Using Theorem 4.5.2 and then (4.21) we see that

$$\begin{aligned}\sum_{n\geq 0} a_n \frac{x^n}{n!} &= \sinh F_{\bar{c}}(x)\\ &= \frac{e^{\ln(1/(1-x))} - e^{-\ln(1/(1-x))}}{2}\\ &= \frac{1}{2}\left(\frac{1}{1-x} - (1-x)\right)\\ &= x + \frac{1}{2}\sum_{n\geq 2} x^n.\end{aligned}$$

Extracting the coefficient of $x^n/n!$ from the first and last sums above yields

$$a_n = \begin{cases} n!/2 & \text{if } n \geq 2,\\ 1 & \text{if } n = 1.\end{cases} \tag{4.22}$$

Of course, once one has obtained such a simple answer, one would like a purely combinatorial explanation and the reader is encouraged to find one in Exercise 12(c) of this chapter.

Exercises

(1) (a) Use the recursion for the derangement numbers in Exercise 4 of Chapter 2 to reprove Theorem 4.1.1.
   (b) Use the recursion for the derangement numbers in Exercise 5 of Chapter 2 to reprove Theorem 4.1.1.

(2) (a) Finish the proof of Theorem 4.2.1.
   (b) Finish the proof of Corollary 4.2.3.
   (c) Give two proofs of the identity

$$(m+1)^n = \sum_{k\geq 0} A(n,k)\binom{m+n-k}{n},$$

one using equation (4.4) and one using descent partitioned permutations.

(d) Give a combinatorial proof of the following formula for $A(n,k)$:

$$A(n,k) = \sum_{i=0}^{k+1} (-1)^i \binom{n+1}{i} (k-i+1)^n.$$

Hint: Use the Principle of Inclusion and Exclusion and descent partitioned permutations.

(e) Give a combinatorial proof of the following recursion for $n \geq 1$:

$$A_n(q) = A_{n-1}(q) + q \sum_{i=0}^{n-2} \binom{n-1}{i} A_i(q) A_{n-i-1}(q).$$

Hint: Factor each $\pi \in \mathfrak{S}_n$ as $\pi = \sigma n \tau$.

(f) Use (e) to give a second proof of (4.5).

(3) Let $I \subset \mathbb{P}$ be finite and let $m = \max I$ if I is nonempty or $m = 0$ if $I = \emptyset$. For $n > m$ define the corresponding *descent polynomial* $d(I;n)$ to be the number of $\pi \in \mathfrak{S}_n$ such that $\operatorname{Des} \pi = I$.

(a) Prove that $d([k];n) = \binom{n-1}{k}$.

(b) If $I \neq \emptyset$, then let $I^- = I - \{m\}$. Prove that

$$d(I;n) = \binom{n}{m} d(I^-;m) - d(I^-;n).$$

Hint: Consider the set of $\pi \in \mathfrak{S}_n$ such that $\operatorname{Des}(\pi_1\pi_2 \ldots \pi_m) = I^-$ and $\pi_{m+1} < \pi_{m+2} < \cdots < \pi_n$.

(c) Use part (b) to show that $d(I;n)$ is a polynomial in n having degree $\deg(I;n) = m$.

(d) Reprove the fact that $d(I;n)$ is a polynomial in n using the Principle of Inclusion and Exclusion.

(e) Since $d(I;n)$ is a polynomial in n, its domain of definition can be extended to all $n \in \mathbb{C}$. Show that if $i \in I$, then $d(I;i) = 0$.

(f) Show that the complex roots of $d(I;n)$ all lie in the circle $|z| \leq m$ in the complex plane and also all have real part greater than or equal to -1. Note: This seems to be a difficult problem.

(g) (Conjecture) Show that the complex roots of $d(I;n)$ all lie in the circle

$$\left| z - \frac{m+1}{2} \right| \leq \frac{m-1}{2}.$$

Note that this conjecture implies part (f).

(4) (a) Derive the generating function

$$\sum_{n,k \geq 0} S(n,k) t^k \frac{x^n}{n!} = e^{t(e^x - 1)}$$

in two ways: using the recursion for the $S(n,k)$ and using the generating functions in Table 4.1.

(b) Rederive the egf for the Bell numbers $B(n)$ using part (a).

(5) (a) Find a formula for $\sum_{n,k \geq 0} c(n,k) t^k x^n / n!$ and prove it in two ways: using the recursion for the $c(n,k)$ and using the generating functions in Table 4.1.

(b) Rederive the egf for the permutation structure $\mathfrak{S}(\cdot)$ using part (a).

(6) (a) Let i_n be the number of involutions in $\mathfrak{S}_n$. Show that $i_0 = i_1 = 1$ and for $n \ge 2$

$$i_n = i_{n-1} + (n-1)i_{n-2}.$$

(b) Show that

$$\sum_{n\ge 0} i_n \frac{x^n}{n!} = e^{x+x^2/2}$$

in two ways: using the recursion in part (a) and using the Exponential Formula.

(c) Given $A \subseteq \mathbb{P}$, let $S(n,A)$ be the number of partitions of $[n]$ all of whose block sizes are elements of A. Use the Exponential Formula to find and prove a formula for $\sum_{\ge 0} S(n,A)x^n/n!$.

(d) Repeat part (c) for $c(n,A)$, the number of permutations of $[n]$ all of whose cycles have lengths which are elements of A.

(7) Fill in the details for finding the egf and solving the differential equation in the proof of Theorem 4.1.3.

(8) (a) Use (4.14) and the Product Rule for labeled structures to show that

$$S(n,2) = 2^{n-1} - 1.$$

(b) Use (4.15) and the Product Rule for labeled structures to show that

$$c(n+1,2) = n! \sum_{k=1}^{n} \frac{1}{k}.$$

(9) (a) Use the Theorem 4.4.2 to derive the egfs in Table 4.1 for the structures $c_o(\cdot,k)$ and $c(\cdot,k)$.

(b) Use part (a) to rederive the egf for the structure $\mathfrak{S}(\cdot)$.

(10) (a) Suppose $\mathcal{S}$ is a labeled structure satisfying (4.16) and $\mathcal{T}$ is any labeled structure. Their *composition*, $\mathcal{T} \circ \mathcal{S}$, is the structure such that

$$\begin{aligned}(\mathcal{T} \circ \mathcal{S})(L) = \{(\{S_1, S_2, \dots\}, T) \mid & \text{ for all } L_1/L_2/\dots \vdash L \\ & \text{with } S_i \in \mathcal{S}(L_i) \text{ for all } i \text{ and } T \in \mathcal{T}(\{S_1, S_2, \dots\})\}.\end{aligned}$$

Prove that

$$F_{\mathcal{T}\circ\mathcal{S}}(x) = F_{\mathcal{T}}(F_{\overline{\mathcal{S}}}(x)).$$

(b) Use (a) to reprove the Exponential Formula.

(11) Let $\mathcal{F}(L)$ be the labeled structures consisting of all forests with L as vertex set. Show that

$$\sum_{n\ge 0} \#\mathcal{F}([n]) \frac{x^n}{n!} = \exp\left(\sum_{n\ge 1} n^{n-2} x^n/n!\right).$$

(12) (a) Prove the identities (4.20).

(b) Prove Theorem 4.5.2.

(c) Reprove (4.22) by finding a bijection between the permutations of $[n]$, $n \ge 2$, which have an odd number of cycles and those which have an even number of cycles.

(d) Find a formula for the number of permutations of $[n]$ having an even number of cycles in two ways: by using Theorem 4.5.2 and by using (4.22).

(13) Let a_n be the number of permutations in $\mathfrak{S}_n$ that have an even number of cycles, all of them of odd length.

(a) Use a parity argument to show that if n is odd, then $a_n = 0$.

(b) Use egfs to show that if n is even, then

$$a_n = \binom{n}{n/2}\frac{n!}{2^n}.$$

(c) Use part (b) to show that if n is even, then the probability that in tossing a fair coin n times exactly $n/2$ heads occur is the same as the probability that a permutation chosen uniformly at random from $\mathfrak{S}_n$ has an even number of cycles, all of them of odd length.

(d) Reprove part (c) by giving, when n is even, a bijection between pairs (S, π) where $S \in \binom{[n]}{n/2}$ and $\pi \in \mathfrak{S}_n$ and pairs (T, σ) where $T \in 2^{[n]}$ and $\sigma \in \mathfrak{S}_n$ has an even number of cycles, all of them of odd length.

(14) Let j_n be the number of involutions in $\mathfrak{S}_n$ which have no fixed points.

(a) Give a combinatorial proof that $j_{2n+1} = 0$ and that $j_{2n} = 1 \cdot 3 \cdot 5 \cdots (2n-1)$.

(b) Use the Exponential Formula to find a simple expression for the exponential generating function $\sum_{n\geq 0} j_n \frac{x^n}{n!}$.

(c) Use the exponential generating function from (b) to give a second derivation of the formula in (a).

Chapter 5

Counting with Partially Ordered Sets

Partially ordered sets, known as "posets" for short, give a fruitful way of ordering combinatorial objects. In this way they provide new perspectives on objects we have already studied, such as interpreting various combinatorial invariants as rank functions. They also give us new and powerful tools to do enumeration such as the Möbius Inversion Theorem which generalizes the Principle of Inclusion and Exclusion.

5.1. Basic properties of partially ordered sets

A *partially ordered set* or *poset* is a pair $(P, \leq)$ where "P" is a set and "$\leq$" is a binary relation on P satisfying the following axioms for all $x, y, z \in P$:

(a) (reflexivity) $x \leq x$,

(b) (antisymmetry) if $x \leq y$ and $y \leq x$, then $x = y$, and

(c) (transitivity) if $x \leq y$ and $y \leq z$, then $x \leq z$.

Often we will refer to the poset as just P since the partial order will be obvious from context. We will also use notation for the standard order on $\mathbb{R}$ in the setting of posets in the obvious way. For example $x < y$ means $x \leq y$ but $x \neq y$, or using $y \geq x$ as equivalent to $x \leq y$. We say that $x, y \in P$ are *comparable* if $x \leq y$ or $y \leq x$. Otherwise they are *incomparable*. A poset where every pair of elements is comparable is called a *total order*.

There are standard partial orders on many of the combinatorial objects we have already studied and we list some of them here.

- The *chain of length* n is $(C_n, \leq)$ where $C_n = \{0, 1, \ldots, n\}$ and $i \leq j$ is the usual ordering of the integers. So C_n is a total order. Note that C_n has $n+1$ elements and in some texts this would be referred to as an $(n + 1)$-chain. Although

C_n was also used as the notation for the graphical cycle with n vertices, the context should make it clear which object is meant.

- The *Boolean algebra* is $(B_n, \subseteq)$ where $B_n = 2^{[n]}$ and $S \subseteq T$ is set containment. Be sure not to confuse B_n with the nth Bell number $B(n)$.
- The *divisor lattice* is $(D_n, |)$ where D_n consists of all the positive integers which divide evenly into n and $a \mid b$ means that a divides evenly into b (in that b/a is an integer). So in D_{12} we have $2 \leq 6$ but $2 \not\leq 3$. Note the distinction between D_n and the derangement number $D(n)$.
- The *lattice of partitions* is $(\Pi_n, \leq)$ where Π_n is the set of all partitions of $[n]$ and $\rho \leq \tau$ means that every block of ρ is contained in some block of τ, called the *refinement* ordering. For example, in Π_6 we have $14/2/36/5 \leq 1245/36$ because $\{1, 4\}$, $\{2\}$, and $\{5\}$ are all contained in $\{1, 2, 4, 5\}$ and $\{3, 6\}$ is contained in itself.
- *Young's lattice* is $(Y, \leq)$ where Y is the set of all integer partitions and $\lambda \leq \mu$ is containment of Young diagrams as defined in Section 3.2.
- The *lattice of compositions* is $(K_n, \leq)$ where K_n is the set of all compositions of n and $\alpha \leq \beta$ is refinement of compositions: α can be obtained from β by replacing each β_i by a composition $[\alpha_j, \alpha_{j+1}, \ldots, \alpha_k] \models \beta_i$. For example, in K_{11} we have $[2, 3, 2, 1, 3] \leq [2, 5, 4]$ because the first 2 in $[2, 5, 4]$ was replaced by itself, the 5 was replaced by $[3, 2]$, and the 4 by $[1, 3]$. On the other hand $[2, 3, 1, 2, 3] \not\leq [2, 5, 4]$. Again, there is a notational overlap with K_n as the complete graph on n vertices, but the two will never appear together.
- The *pattern poset* is $(\mathfrak{S}, \leq)$ where $\mathfrak{S}$ is the set of all permutations and $\pi \leq \sigma$ means that σ contains π as a pattern.
- The *subspace lattice* is $(L(V), \leq)$ where $L(V)$ is the set of subspaces of a finite vector space V over $\mathbb{F}_q$ and $U \leq W$ means U is a subspace of W. If $V = \mathbb{F}_q^n$, then we often denote this poset by $L_n(q)$.

There are also important partial orders on permutations which reflect their group structure (strong and weak Bruhat order) and on certain subgraphs of a graph (the bond lattice). We will define the latter later when it is needed.

Often a poset $(P, \leq)$ is represented by a certain (di)graph which can be easier to work with than just using the axioms. If $x, y \in P$, then we say that *x is covered by y* or *y covers x*, written either $x \lessdot y$ or $y \gtrdot x$, if $x < y$ and there is no $z \in P$ with $x < z < y$. The *Hasse diagram* of P is the graph with vertices P and an edge from x up to y if $x \lessdot y$. Note that this is actually a digraph where all the arcs are directed up and so are just written as edges with this understanding. Hasse diagrams for examples from the above list are given in Figure 5.1. For those which are infinite, only the bottom of the poset is displayed. Note that in the case of the subspace lattice $V = \mathbb{F}_3^2$ the subspaces are listed using their row-echelon forms. We will make no distinction between a poset and its Hasse diagram if no confusion will result by blurring the distinction. Finally, certain posets such as the real numbers under their normal ordering have no covers. So in such cases it does not make sense to try to draw a Hasse diagram.

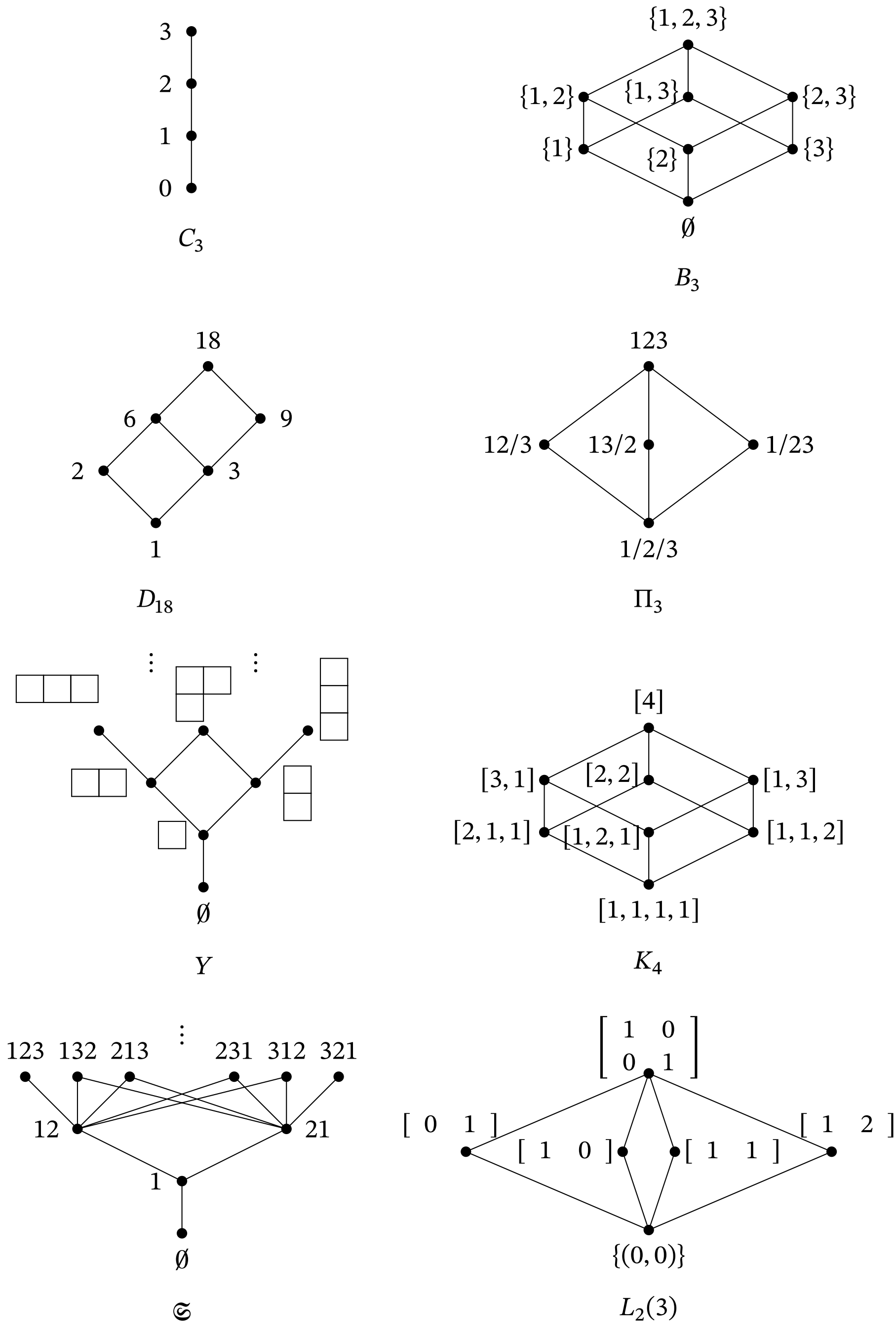


Figure 5.1. A zoo of Hasse diagrams

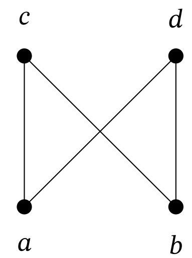


Figure 5.2. Minimal and maximal elements

There are certain parts of a poset P to which we will often refer. A *minimal element* of P is $x \in P$ such that there is no $y \in P$ with $y < x$. Note that P can have multiple minimal elements. The poset in Figure 5.2 has minimal elements a, b. Dually, define a *maximal element* of P to be $x \in P$ with no $y > x$ in P. The example poset just cited has maximal elements c, d. By way of contrast, P has a *minimum element* if there is $x \in P$ such that $x \leq y$ for every $y \in P$. A minimum element is unique if it exists because if x and x' are both minimum elements, then $x \leq x'$ and $x' \leq x$ which forces $x = x'$ by antisymmetry. In this case the minimum element is often denoted $\hat{0}$. All of the posets in Figure 5.1 have a $\hat{0}$. In fact, in D_n we have $\hat{0} = 1$, a rare instance where one can write that zero equals one and be mathematically correct! Again, there is the dual notion of a *maximum element* $x \in P$ which satisfies $x \geq y$ for all $y \in P$. A maximum is unique if it exists and is denoted $\hat{1}$. The following result sums up the existence of minimum and maximum elements for the posets in Figure 5.1. Its proof is sufficiently easy that it is left as an exercise.

Proposition 5.1.1. *We have the following minimum and maximum elements.*

- *In C_n we have $\hat{0} = 0$, $\hat{1} = n$.*
- *In B_n we have $\hat{0} = \emptyset$, $\hat{1} = [n]$.*
- *In D_n we have $\hat{0} = 1$, $\hat{1} = n$.*
- *In Π_n we have $\hat{0} = 1/2/\ldots/n$, $\hat{1} = [n]$.*
- *In Y we have $\hat{0} = \emptyset$ and no $\hat{1}$.*
- *In K_n we have $\hat{0} = [1^n]$, $\hat{1} = [n]$.*
- *In $\mathfrak{S}$ we have $\hat{0} = \emptyset$ and no $\hat{1}$.*
- *In $L(V)$ we have $\hat{0}$ is the zero subspace, $\hat{1} = V$.* □

As one can tell from the previous set of definitions, it is sometimes useful to reverse inequalities in a poset. So if P is a poset, then we define its *dual* P^* to have the same underlying set with $x \leq y$ in P^* if and only if $y \leq x$ in P. The Hasse diagram of P^* is thus obtained by reflecting the one for P in a horizontal axis.

As is often the case in mathematics, we analyze a structure by looking at its substructures. In posets P, these come in several varieties. A *subposet* of P is a subset $Q \subseteq P$ with the inherited partial order; namely $x \leq y$ for $x, y \in Q$ if and only if $x \leq y$

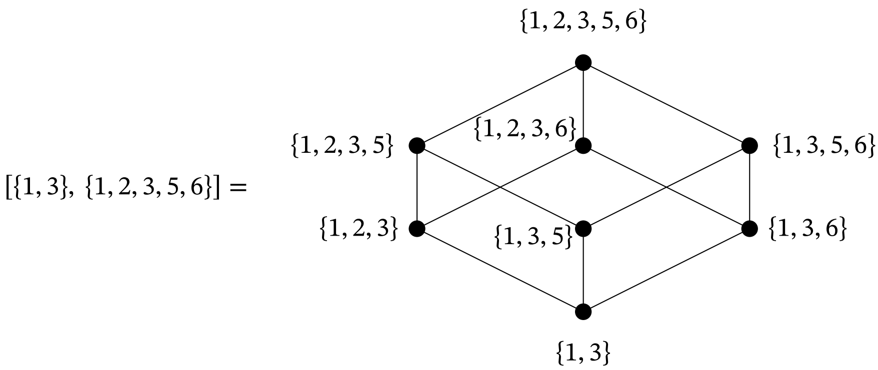


Figure 5.3. An interval in B_7

in P. In such a case we will sometimes use a subscript to make precise which poset is being considered, as in $x \leq_P y$. Note that some authors call this an *induced subposet* and use the term "subposet" when Q satisfies the weaker condition that $x \leq_Q y$ implies $x \leq_P y$ (but not necessarily conversely). There are several especially important subposets. Assume $x, y \in P$. Then the corresponding *closed interval* is

$$[x, y] = \{z \in P \mid x \leq z \leq y\}.$$

Note that $[x, y] = \emptyset$ unless $x \leq y$. For example, the Hasse diagram of the interval $[\{1, 3\}, \{1, 2, 3, 5, 6\}]$ in B_7 is displayed in Figure 5.3. Note that if one removes the labels, then this diagram is exactly like the one for B_3 in Figure 5.1. We will explain this formally when we introduce the concept of isomorphism below. Open and half-open intervals in a poset are defined as expected. A subset $I \subseteq P$ is a *lower-order ideal* if $x \in I$ and $y \leq x$ imply that $y \in I$. For example, if P has a $\hat{0}$, then any interval of the form $[\hat{0}, x]$ is a lower-order ideal. If $S \subseteq P$ is any subset, then the *lower-order ideal generated by* S is

$$I(S) = \{y \in P \mid y \leq x \text{ for some } x \in S\}.$$

We often leave out the set braces in S, for example, writing $I(x, y)$ for the ideal generated by $S = \{x, y\} \subseteq P$. If $\#S = 1$, then the order ideal is called *principal*. If P has a $\hat{0}$, then $I(x) = [\hat{0}, x]$. In Young's lattice, $I(\lambda)$ is just all the partitions contained in λ. So when we have a rectangle $\lambda = (k^l)$, then $I(\lambda) = \mathcal{R}(k, l)$, the set which came into play when discussing the q-binomial coefficients in Section 3.2. *upper-order ideals* U as well as those generated by a set, $U(S)$, are defined by reversing all the inequalities. We will sometimes abbreviate "lower-order ideal" to "order ideal" or even just "ideal," whereas for upper-order ideals both adjectives will always be used.

Some simple properties of ideals are given in the next proposition.

Proposition 5.1.2. *Let P be a poset.*

(a) *We have that $I \subseteq P$ is a lower-order ideal if and only if $P - I$ is an upper-order ideal.*

(b) *If P is finite and I is a lower-order ideal, then $I = I(S)$ where S is the set of maximal elements of I.*

(c) *If P is finite and U is an upper-order ideal, then $U = U(S)$ where S is the set of minimal elements of U.*

Proof. We will prove (b) and leave the other two parts to the reader. We will prove the equality by proving the two corresponding set containments. Suppose $x \in I$ and consider the set $X = \{y \in I \mid y \geq x\}$. This subset of P is nonempty since $x \in X$. So X has at least one maximal element y since P is finite. In fact, y must be maximal in I since, if not, there is $z > y$ with $z \in I$. But then by transitivity $z > x$ so that $z \in X$. This contradicts the maximality of y in X. So $y \in S$ and $x \in I(S)$ showing that $I \subseteq I(S)$.

To show $I(S) \subseteq I$, take $y \in I(S)$. By definition $y \leq x$ for some $x \in S$ and $S \subseteq I$. So $y \leq x$ where $x \in I$ which forces $y \in I$ by definition of lower-order ideal. □

To define isomorphism, we need to consider maps on posets. Assume posets P, Q. Then a function $f : P \to Q$ is *order preserving* if

$$x \leq_P y \implies f(x) \leq_Q f(y).$$

For example, the map $f : C_n \to B_n$ by $f(i) = [i]$ is order preserving because $i \leq j$ implies that $f(i) = [i] \subseteq [j] = f(j)$. We say that f is an *isomorphism* or that P and Q are *isomorphic*, written $P \cong Q$, if f is bijective and both f and f^{-1} are order preserving. It is important to show that f^{-1}, and not just f, is order preserving. For consider the two posets in Figure 5.4. Define a bijection $f : P \to Q$ by $f(a) = a'$ and $f(b) = b'$. Then f is order preserving vacuously since there are no order relations in P. But clearly we do not want f to be an isomorphism since P and Q have different (unlabeled) Hasse diagrams. This is witnessed by the fact that f^{-1} is not order preserving: we have $a' \leq b'$ but $f^{-1}(a') = a \nleq b = f^{-1}(b')$.

There are a number of isomorphisms involving the posets in Figure 5.1. Some of them are collected in the next result.

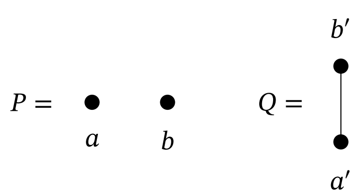


Figure 5.4. Two nonisomorphic posets

Proposition 5.1.3. *We have the following isomorphisms. In all cases we assume the intervals used are nonempty.*

(a) $C_n^* \cong C_n$.
(b) *If* $i, j \in C_n$, *then* $[i, j] \cong C_{j-i}$.
(c) $B_n^* \cong B_n$.
(d) *If* $S, T \in B_n$, *then* $[S, T] \cong B_{|T-S|}$.
(e) $D_n^* \cong D_n$.
(f) *If* $l, m \in D_n$, *then* $[l, m] \cong D_{m/l}$.
(g) *If* n *is a product of* k *distinct primes, then* $D_n \cong B_k$.
(h) *If* $\rho \in \Pi_n$ *has* k *blocks, then* $[\rho, \hat{1}] \cong \Pi_k$.
(i) *For all* n *we have* $K_n \cong B_{n-1}$.
(j) *If* V, W *are vector spaces over* $\mathbb{F}_q$ *of the same dimension, then* $L(V) \cong L(W)$.
(k) *If* $X, Y \in L(V)$, *then* $[X, Y] \cong L(X/Y)$ *where* X/Y *is the quotient vector space.*

Proof. We will prove the statement about $[S, T]$ in B_n and leave the rest of the isomorphisms as exercises. Let $T - S = \{t_1, t_2, \ldots, t_n\}$ where $n = |T - S|$. Define a map $f : [S, T] \to B_n$ as follows. If $X \in [S, T]$, then $X = S \uplus X'$ where $X' \subseteq T - S$. If $X' = \{t_i, \ldots, t_j\}$, then let $f(X) = \{i, \ldots, j\}$. This is well-defined since, by definition, $i, \ldots, j \in [n]$.

To show that f is a bijection, we construct its inverse. Assume $I = \{i, \ldots, j\} \in B_n$. Then let $f^{-1}(I) = S \uplus \{t_i, \ldots, t_j\}$. The proof that this is well-defined is similar to that of f, and the fact that these are inverses is clear from their definitions.

Finally, we need to show that f and f^{-1} are order preserving. If $X \le Y$ in $[S, T]$, then $X' \subseteq Y'$ in $T - S$. It follows that $f(X) \le f(Y)$ in B_n. Thus f is order preserving. As far as f^{-1}, take $I \le J$ in B_n. Then the corresponding sets T_I and T_J in $T - S$ gotten by using I and J for subscripts satisfy $T_I \subseteq T_J$. It follows that

$$f^{-1}(I) = S \uplus T_I \subseteq S \uplus T_J = f^{-1}(J)$$

which shows that f^{-1} is order preserving. □

5.2. Chains, antichains, and operations on posets

We will now consider three operations for building posets. Chains and the related notion of antichains will play important roles.

Given posets $(P, \le_P)$ and $(Q, \le_Q)$ with $P \cap Q = \emptyset$, their *disjoint union* is the poset whose elements are $P \uplus Q$ with the partial order $x \le_{P \uplus Q} y$ if

(a) $x, y \in P$ and $x \le_P y$ or
(b) $x, y \in Q$ and $x \le_Q y$.

So one just takes the relations in P and Q and does not add any new ones. To illustrate, the poset on the left in Figure 5.4 is the disjoint union $P = \{a\} \uplus \{b\}$. If one takes both posets in this figure, then $P \uplus Q$ is the poset on $\{a, b, a', b'\}$ with $a' < b'$ being the only strict order relation. An important example of disjoint union is the *n-element antichain* A_n which consists of a set of n elements with no strict order relations. So $P = \{a\} \uplus \{b\}$ is a copy of A_2.

Another way to combine disjoint posets is their *ordinal sum* $P + Q$ which has elements $P \uplus Q$ and $x \le_{P+Q} y$ if one of (a), (b), (c) hold where (a) and (b) are as in the previous paragraph and the third possibility is

(c) $x \in P, y \in Q$.

Intuitively, one takes the relations in P and Q and also makes everything in P smaller than everything in Q. As an example using chains, we have $C_m + C_n \cong C_{m+n+1}$. Note that in general $P + Q \not\cong Q + P$ as the use of the adjective "ordinal" suggests.

Our third method to produce new posets from old ones is via products. Given two (not necessarily disjoint) posets $(P, \le_P)$ and $(Q, \le_Q)$, their (*direct or Cartesian*) *product* has underlying set

$$P \times Q = \{(x, y) \mid x \in P, y \in Q\}$$

together with the partial order

$$(x, y) \le_{P\times Q} (x', y') \text{ if } x \le_P x' \text{ and } y \le_Q y'.$$

We let P^n denote the n-fold product of P. One can obtain the Hasse diagram for $P \times Q$ by replacing each vertex of Q with a copy of P and then, for each edge between two vertices of Q, connecting each pair of vertices having the same first coordinate in the corresponding two copies of P. Illustrations of this can be found in Figure 5.1. For example, D_{18} looks like a rectangle because it is isomorphic to $C_1 \times C_2$. Also, $B_3 \cong C_1^3$, which is why the Hasse diagram looks like the projection of a 3-dimensional cube into the plane. Both of these isomorphisms generalize, and there is one for Π_n as well.

Proposition 5.2.1. *We have the following product decompositions.*

(a) *We have*

$$B_n \cong C_1^n.$$

(b) *If the prime factorization of n is $n = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$, then*

$$D_n \cong C_{n_1} \times C_{n_2} \times \cdots \times C_{n_k}.$$

(c) *If $\rho \le \tau$ in Π_n, then*

$$[\rho, \tau] \cong \Pi_{n_1} \times \Pi_{n_2} \times \cdots \times \Pi_{n_k}$$

where $\tau = T_1/T_2/\dots/T_k$ and n_i is the number of blocks of ρ contained in T_i for all i.

Proof. (a) Consider the map f used in the proof of Theorem 1.3.1. Stated in the language of posets, we see that $f : B_n \to C_1^n$. And we have already shown that f is bijective. So we only need to prove that f and f^{-1} are order preserving. Suppose $S, T \subseteq [n]$ with $f(S) = (v_1, \dots, v_n)$ and $f(T) = (w_1, \dots, w_n)$. Now $S \subseteq T$ if and only if $i \in S$ implies $i \in T$. By the definition of f, this is equivalent to $v_i = 1$ implying $w_i = 1$. But this means that $v_i \le w_i$ for all i since if $v_i = 0$, then $v_i \le w_i$ is automatic. Thus we have shown that $S \le T$ in B_n if and only if $(v_1, \dots, v_n) \le (w_1, \dots, w_n)$ in C_1^n which is what we wished to prove.

(b) The map for D_n is similar. Define $g : D_n \to \bigtimes_i C_{n_i}$ by $g(d) = (d_1, d_2, \ldots, d_k)$ where $d = \prod_i p_i^{d_i}$. The reader can now verify that this is a well-defined isomorphism of posets.

(c) The construction of the isomorphism is messy but not conceptually difficult. Consider blocks of ρ as being single elements and then aggregate all those lying in a given block of τ together. Again, the reader can fill in the details. □

So chains can help us understand various posets by taking products. There is another important way in which chains can be used to decompose certain posets. Let P be a poset and $C : x_0 < x_1 < \cdots < x_n$ be a chain in P. We say that C is a chain *of length n from x_0 to x_n* and also use the term x_0–x_n *chain*. We let $\ell(C)$ denote the length of C. Call C *maximal* if it is not strictly contained in a larger chain of P. If $[x_0, x_n]$ is finite, this is equivalent to $x_i < x_{i+1}$ being a cover for all $0 \le i < n$. The chain C is *saturated* if it is not strictly contained in a larger chain from x_0 to x_n. Equivalently, C is saturated if it is maximal in $[x_0, x_n]$. For example, in D_{18} the chain $3 < 6 < 18$ is a chain of length 2 from 3 to 18 which is saturated since each inequality is a cover. But this chain is not maximal since it is contained in the larger chain $1 < 3 < 6 < 18$.

Some posets can be written as a disjoint union of certain subposets called ranks as follows. Suppose P is a poset which is *locally finite* in that the cardinality of any interval $[x, y]$ of P is finite. All the posets in Figure 5.1 are locally finite even though Y and $\mathfrak{S}$ are not finite. The poset of real numbers with its usual total order is not locally finite. Let P be locally finite and have a $\hat{0}$. Then P is *ranked* if for any $x \in P$ all saturated chains from $\hat{0}$ to x have the same length. In this case we call this common length the *rank of x* and it is denoted $\operatorname{rk} x$ or $\operatorname{rk}_P x$ if we wish to be specific about the poset. For $k \in \mathbb{N}$, the *kth rank set* of P is

$$\operatorname{Rk}_k P = \{x \in P \mid \operatorname{rk} x = k\}. \tag{5.1}$$

If P is finite, then we define its *rank* to be

$$\operatorname{rk} P = \max\{k \mid \operatorname{Rk}_k P \neq \emptyset\}.$$

All the posets in Figure 5.1 are ranked and we will describe their rank sets shortly. An example of a poset which is not ranked is in Figure 5.5. This is because there are two saturated $\hat{0}$–$\hat{1}$ chains, namely $\hat{0} < b < \hat{1}$ which is of length 2 and $\hat{0} < a < c < \hat{1}$ which is of length 3.

We will now list the rank information for the posets in Figure 5.1. Note how many of the combinatorial concepts which were introduced in earlier chapters occur naturally in this context. These results are easily proved, so the demonstrations can be filled in by the reader.

Proposition 5.2.2. *All the posets in Figure* 5.1 *are ranked with the following rank functions.*

(a) *If $k \in C_n$, then* $\operatorname{rk}(k) = k$, *so* $\operatorname{Rk}_k(C_n) = \{k\}$. *Also,* $\operatorname{rk}(C_n) = n$.

(b) *If $S \in B_n$, then* $\operatorname{rk}(S) = \#S$, *so* $\operatorname{Rk}_k(B_n) = \binom{[n]}{k}$. *Also,* $\operatorname{rk}(B_n) = n$.

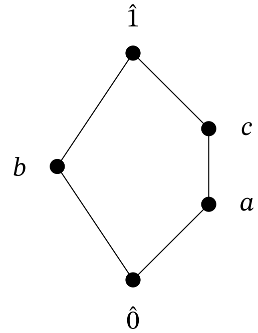


Figure 5.5. An unranked poset

(c) *If $d \in D_n$ has prime factorization $d = p_1^{d_1} \cdots p_r^{d_r}$, then* $\mathrm{rk}(d) = d_1 + \cdots + d_r$. *So* $\mathrm{Rk}_k(D_n)$ *is the set of $d \mid n$ with k primes in their prime factorization, counted with multiplicity. Also,* $\mathrm{rk}(D_n)$ *is the total number of primes dividing n, counted with multiplicity.*

(d) *If $\rho \in \Pi_n$ has b blocks, then* $\mathrm{rk}(\rho) = n - b$, *so* $\mathrm{Rk}_k(\Pi_n) = S([n], n-k)$. *Also,* $\mathrm{rk}(\Pi_n) = n - 1$.

(e) *If $\lambda \in Y$, then* $\mathrm{rk}(\lambda) = |\lambda|$, *so* $\mathrm{Rk}_k(Y) = P(k)$.

(f) *If $\alpha \in K_n$ has c parts, then* $\mathrm{rk}(\alpha) = n - c$, *so* $\mathrm{Rk}_k(K_n) = Q(n, n-k)$. *Also,* $\mathrm{rk}(K_n) = n - 1$.

(g) *If $\pi \in \mathfrak{S}$, then* $\mathrm{rk}(\pi) = |\pi|$, *so* $\mathrm{Rk}_k(\mathfrak{S}) = \mathfrak{S}_k$.

(h) *If $W \in L(V)$, then* $\mathrm{rk}(W) = \dim W$, *so* $\mathrm{Rk}_k(V) = \begin{bmatrix} V \\ k \end{bmatrix}$. *Also,* $\mathrm{rk}(L(V)) = \dim V$. □

5.3. Lattices

The reader will have noticed that several of our example posets are called "lattices". This is an important class of partially ordered sets with the property that pairs of elements have greatest lower bounds and least upper bounds. It is also common to study lattices whose elements satisfy certain identities using these two operations. In this section we will prove a theorem characterizing the lattices satisfying a distributive law.

If P is a poset and $x, y \in P$, then a *lower bound* for x, y is a $z \in P$ such that $z \leq x$ and $z \leq y$. For example, if $S, T \in B_n$, then any set contained in both S and T is a lower bound. We say that x, y have a *greatest lower bound* or *meet* if there is an element in P, denoted $x \wedge y$, which is a lower bound for x, y, and $x \wedge y \geq z$ for all lower bounds z of x and y. Returning to B_n, we have $S \wedge T = S \cap T$. In fact, one can remember the notation for meet as just a squared-off intersection symbol. Note that if the meet of x, y exists, then it is unique. Indeed, if z, z' are both greatest lower bounds of x, y, then we have $z \geq z'$ since z' is a lower bound and z is the greatest lower bound. But interchanging the roles of z and z' also gives $z' \geq z$. So $z = z'$ by antisymmetry. Note also that it is possible for the meet not to exist. For example, in the poset of Figure 5.2, $a \wedge b$ does not exist because this pair has no lower bound. Also, $c \wedge d$ does not exist but this is because the pair has both a and b as lower bounds, but there is no lower bound larger than both a and b. One can extend these definitions in the obvious way from pairs of elements to any nonempty set of elements $X = \{x_1, \ldots, x_n\} \subseteq P$. In this case the meet is denoted

$$\bigwedge X = \bigwedge_{x \in X} x.$$

One can also reasonably define the meet of the empty set as long as P has a $\hat{1}$. Indeed, for any $x \in P$ we would want

$$x \wedge \left(\bigwedge \emptyset\right) = \bigwedge(\{x\} \cup \emptyset) = \bigwedge\{x\} = x.$$

But the only element y of P such that $x \wedge y = x$ for all x is $y = \hat{1}$. So we let $\wedge\emptyset = \hat{1}$.

The concepts of *upper bound* and *least upper bound* are obtained by reversing the inequalities in the definitions of the previous paragraph. If the least upper bound of x, y exists, then it is denoted $x \vee y$ and is also called their *join*. A *lattice* is a poset such that every pair of elements has a meet and a join. Note that this is different from the use of the term "lattice" as in "lattice path" which in this case refers to the discrete subgroup $\mathbb{Z}^2$ of $\mathbb{R}^2$. The context should make it clear which meaning is intended.

All the poset families in Figure 5.1 are lattices except for the pattern poset $\mathfrak{S}$ which has subposets isomorphic to Figure 5.2 between its second and third ranks. To describe the meets and joins we need the following terminology. If $c, d \in \mathbb{P}$, then $\gcd(c, d)$ and $\operatorname{lcm}(c, d)$ denote their greatest common divisor and least common multiple, respectively. Given two Young diagrams λ, μ, we then take their intersection $\lambda \cap \mu$ or union $\lambda \cup \mu$ by aligning them as in Figure 3.1 and then taking the intersection or union of their sets of squares, respectively. If U, W are vector subspaces of V, then their *sum* is

$$U + W = \{u + w \mid u \in U, w \in W\}.$$

We leave the verification of the next result to the reader.

Proposition 5.3.1. *The posets $C_n, B_n, D_n, \Pi_n, Y, K_n$, and $L(V)$ are all lattices for all n and V of finite dimension over some $\mathbb{F}_q$. In addition, we have the following descriptions of their meets and joins.*

(a) *If $i, j \in C_n$, then $i \wedge j = \min\{i, j\}$ and $i \vee j = \max\{i, j\}$.*
(b) *If $S, T \in B_n$, then $S \wedge T = S \cap T$ and $S \vee T = S \cup T$.*
(c) *If $c, d \in D_n$, then $c \wedge d = \gcd(c, d)$ and $c \vee d = \operatorname{lcm}(c, d)$.*
(d) *Suppose $\rho, \tau \in \Pi_n$. Then $\rho \wedge \tau$ is the partition whose blocks are the nonempty intersections of the form $B \cap C$ for blocks $B \in \rho$, $C \in \tau$. Also, $\rho \vee \tau$ is the partition such that b, c are in the same block of the join if and only if there is a sequence of blocks $D_1, \ldots, D_m$ where each D_i is a block of either ρ or τ, $b \in D_1$, $c \in D_m$, and $D_i \cap D_{i+1} \neq \emptyset$ for all i.*
(e) *If $\lambda, \mu \in Y$, then $\lambda \wedge \mu = \lambda \cap \mu$ and $\lambda \vee \mu = \lambda \cup \mu$.*
(f) *If $U, W \in L(V)$, then $U \wedge W = U \cap W$ and $U \vee W = U + W$.* □

Next we will give a list of some elementary properties of lattices.

Proposition 5.3.2. *Let L be a lattice. Then the following are true for all $x, y, z \in L$.*

(a) (*Idempotent law*) $x \wedge x = x \vee x = x$.

(b) (*Commutative law*) $x \wedge y = y \wedge x$ *and* $x \vee y = y \vee x$.

(c) (*Associative law*) $(x \wedge y) \wedge z = x \wedge (y \wedge z)$ *and* $(x \vee y) \vee z = x \vee (y \vee z)$.

(d) (*Absorption law*) $x \wedge (x \vee y) = x = x \vee (x \wedge y)$.

(e) $x \le y \iff x \wedge y = x \iff x \vee y = y$.

(f) *If* $x \le y$, *then* $x \wedge z \le y \wedge z$ *and* $x \vee z \le y \vee z$.

(g) $x \wedge (y \vee z) \ge (x \wedge y) \vee (x \wedge z)$ *and* $x \vee (y \wedge z) \le (x \vee y) \wedge (x \vee z)$.

(h) *If X is a finite, nonempty subset of L, then* $\bigwedge X$ *and* $\bigvee X$ *exist.*

(i) *If L is finite, then L has a* $\hat{0}$ *and a* $\hat{1}$.

(j) *The dual* L^* *is a lattice.*

(k) *If M is a poset and there is an isomorphism* $f : L \to M$, *then M is also a lattice. Furthermore, for* $x, y \in L$,

$$f(x \wedge y) = f(x) \wedge f(y) \quad \textit{and} \quad f(x \vee y) = f(x) \vee f(y).$$

Proof. The proofs of these results are straightforward. So we will give a demonstration for the first inequality in (g) and leave the rest to the reader. By definition of lower bound, $x \wedge y \le x$. And similarly $x \wedge y \le y \le y \vee z$. So by definition of greatest lower bound, $x \wedge y \le x \wedge (y \vee z)$. Similar reason gives $x \wedge z \le x \wedge (y \vee z)$. Using the previous two inequalities and the definition of least upper bound yields $(x \wedge y) \vee (x \wedge z) \le x \wedge (y \vee z)$ which is what we wished to prove. □

It is certainly possible for the inequalities in (g) of the previous proposition to be strict. For example, using the lattice in Figure 5.5 we have

$$c \wedge (a \vee b) = c \wedge \hat{1} = c$$

while

$$(c \wedge a) \vee (c \wedge b) = a \vee \hat{0} = a.$$

However, when we have equality in one of these two inequalities, it is forced in the other.

Proposition 5.3.3. *Let L be a lattice. Then*

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) \ \textit{for all } x, y, z \in L$$

if and only if

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z) \ \textit{for all } x, y, z \in L.$$

Proof. For the forward direction we have

$$\begin{aligned}(x \vee y) \wedge (x \vee z) &= [(x \vee y) \wedge x] \vee [(x \vee y) \wedge z] \\ &= x \vee [(x \vee y) \wedge z] \\ &= x \vee [(x \wedge z) \vee (y \wedge z)] \\ &= [x \vee (x \wedge z)] \vee (y \wedge z) \\ &= x \vee (y \wedge z),\end{aligned}$$

which is the desired equality. The proof of the other direction is similar. □

A lattice which satisfies either of the two equalities in Proposition 5.3.3 is called a *distributive lattice*, and these equations are called the *distributive laws*. Of the lattices in Proposition 5.3.1, Π_n is not distributive for $n \geq 3$. In fact, taking z, y, z to be the three elements of rank 1 in Π_3 gives a strict inequality in the defining relation. Similarly, $L(V)$ is not distributive for $\dim V \geq 2$. All the other lattices are distributive as the reader will be asked to show in the exercises.

Proposition 5.3.4. *The posets C_n, B_n, D_n, Y, K_n are distributive lattices for all n.* □

We will now prove a beautiful theorem characterizing finite distributive lattices due to Birkhoff [**14**]. It says that every such lattice is essentially a set of lower-order ideals partially ordered by inclusion. Given a poset P, consider

$$\mathcal{J}(P) = \{I \mid I \text{ is a lower-order ideal of } P\}.$$

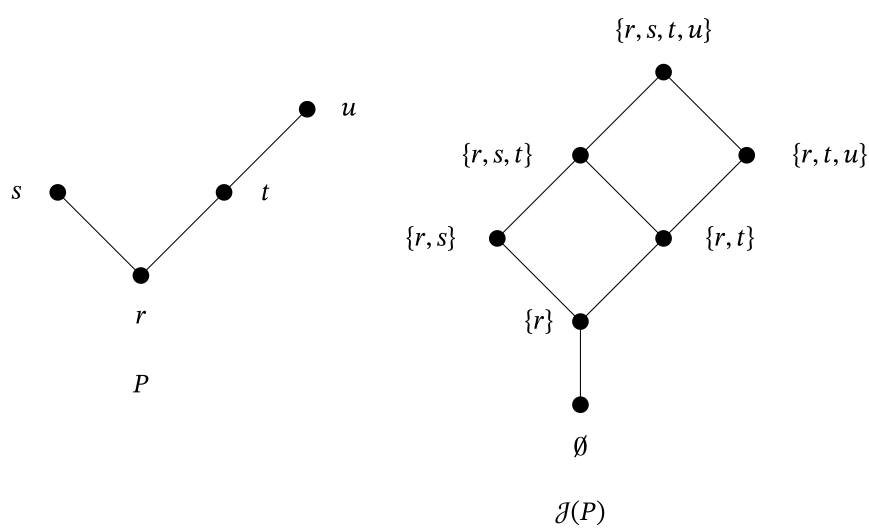


Figure 5.6. A poset and its associated distributive lattice

Turn $\mathcal{J}(P)$ into a poset by letting $I \leq J$ if $I \subseteq J$. For example, a poset P and the corresponding poset $\mathcal{J}(P)$ are shown in Figure 5.6. Our first order of business is to show that $\mathcal{J}(P)$ is always a distributive lattice.

Proposition 5.3.5. *If P is any poset, then $\mathcal{J}(P)$ is a distributive lattice.*

Proof. It suffices to show that if $I, J \in \mathcal{J}(P)$, then so are $I \cap J$ and $I \cup J$. Indeed, this suffices because these are the greatest lower bound and least upper bound if one considers all subsets of P, and set intersection and union satisfy the distributive laws. We will demonstrate this for $I \cap J$ since the argument for union is similar. We need to show that $I \cap J$ is a lower-order ideal of P. So take $x \in I \cap J$ and $y \leq x$. Now $x \in I$ and $x \in J$. Since both sets are ideals, this implies $y \in I$ and $y \in J$. So $y \in I \cap J$ which is what we needed to show. □

The amazing thing is that every finite distributive lattice is of the form $\mathcal{J}(P)$ for some poset P. In order to prove this, we need a way that, given a distributive lattice L, we can identify the P from which it was built. This is done using certain elements of P which we now define. Let L be a finite lattice so that L has a $\hat{0}$ by Proposition 5.3.2(i). Call $x \in L - \{\hat{0}\}$ *join irreducible* if x cannot be written as $x = y \vee z$ where $y, z < x$. Equivalently, if $x = y \vee z$, then $y = x$ or $z = x$. Let

$$\operatorname{Irr}(L) = \{x \in L \mid x \text{ is join irreducible}\}.$$

It turns out the join irreducibles are easy to spot in the Hasse diagram of L. Also, the join irreducibles under a given element join to give that element.

Proposition 5.3.6. *Let L be a finite lattice.*

(a) *Element $x \in L$ is join irreducible if and only if x covers exactly one element.*

(b) *For any $x \in L$, if we let*

$$I_x = \{r \leq x \mid r \in \operatorname{Irr}(L)\}, \tag{5.2}$$

then $x = \bigvee I_x$.

Proof. (a) First note that, by definition, $\hat{0}$ is not join irreducible. And $\hat{0}$ covers no elements, so the proposition is true in that case. Now assume $x \neq \hat{0}$.

For the forward direction suppose, towards a contradiction, that x covers (at least) two elements y, z. But then $x = y \vee z$ since x is clearly an upper bound for y, z and there can be no smaller bound because of the covering relations. This contradicts the fact that x is join irreducible.

For the converse, let x cover a unique element x'. If $x = y \vee z$ with $y, z < x$, then we must have $y, z \leq x'$ because x covers no other element. This forces $y \vee z \leq x' < x$ which is the desired contradiction.

(b) We induct on the number of elements in $[\hat{0}, x]$. If $x = \hat{0}$, then the statement is true because $I_x = \emptyset$ and the empty join equals $\hat{0}$ just as the empty meet equals $\hat{1}$. If $x > \hat{0}$, then there are two cases. If $x \in \operatorname{Irr}(L)$, then x is the maximum element of I_x so $\bigvee I_x = x$ by Proposition 5.3.2(e). If $x \notin \operatorname{Irr}(L)$, then, by part (a), x covers more than one element. So we can write $x = y \vee z$ for any pair of elements y, z covered by x. By induction, $y = \bigvee I_y$ and $z = \bigvee I_z$ where $I_y, I_z \subseteq I_x$. It follows that

$$x = y \vee z = \left(\bigvee I_y\right) \vee \left(\bigvee I_z\right) = \bigvee (I_y \cup I_z) \leq \bigvee I_x \leq x$$

where the last inequality follows since $r \leq x$ for all $r \in I_x$. The previous displayed inequalities force $x = \bigvee I_x$ as desired. □

Using this proposition, the reader can see immediately in Figure 5.6 that $\mathcal{J}(P)$ has four join irreducibles which form a subposet isomorphic to P. Birkhoff's theorem says this always happens.

Theorem 5.3.7 (Fundamental Theorem of Finite Distributive Lattices). *If L is a finite distributive lattice, then $L \cong \mathcal{J}(P)$ where $P = \operatorname{Irr}(L)$.*

Proof. We need to define two order-preserving maps $f : L \to \mathcal{J}(P)$ and $g : \mathcal{J}(P) \to L$ which are inverses of each other. If $x \in L$, then let $f(x) = I_x$ as defined by (5.2). Note that this is well-defined since I_x is an ideal: if $r \in I_x$ and $s \leq r$ is irreducible, then $s \leq r \leq x$ so that $s \in I_x$. Also, f is order preserving since $x \leq y$ implies $I_x \subseteq I_y$ by an argument similar to the one just given. For the inverse map, we let $g(I) = \bigvee I$ for $I \in \mathcal{J}(P)$. Clearly this is an element of L since $P \subset L$ and so g is well-defined. It is also order preserving since if $I \subseteq J$, then

$$g(J) = \bigvee J = \left(\bigvee I\right) \vee \left(\bigvee (J - I)\right) \geq \bigvee I = g(I).$$

It remains to prove that f and g are inverses. If $x \in L$, then, using Proposition 5.3.6(b),

$$g(f(x)) = g(I_x) = \bigvee I_x = x.$$

Now consider $I \in \mathcal{J}(P)$ and let $x = g(I) = \bigvee I$. Then $f(g(I)) = I_x$ and we must show $I = I_x$. For the containment $I \subseteq I_x$, take $r \in I$. So $r \leq \bigvee I = x$. But this means $r \in I_x$ by definition (5.2), which gives the desired subset relation.

To show $I_x \subseteq I$, take $r \in I_x$. We have that $\bigvee I = x = \bigvee I_x$. So $r \wedge (\bigvee I) = r \wedge (\bigvee I_x)$ and, applying the distributive law,

$$\bigvee \{r \wedge s \mid s \in I\} = \bigvee \{r \wedge s \mid s \in I_x\}. \tag{5.3}$$

Since $r \in I_x$, the set on the right in (5.3) contains $r \wedge r = r$. Furthermore, every element of that set is of the form $r \wedge s \leq r$. It follows that the right-hand side of the equality in (5.3) is just r. But r is join irreducible, so there must be some element $s \in I$ on the left in (5.3) such that $r \wedge s = r$. By Proposition 5.3.2(e) this forces $r \leq s \in I$. Since I is an ideal, we have $r \in I$ which is the final step of the proof. □

5.4. The Möbius function of a poset

The Möbius function is a fundamental invariant of any locally finite poset. A special case of this function was first studied in number theory. But the generalization to partially ordered sets is both more powerful and also in some ways more intuitive. It is of great use to enumerators because, as we will see in the next section, it permits one to invert certain summation formulas to get information about the summands.

Let P be a locally finite poset with a $\hat{0}$. The (*one-variable*) *Möbius function* of P is a map $\mu : P \to \mathbb{Z}$ defined inductively by

$$\mu(x) = \begin{cases} 1 & \text{if } x = \hat{0}, \\ -\sum_{y<x} \mu(y) & \text{otherwise.} \end{cases} \tag{5.4}$$

Note that since P is locally finite, the number of summands is finite so that μ is well-defined. By moving the terms in the sum to the left-hand side of the equation, we get the following equivalent definition: for any $x \in P$ we have

$$\sum_{y \le x} \mu(y) = \delta_{\hat{0},x} \tag{5.5}$$

where $\delta_{\hat{0},x}$ is the Kronecker delta. We will write μ_P if we wish to be specific about the poset whose Möbius function is under consideration. Also, if P has a $\hat{1}$, then we will write

$$\mu(P) = \mu(\hat{1}).$$

Let us now calculate the Möbius function for some of our example posets. First consider C_3 as displayed in Figure 5.1. Using (5.4) we see that

$$\begin{aligned} \mu(0) &= 1, \\ \mu(1) &= -\mu(0) = -1, \\ \mu(2) &= -(\mu(0) + \mu(1)) = -0 = 0, \\ \mu(3) &= -(\mu(0) + \mu(1) + \mu(2)) = -0 = 0. \end{aligned}$$

The next result should now be obvious.

Proposition 5.4.1. *In C_n we have*

$$\mu(i) = \begin{cases} 1 & \text{if } i = 0, \\ -1 & \text{if } i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

So $\mu(C_n) = 1, -1,$ or 0 depending on whether we have $n = 0, 1,$ or $n \ge 2$, respectively. □

Now have a look at B_3. Similarly to C_3 we have $\mu(\emptyset) = 1$ and $\mu(\{1\}) = \mu(\{2\}) = \mu(\{3\}) = -1$. Using (5.4) we see that

$$\mu(\{1, 2\}) = -(\mu(\emptyset) + \mu(\{1\}) + \mu(\{2\})) = -(1 - 1 - 1) = 1.$$

Analogous computations show that $\mu(\{1,3\}) = \mu(\{2,3\}) = 1$. Finally

$$\mu(\{1,2,3\}) = -\sum_{S\subset\{1,2,3\}} \mu(S) = -(1-1-1-1+1+1+1) = -1.$$

The next result should not be hard to guess.

Proposition 5.4.2. *If $S \in B_n$, then*

$$\mu(S) = (-1)^{\#S}. \tag{5.6}$$

So $\mu(B_n) = (-1)^n$.

Proof. It will suffice to show that the function $(-1)^{\#S}$ satisfies (5.5) since that equation uniquely defines μ. So suppose $T \in B_n$ and let $\#T = k$. Then, using Theorem 1.3.3(d),

$$\sum_{S\subseteq T}(-1)^{\#S} = \sum_{i=0}^{k}\sum_{S\in\binom{T}{i}}(-1)^i = \sum_{i=0}^{k}\binom{k}{i}(-1)^i = \delta_{0,k} = \delta_{\emptyset,T}$$

which is the desired equality. □

In the divisor lattice D_{18}, the reader should now find it easy to verify that

$$\mu(1) = \mu(6) = 1, \quad \mu(2) = \mu(3) = -1, \quad \mu(9) = \mu(18) = 0.$$

Now the pattern is not as clear. To help us, we will need a result about how the Möbius function interacts with the product operation on posets. But first, it will be useful to have a result about isomorphism and μ.

Theorem 5.4.3. *Let P be a locally finite poset with $\hat{0}$ and let $f : P \to Q$ be in isomorphism. Then for all $x \in P$ we have*

$$\mu_P(x) = \mu_Q(f(x)).$$

Proof. We induct on the cardinality of the ideal $I(x)$. If $\#I(x) = 1$, then $x = \hat{0}_P$ and $f(x) = \hat{0}_Q$. So $\mu_P(x) = 1 = \mu_Q(f(x))$. Now assume $\#I(x) > 1$ so that $x > \hat{0}_P$ and $f(x) > \hat{0}_Q$. Now, by induction,

$$\mu_P(x) = -\sum_{y<x}\mu_P(y) = -\sum_{f(y)<f(x)}\mu_Q(f(y)) = \mu_Q(f(x))$$

as we wished. □

The Möbius function also plays well with poset products.

Theorem 5.4.4. *Let P and Q be locally finite posets containing $\hat{0}_P$ and $\hat{0}_Q$, respectively. Then for all $s \in P$ and $x \in Q$ we have*

$$\mu_{P\times Q}(s,x) = \mu_P(s)\mu_Q(x).$$

Proof. It suffices to show that the right-hand side of the displayed equation satisfies (5.5). But given $(s,x) \in P \times Q$, we have

$$\sum_{(t,y)\le(s,x)}\mu_P(t)\mu_Q(y) = \sum_{t\le s}\mu_P(t)\sum_{y\le x}\mu_Q(y) = \delta_{\hat{0}_P,s}\delta_{\hat{0}_Q,x} = \delta_{(\hat{0}_P,\hat{0}_Q),(s,x)}$$

as desired. □

We can now compute the Möbius function of the divisor lattice.

Proposition 5.4.5. *The Möbius function of D_n is*

$$\mu(d) = \begin{cases} (-1)^m & \text{if } d \text{ is a product of } m \text{ distinct primes,} \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

Proof. We will use the notation and definitions in Proposition 5.2.1 and its proof as well as letting $P = \bigtimes_i C_{n_i}$. Using Theorems 5.4.3 and 5.4.4

$$\mu_{D_n}(d) = \mu_P(g(d)) = \prod_i \mu_{C_{n_i}}(d_i).$$

Recalling the Möbius function for a chain as determined in Proposition 5.4.1, we see that the product is zero if any $d_i \geq 2$ and otherwise equals $(-1)^m$ where m is the number of $d_i = 1$. Since the d_i are the exponents in the prime factorization of d (with $d_i = 0$ if p_i is a prime factor of n but not d), the proposition follows. □

The reader can now see the power of the poset viewpoint in this context. Most number theory texts take (5.7) as the *definition* of the Möbius function, which is not at all intuitive. But from our perspective, this equation is a natural consequence of the fact that D_n is a product of chains. We also note that Theorems 5.4.3 and 5.4.4 can be used to rederive the formula for μ in B_n as the reader is asked to do in the exercises.

We end this section by noting that in a ranked poset P one can get interesting results by looking at the Möbius values at a given rank. Recalling (5.1), define the *Whitney numbers of the second kind* for P to be $W_k(P) = \# \operatorname{Rk}_k(P)$. Equivalently

$$W_k(P) = \sum_{x \in \operatorname{Rk}_k(P)} 1.$$

Also define P's *Whitney numbers of the first kind* as

$$w_k(P) = \sum_{x \in \operatorname{Rk}_k(P)} \mu(x).$$

For example we have $W_k(B_n) = \#\binom{[n]}{k} = \binom{n}{k}$ and, by (5.6),

$$w_k(B_n) = (-1)^k \binom{n}{k}. \tag{5.8}$$

As another illustration, from Proposition 5.2.2(d) we see that

$$W_k(\Pi_n) = \#S([n], n-k) = S(n, n-k)$$

which are the Stirling numbers of the second kind. We will now show that there is a similar relationship between $w_k(\Pi_n)$ and the (signed) Stirling numbers of the first kind. To prove this we need a definition. If $\pi \in \mathfrak{S}_n$ has cycle decomposition $\pi = c_1 \cdots c_k$, then π has *corresponding partition* $\rho = B_1/\ldots/B_k$ where B_i is the set of elements in c_i for all i. Note that, by Proposition 4.3.1, the number of permutations corresponding to a given partition ρ is $\prod_i (|B_i| - 1)!$. Using this fact, the equality

$$w_k(\Pi_n) = s(n, n-k) \tag{5.9}$$

follows immediately from the next proposition.

Proposition 5.4.6. *If* $\rho = B_1/\dots/B_k \in \Pi_n$, *then*

$$\mu(\rho) = (-1)^{n-k}(|B_1|-1)! \cdots (|B_k|-1)!\,. \tag{5.10}$$

So $\mu(\Pi_n) = (-1)^{n-1}(n-1)!$.

Proof. We induct on n, where the case $n = 1$ is easy to verify. Assume that $\mu(\Pi_m) = (-1)^{m-1}(m-1)!$ for $m < n$. It follows from Proposition 5.2.1(c), Theorem 5.4.4, and induction that (5.10) holds for all $\rho < \hat{1}$ in Π_n. To verify that it continues to be true for $\rho = \hat{1}$, it suffices to show that summing the right-hand side of this equation over all of Π_n satisfies (5.5). Using the observation about the number of permutations corresponding to a partition as well as Corollary 1.5.3 gives

$$\begin{aligned}
\sum_{\rho=B_1/\dots/B_k\in\Pi_n} (-1)^{n-k}\prod_{i=1}^{k}(|B_k|-1)! &= \sum_{\pi=c_1\cdots c_k\in\mathfrak{S}_n} (-1)^{n-k} \\
&= \sum_{k=0}^{n}\ \sum_{\pi\in c([n],k)} (-1)^{n-k} \\
&= \sum_{k=0}^{n} s(n,k) \\
&= \delta_{0,n}
\end{aligned}$$

and this finishes the proof. □

5.5. The Möbius Inversion Theorem

In this section we will prove the Möbius Inversion Theorem, which is a very general method for inverting sums over posets P. In fact, we will show that special cases of this result include the Fundamental Theorem of the Difference Calculus ($P = C_n$), the Principle of Inclusion and Exclusion ($P = B_n$), and the Möbius Inversion Theorem in number theory ($P = D_n$). A useful perspective will be to consider a certain algebra associated with P called the incidence algebra and which permits linear algebra techniques to be employed.

Our first step will be to generalize the Möbius function to a map having two arguments. Let P be a locally finite poset and let $\operatorname{Int}(P)$ be the set of closed intervals of P. Note that every $[x,z] \in \operatorname{Int}(P)$ has a minimum element; namely $\hat{0}_{[x,y]} = x$. The *Möbius function* of P is the map $\mu : \operatorname{Int}(P) \to \mathbb{Z}$ defined inductively on $[x,z]$ by

$$\mu(x,z) = \begin{cases} 1 & \text{if } x = z, \\ -\sum_{x\le y<z} \mu(x,y) & \text{otherwise.} \end{cases} \tag{5.11}$$

Note that $\mu(x,z)$ denotes the value of μ on the closed interval $[x,z]$ even though the square brackets have been dropped in the notation. Also note that this is essentially the same definition as (5.4). Indeed, as a poset $[x, z]$ has a minimum element $\hat{0} = x$ and

$$\mu(x, z) = \mu_{[x,z]}(z) \tag{5.12}$$

where the latter is the Möbius function of one variable. The readers may wish to convince themselves of this by computing $\mu(\{1, 3\}, \{1, 2, 3, 5, 6\})$ in the poset of Figure 5.3 and then comparing this with the computation done in the previous section for B_3. Just as in the one-variable case, we have the alternative definition

$$\sum_{x \le y \le z} \mu(x, y) = \delta_{x,z}. \tag{5.13}$$

It turns out that μ is only one of an important family of functions on $\operatorname{Int}(P)$. If P is a locally finite poset, then its *incidence algebra*, $\mathcal{I}(P)$, is the set of all functions $\phi : \operatorname{Int}(P) \to \mathbb{R}$ under the operations of *addition*

$$(\phi + \psi)(x, z) = \phi(x, z) + \psi(x, z),$$

scalar multiplication of $c \in \mathbb{R}$

$$(c \cdot \phi)(x, z) = c(\phi(x, z)),$$

and *convolution product*

$$(\phi * \psi)(x, z) = \sum_{x \le y \le z} \phi(x, y)\psi(y, z).$$

It will often be convenient to extend the domain of $\phi \in \mathcal{I}(P)$ to all of $P \times P$ by letting $\phi(x, z) = 0$ whenever $x \not\le z$. Using this convention the sum in the convolution product can take place over all $y \in P$. We will now show that the incidence algebra lives up to its name.

Theorem 5.5.1. *If P is a locally finite poset, then $(\mathcal{I}(P), +, \cdot, *)$ is an associative algebra over $\mathbb{R}$.*

Proof. We will prove the associative law for convolution, leaving the check of the other algebra axioms as an exercise. If $\phi, \psi, \omega \in \mathcal{I}(P)$ and $[x, z] \in \operatorname{Int}(P)$, then, using the fact that $\mathbb{R}$ itself is associative,

$$\begin{aligned}
((\phi * \psi) * \omega)(x, z) &= \sum_s (\phi * \psi)(x, s)\omega(s, z) \\
&= \sum_{r,s} (\phi(x, r)\psi(r, s))\omega(s, z) \\
&= \sum_{r,s} \phi(x, r)(\psi(r, s)\omega(s, z)) \\
&= \sum_r \phi(x, r)(\psi * \omega)(r, z) \\
&= (\phi * (\psi * \omega))(x, z)
\end{aligned}$$

as desired. □

We have already met one element of $\mathcal{I}(P)$, namely μ. But there are others which are important. Consider the analogue of the Kronecker delta, which is $\delta \in \mathcal{I}(P)$ defined by

$$\delta(x,z) = \begin{cases} 1 & \text{if } x = z, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $\delta(x,z) = \delta_{x,z}$.

Proposition 5.5.2. *The incidence algebra $\mathcal{I}(P)$ has identity element δ; that is, for any $\phi \in \mathcal{I}(P)$ we have*

$$\delta * \phi = \phi * \delta = \phi.$$

Proof. We will prove $\delta * \phi = \phi$ as the other equality is entirely analogous. Since δ is only nonzero when its two arguments are equal,

$$(\delta * \phi)(x,z) = \sum_y \delta(x,y)\phi(y,z) = \delta(x,x)\phi(x,z) = \phi(x,z)$$

as we wished to show. □

Another useful element of $\mathcal{I}(P)$ is the *zeta function* which satisfies $\zeta(x,z) = 1$ for all $[x,z] \in \operatorname{Int}(P)$. In Section 5.9 we will see how ζ is related to the Riemann zeta function. Recall that if A is an associative algebra with identity element e, then $a \in A$ has a (*multiplicative*) *inverse* if there is an element denoted a^{-1} such that both $a^{-1}a = e$ and $aa^{-1} = e$. If $n \in \mathbb{N}$, then the algebra of $n \times n$ matrices over $\mathbb{R}$ has the property that to prove the existence of a^{-1} it suffices to show that it satisfies $a^{-1}a = e$. As we will see shortly, there is a correspondence between incidence algebras of finite posets and matrix algebras which will show that the same implication holds for $\mathcal{I}(P)$. It turns out that ζ and μ are inverses in $\mathcal{I}(P)$.

Proposition 5.5.3. *We have*

$$\mu = \zeta^{-1}.$$

Proof. Using (5.13) and the definition of ζ we see that

$$(\mu * \zeta)(x,z) = \sum_{x \le y \le z} \mu(x,y)\zeta(y,z) = \sum_{x \le y \le z} \mu(x,y) \cdot 1 = \delta(x,z).$$

By the discussion preceding this proposition, this is enough to prove that $\mu = \zeta^{-1}$. □

By the discussion just before the previous proposition, we can conclude that $\zeta * \mu = \delta$. Evaluating this equality on an interval gives

$$\sum_{x \le y \le z} \zeta(x,y)\mu(y,z) = \delta_{x,z}$$

or

$$\sum_{x \le y \le z} \mu(y,z) = \delta_{x,z}. \tag{5.14}$$

This looks very much like (5.13) except that in one the first argument of μ is fixed while the second varies, while in the other the roles are reversed. So (5.14) could also be used to uniquely define the Möbius function except in a dual manner from the original. This equation can be used to calculate μ in a "top-down" fashion. It is not at all obvious a priori that this computation and the one proceeding "bottom-up" give the same value for $\mu(x, z)$, although they must.

One can make the incidence algebra more concrete by identifying it with an algebra of matrices. Let P be a finite poset. A *linear extension* of P is a permutation $L = x_1 x_2 \ldots x_n$ of the elements of P such that $x_i \leq_P x_j$ implies $i \leq j$; that is, x_i comes before x_j in the permutation. One can think of a linear extension as a listing of the elements of P which respects the partial order in that smaller elements must come before larger ones. For example, B_2 has two linear extensions, namely

$$\emptyset, \{1\}, \{2\}, \{1, 2\} \quad \text{and} \quad \emptyset, \{2\}, \{1\}, \{1, 2\}.$$

Given a linear extension L and $\phi \in \mathcal{I}(P)$, the *matrix of f with respect to L* is

$$M_\phi = (\phi(x_i, x_j))_{1 \leq i, j \leq n}$$

recalling that $\phi(x, y) = 0$ if $x \not\leq y$. Returning to our example and using the first linear extension above we have

$$M_\zeta = \begin{array}{c c} & \begin{array}{cccc} \emptyset & \{1\} & \{2\} & \{1,2\} \end{array} \\ \begin{array}{c} \emptyset \\ \{1\} \\ \{2\} \\ \{1,2\} \end{array} & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

where the elements of B_2 indexing the rows and columns are shown in the margins. For any linear extension L and any $\phi \in \mathcal{I}(P)$ the matrix M_ϕ must be upper triangular since if $i > j$, then $x_i \not\leq x_j$ and so $(M_\phi)_{i,j} = \phi(x_i, x_j) = 0$.

Theorem 5.5.4. *Let P be a finite poset and fix a linear extension $L = x_1 \ldots x_n$ of P. Then $\mathcal{I}(P)$ is isomorphic to the algebra of $n \times n$ real matrices M such that $M_{i,j} = 0$ if $x_i \not\leq x_j$.*

Proof. The map $\phi \mapsto M_\phi$ is a bijection between the two algebras since M_ϕ contains all the values of ϕ. It is easy to prove that this bijection preserves addition and scalar multiplication. For multiplication of algebra elements, we must show that $M_\phi M_\psi = M_{\phi * \psi}$ for any $\phi, \psi \in \mathcal{I}(P)$. To do this, it suffices to prove that these matrices have the same (i, j) entry for any i, j. But

$$\begin{aligned} (M_\phi M_\psi)_{i,j} &= \sum_{1 \leq k \leq n} (M_\phi)_{i,k} (M_\psi)_{k,j} \\ &= \sum_{x_k \in P} \phi(x_i, x_k) \psi(x_k, x_j) \\ &= (\phi * \psi)(x_i, x_j) \\ &= (M_{\phi * \psi})_{i,j} \end{aligned}$$

as desired. □

We can now prove the Möbius Inversion Theorem as well as its dual version. Note that each of the four conditions are required to hold for all $x \in P$, not just for a specific element.

Theorem 5.5.5 (Möbius Inversion Theorem). *Let P be a finite poset, let V be a real vector space, and let $f, g : P \to V$ be two functions.*

(a) *We have*

$$f(x) = \sum_{y \geq x} g(y) \text{ for all } x \in P \iff g(x) = \sum_{y \geq x} \mu(x,y) f(y) \text{ for all } x \in P.$$

(b) *We have*

$$f(x) = \sum_{y \leq x} g(y) \text{ for all } x \in P \iff g(x) = \sum_{y \leq x} \mu(y,x) f(y) \text{ for all } x \in P.$$

Proof. We will prove (a), leaving (b) as an exercise. In fact, we will give two proofs of (a), one working directly with the elements of $\mathcal{I}(P)$ and one using linear algebra.

Let us assume that $f(x) = \sum_{y \geq x} g(y)$ for all $x \in P$. Plugging this into summation involving μ and using (5.13) yields

$$\begin{aligned}
\sum_{y \geq x} \mu(x,y) f(y) &= \sum_{y \geq x} \mu(x,y) \sum_{z \geq y} g(z) \\
&= \sum_{z \geq x} g(z) \sum_{x \leq y \leq z} \mu(x,y) \\
&= \sum_{z \geq x} g(z) \delta_{x,z} \\
&= g(x).
\end{aligned}$$

The proof of the reverse implication follows the same strategy and so is safely left to the reader.

For the linear algebraic proof, we will fix a linear extension $L = x_1 \dots x_n$ of P. Then any $f : P \to V$ has associated column vector

$$v_f = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}.$$

Note that the first condition in (a) can be written $f(x) = \sum_{y \in P} \zeta(x,y) g(y)$ since

$$\zeta(x,y) = \begin{cases} 1 & \text{if } x \leq y, \\ 0 & \text{otherwise.} \end{cases}$$

The summation for f says that the entry in row x of v_f is the same as the entry in row x of the product $M_\zeta v_g$. And since this must hold for all x, the first condition in (a) is equivalent to the matrix equation $v_f = M_\zeta v_g$. But by Theorem 5.5.4, M_ζ has an inverse which is M_μ. So $v_g = M_\mu v_f$. This is equivalent to the summation condition for g since we have, taking the entry in row x on both sides,

$$g(x) = \sum_{y \in P} \mu(x,y) f(y) = \sum_{y \geq x} \mu(x,y) f(y)$$

for any x. $\square$

This theorem is useful when the function f is easy to compute, but one is really interested in g. If the two maps under consideration are related by an appropriate summation condition, then we can express g in terms of f by inversion. We will now give several examples, starting with the ones mentioned in the first paragraph of this section.

Our first application will be to the theory of finite differences, which is a discrete analogue of the calculus. A function $f : \mathbb{N} \to \mathbb{R}$ has as (*forward*) *difference* the function $\Delta f : \mathbb{N} \to \mathbb{R}$ defined by

$$\Delta f(n) = f(n+1) - f(n).$$

This corresponds to differentiation. Indeed, the derivative of $f : \mathbb{R} \to \mathbb{R}$ is

$$f'(x) = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon}$$

and at $\epsilon = 1$ the function inside the limit is just $f(x+1) - f(x)$. For example, if $f(n) = n^2$, then $\Delta f(n) = (n+1)^2 - n^2 = 2n + 1$ which bears a strong resemblance to $(x^2)' = 2x$. There is also a version of the definite integral in this context. The *definite summation* of $f : \mathbb{N} \to \mathbb{R}$ is the function $Sf : \mathbb{N} \to \mathbb{R}$ where

$$Sf(n) = \sum_{i=0}^{n} f(i).$$

The analogue of the Fundamental Theorem of Calculus is as follows. It will be convenient to extend the domain of any $f : \mathbb{N} \to \mathbb{R}$ to $\mathbb{Z}$ by letting $f(i) = 0$ for $i < 0$.

Theorem 5.5.6 (Fundamental Theorem of Difference Calculus)**.** *Given two function $f, g : \mathbb{N} \to \mathbb{R}$, we have*

$$f(n) = Sg(n) \text{ for all } n \geq 0 \iff g(n) = \Delta f(n-1) \text{ for all } n \geq 0.$$

Proof. It is easy to compute that in the chain C_n we have

$$\mu(i,n) = \begin{cases} 1 & \text{if } i = n, \\ -1 & \text{if } i = n-1, \\ 0 & \text{otherwise.} \end{cases}$$

Now for all $n \geq 0$, the first condition in the theorem can be translated as

$$f(n) = Sg(n) = \sum_{i=0}^{n} g(i) = \sum_{i \leq n} g(i)$$

where the inequality indexing the last summation is taking place in C_n. Using Theorem 5.5.5(b) and the Möbius values in C_n above, this is equivalent to

$$g(n) = \sum_{i \leq n} \mu(i,n) f(i) = (1)f(n) + (-1)f(n-1) = \Delta f(n-1)$$

for all $n \geq 0$. □

It turns out that the Principle of Inclusion and Exclusion is just the Möbius Inversion Theorem applied to the poset B_n. We restate it here for ease of reference.

Theorem 5.5.7. *Given a finite set S and subsets $S_1, \dots, S_n$, we have*

$$\left|S - \bigcup_{i=1}^{n} S_i\right| = |S| - \sum_{1 \le i \le n} |S_i| + \sum_{1 \le i < j \le n} |S_i \cap S_j| - \cdots + (-1)^n \left|\bigcap_{i=1}^{n} S_i\right|.$$

Proof. Define two functions $f, g : B_n \to \mathbb{N}$ by

$$f(I) = \left|\bigcap_{i \in I} S_i\right|$$

and

$$g(I) = \left|\bigcap_{i \in I} S_i - \bigcup_{j \notin I} S_j\right|.$$

In words, $f(I)$ counts the number of elements of S which are in all the S_i for $i \in I$ and possibly in other S_j. On the other hand, $g(I)$ is the number of elements which are in exactly the S_i for $i \in I$ and no others. From this description we see that, for all $I \in B_n$,

$$f(I) = \sum_{J \supseteq I} g(J)$$

since any element in the S_i for $i \in I$ must be in exactly the S_j for the elements j of some $J \supseteq I$. Applying Theorem 5.5.5(a) together with Proposition 5.1.3(d) and (5.6) we obtain

$$g(I) = \sum_{J \supseteq I} \mu(I, J) f(J) = \sum_{J \supseteq I} (-1)^{|J-I|} \left|\bigcap_{j \in J} S_j\right|.$$

Specializing to the case $I = \emptyset$ we obtain

$$\left|S - \bigcup_{i=1}^{n} S_i\right| = g(\emptyset) = \sum_{J \in B_n} (-1)^{|J|} \left|\bigcap_{j \in J} S_j\right|$$

which is what we wished to prove. □

The Möbius Inversion Theorem originated in number theory. Here is that version.

Theorem 5.5.8. *Given two functions $f, g : \mathbb{P} \to \mathbb{R}$, we have*

$$f(n) = \sum_{d|n} g(d) \text{ for all } n \in \mathbb{P} \iff g(n) = \sum_{d|n} \mu(d) f(n/d) \text{ for all } n \in \mathbb{P}.$$

Proof. The first condition is an exact translation of the first condition in Theorem 5.5.5(b) for the poset D_n. Inverting using Theorem 5.1.3(f) as well as the relationship (5.12) between the one- and two-variable forms of μ gives

$$g(n) = \sum_{d|n} \mu(d, n) f(d) = \sum_{d|n} \mu(n/d) f(d) = \sum_{d'|n} \mu(d') f(n/d')$$

where $d' = n/d$. □

5.6. Characteristic polynomials

As was made abundantly clear in Chapter 3, one way to get insight into a combinatorial object is to study its generating function. This is also true of the Möbius function and the corresponding generating function is called the characteristic polynomial. In particular, we will show that using this polynomial one can get an interesting connection between a particular lattice associated to a graph and its chromatic polynomial.

Let P be a finite ranked poset with $\operatorname{rk} P = n$. The *characteristic polynomial* of P is

$$\chi(P) = \chi(P;t) = \sum_{x \in P} \mu(x) t^{n - \operatorname{rk} x}. \tag{5.15}$$

where we are using the one variable form of the Möbius function. We also used χ for the chromatic number of a graph, but this should cause no confusion since here we are dealing with posets. The quantity $n - \operatorname{rk} x$ appearing in the power on t is called the *corank* of x and the reader may be wondering why we are using this rather than just the rank. One reason is that this makes $\chi(P)$ monic: the highest power of t appears when $x = \hat{0}$ and $\mu(\hat{0}) = 1$. Also, as will be seen, this choice of exponent results in $\chi(P)$ having some interesting properties. Note that collecting terms in (5.15) for x at the same rank shows that

$$\chi(P) = \sum_{k=0}^{n} w_k(P) t^{n-k} \tag{5.16}$$

where the $w_k(P)$ are the Whitney numbers of the first kind for P.

Let us begin by computing the characteristic polynomials for some of our standard example posets.

Proposition 5.6.1. *We have the following characteristic polynomials:*

(a) *For C_n,*
$$\chi(C_n) = t^{n-1}(t-1).$$

(b) *For B_n,*
$$\chi(B_n) = (t-1)^n.$$

(c) *If n has prime factorization $n = p_1^{m_1} \cdots p_k^{m_k}$ and $m = \sum_i m_i$, then*
$$\chi(D_n) = t^{m-k}(t-1)^k.$$

(d) *For Π_n,*
$$\chi(\Pi_n) = (t-1)(t-2)\cdots(t-n+1).$$

(e) *For $L_n(q)$,*
$$\chi(L_n(q)) = (t-1)(t-q)(t-q^2)\cdots(t-q^{n-1}).$$

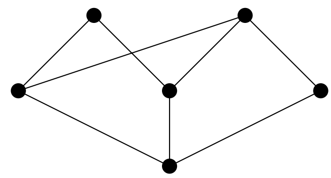
$P =$ 

Figure 5.7. A poset P with $\chi(P)$ having complex roots

Proof. We will prove the results for B_n, leaving the others as exercises.

In the case of B_n, one can plug (5.8) into (5.16) and then use reindexing, the symmetry of the binomial coefficients, as well as the Binomial Theorem to obtain

$$\begin{aligned}\chi(B_n) &= \sum_{k=0}^{n} (-1)^k \binom{n}{k} t^{n-k} \\ &= \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{n-k} t^k \\ &= (-1)^n \sum_{k=0}^{n} \binom{n}{k} (-t)^k \\ &= (-1)^n (1-t)^n \\ &= (t-1)^n\end{aligned}$$

which is the desired conclusion. □

It is striking that all the characteristic polynomials in Proposition 5.6.1 have nonnegative integer roots. This is not always the case. For example, consider the poset in Figure 5.7. Then an easy computation gives $\chi(P) = t^2 - 3t + 3$ which has complex roots. We also note that if all the roots of a polynomial are nonnegative reals, then the coefficient sequence is log-concave. However, we will postpone the proof of this until we have introduced the elementary symmetric functions in Section 7.1.

One way to explain some of the factorizations in Proposition 5.6.1 is via the following result.

Theorem 5.6.2. *Let P, Q be finite ranked posets.*

(a) *If there is an isomorphism $f : P \to Q$, then $\chi(P) = \chi(Q)$.*

(b) *We have*

$$\chi(P \times Q) = \chi(P)\chi(Q).$$

Proof. (a) From Exercise 8(b) at the end of the chapter we have $\operatorname{rk}_P x = \operatorname{rk}_Q f(x)$ for all $x \in P$. In particular, $\operatorname{rk} P = \operatorname{rk} Q = n$ for some n. Thus, using Theorem 5.4.3 and the fact that f is a bijection

$$\begin{aligned}\chi(P) &= \sum_{x\in P} \mu_P(x)t^{n-\mathrm{rk}_P x}\\ &= \sum_{x\in P} \mu_P(f(x))t^{n-\mathrm{rk}_P f(x)}\\ &= \sum_{y\in Q} \mu_Q(y)t^{n-\mathrm{rk}_Q y}\\ &= \chi(Q).\end{aligned}$$

(b) The result of Exercise 8(c) is that $\mathrm{rk}_{P\times Q}(x,y) = \mathrm{rk}_P x + \mathrm{rk}_Q y$ for all $(x,y)\in P\times Q$. So if $\mathrm{rk}\, P = m$ and $\mathrm{rk}\, Q = n$, then $\mathrm{rk}(P\times Q) = m+n$. Applying Theorem 5.4.4

$$\begin{aligned}\chi(P\times Q) &= \sum_{(x,y)\in P\times Q} \mu_{P\times Q}(x,y)t^{m+n-\mathrm{rk}_{P\times Q}(x,y)}\\ &= \sum_{x\in P} \mu_P(x)t^{m-\mathrm{rk}_P x} \sum_{y\in Q} \mu_Q(y)t^{n-\mathrm{rk}_Q y}\\ &= \chi(P)\chi(Q)\end{aligned}$$

which is what we wished to prove. □

This theorem can be used to explain a couple of the factorizations in Proposition 5.6.1. For example $B_n \cong C_1^n$ so

$$\chi(B_n) = \chi(C_1^n) = \chi(C_1)^n = (t-1)^n.$$

However, neither Π_n nor $L_n(q)$ decompose as a product of smaller posets. So to understand the factorization of their characteristic polynomials, we will have to use poset quotients as discussed in the next section.

We end this section by making a connection between the characteristic polynomial of a lattice associated with a graph G and the chromatic polynomial of G. A subgraph H of G is *induced* if $vw \in E(G)$ implies $vw \in E(H)$ for all $v, w \in V(H)$. In words, every

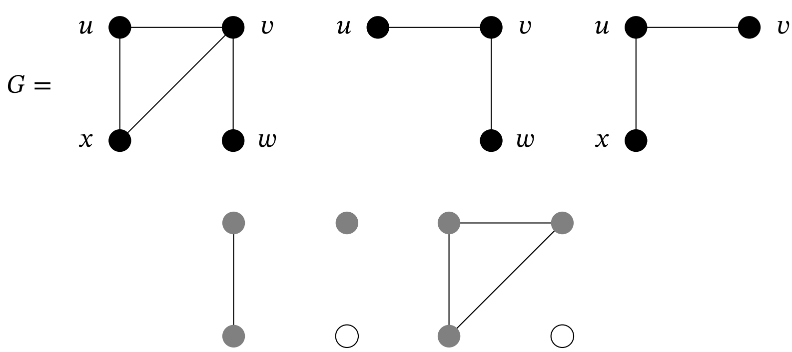


Figure 5.8. A graph G and two subgraphs on the top, as well as two spanning subgraph colorings on the bottom

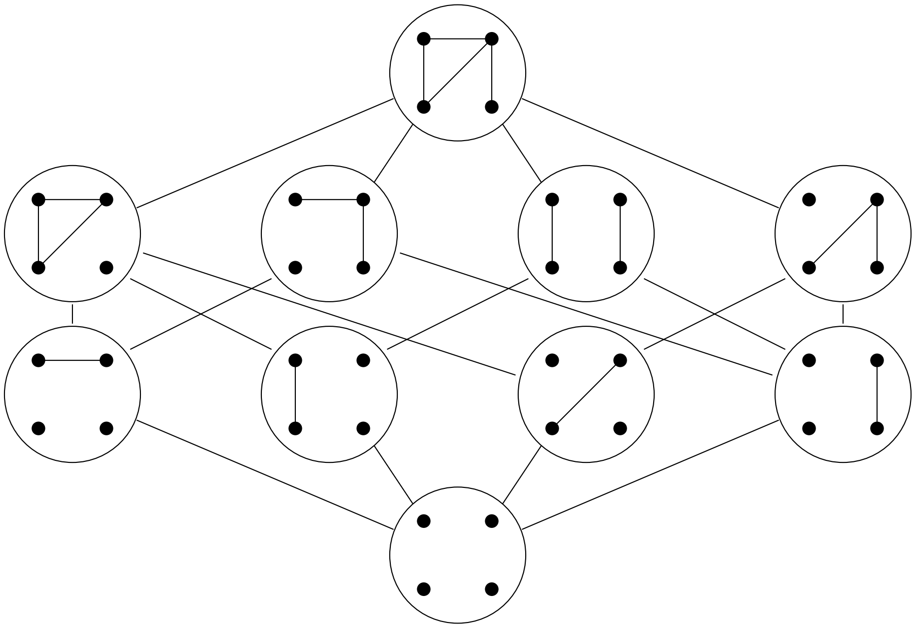


Figure 5.9. The bond lattice of the graph G in Figure 5.8

edge of G between vertices of H must be in H. By way of illustration, given the graph G in Figure 5.8, the first subgraph in that figure is induced but the second is not because it is missing the edge $vx \in E(G)$. A *bond* of G is a spanning subgraph such that each component is induced. The *bond lattice* of G, denoted $\mathcal{L}(G)$, is the set of bonds of G ordered by containment. The bond lattice for the graph G of Figure 5.8 is displayed in Figure 5.9. It is not hard to show that $\mathcal{L}(G)$ is a ranked lattice with rank function

$$\operatorname{rk} H = n - k(H) \tag{5.17}$$

where n is the number of vertices of G and $k(H)$ is the number of components of H.

We need a couple of other definitions before we can connect bond lattices with chromatic polynomials. Let $c : V \to S$ be a (not necessarily proper) coloring of G. If H is a spanning subgraph of G, then we say that c is *H-improper* if every component of H is monochromatic, that is, has all vertices of the same color. The two rightmost graphs in Figure 5.8 are two subgraphs which are both H-improper for the same coloring c. The *subgraph induced by c* is the spanning subgraph of G such that $uv \in E(H)$ if and only if $uv \in E(G)$ and $c(u) = c(v)$. Of the two subgraphs colored white and gray in Figure 5.8, only the second is induced by the coloring of G. We note that the subgraph induced by c is unique and is a bond. Furthermore it follows directly from the definitions that c is proper if and only if its induced subgraph is the spanning subgraph with no edges.

Theorem 5.6.3. *Let G be a graph with $k(G) = k$ components. We have*

$$P(G) = t^k \chi(\mathcal{L}(G)).$$

Proof. Let G have $\#V = n$ and consider all possible colorings $c\colon V \to [t]$ for $t \in \mathbb{N}$. Given a bond H of G, we will be interested in two associated subsets:

$$S(H) = \{c \mid c \text{ is } H\text{-improper}\}$$

and

$$T(H) = \{c \mid H \text{ is induced by } c\}.$$

Note that if c is H-improper, then there is a unique bond $K \supseteq H$ such that K is the bond induced by c. It follows that for all $H \in \mathcal{L}(G)$ we have

$$S(H) = \biguplus_{K \supseteq H} T(K).$$

Now define $f, g\colon \mathcal{L}(G) \to \mathbb{N}$ by $f(H) = \#S(H)$ and $g(H) = \#T(H)$. Note that $f(H) = t^{k(H)}$ since there are t ways to choose the color of each component of H. Also, the previous displayed equation yields $f(H) = \sum_{K \supseteq H} g(H)$ for all $H \in \mathcal{L}(G)$. Applying the Möbius Inversion Theorem gives

$$g(H) = \sum_{K \supseteq H} \mu(H, K) f(K).$$

Recall that c is proper if and only if it induces the spanning subgraph of G with no edges. And this subgraph is the $\hat{0}$ element of $\mathcal{L}(G)$. Thus, using in turn the formula for $f(H)$ in terms of $k(H)$, the rank function of $\mathcal{L}(G)$ as given in (5.17), and the fact that $P(G)$ counts proper colorings,

$$\begin{aligned}
t^k \chi(\mathcal{L}(G)) &= t^{k(G)} \sum_{K \in \mathcal{L}(G)} \mu(K) t^{(n-k(G))-(n-k(K))} \\
&= \sum_{K \in \mathcal{L}(G)} \mu(K) t^{k(K)} \\
&= g(\hat{0}) \\
&= P(G)
\end{aligned}$$

which is what we wanted. □

5.7. Quotients of posets

In many areas of mathematics one studies the objects under consideration by taking quotients which can have a simpler structure than the original entity. In this and the next section we will present a concept of quotient for posets P. We will see that it is useful for proving that the characteristic polynomial of P factors even though P may not be a product of smaller posets. This notion can also be used to give inductive proofs of various well-known theorems about μ. Quotients of the type we will consider first appeared in the work of Hallam and Sagan [**41**].

A number of techniques have been proposed for proving that the characteristic polynomial factors over the integers. See [**78**] for a survey. The method we will use proceeds as follows. We wish to show that a poset Q has characteristic polynomial which factors as $\chi(Q) = \prod_i \chi_i$ for certain polynomials χ_i. Suppose we can construct posets P_i with $\chi(P_i) = \chi_i$ for all i and let $P = \bigtimes_i P_i$. We wish to find an equivalence relation $\sim$ on P and a partial order on the set of equivalence classes $P/\sim$ such that

(i) $(P/\sim) \cong Q$ and

(ii) $\chi(P/\sim) = \chi(P)$.

From this and Theorem 5.6.2 we get

$$\chi(Q) = \chi(P/\sim) = \chi(P) = \chi(\times_i P_i) = \prod_i \chi_i$$

as we wished to show.

Since we will be particularly interested in the case where $\chi(Q)$ has integer roots, we introduce a simple family of posets with this property. Suppose P has a $\hat{0}$. Then the elements covering $\hat{0}$ are the *atoms* of P. We will use the notation

$$\mathcal{A}(P) = \{x \in P \mid x \text{ is an atom of } P\}.$$

If P is ranked, then the atoms are just the elements of rank one. Define the *n-claw*, CL_n, to be the poset consisting solely of a $\hat{0}$ and n atoms. For example, in Figure 5.10 the two posets in the direct product are CL_1 and CL_2. Clearly

$$\chi(CL_n; t) = t - n.$$

As our running example, we will use the partition lattice Π_3 in Figure 5.1 and the product poset on the right in Figure 5.10. Note that

$$\chi(CL_1 \times CL_2) = \chi(CL_1)\chi(CL_2) = (t-1)(t-2) = \chi(\Pi_3).$$

Unfortunately, $\Pi_3 \ncong CL_1 \times CL_2$. But they are close to being isomorphic. In particular, if we could merge the two maximal elements of $CL_1 \times CL_2$ into one, then the resulting poset would be a copy of Π_3. Poset quotients are designed to make this sort of identification of elements precise.

Let P be a finite poset with a $\hat{0}$ and let $\sim$ be an equivalence relation on P. Define the *quotient* $P/\sim$ to be the set of equivalence classes together with the binary relation $X \le Y$ in $P/\sim$ if and only if $x \le_P y$ for some $x \in X$ and $y \in Y$. It is important to note that this binary relation is *not* necessarily a partial order. To see what can go wrong, consider the chain C_2 with equivalence classes $X = \{0, 2\}$ and $Y = \{1\}$. Then we have $X \le Y$ since $0 \le 1$. But we also have $Y \le X$ since $1 \le 2$. And clearly $X \ne Y$, so the

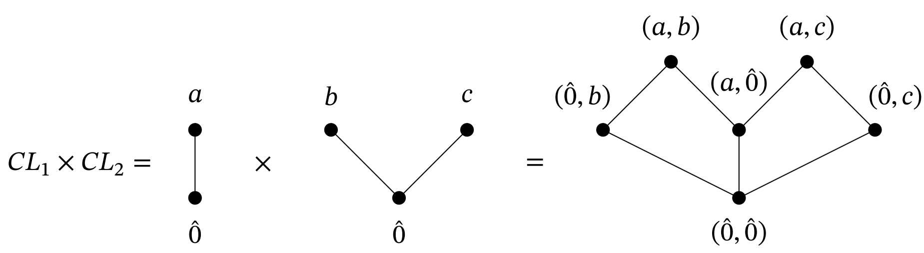


Figure 5.10. A product of claws

antisymmetry relation fails. To fix this, define, for P a finite poset with $\hat{0}$, $P/\sim$ to be a *homogeneous quotient* if

(1) the equivalence class containing $\hat{0}$ is $\{\hat{0}\}$ and

(2) if $X \le Y$, then for any $x \in X$ there is a $y \in Y$ with $x \le_P y$.

Homogeneous quotients yield posets.

Lemma 5.7.1. *If $P/\sim$ is a homogeneous quotient, then $P/\sim$ is a poset.*

Proof. Verifying the reflexive and transitive laws is easy and so it is left as an exercise. For antisymmetry, suppose $X \le Y$ and $Y \le X$. Let x be a maximal element of X. Then since $X \le Y$ and the quotient is homogeneous, there is $y \in Y$ with $x \le y$. Similarly $Y \le X$ implies there is $x' \in X$ with $y \le x'$. So $x \le y \le x'$. But x was picked to be maximal in X which forces $x = x'$. This in turn yields $y = x$. Since $x \in X$ and $y \in Y$ we have found an element of $X \cap Y$. Equivalence classes are either disjoint or equal, so this implies $X = Y$ as we wished to prove. □

Returning to the product $P = CL_1 \times CL_2$ in Figure 5.10, we impose the equivalence relation $\sim$ where every element is in an equivalence class by itself except for the two maximal elements which are in a class together. It is easy to see that this is a homogeneous quotient since any time we have $X < Y$, the class X is a singleton. Furthermore $(P/\sim) \cong \Pi_3$ which is our desired condition (i). As far as (ii), one can verify by direct computation that χ does not change in passing from P to $P/\sim$. In fact, more is true. Note that the equivalence class $\{(a,b),(a,c)\}$ of P becomes the $\hat{1}$ of $P/\sim$. Furthermore

$$\mu_P(a,b) + \mu_P(a,c) = 1 + 1 = 2 = \mu_{P/\sim}(\hat{1}).$$

Our next order of business is to give a condition under which this always happens.

Lemma 5.7.2. *Let $P/\sim$ be a homogeneous quotient. Suppose that for all nonzero $X \in P/\sim$ we have*

$$\sum_{y \in I(X)} \mu(y) = 0 \tag{5.18}$$

where $I(X)$ is the lower-order ideal generated by X as a subset of P. Then for all $X \in P/\sim$ we have

$$\mu(X) = \sum_{x \in X} \mu(x).$$

Proof. We will induct on the length of the longest $\hat{0}$–X chain in $P/\sim$. When the length is zero we have, by the first requirement for a homogeneous quotient, $X = \{\hat{0}_P\}$ and $\mu(X) = 1 = \mu(\hat{0}_P)$.

For a nonzero X we have, by induction,

$$\mu(X) = -\sum_{Y<X} \mu(Y) = -\sum_{Y<X} \sum_{y \in Y} \mu(y).$$

We claim that $\{y \in Y \mid Y < X\} = I(X) - X$. Indeed, $y \in I(X) - X$ means that $y \notin X$ and $y < x$ for some $x \in X$. And by the second condition for a homogeneous quotient, this is equivalent to being in $\{y \in Y \mid Y < X\}$. So the previous displayed equation becomes

$$\mu(X) = -\sum_{y\in I(X)-X} \mu(y) = \sum_{x\in X} \mu(x)$$

where the second equality comes from solving for the terms when $x \in X$ in (5.18). This completes the proof. $\square$

We will call (5.18) the *Summation Condition*. We also need to know how the rank function behaves when taking a quotient of a ranked poset. This is taken care of by the next result, where (5.19) will be called the *Rank Condition*.

Lemma 5.7.3. *Let $P/\sim$ be a homogeneous quotient of a ranked poset. Suppose that for all $x, y \in P$*

$$x \sim y \implies \operatorname{rk} x = \operatorname{rk} y. \tag{5.19}$$

Then $P/\sim$ is ranked and $\operatorname{rk} X = \operatorname{rk} x$ for all $x \in X$.

Proof. First we claim that if we have a cover $X \lessdot Y$, then for any $x \in X$ there is $y \in Y$ with $x \lessdot y$. We know that we can pick y with $x < y$. Suppose, towards a contradiction, that there is a z with $x < z < y$. Letting Z be the equivalence class of z, we must have $X < Z < Y$ where the inequalities are strict because (5.19) forces an equivalence class to consist of elements at a given rank. But this contradicts $X \lessdot Y$.

To show that $P/\sim$ is ranked, suppose we have two saturated chains $\hat{0} = X_0 \lessdot X_1 \lessdot \cdots \lessdot X_m$ and $\hat{0} = Y_0 \lessdot Y_1 \lessdot \cdots \lessdot Y_n$ where $X_m = Y_n$. Then, by the claim, we have corresponding chains $\hat{0} = x_0 \lessdot x_1 \lessdot \cdots \lessdot x_m$ and $\hat{0} = y_0 \lessdot y_1 \lessdot \cdots \lessdot y_n$. Since x_m and y_n are in the same equivalence class, they must have the same rank by (5.19). This forces $m = n$ and that $P/\sim$ must be ranked. This also shows that $\operatorname{rk} X = \operatorname{rk} x$ for all $x \in X$. $\square$

It is now a short step to our desired conclusion.

Theorem 5.7.4. *Let $P/\sim$ be a homogeneous quotient of a ranked poset satisfying the Summation Condition* (5.18) *and Rank Condition* (5.19)*. Then*

$$\chi(P/\sim) = \chi(P).$$

Proof. Using the previous two lemmas gives

$$\chi(P/\sim) = \sum_{X\in P/\sim} \mu(X)t^{\operatorname{rk}(P/\sim)-\operatorname{rk}(X)} = \sum_{x\in P} \mu(x)t^{\operatorname{rk}(P)-\operatorname{rk}(x)} = \chi(P)$$

as desired. $\square$

As an application, we will use this theorem to calculate $\chi(\Pi_n)$. Before doing so, we will return to the case of Π_3 to motivate the equivalence relation used. We will label the atoms of the two claws in Figure 5.10 with atoms of Π_3 as follows. A block of a partition of $[n]$ is *trivial* if it contains a single element. If $1 \le i < j \le n$, then let ij denote the atom of Π_n whose only nontrivial block is $\{i, j\}$. Now consider the claws as labeled in Figure 5.11. Finally, replace each pair labeling an element of the product with the label which is the join in Π_n of its two components to obtain the poset on the

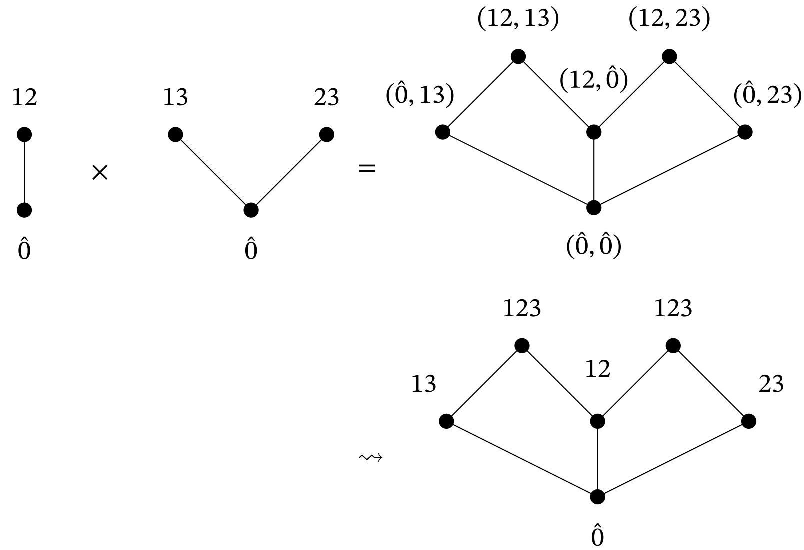


Figure 5.11. A product of claws with partition labels

right in the figure. Note that two elements of the product are in the same equivalence class of $P/\sim$ if and only if they have the same label in the right-hand poset. This is the key to defining the equivalence relation.

Now consider Π_n for any n. It will be convenient to make the convention just for this proof that Π_n is the poset of partitions of $\{0, 1, \ldots, n-1\}$. We let

$$P = CL_1 \times CL_2 \times \cdots \times CL_{n-1}$$

where for $1 \leq j \leq n-1$ the atoms of CL_j are labeled with the atoms ij, $i < j$, of Π_n. Clearly

$$\chi(P) = (t-1)(t-2)\ldots(t-n+1).$$

Put an equivalence relation on elements $x = (x_1, \ldots, x_{n-1}) \in P$ by $x \sim y$ if and only if $\bigvee x = \bigvee y$ in Π_n. So for an equivalence class X, we can define the partition $\rho(X) = \bigvee x$ for any $x \in X$. To finish, we need to show that $P/\sim$ is a homogeneous quotient satisfying the Summation and Rank Conditions and that $(P/\sim) \cong \Pi_n$. We claim that $X \leq Y$ in the binary relation on $P/\sim$ if and only if $\rho(X) \leq \rho(Y)$ in Π_n. Indeed, take $x \in X$ and $y \in Y$ with $x \leq_P y$. Then $x_i \leq y_i$ in Π_n for all i. It follows that

$$\rho(X) = \bigvee x \leq \bigvee y = \rho(Y). \tag{5.20}$$

The proof of the backwards direction involves similar ideas and so is left as an exercise

For homogeneity, clearly $(\hat{0}, \ldots, \hat{0})$ is the only element whose join is $\hat{0}$ which gives the first condition in the definition. For the second condition, we use the assumption $X \leq Y$ and (5.20) to induct on the length l of a saturated chain in the interval $[\rho(X), \rho(Y)]$. If $l = 0$, then $X = Y$ and the conclusion is trivial. So assume $X < Y$, $x \in X$. Take two blocks B, C of $\bigvee x = \rho(X)$ which are contained in the same block of $\rho(Y)$. Let $i = \min B$ and $j = \min C$ where, without loss of generality, $i < j$. We claim that coordinate x_j of x must be $\hat{0}$. For suppose that $x_j = kj$ for some $k < j$. But then $k, j \in C$ contradicting $j = \min C$. Now define z to agree with x except in the jth coordinate where $z_j = ij$ so that $x < z$. By construction, $\bigvee x < \bigvee z \leq \rho(Y)$. So if Z is the equivalence class of z, then $X < Z \leq Y$ by the previous claim. And by induction, there is $y \in Y$ with $z \leq y$ so that $x < z \leq y$ as desired.

To obtain the Summation Condition, it is easy to see from Theorem 5.4.4 that for any $x \in P$ we have

$$\mu(x) = (-1)^{\operatorname{supp} x}$$

where $\operatorname{supp} x$ is the cardinality of the *support set* of x

$$\operatorname{Supp} x = \{i \mid x_i \neq \hat{0}\}. \tag{5.21}$$

So to get (5.18) it suffices to find a sign-reversing involution on $I(X)$ which has no fixed points, where the sign of x is $\mu(x)$. Since $X \neq \hat{0}$, the partition $\rho = \rho(X)$ has a nontrivial block B, say $B = \{i < j < \cdots\}$. Take $y \in I(x)$. We claim that $y_j = \hat{0}$ or $y_j = ij$. For suppose $y_j = kj$ for some $k \neq i$. But then $\bigvee y \not\leq \rho$ since the block containing j in $\bigvee y$ contains $k < j$ and so cannot be B. Now define an involution $\iota\colon I(X) \to (X)$ so that $\iota(y)$ is y with y_j either changed from $\hat{0}$ to ij or from ij to $\hat{0}$. It is easy to check that this map has the desired properties.

For the Rank Condition, note that if $\rho, a \in \Pi_n$ where $a = ij$ is an atom, then either $\rho \vee a = \rho$ if i and j are in the same block of ρ, or $\rho \vee a$ covers ρ if i and j are in different blocks of ρ since those two blocks get merged in the join. So given $x \in P$, let $i_1j_1, \ldots, i_kj_k$ be the elements of $\operatorname{Supp} x$ listed so that $j_1 < \cdots < j_k$ and consider $\bigvee x = i_1j_1 \vee \cdots \vee i_kj_k$. Let $\rho_l = i_1j_1 \vee \cdots \vee i_lj_l$ for any $0 \leq l \leq k$. For $l \geq 1$ we see that ρ_l must cover ρ_{l-1} since the ordering of the i's and j's forces $j_l > i_1, j_1, \ldots, i_l$ so that $\{j_l\}$ is a trivial block of ρ_{l-1}. Thus $\operatorname{rk} x = k = n - |\rho|$. Hence if $x \sim y$, then $\bigvee x = \rho = \bigvee y$ and so $\operatorname{rk} x = n - |\rho| = \operatorname{rk} y$ which is what we wished to prove.

Finally, we need an isomorphism between $(P/\sim)$ and Π_n. The function we have already defined, $\rho(X) = \bigvee x$ for any $x \in X$, is such a map. We leave the proof that this is a well-defined isomorphism to the reader. This completes the proof that $\chi(\Pi_n)$ has the desired form.

One can motivate the choice of atoms used to label the claws in the Π_n example as follows. Consider the maximal chain in Π_n which is $\hat{0} = \rho_1 \lessdot \rho_2 \lessdot \cdots \lessdot \rho_n = \hat{1}$ where, for all j, ρ_j is the partition with a block $[j]$ and all other blocks trivial. Considering the set of atoms a such that $a \leq \rho_j$ but $a \not\leq \rho_{j-1}$, we see that these are exactly the atoms used to label the claw CL_j. This technique of partitioning the atom set of a poset has been used before in proving that the characteristic polynomial factors over $\mathbb{Z}$. See, for example, Stanley's work on supersolvable lattices [**83**].

To end this section, we note that there is a quicker method for proving these factorizations when the poset is a finite lattice L. Let A_L and A_x be the set of atoms of L and the set of atoms of L less than or equal to $x \in L$, respectively. Say that a $\hat{0}$–$\hat{1}$ chain $C : \hat{0} = x_0 < x_1 < \cdots < x_n = \hat{1}$ *induces* a partition $A_1/\ldots/A_n$ of A_L by defining

$$A_i = \{a \in A_L \mid a \leq x_i \text{ but } a \not\leq x_{i-1}\}.$$

For Π_n, this is exactly the partition described in the previous paragraph for the given saturated chain. Let CL_{A_i} be the claw whose atoms are labeled by A_i and consider $\mathbf{t} = (x_1, \ldots, x_n) \in CL_{A_1} \times \cdots \times CL_{A_n}$. The vector $\mathbf{t}$ is called a *transversal* and it has *support* defined by equation (5.21) but in this more general setting. For $x \in L$, define

$$\mathcal{T}_x = \left\{ \mathbf{t} = (x_1, \ldots, x_n) \;\middle|\; \bigvee_i x_i = x \right\}$$

where the join is taken in L.

Theorem 5.7.5. *Let L be a finite ranked lattice and let $A_1/\ldots/A_n$ be induced by a $\hat{0}$–$\hat{1}$ chain. Suppose that for all $x \in L$ and $\mathbf{t} \in \mathcal{T}_x$ we have*

$$|\operatorname{Supp} \mathbf{t}| = \operatorname{rk} x.$$

Then the following are equivalent.

(1) *For every nonzero $x \in L$ there is an index i such that $|A_i \cap A_x| = 1$.*

(2) *The polynomial $\chi(L;t)$ factors with nonegative integral roots as*

$$\chi(L;t) = t^{\operatorname{rk} L - n} \prod_{i=1}^{n} (t - |A_i|). \qquad \square$$

Note that to check factorization using the previous theorem there are only two conditions to check. And these are usually rather easy to verify. So this is much more efficient than using the previous method. Theorem 5.7.5 is also the only one we know of which gives conditions equivalent to (rather than just implying) factorization of χ over the nonnegative integers. Proving this result would consume too much of our time, but the reader can find details in [**41**].

5.8. Computing the Möbius function

We will now use quotient posets to prove three classic theorems about the Möbius function. Each of these results gives a different way to compute μ. One of the advantages of using quotients is that all three can be proven inductively using a lemma about a very simple equivalence relation. The proofs in this section are based on the arXiv version of a paper of Hallam [**38, 39**].

Let P be a poset with a $\hat{1}$. A *coatom* of P is an element c covered by $\hat{1}$. We wish to examine what happens when a coatom and $\hat{1}$ are identified by an equivalence relation. If $x \in P$, then we use $[x]$ for the equivalence class of x.

Lemma 5.8.1. *Let P be a finite poset with a $\hat{0}$ and a $\hat{1}$ such that $\#P \geq 3$. Let c be a coatom and let $\sim$ be the equivalence relation with classes $\{c, \hat{1}\}$ and all others having only one element. In this case, the following hold.*

(a) *$P/\sim$ is homogeneous.*

(b) *We have*

$$\mu([\hat{1}]) = \mu(c) + \mu(\hat{1}).$$

(c) *If P is a lattice, then so is $P/\sim$ with $[x] \vee [y] = [x \vee y]$ for all $x, y \in P$ and $[x] \wedge [y] = [x \wedge y]$ for all $[x], [y] \neq [\hat{1}]$.*

Proof. (a) Since $\#P \geq 3$, the definition of $\sim$ shows that $[\hat{0}] = \{\hat{0}\}$ which is the first condition for a homogeneous quotient. For the second, if $X = Y$, then the conclusion is trivial. And if $X < Y$, then $X = \{x\}$ for some x. So $x \leq y$ for some $y \in Y$ by the definition of the binary relation.

(b) It suffices to show that the Summation Condition (5.18) holds, so assume $X \neq \{\hat{0}\}$. We either have $X = \{x\}$ or $X = [x]$ where $x = \hat{1}$. In either case

$$\sum_{y \in I(X)} \mu(y) = \sum_{y \leq x} \mu(y) = 0$$

by the definition of the Möbius function in P.

(c) It is not hard to show that $(P/\sim) \cong P - \{c\}$ and we will use the latter description. Consider three cases:

(1) one or both of x, y is equal to c or
(2) $x \vee y \in \{c, \hat{1}\}$ with $x, y \neq c$ or
(3) $x \vee y \in P - \{c, \hat{1}\}$ with $x, y \neq c$.

We will do the second one and leave the others to the reader. If $x \vee y \in \{c, \hat{1}\}$ with $x, y \neq c$, then in $P - \{c\}$ we have that x, y have a unique upper bound, namely $\hat{1}$. It follows that $[x] \vee [y]$ exists and

$$[x] \vee [y] = [\hat{1}] = [c] = [x \vee y]$$

as desired. Checking the meets is similar. □

We will now demonstrate a theorem of Hall [**37**] which gives an interesting relationship between the Möbius function of a poset and its chains.

Theorem 5.8.2. *Let P be a finite poset with a $\hat{0}$ and a $\hat{1}$. We have*

$$\mu(\hat{1}) = \sum_{i \geq 0} (-1)^i c_i \tag{5.22}$$

where c_i is the number of $\hat{0}$–$\hat{1}$ chains of length i in P.

Proof. We will induct on $\#P$. The result is easy to verify if $\#P \leq 2$ so assume $\#P \geq 3$. Let c be a coatom of P and let $\sim$ be the equivalence relation of Lemma 5.8.1. Let a_i, respectively b_i, be the number of $\hat{0}$–$\hat{1}$ chains of P of length i which do not, respectively do, contain c. Clearly

$$\sum_{i \geq 0} (-1)^i c_i = \sum_{i \geq 0} (-1)^i a_i + \sum_{i \geq 0} (-1)^i b_i. \tag{5.23}$$

There is a length-preserving bijection between the $\hat{0}$–$\hat{1}$ chains of P which do not contain c and the $[\hat{0}]$–$[\hat{1}]$ chains of $P/\sim$ which sends x to $[x]$ for each x in the chain. Since $|P/\sim| < |P|$ we have, by induction,

$$\mu([\hat{1}]) = \sum_{i \geq 0} (-1)^i a_i. \tag{5.24}$$

There is also a bijection between $\hat{0}$–$\hat{1}$ chains of P containing c and $\hat{0}$–c chains of $[\hat{0}, c]$ gotten by removing $\hat{1}$ from such chains in P. Since this bijection changes length by one we have, again by induction,

$$\mu(c) = -\sum_{i\geq 0}(-1)^i b_i. \tag{5.25}$$

Plugging (5.24) and (5.25) into (5.23) and using Lemma 5.8.1 gives

$$\sum_{i\geq 0}(-1)^i c_i = \mu([\hat{1}]) - \mu(c) = \mu(\hat{1})$$

as desired. □

The reader familiar with algebraic topology may have noticed that the sum in equation (5.22) looks like a reduced Euler characteristic. This can be made precise as follows. Let P be a poset with a $\hat{0}$ and a $\hat{1}$ which are distinct. The set of $\hat{0}$–$\hat{1}$ chains in P are in bijection with the set of all chains in $\overline{P} \coloneqq P - \{\hat{0}, \hat{1}\}$: merely remove the $\hat{0}$ and $\hat{1}$ from each chain of P. So define the *order complex of* P, $\Delta(P)$, to be the set of all chains in the open interval $(\hat{0}, \hat{1})$. This is clearly an (abstract) simplicial complex since a subset of a chain is still a chain. Since the lengths of a $\hat{0}$–$\hat{1}$ chain in P and of its image in $\Delta(P)$ differ by two, the right-hand side of (5.22) is the reduced Euler characteristic $\tilde{\chi}(\Delta(P))$. So one can bring the tools of algebraic topology to bear on questions about the Möbius function. For more information on this approach, see the survey articles of Björner [**15**] and Wachs [**96**].

The next result is due to Weisner [**98**]. It gives an expression for μ similar to the one given in its definition but with what could be a substantially smaller number of terms. Recall from Proposition 5.3.2(i) that a finite lattice has a $\hat{0}$ and a $\hat{1}$.

Theorem 5.8.3. *If L is a finite lattice and $a \in L - \{\hat{0}\}$, then*

$$\mu(\hat{1}) = -\sum_{\substack{x\neq\hat{1}\\ x\vee a=\hat{1}}} \mu(x). \tag{5.26}$$

Proof. We induct on $\#L$, just doing the induction step when $\#L \geq 3$. When $a = \hat{1}$, the sum in (5.26) is over all $x < \hat{1}$. So the equation is true by definition of $\mu(\hat{1})$. Now assume $a \neq \hat{1}$ and pick a coatom c with $a \leq c$. Let $\sim$ be the equivalence relation from Lemma 5.8.1. By induction we have

$$\mu([\hat{1}]) = -\sum_{\substack{[x]\neq[\hat{1}]\\ [x]\vee[a]=[\hat{1}]}} \mu([x]).$$

Now $[\hat{1}] = \{c, \hat{1}\}$, $[x] \vee [a] = [x \vee a]$ by Lemma 5.8.1, and $\mu([x]) = \mu(x)$ for $[x] \neq [\hat{1}]$. So the previous displayed equation becomes

$$\mu([\hat{1}]) = -\sum_{\substack{x\neq c,\hat{1}\\ x\vee a=c,\hat{1}}} \mu(x) = -\sum_{\substack{x\neq c,\hat{1}\\ x\vee a=c}} \mu(x) - \sum_{\substack{x\neq c,\hat{1}\\ x\vee a=\hat{1}}} \mu(x).$$

If $x \vee a = c$, then clearly $x \neq \hat{1}$. And if $x \vee a = \hat{1}$, then $x \neq c$ since $a \leq c$. So we can write

$$\mu([\hat{1}]) = -\sum_{\substack{x\neq c\\ x\vee a=c}} \mu(x) - \sum_{\substack{x\neq\hat{1}\\ x\vee a=\hat{1}}} \mu(x).$$

If $x \vee a = c$, then $x \leq c$. So the first sum above can be viewed as taking place in $[\hat{0}, c]$ which has fewer elements than P. By induction

$$\mu([\hat{1}]) = \mu(c) - \sum_{\substack{x \neq \hat{1} \\ x \vee a = \hat{1}}} \mu(x).$$

Rearranging terms and using Lemma 5.8.1 finishes the proof. □

To illustrate how this result can be used to easily compute μ, consider B_n for $n \geq 2$. Consider the atom $a = \{n\}$. To satisfy $x \cup a = [n]$ where $x \neq [n]$ we must have $x = [n-1]$. So using (5.26) and induction gives

$$\mu(B_n) = -\mu([n-1]) = -\mu(B_{n-1}) = -(-1)^{n-1} = (-1)^n.$$

We end this section with a theorem of Rota [**76**]. To state it, we need a new definition. Let P be a finite poset with a $\hat{0}$ and a $\hat{1}$. A *crosscut* of P is $K \subset P$ with the following properties.

(1) $\hat{0}, \hat{1} \notin K$.

(2) K is an antichain.

(3) Every maximal chain of P intersects K.

For example, if P is ranked, then $\mathrm{Rk}_k(P)$ is a crosscut for $0 < k < \operatorname{rk} P$.

Theorem 5.8.4 (Crosscut Theorem)**.** *Let L be a finite lattice and let K be a crosscut. Then*

$$\mu(\hat{1}) = \sum_{\substack{\bigvee B = \hat{1} \\ \bigwedge B = \hat{0}}} (-1)^{\#B}$$

where the sum is over all $B \subseteq K$ satisfying the meet and join conditions.

Proof. First consider the case where every atom of L is also a coatom. This forces $K = L - \{\hat{0}, \hat{1}\}$. And the meet and join conditions are satisfied for all $B \subseteq K$ with $\#B \geq 2$. If $\#K = n$, then, using Theorem 1.3.3(d),

$$\sum_{\substack{\bigvee B = \hat{1} \\ \bigwedge B = \hat{0}}} (-1)^{\#B} = \sum_{k=2}^{n} \binom{n}{k} (-1)^k = -\sum_{k=0}^{1} \binom{n}{k} (-1)^k = n - 1 = \mu(\hat{1}).$$

Now assume that the atom and coatom sets of L do not coincide. By Exercises 12(c) and 23(c), the theorem holds for L if and only if it holds for L^*. So, by taking the dual if necessary, we can assume that there is a coatom $c \notin K$. We now induct on $\#L$. The previous paragraph takes care of the case $\#L = 3$ and smaller lattices do not have crosscuts. Suppose that $\#L \geq 4$ and let $\sim$ be the equivalence relation of Lemma 5.8.1. Since $c, \hat{1} \notin K$, the $[x]$ for $x \in K$ form a crosscut for $P/\sim$ which we will denote by $[K]$ and its subsets by $[B]$. By induction,

$$\mu([\hat{1}]) = \sum_{\substack{\bigvee [B] = [\hat{1}] \\ \bigwedge [B] = [\hat{0}]}} (-1)^{\#[B]}.$$

By Lemma 5.8.1 we have $\bigvee[B] = [\hat{1}]$ if and only if $\bigvee B = c$ or $\hat{1}$. By the same token and the fact that $c, \hat{1} \notin K$ we get that $\bigwedge[B] = [\hat{0}]$ is equivalent to $\bigwedge B = \hat{0}$. So the previous displayed equation becomes

$$\mu([\hat{1}]) = \sum_{\substack{\bigvee B = c \\ \bigwedge B = \hat{0}}} (-1)^{\#B} + \sum_{\substack{\bigvee B = \hat{1} \\ \bigwedge B = \hat{0}}} (-1)^{\#B}.$$

Note that since K is a crosscut of L not containing c we have that $K' := K \cap [0, c]$ is a crosscut of $[0, c]$. Furthermore, $\bigvee B = c$ implies that $B \subseteq K'$. So applying induction to the sum with this restriction gives

$$\mu([\hat{1}]) = \mu(c) + \sum_{\substack{\bigvee B = \hat{1} \\ \bigwedge B = \hat{0}}} (-1)^{\#B}.$$

A rearrangement of terms and Lemma 5.8.1 completes the demonstration. □

As an application of this result, consider B_n, $n \geq 2$, with the crosscut K consisting of its atoms. But for $B \subseteq K$ we can only have $\bigvee B = [n]$ if $B = K$. And in this case $\bigwedge B = \emptyset$. So $\mu(B_n) = (-1)^{\#K} = (-1)^n$.

5.9. Binomial posets

Binomial posets were introduced by Doubilet, Rota, and Stanley [**23**] and further studied in [**87**]. They provide an explanation about why certain types of generating functions arise in practice while others do not. For example, we have already met ordinary generating functions $\sum_n a_n t^n$ and exponential generating functions $\sum_n a_n t^n / n!$. There are also *Eulerian generating functions* which are of the form $\sum_n a_n t^n / [n]_q!$. We have seen one example of this in the q-Binomial Theorem where the right-hand side of equation (3.6) can be written

$$\sum_k q^{\binom{k}{2}} \left[\begin{array}{c} n \\ k \end{array} \right]_q t^k = \sum_k \left(q^{\binom{k}{2}} [n][n-1] \cdots [n-k+1] \right) \frac{t^k}{[k]!}.$$

Why do such generating functions appear while others, say of the form $\sum_n a_n t^n / C(n)$ where $C(n)$ is the nth Catalan number, do not? Binomial posets provide one possible explanation.

A poset P is called *binomial* if it satisfies the following conditions:

- BP1 P is locally finite and contains arbitrarily long chains.
- BP2 Every interval $[x, z]$ is ranked. The interval is called an *n-interval* if, considered as a poset, $\operatorname{rk}[x, z] = n$.
- BP3 Any two n-intervals contain the same number of maximal chains. This number is denoted $F(n) = F_P(n)$ and called the *factorial function* of P.

We will consider three examples corresponding to the three types of generating functions mentioned at the beginning of this section. Let C_∞ be the nonnegative integers under the usual total order. We also have B_∞ which consists of all finite subsets of positive integers partially ordered by set containment. Finally, let V_∞ be the vector space over $\mathbb{F}_q$ with countable basis $e_1, e_2, \ldots$ and denote by $L_\infty(q)$ the poset of all finite-dimensional subspaces of V_∞ with containment as the partial order.

Proposition 5.9.1. *The posets C_∞, B_∞, and $L_\infty(q)$ are all binomial. Their factorial functions are*

$$F_{C_\infty}(n) = 1, \quad F_{B_\infty}(n) = n!\,, \quad \textit{and} \quad F_{L_\infty(q)}(n) = [n]_q!\,.$$

Proof. We will prove this for B_∞ and leave the other two cases as exercises. If $[S,T]$ is an interval in B_∞, then $[S,T] \cong B_n$ for some n. So P is locally finite with ranked intervals. Subchains of the infinite chain $\emptyset \subset \{1\} \subset \{1,2\} \subset \cdots$ can be arbitrarily long. Since any two n-intervals are isomorphic to B_n, they contain the same number of maximal chains. To find the factorial function, note that a maximal chain in B_n has the form

$$\emptyset \subset \{s_1\} \subset \{s_1, s_2\} \subset \cdots \subset [n].$$

There are n choices for s_1, and after that $n-1$ for s_2, etc. So the total number of chains is $n!$. □

An important property of binomial posets is that the number of elements at a given rank in an n-interval $[x,z]$ does not depend on x, z.

Lemma 5.9.2. *If $[x,z]$ is an n-interval in a binomial poset P and $0 \le k \le n$, then*

$$\#\operatorname{Rk}_k[x,z] = \frac{F(n)}{F(k)F(n-k)}.$$

Proof. Given $y \in \operatorname{Rk}_k[x,z]$, we first count the number of maximal chains C of $[x,z]$ passing through y. Such a C must be the concatenation of a maximal chain in $[x,y]$ with a maximal chain in $[y,z]$. Since $[x,y]$ is a k-interval and $[y,z]$ is an $(n-k)$-interval, the number of C must be $F(k)F(n-k)$. But this expression is independent of y. So the total number of maximal chains in $[x,z]$ is $F(k)F(n-k)\cdot\#\operatorname{Rk}_k[x,z]$. Since $[x,z]$ is an n-interval, this number is also $F(n)$. Setting the two expressions equal and solving for $\#\operatorname{Rk}_k[x,z]$ completes the proof. □

To make the connection between binomial posets P and generating functions, we must consider a subalgebra of the incidence algebra $\mathcal{I}(P)$. The *reduced incidence algebra* of a binomial poset P is

$$\mathcal{R}(P) = \{\phi \in \mathcal{I}(P) \mid \phi \text{ is constant on } n\text{-intervals}\}.$$

Equivalently the $\phi \in \mathcal{R}(P)$ are precisely those such that $\phi(x,z) = \phi(x',z')$ whenever $[x,z]$ and $[x',z']$ are both n-intervals. We let $\phi(n)$ denote this common value. So, for example, $\zeta \in \mathcal{R}(P)$ since $\zeta(x,z) = 1$ on all intervals $[x,z]$. It is not clear a priori that $\mu \in \mathcal{R}(P)$. But this will follow from the next result.

Theorem 5.9.3. *Let P be a binomial poset.*

(a) *$\mathcal{R}(P)$ is a subalgebra of $\mathcal{I}(P)$.*

(b) *If $\phi \in \mathcal{R}(P)$ and ϕ^{-1} exists in $\mathcal{I}(P)$, then $\phi^{-1} \in \mathcal{R}(P)$.*

Proof. (a) We need to show that $\mathcal{R}(P)$ is closed under addition, scalar multiplication, and convolution. We will prove the last and leave the other two as exercises. Suppose $\phi, \psi \in \mathcal{R}(P)$. Using the previous lemma, we can write

$$\begin{aligned}(\phi * \psi)(x,z) &= \sum_{x \le y \le z} \phi(x,y)\psi(y,z)\\ &= \sum_{k=0}^{n} \sum_{y \in \mathrm{Rk}_k[x,z]} \phi(x,y)\psi(y,z)\\ &= \sum_{k=0}^{n} \sum_{y \in \mathrm{Rk}_k[x,z]} \phi(k)\psi(n-k)\\ &= \sum_{k=0}^{n} \frac{F(n)}{F(k)F(n-k)}\phi(k)\psi(n-k).\end{aligned}$$

But this last expression is clearly independent of x, z and so we are done.

(b) We must show that if $[x,z]$ is an n-interval, then $\phi^{-1}(x,z)$ depends only on n. We will induct on n. If $n = 0$, then $x = z$ and $\phi * \phi^{-1}(x,x) = \delta(x,x) = 1$. So, using the fact that $\phi \in \mathcal{R}(P)$,

$$1 = \sum_{x \le y \le x} \phi(x,y)\phi^{-1}(y,x) = \phi(x,x)\phi^{-1}(x,x) = \phi(0)\phi^{-1}(x,x).$$

This can be rewritten $\phi^{-1}(x,x) = 1/\phi(0)$ and the right-hand side does not depend on x as desired.

Now suppose $n > 0$. Similar to the base case we have

$$(\phi * \phi^{-1})(x,z) = \delta(x,z) = 0.$$

Expanding the convolution and moving the first term outside the sum gives

$$0 = \phi(x,x)\phi^{-1}(x,z) + \sum_{x < y \le z} \phi(x,y)\phi^{-1}(y,z) = \phi(0)\phi^{-1}(x,z) + S$$

where S is the sum in this displayed equation. But, by induction, S only depends on n. So $\phi^{-1}(x,z) = -S/\phi(0)$ is also solely a function of n as desired. □

We can now draw a concrete relation between binomial posets and generating functions.

Theorem 5.9.4. *If P is binomial, then $\mathcal{R}(P) \cong \mathbb{R}[[t]]$ as algebras via the map*

$$\phi \mapsto F_\phi(t) := \sum_{n \ge 0} \phi(n)\frac{t^n}{F(n)}.$$

Proof. This function is a bijection since it has an inverse. In particular, if $F(t) \in \mathbb{R}[[t]]$, then we can write $F(t) = \sum_n a_n t^n / F(n)$ for some $a_n \in \mathbb{R}$. So the inverse maps $F(t)$ to $\phi \in \mathcal{R}(P)$ defined by $\phi(n) = a_n$.

Showing that the bijection preserves addition and scalar multiplication is left as an exercise. For convolution we want $F_{\phi*\psi}(t) = F_\phi(t)F_\psi(t)$. Using the expression for $(\phi * \psi)(x,z)$ obtained in the proof of the previous theorem

$$\begin{aligned}
F_\phi(x)F_\psi(t) &= \sum_{n\ge 0} \phi(n)\frac{t^n}{F(n)} \cdot \sum_{n\ge 0} \psi(n)\frac{t^n}{F(n)} \\
&= \sum_{n\ge 0}\left(\sum_{k=0}^{n} \frac{\phi(k)}{F(k)} \cdot \frac{\psi(n-k)}{F(n-k)}\right)t^n \\
&= \sum_{n\ge 0}\left(\sum_{k=0}^{n} \frac{F(n)}{F(k)F(n-k)}\phi(k)\psi(n-k)\right)\frac{t^n}{F(n)} \\
&= \sum_{n\ge 0}(\phi * \psi)(n)\frac{t^n}{F(n)} \\
&= F_{\phi*\psi}(t)
\end{aligned}$$

as we wished to conclude. □

Note that in C_∞ this map becomes $\phi \mapsto \sum_n \phi(n)t^n$ which is an ordinary generating function. Similarly, the images in B_∞ and $L_\infty(q)$ are exponential and Eulerian generating functions, respectively.

There is one of our important initial example posets which does not seem to be covered by this theory. Consider D_∞ which is the positive integers ordered by divisibility. Then D_∞ is not binomial. For consider the intervals $[1,4]$ and $[1,6]$. We have $\mathrm{rk}[1,4] = \mathrm{rk}[1,6] = 2$. But $[1,4]$ contains a single maximal chain whereas $[1,6]$ has two. But there is a way around this difficulty. Let $P_1, P_2, P_3, \ldots$ be posets each with a $\hat{0}$. We will use subscripts to indicate which poset an element belongs to, for example $\hat{0}_i$ is the $\hat{0}$ in P_i. The *direct sum* of the P_i has as underlying set

$$\bigoplus_{i\ge 1} P_i = \{(x_1, x_2, \ldots) \mid x_i \in P_i \text{ for all } i,\ x_i \ne \hat{0}_i \text{ for only finitely many } i\}.$$

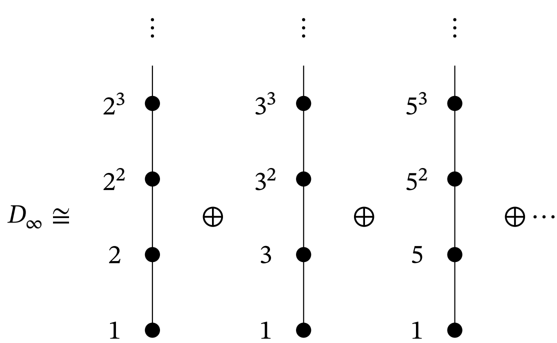


Figure 5.12. D_∞ as a direct sum of chains.

Letting $(x_i) = (x_1, x_2, \ldots)$, we impose the partial order $(x_i) \le (y_i)$ if and only if $x_i \le y_i$ for all i. It is not hard to see that D_∞ is isomorphic to the direct sum of chains as illustrated in Figure 5.12.

A *Dirichlet poset* is $P = \bigoplus_{i\ge1} P_i$ where each P_i is a binomial poset with a $\hat{0}$. So D_∞ is Dirichlet. An interval $[(x_i),(z_i)]$ in P is called an (n_i)*-interval* where $(n_i) = (n_1, n_2, \ldots)$ if $[x_i, z_i]$ is an n_i-interval in P_i for all i. The corresponding *reduced incidence algebra* is

$$\mathcal{R}(P) = \{\phi \in \mathcal{I}(P) \mid \phi \text{ is constant on } (n_i)\text{-intervals}\}.$$

The notation $\phi((n_i))$ should be self-explanatory. Note that $\zeta \in \mathcal{R}(P)$ for P Dirichlet. Theorem 5.9.3 remains true if P is replaced by a Dirichlet poset. Theorem 5.9.4 generalizes as follows, where

$$\mathbf{t} = \{t_1, t_2, \ldots\}$$

is a countably infinite sequence of variables.

Theorem 5.9.5. *If $P = \bigoplus_{i\ge1} P_i$ is Dirichlet, then $\mathcal{R}(P) \cong \mathbb{R}[[\mathbf{t}]]$ as algebras via the map*

$$\phi \mapsto F_\phi(\mathbf{t}) := \sum_{(n_i)} \phi((n_i)) \prod_{i\ge1} \frac{t_i^{n_i}}{F(n_i)}$$

where the sum is over all (n_i) with only finitely many $n_i \ne 0$, and $F(n_i)$ is the factorial function of P_i. □

Now suppose $P = D_\infty$ so $F(n_i) = 1$ for all i. Let $t_i = 1/p_i^s$ where p_i is the ith prime and $s \in \mathbb{C}$. Using the unique factorization of the integers we see that

$$F_\zeta(\mathbf{t}) = \sum_{(n_i)} \zeta((n_i)) \prod_{i\ge1} t_i^{n_i} = \sum_{(n_i)} \prod_{i\ge1} \frac{1}{(p_i^s)^{n_i}} = \sum_{(n_i)} \frac{1}{\left(\prod_{i\ge1} p_i^{n_i}\right)^s} = \sum_{n\ge1} \frac{1}{n^s}$$

where the last sum is the Riemann ζ-function $\zeta(s) = \sum_{n\ge1} 1/n^s$. For the rest of this section s will be a complex number so that $\zeta(s)$ will always refer to Riemann's function rather than the value of the reduced incidence algebra element of the same name on an n-interval. As a function of a complex variable, one can show that $\zeta(s)$ has zeros at the negative even integers which are sometimes called its *trivial zeros*. Perhaps the most famous conjecture in all of mathematics is the following by Riemann [**74**].

Conjecture 5.9.6 (Riemann Hypothesis)**.** *All the nontrivial zeros of $\zeta(s)$ have real part* 1/2.

One can restate the Riemann Hypothesis in terms of the Möbius function μ of D_∞. To do so, we need a concept from asymptotic combinatorics. Suppose we have two functions $f, g\colon \mathbb{N} \to \mathbb{R}$. We say that *$f$ is big oh of g*, written $f = O(g)$, if there are constants C, N such that $|f(n)| \le C|g(n)|$ for all $n \ge N$. To see why this might be relevant to Conjecture 5.9.6, note that if $f(z) = 1/q(z)$ is a rational function of $z \in \mathbb{C}$, then the zeros of the polynomial $q(z)$ are called the *poles* of $f(z)$. These poles control the growth rate of the coefficients of the (not formal) power series expansion $f(z) = \sum_{n\ge0} a_n z^n$. For example, if $f(z) = 1/(1-2z)$, then $f(z)$ has a pole at $r = 1/2$. And we also know that for sufficiently small $|z|$ we have $f(z) = \sum_n 2^n z^n$ whose coefficients grow like, in fact are exactly, $2^n = (1/r)^n$.

It is not hard to show using Theorem 5.9.5 that

$$\frac{1}{\zeta(s)} = \sum_{n\geq 1} \frac{\mu(n)}{n^s} \tag{5.27}$$

where μ is taken in D_∞. Given the relation between roots and poles just discussed, it makes sense to consider the *Mertens function*

$$M(n) = \sum_{1\leq k\leq n} \mu(k).$$

Then a conjecture equivalent to Conjecture 5.9.6 is as follows.

Conjecture 5.9.7. *For all real* $\epsilon > 0$ *we have*

$$M(n) = O(n^{1/2+\epsilon}).$$

There was an earlier conjecture of Mertens [**62**] that $|M(n)| \leq n^{1/2}$ for all n. But this was disproved by Odlyzko and te Riele [**66**].

Exercises

(1) This exercise refers to the list of examples just after the definition of a poset.
   (a) Verify that they satisfy the definition of a poset.
   (b) Show that the partial order in Π_n is equivalent to defining $\rho \leq \tau$ if every block of τ is a union of blocks of ρ.
   (c) Describe the cover relations in the list. For example, in C_n the covers are of the form $i < i+1$ for $0 \leq i < n$.

(2) Prove Proposition 5.1.1.

(3) Complete the proof of Proposition 5.1.2. For part (c) give two proofs: one by mimicking the proof of part (b) and one using P^*.

(4) Complete the proof of Proposition 5.1.3. To show that $K_n \cong B_{n-1}$ it may be simpler to show that $K_n \cong B_{n-1}^*$ using the map ϕ from Section 1.7.

(5) Let $f : P \to Q$ be an isomorphism of posets.
   (a) Show that f is also an isomorphism of P^* with Q^*.
   (b) Show that if P has a $\hat{0}$, then so does Q.
   (c) Show in two ways that if P has a $\hat{1}$, then so does Q: by mimicking the proof of part (b) and by using the result of (b) together with part (a).

(6) (a) Show that the axioms for a partially ordered set are satisfied by $P \uplus Q$, $P + Q$, and $P \times Q$.
   (b) Show that $P \times Q \cong Q \times P$.

(7) Complete the proof of Proposition 5.2.1.

(8) (a) Show that if P is a ranked poset, then for any k we have $\mathrm{Rk}_k(P)$ is an antichain.
(b) Let P be a ranked poset and assume $f : P \to Q$ is an isomorphism. Show that Q is also ranked and for all $x \in P$ we have
$$\mathrm{rk}_P\, x = \mathrm{rk}_Q\, f(x).$$
(c) Show that if P, Q are ranked posets, then so is $P \times Q$ with rank function
$$\mathrm{rk}_{P\times Q}(x, y) = \mathrm{rk}_P\, x + \mathrm{rk}_Q\, y.$$
(9) Prove Proposition 5.2.2.
(10) (a) Prove Proposition 5.3.1
(b) Find, with proof, a description of the meet and join operations in K_n.
(c) Show that $\mathfrak{S}$ is not a lattice.
(11) Fill in the proof of Proposition 5.3.2.
(12) (a) Show that if L is a lattice, then so is its dual L^* with
$$x \wedge_{L^*} y = x \vee_L y \quad \text{and} \quad x \vee_{L^*} y = x \wedge_L y.$$
(b) Show that if L, M are lattices, then so is $L \times M$ with
$$(a, x) \wedge (b, y) = (a \wedge b, x \wedge y) \quad \text{and} \quad (a, x) \vee (b, y) = (a \vee b, x \vee y).$$
(c) A *meet semilattice* is a poset P where every pair of elements has a meet. Show that if a finite meet semilattice has a $\hat{1}$, then it is a lattice. Hint: Given $x, y \in L$, let U be the set of upper bounds of x, y. Show that $\bigwedge U$ exists and is their join.
(d) Let P be a finite poset with a $\hat{0}$ and a $\hat{1}$. Suppose that for any $x, y \in P$ which both cover an element z, the join $x \vee y$ exists. Prove that P is a lattice. Hint: Use the previous part in its dual form and induct on $\#P$.
(13) A set partition $\pi = B_1/\ldots/B_k$ of $[nd]$ is *d-divisible* if d divides $|B_i|$ for all i. Let
$$\Pi_{nd,d} = \{\pi \mid \pi \text{ is a } d\text{-divisible partition of } [nd] \text{ or } \pi = 1/2/\ldots/nd\}$$
be partially ordered by refinement of set partitions.
(a) Show that if $\pi, \sigma \in \Pi_{nd,d}$, then $\pi \vee \sigma$ exists and is the same as the join in the ordinary partition lattice Π_{nd}.
(b) Show that $\Pi_{nd,d}$ is a lattice but that $\pi \wedge \sigma$ may not be the same in $\Pi_{nd,d}$ and in the ordinary partition lattice Π_{nd}. Hint: Use the dual form of Exercise 12(c).
(14) Prove the backward direction of Proposition 5.3.3.
(15) Prove Proposition 5.3.4.
(16) Finish the proof of Proposition 5.3.5.
(17) Let P be a finite poset and let $L = \mathcal{J}(P)$ be the corresponding distributive lattice. If $X \subseteq P$ is a lower-order ideal, then use the corresponding lowercase letter x to denote the associated element of L.
(a) Show that x covers y in L if and only if $Y = X - \{m\}$ where m is a maximal element of X.
(b) Show that x is join irreducible in L if and only if X is a principal ideal of P.
(18) (a) Given a poset P, let $\mathcal{A}(P)$ be the set of antichains of P. Show that the map $f : \mathcal{A}(P) \to \mathcal{J}(P)$ given by $f(A) = I(A)$ (where $I(A)$ is the order ideal generated by A) is a bijection.

(b) Show that $\lambda \in Y$ is join irreducible if and only if $\lambda = (k^l)$ for some $k, l \in \mathbb{P}$.
(c) Show that $C_n \cong \mathcal{J}(C_{n-1})$, $B_n \cong \mathcal{J}(A_n)$, and $D_n \cong \bigtimes_i \mathcal{J}(C_{n_i-1})$ where $n = \prod_i p_i^{n_i}$ is the prime factorization of n.
(d) Show that if L, M are finite, disjoint lattices, then

$$\operatorname{Irr}(L \times M) \cong \operatorname{Irr}(L) \uplus \operatorname{Irr}(M).$$

(e) Show that if P, Q are finite, disjoint posets, then

$$\mathcal{J}(P \uplus Q) \cong \mathcal{J}(P) \times \mathcal{J}(Q).$$

Use this to rederive the statements about B_n and D_n in part (c). For D_n you will also need part (d).

(19) (a) Rederive the formula for μ in B_n, equation (5.6), in two ways: by mimicking the proof of (5.7) and by constructing an $m \in \mathbb{P}$ such that $D_m \cong B_n$ and then applying (5.7).
(b) Prove that if $W \in L(\mathbb{F}_q^n)$ has dimension k, then

$$\mu(W) = (-1)^k q^{\binom{k}{2}}.$$

Hint: Use the q-Binomial Theorem, Theorem 3.2.4.

(20) (a) Let P be a locally finite poset with a $\hat{0}$. Show that if x covers exactly one element of P, then

$$\mu(x) = \begin{cases} -1 & \text{if } x \text{ covers } \hat{0}, \\ 0 & \text{otherwise.} \end{cases}$$

(b) Given any $n \in \mathbb{Z}$, construct a poset containing an element x with $\mu(x) = n$.

(21) Complete the proof of Theorem 5.5.1.

(22) Complete the proof of Theorem 5.5.4.

(23) Throughout this exercise, P is a locally finite poset.
(a) Show that $\zeta^2(x, z) = \#[x, z]$ where $\zeta^2 = \zeta * \zeta$.
(b) Show in two ways that f has an inverse if and only if $f(x, x) \neq 0$ for all $x \in P$: by working directly with elements of $\mathcal{I}(P)$ and by using linear algebra.
(c) Prove that $\mu_P(x, y) = \mu_{P^*}(y, x)$ where, as usual, P^* is the dual of P.
(d) Give three proofs of Theorem 5.5.5(b): by working directly with elements of $\mathcal{I}(P)$, by using linear algebra, and by using part (c) of this exercise.

(24) (a) Recall that the Euler phi function $\phi : \mathbb{P} \to \mathbb{P}$ from Exercise 6 in Chapter 2 is defined by

$$\phi(n) = \#\{m \in [n] \mid \gcd(m, n) = 1\}.$$

Use a counting argument to prove that for all $n \in \mathbb{P}$ we have

$$n = \sum_{d \mid n} \phi(d).$$

Hint: Consider the fractions $1/n, 2/n, \ldots, n/n$. Reduce each fraction to lowest terms and show that there will be exactly $\phi(d)$ fractions with denominator d for each $d \mid n$.

(b) Show that for all $n \in \mathbb{P}$ we have

$$\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

where the product is over all primes p dividing n. Hint: Use part (a) of this exercise.

(25) Prove that if P is a finite ranked poset with a $\hat{0}$ and a $\hat{1}$ which are distinct, then $\chi(P)$ has $t-1$ as a factor.

(26) Prove the formula for $\chi(D_n)$ in Proposition 5.6.1 in two ways: using the definition of χ and using Theorem 5.6.2.

(27) Finish the proof of Proposition 5.6.1. Hint: For Π_n, remember that $\operatorname{rk}\Pi_n = n-1$ and use Theorem 3.6.1. For $L_n(q)$ use Theorem 3.2.4.

(28) (a) Let G be a graph with n vertices. Show that $\mathcal{L}(G)$ is a ranked lattice with rank function given by (5.17).
(b) Show that the subgraph induced by a coloring of G is unique and is a bond.
(c) Show that if K_n is the complete graph on n vertices, then

$$\mathcal{L}(K_n) \cong \Pi_n.$$

(d) Show that if T is a tree on n vertices, then

$$\mathcal{L}(T) \cong B_{n-1}.$$

(29) Fill in the details of the proof of Lemma 5.7.1.

(30) This exercise refers to the proof that $\chi(\Pi_n) = (t-1)(t-2)\cdots(t-n+1)$ at the end of Section 5.7.
(a) Prove the reverse direction of the equivalence which lead to equation (5.20).
(b) Show that the map $\iota : I(X) \to I(X)$ is a well-defined sign-reversing involution without fixed points.
(c) Show that the map $\rho : (P/\sim) \to \Pi_n$ is a well-defined isomorphism of posets.

(31) Prove that $\chi(L_n(q);t)$ factors over the integers in two ways: by using quotient posets and by using Theorem 5.7.5.

(32) Reprove the factorization for $\chi(\Pi_n)$ using Theorem 5.7.5.

(33) This exercise is about Lemma 5.8.1(c).
(a) Finish the proof.
(b) Show that if $[x] = [\hat{1}]$, then it is possible to have $[x] \wedge [y] \neq [x \wedge y]$.

(34) (a) State and prove a dual version of Theorem 5.8.3.
(b) Using Weisner's Theorem or its dual from part (a), rederive the formula for $\mu(P)$ when $P = D_n, \Pi_n$, and $L_n(q)$.

(35) (a) Let L be a finite lattice with atom set $\mathcal{A}$. Show that if $\bigvee \mathcal{A} \neq \hat{1}$, then $\mu(L) = 0$.
(b) Use the Crosscut Theorem, Theorem 5.8.4, to rederive the formula for $\mu(D_n)$.

(36) Let L be a finite distributive lattice with atom set $\mathcal{A}(L)$ and join irreducibles $\operatorname{Irr}(L)$.
(a) Show that $\mathcal{A}(L) \subseteq \operatorname{Irr}(L)$ and that this is in fact true for any finite lattice.
(b) Show that if $\mathcal{A}(L) \subset \operatorname{Irr}(L)$ (proper subset), then $\bigvee \mathcal{A}(L) \neq \hat{1}$.
(c) Show that if $\mathcal{A}(L) = \operatorname{Irr}(L)$, then $L \cong B_n$ where $n = \#\mathcal{A}(L)$.

(d) Show that in L we have

$$\mu(\hat{0}, \hat{1}) = \begin{cases} (-1)^{\#\mathcal{A}(L)} & \text{if } \mathcal{A}(L) = \operatorname{Irr}(L), \\ 0 & \text{if } \mathcal{A}(L) \subset \operatorname{Irr}(L). \end{cases}$$

(37) Consider Young's lattice Y of all integer partitions λ ordered by containment of Young diagrams. Given λ, consider the interval $P_\lambda = [(1), \lambda]$ as a subposet of Y. Recall that $|\lambda| = \sum_i \lambda_i$ where $\lambda = (\lambda_1, \ldots, \lambda_k)$.
(a) Compute $\mu(P_\lambda) = \mu((1), \lambda)$ for all λ with $1 \le |\lambda| \le 3$.
(b) Show that $\mu(P_\lambda) = 0$ for $|\lambda| \ge 4$ in two ways: by using Weisner's Theorem (Theorem 5.8.3) and by using the Crosscut Theorem (Theorem 5.8.4).

(38) Complete the proof of Proposition 5.9.1.

(39) Complete the proof of Theorem 5.9.3(a).

(40) Complete the proof of Theorem 5.9.4.

(41) Show by direct computation (without using Theorem 5.9.4) that we have $F_\mu(t) = F_\zeta(t)^{-1}$
(a) in C_∞,
(b) in B_∞.

(42) (a) Prove that the direct sum of posets satisfies the axioms for a poset.
(b) Show that D_∞ is isomorphic to a direct sum.

(43) Prove the analogue of Theorem 5.9.3 in the case when P is Dirichlet.

(44) Prove Theorem 5.9.5.

(45) Prove (5.27). Hint: Use Theorem 5.9.5.

Chapter 6

Counting with Group Actions

Sometimes we want to count objects up to some symmetry. As an example, consider counting necklaces with colored beads where two necklaces are considered the same if one is a rotation of the other. We can deal with such situations by considering a group acting on the underlying set. These tools can also be used in other contexts, such as in proving congruences from number theory or enumeration using roots of unity.

6.1. Groups acting on sets

In this section we introduce the basic notions which will be used throughout this chapter. We wish to have a formal way to talk about symmetries of objects.

Let G be a group with identity element e, and let X be a set. We say that G *acts* on X if, for each $g \in G$, there is a map $g : X \to X$ such that for all $x \in X$

(a) $h(g(x)) = (hg)(x)$ for all $h, g \in G$,

(b) $e(x) = x$.

The reader should be careful to distinguish between when g is being used to denote a group element and when it is being used as a map on X. For example, on the left in property (a), $h(g(x))$ means apply the map g to x and then apply the map h; i.e., compose the two maps. But on the right, hg refers to the product in G and $(hg)(x)$ applies the corresponding map to x. It is common to write gx for $g(x)$. This should cause no confusion with group multiplication since x is an element of X, not G.

By way of illustration, consider the 4-cycle $(1, 2, 3, 4) \in \mathfrak{S}_4$ and the group $G = \langle(1, 2, 3, 4)\rangle$ where the angle brackets denote the group generated by the elements inside. So in our case

$$G = \{e = (1)(2)(3)(4),\ (1, 2, 3, 4),\ (1, 3)(2, 4),\ (1, 4, 3, 2)\}. \tag{6.1}$$

Of course, G acts on $[4]$ in the usual way. But we wish to consider an action on

$$X = \binom{[4]}{2} = \{\{1,2\},\ \{1,3\},\ \{1,4\},\ \{2,3\},\ \{2,4\},\ \{3,4\}\}$$

given by

$$g\{x,y\} = \{gx, gy\} \tag{6.2}$$

for $g \in G$ and $\{x,y\} \in X$. For example,

$$(1,2,3,4)\{1,3\} = \{(1,2,3,4)1, (1,2,3,4)3\} = \{2,4\}.$$

For our first result, we will show that the maps $g : X \to X$ have special properties.

Proposition 6.1.1. *Suppose G acts on X. Then $g : X \to X$ is a bijection for all $g \in G$ and $e : X \to X$ is the identity map on X.*

Proof. The statement about the map e follows immediately from part (b) of the definition of a group action. To prove that $g : X \to X$ is always a bijection, it suffices to show that $g^{-1} : X \to X$ is its compositional inverse. So we must prove that, for all $x \in X$, we have

$$g^{-1}(g(x)) = x = g(g^{-1}(x)).$$

We will prove the first equality, leaving the second to the reader. But using requirements (a) and (b) in the definition of a group action in that order gives

$$g^{-1}(g(x)) = (g^{-1}g)(x) = e(x) = x$$

as required. □

When counting under the action of a group, all the elements of X which can be obtained by acting on a given $x \in X$ by the elements of the group will be considered the same. This leads to the following definition. If G acts on X, then the *orbit* of $x \in X$ is

$$\mathcal{O}_x = \{gx \mid g \in G\}.$$

Note that $\mathcal{O}_x \subseteq X$. It is important to keep in mind when we are talking about elements of G and when we are talking about elements of X. Continuing the example above, if $x = \{1,2\}$, then

$$\begin{aligned}\mathcal{O}_{\{1,2\}} &= \{\,\{1,2\},\ (1,2,3,4)\{1,2\},\ (1,3)(2,4)\{1,2\},\ (1,4,3,2)\{1,2\}\,\}\\ &= \{\,\{1,2\},\ \{2,3\},\ \{3,4\},\ \{1,4\}\,\}.\end{aligned}$$

Similar computations show that $\mathcal{O}_{\{1,2\}} = \mathcal{O}_{\{2,3\}} = \mathcal{O}_{\{3,4\}} = \mathcal{O}_{\{1,4\}}$. That is, if we let $\mathcal{O}$ be the second line of the displayed equations above, then $\mathcal{O}_x = \mathcal{O}$ for all $x \in \mathcal{O}$. Now consider what happens if $x = \{1,3\}$:

$$\begin{aligned}\mathcal{O}_{\{1,3\}} &= \{\,\{1,3\},\ (1,2,3,4)\{1,3\},\ (1,3)(2,4)\{1,3\},\ (1,4,3,2)\{1,3\}\,\}\\ &= \{\,\{1,3\},\ \{2,4\}\,\}.\end{aligned}$$

As before $\mathcal{O}_{\{1,3\}} = \mathcal{O}_{\{2,4\}}$. Also note that we have the set partition $X = \mathcal{O}_{\{1,2\}} \uplus \mathcal{O}_{\{1,3\}}$. And for all orbits $\mathcal{O}$ we have $\#\mathcal{O} \mid \#X$ where the vertical bar indicates divisibility of integers. These observations will be explained shortly.

Another important concept for analyzing group actions is the stabilizer. If G acts on X, then the *stabilizer* of $x \in X$ is

$$G_x = \{g \in G \mid gx = x\}.$$

We clearly have $G_x \subseteq G$. Returning to our running example, it is easy to check by considering each $g \in G$ that

$$G_{\{1,2\}} = \{g \mid g\{1,2\} = \{1,2\}\} = \{e\}$$

and

$$G_{\{1,3\}} = \{g \mid g\{1,3\} = \{1,3\}\} = \{e,\ (1,3)(2,4)\}.$$

Observe that both of these stabilizers are subgroups of G. Furthermore

$$\frac{\#G}{\#G_{\{1,3\}}} = \frac{4}{2} = 2 = \#\mathcal{O}_{\{1,3\}}$$

and similarly for $\#G_{\{1,2\}}$ and $\#\mathcal{O}_{\{1,2\}}$. It is time to explain these patterns. The notation $H \leq G$ means that H is a subgroup of G.

Lemma 6.1.2. *Let G act on X.*

(a) *The distinct orbits form a set partition of X.*

(b) *For any $x \in X$ we have $G_x \leq G$.*

(c) *If G and X are finite and $x \in X$, then*

$$\#\mathcal{O}_x = \frac{\#G}{\#G_x}.$$

Proof. (a) Let $\mathcal{O}^{(1)}, \ldots, \mathcal{O}^{(k)}$ be the distinct orbits of G acting on X. We first need to show $X = \bigcup_i \mathcal{O}^{(i)}$. Since $\mathcal{O}^{(i)} \subseteq X$ for all i we clearly have $\bigcup_i \mathcal{O}^{(i)} \subseteq X$. For the reverse containment, if $x \in X$, then $x = ex \in \mathcal{O}_x$. Also $\mathcal{O}_x = \mathcal{O}^{(i)}$ for some i. So x is in the union as desired.

To show that the union is disjoint, it suffices to prove that if $\mathcal{O}_x \cap \mathcal{O}_y \neq \emptyset$ for two orbits $\mathcal{O}_x, \mathcal{O}_y$, then $\mathcal{O}_x = \mathcal{O}_y$. We will show that $\mathcal{O}_x \subseteq \mathcal{O}_y$ as then the reverse containment follows by just interchanging the roles of $\mathcal{O}_x$ and $\mathcal{O}_y$. So let $z \in \mathcal{O}_x$. Then there is $g \in G$ with $z = gx$. By assumption, there is some $u \in \mathcal{O}_x \cap \mathcal{O}_y$ and thus there exist $h, k \in G$ with $u = hx$ and $u = ky$. It is an easy exercise to show that $u = hx$ is equivalent to $x = h^{-1}u$. It follows that

$$z = gx = g(h^{-1}u) = g(h^{-1}(ky)) = (gh^{-1}k)(y).$$

This means $z \in \mathcal{O}_y$.

(b) We must show that G_x is a group, where the associative law is inherited from G. We have $e \in G_x$ because $ex = x$. If $g \in G_x$, then $gx = x$ and so, as noted in (a), $g^{-1}x = x$. This gives $g^{-1} \in G_x$ so we have closure under taking inverses. Finally, if $g, h \in G_x$, then $gx = x$ and $hx = x$. This gives

$$(gh)(x) = g(h(x)) = g(x) = x$$

which gives closure under taking products.

(c) By part (b) we can apply Lagrange's Theorem which tells us that $\#G/\#G_x = \#(G/G_x)$ where G/G_x is the set of left cosets of G_x in G. So it suffices to find a bijection $f : G/G_x \to \mathcal{O}_x$. Define $f(gG_x) = gx$. We must show that this map is well-defined in that $gG_x = hG_x$ implies $gx = hx$. The hypothesis implies that $g = hk$ where $k \in G_x$ so that $kx = x$. Now

$$gx = (hk)x = h(kx) = hx$$

as we wished. To show f is bijective, it suffices to construct an inverse. Define $f^{-1} : \mathcal{O}_x \to G/G_x$ by $f^{-1}(gx) = gG_x$ for each $gx \in \mathcal{O}_x$. This is clearly the inverse as long as it is well-defined. So we must show that if $gx = hx$, then $gG_x = hG_x$. The hypothesis implies that $x = (g^{-1}h)x$ so that $g^{-1}h \in G_x$. This is equivalent to the desired conclusion. □

6.2. Burnside's Lemma

As remarked in the previous section, when a group G acts on a set X one sometimes wishes to consider all the elements in a given orbit of G the same. So it would be useful to have a formula for the number of orbits of the action. This result is usually referred to as Burnside's Lemma because it was proved in his 1897 book, reprinted in [**21**], although Burnside himself was aware that the formula was already known.

For computations, it is best to express the number of orbits in terms of a concept dual to the notion of a stabilizer of an elements of X. If G acts on X, then the *fixed point set* of $g \in G$ is

$$X^g = \{x \in X \mid gx = x\}.$$

We have $X^g \subseteq X$ while for the stabilizer of $x \in X$ we have $G_x \le G$. To remember this notation, note that in both cases the base of the expression (rather than the superscript or subscript) indicates whether we are dealing with a subset of X or G. Continuing the example from the previous section $X^{(1,2,3,4)} = \emptyset$ and

$$X^{(1,3)(2,4)} = \{\{1,3\}, \{2,4\}\}.$$

For any G and X we have

$$X^e = X. \tag{6.3}$$

Lemma 6.2.1 (Burnside's Lemma)**.** *Let G act on X with G, X finite. Then*

$$\textit{number of orbits} = \frac{1}{\#G}\sum_{g\in G} \#X^g.$$

Proof. We will use the fact that for any positive integer n

$$1 = \frac{n}{n} = \overbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}^{n}.$$

It follows that for any finite set $\mathcal{O}$ we have

$$\sum_{x\in\mathcal{O}} \frac{1}{\#\mathcal{O}} = 1.$$

From Lemma 6.1.2(a) and the finiteness hypothesis we can write $X = \mathcal{O}^{(1)} \uplus \cdots \uplus \mathcal{O}^{(k)}$ where $\mathcal{O}^{(i)}$, $1 \le i \le k$, are the orbits. Using this, the previous displayed equation, and Lemma 6.1.2(c) gives

$$\begin{aligned}
\text{number of orbits } &= k \\
&= \sum_{x \in \mathcal{O}^{(1)}} \frac{1}{\#\mathcal{O}^{(1)}} + \cdots + \sum_{x \in \mathcal{O}^{(k)}} \frac{1}{\#\mathcal{O}^{(k)}} \\
&= \sum_{x \in X} \frac{1}{\#\mathcal{O}_x} \\
&= \frac{1}{\#G} \sum_{x \in X} \#G_x.
\end{aligned}$$

To express this last summation as a sum over G, consider the matrix M with rows indexed by G, columns indexed by X, and entries

$$M_{g,x} = \begin{cases} 1 & \text{if } gx = x, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $\#G_x$ is the sum of the entries in column x of M. Similarly $\#X^g$ is the sum of the entries in row g of M. So $\sum_{x \in X} \#G_x$ and $\sum_{g \in G} \#X^g$ are equal since both give the sum of the entries of M. This completes the proof. □

One standard application of Burnside's Lemma is to count colored objects up to symmetry. To do this, we have to consider group actions on functions. Given sets X, Y, we then let Y^X denote the set of all functions $f : X \to Y$. We know from Table 1.1 that $|Y^X| = |Y|^{|X|}$. Suppose group G acts on the domain X. Then the *induced action of G on Y^X* is defined by sending the function f to the function gf such that

$$(gf)(x) = f(g^{-1}x)$$

for all $x \in X$. Equivalently $gf = f \circ g^{-1}$ where the circle indicates composition of functions. The reason that f is composed with g^{-1} rather than g is to make sure that condition (a) in the definition of a group action is satisfied. Indeed,

$$h(g(f)) = g(f) \circ h^{-1} = f \circ g^{-1} \circ h^{-1} = f \circ (hg)^{-1} = (hg)(f).$$

And condition (b) is easy to verify as $ef = f \circ e^{-1} = f \circ e = f$ since $e : X \to X$ is the identity map. Note that if we defined $gf = f \circ g$, then the two sides of (a) would not be equal.

As our first application of Burnside's Lemma, we consider colorings of a 4-bead necklace with two colors: black (B) and white (W). We wish to count the number of different necklaces if two necklaces are considered the same if one is a rotation of the other. If rotation is not considered, then each of the 4 beads can be colored in two ways, resulting in $4^2 = 16$ necklaces which are displayed in Figure 6.1. Putting all necklaces which are rotations of a given necklace into a set together, we obtain a partition of the set of all necklaces. So we wish to count the number of blocks, which can easily be seen to be six in this case. But we would like to take an approach that could be generalized to more colors or more beads.

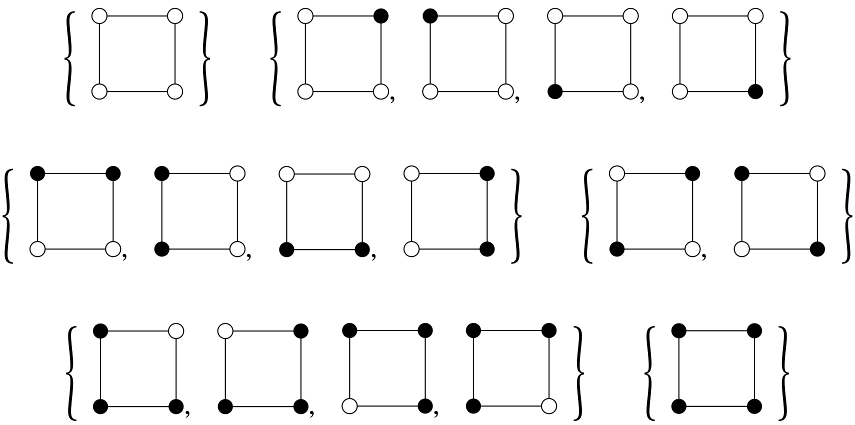


Figure 6.1. The orbits of 2-colored, 4-bead necklaces under rotation

To get group actions into the act, label the four corners of the necklace using the set $X = \{1, 2, 3, 4\}$ as shown in Figure 6.2. Note that $G = \langle(1, 2, 3, 4)\rangle$ acts on X in a way that corresponds to rotation of the necklace. Letting $Y = \{B, W\}$, each necklace coloring defines a function $f : X \to Y$ and the action of G on Y^X rotates colored necklaces. The blocks in Figure 6.1 are exactly the orbits of this action. So we can use Burnside's Lemma as long as we can find a way to compute the fixed points $(Y^X)^g$ for each $g \in G$. To do that, the following lemma will be crucial. It shows that the fixed points of g acting on Y^X are exactly the functions f which are constant on the cycles of g acting on X. An example will be found in Figure 6.2.

Lemma 6.2.2. *Let G act on a set X and let Y be another set with G, X, Y finite. Let $g \in G$ and*

$$c(g) = \text{number of cycles of } g \text{ acting on } X.$$

(a) *For $f \in Y^X$ we have $gf = f$ if and only if $f(x) = f(x')$ whenever x, x' are in the same cycle of g acting on X.*

(b) *We have*

$$\#(Y^X)^g = |Y|^{c(g)}.$$

Proof. (a) We will prove the forward direction as the reverse is similar. Since x, x' are in the same cycle of g there must be an i such that $g^i x = x'$. And since $gf = f$ we also have $g^i f = f$. It follows that

$$f(x) = f(g^{-i}x') = (g^i f)(x') = f(x').$$

(b) From part (a), the fixed points of g acting on Y^X are obtained as follows. Choose an element y for each cycle of g acting on X and let $g(x) = y$ for all x in that cycle. The number of ways of doing this is clearly $|Y|^{c(g)}$. □

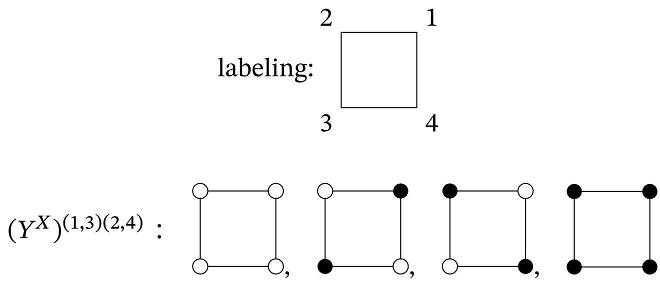


Figure 6.2. Labeling the necklace and the fixed points of $g = (1,3)(2,4)$

Returning to our example where $|Y| = 2$, we can use part (b) of the previous lemma to construct the following table for the $g \in \langle(1,2,3,4)\rangle$:

g	$c(g)$	$\#(Y^X)^g$
$(1)(2)(3)(4)$	4	2^4
$(1,2,3,4)$	1	2^1
$(1,3)(2,4)$	2	2^2
$(1,4,3,2)$	1	2^1

Applying Burnside's Lemma, we see that the number of distinct necklaces is

$$\frac{1}{\#G}\sum_{g\in G}\#(Y^X)^g = \frac{1}{4}(2^4 + 2^1 + 2^2 + 2^1) = 6$$

as before. An immediate benefit of using this approach is that little work is needed to do the more general case where there are r colors for the beads. Because of Lemma 6.2.2 the powers of 2 just get replaced by powers of r so that for 4-bead, r-color necklaces under rotation

$$\text{number of orbits} = \frac{1}{4}(r^4 + r^2 + 2r). \tag{6.4}$$

It follows that 4 must divide evenly into $r^4 + r^2 + 2r$ for all $r \in \mathbb{P}$, a fact that is not obvious a priori.

For a 3-dimensional example, let us find the number of distinct colorings of faces of a cube with r colors if two colorings are equivalent when one is a rotation of the other. Label the faces with $X = [6]$ as shown in Figure 6.3 where arrows indicate labels for faces which cannot be seen. Colorings will be functions $f \in Y^X$ where $\#Y = r$. Rotations can be classified by the axis of rotation and the angle through which one rotates, the exception being the identity whose cycle structure acting on X is $e = (1)(2)(3)(4)(5)(6)$. If the rotation is through an axis bisecting opposite faces, then one will get the same cycle decomposition for angles of both $\pm 90°$. Using the axis and direction given for the first rotation in Figure 6.3, one gets the permutation $(1)(2,3,4,5)(6)$. Furthermore, there are three possible pairs of opposite faces to use, giving a total of (3 axes)(2 rotations per axis) $= 6$ rotations of this type. A complete list of possible rotations g is summarized in the following table, where the example

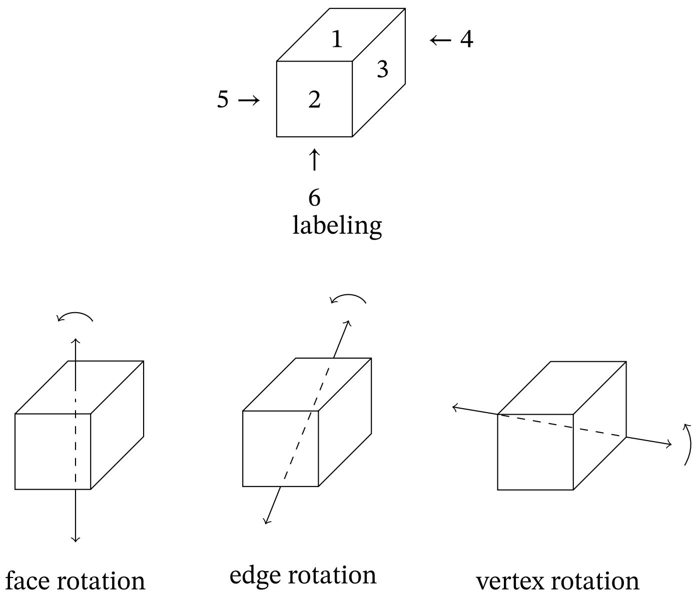


Figure 6.3. Labeling and rotating a cube

permutations are all as indicated in Figure 6.3:

type of rotation g	number of g	example g	$c(g)$	$\#(Y^X)^g$
identity	1	$(1)(2)(3)(4)(5)(6)$	6	r^6
face by $\pm 90°$	$3 \cdot 2 = 6$	$(1)(2,3,4,5)(6)$	3	r^3
face by $180°$	$3 \cdot 1 = 3$	$(1)(2,4)(3,5)(6)$	4	r^4
edge by $180°$	$6 \cdot 1 = 6$	$(1,4)(2,6)(3,5)$	3	r^3
vertex by $\pm 120°$	$4 \cdot 2 = 8$	$(1,2,5)(3,6,4)$	2	r^2

From this table we see that the total number of rotations is $\#G = 24$. Applying Burnside's Lemma to the information given in the second and fourth columns gives the number of colorings as

$$\frac{1}{\#G} \sum_{g \in G} \#(Y^X)^g = \frac{1}{24}(r^6 + 3r^4 + 12r^3 + 8r^2).$$

As a check, consider the case $r = 2$ with $X = \{B, W\}$. Then by inspection of the small number of possibilities one can verify the following data where $B^i W^j$ indicates having i faces colored black and j colored white:

color distribution	number of colorings
B^6 or W^6	$1 + 1 = 2$
B^5W or WB^5	$1 + 1 = 2$
B^4W^2 or W^2B^4	$2 + 2 = 4$
B^3W^3	2

So the total number of colorings is 10. On the other hand

$$\frac{1}{24}(2^6 + 3 \cdot 2^4 + 12 \cdot 2^3 + 8 \cdot 2^2) = 10$$

as well.

6.3. The cycle index

In the previous section we saw in equation (6.4) that the number of orbits of $G = \langle(1,2,3,4)\rangle$ acting on $Y^{[4]}$ where $\#Y = r$ is given by $(r^4 + r^2 + 2r)/4$, a polynomial in a single variable. By permitting more variables, we can encode more information about the action of a group G on a set X. This will permit us to find generating functions for the number of orbits of the induced actions of G on subsets $\binom{X}{k}$ and on permutations $P(X,k)$ for $k \geq 0$.

Suppose G is a finite group acting on a set X with $\#X = n$. If $g \in G$, then let

$$c_i = c_i(g) = \text{ number of cycles of g of length } i.$$

Given variables $t_1, \ldots, t_n$, the associated *cycle index* of g is the monomial

$$z(g) = z(g; t_1, t_2, \ldots, t_n) = t_1^{c_1} t_2^{c_2} \cdots t_n^{c_n}.$$

The *cycle index* or *cycle indicator* of G acting on X is

$$Z(G) = Z(G; t_1, t_2, \ldots, t_n) = \frac{1}{\#G} \sum_{g \in G} z(g).$$

To illustrate, if $X = [4]$ and $G = \langle(1,2,3,4)\rangle$, then by (6.1) we have

g	$z(g)$
$(1)(2)(3)(4)$	t_1^4
$(1,2,3,4)$	t_4
$(1,3)(2,4)$	t_2^2
$(1,4,3,2)$	t_4

so that

$$Z(G) = \frac{1}{4}(t_1^4 + t_2^2 + 2t_4). \tag{6.5}$$

Note that setting $t_1 = t_2 = t_3 = t_4 = r$ in $Z(G)$ we obtain the count in (6.4) for the orbits of G acting on $[r]^X$. It turns out that other specializations of $Z(G)$ give orbit-generating functions for other actions.

If G acts on X and $k \in \mathbb{N}$, then there is an *induced action on* $\binom{X}{k}$ given by

$$g\{x_1, x_2, \ldots, x_k\} = \{gx_1, gx_2, \ldots, gx_k\}.$$

The special case of this action when $G = \langle(1,2,3,4)\rangle$ and $X = [4]$ was the running example in Section 6.1. Instead of subsets, one can consider the set $P(X,k)$ of k-permutations of X. The *induced action on* $P(X,k)$ is

$$g(x_1 x_2 \ldots x_k) = g(x_1)g(x_2)\ldots g(x_k).$$

In Section 6.1, we say that the orbits of $G = \langle(1,2,3,4)\rangle$ acting on $\binom{[4]}{2}$ were

$$\{\{1,2\}, \{2,3\}, \{3,4\}, \{1,4\}\} \quad \text{and} \quad \{\{1,3\}, \{2,4\}\}.$$

Similarly, one can compute that the orbits of G acting on

$$P([4],2) = \{12,21,13,31,14,41,23,32,24,42,34,43\}$$

are

$$\{12,23,34,41\},\ \{21,32,43,14\},\ \text{and}\ \{13,24,31,42\}.$$

In order to apply Burnside's Lemma, we need an analogue of Lemma 6.2.2 in this context.

Lemma 6.3.1. *Let G act on X with X finite and consider $g \in G$.*

(a) *For $S \in \binom{X}{k}$ we have $gS = S$ if and only if S is a union of cycles of g (where to take the union we use the underlying set of each cycle).*

(b) *For $\pi = x_1 \ldots x_k \in P(X,k)$ we have $g\pi = \pi$ if and only if each $gx_i = x_i$ for all $i \in [k]$.*

Proof. (a) We prove the forward direction and leave the converse as an exercise. It suffices to prove that if $x \in S$ and x' is any element of the cycle of g containing x, then $x' \in S$. Since x, x' are in the same cycle of g there is some i with $g^i x = x'$. Also $x \in S$ and $gS = S$ so that

$$x' = g^i x \in g^i S = S$$

and we are done.

(b) By definition of the action on $P(X,k)$, $g\pi = \pi$ means that we have $g(x_1)\ldots g(x_k) = x_1 \ldots x_k$. Since permutations are ordered collections of elements, this is equivalent to $g(x_i) = x_i$ for all i as desired. □

We can now obtain expressions for the number of orbits of the induced actions of G on $\binom{X}{k}$ and on $P(X,k)$ from the cycle indicator for G's action on X itself. Note that the first generating polynomial is ordinary while the second is exponential.

Theorem 6.3.2. *Let G be finite acting on X with $\#X = n$. Also let*

$$b_k = \text{ number of orbits of } G \text{ acting on } \binom{X}{k},$$
$$p_k = \text{ number of orbits of } G \text{ acting on } P(X,k).$$

(a) $\displaystyle\sum_{k=0}^{n} b_k t^k = Z(G; 1+t, 1+t^2, \ldots, 1+t^n).$

(b) $\displaystyle\sum_{k=0}^{n} p_k \frac{t^k}{k!} = Z(G; 1+t, 1, \ldots, 1).$

Proof. (a) Applying Burnside's Lemma to $\binom{X}{k}$ for each k and interchanging summations gives

$$\sum_{k=0}^{n} b_k t^k = \frac{1}{\#G} \sum_{g\in G} \sum_{k=0}^{n} \#\binom{X}{k}^g t^k.$$

So it suffices to show that for all $g \in G$ we have

$$z(g; 1+t, \ldots, 1+t^n) = \sum_{k=0}^{n} \binom{X}{k}^g t^k \tag{6.6}$$

since then

$$\sum_{k=0}^{n} b_k t^k = \frac{1}{\#G} \sum_{g \in G} z(g; 1+t, \ldots, 1+t^n) = Z(G; 1+t, \ldots, 1+t^n).$$

To prove (6.6), we use weight-generating functions as in Section 3.4. Weight a set $S \in \binom{X}{k}^g$ by $\operatorname{wt} S = t^{\#S}$. Then the weight ogf $f_{\mathcal{S}}(t)$ for $\mathcal{S} = \binom{X}{0}^g \uplus \cdots \uplus \binom{X}{n}^g$ is precisely the right-hand side of (6.6). To obtain the left side, write $g = g_1 g_2 \ldots g_m$ where the g_i are the cycles of g. By Lemma 6.3.1(a), $S \in \mathcal{S}$ if and only if S is a union of some of the g_i. So $\mathcal{S}$ can be thought of as a product $\mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_m$ where the ith coordinate can be either g_i or $\emptyset$ if S does or does not contain g_i, respectively. And the weighting on $\mathcal{S}$ can be obtained by letting the weight of that coordinate be $t^{\#g_i}$ or $t^{\#\emptyset} = 1$ for the two respective cases and then using the usual product weighting. Using the Sum and Product Rules for weight ogfs (Lemma 3.4.1) yields

$$\begin{aligned} f_{\mathcal{S}}(t) &= (1+t^{\#g_1})(1+t^{\#g_2}) \cdots (1+t^{\#g_m}) \\ &= (1+t)^{c_1(g)}(1+t^2)^{c_2(g)} \ldots (1+t^n)^{c_n(g)} \\ &= z(g; 1+t, 1+t^2, \ldots, 1+t^n) \end{aligned}$$

which completes the proof.

(b) Let $c_1 = c_1(g)$. By Lemma 6.3.1(b), $\pi = x_1 \ldots x_k \in P(X,k)^g$ if and only if x_i is a fixed point of g for all $i \in [k]$. Since fixed points are cycles of length one, if we choose the elements of π in the order $x_1, \ldots, x_k$, then the number of choices for x_i is $c_1 - i + 1$. It follows that $|P(X,k)^g| = c_1 \downarrow_k$ and so

$$\sum_{k=0}^{n} |P(X,k)^g| \frac{t^k}{k!} = \sum_{k=0}^{n} \frac{c_1 \downarrow_k}{k!} t^k = \sum_{k=0}^{n} \binom{c_1}{k} t^k = (1+t)^{c_1}.$$

Finally, applying Burnside's Lemma similarly to (a) yields

$$\begin{aligned} \sum_{k=0}^{n} p_k \frac{t^k}{k!} &= \frac{1}{\#G} \sum_{g \in G} \sum_{k=0}^{n} |P(X,k)^g| \frac{t^k}{k!} \\ &= \frac{1}{\#G} \sum_{g \in G} (1+t)^{c_1(g)} \\ &= Z(G; 1+t, 1, \ldots, 1) \end{aligned}$$

as desired. □

As a reality check, let's compute the generating functions for $G = \langle(1,2,3,4)\rangle$ and $X = [4]$ and compare the results with the computations of the orbits for $\binom{X}{2}$ and $P(X,2)$

above. Using part (a) of the previous result and (6.5) gives

$$\begin{aligned}\sum_k b_k t^k &= Z(G; 1+t, \dots, 1+t^4)\\ &= \frac{1}{4}((1+t)^4 + (1+t^2)^2 + 2(1+t^4))\\ &= 1 + t + 2t^2 + t^3 + t^4.\end{aligned}$$

Note that the coefficient of t^2 is 2 which is the number of orbits we found previously for this case. Note also that this generating function gives you much more information as it gives the number of orbits for all k, not just $k = 2$. We now make the analogous computation for G's action on $P(X, k)$:

$$\begin{aligned}\sum_k p_k \frac{t^k}{k!} &= Z(G; 1+t, 1, 1, 1) = \frac{1}{4}((1+t)^4 + 1 + 2)\\ &= 1 + t + \frac{3}{2}t^2 + t^3 + \frac{1}{4}t^4\\ &= 1 + \frac{t}{1!} + 3\frac{t^2}{2!} + 6\frac{t^3}{3!} + 6\frac{t^4}{4!}.\end{aligned}$$

The coefficient of $t^2/2!$ is 3 which agrees with our earlier computations.

6.4. Redfield–Pólya theory

We can use the cycle index to give more refined information about orbit counts. For example, it can be used to compute the number of necklaces up to rotation which have a given number of beads of each color. This approach was developed by Redfield [**71**]. It was rediscovered and popularized by Pólya [**70**].

Let G act on X with G, X finite, and let Y be a set of variables. The *weight* of $f \in Y^X$ is the monomial

$$\operatorname{wt} f = \prod_{x \in X} f(x).$$

Consider the example of the 4-bead necklace with two colors $Y = \{B, W\}$ discussed in Section 6.2. Then the second necklace on the first line of Figure 6.1 would have weight

$$\operatorname{wt} f = f(1)f(2)f(3)f(4) = BWWW = BW^3.$$

Note that every other necklace in the orbit of the given one has the same weight. This is not an accident.

Proposition 6.4.1. *Let G act on X with G, X finite, and let Y be a set of variables. If f, f' are in the same orbit of G acting on Y^X, then* $\operatorname{wt} f = \operatorname{wt} f'$.

Proof. Since f, f' are in the same orbit, there is some $g \in G$ with $f' = gf$. By definition of G's action on Y^X and the fact that $g : X \to X$ is a bijection

$$\operatorname{wt} f' = \operatorname{wt}(gf) = \prod_{x \in X} (gf)(x) = \prod_{x \in X} f(g^{-1}x) = \prod_{x' \in X} f(x') = \operatorname{wt} f$$

as desired. □

Because of this result, if $\mathcal{O}$ is an orbit of G acting on Y^X, then we have a well-defined *weight of* $\mathcal{O}$ given by $\operatorname{wt}\mathcal{O} = \operatorname{wt} f$ for any $f \in \mathcal{O}$. So the second orbit in Figure 6.1 would have $\operatorname{wt}\mathcal{O} = BW^3$.

We can express the weight-generating function for the orbits of G acting on Y^X by making certain substitutions into the cycle index for G acting on X. The proof will be a weighted version of the demonstration of Burnside's Lemma combined with some ideas in the proof of Theorem 6.3.2(a). We note that the theory of weight-generating functions from Section 3.4 carries over easily to generating functions with many variables.

Theorem 6.4.2 (Redfield–Pólya Theorem). *Let G be a finite group acting on X where $\#X = n$. Suppose Y is a set of variables. Then*

$$\sum_{\mathcal{O}} \operatorname{wt}\mathcal{O} = Z\left(G; \sum_{y\in Y} y, \sum_{y\in Y} y^2, \ldots, \sum_{y\in Y} y^n\right)$$

where the left-hand sum is over the orbits of G acting on Y^X.

Proof. Recall from the proof of Burnside's Lemma that for any orbit $\mathcal{O}$ we have

$$\sum_{f\in\mathcal{O}} \frac{1}{\#\mathcal{O}_f} = 1$$

since $\mathcal{O}_f = \mathcal{O}$ for all $f \in \mathcal{O}$. It follows that

$$\sum_{f\in\mathcal{O}} \frac{\operatorname{wt} f}{\#\mathcal{O}_f} = \sum_{f\in\mathcal{O}} \frac{\operatorname{wt}\mathcal{O}}{\#\mathcal{O}_f} = \operatorname{wt}\mathcal{O}.$$

Now using Lemma 6.1.2(a) and (c) yields

$$\sum_{\mathcal{O}} \operatorname{wt}\mathcal{O} = \sum_{\mathcal{O}} \sum_{f\in\mathcal{O}} \frac{\operatorname{wt} f}{\#\mathcal{O}_f} = \sum_{f\in Y^X} \frac{\operatorname{wt} f}{\#\mathcal{O}_f} = \frac{1}{\#G} \sum_{f\in Y^X} |G_f| \operatorname{wt} f. \tag{6.7}$$

Again taking a tip from the demonstration of Lemma 6.2.1, consider a matrix M with rows indexed by G, columns indexed by Y^X, and entries

$$M_{g,f} = \begin{cases} \operatorname{wt} f & \text{if } gf = f, \\ 0 & \text{otherwise.} \end{cases}$$

where $g \in G$ and $f \in Y^X$. The sum of column f of M is $|G_f| \operatorname{wt} f$ while the sum of row g is

$$\sum_{f\in(Y^X)^g} \operatorname{wt} f.$$

Using this and (6.7) gives

$$\sum_{\mathcal{O}} \operatorname{wt}\mathcal{O} = \frac{1}{\#G} \sum_{g,f} M_{g,f} = \frac{1}{\#G} \sum_{g\in G} \sum_{f\in(Y^X)^g} \operatorname{wt} f.$$

To finish the proof, we just need to show that

$$z\left(g; \sum_{y\in Y} y, \sum_{y\in Y} y^2, \ldots, \sum_{y\in Y} y^n\right) = \sum_{f\in(Y^X)^g} \operatorname{wt} f \tag{6.8}$$

because then, combining the previous displayed equations,

$$\sum_{\mathcal{O}} \operatorname{wt} \mathcal{O} = \frac{1}{\#G} \sum_{g\in G} z\left(g;\ \sum_{y\in Y} y,\ \sum_{y\in Y} y^2,\ \ldots,\ \sum_{y\in Y} y^n\right)$$

$$= Z\left(G;\ \sum_{y\in Y} y,\ \sum_{y\in Y} y^2,\ \ldots,\ \sum_{y\in Y} y^n\right).$$

Let $\mathcal{S} = (Y^X)^g$. By definition, the right side of (6.8) is the weight-generating function of $\mathcal{S}$. Let $g = g_1 \cdots g_m$ be the decomposition of g into disjoint cycles. Similarly to the proof of Theorem 6.3.2(a), we can decompose $\mathcal{S}$ as a product $\mathcal{S}_1 \times \cdots \times \mathcal{S}_m$ where the ith coordinate contains the possible image sets $f(g_i)$ of f on the ith cycle. But from Lemma 6.2.2 we have that g fixes f if and only if f is constant on the g_i. So $f(g_i)$ must consist of $\#g_i$ copies of some element of Y. It follows that the weight-generating function for these sets is $\sum_{y\in Y} y^{\#g_i}$. Now the Product Rule for ogfs yields

$$\sum_{f\in(Y^X)^g} \operatorname{wt} f = \left(\sum_{y\in Y} y^{\#g_1}\right)\left(\sum_{y\in Y} y^{\#g_2}\right)\cdots\left(\sum_{y\in Y} y^{\#g_m}\right)$$

$$= \left(\sum_{y\in Y} y\right)^{c_1(g)}\left(\sum_{y\in Y} y^2\right)^{c_2(g)}\cdots\left(\sum_{y\in Y} y^n\right)^{c_n(g)}$$

$$= z\left(g;\ \sum_{y\in Y} y,\ \sum_{y\in Y} y^2,\ \ldots,\ \sum_{y\in Y} y^n\right)$$

which is what we wished to show. □

Let us look again at the 4-bead necklaces under the rotation group $G = \langle(1,2,3,4)\rangle$ and with color set $Y = \{B, W\}$. The cycle index for G was computed in (6.5). So, by the result just proved, the weight-generating function for the orbits is

$$Z(G;\ B+W,\ B^2+W^2,\ B^3+W^3,\ B^4+W^4)$$

$$= \frac{1}{4}\left[(B+W)^4 + (B^2+W^2)^2 + 2(B^4+W^4)\right]$$

$$= B^4 + B^3W + 2B^2W^2 + BW^3 + W^4.$$

Of course, in this simple example we could have gotten the same result by just looking at the orbit list in Figure 6.1.

For a more substantial example, we return to the problem, first raised in Section 1.9, of counting unlabeled graphs with n vertices. This can be handled using either Theorem 6.3.2(a) or Theorem 6.4.2. We will use the former as it will make the calculations slightly easier. In particular, we will find the generating function

$$\sum_{m\geq 0} g_m t^m \tag{6.9}$$

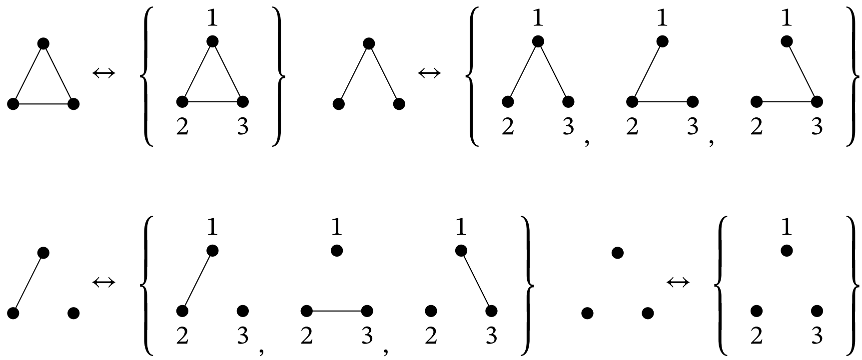


Figure 6.4. The orbits of labeled graphs and the corresponding unlabeled graphs

where g_m is the number of unlabeled graphs with n vertices and m edges. An unlabeled graph on n vertices can be considered as the orbit of a set of labeled graphs with vertex set $[n]$ under the action of the symmetric group $\mathfrak{S}_n$ in the following way. If E is the edge set of a graph with $V = [n]$, then $E \subseteq \binom{[n]}{2}$. Since the vertex set is fixed, we can identify the graph with its subset of edges. The orbits of this action when $n = 3$ are shown in Figure 6.4. So to use Theorem 6.3.2(a) we need to take $X = \binom{[n]}{2}$. In order to distinguish the action of $\pi \in \mathfrak{S}_n$ on $V = [n]$ from its action on X we will use $\pi^{(2)}$ and $\mathfrak{S}_n^{(2)}$ for the group elements and the group in the latter case, where the action on $\{i, j\} \in X$ is the usual one,

$$\pi\{i, j\} = \{\pi(i), \pi(j)\}.$$

It follows that

$$g_m = \text{number of orbits of the induced action of } \mathfrak{S}_n^{(2)} \text{ acting on } \binom{\binom{[n]}{2}}{m}$$

as needed for the theorem.

We must first compute the cycle index of $\mathfrak{S}_n^{(2)}$ acting on $X = \binom{[n]}{2}$. As usual when dealing with graphs, we will write $\{i, j\} \in X$ as ij. Either i, j are in the same cycle of π or they are in different cycles. Consider first when $i, j \in \kappa$, a cycle of π with $|\kappa| = k$. Consider the elements of κ as lying clockwise on a circle and breaking it up into k arcs of length one. An orbit of $\kappa^{(2)}$ consists of all pairs $\kappa^p(i)\kappa^p(j)$ for all possible powers p. This gives all pairs at the same distance d around the cycle where $1 \le d \le k/2$. See the orbit on the left in Figure 6.5 for an example when $\kappa = (1, 2, 3, 4, 5, 6)$ and $d = 2$. If $1 \le d < k/2$, then this orbit has k elements and so contributes a t_k to $z(\pi^{(2)})$. Since the number of such orbits is the floor function $\lfloor (k-1)/2 \rfloor$, these orbits together give a factor of $t_k^{\lfloor (k-1)/2 \rfloor}$. If k is even, then the orbit when $d = k/2$ contains $k/2$ edges for a factor of $t_{k/2}$. The orbit on the right in Figure 6.5 is an illustration. If we make the convention that $t_q = 1$ when q is not an integer, then we can write the total contribution of $\kappa^{(2)}$ as $t_k^{\lfloor (k-1)/2 \rfloor} t_{k/2}$ regardless of the parity of k.

Now consider the case where $i \in \kappa$ and $j \in \gamma$ for two different cycles of π, and suppose $\#\kappa = k, \#\gamma = l$. Now the orbit consists of edges of the form $\kappa^p(i)\gamma^p(j)$. So

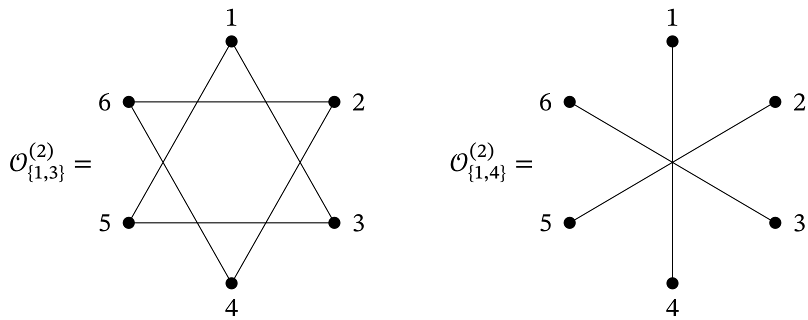


Figure 6.5. Two orbits of $\kappa^{(2)}$ when $\kappa = (1, 2, 3, 4, 5, 6)$

the number of edges in the orbit will be the smallest positive p such that $\kappa^p(i) = i$ and $\gamma^p(j) = j$. We have $\kappa^p(i) = i$ if and only if p is divisible by k and similarly for the second condition. Thus the smallest p satisfying both is $p = \operatorname{lcm}(k, l)$. Since this is independent of i, j all such orbits have the same size. The total number of possible pairs ij in the given two cycles is kl, which means that the number of orbits is $kl/\operatorname{lcm}(k, l) = \gcd(k, l)$. Hence these orbits contribute $t_{\operatorname{lcm}(k,l)}^{\gcd(k,l)}$ to $z(\pi^{(2)})$.

We are now in a position to calculate $z(\pi^{(2)})$. Suppose the cycle type of π is given in multiplicity notation as $(1^{m_1}, \ldots, n^{m_n})$ so that π has m_k cycles of length k. We have

$$z(\pi^{(2)}) = \prod_{k=1}^{n} t_k^{m_k \lfloor (k-1)/2 \rfloor}\, t_{k/2}^{m_k}\, t_k^{k\binom{m_k}{2}} \prod_{1 \le k < l \le n} t_{\operatorname{lcm}(k,l)}^{m_k m_l \gcd(k,l)}, \tag{6.10}$$

where the exponent factors of m_k, $\binom{m_k}{2}$, and $m_k m_l$ count the number of ways to choose a single cycle of length k, a pair of cycles both of length k, and two cycles one of length k and one of length l, respectively. Note that this expression depends only on the cycle type of π and not directly on π itself. So we will be able to combine terms with the same index by using the following result.

Proposition 6.4.3. *If $\lambda = (1^{m_1}, \ldots, n^{m_n})$, then the number of $\pi \in \mathfrak{S}_n$ with cycle type λ is*

$$k_\lambda = \frac{n!}{\prod_{k=1}^{n} k^{m_k}\, m_k!}.$$

Proof. Consider a template of n blank spaces arranged in cycles corresponding to a product of the given cycle type. An example follows this proof. There are $n!$ ways to fill the spaces with the elements of $[n]$. So to complete the count we just need to divide by the number of fillings that give the same permutation. If we fill a blank k-cycle with any of $(a_1, a_2, \ldots, a_k)$, $(a_2, a_3, \ldots, a_k, a_1)$, $\ldots$, $(a_k, a_1, \ldots, a_{k-1})$, then the permutation will not change. This gives a total of k^{m_k} possibilities for the m_k cycles of length k. Also, since disjoint cycles commute, we can permute any of these m_k cycles among themselves. This explains the factor of $m_k!$ and finishes the proof. □

As an example, suppose $\lambda = (3^2)$. Then the template would look like $(\_,\_,\_)(\_,\_,\_)$. Of the 6! ways to fill the template, $(1,2,3)(4,5,6)$ would be the same as $(2,3,1)(4,5,6)$ since $(1,2,3) = (2,3,1)$. Also $(1,2,3)(4,5,6)$ corresponds to the same permutation as $(4,5,6)(1,2,3)$ since cycles commute.

Combining (6.10) and Proposition 6.4.3, as well as canceling the $1/n!$ in the cycle index into the $n!$ in k_λ, gives

$$Z(\mathfrak{S}_n^{(2)}) = \sum_{\lambda \vdash n} \prod_{k=1}^{n} \frac{1}{k^{m_k}\, m_k!}\, t_k^{m_k\lfloor (k-1)/2\rfloor + k\binom{m_k}{2}}\, t_{k/2}^{m_k} \prod_{1\le k<l\le n} t_{\operatorname{lcm}(k,l)}^{m_k m_l \gcd(k,l)} \tag{6.11}$$

where $\lambda = (1^{m_1}, \ldots, n^{m_n})$. By Theorem 6.4.2, we obtain the generating function (6.9) for unlabeled graphs by number of edges using the substitution

$$\sum_{m\ge 0} g_m t^m = Z(\mathfrak{S}_n^{(2)},\ 1+t,\ 1+t^2,\ \ldots,\ 1+t^n).$$

As a check, suppose $n = 3$. Then there are three summands in (6.11) given in the chart

λ	summand
(1^3)	$t_1^3/6$
$(1,2)$	$t_1t_2/2$
(3)	$t_3/3$

so that

$$Z(\mathfrak{S}_n^{(2)}) = \frac{1}{6}t_1^3 + \frac{1}{2}t_1t_2 + \frac{1}{3}t_3.$$

Thus the generating function for unlabeled graphs on three vertices by number of edges is

$$\frac{1}{6}(1+t)^3 + \frac{1}{2}(1+t)(1+t^2) + \frac{1}{3}(1+t^3) = 1 + t + t^2 + t^3,$$

a fact which can be easily verified using Figure 6.4.

6.5. An application to proving congruences

We now show how group actions and Möbius inversion can be combined to prove various congruences from number theory. The advantages of this approach are twofold. One is that it can be used to prove a large array of congruences; see [**77**] for many examples. The other is that these demonstrations can be approached in a uniform manner rather than having to use ad hoc techniques for each of them.

Suppose $a, b \in \mathbb{Z}$ and $m \in \mathbb{P}$. Recall that a and b are *congruent modulo m*, which we write as $a \equiv b \pmod{m}$, if a and b both leave the same remainder on division by m. Equivalently, $m \mid a - b$. We will start with the easiest case, which is when the modulus m is a prime.

Lemma 6.5.1. *Let p be a prime. Let $G = \langle g \rangle$ be a group with $\#G = p$. Then for any finite set X on which G acts*

$$\#X \equiv \#X^g \pmod{p}.$$

Proof. By Lemma 6.1.2(c), for any orbit $\mathcal{O}_x$ of the action, we have $\#G_x = \#G/\#\mathcal{O}_x$. So $\mathcal{O}_x$ must divide evenly into $\#G = p$. Since p is prime, the only possibilities are $\#\mathcal{O}_x = 1$ or p. In the latter case, $\#\mathcal{O}_x \equiv 0 \pmod{p}$. And in the former, $x \in X^g$. Since the orbits partition X by Lemma 6.1.2(a), we have

$$\#X = \sum_{\mathcal{O}} \#\mathcal{O} \equiv \sum_{x \in X^g} 1 = \#X^g \pmod{p}$$

where the first sum is over the distinct orbits of G acting on X. □

We will now use this lemma to prove three well-known congruences. The first is named Fermat's Little Theorem. As described in Burton [**22**, p. 514], Pierre de Fermat stated this theorem in a letter to his friend Frénicle de Bessy in 1640. However, the first published proof seems to be in a 1736 paper of Euler according to Ore's book [**67**, p. 273]. The proof we give here, as well as the one for Wilson's Theorem which follows, are due to Julius Petersen in a paper from 1872 [**68**].

Theorem 6.5.2 (Fermat's Little Theorem)**.** *Let $a \in \mathbb{Z}$ and let p be prime. Then*

$$a^p \equiv a \pmod{p}.$$

Proof. It suffices to prove this for one element out of every congruence class modulo p, so we can assume $a > 0$. Let $X = [p]$, $Y = [a]$ and consider the action of $G = \langle(1, 2, \ldots, p)\rangle$ on Y^X. Of course, we picked this action because $|Y^X| = a^p$. And according to Lemma 6.2.2, the fixed points of $g = (1, 2, \ldots, p)$ are the f which are constant on this cycle. But this means $f(1) = f(2) = \cdots = f(p) \in [a]$. It follows that $\#(Y^X)^g = a$. The congruence now follows from Lemma 6.5.1. □

We next turn to Wilson's Congruence. It was stated around 1000 AD by Ibn al-Haytham; see O'Connor and Robertson [**65**]. In 1770, Waring [**97**] mentioned that the result had been found by his student, Wilson, but neither of them could prove it. A demonstration was give by Lagrange [**56**] one year later. To prove this result, we will need to consider that action of $\mathfrak{S}_n$ on itself by conjugation.

Lemma 6.5.3. *Suppose $\pi, \sigma \in \mathfrak{S}_n$ and let $\tau = \sigma\pi\sigma^{-1}$. Then the cycles of τ are exactly those of the form $(\sigma(i), \sigma(j), \ldots, \sigma(k))$ where $(i, j, \ldots, k)$ is a cycle of π.*

Proof. Since π is a bijection, it suffices to show that if $(i, j, \ldots, k)$ is a cycle of π, then $(\sigma(i), \sigma(j), \ldots, \sigma(k))$ is a cycle of τ. Equivalently, we must demonstrate that if $\pi(i) = j$, then $\tau(\sigma(i)) = \sigma(j)$. But

$$\tau(\sigma(i)) = \sigma\pi\sigma^{-1}(\sigma(i)) = \sigma\pi(i) = \sigma(j)$$

so we are done. □

It is not hard to show, and so left as an exercise, that there is an action of $\mathfrak{S}_n$ itself where $\sigma : \mathfrak{S}_n \to \mathfrak{S}_n$ sends π to $\sigma\pi\sigma^{-1}$. From this definition it is clear that the action can be restricted to any subset of $\mathfrak{S}_n$ which is a union of conjugacy classes. And by the previous result, a conjugacy class consists of all permutations of the same cycle type since given two cycles, one can always construct a σ such that the first is obtained from the second by applying σ to each of its elements. We can also obviously act with a subgroup of $\mathfrak{S}_n$ rather than the whole group.

Theorem 6.5.4 (Wilson's Congruence). *If p is prime, then*

$$(p-1)! \equiv -1 \pmod{p}.$$

Proof. Let $G = \langle\sigma\rangle$ where $\sigma = (1, 2, \ldots, p)$ acts on the conjugacy class X of $\mathfrak{S}_n$ consisting of all p-cycles. By Proposition 4.3.1 we have $\#X = (p-1)!$. It suffices to show that $\#X^\sigma = p - 1$ since then, by Lemma 6.5.1

$$(p-1)! = \#X \equiv \#X^\sigma \equiv -1 \pmod{p}.$$

In fact, we claim that $X^\sigma = \{\sigma^i \mid 0 < i < p\}$. Note that σ^i is in X for $0 < i < p$ since p is prime and so these powers are all p-cycles. Also, $\sigma^i \in X^g$ since $\sigma\sigma^i\sigma^{-1} = \sigma^i$. To show that these are the only elements fixed by σ, suppose $\pi \in X^\sigma$. Since π is a single cycle we must have $\pi(1) = 1 + i$ for some i with $0 < i < p$. We will show that $\pi = \sigma^i$. Since $\pi \in X^\sigma$ we have $\sigma^j\pi\sigma^{-j} = \pi$ for any j. So, by Lemma 6.5.3, π must also send $\sigma^j(1) = 1+j$ to $\sigma^j(1+i) = 1+i+j$ for any j. In particular, π sends $1+i$ to $1+2i$, $1+2i$ to $1+3i$, and so forth, where all values are taken modulo p. It follows that $\pi = \sigma^i$ as desired. □

Our next goal is to prove a congruence of Lucas [**59**]. It gives a way of evaluating a binomial coefficients modulo a prime in terms of the digits in the p-ary expansions of the two arguments. First, we will prove a warm-up result which gives a recursion for the binomial coefficients modulo p. Note how this recurrence relation is the same as the one in Theorem 1.3.3(a) except that every -1 has been replaced by a $-p$.

Lemma 6.5.5. *Let p be prime and let $n \geq p$. We have*

$$\binom{n}{k} \equiv \binom{n-p}{k-p} + \binom{n-p}{k} \pmod{p}.$$

Proof. When $k < 0$ or $k > n$, then it is easy to check that both sides are zero. If $0 \leq k \leq n$, then let $\sigma = (1, 2, \ldots, p)(p+1)(p+2)\ldots(n)$ and consider the action of $G = \langle\sigma\rangle$ on $X = \binom{[n]}{k}$. So $\#X = \binom{n}{k}$. Because of the cycle structure of σ, Lemma 6.3.1(a) implies that $S \in X^\sigma$ if and only if $[p] \subseteq S$ or $[p] \subseteq [n] - S$. In the first case, the number of ways to choose the remaining $k - p$ elements of S from the elements of $[n] - [p]$ is $\binom{n-p}{k-p}$. In the second, we must choose k elements from $[n] - [p]$ to be in S for a total of $\binom{n-p}{k}$ choices. Adding the two counts and using Lemma 6.5.1 finishes the proof. □

Theorem 6.5.6 (Lucas's Congruence). *Let p be prime and let $0 \leq k \leq n$. Consider the base p expansions $n = \sum_{i\geq 0} n_i p^i$ and $k = \sum_{i\geq 0} k_i p^i$ where $0 \leq n_i, k_i < p$ for all i. We have*

$$\binom{n}{k} \equiv \prod_{i\geq 0} \binom{n_i}{k_i} \pmod{p}.$$

Proof. Dividing by p we can write $n = pn' + n_0$ and $k = pk' + k_0$. We will prove that

$$\binom{n}{k} \equiv \binom{n'}{k'}\binom{n_0}{k_0} \pmod p$$

from which Lucas's result follows by induction on n. Let $G = \langle\sigma\rangle$ for the permutation

$$\sigma = \sigma_1\sigma_2 \cdots \sigma_{n'}(pn'+1)(pn'+2)\cdots(n),$$

where $\sigma_i = (p(i-1)+1, p(i-1)+2, \ldots, pi)$. Then G acts on $X = \binom{[n]}{k}$. Since $k_0 < p$, for $S \in X$ to be fixed by σ we must have that S is the union of k_0 elements from $pn'+1, pn'+2, \ldots, n$ together with k' of the cycles σ_i. So the number of ways of choosing the elements of the first type is $\binom{n_0}{k_0}$, and of the second $\binom{n'}{k'}$. This completes the proof. □

To prove congruences with a nonprime modulus, we need to use Möbius inversion. Let G be a group acting on X with G, X finite. Call $x \in X$ *aperiodic* if $G_x = e$, the identity element of G. (For sets of one element we sometimes dispense with the curly braces.) By Lemma 6.1.2(c), this is equivalent to x lying in an orbit of size $\#G$. Since distinct orbits are disjoint by Lemma 6.1.2(a), the number of aperiodic elements is divisible by $\#G$.

Now consider $L(G)$, the lattice of subgroups of G ordered by inclusion. If $H \in L(G)$, then we define two functions

$$\alpha(H) = \#\{x \in X \mid G_x = H\}$$

and

$$\beta(H) = \#\{x \in X \mid G_x \geq H\}.$$

It follows immediately that $\beta(H) = \sum_{K \geq H} \alpha(K)$. Furthermore $\alpha(e)$ is, by definition, the set of aperiodic elements. So, by the previous paragraph, $\alpha(e) \equiv 0 \pmod{\#G}$. Applying the Möbius Inversion Theorem, Theorem 5.5.5(a), and using the fact that $\{e\}$ is the $\hat{0}$ element of $L(G)$, we have proved the following result.

Theorem 6.5.7. *Suppose G acts on X with G, X finite. We have*

$$\sum_{H \in L(G)} \mu(H)\beta(H) \equiv 0 \pmod{\#G}.$$ □

Before applying this result, we note that it has Lemma 6.5.1 as a corollary. Indeed, by Lagrange's Theorem, any $H \leq G$ has $\#H \mid \#G$. So if $\#G = p$ is prime, then $\#H = 1$ or $\#H = p$. It follows that $H = e$ and $H = G$ are the only subgroups of G and $L(G) \cong C_1$, the chain with two elements. Thus $\mu(e) = 1$, $\mu(G) = -1$, and Theorem 6.5.7 becomes

$$\beta(e) - \beta(G) \equiv 0 \pmod p.$$

But $\beta(e) = \#X$ since every $x \in X$ satisfies $G_x \geq e$. Furthermore, $\beta(G) = \#X^g$ where $G = \langle g\rangle$. Indeed, $G_x \geq G$ implies $G_x = G$, which in turn is equivalent to $gx = x$ since $\langle g\rangle = G$. Plugging these values into the previous displayed equation yields Lemma 6.5.1.

We now give an application of Theorem 6.5.7. We first need to characterize $L(G)$ when G is an arbitrary cyclic group.

Proposition 6.5.8. *If $n \in \mathbb{P}$ and $G = \langle(1,2,\ldots,n)\rangle$, then $L(G) \cong D_n$, the lattice of divisors of n.*

Proof. Let $g = (1,2,\ldots,n)$. Since G is cyclic, so is every subgroup. For $d \mid n$ let $H_d = \langle g^d\rangle$ so that $\#H_d = n/d$. We now have a bijection from the set of H_d to D_n given by $H_d \mapsto n/d$. Clearly this map and its inverse are order preserving. So we will be done if we can show that every subgroup of G is one of the H_d.

Suppose $H \le G$. Since H is cyclic, we can write $H = \langle g^d\rangle$ and choose d with $0 \le d < n$ that is minimum over all generators of H. We claim $H = H_d$ which will finish the proof. For this, it suffices to show that $d \mid n$. Suppose, towards a contradiction, that this is not the case. Then dividing n by d gives $n = qd + r$ where $0 \le r < d$. Now $g^r = g^{n-qd} = (g^{-q})^d \in H$. But this contradicts the fact that d was the smallest possible exponent of an element of H. So the proof is complete. □

We can now prove an analogue of Lemma 6.5.5 modulo p^2.

Proposition 6.5.9. *Let p be prime and $n \ge p^2$. We have*

$$\binom{n}{k} \equiv \sum_{i=0}^{p} \binom{p}{i}\binom{n-p^2}{k-ip} \pmod{p^2}.$$

Proof. Let $g = (1,2,\ldots,p^2) \in \mathfrak{S}_n$ where we do not write down any cycles of length one. If $G = \langle g\rangle$, then, by the previous proposition and its proof, $L(G)$ consists of three groups $e, H = \langle g^p\rangle$, and G, with Möbius values $\mu(e) = 1$, $\mu(H) = -1$, and $\mu(G) = 0$. So from Theorem 6.5.7 we have $\beta(e) \equiv \beta(H) \pmod{p^2}$ for any X on which G acts. Let $X = \binom{[n]}{k}$. Then $\beta(e) = \#X = \binom{n}{k}$. So we will be done if we can show that $\beta(H)$ is the right-hand sum in the statement of the proposition.

Note that

$$g^p = (1, 1+p, 1+2p, \ldots)(2, 2+p, 2+2p, \ldots)\cdots(p, 2p, 3p, \ldots)$$

and so consists of p cycles each with p elements. By Lemma 6.3.1(a), $g^pS = S$ if and only if each of these cycles is either entirely in S or in its complement. If i of the cycles are in S, $0 \le i \le p$, then there are $\binom{p}{i}$ ways to choose which cycles. Once these cycles are chosen, we must choose $k - ip$ other elements to be in S from the $n - p^2$ elements of $[n] - [p^2]$. Since this can be done in $\binom{n-p^2}{k-ip}$ ways, we are done. □

6.6. The cyclic sieving phenomenon

As we saw in Chapter 2, one can often express the solution to a counting problem as a sum where the summands have positive and negative coefficients. So one might expect that there are also situations where nth roots of unity come into play for $n > 2$. This is indeed the case with instances of the cyclic sieving phenomenon, so called because these roots of unity form a cyclic group. This concept was introduced by Reiner, Stanton, and White [**72**]. For a survey of such results see [**80**].

Table 6.1. The action of $(1,2,3)$ on $\left(\!\binom{[3]}{2}\!\right)$

$$(1,2,3)11 = 22, \quad (1,2,3)22 = 33, \quad (1,2,3)33 = 11,$$
$$(1,2,3)12 = 23, \quad (1,2,3)13 = 12, \quad (1,2,3)23 = 13.$$

Let G be a multiplicative group with identity element e. If $g \in G$, then we let $o(g)$ be the *order* of g, that is, the smallest positive integer such that $g^{o(g)} = e$. In particular, we will be interested in the cyclic group $\langle e^{2\pi i/n}\rangle \subset \mathbb{C}$ consisting of all nth roots of unity. An nth root of unity ω is *primitive* if $o(\omega) = n$. So the primitive nth roots are exactly those of the form $e^{2k\pi i/n}$ where $\gcd(k,n) = 1$. We will use the notation ω_n for a primitive nth root of unity. We need the following facts.

Lemma 6.6.1. *Let $\omega \neq 1$ be an nth root of unity.*

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

(b) *If ω is primitive, then $1 + \omega + \omega^2 + \cdots + \omega^i \neq 0$ for $0 \leq i < n-1$.*

Proof. We will prove (a) and leave (b) as an exercise. Since ω is an nth root of unity, $\omega^n = 1$ which can be rewritten as

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

Since $\omega \neq 1$ we have $1 - \omega \neq 0$. It follows that the second factor in the displayed equation above is zero, which completes the proof. □

Suppose we are given a cyclic group G, a set X on which G acts, and a polynomial $f(q) \in \mathbb{N}[q]$. The triple $(X, G, f(q))$ exhibits the *cyclic sieving phenomenon* or *CSP* if for all $g \in G$ we have

$$\#X^g = f(\omega_{o(g)}). \tag{6.12}$$

So we can count the fixed points of g by plugging into $f(q)$ a root of unity which has the same order. This is quite surprising since there is, a priori, no promise that substituting a complex number into $f(q)$ will yield an integer, much less that it will count something! Nevertheless, many examples of the CSP have been found and we will explore one in this section. Before we do this, note that a special case of (6.12) is

$$f(1) = \#X^e = \#X.$$

So $f(q)$ will be a q-analogue of $\#X$.

For our running example we will take $G = \langle(1,2,\ldots,n)\rangle$ acting on the set of multisets $X = \left(\!\binom{[n]}{k}\!\right)$ by

$$g\{\{x_1,\ldots,x_k\}\} = \{\{gx_1,\ldots,gx_k\}\}.$$

For ease of notation, in this section we will dispense with the curly braces and commas and just write $x_1 \ldots x_n$ for a multiset with the understanding that this is not a permutation. To be really concrete, let $n = 3$ and $k = 2$. So

$$X = \{11, 22, 33, 12, 13, 23\}.$$

The action of $(1,2,3)$ on X is given in Table 6.1. Recalling Theorem 1.3.4 and the remark at the end of the previous paragraph, a natural choice for our polynomial is

$$f(q) = \left[\begin{array}{c} n+k-1 \\ k \end{array} \right]_q .$$

In the special case under consideration

$$f(q) = \left[\begin{array}{c} 4 \\ 2 \end{array} \right] = 1 + q + 2q^2 + q^3 + q^4.$$

Note that $(1,2,3)$ has the same order as a root $\omega = \omega_3$. And in this case, using Lemma 6.6.1(a),

$$f(\omega) = 1 + \omega + 2\omega^2 + \omega^3 + \omega^4 = (1 + \omega + \omega^2)(1 + \omega^2) = 0 = \#X^{(1,2,3)}$$

where the last equality can be seen from Table 6.1. The rest of this section will be devoted to proving the following result. Another example of the CSP will be found in the exercises.

Theorem 6.6.2. *The cyclic sieving phenomenon is exhibited by the triple*

$$\left(\left(\!\!\binom{[n]}{k}\!\!\right), \langle (1,2,\ldots,n) \rangle, \left[\begin{array}{c} n+k-1 \\ k \end{array} \right]_q \right).$$

Theorem 6.6.2 will be proved by a sequence of results which will permit us to explicitly evaluate the two sides of (6.12). We will start on the left. We first need an analogue of Lemma 6.3.1 for multisets. The *disjoint union* of two multisets is defined by

$$a^{l_a} \ldots c^{l_c} \uplus a^{m_a} \ldots c^{m_c} = a^{l_a+m_a} \ldots c^{l_c+m_c}.$$

Note that this definition also applies to sets, where a set is just a multiset with all multiplicities zero or one. So, for example,

$$123 \uplus 123 \uplus 25^4 = 1^2 2^3 3^2 5^4.$$

The proof of the next lemma is similar to that of Lemma 6.3.1(a) and so is left as an exercise.

Lemma 6.6.3. *Let G act on X with both finite and let $g \in G$. For $M \in \left(\!\binom{X}{k}\!\right)$ we have $gM = M$ if and only if M is a disjoint union of (not necessarily distinct) cycles of g.* □

It is now easy to count fixed points in this situation. To simplify notation, let $C_n = \langle (1,2,\ldots,n) \rangle$.

Corollary 6.6.4. *Let $X = \left(\!\binom{[n]}{k}\!\right)$ and suppose $g \in C_n$ has $o(g) = d$. We have*

$$\#X^g = \begin{cases} \left(\!\!\binom{n/d}{k/d}\!\!\right) & \text{if } d \mid k, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Since g is a power of $(1, 2, \ldots, n)$ its disjoint cycle decomposition must consist of n/i cycles of length i for some $i \mid n$. It follows that if $o(g) = d$, then g is a product of n/d cycles of length d. So if d does not divide k, then a multiset with k elements cannot be written as a disjoint union of cycles of g. Thus Lemma 6.6.3 forces $\#X^g = 0$. On the other hand, if $d \mid k$, then, by the same lemma, the fixed points are those M which are a disjoint union of k/d of the n/d cycles of g. Since cycles can be chosen with repetition, we have now verified the count for $\#X^g$ in this case as well. □

For the right-hand side of (6.12), we need the following lemma.

Lemma 6.6.5. *Suppose $m \equiv n \pmod d$ and $\omega = \omega_d$. We have*

$$\lim_{q\to\omega} \frac{[m]_q}{[n]_q} = \begin{cases} \dfrac{m}{n} & \text{if } d \mid n, \\ 1 & \text{otherwise.} \end{cases} \tag{6.13}$$

Proof. By the assumption about m, n we can write $m = kd + r$ and $n = ld + r$ for some $k, l \in \mathbb{N}$ and $0 \le r < d$. Using the definition of $[n]_q$ in equation (3.2), we see that

$$[n]_q = [r]_q + q^r [d]_q [l]_{q^d} \tag{6.14}$$

where the reader will note the substitution of q^d for q in the last factor. A similar expression holds for $[m]_q$. So if $r \neq 0$, then Lemma 6.6.1 gives $[m]_\omega = [r]_\omega = [n]_\omega$ where $[r]_\omega \neq 0$. The "otherwise" case of the lemma follows immediately. If $r = 0$, then we can use the displayed equation and the fact that $\omega^d = 1$ to write

$$\lim_{q\to\omega} \frac{[m]_q}{[n]_q} = \lim_{q\to\omega} \frac{[d]_q [k]_{q^d}}{[d]_q [l]_{q^d}} = \frac{[k]_1}{[l]_1} = \frac{k}{l} = \frac{m}{n}$$

as desired. □

As a corollary, we can evaluate certain q-binomial coefficients when substituting ω.

Corollary 6.6.6. *Suppose $\omega = \omega_d$ where $d \mid n$. We have*

$$\left[\begin{matrix} n+k-1 \\ k \end{matrix} \right]_\omega = \begin{cases} \dbinom{n/d + k/d - 1}{k/d} & \text{if } d \mid k, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. After canceling $[n-1]_q!$ we have

$$\left[\begin{matrix} n+k-1 \\ k \end{matrix} \right]_q = \frac{[n]_q [n+1]_q \cdots [n+k-1]_q}{[1]_q [2]_q \cdots [k]_q}.$$

Since $d \mid n$ we have, using Lemma 6.6.1 and equation (6.14), that in the product $[n]_\omega [n+1]_\omega \cdots [n+k-1]_\omega$ the first factor and every dth factor after that is zero while the other factors are nonzero. Furthermore, the factors which become zero have ω as a root with multiplicty one. By the same token, in $[1]_\omega [2]_\omega \cdots [k]_\omega$ the zero factors have period d but one starts with $d-1$ nonzero factors, and again each zero factor has ω as a root exactly once. It follows that the number of times ω is a root in the numerator is always greater than or equal to the number in the denominator, with equality if and only if $d \mid k$. So if d does not divide k, then, because the q-binomial coefficient is a polynomial in q, this polynomial will have a factor whose root is ω. Thus the second case of the corollary is proved.

To see what happens when $d \mid k$, we use the previous lemma to obtain

$$\begin{aligned}
\begin{bmatrix} n+k-1 \\ k \end{bmatrix}_\omega &= \lim_{q\to\omega}\left(\frac{[n]_q}{[k]_q}\cdot\frac{[n+1]_q}{[1]_q}\cdot\frac{[n+2]_q}{[2]_q}\cdots\frac{[n+k-1]_q}{[k-1]_q}\right)\\
&= \frac{n}{k}\cdot 1\cdots 1\cdot\frac{n+d}{d}\cdot 1\cdots 1\cdot\frac{n+2d}{2d}\cdot 1\cdots\\
&= \frac{n/d}{k/d}\cdot\frac{n/d+1}{1}\cdot\frac{n/d+2}{2}\cdots\\
&= \binom{n/d+k/d-1}{k/d}
\end{aligned}$$

which is what we wished to prove in this case. □

Comparing Corollaries 6.6.4 and 6.6.6 while remembering Theorem 1.3.4 completes the proof of Theorem 6.6.2. Another way to prove this result using symmetric functions and representation theory will be found in Section 7.9.

Exercises

(1) Complete the proof of Proposition 6.1.1.

(2) (a) Prove that (6.2) satisfies the definition of a group action.
 (b) Consider the action of $G = \langle(1, 2, 3, 4)\rangle$ on the set $P([4], 2)$ of 2-permutations of $[4]$ given by
 $$g(xy) = g(x)g(y).$$
 Compute the orbits and stabilizers of this action and verify that the parts of Lemma 6.1.2 are satisfied.

(3) Show that if G acts on X, then $gx = y$ if and only if $x = g^{-1}y$.

(4) (a) Let G act on X and let Y be another set. For $f \in Y^X$, define $gf = f \circ g$ and show that this definition does not satisfy part (a) of the definition of a group action.
 (b) Let G act on Y and let X be another set. For $f \in Y^X$, define $gf = g \circ f$ and show that this defines a group action on Y^X.

(5) Complete the proof of Lemma 6.2.2(a).

(6) Prove that $4 \mid r^4 + r^2 + 2r$ for all $r \in \mathbb{Z}$ by considering the possible congruence classes of r modulo 4.

(7) (a) Show that the number of distinct n-bead, r-color necklaces under rotation is

$$\frac{1}{n}\sum_{i=1}^{n} r^{\gcd(i,n)} = \frac{1}{n}\sum_{d|n}\phi(n/d)r^d$$

where ϕ is Euler's function. Hint: For the second sum, use the hint for Exercise 24(a) in Chapter 5.

(b) Use part (a) to give two new proofs of the formula obtained in the text when $n = 4$.

(8) (a) The group of symmetries of a regular n-gon is called a *dihedral group* and consists of the n rotations and n reflections which map the n-gon to itself. Find the number of different 4-bead, r-color necklaces if necklaces are considered the same when one is a rotation or reflection of the other.

(b) Find an expression for the number of distinct n-bead, r-color necklaces if two are the same when one is a rotation or a reflection of the other.

(9) (a) How many distinct cubes are there under rotation if the edges are colored from a set with r colors?

(b) Repeat part (a) if you are coloring the vertices.

(10) (a) How many distinct regular tetrahedra are there under rotation if the faces are colored from a set with r colors?

(b) Show in two ways that you get the same answer in (a) if you color vertices: one using Burnside's Lemma and one without using it.

(11) Calculate the cycle index of G acting on X for the following pairs.

(a) $G = \langle(1, 2, \ldots, n)\rangle$ and $X = [n]$.

(b) G is the dihedral group of a regular n-gon (see Exercise 8(a)) and $X = [n]$.

(c) G is the group of rotations of the cube and X is the faces of the cube.

(d) Repeat part (c) for X being the edges and vertices of the cube.

(12) Complete the proof of Lemma 6.3.1(a).

(13) Using the notation of Theorem 6.3.2, give two proofs of each of the following facts about the b_k and p_k, one using their definition in terms of orbits and one using the expression for their generating functions in terms of $Z(G)$.

(a) $b_0 = p_0 = 1$.

(b) $p_n = p_{n-1}$.

(c) $b_n = 1$.

(14) Call a sequence $a_0, \ldots, a_n$ *symmetric* or *palindromic* if $a_k = a_{n-k}$ for all k with $0 \le k \le n$. In this case also call the associated generating function $\sum_{k=0}^{n} a_k t^k$ symmetric or palindromic.

(a) Give three proofs that the sequence $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$ is palindromic: one using (1.5), one inductive, and one combinatorial.

(b) Prove that the product of palindromic unimodal polynomials is palindromic and unimodal.

(c) Use (b) to give another proof of (a).

(d) Prove that the generating function in Theorem 6.3.2(a) is palindromic.

(15) (a) Consider 4-bead r-colored necklaces under rotation. Find the number of distinct necklaces which have 2 beads of one color and 2 beads of another color in two ways: by using Theorem 6.4.2 and by making a direct count.

(b) Consider the cube under rotation where the faces have been colored black and white. Find the generating function for the number of orbits by the number of white and number of black faces in two ways: by using Theorem 6.4.2 and by making a direct count.

(c) Repeat part (b) for coloring the edges and for coloring the vertices.

(16) Find the generating function for unlabeled digraphs on n vertices by the number of arcs.

(17) Let K_n denote the unlabeled complete graph on n vertices.

(a) Find a polynomial $p(r)$ which counts the number of colorings of the edges of K_n with r colors.

(b) Show that $p(2)$ equals the result of plugging $t = 1$ into equation (6.9).

(18) Use the fact that every integer can be written uniquely as a product of primes to show that if $k, l \in \mathbb{P}$, then

$$\operatorname{lcm}(k, l) = \frac{kl}{\gcd(k, l)}.$$

(19) Call $a, b \in \mathbb{Z}$ *relatively prime* if $\gcd(a, b) = 1$. Recall that every integer can be written uniquely as a product of primes.

(a) Show that if $\gcd(a, m) = \gcd(b, m) = 1$, then $\gcd(ab, m) = 1$.

(b) Let $m \in \mathbb{P}$ and let $[a]$ denote the congruence class of a modulo m. Use part (a) to show that

$$G_m = \{[a] \mid \gcd(a, m) = 1\}$$

is a group.

(c) Use part (b) to give two proofs that if p is prime and $\gcd(a, p) = 1$, then

$$a^{p-1} \equiv 1 \pmod{p},$$

one demonstration using Fermat's Little Theorem and one using Lagrange's Theorem from group theory.

(d) Prove Euler's Theorem: if a and n are relatively prime, then

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

where ϕ is the Euler phi function from Exercise 6 of Chapter 2. Hint: Use the ideas in the second proof of part (c).

(e) Show that Fermat's Little Theorem is a special case of part (d).

(20) (a) Use Theorem 6.5.7 to prove that if $a, n \in \mathbb{P}$, then

$$\sum_{d|n} \mu(n/d) a^d \equiv 0 \pmod{n}.$$

(b) Show that Fermat's Little Theorem can be derived from part (a).

(21) (a) Show that the map $\sigma : \mathfrak{S}_n \to \mathfrak{S}_n$ which sends π to $\sigma\pi\sigma^{-1}$ defines an action of $\mathfrak{S}_n$ on itself.

(b) Show that the converse of Wilson's Theorem is true: for $n > 1$ we have that $(n-1)! \equiv -1 \pmod{n}$ implies n is prime.

(22) (a) Let p be prime and let $n \geq p$. Show for the Stirling numbers of the first kind that
$$c(n,k) \equiv c(n-p,k-p) - c(n-p,k-1) \ (\mathrm{mod}\ p).$$
(b) Let p be prime and let $n > kp$. Show that
$$c(n,k) \equiv 0 \ (\mathrm{mod}\ p).$$
Hint: Use part (a).
(c) Let p be prime. Given $n \geq k \geq 0$, write $n = n''p + n'$ where $0 \leq n' < p$ and $k = k''(p-1) + k'$ where $0 \leq k' < p-1$. Then
$$c(n,n-k) \equiv (-1)^{k''}\binom{n''}{k''}c(n',n'-k') \ (\mathrm{mod}\ p).$$
Hint: Use part (a).
(d) Consider two polynomials $f(x), g(x) \in \mathbb{Z}[x]$ and let $m \in \mathbb{P}$. We say that $f(x)$ *is congruent to* $g(x)$ *modulo* m if every coefficient of $f(x) - g(x)$ is divisible by m. If p is prime, show that
$$x\downarrow_p \equiv x^p - x \ (\mathrm{mod}\ p)$$
where $x\downarrow_p = x(x-1)\cdots(x-p+1)$. Hint: Use Theorem 3.1.2 and part (a).

(23) Finish the proof of Theorem 6.5.6

(24) (a) Show that if G acts on X, then it also acts on $S(X,k)$, the set of partitions of X into k blocks.
(b) Let p be prime and let $n \geq p$. Show for the Stirling numbers of the second kind that
$$S(n,k) \equiv S(n-p,k-p) + S(n-p+1,k) \ (\mathrm{mod}\ p).$$
(c) If p is prime and $k \in \mathbb{Z}$, then show
$$S(p,k) \equiv \begin{cases} 1 & \text{if } k = 1 \text{ or } p \\ 0 & \text{otherwise} \end{cases} \quad (\mathrm{mod}\ p).$$
(d) Suppose p is prime and $0 \leq k \leq n$. If j satisfies $p^j \leq k < p^{j+1}$, then
$$S(n + p^j(p-1), k) \equiv S(n,k) \ (\mathrm{mod}\ p).$$
Hint: Use part (b).

(25) (a) Recall from Exercise 10 in Chapter 1 that Pascal's triangle is fractal modulo 2. Give a second proof of this by using Lucas's Congruence (Theorem 6.5.6) to show that if $0 \leq n < 2^m$ and $0 \leq k \leq n + 2^m$, then
$$\binom{n+2^m}{k} \equiv \begin{cases} \binom{n}{k} & \text{if } 0 \leq k \leq n \\ 0 & \text{if } n < k < 2^m \\ \binom{n}{k-2^m} & \text{if } 2^m \leq k \leq n + 2^m \end{cases} \quad (\mathrm{mod}\ 2).$$
(b) Find and prove the analogue of part (a) for an arbitrary prime $p \geq 2$.

(26) Use Proposition 6.5.8 to prove an analogue of Proposition 6.5.9 for any $n \in \mathbb{P}$.

(27) (a) Prove that for any group G the poset $L(G)$ is a lattice. In particular, if $H, K \in L(G)$, then $H \wedge K = H \cap K$ and $H \vee K$ is the subgroup of G generated by H, K.

(b) Let $\mathbb{Z}_p$ be the integers modulo p and consider the direct sum $\mathbb{Z}_p^n$ of n copies of $\mathbb{Z}_p$. Show that if p is prime, then $L(\mathbb{Z}_p^n) \cong L_n(p)$, the subspace lattice of dimension n over the Galois field with p elements.

(c) Use part (b) to prove a congruence modulo p^2 for the binomial coefficients.

(28) Prove Lemma 6.6.1(b).

(29) Prove Lemma 6.6.3.

(30) Let $p \in \mathbb{P}$ and $g = (1, 2, \ldots, p)$. Consider the group $G = \langle g \rangle$ acting on the set of multisets $X = \left(\!\binom{[n]}{k}\!\right)$.

(a) Suppose $M = \{\{1^{m_1}, 2^{m_2}, \ldots, n^{m_n}\}\}$. Show that $gM = M$ if and only if

$$m_1 = m_2 = \cdots = m_p.$$

(b) Show that if p is prime, then

$$\left(\!\!\binom{n}{k}\!\!\right) \equiv \sum_{m \geq 0} \left(\!\!\binom{n-p}{k-mp}\!\!\right) \pmod{p}.$$

(31) (a) Prove that $e^{2k\pi i/n}$ is a primitive nth root of unity if and only if $\gcd(k, n) = 1$.

(b) Prove that the CSP is exhibited by the triple

$$\left(\left(\!\!\binom{[n]}{k}\!\!\right), \langle(1, 2, \ldots, n)\rangle, \begin{bmatrix} n \\ k \end{bmatrix}_q\right).$$

Counting with Symmetric Functions

A formal power series is symmetric if it is invariant under permutation of its variables. We have already seen generating functions of this type arise naturally in Theorem 6.4.2. Symmetric functions have many other connections to combinatorics some of which we will discuss in this chapter. In particular, we will see that they arise when studying log-concavity, Young tableaux, various posets, chromatic polynomials, and the cyclic sieving phenomenon. They are also intimately connected with group representations. The appendix to this book contains a summary of the facts we will need in this regard, and more information can be found in Sagan's text [**79**]. For a wealth of information about symmetric functions in general, see Macdonald's book [**60**].

7.1. The algebra of symmetric functions, Sym

In this section we formally define the algebra of symmetric functions and introduce some of its standard bases. Along the way, we prove the Fundamental Theorem of Symmetric Functions and show how the coefficients of a polynomial can be expressed as a symmetric function of its roots.

Let $\mathbf{x} = \{x_1, x_2, x_3, \dots\}$ be a countably infinite set of commuting variables. Consider the algebra of formal power series $\mathbb{C}[[\mathbf{x}]]$. A monomial $m = x_{i_1}^{\lambda_1} x_{i_2}^{\lambda_2} \cdots x_{i_l}^{\lambda_l}$ has *degree* $\deg m = \sum_i \lambda_i$. For example, $\deg(x_2^5 x_4 x_8^6) = 5 + 1 + 6 = 12$. We say that $f(\mathbf{x}) \in \mathbb{C}[[\mathbf{x}]]$ is *homogeneous of degree n* if $\deg m = n$ for all monomials m appearing in $f(\mathbf{x})$. A weaker condition is that $f(\mathbf{x})$ have *bounded degree* which means that there is an n with $\deg m \le n$ for all m appearing in $f(\mathbf{x})$. To illustrate, $f(\mathbf{x}) = \sum_{i<j} x_i x_j^2$ is homogeneous of degree 3. On the other hand

$$f(\mathbf{x}) = \prod_{i \ge 1}(1 + x_i) \tag{7.1}$$

is not of bounded degree.

There is an action of $\mathfrak{S}_m$ on $\mathbb{C}[[\mathbf{x}]]$. Specifically, if we have $\pi \in \mathfrak{S}_m$ and $f(x_1, x_2, x_3, \dots) \in \mathbb{C}[[\mathbf{x}]]$, then we let

$$\pi f(x_1, x_2, x_3, \dots) = f(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)}, \dots) \tag{7.2}$$

where $\pi(i) = i$ for $i > m$. For example,

$$(1,2)(x_1^2 + 2x_1x_2^3 + 5x_1^4x_3 - x_3x_4) = x_2^2 + 2x_1^3x_2 + 5x_2^4x_3 - x_3x_4. \tag{7.3}$$

Call $f(\mathbf{x})$ *symmetric* if $\pi f = f$ for all π in every symmetric group $\mathfrak{S}_m$. Equivalently, any two monomials $x_{i_1}^{\lambda_1} x_{i_2}^{\lambda_2} \cdots x_{i_l}^{\lambda_l}$ and $x_{j_1}^{\lambda_1} x_{j_2}^{\lambda_2} \cdots x_{j_l}^{\lambda_l}$, where $i_1, \dots, i_l$ are distinct and similarly for $j_1, \dots, j_l$, have the same coefficient in $f(\mathbf{x})$ since one can always find a permutation π such that $\pi(i_k) = j_k$ for all k. Another equivalent description is that any monomial $x_{i_1}^{\lambda_1} x_{i_2}^{\lambda_2} \cdots x_{i_l}^{\lambda_l}$ has the same coefficient as $x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_l^{\lambda_l}$ in $f(\mathbf{x})$. To illustrate,

$$f(\mathbf{x}) = 4x_1^5 + 4x_2^5 + 4x_3^5 + \cdots - 6x_1^2x_2^2 - 6x_1^2x_3^2 - 6x_2^2x_3^2 - \cdots \tag{7.4}$$

is symmetric. Let

$$\begin{aligned} \operatorname{Sym}_n &= \operatorname{Sym}_n(\mathbf{x}) \\ &= \{f \in \mathbb{C}[[\mathbf{x}]] \mid f \text{ is symmetric and homogeneous of degree } n\}. \end{aligned}$$

The *algebra of symmetric functions* is

$$\operatorname{Sym} = \operatorname{Sym}(\mathbf{x}) = \bigoplus_{n \geq 0} \operatorname{Sym}_n(\mathbf{x}).$$

Alternatively, $\operatorname{Sym}(\mathbf{x})$ is the set of all symmetric power series in $\mathbb{C}[[\mathbf{x}]]$ of bounded degree. This is because elements of the direct sum can only have a finite number of components which are nonzero. So the series in (7.4) is in Sym, but the ones in (7.1) and (7.3) are not.

There are a number of interesting bases for Sym. We start with those functions obtained by symmetrizing a monomial. If $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_l)$ is a partition, then the associated *monomial symmetric function* is

$$m_\lambda = m_\lambda(\mathbf{x}) = \sum x_{i_1}^{\lambda_1} x_{i_2}^{\lambda_2} \cdots x_{i_l}^{\lambda_l}$$

where the sum is over all distinct monomials having exponents $\lambda_1, \lambda_2, \cdots, \lambda_l$. We will often drop the parentheses and commas in the subscript λ as well as use multiplicity notation. As examples, in (7.4) we have $f = 4m_{(5)} - 6m_{(2,2)} = 4m_5 - 6m_{2^2}$, and

$$m_{21} = x_1^2x_2 + x_1x_2^2 + x_1^2x_3 + x_1x_3^2 + x_2^2x_3 + x_2x_3^2 + \cdots. \tag{7.5}$$

We must verify that we have defined a basis for Sym.

Theorem 7.1.1. *The m_λ as λ varies over all partitions form a basis for* Sym. *Consequently*

$$\dim \operatorname{Sym}_n = p(n),$$

the number of partitions of n.

Proof. The second sentence follows immediately from the first. And it is clear that the m_λ are independent since no two contain a monomial in common. So it suffices to show that every $f \in \operatorname{Sym}$ can be written as a linear combination of the m_λ. Suppose $x_{i_1}^{\lambda_1} x_{i_2}^{\lambda_2} \cdots x_{i_l}^{\lambda_l}$ is a monomial appearing in f and having coefficient $c \in \mathbb{C}$. Without loss of generality we can assume the indices have been arranged so that $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_l$ and let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$. Since f is symmetric, every monomial in f with exponents $\lambda_1, \lambda_2, \ldots, \lambda_l$ appears with coefficient c. So $f - cm_\lambda$ is still symmetric and contains no monomials with these exponents. Since f is of bounded degree, we can repeat this process a finite number of times until we reach the zero power series. It follows that f will be a linear combination of the monomial symmetric functions which appear during this algorithm. □

There are three bases which are formed multiplicatively in that one first defines f_n for $n \in \mathbb{P}$ and then sets

$$f_\lambda = f_{\lambda_1} f_{\lambda_2} \cdots f_{\lambda_l} \tag{7.6}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$. Specifically, for $n \ge 1$ we define the *nth power sum symmetric function*

$$p_n = m_{(n)} = \sum_{i \ge 1} x_i^n,$$

the *nth elementary symmetric function*

$$e_n = m_{(1^n)} = \sum_{i_1 < \cdots < i_n} x_{i_1} \cdots x_{i_n},$$

and the *nth complete homogeneous symmetric function*

$$h_n = \sum_{\lambda \vdash n} m_\lambda = \sum_{i_1 \le \ldots \le i_n} x_{i_1} \cdots x_{i_n}.$$

We also let $e_0 = h_0 = 1$ because of the empty product. To illustrate, when $n = 3$ we have

$$\begin{aligned}
p_3 &= x_1^3 + x_2^3 + x_3^3 + \cdots, \\
e_3 &= x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + \cdots, \\
h_3 &= x_1^3 + x_2^3 + \cdots + x_1^2x_2 + x_1x_2^2 + \cdots + x_1x_2x_3 + x_1x_2x_4 + \cdots.
\end{aligned}$$

Note that we have already met the power sum symmetric functions p_n since they occurred as the substitutions made for the variables of the cycle index polynomial in Theorem 6.4.2. Also note that e_n can be thought of as the sum of all square-free monomials of degree n, while h_n is the sum of all monomials of degree n. We now define p_λ, e_λ, and h_λ using (7.6). So, for example,

$$p_{(4,2)} = (x_1^4 + x_2^4 + x_3^4 + \cdots)(x_1^2 + x_2^2 + x_3^2 + \cdots).$$

To show that these are bases for Sym, it will be helpful to use generating functions. Define the following elements of $\mathbb{C}[[\mathbf{x}, t]]$:

$$\begin{aligned}
P(t) &= \sum_{n \ge 1} p_n(\mathbf{x}) \frac{t^n}{n}, \\
E(t) &= \sum_{n \ge 0} e_n(\mathbf{x}) t^n, \\
H(t) &= \sum_{n \ge 0} h_n(\mathbf{x}) t^n.
\end{aligned}$$

Note that $E(t)$ and $H(t)$ are ogfs, while we have not dealt with a generating function like $P(t)$ previously.

Proposition 7.1.2. *We have the following identities.*

(a) $E(t) = \prod_{i\ge1}(1 + x_i t)$.

(b) $H(t) = \prod_{i\ge1} \frac{1}{1 - x_i t}$.

(c) $P(t) = \ln \prod_{i\ge1} \frac{1}{1 - x_i t}$.

Proof. (a) We will use weight-generating functions where the set is the same one used in the proof of Theorem 3.5.5; namely S is all partitions λ with distinct parts. We weight $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in S$ by

$$\operatorname{wt}\lambda = t^n x_{\lambda_1} x_{\lambda_2} \cdots x_{\lambda_n}.$$

Since e_n is the sum of all square-free monomials of degree n, we have the weight-generating function

$$f_S(\mathbf{x}, t) = \sum_{\lambda\in S} \operatorname{wt}\lambda = \sum_{n\ge0} t^n \sum_{l(\lambda)=n} x_{\lambda_1} x_{\lambda_2} \cdots x_{\lambda_n} = \sum_{n\ge0} e_n(\mathbf{x}) t^n$$

where $\ell(\lambda)$ is λ's length. On the other hand, we have the decomposition of S in the demonstration of Theorem 3.5.5

$$S = (\{1^0\} \uplus \{1^1\}) \oplus (\{2^0\} \uplus \{2^1\}) \oplus (\{3^0\} \uplus \{3^1\}) \oplus \cdots.$$

Applying the Sum and Product Rules for weight-generating functions gives

$$f_S(\mathbf{x}, t) = (1 + x_1 t)(1 + x_2 t)(1 + x_3 t)\cdots$$

so we are done.

(b) This proof is similar to the one for (a) and so is left to the reader.

(c) Using the expansion of $\ln \frac{1}{1-x}$ gives

$$\ln \prod_{i\ge1} \frac{1}{1 - x_i t} = \sum_{i\ge1} \ln \frac{1}{1 - x_i t} = \sum_{i\ge1}\sum_{n\ge1} \frac{(x_i t)^n}{n} = \sum_{n\ge1} \frac{t^n}{n} \sum_{i\ge1} x_i^n = \sum_{n\ge1} p_n(\mathbf{x}) \frac{t^n}{n}$$

as desired. □

In order to prove that the p_λ and e_λ are bases, we will want to encode their expressions as linear combinations of the m_λ. For that we will need a total order on the $\lambda \vdash n$ to index the rows and columns of the corresponding matrix. Say that $(\lambda_1, \ldots, \lambda_l) < (\mu_1, \ldots, \mu_k)$ in *lexicographic order* if, for the smallest index i where λ and μ differ, we have $\lambda_i < \mu_i$.

Theorem 7.1.3. *We have the following bases for* Sym_n.

(a) $\{p_\lambda \mid \lambda \vdash n\}$.

(b) $\{e_\lambda \mid \lambda \vdash n\}$.

(c) $\{h_\lambda \mid \lambda \vdash n\}$.

Proof. (a) The set of p_λ has cardinality $p(n) = \dim \mathrm{Sym}_n$ by Theorem 7.1.1. So we only need to show that the p_λ span Sym_n. Express each p_λ as a linear combination of the m_μ basis and let $A = [a_{\lambda,\mu}]$ be the matrix of coefficients where the rows and columns are listed in lexicographic order. It suffices to show that A is upper triangular with nonzero diagonal elements. For then A^{-1} exists and so we can write each m_μ in terms of the p_λ. So consider a monomial $m = x_1^{\mu_1} \cdots x_k^{\mu_k}$ occurring in the expansion of

$$p_\lambda = (x_1^{\lambda_1} + x_2^{\lambda_1} + \cdots)(x_1^{\lambda_2} + x_2^{\lambda_2} + \cdots) \cdots .$$

(By symmetry, our choice of subscripts for m is without loss of generality.) It follows that each μ_i is a sum of λ_j. But it is easy to see that adding parts of a partition make it larger in lexicographic order. So m_λ will have smallest subscript if it occurs. But we can obtain the given monomial by picking $x_1^{\lambda_1}$ out of the first factor, $x_2^{\lambda_2}$ out of the second, and so forth. So the proof is complete.

(b) This demonstration is similar to the one in part (a) except that one shows

$$e_{\lambda^t} = m_\lambda + \sum_{\mu<\lambda} b_{\lambda,\mu} m_\mu$$

where λ^t is the conjugate of λ.

(c) It suffices to show that each e_n can be written as a polynomial in the h_k. Indeed, by multiplicativity this implies that the e_μ are linearly spanned by the h_λ. And since the number of h_λ is the dimension of Sym_n, they must form a basis.

From Proposition 7.1.2 we see that

$$H(t)E(-t) = 1.$$

Taking the coefficient of t^n on both sides for $n \geq 1$ yields

$$\sum_{i=0}^{n} (-1)^i h_{n-i} e_i = 0.$$

Solving for e_n gives

$$e_n = h_1 e_{n-1} - h_2 e_{n-2} + \cdots .$$

By induction on n, the e_i on the right in this sum are polynomials in the h_k, so the same is true of e_n. □

We note that (b) of the previous theorem is sometimes called the Fundamental Theorem of Symmetric Functions and is expressed in the following manner: any symmetric function can be written uniquely as a polynomial in the e_n. Also note that since the triangular transition matrix from the m_λ to the e_λ in (b) has ones down the diagonal, we have actually proved that any symmetric function with integer coefficients is in fact an integral linear combination of the e_λ.

We wish to examine a corollary of Proposition 7.1.2(a) which shows that the coefficients of a polynomial can be expressed as elementary symmetric functions of its roots. This will be useful for proving a log-concavity result in the next section. This result is well known for quadratic polynomials (and follows easily from the quadratic formula) but holds in general. In what follows we will specialize our symmetric functions to the first m variables by setting $x_i = 0$ for $i > m$. For example,

$$e_2(x_1, x_2, x_3, x_4) = x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4.$$

Lemma 7.1.4. *Let*

$$f(t) = a_0t^n + a_1t^{n-1} + a_2t^{n-2} + \cdots + a_n \tag{7.7}$$

be a monic polynomial (so $a_0 = 1$*) with complex coefficients. Let the roots of* $f(t)$ *be* $r_1, \ldots, r_n$*. Then for all* $k \geq 0$ *we have*

$$a_k = e_k(-r_1, -r_2, \ldots, -r_n).$$

Proof. From the definitions

$$f(t) = (t - r_1)(t - r_2)\cdots(t - r_n).$$

Using this and Proposition 7.1.2(a) gives

$$t^n f(1/t) = (1 - r_1t)(1 - r_2t)\cdots(1 - r_nt) = \sum_{k\geq 0} e_k(-r_1, -r_2, \ldots, -r_n)t^k.$$

On the other hand, because of (7.7),

$$t^n f(1/t) = a_nt^n + \cdots + a_1t + 1.$$

Comparing the last two displayed equations finishes the proof. □

7.2. The Schur basis of Sym

There is an important basis for Sym_n whose elements are called the Schur functions. To construct these functions, we will use certain tableaux built out of Young diagrams. Expressing the Schur functions in the elementary, complete homogeneous, and power sum bases will lead to interesting connections with the Lindström–Gessle–Viennot technique and representations of symmetric groups. This will also permit us to prove a result noted in Section 5.6 relating log-concavity to the roots of the corresponding generating polynomial.

Let λ be a partition of n. A *standard Young tableau* (*SYT*) *of shape* λ is a bijective filling T of the boxes of the Young diagram for λ with the elements of $[n]$ such that rows increase from left to right and columns increase from top to bottom. We let

$$\mathrm{SYT}(\lambda) = \{T \mid T \text{ is a standard Young tableau of shape } \lambda\}$$

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 4 \\ \hline 5 \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 5 \\ \hline 4 \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 4 \\ \hline 5 \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 5 \\ \hline 4 \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|}\hline 1 & 4 \\ \hline 2 & 5 \\ \hline 3 \\ \cline{1-1}\end{array}$$

Figure 7.1. The standard Young tableaux of shape $\lambda = (2, 2, 1)$

$$
\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|}
\hline
1 & 1 & 1 & 1 & \underline{2} & \underline{2} & \overline{2} & \overline{3} & \overline{3} & \overline{3} & \overline{3} \\
\hline
\underline{2} & \overline{2} & \overline{2} & \overline{3} & \underline{3} & \underline{3} & 4 & 4 \\
\cline{1-8}
\underline{3} & 4 & 6 \\
\cline{1-3}
\end{array}
\quad \xrightarrow{\iota}
$$

$$
\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|}
\hline
1 & 1 & 1 & 1 & \underline{2} & \underline{2} & \overline{2} & \overline{2} & \overline{2} & \overline{2} & \overline{3} \\
\hline
\underline{2} & \overline{2} & \overline{3} & \overline{3} & \underline{3} & \underline{3} & 4 & 4 \\
\cline{1-8}
\underline{3} & 4 & 6 \\
\cline{1-3}
\end{array}
$$

Figure 7.2. Two semistandard Young tableaux illustrating a Bender–Knuth interchange

and

$$f^\lambda = \#\,\mathrm{SYT}(\lambda).$$

We also write sh T for the shape of T and call T a *standard λ-tableau* if sh $T = \lambda$. The SYT of shape $\lambda = (2, 2, 1)$ are listed in Figure 7.1 so $f^{(2,2,1)} = 5$.

A *semistandard Young tableau* (*SSYT*) *of shape* λ is a filling T of the boxes of the Young diagram for λ with elements of $\mathbb{P}$ such that rows weakly increase and columns strictly increase. As expected, we let

$$\mathrm{SSYT}(\lambda) = \{T \mid T \text{ is a semistandard Young tableau of shape } \lambda\}$$

and call the elements of this set *semistandard λ-tableaux*. Two semistandard Young tableaux of shape $(11, 8, 3)$ are displayed in Figure 7.2. (The reader should ignore the underlines and overlines for now.) We denoted by $c = (i, j)$ the square, also called a *cell*, in row i and column j of the Young diagram of λ where rows and columns are indexed as in a matrix. The entry in cell (i, j) of T is denoted $T_{i,j}$. For example, the first tableau in Figure 7.2 has $T_{2,7} = 4$. The *content* of an SSYT T is the weak composition $\alpha = \mathrm{co}\, T$ where α_i is the number of occurrences of i in T. The first tableau in Figure 7.2 has $\mathrm{co}\, Y = [4, 6, 8, 3, 0, 1]$. Strictly speaking, one could add as many zeros as one liked to the end of co T, but usually we will terminate the composition with a positive entry. The *Kostka numbers* are

$$K_{\lambda,\alpha} = \#\{T \in \mathrm{SSYT}(\lambda) \mid \mathrm{co}\, T = \alpha\}.$$

If we wish to use a content which is a partition μ and not just a weak composition, we will write $K_{\lambda,\mu}$. Note that if $\lambda \vdash n$, then $K_{\lambda,(1^n)} = f^\lambda$.

To define the Schur functions, we will weight a $T \in \mathrm{SSYT}(\lambda)$ by letting

$$\mathbf{x}^T = \prod_{(i,j)\in\lambda} x_{T_{i,j}}.$$

Note that if $\mathrm{co}\, T = [\alpha_1, \alpha_2, \ldots, \alpha_k]$, then $\mathbf{x}^T = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$. The tableau on the left in Figure 7.2 has $\mathbf{x}^T = x_1^4 x_2^6 x_3^8 x_4^3 x_6$. The *Schur function* corresponding to a partition λ is

$$s_\lambda = \sum_{T\in\mathrm{SSYT}(\lambda)} \mathbf{x}^T.$$

For example, if $\lambda = (2,1)$, then a partial list of the semistandard tableaux of shape λ is

$$T: \begin{array}{|c|c|}\hline 1&1\\ \hline 2 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1&2\\ \hline 2 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1&1\\ \hline 3 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1&3\\ \hline 3 \\ \cline{1-1}\end{array} \quad \cdots \quad \begin{array}{|c|c|}\hline 1&2\\ \hline 3 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1&3\\ \hline 2 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1&2\\ \hline 4 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 1&4\\ \hline 2 \\ \cline{1-1}\end{array} \quad \cdots$$

so that

$$s_{(2,1)} = x_1^2x_2 + x_1x_2^2 + x_1^2x_3 + x_1x_3^2 + \cdots + 2x_1x_2x_3 + 2x_1x_2x_4 + \cdots.$$

Note that it is not obvious from the definition that s_λ is even symmetric, but we will prove this shortly. As special cases, if λ is a single row, then the corresponding T are just a weakly increasing sequences of integers so that

$$s_{(n)} = h_n.$$

Similarly, if λ is a single column, then the T are strictly increasing sequences, which gives

$$s_{(1^n)} = e_n.$$

We will see generalizations of these equations shortly when we study the Jacobi–Trudi Determinants. For now, we must show that $s_\lambda \in \mathrm{Sym}$. We will use a clever combinatorial involution of Bender and Knuth [**6**] for the proof.

Proposition 7.2.1. *The function $s_\lambda(\mathbf{x})$ is symmetric.*

Proof. Since the adjacent transpositions generate the symmetric group, it suffices to show that

$$(i, i+1)s_\lambda(\mathbf{x}) = s_\lambda(\mathbf{x})$$

where the action is the one given by (7.2). To do this, we will define an involution $\iota\colon \mathrm{SSYT}(\lambda) \to \mathrm{SSYT}(\lambda)$ such that if $\iota(T) = T'$, then the number of i's and $(i+1)$'s are exchanged in passing from T to T' while all other entries are unchanged. If a column of T contains both i and $i+1$, then such pairs are called *fixed*. All other entries equal to i or $i+1$ are called *free*. See Figure 7.2 where $i = 2$, the fixed entries are underlined, and the free ones are overlined. The map ι takes each row containing k free i's followed by l free $(i+1)$'s and replaces these entries by l free i's followed by k free $(i+1)$'s. This clearly preserves the weakly increasing condition on the rows. And the columns are still strictly increasing because of the definition of free. Also, the number of i's and $(i+1)$'s are interchanged since this is true for the free elements by construction and the fixed elements come in pairs. Finally, it is clear that this map is its own inverse and hence an involution. This finishes the proof. $\square$

Theorem 7.2.2. *For $\lambda \vdash n$ we have*

$$s_\lambda = \sum_{\mu \le \lambda} K_{\lambda,\mu} m_\mu$$

where the sum is over partitions μ of n and $K_{\lambda,\lambda} = 1$. So the set

$$\{s_\lambda \mid \lambda \vdash n\}$$

is a basis for Sym_n.

Proof. The second sentence follows from the first in the same way as the proof of Theorem 7.1.3(a). The fact that $K_{\lambda,\mu}$ is the coefficient of m_μ in the expansion of s_λ comes from the previous proposition and the definitions of s_λ and $K_{\lambda,\mu}$. Clearly there is only one element of $\mathrm{SSYT}(\lambda, \lambda)$, namely the tableau whose ith row consists completely of i's for all $i \geq 1$. So it remains to show that if $K_{\lambda,\mu} \neq 0$, then $\mu \leq \lambda$. Suppose $\lambda \neq \mu$ since equality has already been considered, and pick $T \in \mathrm{SSYT}(\lambda, \mu)$. Let j be the first index where $\lambda_j \neq \mu_j$. Then for $i < j$, the ith row of T is all i's. It follows from column strictness that the jth row must contain all the j's. So

$$\mu_j = \text{number of } j\text{'s} < \text{number of boxes in row } j = \lambda_j$$

as desired. $\square$

We now wish to find the expansion of s_λ in the elementary and complete homogeneous bases. These are best expressed as determinants which were discovered by Jacobi [**44**] and subsequently simplified by his student Trudi [**94**]. For the proof, we will use a weighted version of the Lindström–Gessle–Viennot Lemma, Lemma 2.5.4. In it, all cardinalities are replaced by the corresponding weight-generating function over the set being counted. The only extra information which needs to be checked in this case is that the involution Ω in (2.12) is weight preserving in the sense that $\mathrm{wt}(P_i, P_j) = \mathrm{wt}(P_i', P_j')$ where, as usual, the weight of a Cartesian product is the product of the weights.

Theorem 7.2.3 (Jacobi–Trudi Determinants)**.** *Suppose* $\lambda = (\lambda_1, \dots, \lambda_l)$.

(a) $s_\lambda = \det[h_{\lambda_i - i + j}]_{1 \leq i,j \leq l}$.

(b) $s_{\lambda^t} = \det[e_{\lambda_i - i + j}]_{1 \leq i,j \leq l}$.

Before beginning the proof, we note that a good way of remembering the subscripts in these determinants is to put the parts of λ down the diagonal and then in each row add 1 or subtract 1 as one moves right or left, respectively. So, for example,

$$s_{(7,4,1)} = \det\begin{bmatrix} h_7 & h_8 & h_9 \\ h_3 & h_4 & h_5 \\ h_{-1} & h_0 & h_1 \end{bmatrix} = \det\begin{bmatrix} h_7 & h_8 & h_9 \\ h_3 & h_4 & h_5 \\ 0 & 1 & h_1 \end{bmatrix}.$$

Proof. (a) We will use northeast lattice paths P in the extension of the integer lattice $\mathbb{Z}^2$ obtained by adding a vertex (i, ∞) for each $i \in \mathbb{Z}$. One can only reach (i, ∞) by taking an infinite number of north steps along the line $x = i$. We label the east steps of $P : s_1 s_2 s_3 \dots$ by letting $L(s_i)$ be the y-coordinate of s_i if $s_i = E$. See, for example, the path on the left in Figure 7.3 where we are assuming the path starts at the point $(1, 1)$. If P only has a finite number of east steps all on or above $y = 1$, we weight it by

$$\mathrm{wt}\, P = \prod_{s_i} x_{L(s_i)}$$

where the product is over all s_i which are east steps of P. In Figure 7.3 we have $\mathrm{wt}\, P = x_1 x_3^2 x_4$.

Now let $u = (i, 1)$ and $v = (i + n, \infty)$ where $n \geq 0$. Then all $P \in \mathcal{P}(u; v)$ have exactly n east steps. Furthermore, as P varies over all elements of $\mathcal{P}(u; v)$ we see that

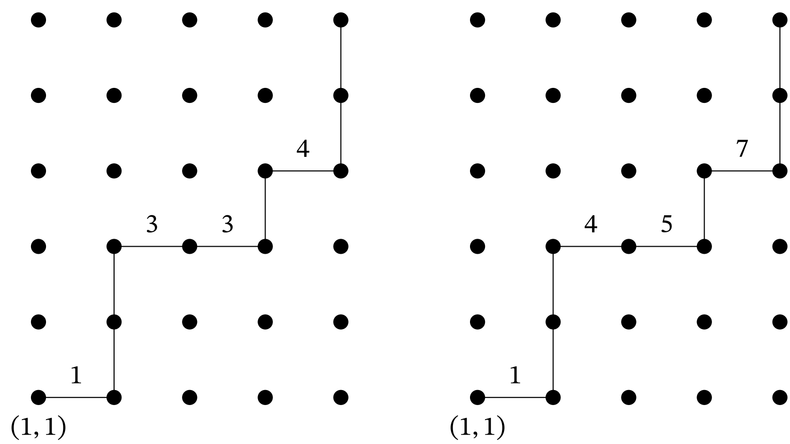


Figure 7.3. A path with the h-labeling on the left and the e-labeling on the right

$\operatorname{wt} P$ varies over all products $x_{j_1}x_{j_2}\cdots x_{j_n}$ with $1 \le j_1 \le j_2 \le \cdots \le j_n$. It follows that $\operatorname{wt}\mathcal{P}(u;v) = h_n$. To apply Lemma 2.5.4, let the initial and final vertices be

$$u_i = (1-i, 1) \quad \text{and} \quad v_i = (\lambda_i - i + 1, \infty)$$

for $i \in [l]$. See Figure 7.4 for an example where $\lambda = (3,3,1)$. With this choice of vertices, the weighted entries of the Lindström–Gessle–Viennot matrix give (up to transposition which does not affect the determinant) those on the right in part (a) of this theorem. Indeed, if P goes from u_i to v_j, then it has

$$(\lambda_j - j + 1) - (1 - i) = \lambda_j + i - j \tag{7.8}$$

east steps so that the set of such paths has weight h_{λ_j+i-j}. Also note that since the weight of a step only depends upon its height, the map Ω is weight preserving. We also need to show that any path family P whose associated permutation is not the identity must be intersecting. This will be left as an exercise.

Finally, to complete the proof, we merely need to show that the weight-generating function for the nonintersecting path families is s_λ. For this it suffices to give a weight-preserving bijection from such paths to $\mathrm{SSYT}(\lambda)$. Map such a path family $(P_1, \ldots, P_l)$ to the tableau T whose ith row consists of the labels on P_i read left to right. An example is in Figure 7.4. Since P_i goes from u_i to v_i it has λ_i east steps by (7.8), so T has shape λ. Further, the definition of the map and the labeling of the paths show that the rows are weakly increasing. To show that the columns are strictly increasing, we need to check that for all i and j, the jth step on P_i is lower than the jth step on P_{i+1}. But this is forced by the nonintersecting condition. It is easy to describe an inverse sending a tableau back to a path family, so we leave this detail to the reader.

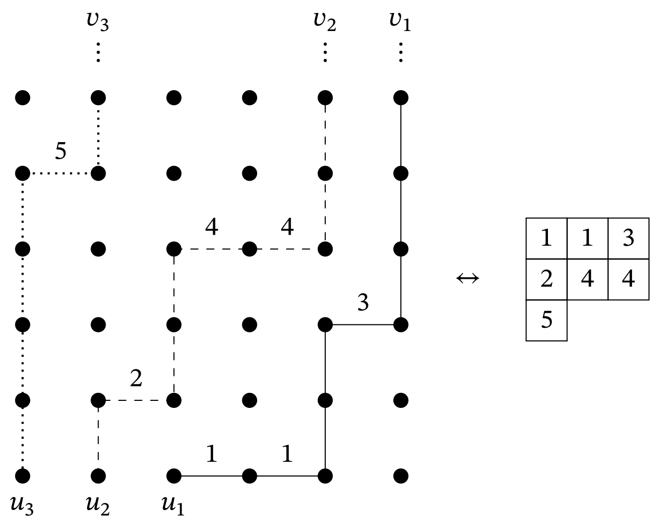


Figure 7.4. Nonintersecting paths and the associated semistandard Young tableau

(b) The proof is similar to that of (a) except that we label the east steps of P by $L'(s_i) = i$. See the path on the right in Figure 7.3 for an example. The reader will find it a good exercise to fill in the rest of the proof. □

The expansion of s_λ in the power sum basis is also important. But the proof is beyond the scope of this book. See [**79**, Theorem 4.6.4] for a demonstration of the next result. In it, we let $p_\pi = p_\lambda$ if $\pi \in \mathfrak{S}_n$ has cycle type λ. Also, χ^λ is the character of the irreducible representation of $\mathfrak{S}_n$ corresponding to λ.

Theorem 7.2.4. *If $\lambda \vdash n$, then*

$$s_\lambda = \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \chi^\lambda(\pi) p_\pi. \qquad \square$$

We now have the tools we need to prove a result postponed from Section 5.6.

Theorem 7.2.5. *Let $a_0, a_1, \ldots, a_n$ be a sequence of real numbers with generating function $f(t) = \sum_{k\geq 0} a_k t^k$. If $f(t)$ has only real roots none of which are positive, then the sequence is log-concave.*

Proof. Clearly $a_0, a_1, \ldots, a_n$ is log-concave if and only if $a_n, \ldots, a_1, a_0$ is. And $f(t)$ has only real, nonpositive roots if and only if $g(t) = t^n f(1/t) = \sum_{k\geq 0} a_{n-k} t^k$ does. We can also assume, without loss of generality, that $a_0 \neq 0$. Now write $g(t) = a_0 h(t)$ where $h(t)$ is monic and has the same roots $r_1, \ldots, r_n$ as $g(t)$. So applying Lemma 7.1.4 to $h(t)$ and multiplying by a_0 we get

$$a_k = a_0\, e_k(-r_1, \ldots, -r_n)$$

for $0 \le k \le n$. Now let $\lambda = (k,k)$ and use Theorem 7.2.3(b) to obtain

$$\begin{aligned}
a_k^2 - a_{k-1}a_{k+1} &= \det\begin{bmatrix} a_k & a_{k+1} \\ a_{k-1} & a_k \end{bmatrix} \\
&= \det\begin{bmatrix} a_0\, e_k(-r_1,\dots,-r_n) & a_0\, e_{k+1}(-r_1,\dots,-r_n) \\ a_0\, e_{k-1}(-r_1,\dots,-r_n) & a_0\, e_k(-r_1,\dots,-r_n) \end{bmatrix} \\
&= a_0^2\, s_{(k,k)^t}(-r_1,\dots,-r_n).
\end{aligned}$$

Since $s_\lambda(\mathbf{x})$ has nonnegative coefficients and $-r_i \ge 0$ for all i, we have $a_k^2 - a_{k-1}a_{k+1} \ge 0$, which is what we wished to show. □

7.3. Hooklengths

In this section we will derive formulae for counting standard Young tableaux and semi-standard Young tableaux of a given shape. These expressions will be based on the sizes of certain subsets of the Young diagram called hooks.

Given a Young diagram λ, a cell $c = (i,j) \in \lambda$ has *hook*

$$H_c = H_{i,j} = \{(i',j) \in \lambda \mid i' \ge i\} \cup \{(i,j') \in \lambda \mid j' \ge j\}.$$

The partition $\lambda = (7,7,6,6,4)$ is displayed on the left in Figure 7.5 and the cells in the hook $H_{2,3}$ are marked with dotted lines. The *hooklength* of $c = (i,j)$ is

$$h_c = h_{i,j} = \#H_c.$$

Using the previous example, we have $h_{2,3} = 8$. And on the right in the same figure, we have displayed the hooklengths of all the cells for the shape $(2,2,1)$. It is not hard to show that

$$h_{i,j} = \lambda_i + \lambda_j^t - i - j + 1. \tag{7.9}$$

There is a beautiful formula for the number of standard Young tableau of given shape due to Frame, Robinson, and Thrall [**29**]. We will give a probabilistic proof of

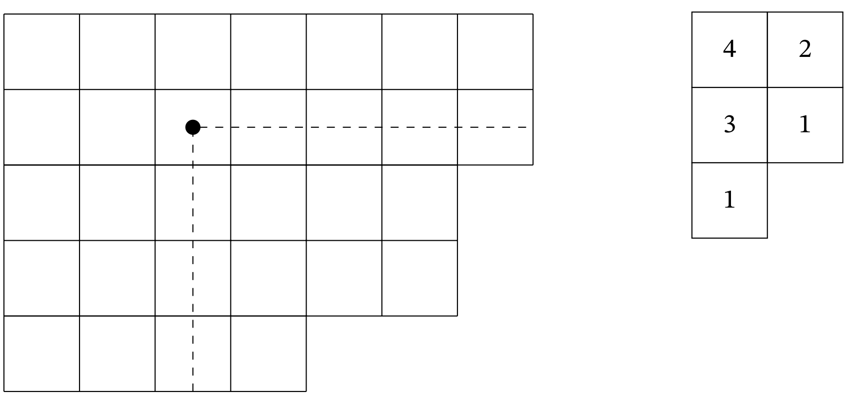


Figure 7.5. The hook $H_{2,3}$ in $\lambda = (7^2, 6^2, 4)$ and the hooklengths of $\lambda = (2^2, 1)$

this result discovered by Greene, Nijenhuis, and Wilf [**35**], which has the added benefit of providing an algorithm for choosing an SYT of shape λ uniformly at random. For the demonstration we will need the concept of an *inner corner* of a Young diagram λ which is a cell c at the end of its row and column. Equivalently $h_c = 1$. The inner corners of the $\lambda = (7^2, 6^2, 4)$ in Figure 7.5 are $(2,7)$, $(4,6)$, and $(5,4)$. Note that in any SYT of shape $\alpha \vdash n$ one must have n in one of the inner corners of λ.

Theorem 7.3.1 (Hook Formula)**.** *If $\lambda \vdash n$, then*

$$f^\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h_{i,j}}. \tag{7.10}$$

Before proving this result, let us verify it for $\lambda = (2,2,1) \vdash 5$. Using the hooklengths in Figure 7.5 we see that

$$\frac{n!}{\prod_{(i,j)\in\lambda} h_{i,j}} = \frac{5!}{4\cdot 3\cdot 2\cdot 1^2} = 5$$

which agrees with the count in Figure 7.1.

Proof. Consider the following algorithm for constructing a standard Young tableau T of shape λ. In it, $A \coloneqq B$ means that A is to be replace by B.

GNW1 Pick $c \in \lambda$ with probability $1/n$.
GNW2 While c is not an inner corner, pick $c' \in H_c - \{c\}$ with probability $1/(h_c - 1)$ and update $c' \coloneqq c$.
GNW3 Let $T_c = n$ and update $n \coloneqq n-1$, $\lambda \coloneqq \lambda - \{c\}$. If $n > 0$, then return to GNW1; otherwise terminate.

The sequence of cells chosen in GNW2 is called a *trial* t. In Figure 7.6 the solid dots and lines show a possible trial in $\lambda = (7^2, 6^2, 4)$ with probability

$$\Pr(t) = \frac{1}{30}\cdot\frac{1}{7}\cdot\frac{1}{4}\cdot\frac{1}{1} = \frac{1}{840}.$$

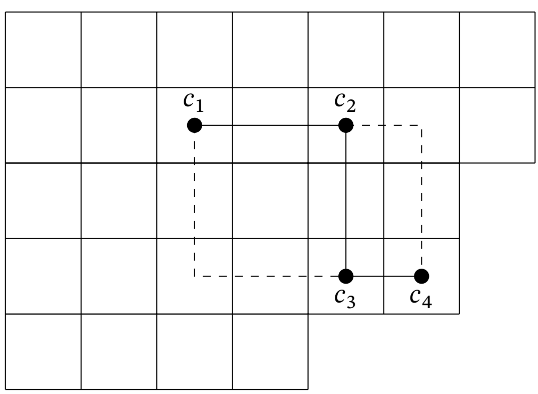


Figure 7.6. A trial in $\lambda = (7^2, 6^2, 4)$

In order to prove (7.10) it suffices to show that for any SYT T of shape $\lambda \vdash n$, the probability that GNW1–GNW3 will produce T is

$$\Pr(T) = \frac{\prod_{(i,j)\in\lambda} h_{i,j}}{n!}. \tag{7.11}$$

We will induct on n, where the case $n = 1$ is trivial. Suppose (α, ω) is the cell of T containing n. Also let $\lambda' = \lambda - \{(\alpha, \omega)\}$ and let T' be the tableau of shape λ' obtained by removing n from T. Then $\Pr(T) = \Pr(\alpha, \omega) \cdot \Pr(T')$ where $\Pr(\alpha, \omega)$ is the probability that a trial ends at (α, ω). Note that the hooklengths of T' are the same as those in T except for the ones in row α or column ω which have each been decreased by one. Also, we can assume by induction that $\Pr(T')$ has the desired form. So it suffices to show that, using $h'_{i,j}$ for the hooklengths in T',

$$\begin{aligned}
\Pr(\alpha,\omega) &= \frac{\prod_{(i,j)\in\lambda} h_{i,j}/n!}{\prod_{(i,j)\in\lambda'} h'_{i,j}/(n-1)!}\\
&= \frac{1}{n}\prod_{1\le i<\alpha}\frac{h_{i,\omega}}{h_{i,\omega}-1}\prod_{1\le j<\omega}\frac{h_{\alpha,j}}{h_{\alpha,j}-1}\\
&= \frac{1}{n}\prod_{1\le i<\alpha}\left(1+\frac{1}{h_{i,\omega}-1}\right)\prod_{1\le j<\omega}\left(1+\frac{1}{h_{\alpha,j}-1}\right)\\
&= \frac{1}{n}\sum_{\substack{I\subseteq[\alpha-1]\\ J\subseteq[\omega-1]}}\prod_{i\in I}\frac{1}{h_{i,\omega}-1}\prod_{j\in J}\frac{1}{h_{\alpha,j}-1}.
\end{aligned}$$

We will prove that this last expression equals $\Pr(\alpha, \omega)$ by giving a probabilistic interpretation to each summand as follows. Given a trial $c_1 = (i_1, j_1), c_2 = (i_2, j_2), \ldots, c_m = (i_m, j_m) = (\alpha, \omega)$ we define its *row* and *column* projections to be the sets $I' = \{i_1, i_2, \ldots, i_m\}$ and $J' = \{j_1, j_2, \ldots, j_m\}$, respectively. For the trial in Figure 7.6 we have $I' = \{2, 4\}$ and $J' = \{3, 5, 6\}$ corresponding to the solid and dotted lines in the diagram. Note that since we assume the trial ends at (α, ω) we always have $\alpha = \max I'$ and $\omega = \max J'$. Let $I = I' - \{\alpha\}$ and $J = J' - \{\omega\}$. Let $\Pr(I', J')$ be the probability that a trial ending at (α, ω) has row and column projections I' and J', respectively. We claim that

$$\Pr(I',J') = \frac{1}{n}\prod_{i\in I}\frac{1}{h_{i,\omega}-1}\prod_{j\in J}\frac{1}{h_{\alpha,j}-1}. \tag{7.12}$$

If this is true, then we are done since, by definition of the probabilities which are involved, $\Pr(\alpha, \omega) = \sum_{I',J'} \Pr(I', J')$ which is the same as the sum at the end of the previous paragraph.

To prove the claim, we induct on m, the number of cells in the trial. If $m = 1$, then this is clearly true since then $I = J = \emptyset$ and $1/n$ is the probability of picking (α, ω) as the only cell of the trial. If $m > 1$, then the trial must begin by going from (i_1, j_1) to either (i_2, j_1) or (i_1, j_2). So

$$\Pr(I',J') = \frac{1}{h_{i_1,j_1}-1}[\Pr(I'-i_1,J') + \Pr(I',J'-j_1)].$$

Letting P be the right-hand side of (7.12) we have, by induction, that

$$\Pr(I' - i_1, J') = (h_{i_1,\omega} - 1)P$$

and

$$\Pr(I', J' - j_1) = (h_{\alpha,j_1} - 1)P.$$

It is also easy to show, using (7.9), that

$$h_{i_1,j_1} - 1 = (h_{i_1,\omega} - 1) + (h_{\alpha,j_1} - 1). \tag{7.13}$$

Thus

$$\Pr(I', J') = \frac{1}{h_{i_1,j_1} - 1}[(h_{i_1,\omega} - 1)P + (h_{\alpha,j_1} - 1)P] = P$$

as desired. □

We now derive a generating function for semistandard Young tableaux T of a given shape λ by the sum of the parts which is denoted $|T| = \sum_{(i,j)\in\lambda} T_{i,j}$. It will be convenient to consider a related type of array. A *reverse plane partition* (*RPP*) *of shape* λ is a filling R of the Young diagram of λ with elements of $\mathbb{N}$ such that the rows and columns weakly increase. (The term "reverse" comes from the fact that ordinary partitions of n are written in weakly decreasing, rather than increasing, order.) The first row of Figure 7.7 contains a list of six RPPs. We use notation for semistandard Young tableaux in the obvious way applied to reverse plane partitions. Let $\operatorname{rpp}_n(\lambda)$ be the number of reverse plane partitions R with $\operatorname{sh} R = \lambda$ and $|R| = n$. We note that there is a bijection $T \mapsto R$ where T is an SSYT, R is an RPP, and $\operatorname{sh} T = \operatorname{sh} R = \lambda$, given by letting $R_{i,j} = T_{i,j} - i$ for all $(i,j) \in \lambda$. Notice that in this case $|R| = |T| + \sum_i i\lambda_i$. So finding the generating function for RPPs is equivalent to finding the one for SSYT. The former generating function was first derived by Stanley [**84**]. The algorithmic proof we give is due to Hillman and Grassl [**43**].

Theorem 7.3.2. *For any partition λ we have*

$$\sum_{n\geq 0} \operatorname{rpp}_n(\lambda) x^n = \prod_{(i,j)\in\lambda} \frac{1}{1 - x^{h_{i,j}}}. \tag{7.14}$$

Proof. By Lemma 3.4.1, the right-hand side of (7.14) is the weight-generating function for the product

$$S = \bigtimes_{(i,j)\in\lambda} \{\{h_{i,j}^{m_{i,j}} \mid m_{i,j} \geq 0\}\}$$

where if we have a multiset $M \in S$, then $\operatorname{wt} M = x^{\sum_{(i,j)\in\lambda} m_{i,j} h_{i,j}}$. Note that even if two hooks have the same length, they contribute to different components of the product. So we need a weight-preserving bijection between RPPs R and multisets of hooklengths M.

Given R, we find the hooklengths in M by producing a series of RPPs

$$R = R_0, R_1, \ldots, R_m \tag{7.15}$$

where R_m is the all-zero RPP and, at each stage, R_k is obtained from R_{k-1} by subtracting one from all the elements of R_{k-1} along a path p_k which is constructed so that $|p_k| = h_{i_k,j_k}$ for some (i_k, j_k).

$$
\begin{array}{cccccccc}
R_k : & \begin{array}{|c|c|c|c|}\hline 1&2&2&2\\\hline 3&3&3\\\cline{1-3} 3\\\cline{1-1}\end{array} & \begin{array}{|c|c|c|c|}\hline 1&1&1&1\\\hline 2&2&3\\\cline{1-3} 2\\\cline{1-1}\end{array} & \begin{array}{|c|c|c|c|}\hline 0&0&0&0\\\hline 1&2&3\\\cline{1-3} 1\\\cline{1-1}\end{array} & \begin{array}{|c|c|c|c|}\hline 0&0&0&0\\\hline 1&2&2\\\cline{1-3} 1\\\cline{1-1}\end{array} & \begin{array}{|c|c|c|c|}\hline 0&0&0&0\\\hline 1&1&1\\\cline{1-3} 1\\\cline{1-1}\end{array} & \begin{array}{|c|c|c|c|}\hline 0&0&0&0\\\hline 0&0&0\\\cline{1-3} 0\\\cline{1-1}\end{array} \\
& & & & & & \\
h_{i_k,j_k} : & & h_{1,1} & h_{1,1} & h_{2,3} & h_{2,2} & h_{2,1}
\end{array}
$$

Figure 7.7. The Hillman–Grassl Algorithm

Given R, we find the path $p = p_1$ as follows.

HG1 Start p at (a, b) which is the northeastmost cell in R such that $R_{a,b} \neq 0$.
HG2 Continue p by

$$(i, j) \in p \implies \begin{cases} (i, j-1) \in p & \text{if } T_{i,j-1} = T_{i,j}, \\ (i+1, j) \in p & \text{otherwise.} \end{cases}$$

HG3 Terminate p when trying to apply HG2 leads to $(i + 1, j) \notin \lambda$.

Note that HG2 amounts to saying that p moves down unless forced to move left so as not to violate the weakly increasing condition on the rows once the ones are subtracted from R. Note also that the termination condition in HG3 forces p to be at the bottom of some column c. Since all southwest lattice paths from (a, b) to the bottom of column c have the same length, we must have $|p| = h_{a,c}$. One also needs to check that the algorithm is well-defined in that the output array is actually an RPP; that is, it has weakly increasing rows and columns. But this verification will be left as an exercise.

To illustrate, let R be the first RPP in Figure 7.7. Then using HG1–HG3 returns the path given by the dots in

$$\begin{array}{|c|c|c|c|}\hline & \bullet & \bullet & \bullet \\\hline \bullet & \bullet & \\\cline{1-3} \bullet \\\cline{1-1}\end{array}$$

and upon subtraction one obtains the second RPP in the figure. Notice that we have subtracted a total of $h_{1,1} = 6$ as indicated on the bottom line. The rest of the figure illustrates finding the other RPPs in the sequence (7.15). So, in this case,

$$R \mapsto M = \{\{h_{1,1}^2, h_{2,3}, h_{2,2}, h_{2,1}\}\}.$$

To reverse the procedure and find an inverse map, we first need to determine in what order hooklengths are removed from R to form M. We claim that $h_{i',j'}$ was removed before $h_{i'',j''}$ in the hook decomposition of R if and only if

(7.16) $$i' < i'' \quad \text{or} \quad i' = i'' \text{ and } j' \geq j''.$$

It is easy to see that this is a total order on the cells of λ, so it suffices to prove the forward direction. And by transitivity, one can reduce to the case when $h_{i'',j''}$ is removed directly after $h_{i',j'}$. Let R' and R'' be the reverse plane partitions from which $h_{i',j'}$ and $h_{i'',j''}$ were removed using paths p' and p'', respectively. Since entries decrease in passing from R' to R'', the initial condition in HG1 forces $i' \leq i''$. If this inequality is strict, then we are done. If $i' = i''$, then we assert that every cell on p'' is weakly left of a cell of p'. So if p' ends in column j', then p'' must end in a column $j'' \leq j'$ as claimed.

The assertion is proven by induction, for suppose $(i,h) \in p''$ is weakly left of $(i,j) \in p'$ so that $h \le j$. If the next step of p' is to $(i+1,j)$, then p'' will enter row $i+1$ still weakly left of column j. If the next step of p' is to $(i,j-1)$, then the given cell of p'' will still be weakly left if $h < j$. And if $h = j$, then p'' must move to $(i,j-1)$ as well since p' only moves left if $T'_{i,j-1} = T'_{i,j}$, and after subtraction we will also have $T''_{i,j-1} = T''_{i,j}$. So the assertion, and hence (7.16), holds.

To construct the inverse map, given a multiset M, we arrange its elements in a sequence according to (7.16):

$$h_{i_1,j_1}, h_{i_2,j_2}, \dots, h_{i_m,j_m}.$$

From this, we construct a sequence of RPPs

$$R_m, R_{m-1}, \dots, R_0 = R$$

where R_m is the all-zero RPP and R_{k-1} is obtained by adding a one to a total of h_{i_k,j_k} elements of R_k for $k = m, m-1, \dots, 1$. Given an RPP R, we add back $h_{a,c}$ ones along a *reverse path* r defined as follows.

GH1 Start r at the bottom cell in column c of R.
GH2 Continue r by

$$(i,j) \in r \implies \begin{cases} (i,j+1) \in r & \text{if } T_{i,j+1} = T_{i,j}, \\ (i-1,j) \in r & \text{otherwise.} \end{cases}$$

GH3 Terminate r when it passes through the rightmost cell in row a.

Note that this is a step-by-step reversal of the construction of the (forward) path in HG1–HG3. So this will be an inverse map provided that it is well-defined, that is, provided that in GH3 the reverse path actually reaches the target cell in row a. This is forced by (7.16) and the proof of this implication is left to the reader. We also leave as an exercise the check that adding back ones along the reverse path yields an RPP. □

7.4. P-partitions

It is natural to wonder if there is any relationship between equations (7.10) and (7.14). In fact, the latter can be used to derive the former. In order to do this, we will need to develop the theory of P-partitions, P being a poset, which is due to Stanley [**84**].

We start with the central concept of compatibility of a permutation and a function. A function $f : [n] \to \mathbb{N}$ is *compatible* with a permutation $\pi = \pi_1 \dots \pi_n \in \mathfrak{S}_n$ if

C1 $f(\pi_1) \ge f(\pi_2) \ge \cdots \ge f(\pi_n)$ and
C2 $f(\pi_i) > f(\pi_{i+1})$ whenever $i \in \operatorname{Des} \pi$.

By way of example, suppose $\pi = 37814526$. It is easy to check that $f : [8] \to \mathbb{N}$ defined by

$$f(3) = f(7) = f(8) = 21,\ f(1) = f(4) = 20,\ f(5) = 10,\ f(2) = f(6) = 0 \tag{7.17}$$

is compatible with π since

$$f(\pi_1), f(\pi_2), \dots, f(\pi_8) = 21 \ge 21 \ge 21 > 20 \ge 20 \ge 10 > 0 \ge 0.$$

For $\pi \in \mathfrak{S}_n$, let

$$\mathcal{C}(\pi) = \{f : [n] \to \mathbb{N} \mid f \text{ is compatible with } \pi\}.$$

These sets partition the set of all functions from $[n]$ to $\mathbb{N}$.

Lemma 7.4.1. *Every $f : [n] \to \mathbb{N}$ is compatible with a unique $\pi \in \mathfrak{S}_n$. Thus*

$$\{f \mid f : [n] \to \mathbb{N}\} = \biguplus_{\pi \in \mathfrak{S}_n} \mathcal{C}(\pi). \tag{7.18}$$

Proof. We first show how, given f, we can construct a π with which f is compatible. The reader may wish to follow the construction using the example f in (7.17). Let the image of f be the set $S = \{s_1 > s_2 > \cdots > s_k\} \subset \mathbb{N}$. Since f is weakly decreasing on π by condition C1, those r with $f(r) = s_1$ must come first in π. Furthermore, such r must be arranged in increasing order since, if not, then there would be a descent which would force two of these r to have distinct images by C2. Similar considerations show that the next elements in π must be those such that $f(r) = s_2$ in increasing order, and so forth. Since all of the choices made in constructing π are forced on us by the definition, the permutation is unique and we have proved the first statement of the proposition.

As for (7.18), the uniqueness statement just proved shows that the union is disjoint. And existence of a compatible π for each f shows containment of the left-hand side in the right. The other containment is trivial since each $\mathcal{C}(\pi)$ consists of functions $f : [n] \to \mathbb{N}$. □

Define the *size* of $f : [n] \to \mathbb{N}$ to be

$$|f| = \sum_{i=1}^{n} f(i). \tag{7.19}$$

Continuing our example

$$|f| = 21 + 21 + 21 + 20 + 20 + 10 + 0 + 0 = 113.$$

It will also be instructive to consider the following subsets of $\mathcal{C}(\pi)$ where the maximum of a function is the maximum of the values in its image

$$\mathcal{C}_m(\pi) = \{f \in \mathcal{C}(\pi) \mid \max f \le m\}.$$

There are nice generating functions associated with $\mathcal{C}(\pi)$ and $\mathcal{C}_m(\pi)$.

Lemma 7.4.2. *For any $\pi \in \mathfrak{S}_n$ we have*

$$\sum_{f \in \mathcal{C}(\pi)} x^{|f|} = \frac{x^{\operatorname{maj} \pi}}{(1-x)(1-x^2)\cdots(1-x^n)} \tag{7.20}$$

and

$$\sum_{m \ge 0} \#\mathcal{C}_m(\pi) x^m = \frac{x^{\operatorname{des} \pi}}{(1-x)^{n+1}}. \tag{7.21}$$

Proof. We will prove the first equality as the demonstration of the second is similar and so is left as an exercise. The basic idea behind the proof is the same as used in the demonstration of Theorem 1.3.4 where one adds or subtracts sufficient amounts to turn weak inequalities into strict ones or vice versa. An example will follow the proof.

Let $\pi = \pi_1\pi_2\dots\pi_n$ with $\operatorname{Des}\pi = \{d_1 < d_2 < \cdots < d_k\}$. We will construct a bijection $\phi\colon \mathcal{C}(\pi) \to \Lambda_n$ where Λ_n is the set of all partitions λ satisfying the length restriction $\ell(\lambda) \le n$. Given $f \in \mathcal{C}(\pi)$, we will construct a sequence of functions $f = f_0, \dots, f_k$ where f_i will remove the strict inequality restriction at index d_i from f_{i-1}. So let f_1 be obtained from f by subtracting one from each of $f(\pi_1), \dots, f(\pi_{d_1})$ and leaving the other f values the same. Similarly, f_2 is constructed from f_1 by subtracting one from $f_1(\pi_1), \dots, f_1(\pi_{d_2})$ with other values constant, and so forth. By the end, the only restrictions on f_k are that we have $f_k(\pi_1) \ge \cdots \ge f_k(\pi_n) \ge 0$. So the nonzero images of f_k form a partition $\lambda \in \Lambda_n$ and we let $\phi(f) = \lambda$. This is a bijection as its inverse is easy to construct. Furthermore, from the definition of the algorithm, it follows that

$$|f| = |f_1| + d_1 = \cdots = |\lambda| + \sum_{i=1}^{k} d_i = |\lambda| + \operatorname{maj}\pi.$$

Now appealing to Corollary 3.5.4 we have

$$\sum_{f \in \mathcal{C}(\pi)} x^{|f|} = \sum_{\lambda \in \Lambda_n} x^{|\lambda| + \operatorname{maj}\pi} = \frac{x^{\operatorname{maj}\pi}}{(1-x)(1-x^2)\cdots(1-x^n)}$$

as desired. □

Using our running example, the vector of values of the initial function on π is given by $f = (21, 21, 21, 20, 20, 10, 0, 0)$. Since $\operatorname{Des}\pi = \{3, 6\}$ our first step is to subtract one from the first 3 values to obtain $f_1 = (20, 20, 20, 20, 20, 10, 0, 0)$. Next we subtract one from the first 6 values so that $f_2 = (19, 19, 19, 19, 19, 9, 0, 0)$. Taking the nonzero components gives $\lambda = (19, 19, 19, 19, 19, 9)$.

We now have all the tools needed to find generating functions for partitions whose parts are distributed over a poset. Let P be a partial order on the set $[n]$. In order to distinguish the usual total order on integers from the partial order in P, we will use $i \le j$ for the former and $i \trianglelefteq j$ for the latter. So in the poset on the left in Figure 7.8 we have $3 \triangleleft 2$, but $2 < 3$ as integers. If P is a poset on $[n]$, then a *P-partition* is a map

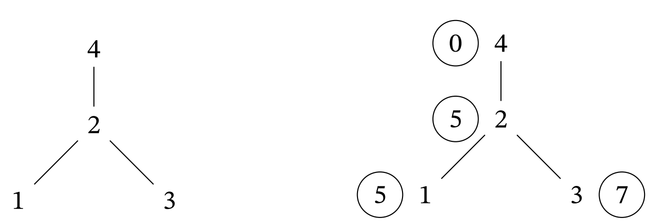


Figure 7.8. A poset P on [4] on the left and a P-partition on the right

$f : P \to \mathbb{N}$ such that

PP1 $i \trianglelefteq j$ implies $f(i) \geq f(j)$ and
PP2 $i \trianglelefteq j$ and $i > j$ implies $f(i) > f(j)$.

So PP1 says that f is weakly decreasing on P, while PP2 means that f is strictly decreasing on "descents" of P. A P-partition for the poset in Figure 7.8 is shown on the right with the values of f circled. Note that by transitivity, it suffices to assume that PP1 and PP2 hold when i is covered by j. Let

$$\operatorname{Par} P = \{f : P \to \mathbb{N} \mid f \text{ is a } P\text{-partition}\}.$$

For the poset in Figure 7.8 we have

$$\operatorname{Par} P = \{f : [4] \to \mathbb{N} \mid f(1) \geq f(2), f(3) > f(2), f(2) \geq f(4)\}.$$

We will also need the *Jordan–Hölder set* of an (arbitrary) poset P which is

$$\mathcal{L}(P) = \{\pi \mid \pi \text{ is a linear extension of } P\}.$$

Note that if P is a poset on $[n]$, then $\mathcal{L}(P) \subseteq \mathfrak{S}_n$. Continuing the Figure 7.8 example, $\mathcal{L}(P) = \{1324, 3124\}$. The next result, while not hard to prove, is crucial, as its name suggests.

Lemma 7.4.3 (Fundamental Lemma of P-Partitions). *Let P be a poset on $[n]$. Then we have $f \in \operatorname{Par} P$ if and only if $f \in \mathcal{C}(\pi)$ for some $\pi \in \mathcal{L}(P)$. Thus*

$$\operatorname{Par} P = \biguplus_{\pi \in \mathcal{L}(P)} \mathcal{C}(\pi).$$

Proof. We will just prove the forward implication as the reverse is similar. And the proof of the equation for $\operatorname{Par} P$ is also omitted as it follows the same lines as for (7.18). So suppose $f \in \operatorname{Par} P$. We know from the first part of Lemma 7.4.1 that f is compatible with a unique $\pi \in \mathfrak{S}_n$. So we just need to show that $\pi \in \mathcal{L}(P)$; that is, if $i \triangleleft j$, then i should appear before j in π. Assume, to the contrary, that we have

$$\pi = \ldots j \ldots i \ldots \tag{7.22}$$

being the order of the two elements in π. Since $f \in \operatorname{Par} P$ and $i \triangleleft j$ we must have $f(i) \geq f(j)$ by condition PP1. But C1 and the form of π in (7.22) force $f(j) \geq f(i)$. Thus $f(i) = f(j)$. This equality together with (7.22) and C2 imply $j < i$. But now we have $i \triangleleft j$ and $i > j$ as well as $f(i) = f(j)$ which contradicts P2. This is the desired contradiction. □

We now translate the previous result in terms of generating functions. Just as with compatible functions, use the notation

$$\operatorname{Par}_m P = \{f \in \operatorname{Par} P \mid \max f \leq m\}.$$

Theorem 7.4.4. *For any poset P on $[n]$ we have*

$$\sum_{f \in \operatorname{Par} P} x^{|f|} = \frac{\sum_{\pi \in \mathcal{L}(P)} x^{\operatorname{maj} \pi}}{(1-x)(1-x^2)\cdots(1-x^n)} \tag{7.23}$$

and

(7.24)
$$\sum_{m\ge 0} |\operatorname{Par}_m P|\, x^m = \frac{\sum_{\pi\in\mathcal{L}(P)} x^{\operatorname{des}\pi}}{(1-x)^{n+1}}.$$

Proof. We prove (7.23), leaving (7.24) as an exercise. Using the previous lemma and then (7.20) yields

$$\sum_{f\in\operatorname{Par}P} x^{|f|} = \sum_{\pi\in\mathcal{L}(P)} \sum_{f\in\mathcal{C}(\pi)} x^{|f|} = \frac{\sum_{\pi\in\mathcal{L}(P)} x^{\operatorname{maj}\pi}}{(1-x)(1-x^2)\cdots(1-x^n)}$$

which is what we wished to demonstrate. □

As a check, consider the poset P which is the chain $1 \lhd 2 \lhd \ldots \lhd n$. Then the only inequalities satisfied by $f \in \operatorname{Par}P$ are $f(1) \ge f(2) \ge \cdots \ge f(n) \ge 0$. So f corresponds to a partition λ with at most n parts. On the other hand $\mathcal{L}(P)$ consists of the single permutation $\pi = 12\ldots n$ with $\operatorname{maj}\pi = 0$. So (7.23) becomes

(7.25)
$$\sum_{\ell(\lambda)\le n} x^{|\lambda|} = \frac{1}{(1-x)(1-x^2)\cdots(1-x^n)}$$

which agrees with Corollary 3.5.4. Of course, this cannot be considered a new proof of the corollary since it was used in the demonstration of (7.20). But at least it suggests we haven't made any mistakes!

Another case is the chain $n \lhd n-1 \lhd \ldots \lhd 1$. Now the $f \in \operatorname{Par}P$ satisfy the inequalities $f(n) > f(n-1) > \cdots > f(1) \ge 0$. The set $\mathcal{L}(P)$ still contains a unique element, but it is $\pi = n\ldots 21$ which has

$$\operatorname{maj}\pi = 1 + 2 + \cdots + (n-1) = \binom{n}{2}.$$

Plugging into (7.23) we see that

(7.26)
$$\sum_{f\in\operatorname{Par}P} x^{|f|} = \frac{x^{\binom{n}{2}}}{(1-x)(1-x^2)\cdots(1-x^n)}.$$

The reader may find it instructive to write down the bijection which permits one to derive this equation from (7.25). This map is a special case of the one used in the proof of (7.20) but it is easier to see what is going on in this simple case.

The time has come to fulfill our promise from the beginning of this section to derive the Hook Formula, (7.10), from the generating function for reverse plane partitions, (7.14). Let λ be the Young diagram of a partition of n. We turn λ into a poset P_λ by partially ordering the cells of λ componentwise: $(i,j) \unlhd (i',j')$ whenever $i \le i'$ and $j \le j'$. See Figure 7.9 for an example where $\lambda = (4,3,1)$. So P_λ is formed from the Young diagram of λ by rotating 135° counterclockwise and imposing a grid of covers. Note that $\#\mathcal{L}(P_\lambda) = f^\lambda$ because there is a simple bijection between SYT T of shape λ and linear extensions of P_λ: each tableau T corresponds to a linear extension of the cells $c_1,\ldots,c_n$ where they are ordered so that c_i is the cell of T containing i for $1 \le i \le n$.

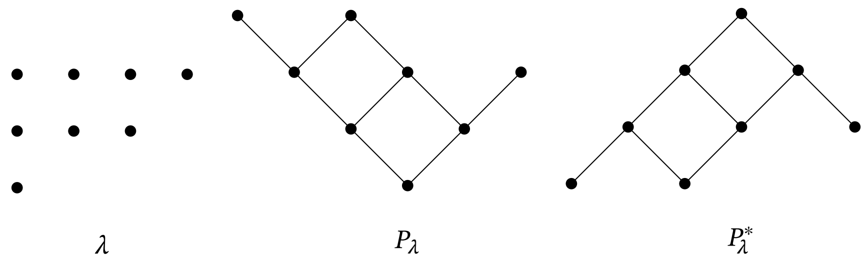


Figure 7.9. A Young diagram and associated posets

Now consider the poset dual P_λ^* and label its elements with the numbers in $[n]$ in any way which corresponds to a linear extension of P_λ^* as described in the previous paragraph for P_λ. It is easy to see that the P_λ^*-partitions are precisely the reverse plane partitions of shape λ. (The use of the dual corresponds to these plane partitions being "reverse".) And clearly we still have $\#\mathcal{L}(P_\lambda^*) = f^\lambda$. Combining (7.14) and (7.23) yields.

$$\prod_{(i,j)\in\lambda} \frac{1}{1-x^{h_{i,j}}} = \sum_{n\geq 0} \operatorname{rpp}_n(\lambda)x^n = \frac{p(x)}{(1-x)(1-x^2)\cdots(1-x^n)}$$

where $p(x) = \sum_{\pi\in\mathcal{L}(P_\lambda^*)} x^{\operatorname{maj}\pi}$. Thus

$$f^\lambda = p(1) = \lim_{x\to 1} \frac{(1-x)(1-x^2)\cdots(1-x^n)}{\prod_{(i,j)\in\lambda} 1-x^{h_{i,j}}} = \frac{n!}{\prod_{(i,j)\in\lambda} h_{i,j}}$$

which is the Hook Formula.

7.5. The Robinson–Schensted–Knuth correspondence

This section will be devoted to proving the following important identity.

Theorem 7.5.1. *For any given $n \geq 0$ we have*

$$\sum_{\lambda\vdash n} (f^\lambda)^2 = n!\,. \tag{7.27}$$

From the point of view of representation theory this is just the special case of equation (A.9) in the appendix where $G = \mathfrak{S}_n$ and the dimensions are given by (A.7). However, we wish to give a bijective proof of (7.27). This map was discovered in two very different forms by Robinson [**75**] and Schensted [**81**]. It is the latter description which will be presented here. We will also see that this algorithm and the corresponding identity can be generalized from standard to semistandard Young tableaux.

To prove (7.27), it suffices to construct a bijection

$$\pi \overset{\text{RS}}{\mapsto} (P, Q) \tag{7.28}$$

between permutations $\pi \in \mathfrak{S}_n$ and pairs of SYT (P, Q) of the same shape $\lambda \vdash n$. The heart of this construction will be a method of inserting a positive integer into a tableau. A *partial Young tableau* (*PYT*) is a filling of a shape with distinct positive integers such that the rows and columns increase. A PYT is standard precisely when its entries are

$$P = \begin{array}{|c|c|c|c|}\hline 1&3&6&9\\ \hline 2&8&10 \\ \cline{1-3} 4 \\ \cline{1-1} 7 \\ \cline{1-1}\end{array}$$

$$\begin{array}{|c|c|c|c|}\hline 1&3&6&9\\ \hline 2&8&10 \\ \cline{1-3} 4 \\ \cline{1-1} 7 \\ \cline{1-1}\end{array} \leftarrow \mathbf{5} \quad \begin{array}{|c|c|c|c|}\hline 1&3&\mathbf{5}&9\\ \hline 2&8&10 \\ \cline{1-3} 4 \\ \cline{1-1} 7 \\ \cline{1-1}\end{array} \leftarrow \mathbf{6} \quad \begin{array}{|c|c|c|c|}\hline 1&3&5&9\\ \hline 2&\mathbf{6}&10 \\ \cline{1-3} 4 \\ \cline{1-1} 7 \\ \cline{1-1}\end{array} \leftarrow \mathbf{8} \quad \begin{array}{|c|c|c|c|}\hline 1&3&5&9\\ \hline 2&6&10 \\ \cline{1-3} 4&\mathbf{8} \\ \cline{1-2} 7 \\ \cline{1-1}\end{array} = r_5(P).$$

Figure 7.10. Inserting $x = 5$ into a partial tableau P

$[n]$ for some n. A partial tableau P is shown at the top in Figure 7.10. Now given a PYT P and a positive integer $x \notin P$, we *insert x into P* using the following algorithm:

RS1 Set $R :=$ the first row of P.

RS2 While x is less than some element of row R, let y be the leftmost such element and replace y by x in R. Repeat this step with $R :=$ the row below R and $x := y$.

RS3 Now x is greater than every element of R, so place x at the end of this row and terminate.

In step RS2 we say that *x bumps y*.

An example of inserting 5 into the PYT in Figure 7.10 is given in the second row of the figure. Elements being bumped are written in boldface and the notation $R \leftarrow x$ means that x is being inserted in row R. If P' is the result of inserting x into P by rows, then we write

$$r_x(P) = P'.$$

The reader should check that this operation is well-defined in that P' is still a PYT.

There is a second operation needed to describe the map (7.28) which will be used for the second component. An *outer corner* of a shape λ (or of a tableau of that shape) is a cell $(i,j) \notin \lambda$ such that $\lambda \cup \{(i,j)\}$ is the Young diagram of a partition. The outer corners of Q in Figure 7.11 are $(1,5)$, $(2,3)$, $(3,2)$, and $(5,1)$. Suppose we have a partial tableau Q, $y > \max Q$, and (i,j) which is an outer corner of Q. The tableau Q' obtained by *placing y in Q at (i,j)* has all the entries of Q together with $Q'_{i,j} = y$. The choice of an outer corner and the condition on y ensure that Q' is still a PYT. See Figure 7.11 for an example of a placement.

We are now ready to describe (7.28). Consider π as being given in two-line notation (1.7)

$$\pi = \begin{array}{cccc} 1 & 2 & \ldots & n \\ \pi_1 & \pi_2 & \ldots & \pi_n \end{array}.$$

We will construct a sequence of pairs of tableaux

(7.29) $$(P_0, Q_0) = (\emptyset, \emptyset),\ (P_1, Q_1),\ (P_2, Q_2),\ \ldots,\ (P_n, Q_n) = (P, Q)$$

$$Q = \begin{array}{cccc} 1 & 2 & 6 & 7 \\ 4 & 5 & & \\ 8 & & & \\ 9 & & & \end{array} \qquad Q' = \begin{array}{cccc} 1 & 2 & 6 & 7 \\ 4 & 5 & & \\ 8 & 10 & & \\ 9 & & & \end{array}$$

Figure 7.11. The result Q' of placing 10 at $(3, 2)$ in Q

by starting with the empty pair and then letting

$$\begin{aligned} P_k &= r_{\pi_k}(P_{k-1}), \\ Q_k &= \text{place } k \text{ in } Q_{k-1} \text{ at the cell where } r_{\pi_k} \text{ terminates,} \end{aligned}$$

for $k = 1, 2, \ldots, n$. We then let $(P, Q) = (P_n, Q_n)$. Note that by construction we have $\operatorname{sh} P_k = \operatorname{sh} Q_k$ for all k. A complete example is worked out in Figure 7.12 where the elements of the lower line of π as well as their counterparts in the P_k are set in bold. If $\mathrm{RS}(\pi) = (P, Q)$, we also write $P(\pi) = P$ and call P the *P-tableau* or *insertion tableau* of π. Similarly we use the notation $Q(\pi) = Q$ for the *Q-tableau* of π which is also called the *recording tableau*.

We now come to our main theorem about this procedure.

Theorem 7.5.2. *The map*

$$\pi \overset{\mathrm{RS}}{\mapsto} (P, Q)$$

is a bijection between permutations $\pi \in \mathfrak{S}_n$ and pairs (P, Q) of SYT of the same shape $\lambda \vdash n$.

Proof. It suffices to construct the inverse. This will be done by reversing the algorithm step by step. So we will build the sequence (7.29) backwards, starting from $(P_n, Q_n) = (P, Q)$ and, in the process, recover π. Assume that we have reached (P_k, Q_k) and let (i, j) be the cell containing k in Q_k. To obtain Q_{k-1} we merely erase k from Q_k. As for finding P_{k-1} and π_k, we note that (i, j) must have been the cell at which the insertion into P_{k-1} terminated. So we use the following *deletion* procedure to undo this insertion.

SR1 Let x be the (i, j) entry of P_k and erase it from P_k. Set $R :=$ the $(i-1)$st row of P_k.

SR2 While R is not the zeroth row of P_k, let y be the rightmost element of R smaller than x and replace y by x in P_k. Repeat this step with $R :=$ the row above R and $x := y$.

SR3 Now R is the zeroth row so let $\pi_k = x$ and terminate.

It should be clear from the constructions that insertion and deletion are inverses of each other. So we are done. □

In order to generalize (7.27), we will consider two sets of variables $\mathbf{x} = \{x_1, x_2, \ldots\}$ and $\mathbf{y} = \{y_1, y_2, \ldots\}$. The next result is called Cauchy's Identity and it can be found in Littlewood's text [**58**].

$$\pi = \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \mathbf{5} & \mathbf{2} & \mathbf{3} & \mathbf{6} & \mathbf{4} & \mathbf{1} & \mathbf{7} \end{array}.$$

$$\mathbf{P_k}: \quad \emptyset \quad \begin{ytableau} \mathbf{5} \end{ytableau} \quad \begin{ytableau} \mathbf{2} \\ \mathbf{5} \end{ytableau} \quad \begin{ytableau} \mathbf{2} & \mathbf{3} \\ \mathbf{5} \end{ytableau} \quad \begin{ytableau} \mathbf{2} & \mathbf{3} & \mathbf{6} \\ \mathbf{5} \end{ytableau} \quad \begin{ytableau} \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{5} & \mathbf{6} \end{ytableau} \quad \begin{ytableau} \mathbf{1} & \mathbf{3} & \mathbf{4} \\ \mathbf{2} & \mathbf{6} \\ \mathbf{5} \end{ytableau} \quad \begin{ytableau} \mathbf{1} & \mathbf{3} & \mathbf{4} & \mathbf{7} \\ \mathbf{2} & \mathbf{6} \\ \mathbf{5} \end{ytableau} \quad = \mathbf{P}$$

$$Q_k: \quad \emptyset, \quad \begin{ytableau} 1 \end{ytableau} \quad \begin{ytableau} 1 \\ 2 \end{ytableau} \quad \begin{ytableau} 1 & 3 \\ 2 \end{ytableau} \quad \begin{ytableau} 1 & 3 & 4 \\ 2 \end{ytableau} \quad \begin{ytableau} 1 & 3 & 4 \\ 2 & 5 \end{ytableau} \quad \begin{ytableau} 1 & 3 & 4 \\ 2 & 5 \\ 6 \end{ytableau} \quad \begin{ytableau} 1 & 3 & 4 & 7 \\ 2 & 5 \\ 6 \end{ytableau} \quad = Q$$

Figure 7.12. The Robinson–Schensted map

Theorem 7.5.3. *We have*

$$\sum_{\lambda} s_\lambda(\mathbf{x}) s_\lambda(\mathbf{y}) = \prod_{i,j \geq 1} \frac{1}{1 - x_i y_j} \tag{7.30}$$

where the sum is over all partitions λ of any nonnegative integer.

To give a bijective proof of this formula, we must interpret each side as a weight-generating function. On the left, we clearly have the weight ogf for pairs (T, U) of semistandard tableaux of the same shape λ where

$$\operatorname{wt}(T, U) = \mathbf{x}^U \mathbf{y}^T.$$

For the right-hand side, consider the set Mat of all infinite matrices M with rows and columns indexed by $\mathbb{P}$, entries in $\mathbb{N}$, and only finitely many entries nonzero. A matrix $M \in$ Mat is shown in the upper-left corner of Figure 7.13 where all entries not shown are zero. Weight these matrices by

$$\operatorname{wt} M = \prod_{i,j \geq 1} (x_i y_j)^{M_{i,j}}.$$

Our example matrix has weight

$$\operatorname{wt} M = (x_1 y_2)^2 (x_1 y_3)^3 (x_2 y_1)(x_2 y_2)^2 (x_3 y_1)(x_3 y_3) = x_1^5 x_2^3 x_3^2 y_1^2 y_2^4 y_3^4.$$

$$M = \begin{bmatrix} 0 & 2 & 3 & 0 & \cdots \\ 1 & 2 & 0 & 0 & \cdots \\ 1 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \mapsto \pi = \begin{array}{cccccccccc} 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 \\ 2 & 2 & 3 & 3 & 3 & 1 & 2 & 2 & 1 & 3 \end{array}$$

$$\overset{\text{RSK}}{\mapsto} (T, U) = \left(\begin{ytableau} 1 & 1 & 2 & 2 & 3 & 3 \\ 2 & 2 & 3 \\ 3 \end{ytableau}, \begin{ytableau} 1 & 1 & 1 & 1 & 1 & 3 \\ 2 & 2 & 2 \\ 3 \end{ytableau} \right)$$

Figure 7.13. The Robinson–Schensted–Knuth map

So, by the Sum and Product Rules for weight ogfs

$$\sum_{M \in \mathrm{Mat}} \mathrm{wt}\, M = \prod_{i,j \geq 1} \sum_{k \geq 0} (x_i y_j)^k = \prod_{i,j \geq 1} \frac{1}{1 - x_i y_j}.$$

Thus we need a weight-preserving bijection

$$M \overset{\mathrm{RSK}}{\mapsto} (T, U) \tag{7.31}$$

between matrices $M \in \mathrm{Mat}$ and pairs $(T, U) \in \mathrm{SSYT}(\lambda) \times \mathrm{SSYT}(\lambda)$ as λ varies over all partitions. Such a map was given by Knuth [**49**].

It will be convenient to reinterpret the elements of Mat as two-line arrays. Given $M \in \mathrm{Mat}$, we create an array π such that

$$M_{i,j} = \text{the number of times a column } \begin{matrix} i \\ j \end{matrix} \text{ occurs in } \pi,$$

and the columns are arranged in lexicographic order with the top row taking precedence. The two-line array π associated with the matrix in Figure 7.13 is displayed at the top right.

We can now define the map (7.31). Given M, construct its two-line array π. Now, starting with the empty tableau, insert the elements of the lower row of π sequentially to form T using exactly the same rules RS1–RS3 as before. Note that this algorithm never used the assumption that $x \notin P$ and so it applies equally well to semistandard tableaux. As one does the insertions, one places the corresponding elements of the upper row of π in U so that the two tableaux always have the same shape. Figure 7.13 displays the final output of this algorithm on the bottom line. The reader should now be able to fill in the details of the proof of the following theorem.

Theorem 7.5.4. *The map*

$$M \overset{\mathrm{RSK}}{\mapsto} (T, U)$$

is a weight-preserving bijection between matrices $M \in \mathrm{Mat}$ *and pairs* (T, U) *of SSYT such that* $\mathrm{sh}\, T = \mathrm{st}\, U$. □

We will use the notation $\mathrm{RSK}(M) = \mathrm{RSK}(\pi) = (T, U)$ where π is the two-line array corresponding to M.

7.6. Longest increasing and decreasing subsequences

One of Schensted's motivations [**81**] for introducing the algorithm which bears his name was to study the lengths of longest increasing and decreasing subsequences of a permutation. He proved that these quantities were given by the length of the first row and the length of the first column, respectively, of the associated tableaux. In this section, we will prove his result. Along the way we will see what effect reversing a sequence has on its insertion tableau.

Consider a permutation $\pi = \pi_1 \pi_2 \ldots \pi_n \in \mathfrak{S}_n$. Then an *increasing subsequence of* π *of length* l is $\pi_{i_1} < \pi_{i_2} < \cdots < \pi_{i_l}$ where $i_1 < i_2 < \cdots < i_l$. A *decreasing subsequence* is defined similarly with the inequalities among the π_{i_j} reversed. We let

$$\mathrm{lis}\, \pi = \text{length of a longest increasing subsequence of } \pi$$

and

$$\operatorname{lds}\pi = \text{length of a longest decreasing subsequence of } \pi.$$

If $\pi = 5236417$ is the permutation in Figure 7.12, then π has increasing subsequences 2347 and 2367 of length 4 and none longer, so $\operatorname{lis}(\pi) = 4$. Similarly, $\operatorname{lds}(\pi) = 3$ because of the subsequence 531, among others. The reader will notice from Figure 7.12 that the first row of the insertion tableau P (or of the recording tableau Q) has length $4 = \operatorname{lis}(\pi)$ and the length of the first column is $3 = \operatorname{lds}\pi$. This is always the case.

Theorem 7.6.1. *If* $\pi \overset{\mathrm{RS}}{\mapsto} (P, Q)$ *with* $\operatorname{sh} P = \operatorname{sh} Q = \lambda$, *then*

$$\operatorname{lis}\pi = \lambda_1.$$

Proof. Let P_{k-1} be the tableau formed after inserting $\pi_1 \dots \pi_{k-1}$. We claim that if π_k enters P_{k-1} in column j, then the length of a longest increasing subsequence of π ending with π_k is j. Note that the claim proves the theorem since after inserting all of π we will have an increasing sequence of length λ_1 ending at the element P_{1,λ_1}. And there is no longer subsequence since there is no element of P in cell $(1, \lambda_1 + 1)$.

To prove the claim, we induct on k, where the case $k = 1$ is trivial. For the induction step, suppose x is the element in cell $(1, j-1)$ of P_{k-1}. Then there is an increasing subsequence σ of $\pi_1 \dots \pi_{k-1}$ of length $j - 1$ ending in x. Since π_k entered P_{k-1} in a column to the right of x we must have $x < \pi_{k-1}$ by RS2 and RS3. It follows that the concatenation $\sigma\pi_k$ is an increasing subsequence of π of length j ending in π_k.

To show that j is the length of a longest such subsequence suppose, towards a contradiction, that $\tau\pi_k$ is increasing of length greater than j. Let y be the last element of τ. Then, by induction, when y was inserted it entered in a column $j' \geq j$. Since $y < \pi_k$ and rows increase, the element in cell $(1, j)$ just after y's insertion must be less than π_k. And since elements only bump elements larger than themselves, the element in cell $(1, j)$ in P_{k-1} must still be smaller than π_k. But this contradicts the fact that π_k enters in column j since it must bump an element larger than itself. $\square$

Note that the previous proof did *not* show that the first row of P is actually an increasing subsequence of π. In fact, this assertion is false as can be seen in Figure 7.12.

To prove our suspicion about $\operatorname{lds}\pi$, we need to do insertion by columns. Define *column insertion* of $x \notin P$, where P is a partial tableau, using RS1–RS3 but with "row" replaced by "column" everywhere and "leftmost" by "uppermost". Denote the result of column insertion by $c_x(P)$. Amazingly, the row and column operators commute.

Lemma 7.6.2. *Let P be a partial tableau and x, y distinct positive integers with $x, y \notin P$. Then*

$$c_y r_x(P) = r_x c_y(P).$$

Proof. Let

$$m = \max(\{x, y\} \uplus P).$$

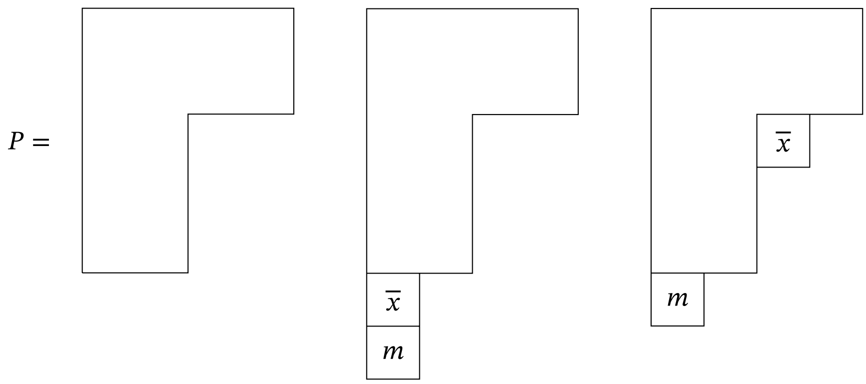


Figure 7.14. The case $y = m$ in the proof of Lemma 7.6.2

Note that by RS2 and RS3, m cannot bump any element during the insertion process. There are two cases depending on which set m comes from.

Case 1: $y = m$. (The case $x = m$ is similar.) Represent P schematically as on the left in Figure 7.14. Since m is the maximum element, c_m will insert m at the end of the first column of whatever tableau to which the operator is applied. Suppose $\overline{x}$ is the last element to be bumped during the insertion $r_x(P)$, and suppose $\overline{x}$ comes to rest in cell u. If u is at the end of the first column, then it is easy to check that $c_m r_x(P)$ and $r_x c_m(P)$ are both the middle diagram in Figure 7.14. Similarly, if u is not at the end of the first column, then both insertions result in the diagram on the right in Figure 7.14.

Case 2: $m \in P$. We induct on $\#P$. The case when $\#P = 1$ is easy to check. Let $\overline{P} = P - \{m\}$, that is, P with m erased from its cell. Using the fact that m never bumps another element as well as induction gives

$$c_y r_x(P) - \{m\} = c_y r_x(\overline{P}) = r_x c_y(\overline{P}) = r_x c_y(P) - \{m\}.$$

So to finish the proof, we need to show that m is in the same position in both $c_y r_x(P)$ and $r_x c_y(P)$. Let $\overline{x}$ be the last element displaced during $r_x(\overline{P})$ and let u be the cell it occupies at the end of the insertion. Similarly define $\overline{y}$ and v for $c_x(\overline{P})$. We now have two subcases depending on the relative locations of u and v.

Subcase 2a: $u = v$. The first two schematic diagrams in Figure 7.15 illustrate $r_x(\overline{P})$ and $c_y(\overline{P})$ in this case. If $(1, j_1)$, $(2, j_2)$, $\ldots$, (k, j_k) are the cells whose elements change during a row insertion, then it is easy to prove that $j_1 \geq j_2 \geq \cdots \geq j_k$. So during $r_x(\overline{P})$ the only columns which are disturbed are those weakly right of the column of u. Similarly, the insertion $c_y(\overline{P})$ only changes rows which are weakly below the row of u. It follows that the insertion paths for r_x and c_y in either order do not intersect until they come to u.

If $\overline{x} < \overline{y}$ (the case $\overline{y} < \overline{x}$ is similar), then $c_y r_x(\overline{P})$ and $r_x c_y(\overline{P})$ will both be as in the last diagram in Figure 7.15. Note also that $\overline{x}$ and $\overline{y}$ must be in the same column since this is clearly true for $c_y r_x(\overline{P})$. Now if m was not in cell u in P, then it will not be bumped

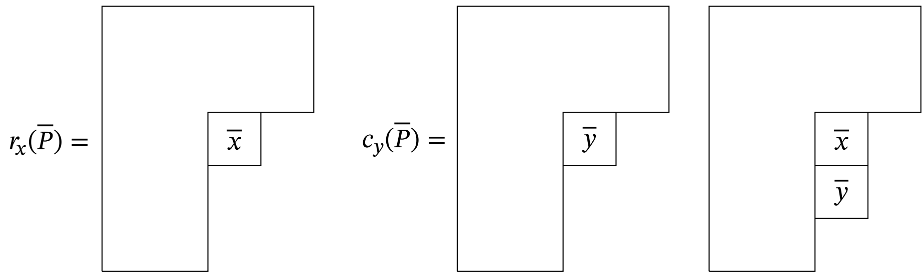


Figure 7.15. The subcase $u = v$ in the proof of Lemma 7.6.2

by either insertion and so will remain in its cell. If m is in cell u, then it is easy to check that it will be bumped into the column just to the right of u in both orders of insertion. This completes this subcase.

Subcase 2b: $u \neq v$. This subcase is taken care of using arguments similar to those in the rest of the proof, so it is left as an exercise. □

If $\pi = \pi_1\pi_2 \ldots \pi_n \in \mathfrak{S}_n$, then its *reversal* (as defined in Exercise 37(a) of Chapter 1) is $\pi^r = \pi_n\pi_{n-1} \ldots \pi_1$. The insertion tableaux of π and π^r are intimately related.

Theorem 7.6.3. *If $P(\pi) = P$, then $P(\pi^r) = P^t$ where t denotes transpose.*

Proof. Clearly, inserting a single element into an empty tableau gives the same result whether it be by rows or columns. Using this and the previous lemma repeatedly

$$\begin{aligned} P(\pi^r) &= r_{\pi_1} \cdots r_{\pi_{n-1}} r_{\pi_n}(\emptyset) \\ &= r_{\pi_1} \cdots r_{\pi_{n-1}} c_{\pi_n}(\emptyset) \\ &= c_{\pi_n} r_{\pi_1} \cdots r_{\pi_{n-1}}(\emptyset) \\ &\vdots \\ &= c_{\pi_n} c_{\pi_{n-1}} \cdots c_{\pi_1}(\emptyset) \\ &= P^t \end{aligned}$$

which is the conclusion we seek. □

We can now characterize $\operatorname{lds}(\pi)$ in terms of the shape of its output tableaux.

Corollary 7.6.4. *If $\pi \overset{\text{RS}}{\mapsto} (P, Q)$ with $\operatorname{sh} P = \operatorname{sh} Q = \lambda$, then*

$$\operatorname{lds} \pi = \lambda_1^t,$$

where λ^t is the transpose of λ.

Proof. Reversing a permutation interchanges increasing and decreasing subsequences so that $\operatorname{lds} \pi = \operatorname{lis} \pi^r$. By Theorem 7.6.1, $\operatorname{lis} \pi^r$ is the length of the first row of $P(\pi^r)$. And $P(\pi^r) = P^t$ by Theorem 7.6.3. So $\operatorname{lds} \pi$ is the length of the first column of P, as desired. □

We note that Greene [**34**] has proved the following extension of Schensted's theorem.

Theorem 7.6.5. *Let* $\operatorname{lis}_k(\pi)$ *be the longest length of a subsequence of* π *which is a union of* k *disjoint increasing subsequences. If* $\pi \overset{\mathrm{RS}}{\mapsto} (P, Q)$ *with* $\operatorname{sh} P = \operatorname{sh} Q = \lambda$*, then*

$$\operatorname{lis}_k(\pi) = \lambda_1 + \lambda_2 + \cdots + \lambda_k$$

and similarly for decreasing subsequences.

Interestingly, there does not seem to be an easy interpretation of the individual λ_i in the shape of the output tableaux. For example, if we consider $\pi = 247951368$, then

$$P(\pi) = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 5 & 6 & 8 \\ \hline 2 & 4 & 9 \\ \cline{1-3} 7 \\ \cline{1-1} \end{array}\ .$$

So $\lambda_1 + \lambda_2 = 5 + 3 = 8$ and $2479 \uplus 1368$ is a union of two increasing subsequences of π and is of length 8. But one can check that there is no length 8 subsequence which is a disjoint union of two increasing subsequences of lengths 5 and 3.

7.7. Differential posets

In this section we will give a second proof of (7.27) based on properties of Young's lattice, Y. This technique can be generalized to a wider class of posets which were introduced and further studied by Stanley [**88**, **90**]. These posets are called differential because of an identity which they satisfy.

To connect the summation side of (7.27) with Y, we will use a simple bijection between standard Young tableaux of shape λ and saturated $\emptyset$–λ chains in Y. Specifically, a $T \in \mathrm{SYT}(\lambda)$ where $\lambda \vdash n$ will be associated with the chain $C : \emptyset = \lambda_0 \lessdot \lambda_1 \lessdot \cdots \lessdot \lambda_n = \lambda$ where λ_k is the shape of the subtableau of T containing the elements $[k]$ for $0 \le k \le n$. An example will be found in Figure 7.16. To go the other way, given a chain C, we define T to be the tableau which has k in the unique cell of the skew partition λ_k/λ_{k-1} where skew partitions were defined in (3.7). It is easy to see that these two maps are inverses of each other. From this discussion, it should be clear that $(f^\lambda)^2$ is the number of pairs of saturated $\emptyset$–λ chains in Y.

In order to work with this observation, we will use a common technique for turning sets into vector spaces. If X is a set, then consider the set of finite formal linear combinations

$$\mathbb{C}X = \left\{ \sum_{x \in X} c_x x \mid c_x \in \mathbb{C} \text{ for all } x \text{ and only finitely many } c_x \neq 0 \right\}. \tag{7.32}$$

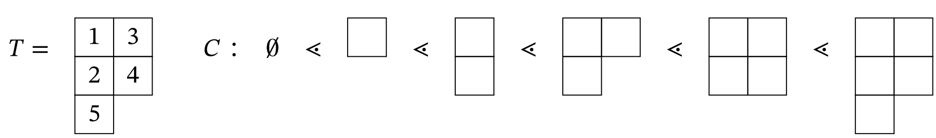


Figure 7.16. The bijection between SYT and saturated chains in Young's lattice

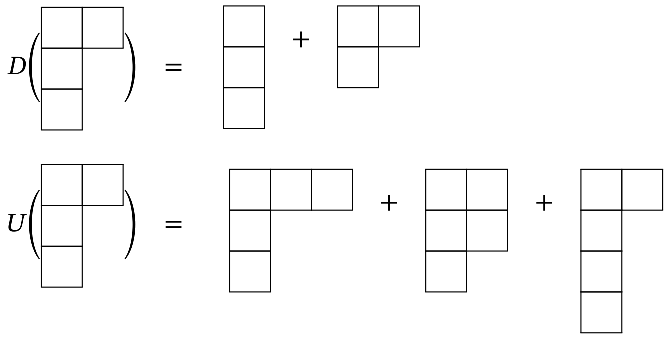


Figure 7.17. The down and up operators in Young's lattice

Now $\mathbb{C}X$ is a vector space with vector addition and scalar multiplication given by

$$\sum_x c_x x + \sum_x d_x x = \sum_x (c_x + d_x)x,$$
$$c\sum_x c_x x = \sum_x c c_x x.$$

Note that X is a basis for $\mathbb{C}X$.

We will define two linear operators on $\mathbb{C}Y$. The *down operator* $D : \mathbb{C}Y \to \mathbb{C}Y$ is defined by

$$D(\lambda) = \sum_{\lambda^- \lessdot \lambda} \lambda^-$$

and linear extension. An example is given in Figure 7.17. It will be useful to think of $D(\lambda)$ as the sum of all partitions which can be reached by taking a walk of length one downward from λ in Y viewed as a graph. Note that $D(\emptyset)$ is the empty sum so that $D(\emptyset) = 0$, the zero vector. Similarly, the *up operator* $U : \mathbb{C}Y \to \mathbb{C}Y$ is

$$U(\lambda) = \sum_{\lambda^+ \gtrdot \lambda} \lambda^+.$$

Again, Figure 7.17 contains an example and a similar walk interpretation holds.

We claim that

$$D^n U^n(\emptyset) = \left(\sum_{\lambda \vdash n} (f^\lambda)^2\right)\emptyset. \tag{7.33}$$

Indeed, the coefficient of $\lambda \vdash n$ in $U^n(\emptyset)$ is the number of walks from $\emptyset$ to λ in Y which always go up. But such a walk is just a saturated $\emptyset$–λ chain so that, by the previous bijection,

$$U^n(\emptyset) = \sum_{\lambda \vdash n} f^\lambda \lambda.$$

By the same token

$$D^n \lambda = f^\lambda \emptyset$$

since walks which always go down also follow saturated chains. So applying D^n to the expression for $U^n(\emptyset)$ and using linearity gives the desired equality.

To make use of (7.33), we need a closer investigation of the structure of Y. Say that $\lambda \in Y$ *covers k elements* if $\#\{\lambda^- \mid \lambda^- \lessdot \lambda\} = k$. Similarly define the phrase "is covered by k elements". Also say that $\lambda, \mu \in Y$ *cover l elements* if $\#\{\nu \mid \nu \lessdot \lambda, \mu\} = l$ and ditto for being covered.

Proposition 7.7.1. *The poset Y has the following two properties for all distinct* $\lambda, \mu \in Y$.

(a) *λ covers k elements if and only if it is covered by $k+1$ elements.*

(b) *λ, μ cover l elements if and only if they are covered by l elements. In this case* $l \leq 1$.

Proof. (a) It suffices to prove the forward direction since then the number of elements which cover λ is uniquely determined by the number which it covers. The elements which λ covers are precisely those obtained by removing an inner corner of λ. And those which cover λ are the partitions obtained by adding an outer corner to λ. But inner corners and outer corners alternate along the southeast boundary of λ, beginning with an outer corner at the end of the first row and ending with an outer corner at the end of the first column. The result follows.

(b) Again, we only need to prove the forward implication. There are two cases. If there is an element $\nu \lessdot \lambda, \mu$, then, since Y is ranked, it must be $\operatorname{rk}\lambda = \operatorname{rk}\mu = n$ for some n and $\operatorname{rk}\nu = n-1$. Since Y is a lattice, it follows that $\nu = \lambda \wedge \mu = \lambda \cap \mu$ is unique and so $l = 1$. Also $|\lambda \cap \mu| = |\nu| = n-1$ so that $|\lambda \vee \mu| = |\lambda \cup \mu| = n+1$. It follows that λ, μ are covered by a unique element, namely $\lambda \vee \mu$.

If there is no element covered by both λ, μ, then similar considerations show that no element covers λ, μ. We leave this verification to the reader. □

We can translate this result in terms of the down and up operators.

Proposition 7.7.2. *The operators D, U on Y satisfy*

$$DU - UD = I \tag{7.34}$$

where I is the identity map.

Proof. By linearity, it suffices to show that this equation is true when applied to a basis element $\lambda \in \mathbb{C}Y$. First consider $DU(\lambda)$. The coefficient of μ in this expression is the number of walks λ to μ which first go up an edge of Y and then come down an edge (possibly the same one). These are precisely the walks of length 2 going through some element covering both λ and μ. From the previous proposition, we get

$$DU(\lambda) = (k+1)\lambda + \sum \mu$$

where $k+1$ elements cover λ and the sum is over all $\mu \neq \lambda$ such that μ, λ are covered by a common element. In a similar way

$$UD(\lambda) = k\lambda + \sum \mu$$

where the sum is over the same set of μ. Subtracting the two equalities gives (7.34). □

Note that (7.34) is reminiscent of an identity from calculus. Consider a differentiable function $f(t)$. Let D stand for differentiation and let U be multiplication by t. Then

$$DU(f(t)) = (tf(t))' = f(t) + tf'(t) = I(f(t)) + UD(f(t))$$

which is just (7.34) with the negative term moved to the other side of the equation. We will need an extension of (7.34) where U is replaced by an arbitrary operator which is a polynomial in U.

Corollary 7.7.3. *For any polynomial $p(t) \in \mathbb{C}[t]$ we have*

$$Dp(U) = p'(U) + p(U)D$$

where $p'(t)$ is the derivative of $p(t)$.

Proof. By linearity it suffices to prove this result for the powers U^n for $n \geq 0$. The base case $n = 0$ is easy to check. Assuming the result is true for n and then applying (7.34) we obtain

$$\begin{aligned} DU^{n+1} &= (DU^n)U \\ &= (nU^{n-1} + U^nD)U \\ &= nU^n + U^n(I + UD) \\ &= (n+1)U^n + U^{n+1}D \end{aligned}$$

as desired. □

We are now ready to reprove (7.27) which we restate here for ease of reference:

$$\sum_{\lambda \vdash n} (f^\lambda)^2 = n!\,. \tag{7.35}$$

Proof. Because of (7.33), it suffices to show that $D^nU^n(\emptyset) = n!\,\emptyset$. We induct on n, where the case $n = 0$ is trivial since $D^0U^0 = I$. Applying Corollary 7.7.3, the fact that $D\emptyset = 0$, and induction gives

$$\begin{aligned} D^nU^n(\emptyset) &= D^{n-1}(DU^n)(\emptyset) \\ &= D^{n-1}(nU^{n-1} + U^nD)(\emptyset) \\ &= nD^{n-1}U^{n-1}(\emptyset) + 0(\emptyset) \\ &= n(n-1)!\;\emptyset \end{aligned}$$

which is what we wished to show. □

Stanley generalized these ideas using the following definition. Call a poset P *differential* if it satisfies the following three properties where x, y are distinct elements of P.

DP1 P is ranked.
DP2 If x covers k elements, then it is covered by $k + 1$ elements.
DP3 If x, y cover l elements, then they are covered by l elements.

From what we have proved, Y is a differential poset. Another example is given in Exercise 29. In fact, Stanley defined a more general type of poset called r-differential which will be studied in Exercise 30.

As with Young's lattice, we can show that the parameter l must satisfy $l \leq 1$.

Lemma 7.7.4. *If poset P satisfies DP1 and DP3, then $l \leq 1$.*

Proof. As noted in the proof of Proposition 7.7.1, DP3 implies that its converse is also true. Suppose the lemma is false and pick a pair x, y with $l \geq 2$. Since P is ranked by DP1, we must have $\operatorname{rk} x = \operatorname{rk} y$. Pick the counterexample pair x, y to be of minimum rank and let x', y' be two of the elements covered by x, y. But since x', y' are covered by at least two elements, they must cover at least two elements. This contradicts the fact that we took a minimum-rank pair. □

We want to define up and down operators in a differential poset. But to make sure they are well-defined, the sums need to be finite.

Lemma 7.7.5. *If P satisfies* DP1 *and* DP2*, then its nth rank* $\operatorname{Rk}_n P$ *is finite for all $n \geq 0$.*

Proof. We induct on n. Since P is ranked by DP1, it has a $\hat{0}$ and so the result holds for $n = 0$. Assume the lemma through rank n. Now any $x \in \operatorname{Rk}_n P$ covers at most $\# \operatorname{Rk}_{n-1} P$ elements. So, by DP2,

$$\# \operatorname{Rk}_{n+1} P \leq (\# \operatorname{Rk}_n P)(1 + \# \operatorname{Rk}_{n-1} P)$$

which forces $\operatorname{Rk}_{n+1} P$ to be finite. □

Thus we can define two operators on $\mathbb{C}P$ by

$$D(x) = \sum_{x^- \lessdot x} x^-$$

and

$$U(x) = \sum_{x^+ \gtrdot x} x^+.$$

The proof of the next result is similar enough to that of Proposition 7.7.2 that it is left as an exercise.

Proposition 7.7.6. *Let P be a ranked poset with* $\operatorname{Rk}_n P$ *finite for all $n \geq 0$. Then*

$$P \text{ is differential} \iff DU - UD = I. \qquad \square$$

Also, the reader should be able to generalize the operator proof of (7.27) to the setting of differential posets and show the following.

Theorem 7.7.7. *In any differential poset P we have*

$$\sum_{x \in \operatorname{Rk}_n P} (f^x)^2 = n!$$

where f^x is the number of saturated $\hat{0}$–x chains. □

The reader might wonder if there is a way to give a bijective proof of this theorem just as the Robinson–Schensted algorithm provides a bijection for (7.27). This has been done by Fomin as part of his theory of duality of graded graphs [**27**, **28**].

7.8. The chromatic symmetric function

Stanley [**91**] defined a symmetric function associated with graph colorings which generalizes the chromatic polynomial. In this section we will prove some of his results about this function, including expressions for its expansion in the monomial and power sum bases for Sym.

Let G be a graph with vertex set $V = \{v_1, \ldots, v_n\}$. Consider a coloring $c\colon V \to \mathbb{P}$ of G using the positive integers as color set. Then c has monomial

$$\mathbf{x}^c = x_{c(v_1)}x_{c(v_2)} \cdots x_{c(v_n)}. \tag{7.36}$$

For example, if G is the graph on the left in Figure 7.18 and c is the coloring on the right, then

$$\mathbf{x}^c = x_{c(u)}x_{c(v)}x_{c(w)}x_{c(z)} = x_1x_2^2x_4.$$

Now define the *chromatic symmetric function* of G to be

$$X(G) = X(G;\mathbf{x}) = \sum_{c\colon V\to\mathbb{P}} \mathbf{x}^c$$

where the sum is over all proper colorings $c\colon V \to \mathbb{P}$. To illustrate, let us return to the graph of Figure 7.18. Because G contains a triangle, we must use three or four colors for a proper coloring. If we use four different colors, then this will give rise to a monomial $x_ix_jx_kx_l$ for some distinct i, j, k, l. And any of the $4! = 24$ ways of assigning these colors to the four vertices is proper. Since this count is independent of which four colors we use, the contribution of such colorings to $X(G)$ is $24m_{1^4}$. If we use three colors, then one of them must be used twice and so correspond to a monomial $x_i^2x_jx_k$ for distinct i, j, k. One copy of color i must go on vertex w and the other can be on u or x, giving two choices. The other two colors can be distributed among the remaining two vertices in two ways. So these colorings give a term $4m_{21^2}$. In total $X(G) = 24m_{1^4} + 4m_{21^2}$ which the reader will note is a symmetric function. We will now prove that this is always the case, as well as showing a connection with the chromatic polynomial of G.

Proposition 7.8.1. *Let G be a graph with vertex set V.*

(a) $X(G) \in \mathrm{Sym}_n$ *where* $n = \#V$.

(b) *If we set* $x_1 = \cdots = x_t = 1$ *and* $x_i = 0$ *for* $i > t$, *written* $\mathbf{x} = 1^t$, *then*

$$X(G;1^t) = P(G;t).$$

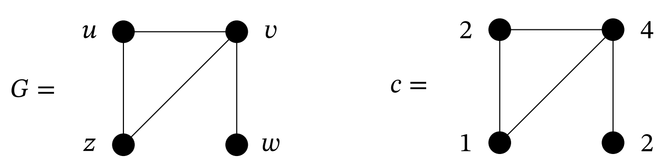


Figure 7.18. A graph and a coloring using $\mathbb{P}$

Proof. (a) It is clear that $\mathbf{x}^c$ has n factors for any coloring c so that $X(G)$ is homogeneous of degree n. To show that it is symmetric, note that any permutation of the colors of a proper coloring is proper. This means that permuting the subscripts in $X(G)$ leaves it invariant; that is, $X(G)$ is symmetric.

(b) The given substitution results in $\mathbf{x}^c = 1$ if c only uses colors from $[t]$ and $\mathbf{x}^c = 0$ otherwise. So $X(G; 1^t)$ is just the number of proper colorings $c\colon V \to [t]$. But this was the definition of $P(G; t)$. □

Since $X(G)$ is symmetric, we can expand it in terms of various bases for the symmetric functions and see if the coefficients have any nice combinatorial interpretation. We start with the monomial basis. To describe the coefficients, we will need some definitions. If $G = (V, E)$ is a graph, then $W \subseteq V$ is *independent* or *stable* if there is no edge of G between any pair of vertices of W. For example, if $G = T_1$ as in Figure 1.9, then $W = \{2, 5, 6\}$ is stable but $W = \{2, 3, 6\}$ is not because of the edge 36. The reason we care about stable sets is that if c is a proper coloring of G, then the set of all vertices with a given color r, in other words the vertices in $c^{-1}(r)$, form a stable set. Similarly, call a partition $\rho = B_1/\ldots/B_k$ of V *independent* or *stable* if each block is. Returning to Figure 1.9, the partition 13/256/4 is stable in T_1. The *type* of a set partition $\rho = B_1/\ldots/B_k$ is the integer partition $\lambda(\rho) = (\lambda_1, \ldots, \lambda_k)$ obtained by arranging $\#B_1, \ldots, \#B_k$ in weakly decreasing order. To illustrate, $\lambda(1456/27/38) = (4, 2, 2)$. For a graph G, let

$$i_\lambda(G) = \text{number of independent partitions of } V \text{ of type } \lambda.$$

For the graph in Figure 7.18 we have $i_{1^4}(G) = 1$, $i_{21^2}(G) = 2$ and all other $i_\lambda(G) = 0$. Any proper coloring c of G induces a stable partition of V whose blocks are the nonempty $c^{-1}(r)$ for the colors r in the color set. Finally, if $\lambda = (1^{m_1}, 2^{m_1}, \ldots, n^{m_n})$ is an integer partition in multiplicity notation, then let

$$\lambda^! = m_1!\, m_2! \cdots m_n!\,.$$

Theorem 7.8.2. *If graph G has $\#V = n$, then*

$$X(G) = \sum_{\lambda \vdash n} i_\lambda(G) \lambda^! m_\lambda.$$

Proof. If $\lambda = (\lambda_1, \ldots, \lambda_k)$, then the coefficient of $x_1^{\lambda_1} \cdots x_k^{\lambda_k}$ is the coefficient of m_λ since $X(G)$ is symmetric. And the coefficient of this monomial is the number of proper colorings $c\colon V \to [k]$ where i gets used λ_i times for all i. By the discussion preceding this theorem, these colorings can be obtained by taking an independent partition $\rho \vdash V$ and then deciding which colors get assigned to which blocks of ρ. The number of choices for ρ is $i_\lambda(G)$. Now any of the m_j colors which are used j times can be used on any of the m_j blocks of ρ of size j. The number of such assignments is $m_j!$ and this is true for all j. This gives the factor of $\lambda^!$. □

The expansion of $X(G)$ in the power sum basis will be found by Möbius inversion. Let $G = (V, E)$ be a graph. Then any spanning subgraph H can be identified with its set of edges $E(H)$ since we have $V(H) = V(G)$. To illustrate, for the graph G in Figure 3.5 we would identify F_1 with the edge set $\{12, 24\}$ and F_2 with the edge set $\{14, 24\}$. Given

$F \subseteq E$, we get a partition $\rho(F)$ of the vertex set where a block of ρ is the set of vertices in a component of the corresponding spanning subgraph. Returning to our example, F_1 and F_2 both have partition $\rho = 124/3$. Let

$$\lambda(F) = \text{type of the partition } \rho(F).$$

In our running example $\lambda(F_1) = \lambda(F_2) = (3, 1)$.

Theorem 7.8.3. *If $G = (V, E)$ is a graph, then*

$$X(G) = \sum_{F \subseteq E} (-1)^{\#F} p_{\lambda(F)}$$

where the sum is over all subsets F of the edge set of G and $\#F$ is the number of edges in F.

Proof. Consider the Boolean algebra B_E of all subsets of E ordered by containment. Given $F \subseteq E$, we define the power series

$$\alpha(F) = \sum_c \mathbf{x}^c$$

where the sum is over all colorings $c\colon V \to \mathbb{P}$ such that $c(u) = c(v)$ for all $uv \in F$. These are the colorings which are monochromatic on each component of F's spanning subgraph. So if $\rho(F) = B_1/\ldots/B_k$, then each B_i can get any of the colors in $\mathbb{P}$. It follows that

$$\alpha(F) = \prod_{i=1}^{k} (x_1^{\#B_i} + x_2^{\#B_i} + \cdots) = p_{\lambda(F)}. \tag{7.37}$$

Also define

$$\beta(F) = \sum_c \mathbf{x}^c$$

where the sum is over all colorings $c\colon V \to \mathbb{P}$ such that $c(u) = c(v)$ for all $uv \in F$ and $c(u) \neq c(v)$ for $uv \in E - F$. So these colorings are constant on the components of F but also cannot have any other edge of E monochromatically colored. Given any spanning subgraph F' and a coloring c appearing in $\alpha(F')$, one can define a unique spanning subgraph F by letting $uv \in F$ if and only if $c(u) = c(v)$. By the definition of α it is clear that $F \supseteq F'$. Also, the definition of F implies that c is a coloring in the sum for $\beta(F)$. It follows that

$$\alpha(F') = \sum_{F \supseteq F'} \beta(F)$$

for all $F' \subseteq E$. Applying Möbius inversion (Theorem 5.5.5) as well as (5.6) and (7.37) gives

$$\beta(\emptyset) = \sum_{F \in B_E} \mu(F)\alpha(F) = \sum_{F \subseteq E} (-1)^{\#F} p_{\lambda(F)}.$$

But $\beta(\emptyset)$ is the generating function for all colorings $d\colon V \to \mathbb{P}$ such that $d(u) \neq d(v)$ for $uv \in E$, and these are exactly the proper colorings. Thus $\beta(\emptyset) = X(G)$ and we are done. □

We end this section with an open question about $X(G)$. To appreciate it, we first prove a result about the chromatic polynomial.

Proposition 7.8.4. *Let T be a graph with $\#V = n$. We have that T is a tree if and only if*

$$P(T;t) = t(t-1)^{n-1}.$$

Proof. We will prove the forward direction and leave the reverse implication as an exercise. If $v \in V$, then we color T by first coloring v, then all the neighbors of v, then all the uncolored vertices which are neighbors of neighbors of v, etc. The number of ways to color v is t. Because T is connected and acyclic, when coloring each vertex $w \in V - \{v\}$ we will have w adjacent to exactly one already colored vertex. So the number of colors available for w is $t-1$. The result follows. □

This proposition is sometimes summarized by saying that the *chromatic polynomial does not distinguish trees* since all trees on n vertices have the same polynomial. Stanley asked if the opposite was true for his chromatic symmetric function.

Question 7.8.5. If T_1 and T_2 are nonisomorphic trees, is it true that $X(T_1) \neq X(T_2)$?

It has been checked for trees with up to 23 vertices that the answer to this question is "yes" and there are other partial results in the literature.

7.9. Cyclic sieving redux

Cyclic sieving phenomena are often associated to results in representation theory. In this section we will give a second proof of Theorem 6.6.2 using this approach. To do so, we will assume the reader is familiar with the material in the appendix in this book. We start by presenting a general paradigm for proving a CSP by using group actions.

Recall that we start with a set X, a cyclic group G acting on X, and a polynomial $f(q) \in \mathbb{N}[q]$. Also, for the rest of this section we let $\omega = e^{2\pi i/n}$. The triple $(X, G, f(q))$ was said to exhibit the cyclic sieving phenomenon if, for all $g \in G$,

$$\#X^g = f(\gamma) \tag{7.38}$$

where γ is a root of unity satisfying $o(\gamma) = o(g)$.

To interpret the left side of (7.38), consider the permutation representation $\mathbb{C}X$ of G. Since $g \in G$ takes each basis element in X to another basis element, the matrix $[g]_X$ consists of zeros and ones. And there is a one on the diagonal precisely when $gx = x$ for $x \in X$. So this representation has character

$$\chi(g) = \operatorname{tr}[g]_X = \#X^g. \tag{7.39}$$

As for the right-hand side of (7.38), let h be a generator of G where $\#G = n$ and let $\omega = \omega_n$. For $i \geq 0$, let $V^{(i)}$ be the irreducible G-module such that the matrix of h is $[\omega^i]$ and let its character be $\chi^{(i)}$. Suppose $f = \sum_{i\geq 0} m_i q^i$ where $m_i \in \mathbb{N}$ for all i. Since the coefficients are nonnegative integers, we can define a corresponding G-module

$$V_f = \bigoplus_{i\geq 0} m_i V^{(i)}$$

with character χ^f. Let $g = h^j$ and $\gamma = \omega^j$. Now using (A.1) and the fact that the $V^{(i)}$ are 1-dimensional,

$$\chi^f(g) = \sum_{i\geq 0} m_i \chi^{(i)}(h^j) = \sum_{i\geq 0} m_i \omega^{ij} = f(\omega^j) = f(\gamma)$$

which is the right side of (7.38). Now appealing to Theorem A.1.4, we have proved the following result or Reiner, Stanton, and White [**72**].

Theorem 7.9.1. *The triple $(X, G, f(q))$ exhibits the cyclic sieving phenomenon if and only if $\mathbb{C}X \cong V_f$ as G-modules.* □

We will now give a second proof that the triple

$$\left(\left(\!\!\binom{[n]}{k}\!\!\right), \langle(1, 2, \ldots, n)\rangle, \begin{bmatrix} n+k-1 \\ k \end{bmatrix}_q\right) \tag{7.40}$$

exhibits the CSP. We will write vectors in boldface to distinguish them from scalars. To use the previous theorem, any G-module isomorphic to $\mathbb{C}X$ will suffice. So we will construct one whose structure is easy to analyze. Let V be a vector space of dimension n and fix a basis $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ for V. Let $P^k(B)$ be the set of all formal polynomials of degree k using the elements of B as variables and the complex numbers as coefficients. Clearly $P^k(B)$ is a vector space with basis

$$B' = \{\mathbf{b}_{i_1}\mathbf{b}_{i_2}\cdots\mathbf{b}_{i_k} \mid 1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n\} \tag{7.41}$$

In particular, if one takes $V = \mathbb{C}[n]$ with the basis $B = \{\mathbf{i} \mid i \in [n]\}$, then we will write $P^k(n)$ for $P^k(B)$. To illustrate,

$$P^2(3) = \{c_1\mathbf{11} + c_2\mathbf{22} + c_3\mathbf{33} + c_4\mathbf{12} + c_5\mathbf{13} + c_6\mathbf{23} \mid c_i \in \mathbb{C} \text{ for } 1 \leq i \leq 6\}.$$

One can turn $P^k(n)$ into a G_n-module for $G_n = \langle(1, 2, \ldots, n)\rangle$ by letting

$$\mathrm{g}(\mathbf{i}_1\mathbf{i}_2\cdots\mathbf{i}_k) = \mathrm{g}(\mathbf{i}_1)\mathrm{g}(\mathbf{i}_2)\cdots\mathrm{g}(\mathbf{i}_k) \tag{7.42}$$

for $g \in G_n$ and extending linearly. It should be clear from the definitions that

$$\mathbb{C}\left(\!\!\binom{[n]}{k}\!\!\right) \cong P^k(n) \tag{7.43}$$

as G_n-modules. So we will use the latter in establishing the CSP.

The advantage of using $P^k(n)$ is that it is also a GL_n-module. In particular, we can define the action of $g \in \mathrm{GL}_n$ exactly as in (7.42), where $\mathbf{i}_j$ is thought of as the jth coordinate vector and the linear combinations $g(\mathbf{i}_1), \ldots, g(\mathbf{i}_k)$ are multiplied together formally to get an element of $P^k(n)$. For example, if $n = 3$, $k = 2$, and

$$\mathrm{g} = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

then

$$\mathrm{g}(\mathbf{13}) = \mathrm{g}(\mathbf{1})\mathrm{g}(\mathbf{3}) = (\mathbf{1} + 3\,\mathbf{2})(-\mathbf{3}) = -\mathbf{13} - 3\,\mathbf{23}.$$

It is not hard to show that this is an action. We will also need the fact that if $\tilde{B}$ is any other basis for $\mathbb{C}[n]$, then the monomials defined by (7.41) (where one replaces each $\mathbf{b}_{i_j}$ by the corresponding element of $\tilde{B}$) also form a basis for $P^k(n)$.

We now compute the character of $P^k(n)$. Notice from Corollary A.1.2 that, since G_n is cyclic, there is a basis $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ for $\mathbb{C}[n]$ which diagonalizes $[g]$ for all $g \in G_n$; say

$$[g]_B = \operatorname{diag}(x_1, x_2, \ldots, x_n).$$

To compute the action of G_n in $P^k(n)$, we use the basis B' in (7.41) and see that

$$g(\mathbf{b}_{i_1}\mathbf{b}_{i_2}\cdots\mathbf{b}_{i_k}) = g(\mathbf{b}_{i_1})g(\mathbf{b}_{i_2})\cdots g(\mathbf{b}_{i_k}) = x_{i_1}x_{i_2}\cdots x_{i_k}\mathbf{b}_{i_1}\mathbf{b}_{i_2}\cdots\mathbf{b}_{i_k}.$$

So B' diagonalizes this action with

$$[g]_{B'} = \operatorname{diag}(x_{i_1}x_{i_2}\cdots x_{i_k} \mid 1 \le i_1 \le i_2 \le \cdots \le i_k \le n).$$

This gives the character

$$\chi'(g) = \sum_{1\le i_1\le i_2\le\ldots\le i_k\le n} x_{i_1}x_{i_2}\cdots x_{i_k} = h_k(x_1, x_2, \ldots, x_n)$$

which is just a complete homogeneous symmetric polynomial in the eigenvalues. For example, if $n = 3$ and $k = 2$, then we would have a basis $B = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ such that $[g]_B = \operatorname{diag}(x_1, x_2, x_3)$. So in $P^2(3)$

$$\begin{array}{lll} g(\mathbf{aa}) = x_1^2\mathbf{aa}, & g(\mathbf{bb}) = x_2^2\mathbf{bb}, & g(\mathbf{cc}) = x_3^2\mathbf{cc}, \\ g(\mathbf{ab}) = x_1x_2\mathbf{ab}, & g(\mathbf{ac}) = x_1x_3\mathbf{ac}, & g(\mathbf{bc}) = x_2x_3\mathbf{bc}, \end{array}$$

which gives

$$\chi'(g) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3.$$

To prove the CSP we will need to relate homogeneous symmetric polynomials to q-binomial coefficients. This is done via the *principal specialization* which sets $x_i = q^{i-1}$ for $i \ge 1$.

Proposition 7.9.2. *We have the principal specializations*

$$e_k(1, q, \ldots, q^{n-1}) = q^{\binom{k}{2}} \begin{bmatrix} n \\ k \end{bmatrix}_q$$

and

$$h_k(1, q, \ldots, q^{n-1}) = \begin{bmatrix} n+k-1 \\ k \end{bmatrix}_q.$$

Proof. We will prove the identity for h_k, leaving the one for e_k as an exercise. From the definition of the complete homogeneous symmetric functions we see that

$$h_k(1, q, \ldots, q^{n-1}) = \sum_{0\le j_1\le j_2\le\cdots\le j_k\le n-1} q^{j_1}q^{j_2}\cdots q^{j_k}.$$

But a sequence $j_1 \le j_2 \le \cdots \le j_k$ corresponds to an integer partition λ obtained by listing the nonzero elements of the sequence in weakly decreasing order. Furthermore $q^{j_1}q^{j_2}\cdots q^{j_k} = q^{|\lambda|}$ and the bounds on the j_i imply that $\lambda \in \mathcal{R}(k, n-1)$, the set of partitions contained in a $k \times (n-1)$ rectangle. The equality now follows from Theorem 3.2.5. □

We are now ready to complete the representation-theoretic demonstration that the triple (7.40) exhibits the cyclic sieving phenomenon. Consider $[(1,2,\ldots,n)]$ acting as a linear transformation on $\mathbb{C}[n]$. Its characteristic polynomial is $x^n - 1$ which has roots $1, \omega, \omega^2, \ldots, \omega^{n-1}$. So, by Corollary A.1.2, there is a diagonalizing basis B for the action of $G_n = \langle(1,2,\ldots,n)\rangle$ with

$$[(1,2,\ldots,n)]_B = \operatorname{diag}(1, \omega, \omega^2, \ldots, \omega^{n-1}).$$

Since any $g \in G_n$ has the form $g = (1,2,\ldots,n)^j$ for some j and since the generator has been diagonalized, we have

$$[g]_B = \operatorname{diag}(1^j, \omega^j, \omega^{2j}, \ldots, \omega^{(n-1)j}) = \operatorname{diag}(1, \gamma, \gamma^2, \ldots, \gamma^{n-1})$$

where $\gamma = \omega^i$ is a primitive $o(g)$th root of unity. But the previous theorem and the discussion just preceding it show that

$$\chi'(g) = h_k(1, \gamma, \gamma^2, \ldots, \gamma^{n-1}) = \left[\begin{matrix} n+k-1 \\ k \end{matrix}\right]_\gamma$$

where χ' is the character of $P^k(n)$. But, by (7.43),

$$\chi'(g) = \#\left(\!\!\binom{[n]}{k}\!\!\right)^g.$$

Equating the last two displayed equations completes the proof.

Exercises

(1) (a) Show that (7.2) satisfies the definition of a group action.
   (b) Show that Sym is an algebra; that is, it is a vector space which is closed under multiplication of symmetric functions.

(2) Prove Proposition 7.1.2(b).

(3) Prove Theorem 7.1.3(b).

(4) (a) Prove that lexicographic order is a total order on partitions.
   (b) Prove that adding parts of a partition makes it larger in lexicographic order.
   (c) Prove that lexicographic order on partitions is a linear extension of dominance order as introduced in Section 1.12.
   (d) Prove that the lexicographic order inequalities in the proof of Theorem 7.1.3(a) and (b) can be strengthened to dominance order inequalities.
   (e) Show that part (d) is also true in the statement of Theorem 7.2.2.

(5) Show that $\mathfrak{S}_n$ is generated by the adjacent transpositions $(i, i+1)$ for $1 \leq i < n$. Hint: Induct on $\operatorname{inv}\pi$ for $\pi \in \mathfrak{S}_n$.

(6) Consider the proof of Theorem 7.2.3.
   (a) Show in part (a) that every path family whose associated permutation is not the identity must intersect.

(b) In part (a), provide a description of the inverse of the map from lattice path families to SSYT. Be sure to prove that it is well-defined and indeed an inverse to the forward map.
(c) Prove part (b).

(7) Verify that the real sequence $a_0, \ldots, a_n$ is log-concave if and only if $a_0/r, \ldots, a_n/r$ is, where $r \in \mathbb{R} - \{0\}$.

(8) Use Theorem 7.2.5 to prove log-concavity of the following sequences.
(a) $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$.
(b) $c(n,0), c(n,1), \ldots, c(n,n)$ (signless first kind Stirling numbers).

(9) Prove equation (7.9).

(10) Prove equation (7.13).

(11) A *plane partition of shape* λ is a filling P of the cells of λ with positive integers so that rows and columns weakly decrease. Let pp_n be the number of plane partitions P (of any shape) such that $|P| = n$. Show that

$$\sum_{n \geq 0} pp_n x^n = \prod_{i \geq 1} \frac{1}{(1 - x^i)^i}$$

in two ways: by taking a limit in (7.14) and by providing a proof in the spirit of the Hillman–Grassl algorithm.

(12) (a) Prove that the output array of HG1–HG3 is a reverse plane partition.
(b) Prove that (7.16) is a total order.
(c) Show that in GH3 of the Hillman–Grassl construction, the reverse path p must reach the rightmost cell in row a.
(d) Prove that after adding ones along the reverse path, the result is still a reverse plane partition.

(13) (a) Construct the inverse of the map used in the proof of (7.20).
(b) Prove (7.21).
(c) Prove (7.24).

(14) Complete the proof of Lemma 7.4.3.

(15) Derive (7.26) from (7.25) using a bijection.

(16) (a) Show that the P_λ^*-partitions are exactly the reverse plane partitions of shape λ.
(b) Show that for any finite poset P we have $\#\mathcal{L}(P) = \#\mathcal{L}(P^*)$.

(17) (a) Let τ be a poset. Call τ a *rooted tree* if it has a $\hat{0}$ and its Hasse diagram is a tree in the graph-theoretic sense of the term. If $\#\tau = n$, then a *natural labeling* of τ is an order-preserving bijection $\tau \to [n]$. See Figure 7.19 for an example of a rooted tree (on the left) and a natural labeling (in the middle). Let f^τ be the number of natural labelings of τ. Define the *hooklength* of $v \in \tau$ to be

$$h_v = \#U(v)$$

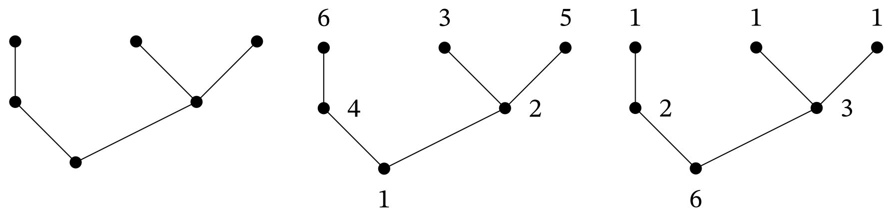


Figure 7.19. A rooted tree poset, a natural labeling, and its hooklengths

where $U(v)$ is the upper-order ideal generated by v. The right-hand tree in Figure 7.19 lists its hooklengths. Prove that if $\#\tau = n$, then

$$f^\tau = \frac{n!}{\prod_{v\in\tau} h_v}$$

in two ways: probabilistically and using induction on n.

(b) The *comb* is the infinite poset on the left in Figure 7.20. Let L_n be the set of lower-order ideals of the comb which have n elements. The three elements of L_3 are displayed on the right in Figure 7.20. Note that the last two order ideals are considered distinct even though they are isomorphic as posets. Show that $\#L_n = f_n$, the Fibonacci numbers defined in (1.2).

(c) Using the notation of part (a), show that

$$\sum_{\tau\in L_n} (f^\tau)^2 = n!\,.$$

(18) (a) Show that if P is a PYT and $x \notin P$, then $P' = r_x(P)$ is still a PYT; that is, the rows and columns of P' still increase.

(b) Show that the RSK map is well-defined in that T and U are both semistandard.

(19) Consider three sets of variables $\mathbf{x} = \{x_i\}_{i\ge1}$, $\mathbf{y} = \{y_j\}_{j\ge1}$, and $\mathbf{xy} = \{x_iy_j\}_{i,j\ge1}$. Prove that

$$\sum_{\lambda\vdash n} s_\lambda(\mathbf{x})s_\lambda(\mathbf{y}) = s_n(\mathbf{xy})$$

where the variables in $\mathbf{xy}$ can be substituted into the Schur function in any order since s_n is symmetric. Hint: Use the Cauchy identity, equation (7.30).

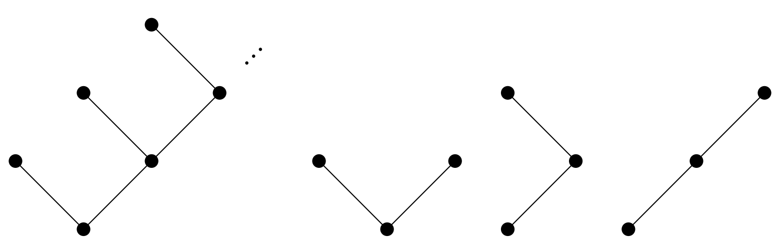


Figure 7.20. The comb and its three lower-order ideals with 3 elements

(20) Prove Theorem 7.5.4. Hint: First prove that if T is an SSYT into which one inserts x and y in that order with $x \le y$, then the box added to the shape by the insertion of y is strictly to the right of the box for the insertion of x.

(21) Show that (7.27) can be derived from (7.30) by taking the coefficient of $x_1 \cdots x_n y_1 \cdots y_n$ on both sides.

(22) Fill in the details of the proof of Lemma 7.6.2.

(23) If π is a two-line array, then let $\hat{\pi}$ be the upper line and let $\check{\pi}$ be the lower line.
   (a) Show that any two-line array π with entries in $\mathbb{P}$ and columns which are lexicographically weakly increasing corresponds to a matrix $M \in \mathrm{Mat}$.
   (b) Show that if $\mathrm{RSK}(\pi) = (T, U)$ with $\operatorname{sh} T = \operatorname{sh} U = \lambda$, then

$$\lambda_1 = \text{the length of a longest weakly increasing subsequence of } \check{\pi}$$

   by mimicking the proof of Theorem 7.6.1.
   (c) Let T be a semistandard Young tableau and suppose that the content of T is $\operatorname{co} T = [\alpha_1, \alpha_2, \ldots, \alpha_k]$. The *standardization* of T is the tableau $\operatorname{std} T$ obtained by replacing the ones in T by the numbers $1, 2, \ldots, \alpha_1$ from left to right, replacing the twos in T by the numbers $\alpha_1 + 1, \alpha_1 + 2, \ldots, \alpha_1 + \alpha_2$ from left to right, and so on. Show that $\operatorname{std} T$ is a standard Young tableau.
   (d) Standardize a two-line array π by using the same left-to-right replacement as in the previous part on $\hat{\pi}$ and then doing so again on $\check{\pi}$. Clearly $\operatorname{std} \pi$ is a permutation in two-line form if the columns of π are lexicographically ordered. Show that in this case

$$\mathrm{RS}(\operatorname{std}(\pi)) = \operatorname{std}(\mathrm{RSK}(\pi)).$$

   (e) Use part (d) and the result (rather than the proof) of Theorem 7.6.1 to give a second proof of part (b).

(24) (a) Extend column insertion to semistandard tableaux T by having an element x bump the uppermost element in a column greater than or equal to x. Show that with this definition $c_x(T)$ is semistandard.
   (b) Give two proofs that for any semistandard Young tableau T and positive integers x, y we have

$$c_y(r_x(T)) = r_x(c_y(T)),$$

   one by mimicking the proof of Lemma 7.6.2 and one by using the standardization operator std.
   (c) Give two proofs that if $\mathrm{RSK}(\pi) = (T, U)$ with $\operatorname{sh} T = \operatorname{sh} U = \lambda$, then

$$\lambda_1^t = \text{the length of a longest decreasing subsequence of } \pi,$$

   one by using part (b) and one using the standardization operator std from the previous exercise.
   (d) Prove the identity

$$\sum_{\lambda} s_\lambda(\mathbf{x}) s_{\lambda^t}(\mathbf{y}) = \prod_{i,j \ge 1} (1 + x_i y_j).$$

   Hint: Use column insertion to define a weight-preserving bijection $M \to (T, U)$ where $M \in \mathrm{Mat}$ is a matrix with all entries zero or one, $\operatorname{sh} T = \operatorname{sh} U$, and T, U^t semistandard.

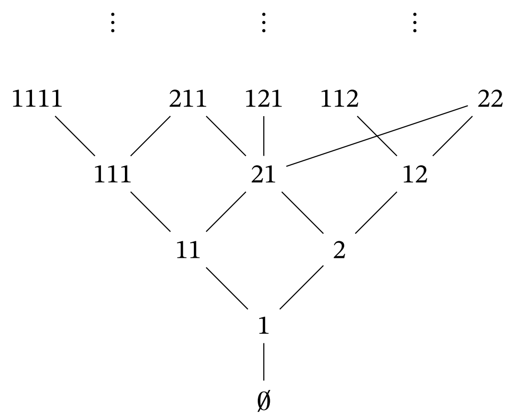


Figure 7.21. Part of the poset $\mathcal{F}$

(25) Finish the proof of Proposition 7.7.1.

(26) Show that the base case holds in Corollary 7.7.3.

(27) Prove Proposition 7.7.6.

(28) Prove Theorem 7.7.7.

(29) (a) Let $\mathcal{F}_n$ be the set of all words w of ones and twos having $\sum_i w_i = n$. Show that for $n \geq 0$ we have

$$\#\mathcal{F}_n = f_n,$$

the Fibonacci numbers defined in (1.2).

(b) Put a partial order on $\mathcal{F} = \biguplus_{n\geq 0} \mathcal{F}_n$ with covers $v \lessdot w$ whenever

F1 v can be obtained from w by removing the first one in w or

F2 w can be obtained from v by changing the first one in v to a two.

The lower ranks of $\mathcal{F}$ are shown in Figure 7.21. Show that $\mathcal{F}$ is differential.

(c) Show that $\mathcal{F}$ is a lattice. Hint: Prove that $v \wedge w$ exists by induction on $\operatorname{rk} v + \operatorname{rk} w$ and then use Exercise 12(c) in Chapter 5.

(d) Give a second proof that $\mathcal{F}$ is a lattice using Exercise 12(d) in Chapter 5.

(30) Let $r \in \mathbb{P}$. Say poset P is *r-differential* if it satisfies DP1, DP3, and the following axiom for any $x \in P$:

DP2r If x covers k elements, then it is covered by $k + r$ elements.

Prove the following statements.

(a) If P is r-differential, then $\#\operatorname{Rk}_n P$ is finite for all n. So there are well-defined D and U operators.

(b) Let P be ranked with $\#\operatorname{Rk}_n P$ finite for all n. We have

$$P \text{ is } r\text{-differential} \iff DU - UD = rI.$$

(c) In any r-differential poset we have
$$\sum_{x\in \mathrm{Rk}_n P} (f^x)^2 = r^n n!$$
where f^x is the number of saturated $\hat{0}$–x chains.

(d) If P is r-differential and Q is s-differential, then $P\times Q$ is $(r+s)$-differential. In particular, if P is differential, then P^r is r-differential.

(31) Show that the backwards implication in Proposition 7.8.4 holds. Hint: Use Theorem 3.8.4.

(32) (a) The *star* S_n is the tree with $V = \{v_1, \dots, v_n\}$ and with edges $E = \{v_1v_2, v_1v_3, \dots, v_1v_n\}$. Find expressions for $X(S_n)$ in the monomial basis and in the power sum basis with coefficients which are, up to sign, products of binomial coefficients. (Of course, every integer is a binomial coefficient since $\binom{n}{1} = n$. But any such factor should come choosing one thing from n things combinatorially.)

(b) Prove that for any graph G
$$P(G;t) = \sum_{F\subseteq E} (-1)^{\#F} t^{\ell(\lambda(F))}$$
in three ways: by using deletion-contraction, by using a sign-reversing involution, and by using $X(G)$.

(33) (a) Show that (7.42) defines an action of GL_n on $P^k(n)$.

(b) Show that if $\tilde{B} = \{\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \dots, \tilde{\mathbf{b}}_n\}$ is any basis for $\mathbb{C}[n]$, then
$$\{\tilde{\mathbf{b}}_{i_1}\tilde{\mathbf{b}}_{i_2}\cdots\tilde{\mathbf{b}}_{i_k} \mid i_1 \le i_2 \le \cdots \le i_k\}$$
is a basis for $P^k(n)$.

(34) (a) Let
$$e_k(n) = e_k(x_1, x_2, \dots, x_n)$$
and similarly for $h_k(n)$. Prove that $e_k(0) = h_k(0) = \delta_{k,0}$ and for $n \ge 1$
$$e_k(n) = e_k(n-1) + x_n e_{k-1}(n-1),$$
$$h_k(n) = h_k(n-1) + x_n h_{k-1}(n).$$

(b) Give three proofs of the first identity in Proposition 7.9.2: one by induction, one using the q-Binomial Theorem (Theorem 3.2.4), and one using Exercise 7(c) in Chapter 3.

(c) Give two more proofs of the second identity in Proposition 7.9.2: one by induction and one which uses the q-Binomial Theorem (Theorem 3.2.4).

(35) (a) Prove that
$$S[n,k] = h_{n-k}([1]_q, [2]_q, \dots, [k]_q),$$
where $S[n,k]$ is the q-Stirling number of the second kind introduced in Chapter 3, Exercise 21. Hint: Use part (a) of Exercise 34.

(b) Prove that
$$c[n,k] = e_{n-k}([1]_q, [2]_q, \dots, [n-1]_q),$$
where $c[n,k]$ is the signless q-Stirling number of the first kind introduced in Chapter 3, Exercise 22. Hint: Use part (a) of Exercise 34.

(36) Define a relation on the polynomial ring $\mathbb{R}[q]$ by letting $f(q) \le g(q)$ if, for all $i \in \mathbb{N}$, the coefficient of q^i in $f(q)$ is less than or equal to the coefficient of q^i in $g(q)$.

(a) Prove that this relation is a partial order on $\mathbb{R}[q]$, but not a total order.

(b) Define a sequence of polynomials $f_0(q), f_1(q), \ldots$ to be *q-log-concave* if

$$f_k(q)^2 \ge f_{k-1}(q) f_{k+1}(q)$$

for all $k \ge 1$. If the polynomial sequence is finite, we let $f_k(q) = 0$ for all sufficiently large k. Prove the following.

(i) For any $n \in \mathbb{N}$ the finite sequence

$$q^{\binom{0}{2}} \begin{bmatrix} n \\ 0 \end{bmatrix}, q^{\binom{1}{2}} \begin{bmatrix} n \\ 1 \end{bmatrix}, \ldots, q^{\binom{n}{2}} \begin{bmatrix} n \\ n \end{bmatrix}$$

is q-log-concave. Hint: Use the first identity in Proposition 7.9.2.

(ii) For any $n \in \mathbb{N}$ the infinite sequence

$$\begin{bmatrix} n \\ 0 \end{bmatrix}, \begin{bmatrix} n+1 \\ 1 \end{bmatrix}, \begin{bmatrix} n+2 \\ 2 \end{bmatrix}, \ldots$$

is q-log-concave. Hint: Use the second identity in Proposition 7.9.2.

(iii) For any $n \in \mathbb{N}$ the finite sequence

$$c[n,0], c[n,1], \ldots, c[n,n]$$

is q-log-concave. Hint: Use part (b) of the previous exercise.

Chapter 8

Counting with Quasisymmetric Functions

While symmetric functions are invariant under arbitrary permutations of variables, quasisymmetric functions only need to be preserved by order-preserving bijections on the variable subscripts. Quasisymmetric functions are implicit in the work of Stanley on P-partitions [**84**] but were first explicitly defined and studied by Gessel [**32**]. As we will see in this chapter, these functions also have interesting connections with chain enumeration in posets, pattern avoidance, and graph coloring.

8.1. The algebra of quasisymmetric functions, QSym

We start by defining what it means for a power series to be quasisymmetric. We will introduce two important bases for the algebra of quasisymmetric functions and discuss their relationship with symmetric functions.

As usual, $\mathbf{x} = \{x_1, x_2, x_3, \ldots\}$ will be a countably infinite variable set. A power series $f(\mathbf{x}) \in \mathbb{C}[[\mathbf{x}]]$ is *quasisymmetric* if any two monomials of the form $x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_l}^{\alpha_l}$ with $i_1 < i_2 < \cdots < i_l$ and $x_{j_1}^{\alpha_1} x_{j_2}^{\alpha_2} \cdots x_{j_l}^{\alpha_l}$ with $j_1 < j_2 < \cdots < j_l$ have the same coefficient. Note the increasing condition on the subscripts which is not present in the definition of a symmetric function. So this is more a restrictive condition; that is, every symmetric function is quasisymmetric but not conversely. For example

$$f(\mathbf{x}) = 5x_1^4x_2 + 5x_1^4x_3 + 5x_2^4x_3 + \cdots - x_1x_2^2x_3 - x_1x_2^2x_4 - x_1x_3^2x_4 - x_2x_3^2x_4 - \cdots \tag{8.1}$$

is quasisymmetric but not symmetric. An equivalent way to define $f(\mathbf{x})$ being quasisymmetric is to say that any monomial of the form $x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_l}^{\alpha_l}$ with $i_1 < i_2 < \cdots < i_l$ has the same coefficient as $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_l^{\alpha_l}$.

To set up notation, let

$$\begin{aligned}\mathrm{QSym}_n &= \mathrm{QSym}_n(\mathbf{x})\\ &= \{f(\mathbf{x}) \in \mathbb{C}[[\mathbf{x}]] \mid f \text{ is quasisymmetric and homogeneous of degree } n\}.\end{aligned}$$

Then the *algebra of quasisymmetric functions* is

$$\mathrm{QSym} = \mathrm{QSym}(\mathbf{x}) = \bigoplus_{n\geq 0} \mathrm{QSym}_n(\mathbf{x}).$$

Note that, as with symmetric functions, since QSym is a direct sum the quasisymmetric power series in it are of bounded degree. Note also that, unlike symmetric functions, it is not obvious that this is an algebra, i.e., that QSym is closed under multiplication and not just under taking linear combinations. The proof of this fact will follow from Theorem 8.3.1. As we will see, bases for QSym are indexed by integer compositions. We will be interested in two particular bases.

Given a composition $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_l]$, the associated *monomial quasisymmetric function* is

$$M_\alpha = \sum_{i_1 < i_1 < \cdots < i_l} x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_l}^{\alpha_l}.$$

So M_α can be thought of as the result of quasisymmetrizing the monomial $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_l^{\alpha_l}$. To illustrate,

$$M_{[1,3]} = x_1 x_2^3 + x_1 x_3^3 + x_2 x_3^3 + \cdots.$$

We will often drop the square brackets and commas in the subscript of M_α. This should cause no confusion with partitions because of the use of capital letters for quasisymmetric bases and lowercase ones for symmetric function bases. We will also use multiplicity notation for α where i^{m_i} denotes a string of m_i consecutive i's. Note that the quasisymmetric function in (8.1) can be written as the linear combination $f(\mathbf{x}) = 5M_{41} - 7M_{121}$. This can always be done as the M_α are a basis. This proof follows the same lines as in the demonstration that the m_λ form a basis for Sym, Theorem 7.1.1.

Theorem 8.1.1. *The M_α as α varies over all compositions form a basis for* QSym. *Consequently, for $n \geq 1$,*

$$\dim \mathrm{QSym}_n = 2^{n-1}.$$

Proof. The dimension statement follows from the basis claim and Theorem 1.7.1. To prove that the M_α are a basis, note first that they are independent since no two monomial quasisymmetric functions contain the same monomial. To show that they span, take an $f \in \mathrm{QSym}$. Consider any term in f, say $c x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_l}^{\alpha_l}$ where $i_1 < i_2 < \cdots < i_l$ and $c \in \mathbb{C}$. Then all monomials $x_{j_1}^{\alpha_1} x_{j_2}^{\alpha_2} \cdots x_{j_l}^{\alpha_l}$ such that $j_1 < j_2 < \cdots < j_l$ appear with coefficient c. So $f - cM_\alpha$ is still quasisymmetric and contains no monomials with ordered exponent sequence α. The fact that f is of bounded degree implies that repeating this process a finite number of times will yield zero. Thus f is a linear combination of the M_α which were subtracted. □

There is a nice relationship between the monomial quasisymmetric functions and their symmetric counterparts. A *rearrangement* of a partition λ is a composition α obtained by listing the parts of λ in a particular order. For example, the rearrangements of

$\lambda = (2,1,1)$ are $\alpha = [2,1,1]$, $\alpha = [1,2,1]$, and $\alpha = [1,1,2]$. The proof of the following result is easy and so is left to the reader.

Proposition 8.1.2. *For any partition λ*

$$m_\lambda = \sum_\alpha M_\alpha$$

where the sum is over all rearrangements α of λ. □

To describe the other basis for QSym which will interest us, it will be convenient to remember that there is a simple bijection ϕ between subsets $S \subseteq [n-1]$ and compositions $\alpha \models n$ defined by (1.8). So we will sometimes write M_S instead of M_α if $\phi(S) = \alpha$. Strictly speaking, M_S is not well-defined since there will be many n such that $S \subseteq [n-1]$. But the context will always make it clear which n is meant. To illustrate, if $n = 2$ and $S = \{1\}$, then $M_S = M_{[1,1]}$, whereas if $n = 3$ with the same S, then $M_S = M_{[1,2]}$. Given $S = \{s_1, \ldots, s_k\} \subseteq [n-1]$, the corresponding *fundamental quasisymmetric function* is

$$F_S = \sum_{\substack{i_1 \le i_2 \le \cdots \le i_n \\ i_s < i_{s+1} \text{ if } s \in S}} x_{i_1} x_{i_2} \cdots x_{i_n}.$$

In words, one sums over all monomials whose indices form a weakly increasing sequence with strict increases at the positions indexed by S. As an example, if $n = 4$ and $S = \{1,3\}$, then

$$(8.2) \quad F_S = \sum_{i<j\le k<l} x_i x_j x_k x_l = x_1 x_2^2 x_3 + x_1 x_2^2 x_4 + \cdots + x_1 x_2 x_3 x_4 + x_1 x_2 x_3 x_5 + \cdots.$$

We let $F_\alpha = F_S$ when $\phi(S) = \alpha$. So if $\alpha = [\alpha_1, \ldots, \alpha_l]$, then the strict inequalities in the subscripts for the monomials in F_α must occur at positions indexed by partial sums $\alpha_1 + \cdots + \alpha_i$ for each $i < l$. To describe the expansion of the F_α in terms of the M_β we will use the partial order in the composition lattice K_n described at the beginning of Section 5.1.

Proposition 8.1.3. *We have*

$$F_\alpha = \sum_{\beta \le \alpha} M_\beta$$

where $\le$ is the partial order in the composition lattice K_n.

Proof. The power series F_α is quasisymmetric since the inequalities impose by α on a sequence $i_1 \le i_2 \le \cdots \le i_n$ depend only on the positions in the sequence and not on the actual choice of the i_j. Furthermore, each monomial appearing in F_α has coefficient one. So the same is true of the expansion of F_α in the M_β basis. The only thing left to prove is that M_β appears in the expansion if and only if $\beta \le \alpha$. The monomials occurring in F_α are those which can be expressed as $x_{i_1} x_{i_2} \cdots x_{i_n}$ where $i_1 \le i_2 \le \cdots \le i_n$ and $i_j < i_{j+1}$ for each j which is a partial sum of α. Collecting together variables with the same subscripts, these are exactly the monomials which can be written as $x_{j_1}^{\beta_1} x_{j_2}^{\beta_2} \cdots x_{j_l}^{\beta_l}$ where $j_1 < j_2 < \cdots < j_l$ and $\beta \le \alpha$. This observation completes the proof. □

We can use the previous result to show that the F_α are a basis for QSym. The proof is much the same as that of Theorem 7.1.3(a) and so it is left to the reader.

Theorem 8.1.4. *The set* $\{F_\alpha \mid \alpha \models n\}$ *is a basis for* QSym_n. □

8.2. Reverse P-partitions

Fundamental quasisymmetric functions can be used to count P-partitions. In fact, they give a more refined generating function which keeps track of the parts used in the partitions in the same way that Schur functions and chromatic symmetric functions do for semistandard Young tableaux and proper colorings, respectively. This permits us to express a Schur function as a sum over standard Young tableaux and to write down a rule for the multiplication of fundamental quasisymmetric functions. But to make the partition conventions align with those for quasisymmetric functions, we will first define a slight variant of P-partitions where the inequalities are reversed.

Consider functions $f : [n] \to \mathbb{P}$ whose range is the positive integers. Say that f is *reverse compatible* with permutation $\pi \in \mathfrak{S}_n$ if

RC1 $f(\pi_1) \le f(\pi_2) \le \cdots \le f(\pi_n)$ and
RC2 $f(\pi_i) < f(\pi_{i+1})$ whenever $i \in \operatorname{Des} \pi$.

Note that there is a bijection between the functions $f : [n] \to \mathbb{P}$ which are reverse compatible with π and the functions $g : [n] \to \mathbb{N}$ which are compatible with π's reverse complement

$$\pi' = \pi'_1\pi'_2 \ldots \pi'_n = n+1-\pi_n,\ n+1-\pi_{n-1},\ \ldots,\ n+1-\pi_1 \tag{8.3}$$

where $g(\pi'_i) = f(\pi_{n+1-i}) - 1$ for all $i \in [n]$. So studying reverse compatibility and compatibility is essentially the same. But, as already mentioned, RC1 and RC2 will play more nicely with fundamental quasisymmetric functions. Let

$$\mathcal{RC}(\pi) = \{f : [n] \to \mathbb{P} \mid f \text{ is reverse compatible with } \pi\}.$$

The proof of the next result is similar to that of Lemma 7.4.1 and so it is omitted.

Lemma 8.2.1. *Every* $f : [n] \to \mathbb{P}$ *is reverse compatible with a unique* $\pi \in \mathfrak{S}_n$. *Thus*

$$\{f \mid f : [n] \to \mathbb{P}\} = \biguplus_{\pi \in \mathfrak{S}_n} \mathcal{RC}(\pi). \qquad \square$$

To make a connection with quasisymmetric functions, associate with any $f : [n] \to \mathbb{P}$ the monomial

$$\mathbf{x}^f = x_{f(1)}x_{f(2)} \cdots x_{f(n)}.$$

Recall that the fundamental quasisymmetric functions can be indexed by subsets $S \subseteq [n-1]$. And given a permutation $\pi \in \mathfrak{S}_n$, we have $\operatorname{Des}\pi \subseteq [n-1]$.

Lemma 8.2.2. *For any* $\pi \in \mathfrak{S}_n$ *we have*

$$\sum_{f \in \mathcal{RC}(\pi)} \mathbf{x}^f = F_{\operatorname{Des}\pi}. \tag{8.4}$$

Proof. Since the monomials on both sides of (8.4) all occur with coefficient one, it suffices to find a bijection between the subscripts which can appear for monomials on the two sides of the equation. By definition, the subscripts of monomials in $F_{\operatorname{Des}\pi}$ are precisely those of the form $i_1 \le i_2 \le \cdots \le i_n$ with $i_k < i_{k+1}$ if $k \in \operatorname{Des}\pi$. Associate with this subscript the function $f : [n] \to \mathbb{P}$ where $f(\pi_j) = i_j$ for $j \in [n]$. We claim that this is well-defined in that $f \in \mathcal{RC}(\pi)$. Indeed, the fact that the i_j are weakly increasing is condition RC1 and the placement of the strict inequalities agrees with RC2. It is now an easy matter to construct a well-defined inverse, completing the proof. □

We now bring the P-partition definitions through the looking glass into the land of reverse. Given a poset P on $[n]$, a *reverse P-partition* is a function $f : P \to \mathbb{P}$ satisfying

RPP1 $i \trianglelefteq j$ implies $f(i) \le f(j)$ and
RPP2 $i \trianglelefteq j$ and $i > j$ implies $f(i) < f(j)$.

We let

$$\operatorname{RPar} P = \{f : P \to \mathbb{P} \mid f \text{ is a reverse } P\text{-partition}\}.$$

The result below follows from Lemma 8.2.1 in much the same way that Lemma 7.4.3 was derived from Lemma 7.4.1.

Lemma 8.2.3 (Fundamental Lemma of Reverse P-Partitions)**.** *Let P be a poset on $[n]$. Then $f \in \operatorname{RPar} P$ if and only if $f \in \mathcal{RC}(\pi)$ for some $\pi \in \mathcal{L}(P)$. Thus*

$$\operatorname{RPar} P = \biguplus_{\pi \in \mathcal{L}(P)} \mathcal{RC}(\pi). \qquad \square$$

To derive a generating function identity we define, for P a poset on $[n]$, the generating function

$$\operatorname{rpar} P = \operatorname{rpar}(P; \mathbf{x}) = \sum_{f \in \operatorname{RPar} P} \mathbf{x}^f.$$

Now using the previous lemma and (8.4) we obtain the following.

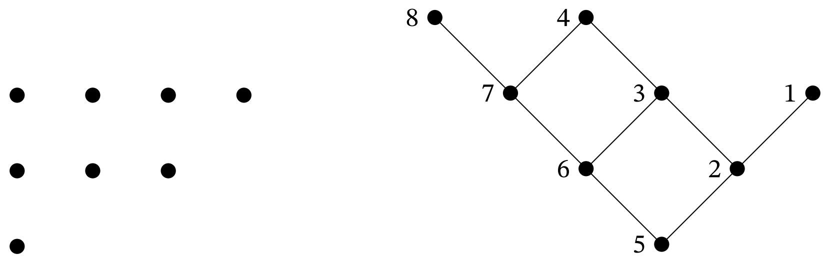
Theorem 8.2.4. *For any poset P on $[n]$ we have*

$$\operatorname{rpar} P = \sum_{\pi \in \mathcal{L}(P)} F_{\operatorname{Des}\pi}. \qquad \square$$

Since every symmetric function is quasisymmetric, one can ask what the expansion of a symmetric function is in one of the bases for QSym. We will now answer this question for the expansion of the Schur functions in terms of the fundamental quasisymmetrics. To do so we need a notion of descent for standard Young tableaux. If Q is an SYT, then let

$$\operatorname{Des} Q = \{k \mid k+1 \text{ is in a row below } k \text{ in } Q\}.$$

For example, the SYT in Figure 7.1 have descent sets $\{2,4\}$, $\{2,3\}$, $\{1,3,4\}$, $\{1,3\}$, and $\{1,2,4\}$, respectively.



λ $\qquad$ P_λ

$$Q = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 4 & 8 \\ \hline 3 & 5 & 7 \\ \cline{1-3} 6 \\ \cline{1-1}\end{array} \qquad \longleftrightarrow \qquad \pi = 56273148$$

Figure 8.1. A Young diagram and the associated labeling of P_λ, along with an SYT Q and the associated linear extension

Theorem 8.2.5. *For any partition $\lambda \vdash n$ we have*

$$s_\lambda = \sum_{Q \in \mathrm{SYT}(\lambda)} F_{\mathrm{Des}\, Q}.$$

Proof. Recall from the end of Section 7.4 that associated with λ is a poset P_λ. We will turn P_λ into a poset on $[n]$ in such a way that the reverse P_λ-partitions are exactly the semistandard Young tableaux of shape λ. To this end, label the vertices of P_λ corresponding to the last row of $\lambda = (\lambda_1, \ldots, \lambda_l)$ from 1 to λ_l so that the labels increase with the partial order. Now similarly label the vertices in the penultimate row with $\lambda_l + 1$ to $\lambda_l + \lambda_{l-1}$, and continue doing so until all of P_λ is labeled. An example is in Figure 8.1. Using the usual coordinates on λ to also refer to the corresponding nodes of P_λ, we will let $l(i,j)$ be the label of node (i,j) in P_λ. In Figure 8.1, $l(2,3) = 4$.

We claim that rotation by 135° is a bijection $\mathrm{SSYT}(\lambda) \to \mathrm{RPar}\, P_\lambda$. As with P-partitions, it suffices to show that RPP1 and RPP2 hold on covers. Now $T \in \mathrm{SSYT}(\lambda)$ if and only if $T_{i,j} \le T_{i,j+1}$ and $T_{i,j} < T_{i+1,j}$ for all cells (i,j). Since $l(i,j) < l(i,j+1)$ and $l(i,j) > l(i+1,j)$, these two conditions on T become exactly RPP1 and RPP2 for the associated reverse P-partition, demonstrating the claim. And from this it follows that

$$s_\lambda = \mathrm{rpar}\, P_\lambda.$$

The SYT Q of shape λ are in bijection with the $\pi \in \mathcal{L}(P_\lambda)$ by letting $\pi_k = l(i,j)$ where $Q_{i,j} = k$. Furthermore, if $Q \longleftrightarrow \pi$ under this bijection, then $\mathrm{Des}\, Q = \mathrm{Des}\, \pi$. To see this, suppose first that $k \in \mathrm{Des}\, Q$ where k is in cell c. Then $k+1$ is in a cell c' in a lower row of Q. It follows that $\pi_k = l(c) > l(c') = \pi_{k+1}$ because of the way the elements of P_λ were labeled. Thus $k \in \mathrm{Des}\, \pi$. Showing the reverse inclusion, $\mathrm{Des}\, \pi \subseteq \mathrm{Des}\, Q$, is similar and so is left as an exercise. Thus if $Q \longleftrightarrow \pi$, then

$$F_{\mathrm{Des}\,Q} = F_{\mathrm{Des}\,\pi}.$$

Combining the two previous displayed equations with Theorem 8.2.4 gives

$$s_\lambda = \mathrm{rpar}\,P_\lambda = \sum_{\pi \in \mathcal{L}(P_\lambda)} F_{\mathrm{Des}\,\pi} = \sum_{Q \in \mathrm{SYT}(\lambda)} F_{\mathrm{Des}\,Q}$$

which is the desired conclusion. □

We will now show that QSym is an algebra by showing that the product of two fundamental quasisymmetric functions is a linear combination of quasisymmetrics. This will give us a second application of Theorem 8.2.4. First note that to apply RPP1 and RPP2 it is not necessary that P be a poset on the set $[n]$: any subset $S \subset \mathbb{P}$ would do in place of $[n]$. Think of such a poset as a pair (P, ω) consisting of an underlying poset P and a labeling $\omega : P \to S$. Two labelings $\omega : P \to S$ and $\omega' : P \to S'$ satisfying $\omega(x) < \omega(y)$ if and only if $\omega'(x) < \omega'(y)$ are said to have the *same relative order*. In this case, the monomials $\mathbf{x}^f$ for reverse (P, ω)-partitions are the same as the monomials for reverse (P, ω')-partitions. Thus $\mathrm{rpar}(P, \omega) = \mathrm{rpar}(P, \omega')$. Given disjoint posets P on a label set S and Q on a label set T where S and T are disjoint, one gives $P \uplus Q$ the label set $S \uplus T$ in the obvious way. The next result follows easily from the definitions.

Proposition 8.2.6. *If P, Q are posets on disjoint subsets of $\mathbb{P}$, then*

$$\mathrm{rpar}(P \uplus Q) = (\mathrm{rpar}\,P)(\mathrm{rpar}\,Q).$$ □

Our final tool involves shuffling sequences of integers. If $\sigma \in P(U)$ and $\tau \in P(V)$ where U, V are disjoint, then the associated set of *shuffles* is

$$\sigma \sqcup\!\sqcup\, \tau = \{\pi \in P(U \uplus V) \mid \sigma \text{ and } \tau \text{ are subwords of } \pi\}.$$

To illustrate, if $\sigma = 14$ and $\tau = \mathbf{52}$, then

$$\sigma \sqcup\!\sqcup\, \tau = \{14\mathbf{52},\ 1\mathbf{5}4\mathbf{2},\ 1\mathbf{52}4,\ \mathbf{5}14\mathbf{2},\ \mathbf{5}1\mathbf{2}4,\ \mathbf{52}14\}.$$

Finally, we note that the definition of a descent set can be extended to any $\pi \in P(S)$ and that Theorem 8.2.4 continues to hold.

Theorem 8.2.7. *If $\sigma \in P(U)$ and $\tau \in P(V)$ where U, V are disjoint subsets of $\mathbb{P}$, then*

$$F_{\mathrm{Des}\,\sigma} F_{\mathrm{Des}\,\tau} = \sum_{\pi \in \sigma \sqcup\!\sqcup \tau} F_{\mathrm{Des}\,\pi}.$$

Proof. Let P be the poset on U which is a chain labeled from bottom to top by the elements of σ read left to right. So $\mathcal{L}(P) = \{\sigma\}$. Similarly define Q so that $\mathcal{L}(Q) = \{\tau\}$. It follows that $\mathcal{L}(P \uplus Q) = \sigma \sqcup\!\sqcup\, \tau$. Now applying Theorem 8.2.4 and Proposition 8.2.6

$$F_{\mathrm{Des}\,\sigma} F_{\mathrm{Des}\,\tau} = (\mathrm{rpar}\,P)(\mathrm{rpar}\,Q) = \mathrm{rpar}(P \uplus Q) = \sum_{\pi \in \sigma \sqcup\!\sqcup \tau} F_{\mathrm{Des}\,\pi}$$

as desired. □

The previous theorem is remarkable for it implies that the multiset $\{\{\operatorname{Des}\pi \mid \pi \in \sigma \sqcup\!\sqcup \tau\}\}$ depends only on the lengths of σ and τ and their descent sets. A function on permutations with this property is called *shuffle compatible* and this concept has been studied by Gessel and Zhuang [**33**], Grinberg [**36**], as well as Baker-Jarvis and Sagan [**4**].

8.3. Chain enumeration in posets

As we have just seen, the fundamental quasisymmetric functions give us information about reverse P-partitions. It turns out that the monomial quasisymmetric functions can be used to model chains in posets, as was shown by Ehrenborg [**24**]. In particular, we will see how multiplication of the M_α corresponds to taking the product of posets. A connection will also be made with the binomial posets studied in Section 5.9.

We begin by deriving a formula for the product of two monomial quasisymmetric functions. Recall that α is a weak composition if it can include parts equal to zero. An *expansion* of a composition α is a weak composition $\bar\alpha$ such that removing the zeros from $\bar\alpha$ one obtains α. For example, one expansion of $\alpha = [1,4,1]$ is $\bar\alpha = [0,0,1,4,0,1,0]$. If α,β,γ are compositions, then we say γ is a *shuffle sum* of the other two compositions if there are expansions $\bar\alpha$ and $\bar\beta$ of α and β, respectively, which have length $\ell(\gamma)$ such that $\gamma = \bar\alpha + \bar\beta$. Here, addition is componentwise. To illustrate, if $\alpha = [1,2]$ and $\beta = [1]$, then there are two ways of writing $\gamma = [1,1,2]$ as a shuffle sum of α and β, namely $[1,0,2] + [0,1,0]$ and $[0,1,2] + [1,0,0]$. The reader can verify that

$$M_{[1,2]}M_{[1]} = M_{[2,2]} + M_{[1,3]} + 2M_{[1,1,2]} + M_{[1,2,1]}$$

where the coefficient 2 of $M_{[1,1,2]}$ corresponds to these two shuffle sums. The general result is as follows.

Theorem 8.3.1. *We have*

$$M_\alpha M_\beta = \sum_\gamma c^\gamma_{\alpha,\beta} M_\gamma$$

where $c^\gamma_{\alpha,\beta}$ is the number of ways of writing γ as a shuffle sum of α and β.

Proof. Since $M_\alpha M_\beta$ is quasisymmetric, it suffices to show that for any $\gamma = [\gamma_1,\dots,\gamma_t]$ we have $x_1^{\gamma_1}\cdots x_t^{\gamma_t}$ occurring in the product with coefficient $c^\gamma_{\alpha,\beta}$. For any, possibly weak, composition $\bar\alpha = [\alpha_1,\dots,\alpha_t]$ we let $\mathbf{x}^{\bar\alpha} = x_1^{\bar\alpha_1}\cdots x_t^{\bar\alpha_t}$. Let $\ell(\alpha) = r$, $\ell(\beta) = s$, and $\ell(\gamma) = t$.

Given any shuffle sum $\gamma = \bar\alpha + \bar\beta$, we clearly have $\mathbf{x}^\gamma = \mathbf{x}^{\bar\alpha}\mathbf{x}^{\bar\beta}$. Conversely, suppose

$$\mathbf{x}^\gamma = (x_{i_1}^{\alpha_1}\cdots x_{i_r}^{\alpha_r})(x_{j_1}^{\beta_1}\cdots x_{j_s}^{\beta_s}) = x_{k_1}^{\gamma_1}\cdots x_{k_t}^{\gamma_t}$$

where $i_1 < \cdots < i_r$ and $j_1 < \cdots < j_s$. Define an expansion $\bar\alpha$ of α of length t by putting α_p in position i_p of $\bar\alpha$ for $p \in [r]$ and placing zeros everywhere else. Similarly define $\bar\beta$. The previous displayed equation implies that $\gamma = \bar\alpha + \bar\beta$. It is not hard to see that the two maps just described are inverses. So we have bijection between shuffle sum decompositions $\gamma = \bar\alpha + \bar\beta$ and ways to write $\mathbf{x}^\gamma$ as a product of a monomial from M_α with a monomial from M_β. The theorem follows. □

To make the connection with chain enumeration, let P be a finite, ranked, poset with a $\hat{1}$. Recall that in this situation, any interval $[x, y] \subseteq P$ can be considered as a poset of rank

$$\operatorname{rk}[x, y] = \operatorname{rk}_P y - \operatorname{rk}_P x.$$

Associate with any chain

$$C : \hat{0} = x_0 < x_1 < \cdots < x_k = \hat{1} \tag{8.5}$$

the composition

$$\alpha(C) = [\operatorname{rk}[x_0, x_1],\ \operatorname{rk}[x_1, x_2],\ \ldots,\ \operatorname{rk}[x_{k-1}, x_k]]. \tag{8.6}$$

For example, in B_6 the chain $C : \emptyset < \{2, 5\} < \{1, 2, 4, 5, 6\} < [6]$ has $\alpha(C) = [2, 3, 1]$. Also associate with P the generating function

$$M(P) = \sum_C M_{\alpha(C)}$$

where the sum is over all chains C of the form (8.5). This generating function respects products of posets.

Theorem 8.3.2. *Let P, Q be finite, ranked posets each having a $\hat{1}$. Then*

$$M(P \times Q) = M(P)M(Q). \tag{8.7}$$

Proof. By definition, the coefficient of M_γ on the left side of (8.7) is the number of $\hat{0}$–$\hat{1}$ chains C in $P \times Q$ with $\alpha(C) = \gamma$. And by Theorem 8.3.1, the coefficient of M_γ on the right is $\sum_{A,B} c^{\gamma}_{\alpha(A),\alpha(B)}$ where the sum is over all $\hat{0}$–$\hat{1}$ chains $A \subseteq P$ and $B \subseteq Q$. So it suffices to find a bijection between $\hat{0}$–$\hat{1}$ chains C in $P \times Q$ and ways of writing $\alpha(C)$ as a shuffle sum of $\alpha(A)$ and $\alpha(B)$ for $\hat{0}$–$\hat{1}$ chains A, B in P, Q, respectively.

Suppose first that we are given

$$C : \hat{0} = (x_0, y_0) < (x_1, y_1) < \cdots < (x_k, y_k) = \hat{1}. \tag{8.8}$$

The *projection of C onto P* is the multichain

$$\overline{A} : \hat{0} = x_0 \le x_1 \le \cdots \le x_k = \hat{1}.$$

This multichain has *underlying chain A* obtained by replacing each maximal string of copies of x in $\overline{A}$ with just x itself. Note that the definition in (8.6) can be applied equally well to multichains, except now the result will be a weak composition. Furthermore $\alpha(\overline{A})$ is an expansion of $\alpha(A)$. Similarly define the projection $\overline{B}$ of C onto Q with its underlying chain B. From Exercise 8(c) in Chapter 5 we know that $\operatorname{rk}(x, y) = \operatorname{rk}(x) + \operatorname{rk}(y)$ for all $(x, y) \in P \times Q$. It follows that $\alpha(C) = \alpha(\overline{A}) + \alpha(\overline{B})$. This completes the definition of the bijection in one direction.

Now suppose we are given $\hat{0}$–$\hat{1}$ chains A in P and B in Q such that there are expansions $\bar{\alpha}$ and $\bar{\beta}$ of $\alpha(A)$ and $\alpha(B)$ satisfying $\gamma = \bar{\alpha} + \bar{\beta}$ for some M_γ appearing on the left in (8.7). Then there is a unique multichain $\overline{A}$ whose underlying chain is A and whose composition is $\alpha(\overline{A}) = \bar{\alpha}$: each $x_i \in A$ is replaced by $m_i + 1$ copies of itself where m_i is the number of zeros in $\bar{\alpha}$ between the elements α_i and α_{i+1} of $\alpha(A)$. Similarly we obtain a multichain $\overline{B}$ from B and $\bar{\beta}$. Finally, we construct C as in (8.8) where the first components are $\overline{A}$ and the second $\overline{B}$. Verifying that $\alpha(C) = \alpha(\overline{A}) + \alpha(\overline{B})$ and that this is the inverse of the map defined in the previous paragraph is left to the reader. □

To end this section, suppose that $I = [x, z]$ is an n-interval as defined in axiom BP2 of Section 5.9 for a binomial poset P. The generating function $M(I)$ has a very nice form in terms of the factorial function $F(n)$ for P. Given a composition $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_k] \models n$, define

$$\binom{n}{\alpha}_P = \frac{F(n)}{F(\alpha_1)F(\alpha_2)\cdots F(\alpha_k)}.$$

Note that when $P = B_\infty$, then $\binom{n}{\alpha}_P$ is just a multinomial coefficient as defined in (1.14).

Theorem 8.3.3. *Let P be a binomial poset and let I be an n-interval in P. Then*

$$M(I) = \sum_{\alpha \models n} \binom{n}{\alpha}_P M_\alpha.$$

Proof. Since P is binomial, there is no loss of generality in assuming that $I = [\hat{0}, z]$ for some z. By definition of $M(I)$, we need to show that the number of chains (8.5) where $\hat{1} = z$ and $\alpha(C) = \alpha$ is given by $\binom{n}{\alpha}_P$. By Lemma 5.9.2,

$$\#\text{ of } x_1 \text{ of rank } \alpha_1 \text{ in } I = \frac{F(n)}{F(\alpha_1)F(n-\alpha_1)}.$$

Similarly

$$\#\text{ of } x_2 \text{ of rank } \alpha_2 \text{ in } [x_1, z] = \frac{F(n-\alpha_1)}{F(\alpha_2)F(n-\alpha_1-\alpha_2)}.$$

So the number of ways to pick x_1 and x_2 is

$$\frac{F(n)}{F(\alpha_1)F(n-\alpha_1)} \cdot \frac{F(n-\alpha_1)}{F(\alpha_2)F(n-\alpha_1-\alpha_2)} = \frac{F(n)}{F(\alpha_1)F(\alpha_2)F(n-\alpha_1-\alpha_2)}.$$

Continuing in this way, we see that the total count is $\binom{n}{\alpha}_P$ as desired. □

8.4. Pattern avoidance and quasisymmetric functions

Given a formal power series $f(\mathbf{x})$ which is a priori a quasisymmetric function, it can be intriguing to see if it is actually symmetric. And, in that case, one could further ask whether $f(\mathbf{x})$ has an expansion with nonnegative coefficients in one of the standard bases for Sym. In this section we are going to be concerned with *Schur nonnegativity*, that is, seeing if $f(\mathbf{x}) = \sum_\lambda c_\lambda s_\lambda$ where $c_\lambda \geq 0$ for all λ. Associated with any set $\Pi \subseteq \mathfrak{S}_n$ of permutations, we can define the quasisymmetric function

$$F_\Pi = \sum_{\pi \in \Pi} F_{\operatorname{Des} \pi} \in \operatorname{QSym}_n. \tag{8.9}$$

It turns out that by taking S to be the set of permutations which satisfy certain avoidance conditions, one gets interesting results. These ideas were first investigated by Hamaker, Pawlowski, and Sagan [**42**], with further results being obtained by Bloom and Sagan [**17**].

If Π is any set of permutations, then we let

$$\mathrm{Av}_n(\Pi) = \{\sigma \in \mathfrak{S}_n \mid \sigma \text{ avoids every } \pi \in \Pi\}.$$

So $\mathrm{Av}_n(\Pi) = \bigcap_{\pi\in\Pi} \mathrm{Av}_n(\pi)$. To illustrate, call a permutation $\pi \in \mathfrak{S}_n$ *reverse layered* if it is of the form

$$\pi = m, m+1, \ldots, n, l, l+1, \ldots, m-1, k, k+1, \ldots \tag{8.10}$$

for certain $n \geq m > l > k > \cdots > 0$. This term is used since the reversal π^r is of a form usually called *layered*. The increasing subsequences $m, m+1, \ldots, n$ and so forth are called its *layers* and their lengths are the *layer lengths*. For example $\pi = 789561234$ is reverse layered with layers 789, 56, 1234 and corresponding layer lengths $3, 2, 4$.

Lemma 8.4.1. *We have*

$$\mathrm{Av}_n(132, 213) = \{\pi \in \mathfrak{S}_n \mid \pi \text{ is reverse layered}\}.$$

Proof. Note that the reverse layered permutations π are exactly the ones such that, for all $a, c \in [n]$, if $a < c$ and a is before c in π, then every element of $[a, c]$ comes between a, c in π. Now π contains 132 if and only if π contains a subsequence acb with $b \in [a, c]$ coming after c. Similarly π containing 213 is equivalent to there being a $b \in [a, c]$ appearing before a. So π avoids both patterns precisely when π is reverse layered. □

To bring in quasisymmetric functions define, for any set Π of permutations, the *pattern quasisymmetric function*

$$Q_n(\Pi) = \sum_{\sigma\in\mathrm{Av}_n(\Pi)} F_{\mathrm{Des}\,\sigma}.$$

Note that $Q_n(\Pi)$ is a sum of fundamental quasisymmetric functions for permutations avoiding Π, while F_Π is a sum over the elements of Π itself. There are times when $Q_n(\Pi)$ being symmetric implies that the same is true for $Q_n(\Pi')$ for certain other Π', similar to what happens with Wilf equivalence. Consider the dihedral group D of the square as defined in (1.11). The following lemma is easy to prove and so its demonstration is left as an exercise.

Lemma 8.4.2. *For any $f \in D$, any set of permutations Π, and any $n \geq 0$ we have*

$$f(\mathrm{Av}_n(\Pi)) = \mathrm{Av}_n(f(\Pi)).$$ □

Recall that if π is a permutation, then its complement and reversal (see Exercise 37 of Chapter 1) are denoted π^c and π^r, respectively.

Proposition 8.4.3. *Suppose that $Q_n(\Pi)$ is symmetric and that it has Schur expansion*

$$Q_n(\Pi) = \sum_\lambda c_\lambda s_\lambda.$$

(a) *We have $Q_n(\Pi^c)$ is symmetric and*

$$Q_n(\Pi^c) = \sum_\lambda c_\lambda s_{\lambda^t}.$$

(b) *We have* $Q_n(\Pi^r)$ *is symmetric and*

$$Q_n(\Pi^r) = \sum_\lambda c_\lambda s_{\lambda^t}.$$

(c) *We have* $Q_n(\Pi^{cr})$ *is symmetric and*

$$Q_n(\Pi^{cr}) = \sum_\lambda c_\lambda s_\lambda.$$

Proof. (a) Applying the definition of $Q_n(\Pi)$ and then Theorem 8.2.5 to the hypothesis we have

$$\sum_{\sigma\in\mathfrak{S}_n(\Pi)} F_{\operatorname{Des}\sigma} = Q_n(\Pi) = \sum_\lambda c_\lambda s_\lambda = \sum_\lambda c_\lambda \sum_{P\in\operatorname{SYT}(\lambda)} F_{\operatorname{Des}P}. \tag{8.11}$$

In passing from σ to σ^c, every descent is changed into an ascent and vice versa. So we have

$$\operatorname{Des}\sigma^c = [n-1] - \operatorname{Des}\sigma. \tag{8.12}$$

Also note that in any SYT P with k in cell (i,j) and $k+1$ in cell (i',j'), either we have $i<i'$ and $j\ge j'$ which implies $k\in\operatorname{Des}P$, or we have $i\ge i'$ and $j<j'$ which implies $k\notin\operatorname{Des}P$. It follows that

$$\operatorname{Des}P^t = [n-1] - \operatorname{Des}P. \tag{8.13}$$

Using Lemma 8.4.2, then equations (8.12), (8.11), and (8.13), and then finally Theorem 8.2.5 in that order yields

$$Q_n(\Pi^c) = \sum_{\sigma\in\mathfrak{S}_n(\Pi)} F_{[n-1]-\operatorname{Des}\sigma} = \sum_\lambda c_\lambda \sum_{P\in\operatorname{SYT}(\lambda)} F_{[n-1]-\operatorname{Des}P} = \sum_\lambda c_\lambda s_{\lambda^t}.$$

(b) The proof is similar to part (a) except that one uses Theorem 7.6.3 in place of equation 8.13. The details are left to the reader.

(c) This is implied by the first two parts and Lemma 8.4.2. □

We will now use our results to prove that two particular $\Pi\subseteq\mathfrak{S}_3$ have $Q_n(\Pi)$ symmetric for all n and determine its Schur expansion. For a complete list of all such Π, as well as examples which are not subsets of $\mathfrak{S}_3$, see [**42**]. A partition λ is a *hook* if $\lambda=(a,1^b)$ for $a\ge 1$ and $b\ge 0$. Let

$$\mathcal{H}_n = \{\lambda\vdash n \mid \lambda \text{ is a hook}\}.$$

Theorem 8.4.4. *We have*

$$Q_n(132,213) = Q_n(231,312) = \sum_{\lambda\in\mathcal{H}_n} s_\lambda.$$

Proof. Since $\{231,312\}=\{132,213\}^c$ it suffices, by the previous proposition, to prove that $Q_n(132,213)$ has the given Schur expansion. First we claim that

$$Q_n(132,213) = \sum_{S\subseteq[n-1]} F_S.$$

To prove the claim, it suffices to show that there is a bijection between subsets $S \subseteq [n-1]$ and permutations $\sigma \in \mathrm{Av}_n(132, 213)$ such that $\operatorname{Des}\sigma = S$. Recall from Lemma 8.4.1 that these σ are exactly the reverse layered permutations. If $\alpha = [\alpha_1, \ldots, \alpha_k]$ is the composition of layer lengths of σ, then $\operatorname{Des}\sigma = S$ where $S = \phi^{-1}(\alpha)$ and ϕ is the bijection (1.8). So it suffices to prove that for any $\alpha \models n$ there is a unique $\sigma \in \mathrm{Av}_n(132, 213)$ having α as its composition of layer lengths. But this is clear since we must put the α_1 largest elements of $[n]$ in the first layer, the next α_2 largest elements in the second layer, and so on.

From the previous paragraph, we will be done if we can show that

$$\sum_{\lambda \in \mathcal{H}_n} s_\lambda = \sum_{S \subseteq [n-1]} F_S.$$

Because of Theorem 8.2.5, it suffices to show that there is a bijection between subsets $S \subseteq [n-1]$ and hook tableaux H such that $\operatorname{Des} H = S$. But given any $S' \subseteq [2, n]$, there is a unique hook tableau H whose first column is $S' \cup \{1\}$. And $\operatorname{Des} H = S' - 1$, the set obtained from S' by subtracting one from each entry. So the bijection $S' \mapsto S = S' - 1$ completes the proof. ☐

8.5. The chromatic quasisymmetric function

Chromatic quasisymmetric functions were introduced by Shareshian and Wachs [**82**] in part to study the $(\mathbf{3}+\mathbf{1})$-Free Conjecture of Stanley and Stembridge [**93**]. These quasisymmetric functions refine the chromatic symmetric functions from Section 7.8 and have many interesting properties, including a connection with Hessenberg varieties from algebraic geometry. Here, we consider what happens when such a function is symmetric as well as making a connection with reverse P-partitions.

Throughout this section, $G = (V, E)$ will be a graph with $V = [n]$. We let

$$\mathcal{PC}(G) = \{c : V \to \mathbb{P} \mid c \text{ is a proper coloring of } G\}.$$

The *set of ascents* of $c \in \mathcal{PC}(G)$ is

$$\operatorname{Asc} c = \{ij \in E \mid i < j \text{ and } c(i) < c(j)\}$$

with corresponding *ascent number* $\operatorname{asc} c = \#\operatorname{Asc} c$. Figure 8.2 displays three proper colorings of a path with edges 13 and 32. They have $\operatorname{Asc} c_1 = \emptyset$, $\operatorname{Asc} c_2 = \{13\}$, and $\operatorname{Asc} c_3 = \{13, 23\}$ so that $\operatorname{asc} c_1 = 0$, $\operatorname{asc} c_2 = 1$, and $\operatorname{asc} c_3 = 2$, respectively. Given

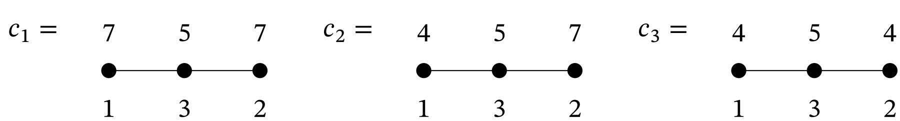


Figure 8.2. Three proper colorings of a graph with $V = [3]$

variable set $\mathbf{x} = \{x_1, x_2, \dots\}$ as well as another parameter q, define the *chromatic quasisymmetric function* of G to be

$$X(G; \mathbf{x}, q) = \sum_{c \in \mathcal{PC}(G)} \mathbf{x}^c q^{\operatorname{asc} c}$$

where $\mathbf{x}^c$ is defined by (7.36). To illustrate with the graph in Figure 8.2, if we consider the colorings c such that $c(1) = c(2) = i$ and $c(3) = j > i$, then each such map contributes $x_i^2 x_j q^2$ to the sum for a total of $M_{21}q^2$. If the inequality between i and j is reversed, then there are no ascents and the contribution is M_{12}. Similar considerations apply to the six ways three distinct colors could be assigned to the vertex set, giving a total of

$$X(G; \mathbf{x}, q) = (M_{12} + 2M_{1^3}) + (2M_{1^3})q + (M_{21} + 2M_{1^3})q^2.$$

Because of the similarity to our notation $X(G; \mathbf{x})$ for the chromatic symmetric function, we will always include the q when referring to its quasisymmetric cousin. Also, it will be convenient to define

$$X_k(G; \mathbf{x}) = \sum_{\substack{c \in \mathcal{PC}(G) \\ \operatorname{asc} c = k}} \mathbf{x}^c$$

so that

$$X(G; \mathbf{x}, q) = \sum_{k \geq 0} X_k(G; \mathbf{x}) q^k.$$

We first show that the chromatic quasisymmetric function lives up to its name and is a refinement of the chromatic symmetric function.

Proposition 8.5.1. *Let G be a graph with $V = [n]$. We have:*

(a) $X_k(G; \mathbf{x}) \in \mathrm{QSym}_n$ *for all* $k \geq 0$.

(b) $X(G; \mathbf{x}, q)$ *has degree* $\#E$ *as a polynomial in* q.

(c) $X(G; \mathbf{x}, 1) = X(G; \mathbf{x})$.

Proof. (a) We have $X_k(G, \mathbf{x})$ is homogeneous of degree n since in $\mathbf{x}^c$ there is a factor x_i for each vertex $i \in V$. To show it is quasisymmetric consider a monomial of $X_k(G, \mathbf{x})$, say $\mathbf{x}^c = x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_l}^{\alpha_l}$ with $i_1 < i_2 < \cdots < i_l$ which arose from a proper coloring $c : V \to \{i_1, i_2, \dots, i_l\}$. Take any other set of positive integers $j_1 < j_2 < \cdots < j_l$ and consider the coloring $c' = f \circ c$ where $f(i_m) = j_m$ for all m. Then c' is proper since f is a bijection and $\operatorname{asc} c' = \operatorname{asc} c = k$ since f is an increasing function. It follows that $X_k(G; \mathbf{x})$ contains a corresponding monomial $\mathbf{x}^{c'} = x_{i_j}^{\alpha_1} x_{i_j}^{\alpha_2} \cdots x_{i_j}^{\alpha_l}$ with $j_1 < j_2 < \cdots < j_l$. Thus $X_k(G; \mathbf{x})$ is quasisymmetric.

(b) Since $\operatorname{asc} c$ counts a subset of E for any coloring c, we have that the degree of $X(G; \mathbf{x}, q)$ is no greater than $\#E$. And the coloring $c(i) = i$ for all $i \in V$ has $\operatorname{asc} c = \#E$, so that is the degree.

(c) This follows trivially from the definitions. $\square$

To explore what happens if $X(G; \mathbf{x}, q)$ is in fact a symmetric function in $\mathbf{x}$, we need to discuss reversal and palindromicity of sequences. Given a sequence $a : a_0, a_1, \dots, a_n$, its *reversal* is then $a^r : a_n, a_{n-1}, \dots, a_0$. We say that a is *palindromic with center* $n/2$

if $a = a^r$. For example, the sequence $0, 7, 2, 2, 7, 0$ is palindromic with center $5/2$. We also say that the generating function $f(q) = \sum_{i=0}^{n} a_i q^i$ has one of these properties if the corresponding sequence does. So, in our example, $7q + 2q^2 + 2q^3 + 7q^4$ is palindromic with center $5/2$. Note that if $f(q)$ is palindromic with center $n/2$, then n need not be the degree of $f(q)$ because of possible initial zeros. Also, since the a_i may contain other variables, we sometimes say that $f(q)$ is palindromic in q to be specific. There is a simple algebraic test for being a palindrome. We leave its proof to the reader.

Lemma 8.5.2. *Let $a : a_0, a_1, \ldots, a_n$ be a sequence with generating function $f(q)$.*

(a) *The generating function for a^r is $q^n f(1/q)$.*

(b) *The given sequence is a palindrome with center $n/2$ if and only if*

$$f(q) = q^n f(1/q). \qquad \square$$

Define an involution ρ on the monomial quasisymmetric functions by letting

$$\rho(M_\alpha) = M_{\alpha^r}.$$

Extend ρ by linearity to $\mathrm{QSym}[q]$, the polynomials in q whose coefficients are quasisymmetric functions, where powers of q are treated as scalars. We will see that this involution reverses $X(G; \mathbf{x}, q)$ as a polynomial in q. To express the resulting power series, define the *descent set* and *descent number* of a coloring c to be

$$\operatorname{Des} c = \{ij \in E \mid i < j \text{ and } c(i) > c(j)\}$$

and $\operatorname{des} c = \# \operatorname{Des} c$, respectively.

Theorem 8.5.3. *Let G be a graph with $V = [n]$ and $\#E = m$.*

(a) $\rho(X(G; \mathbf{x}, q)) = q^m X(G; \mathbf{x}, q^{-1})$.

(b) $\rho(X(G; \mathbf{x}, q)) = \sum_{c \in \mathcal{PC}(G)} \mathbf{x}^c q^{\operatorname{des} c}$.

Proof. (a) For a composition $\alpha = [\alpha_1, \ldots, \alpha_l]$ and $k \in \mathbb{N}$, the coefficient of $x_1^{\alpha_1} \cdots x_l^{\alpha_l} q^k$ on the left side of (a) is

$$\begin{aligned}[x_1^{\alpha_1} \cdots x_l^{\alpha_l} q^k]\rho(X(G; \mathbf{x}, q)) &= [x_1^{\alpha_l} \cdots x_l^{\alpha_1} q^k] X(G; \mathbf{x}, q) \\ &= \#\{\text{proper } c \mid \mathbf{x}^c = x_1^{\alpha_l} \cdots x_l^{\alpha_1} \text{ and } \operatorname{asc} c = k\}. \end{aligned} \tag{8.14}$$

Similarly

$$\begin{aligned}[x_1^{\alpha_1} \cdots x_l^{\alpha_l} q^k](q^m X(G; \mathbf{x}, q^{-1})) &= [x_1^{\alpha_1} \cdots x_l^{\alpha_l} q^{m-k}] X(G; \mathbf{x}, q) \\ &= \#\{\text{proper } c' \mid \mathbf{x}^{c'} = x_1^{\alpha_1} \cdots x_l^{\alpha_l} \text{ and } \operatorname{asc} c' = m-k\}. \end{aligned} \tag{8.15}$$

So it suffices to find a bijection between the c counted by (8.14) and the c' counted by (8.15).

Define $f : [l] \to [l]$ by $f(i) = l - i + 1$ for all i. Now given c as in (8.14), we let $c' = f \circ c$. We have that c' is still proper since f is a bijection. If c sends α_i vertices to color i, then c' sends that many vertices to color $l - i + 1$. So their monomials are related as desired. And since f is order reversing, $\operatorname{asc} c' = m - \operatorname{asc} c$. Finally, f induces a bijection on colorings since $c = f \circ c'$.

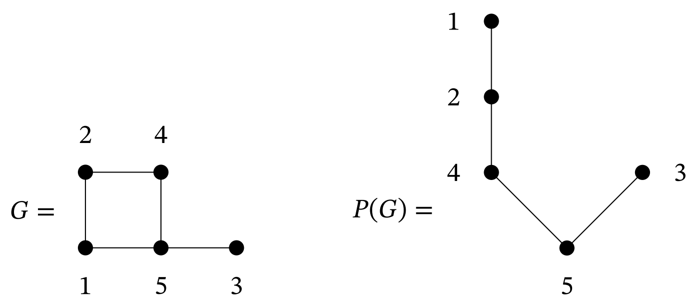


Figure 8.3. A graph and its poset

(b) For any proper coloring we have $\operatorname{Des} c = E - \operatorname{Asc} c$ and so $\operatorname{des} c = m - \operatorname{asc} c$. It follows that

$$q^m X(G; \mathbf{x}, q^{-1}) = \sum_{c \in \mathcal{PC}(G)} \mathbf{x}^c q^{\operatorname{des} c}.$$

The result now follows from part (a). $\square$

Similarly to $\operatorname{QSym}[q]$, define $\operatorname{Sym}[q]$ to be the set of polynomials in q whose coefficients are symmetric functions. The reader should find it easy to supply the details of the demonstration of the next result.

Corollary 8.5.4. *Let G be a graph with $V = [n]$ and $\#E = m$ such that $X(G; \mathbf{x}, q) \in$ $\operatorname{Sym}[q]$.*

(a) $X(G; \mathbf{x}, q) = \sum_{c \in \mathcal{PC}(G)} \mathbf{x}^c q^{\operatorname{des} c}$.

(b) *$X(G; \mathbf{x}, q)$ is palindromic in q with center $m/2$.* $\square$

We end by making a connection with reverse P-partitions. Given any graph G with $V = [n]$, there is an associated poset $P(G)$ on $[n]$ defined as follows. Call a path $i_1, i_2, \ldots, i_l$ in G *decreasing* if $i_1 > i_2 > \cdots > i_l$. Now define $i \trianglelefteq j$ in $P(G)$ if there is a decreasing path from i to j in G. See Figure 8.3 for an example. We must make sure that $P(G)$ satisfies the poset axioms.

Lemma 8.5.5. *If G is a graph with $V = [n]$, then $P(G)$ is a poset on $[n]$.*

Proof. We have $i \trianglelefteq i$ for $i \in [n]$ because of the path of length zero from i to itself which is decreasing. If $i \trianglelefteq j$ and $j \trianglelefteq i$, then the decreasing path from i to j forces $i \geq j$. Similarly $j \geq i$ so that $i = j$. Finally, if $i \trianglelefteq j$ and $j \trianglelefteq k$, then consider the concatenation P of the decreasing paths from i to j and from j to k. Now the vertices on P form a decreasing sequence since the two individual paths are decreasing and the terminal vertex of the first equals the initial vertex of the second. This also shows that P is a path since the decreasing condition makes it impossible to repeat a vertex. So $i \trianglelefteq k$ which finishes transitivity and the proof. $\square$

We can now show that the coefficient of q^0 in $X(G; \mathbf{x}, q)$ is the quasisymmetric generating function for reverse $P(G)$-partitions.

Theorem 8.5.6. *If G is a graph with $V = [n]$, then*

$$X_0(G; \mathbf{x}) = \operatorname{rpar}(P(G); \mathbf{x}).$$

Proof. It suffices to show that any proper coloring of G with no ascents is a reverse $P(G)$-partition and conversely. So let $c : V \to \mathbb{P}$ be a proper coloring with $\operatorname{asc} c = 0$. Recall that it suffices to prove that RPP1 and RPP2 hold for covers $i \lhd j$. But then $i > j$ must be a decreasing path consisting of a single edge. And since this is a proper coloring with no ascents, $c(i) < c(j)$. So both of the desired conditions hold.

For the other direction, let $f : P(G) \to \mathbb{P}$ be a reverse partition. By definition of $P(G)$, any cover $i \lhd j$ comes from an edge ij with $i > j$. By RPP2, we have $f(i) < f(j)$. So f is a proper coloring since the second inequality is strict and, by comparing the last two inequalities, has no ascents. This completes the proof, as well as the book. □

Exercises

(1) Prove that QSym is an algebra; that is, show it is closed under linear combinations and products.

(2) Prove Proposition 8.1.2.

(3) (a) Show that M_α is symmetric if and only if $\alpha = [i^m]$ for some i, m. In addition, show that $M_{[i^m]} = m_{(i^m)}$.
  (b) Show that F_α is symmetric if and only if $\alpha = [n]$ or $\alpha = [1^n]$ for some n. In addition, show that $F_n = h_n$ and $F_{1^n} = e_n$.
  (c) Suppose that $\alpha, \beta \models n$ are distinct compositions where $n > 3$. Show that $F_\alpha + F_\beta$ is symmetric if and only if $\{\alpha, \beta\} = \{[n], [1^n]\}$.

(4) (a) Show that the M_α can be expressed in terms of the F_α as

$$M_\alpha = (-1)^{\ell(\alpha)} \sum_{\beta \le \alpha} (-1)^{\ell(\beta)} F_\beta$$

  where $\ell(\cdot)$ is the length function in two ways: using Theorem 1.3.3(d) and using Möbius inversion.
  (b) Show that part (a) implies Theorem 8.1.4.

(5) Prove Lemma 8.2.1.

(6) Complete the proof of equation (8.4).

(7) (a) Prove that the map between reverse compatible and compatible functions given in the text is a well-defined bijection and that if f maps to g, then $|f| = |g| + n$ where the permutations come from $\mathfrak{S}_n$.
  (b) Given a poset P on $[n]$, find a poset Q on $[n]$ such that there is a bijection $\psi : \operatorname{RPar} P \to \operatorname{Par} Q$ satisfying $|f| = |\psi(f)| + n$ for all $f \in \operatorname{RPar} P$.

(8) (a) Show that if Q is the recording tableau for π under the Robinson–Schensted map, then $\operatorname{Des} Q = \operatorname{Des} \pi$.
(b) Show in the proof of Theorem 8.2.5 that $\operatorname{Des} \pi \subseteq \operatorname{Des} Q$.

(9) Prove Lemma 8.2.3.

(10) Prove Theorem 8.2.4.

(11) Prove Proposition 8.2.6.

(12) (a) Show that if $\pi \in \mathfrak{S}_n$, then we have the principal specialization
$$F_{\operatorname{Des} \pi}(1, q, q^2, \dots) = \frac{q^{\operatorname{maj} \pi'}}{(1-q)(1-q^2)\cdots(1-q^n)}$$
where π' is the reverse complement of π as given by (8.3)
(b) Use part (a) and Theorem 8.2.4 to rederive equation (7.23).

(13) Prove that if $\sigma \in P(U)$ and $\tau \in P(V)$ where U, V are disjoint and $\#U = m$, $\#V = n$, then
$$\sum_{\pi \in \sigma \sqcup\!\sqcup \tau} q^{\operatorname{maj} \pi} = q^{\operatorname{maj} \sigma + \operatorname{maj} \tau} \begin{bmatrix} m+n \\ m \end{bmatrix}_q .$$
Conclude that maj is a shuffle compatible function on permutations. Hint: Use equation (7.23).

(14) Show that the two maps in the proof of Theorem 8.3.1 are inverses.

(15) (a) Let P be a finite, ranked poset with a $\hat{1}$ having $\operatorname{rk} \hat{1} = n$. Associate with any chain $C : \hat{0} = x_0 < x_1 < \cdots < x_k = \hat{1}$ the set $S(C) = \{\operatorname{rk} x_1, \operatorname{rk} x_2, \dots, \operatorname{rk} x_{k-1}\} \subseteq [n-1]$. Show that $\phi(S(C)) = \alpha(C)$ where ϕ is the map in (1.8).
(b) Complete the proof of Theorem 8.3.2.

(16) (a) Show that for the n-chain we have $M(C_n) = h_n$, the complete homogeneous symmetric function of degree n.
(b) Show that if the prime factorization of n is $n = p_1^{\lambda_1} \cdots p_l^{\lambda_l}$ where $\lambda = (\lambda_1, \dots, \lambda_l)$ is a partition, then for the divisor lattice $M(D_n) = h_\lambda$.

(17) Prove that a permutation $\pi \in \mathfrak{S}_n$ is reverse layered if and only if we have $\pi_{i+1} \le \pi_i + 1$ for $1 \le i < n$.

(18) Prove Lemma 8.4.2.

(19) Prove part (b) of Proposition 8.4.3.

(20) Prove that part (c) of Proposition 8.4.3 is implied by parts (a) and (b) and Lemma 8.4.2.

(21) Show that if $i \in [n]$, then
$$s_{(i,1^{n-i})} = \sum_S F_S$$
where the sum is over $S \in \binom{[n-1]}{n-i}$.

(22) (a) The *Knuth class* corresponding to an SYT P is the set
$$K(P) = \left\{\pi \mid \pi \overset{\mathrm{RS}}{\mapsto} (P, Q) \text{ for some SYT } Q\right\}.$$
Prove that if F_Π is as defined in (8.9) and $\operatorname{sh} P = \lambda$, then
$$F_{K(P)} = s_\lambda .$$

(b) The *Knuth aggregate* corresponding to a partition λ is the set

$$K(\lambda) = \biguplus_{P \in \mathrm{SYT}(\lambda)} K(P).$$

Prove that

$$F_{K(\lambda)} = f^\lambda s_\lambda.$$

(c) Show that

$$Q_n(\emptyset) = \sum_{\lambda \vdash n} f^\lambda s_\lambda.$$

(d) If $\iota_k = 12\ldots k$, then show

$$Q_n(\iota_k) = \sum_{\lambda_1 < k} f^\lambda s_\lambda.$$

(e) If $\delta_k = k \ldots 21$, then show

$$Q_n(\delta_k) = \sum_{\lambda_1^t < k} f^\lambda s_\lambda.$$

(23) (a) One can define $X(G; \mathbf{x}, q)$ for any graph whose vertices are distinct positive integers just as we did in the case $V = [n]$. With this extended definition, show that for a disjoint union of graphs we have

$$X(G \uplus H; \mathbf{x}, q) = X(G; \mathbf{x}, q) X(H; \mathbf{x}, q).$$

(b) Show that for the empty graph which has n vertices and no edges

$$X(\emptyset; \mathbf{x}, q) = e_{1^n}(\mathbf{x}).$$

(c) Show that for the complete graph on n vertices

$$X(K_n; \mathbf{x}, q) = e_n(\mathbf{x})[n]_q!\,.$$

(24) Prove Lemma 8.5.2.

(25) Prove Corollary 8.5.4.

Introduction to Representation Theory

This appendix is designed to give just enough information about representation theory to understand some of the material in the body of the text. As such, most proofs are omitted. The reader wanting more information is encouraged to consult the texts of James [**45**], James and Kerber [**46**], or Sagan [**79**].

A.1. Basic notions

Let G be a finite group and let V be a finite-dimensional vector space over the complex numbers. We say that V is a *G-module* or that *V affords a representation of G* if there is an action of G on V such that each map $g : V \to V$ is linear. Since each such function is bijective, g is in fact an element of the *general linear group*, $\mathrm{GL}(V)$, of invertible linear transformations on V. So to say that V is a G-module is equivalent to saying that there is a homomorphism of groups $\rho : G \to \mathrm{GL}(V)$. We will often write $[g]$ for $\rho(g)$. By definition, $g \to [g]$ being a homomorphism means

$$[gh] = [g][h] \tag{A.1}$$

for all $g, h \in G$ where the product on the left is in G and on the right we have composition of linear transformations. The matrix for the linear map $[g]$ in a basis B of V will be denoted $[g]_B$, where we may drop the subscript if the basis is clear from the context. The *dimension* of a representation V is just the usual vector space dimension $\dim V$.

Every group G has the *trivial representation* where $V = \mathbb{C}$ and $gc = c$ for all $g \in G$ and $c \in \mathbb{C}$. Equivalently $[g] = [1]$ for all $g \in G$.

For a less trivial example, we can turn any set X on which G acts into a G-module by considering that vector space $\mathbb{C}X$ generated by X as defined in (7.32). Indeed, since G acts on X which is a basis for $\mathbb{C}X$, the action can be extended to $\mathbb{C}X$ by linearity. Clearly $\dim \mathbb{C}X = \#X$. To be even more concrete, consider the action of $\mathfrak{S}_3$ on $[3]$ and hence on

$$\mathbb{C}[3] = \{c_1\mathbf{1} + c_2\mathbf{2} + c_3\mathbf{3} \mid c_1, c_2, c_3 \in \mathbb{C}\}.$$

Note the distinction between $[n]$ for $n \in \mathbb{N}$ and $[g]$ for $g \in G$. Note also the use of boldface numbers to denote the corresponding vectors. To find the matrix for the permutation $(1, 3, 2)$ in the basis $X = [3]$ we compute the action on each basis vector

$$(1,3,2)\mathbf{1} = \mathbf{3}, \quad (1,3,2)\mathbf{2} = \mathbf{1}, \quad (1,3,2)\mathbf{3} = \mathbf{2}.$$

This corresponds to the matrix

$$[(1,3,2)]_X = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

The reader should find it easy to write out the matrices for the rest of $\mathfrak{S}_3$ and verify that (A.1) holds. If $X = [n]$, then the representation of $\mathfrak{S}_n$ afforded by $\mathbb{C}[n]$ is called its *defining representation*. The matrix $[\pi]_X$ for $\pi \in \mathfrak{S}_n$ is called its corresponding *permutation matrix*. More generally, if G acts on X, then the G-module $\mathbb{C}X$ is called a *permutation representation* of G.

We will be particularly concerned with representations of cyclic groups. Suppose that G is cyclic with $\#G = n$, and let g be a generator of G. Let us find the 1-dimensional representations of G. Suppose that we have a homomorphism $\rho : G \to \mathbb{C}$ which sends g to the matrix $[c]$ for some $c \in \mathbb{C}$. The value of c completely determines ρ since g generates G and, by (A.1),

$$[g^i] = [g]^i = [c]^i = [c^i]$$

for any $i \geq 0$. Furthermore, since $g^n = e$, we must have $[c^n] = [1]$ and so c must be an nth root of unity. It is now easy to check that one obtains n 1-dimensional representations of G by letting $\rho(g^i) = [\omega^i]$ for each nth root of unity ω.

Sometimes we have two G-modules where the action of G is essentially the same. Two G-modules V, W are said to be *G-isomorphic* or *G-equivalent*, written $V \cong W$, if there is an isomorphism of vector spaces $\phi : V \to W$ which respects the action of G in that

$$g\phi(v) = \phi(gv) \tag{A.2}$$

for all $g \in G$ and $v \in V$. Stated in terms of matrices, this definition means that there are bases B for V and $C = \phi(B)$ for W such that

$$[g]_B = [g]_C$$

for all $g \in G$. Otherwise V and W are *G-inequivalent*. We will drop the "G-" modifier if the group is clear from the context.

To illustrate, let us revisit the defining representation of $\mathfrak{S}_3$. Consider the subspace W of $\mathbb{C}[3]$ generated by the vector $\mathbf{1} + \mathbf{2} + \mathbf{3}$:

$$W = \mathbb{C}\{\mathbf{1} + \mathbf{2} + \mathbf{3}\} = \{c(\mathbf{1} + \mathbf{2} + \mathbf{3}) \mid c \in \mathbb{C}\}. \tag{A.3}$$

So for any $w = c(\mathbf{1}+\mathbf{2}+\mathbf{3})$ and $\pi \in \mathfrak{S}_3$, the fact that π is linear and permutes the three vectors $\mathbf{1}, \mathbf{2}, \mathbf{3}$ yields

$$\pi(w) = \pi(c(\mathbf{1} + \mathbf{2} + \mathbf{3})) = c\pi(\mathbf{1} + \mathbf{2} + \mathbf{3}) = c(\mathbf{1} + \mathbf{2} + \mathbf{3}) = w.$$

It follows that $[\pi] = [1]$ for all π and so W is equivalent to the trivial representation of $\mathfrak{S}_3$. More generally, given any permutation representation $\mathbb{C}X$ for a group G, the subspace $W = \mathbb{C}\{w\}$ where $w = \sum_{x\in X} x$ is equivalent to the trivial representation of G.

On the other hand, the n representations of a cyclic group $G = \langle g\rangle$ of order n which we derived above are all inequivalent, for suppose we consider two representations such that

$$\rho(g) = [\omega] \quad \text{and} \quad \rho'(g) = [\omega'] \tag{A.4}$$

for two nth roots of unity ω, ω'. Any vector space isomorphism $\phi : \mathbb{C} \to \mathbb{C}$ is multiplication by some $c \in \mathbb{C} - \{0\}$. So, considering (A.2) with $v = 1$,

$$g\phi(v) = \omega'\phi(v) = \omega'(c1) = c\omega'$$

since g is acting on the representation ρ' in the range of ϕ. On the other hand

$$\phi(gv) = \phi(\omega v) = c(\omega 1) = c\omega$$

for now g is acting by ρ in the domain of ϕ. Setting these two evaluations equal forces $\omega = \omega'$.

If V, W are G-modules, then it is easy to see that $V \oplus W$ is also, where the action is defined by

$$g(v + w) = gv + gw$$

for $g \in G$, $v \in V$, $w \in W$. It turns out that all G-modules can be constructed this way from certain building blocks which are called the irreducible modules. If V is a G-module, then a *submodule* of V is a subspace $W \subseteq V$ which is itself a G-module in that $gw \in W$ for all $g \in G$ and $w \in W$. Any G-module V has the *trivial submodules* consisting of the zero subspace and V itself. All other submodules are *nontrivial*. Note that the usage of the word "trivial" here is different from what we have defined as the trivial representation. Any group G has a unique trivial representation which has dimension 1. On the other hand, any G-module V has two (not necessarily distinct) submodules which are considered trivial. Call V *reducible* if it has nontrivial submodules and call it *irreducible* otherwise.

Clearly every 1-dimensional G-module is irreducible. On the other hand the defining module $\mathbb{C}[n]$ for $\mathfrak{S}_n$ is not irreducible for $n \geq 2$ because of the submodule

$$W = \mathbb{C}\{\mathbf{1} + \mathbf{2} + \cdots + \mathbf{n}\}. \tag{A.5}$$

Of course, W is irreducible since it has dimension 1. Consider the orthogonal complement $W^\perp$ using the inner product on $\mathbb{C}[n]$ given by $\mathbf{i} \cdot \mathbf{j} = \delta_{i,j}$. Now $\mathbb{C}[n] = W \oplus W^\perp$ as vector spaces. And one can show that $W^\perp$ is an irreducible $\mathfrak{S}_n$-module. It turns out that one can write any G-module as a direct sum of irreducibles. We would also like to know how many irreducible modules a group can have up to isomorphism.

Theorem A.1.1. *Let G be a finite group and consider the G-modules which are finite-dimensional vector spaces over $\mathbb{C}$.*

(a) *The number of pairwise inequivalent irreducible G-modules is finite and equals the number of conjugacy classes of G.*

(b) (*Maschke's Theorem*) *Every G-module can be written as a direct sum of irreducible G-modules.* □

We will use the notation

$$V \cong \bigoplus_i m_i V^{(i)} \tag{A.6}$$

to indicate that V is isomorphic to the direct sum of m_i copies of $V^{(i)}$ as i varies. If G is a cyclic group of order n, then, because G is abelian, G has n conjugacy classes each consisting of a single element. So from part (a) of the previous theorem, we know that the n inequivalent irreducible 1-dimensional representations we have found for G are a complete list. Also, because of part (b), any representation of G can be written $V = V^{(1)} \oplus \cdots \oplus V^{(k)}$ for irreducible submodules $V^{(1)}, \ldots, V^{(k)}$ all of dimension 1. Taking v_i to be a basis of $V^{(i)}$ for $1 \le i \le k$, we see that in the basis $B = \{v_1, \ldots, v_k\}$ the matrix $[g]_B$ will be diagonal for any $g \in G$. And because the diagonal elements come from the 1-dimensional representations we found, they are all nth roots of unity. We record this result for future reference.

Corollary A.1.2. *If G is a cyclic group of order n and V is a G-module, then there is a basis for V which simultaneously diagonalizes $[g]$ for all $g \in G$. Furthermore, the diagonal elements are nth roots of unity.* □

In the symmetric group $\mathfrak{S}_n$, a conjugacy class is just all permutations of a given cycle type $\lambda \vdash n$. So the irreducible representations of $\mathfrak{S}_n$ are also indexed by partitions of n. If V^λ is the irreducible module corresponding to λ, then one can show that

$$\dim V^\lambda = f^\lambda. \tag{A.7}$$

So, for example, $V^{(n)}$ is the trivial representation and $\dim V^{(n)} = 1$ which is the number of SYT of shape (n). As another illustration, consider the irreducible module $W^\perp$ where W is the submodule (A.5) of $\mathbb{C}[n]$. Then

$$\dim W^\perp = \dim \mathbb{C}[n] - \dim W = n - 1.$$

In fact, $W^\perp \cong V^{(n-1,1)}$ and it is easy to see that $f^{(n-1,1)} = n - 1$.

It would be nice if there was a natural representation of G which contained all the irreducible representations. This is the case for the regular representation. Any group G acts on the set $X = G$ by left multiplication

$$g(h) = gh \tag{A.8}$$

where on the left we have the action of g on h and on the right the product of g and h in the group. The corresponding G-module $\mathbb{C}G$ is called the (*left*) *regular representation* of G. To illustrate, consider $G = \mathfrak{S}_3$ and the ordered basis

$$B = \{\mathbf{e},\ (\mathbf{1},\mathbf{2}),\ (\mathbf{1},\mathbf{3}),\ (\mathbf{2},\mathbf{3}),\ (\mathbf{1},\mathbf{2},\mathbf{3}),\ (\mathbf{1},\mathbf{3},\mathbf{2})\}$$

of the regular representation. For $g = (1,3,2)$ we have, remembering that we compose permutations right to left,

$$\begin{array}{lll} (1,3,2)\mathbf{e} = (\mathbf{1},\mathbf{3},\mathbf{2}), & (1,3,2)(\mathbf{1},\mathbf{2}) = (\mathbf{2},\mathbf{3}), & (1,3,2)(\mathbf{1},\mathbf{3}) = (\mathbf{1},\mathbf{2}), \\ (1,3,2)(\mathbf{2},\mathbf{3}) = (\mathbf{1},\mathbf{3}), & (1,3,2)(\mathbf{1},\mathbf{2},\mathbf{3}) = \mathbf{e}, & (1,3,2)(\mathbf{1},\mathbf{3},\mathbf{2}) = (\mathbf{1},\mathbf{2},\mathbf{3}), \end{array}$$

with corresponding matrix

$$[(1,3,2)]_B = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We point out again (as we did in Chapter 7) that the sum of squares formula in equation (7.27) is the special case of (A.9) below where $G = \mathfrak{S}_n$.

Theorem A.1.3. *Let G be a finite group and let $V^{(1)}, \dots, V^{(k)}$ be all its pairwise inequivalent irreducible representations. Then the regular representation satisfies*

$$\mathbb{C}G \cong \bigoplus_{i=1}^{k} d_i V^{(i)}$$

where $d_i = \dim V^{(i)}$ for all i. In addition

$$\#G = \sum_{i=1}^{k} d_i^2. \tag{A.9}$$

Proof. For a proof of the first statement, see the demonstration of Proposition 1.10.1 in [**79**]. The second follows directly from the first by taking dimensions on both sides. □

It turns out that a lot of information about a representation can be gleaned from a very simple function. If V is a G-module, then its *character* is the map $\chi : G \to \mathbb{C}$ given by

$$\chi(g) = \operatorname{tr}[g]$$

where tr is the trace function. Note that since the trace of a linear transformation is independent of the basis in which it is computed, $\chi(g)$ is well-defined. Note also that for any representation V of dimension d we must have

$$\chi(e) = \operatorname{tr} I_d = d$$

where I_d is the $d \times d$ identity matrix. We also have that χ is a *class function* in that $\chi(g) = \chi(h)$ if g, h are in the same conjugacy class to G. This is because we must have $g = khk^{-1}$ for some $k \in G$. So, by (A.1) and the fact that the trace is invariant under conjugation,

$$\chi(g) = \chi(khk^{-1}) = \operatorname{tr}[khk^{-1}] = \operatorname{tr}([k][h][k]^{-1}) = \operatorname{tr}[h] = \chi(h).$$

It is also true that equivalent modules have the same character, for suppose $\phi : V \to W$ is an isomorphism of G-modules with characters χ^V and χ^W, respectively. Let B and $C = \phi(B)$ be bases for V and W and suppose that T is the matrix of ϕ with respect to the bases B and C. Since (A.2) holds for all $\in V$ we must have $[g]_C T = T[g]_B$ for all $g \in G$. Since T is invertible we have

$$\chi^W(g) = \operatorname{tr}[g]_C = \operatorname{tr}(T[g]_B T^{-1}) = \operatorname{tr}[g]_B = \chi^V(g).$$

Since this holds for all $g \in G$, it follows that $\chi^W = \chi^V$. The surprising thing is that the converse is also true.

Theorem A.1.4. *Let V and W be G-modules with characters χ^V and χ^W, respectively. We have $V \cong W$ if and only if $\chi^V = \chi^W$.* □

This theorem can provide a quick way of checking whether two representations are equivalent or not. For example, if we are considering a 1-dimensional representation, then $\chi(g)$ is just the single entry of the matrix $[g]$. So if we have two representations of a cyclic group as in (A.4) for distinct ω, ω', then we must have ρ and ρ' inequivalent since $\chi(g) = \omega \neq \omega' = \chi'(g)$.

Exercises

(1) Show that V is a G-module if and only if there is a map $\rho : G \to \mathrm{GL}(V)$ which is a homomorphism of groups.

(2) Show that if $G = \langle g \rangle$ is cyclic of order n and ω is an nth root of unity, then the map $\rho : G \to \mathrm{GL}(\mathbb{C})$ defined by $\rho(g^i) = [\omega^i]$ is a well-defined representation of G.

(3) Let V, W be G-modules. Show that V, W are equivalent if and only if they have bases B, C, respectively, such that
$$[g]_B = [g]_C$$
for all $g \in G$.

(4) Prove that if G acts on X, then the submodule of $\mathbb{C}X$ defined by $V = \langle v \rangle$ where $v = \sum_{x \in X} x$ is equivalent to the trivial representation.

(5) Show that if V, W are G-modules, then so is $V \oplus W$ with the action
$$g(v + w) = gv + gw$$
for $g \in G$, $v \in V$, $w \in W$.

(6) (a) Show that if V is a G-module, then the zero subspace and V itself are submodules.
(b) Consider the submodule $W = \mathbb{C}\{\mathbf{1}+\mathbf{2}+\mathbf{3}\}$ of the defining representation $\mathbb{C}[\mathbf{3}]$ of $\mathfrak{S}_3$. Show that $W^\perp$ is a submodule of $\mathbb{C}[\mathbf{3}]$ and that it is irreducible.

(7) Show that (A.8) defines a group action.

Bibliography

[1] K. Appel and W. Haken, *Every planar map is four colorable. I. Discharging*, Illinois J. Math. **21** (1977), no. 3, 429–490. MR543792

[2] K. Appel, W. Haken, and J. Koch, *Every planar map is four colorable. II. Reducibility*, Illinois J. Math. **21** (1977), no. 3, 491–567. MR543793

[3] Eric Babson and Einar Steingrímsson, *Generalized permutation patterns and a classification of the Mahonian statistics*, Sém. Lothar. Combin. **44** (2000), Art. B44b, 18. MR1758852

[4] Duff Baker-Jarvis and Bruce E. Sagan, *Bijective proofs of shuffle compatibility results*, Adv. in Appl. Math. **113** (2020), 101973, 29, DOI 10.1016/j.aam.2019.101973. MR4032316

[5] Matthias Beck and Raman Sanyal, *Combinatorial reciprocity theorems: An invitation to enumerative geometric combinatorics*, Graduate Studies in Mathematics, vol. 195, American Mathematical Society, Providence, RI, 2018. MR3839322

[6] Edward A. Bender and Donald E. Knuth, *Enumeration of plane partitions*, J. Combinatorial Theory Ser. A **13** (1972), 40–54, DOI 10.1016/0097-3165(72)90007-6. MR299574

[7] Carolina Benedetti and Nantel Bergeron, *The antipode of linearized Hopf monoids*, Algebr. Comb. **2** (2019), no. 5, 903–935, DOI 10.5802/alco.53. MR4023571

[8] Carolina Benedetti, Joshua Hallam, and John Machacek, *Combinatorial Hopf algebras of simplicial complexes*, SIAM J. Discrete Math. **30** (2016), no. 3, 1737–1757, DOI 10.1137/15M1038281. MR3543152

[9] Carolina Benedetti and Bruce E. Sagan, *Antipodes and involutions*, J. Combin. Theory Ser. A **148** (2017), 275–315, DOI 10.1016/j.jcta.2016.12.005. MR3603322

[10] Arthur T. Benjamin and Jennifer J. Quinn, *Proofs that really count: The art of combinatorial proof*, The Dolciani Mathematical Expositions, vol. 27, Mathematical Association of America, Washington, DC, 2003. MR1997773

[11] F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures*, translated from the 1994 French original by Margaret Readdy, with a foreword by Gian-Carlo Rota, Encyclopedia of Mathematics and its Applications, vol. 67, Cambridge University Press, Cambridge, 1998. MR1629341

[12] Nantel Bergeron and Cesar Ceballos, *A Hopf algebra of subword complexes*, Adv. Math. **305** (2017), 1163–1201, DOI 10.1016/j.aim.2016.10.007. MR3570156

[13] George D. Birkhoff, *A determinant formula for the number of ways of coloring a map*, Ann. of Math. (2) **14** (1912/13), no. 1-4, 42–46, DOI 10.2307/1967597. MR1502436

[14] George D. Birkhoff, *On the combination of subalgebras*, Proc. Camb. Phil. Soc. **29** (1933), 441–464.

[15] A. Björner, *Topological methods*, Handbook of combinatorics, Vol. 2, Elsevier Sci. B. V., Amsterdam, 1995, pp. 1819–1872. MR1373690

[16] Andreas Blass and Bruce Eli Sagan, *Bijective proofs of two broken circuit theorems*, J. Graph Theory **10** (1986), no. 1, 15–21, DOI 10.1002/jgt.3190100104. MR830053

[17] Jonathan S. Bloom and Bruce E. Sagan, *Revisiting pattern avoidance and quasisymmetric functions*, Ann. Comb. **24** (2020), no. 2, 337–361, DOI 10.1007/s00026-020-00492-6. MR4110402

[18] Miklós Bóna, *Combinatorics of permutations*, 2nd ed., with a foreword by Richard Stanley, Discrete Mathematics and its Applications (Boca Raton), CRC Press, Boca Raton, FL, 2012. MR2919720

[19] Petter Brändén, *Unimodality, log-concavity, real-rootedness and beyond*, Handbook of enumerative combinatorics, Discrete Math. Appl. (Boca Raton), CRC Press, Boca Raton, FL, 2015, pp. 437–483. MR3409348

[20] Francesco Brenti, *Log-concave and unimodal sequences in algebra, combinatorics, and geometry: an update*, Jerusalem combinatorics ’93, Contemp. Math., vol. 178, Amer. Math. Soc., Providence, RI, 1994, pp. 71–89, DOI 10.1090/conm/178/01893. MR1310575

[21] W. Burnside, *Theory of groups of finite order*, 2nd ed, Dover Publications, Inc., New York, 1955. MR0069818

[22] David M. Burton, *The history of mathematics: An introduction*, 2nd ed., W. C. Brown Publishers, Dubuque, IA, 1991. MR1223776

[23] Peter Doubilet, Gian-Carlo Rota, and Richard Stanley, *On the foundations of combinatorial theory. VI. The idea of generating function*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Univ. California Press, Berkeley, Calif., 1972, pp. 267–318. MR0403987

[24] Richard Ehrenborg, *On posets and Hopf algebras*, Adv. Math. **119** (1996), no. 1, 1–25, DOI 10.1006/aima.1996.0026. MR1383883

[25] Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009. MR2483235

[26] Dominique Foata, *Distributions eulériennes et mahoniennes sur le groupe des permutations* (French), with a comment by Richard P. Stanley, Higher combinatorics (Proc. NATO Advanced Study Inst., Berlin, 1976), NATO Adv. Study Inst. Ser., Ser. C: Math. Phys. Sci., vol. 31, Reidel, Dordrecht-Boston, Mass., 1977, pp. 27–49. MR519777

[27] Sergey Fomin, *Duality of graded graphs*, J. Algebraic Combin. **3** (1994), no. 4, 357–404, DOI 10.1023/A:1022412010826. MR1293822

[28] Sergey Fomin, *Schensted algorithms for dual graded graphs*, J. Algebraic Combin. **4** (1995), no. 1, 5–45, DOI 10.1023/A:1022404807578. MR1314558

[29] J. S. Frame, G. de B. Robinson, and R. M. Thrall, *The hook graphs of the symmetric groups*, Canad. J. Math. **6** (1954), 316–324, DOI 10.4153/cjm-1954-030-1. MR62127

[30] A. M. Garsia and S. C. Milne, *A Rogers-Ramanujan bijection*, J. Combin. Theory Ser. A **31** (1981), no. 3, 289–339, DOI 10.1016/0097-3165(81)90062-5. MR635372

[31] Ira Gessel and Gérard Viennot, *Binomial determinants, paths, and hook length formulae*, Adv. in Math. **58** (1985), no. 3, 300–321, DOI 10.1016/0001-8708(85)90121-5. MR815360

[32] Ira M. Gessel, *Multipartite P-partitions and inner products of skew Schur functions*, Combinatorics and algebra (Boulder, Colo., 1983), Contemp. Math., vol. 34, Amer. Math. Soc., Providence, RI, 1984, pp. 289–317, DOI 10.1090/conm/034/777705. MR777705

[33] Ira M. Gessel and Yan Zhuang, *Shuffle-compatible permutation statistics*, Adv. Math. **332** (2018), 85–141, DOI 10.1016/j.aim.2018.05.003. MR3810249

[34] Curtis Greene, *An extension of Schensted’s theorem*, Advances in Math. **14** (1974), 254–265, DOI 10.1016/0001-8708(74)90031-0. MR354395

[35] Curtis Greene, Albert Nijenhuis, and Herbert S. Wilf, *A probabilistic proof of a formula for the number of Young tableaux of a given shape*, Adv. in Math. **31** (1979), no. 1, 104–109, DOI 10.1016/0001-8708(79)90023-9. MR521470

[36] Darij Grinberg, *Shuffle-compatible permutation statistics II: the exterior peak set*, Electron. J. Combin. **25** (2018), no. 4, Paper No. 4.17, 61. MR3874283

[37] P. Hall, *The Eulerian functions of a group*, Quart. J. Math. **7** (1936), no. 1, 134–151.

[38] Joshua Hallam, *Applications of quotient posets*, Preprint `arXiv:1411.3022`.

[39] Joshua Hallam, *Applications of quotient posets*, Discrete Math. **340** (2017), no. 4, 800–810, DOI 10.1016/j.disc.2016.11.019. MR3603561

[40] Joshua Hallam, Jeremy L. Martin, and Bruce E. Sagan, *Increasing spanning forests in graphs and simplicial complexes*, European J. Combin. **76** (2019), 178–198, DOI 10.1016/j.ejc.2018.09.011. MR3886522

[41] Joshua Hallam and Bruce Sagan, *Factoring the characteristic polynomial of a lattice*, J. Combin. Theory Ser. A **136** (2015), 39–63, DOI 10.1016/j.jcta.2015.06.006. MR3383266

[42] Zachary Hamaker, Brendan Pawlowski, and Bruce Sagan, *Pattern avoidance and quasisymmetric functions*, Preprint `arXiv:1810.11372`.

[43] A. P. Hillman and R. M. Grassl, *Reverse plane partitions and tableau hook numbers*, J. Combinatorial Theory Ser. A **21** (1976), no. 2, 216–221, DOI 10.1016/0097-3165(76)90065-0. MR414387

[44] C. G. J. Jacobi, *De functionibus alternantibus earumque divisione per productum e differentiis elementorum conflatum* (German), J. Reine Angew. Math. **22** (1841), 360–371, DOI 10.1515/crll.1841.22.360. MR1578283

[45] G. D. James, *The representation theory of the symmetric groups*, Lecture Notes in Mathematics, vol. 682, Springer, Berlin, 1978. MR513828

[46] Gordon James and Adalbert Kerber, *The representation theory of the symmetric group*, with a foreword by P. M. Cohn, with an introduction by Gilbert de B. Robinson, Encyclopedia of Mathematics and its Applications, vol. 16, Addison-Wesley Publishing Co., Reading, Mass., 1981. MR644144

[47] André Joyal, *Une théorie combinatoire des séries formelles* (French, with English summary), Adv. in Math. **42** (1981), no. 1, 1–82, DOI 10.1016/0001-8708(81)90052-9. MR633783

[48] Sergey Kitaev, *Patterns in permutations and words*, with a foreword by Jeffrey B. Remmel, Monographs in Theoretical Computer Science. An EATCS Series, Springer, Heidelberg, 2011. MR3012380

[49] Donald E. Knuth, *Permutations, matrices, and generalized Young tableaux*, Pacific J. Math. **34** (1970), 709–727. MR272654

[50] Donald E. Knuth, *Subspaces, subsets, and partitions*, J. Combinatorial Theory Ser. A **10** (1971), 178–180, DOI 10.1016/0097-3165(71)90022-7. MR270933

[51] Donald E. Knuth, *The art of computer programming. Vol. 1*, Fundamental algorithms, third edition [of MR0286317], Addison-Wesley, Reading, MA, 1997. MR3077152

[52] Donald E. Knuth, *The art of computer programming. Vol. 2*, Seminumerical algorithms, third edition [of MR0286318], Addison-Wesley, Reading, MA, 1998. MR3077153

[53] Donald E. Knuth, *The art of computer programming. Vol. 3*, Sorting and searching, second edition [of MR0445948], Addison-Wesley, Reading, MA, 1998. MR3077154

[54] C. Krattenthaler, *Advanced determinant calculus*, The Andrews Festschrift (Maratea, 1998), Sém. Lothar. Combin. **42** (1999), Art. B42q, 67. MR1701596

[55] C. Krattenthaler, *Advanced determinant calculus: a complement*, Linear Algebra Appl. **411** (2005), 68–166, DOI 10.1016/j.laa.2005.06.042. MR2178686

[56] Joseph Louis Lagrange, *Demonstration d'un théorème nouveau concernant les nombres premiers*, Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres (Berlin) **2** (1771), 125–137.

[57] Bernt Lindström, *On the vector representations of induced matroids*, Bull. London Math. Soc. **5** (1973), 85–90, DOI 10.1112/blms/5.1.85. MR335313

[58] Dudley E. Littlewood, *The theory of group characters and matrix representations of groups*, reprint of the second (1950) edition, AMS Chelsea Publishing, Providence, RI, 2006. MR2213154

[59] E. Lucas, *Sur les congruences des nombres eulériens et les coefficients différentiels des functions trigonométriques suivant un module premier* (French), Bull. Soc. Math. France **6** (1878), 49–54. MR1503769

[60] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., with contribution by A. V. Zelevinsky and a foreword by Richard Stanley, reprint of the 2008 paperback edition [MR1354144], Oxford Classic Texts in the Physical Sciences, The Clarendon Press, Oxford University Press, New York, 2015. MR3443860

[61] P. A. MacMahon, *The Indices of Permutations and the Derivation Therefrom of Functions of a Single Variable Associated with the Permutations of any Assemblage of Objects*, Amer. J. Math. **35** (1913), no. 3, 281–322, DOI 10.2307/2370312. MR1506186

[62] F. Mertens, *Über eine zahlentheoretische Funktion*, Sem. ber. Kais. Akad. Wiss. Wien **106** (1897), no. 106, 761–830.

[63] Sri Gopal Mohanty, *Lattice path counting and applications*, Probability and Mathematical Statistics, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London-Toronto, Ont., 1979. MR554084

[64] J. W. Moon, *Counting labelled trees*, from lectures delivered to the Twelfth Biennial Seminar of the Canadian Mathematical Congress (Vancouver), vol. 1969, Canadian Mathematical Congress, Montreal, Que., 1970. MR0274333

[65] John J. O'Connor and Edmund F. Robertson, *Abu Ali al-Hasan ibn al-Haytham*, MacTutor History of Mathematics archive, University of St Andrews. http://www-history.mcs.st-andrews.ac.uk/Biographies/Al-Haytham.html.

[66] A. M. Odlyzko and H. J. J. te Riele, *Disproof of the Mertens conjecture*, J. Reine Angew. Math. **357** (1985), 138–160, DOI 10.1515/crll.1985.357.138. MR783538

[67] Oystein Ore, *Number theory and its history*, reprint of the 1948 original, with a supplement, Dover Publications, Inc., New York, 1988. MR939614

[68] Julius Petersen, *Beviser for Wilsons og Fermats Theoremer*, Tidssk. f. Math. **(3) 2** (1872), 64–65.

[69] T. Kyle Petersen, *Eulerian numbers*, with a foreword by Richard Stanley, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser/Springer, New York, 2015. MR3408615

[70] G. Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen* (German), Acta Math. **68** (1937), no. 1, 145–254, DOI 10.1007/BF02546665. MR1577579

[71] J. Howard Redfield, *The Theory of Group-Reduced Distributions*, Amer. J. Math. **49** (1927), no. 3, 433–455, DOI 10.2307/2370675. MR1506633

[72] V. Reiner, D. Stanton, and D. White, *The cyclic sieving phenomenon*, J. Combin. Theory Ser. A **108** (2004), no. 1, 17–50, DOI 10.1016/j.jcta.2004.04.009. MR2087303

[73] J. B. Remmel, *Bijective proofs of formulae for the number of standard Young tableaux*, Linear and Multilinear Algebra **11** (1982), no. 1, 45–100, DOI 10.1080/03081088208817432. MR647727

[74] Bernhard Riemann, *Gesammelte mathematische Werke, wissenschaftlicher Nachlass und Nachträge* (German), based on the edition by Heinrich Weber and Richard Dedekind, edited and with a preface by Raghavan Narasimhan, BSB B. G. Teubner Verlagsgesellschaft, Leipzig; Springer-Verlag, Berlin, 1990. MR1066697

[75] G. de B. Robinson, *On the Representations of the Symmetric Group*, Amer. J. Math. **60** (1938), no. 3, 745–760, DOI 10.2307/2371609. MR1507943

[76] Gian-Carlo Rota, *On the foundations of combinatorial theory. I. Theory of Möbius functions*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **2** (1964), 340–368 (1964), DOI 10.1007/BF00531932. MR174487

[77] Bruce E. Sagan, *Congruences via abelian groups*, J. Number Theory **20** (1985), no. 2, 210–237, DOI 10.1016/0022-314X(85)90041-1. MR790783

[78] Bruce E. Sagan, *Why the characteristic polynomial factors*, Bull. Amer. Math. Soc. (N.S.) **36** (1999), no. 2, 113–133, DOI 10.1090/S0273-0979-99-00775-2. MR1659875

[79] Bruce E. Sagan, *The symmetric group: Representations, combinatorial algorithms, and symmetric functions*, 2nd ed., Graduate Texts in Mathematics, vol. 203, Springer-Verlag, New York, 2001. MR1824028

[80] Bruce E. Sagan, *The cyclic sieving phenomenon: a survey*, Surveys in combinatorics 2011, London Math. Soc. Lecture Note Ser., vol. 392, Cambridge Univ. Press, Cambridge, 2011, pp. 183–233. MR2866734

[81] C. Schensted, *Longest increasing and decreasing subsequences*, Canadian J. Math. **13** (1961), 179–191, DOI 10.4153/CJM-1961-015-3. MR121305

[82] John Shareshian and Michelle L. Wachs, *Chromatic quasisymmetric functions*, Adv. Math. **295** (2016), 497–551, DOI 10.1016/j.aim.2015.12.018. MR3488041

[83] R. P. Stanley, *Supersolvable lattices*, Algebra Universalis **2** (1972), 197–217, DOI 10.1007/BF02945028. MR309815

[84] Richard P. Stanley, *Ordered structures and partitions*, Memoirs of the American Mathematical Society, No. 119, American Mathematical Society, Providence, R.I., 1972. MR0332509

[85] Richard P. Stanley, *Acyclic orientations of graphs*, Discrete Math. **5** (1973), 171–178, DOI 10.1016/0012-365X(73)90108-8. MR317988

[86] Richard P. Stanley, *Combinatorial reciprocity theorems*, Advances in Math. **14** (1974), 194–253, DOI 10.1016/0001-8708(74)90030-9. MR411982

[87] Richard P. Stanley, *Binomial posets, Möbius inversion, and permutation enumeration*, J. Combinatorial Theory Ser. A **20** (1976), no. 3, 336–356, DOI 10.1016/0097-3165(76)90028-5. MR409206

[88] Richard P. Stanley, *Differential posets*, J. Amer. Math. Soc. **1** (1988), no. 4, 919–961, DOI 10.2307/1990995. MR941434

[89] Richard P. Stanley, *Log-concave and unimodal sequences in algebra, combinatorics, and geometry*, Graph theory and its applications: East and West (Jinan, 1986), Ann. New York Acad. Sci., vol. 576, New York Acad. Sci., New York, 1989, pp. 500–535, DOI 10.1111/j.1749-6632.1989.tb16434.x. MR1110850

[90] Richard P. Stanley, *Variations on differential posets*, Invariant theory and tableaux (Minneapolis, MN, 1988), IMA Vol. Math. Appl., vol. 19, Springer, New York, 1990, pp. 145–165. MR1035494

[91] Richard P. Stanley, *A symmetric function generalization of the chromatic polynomial of a graph*, Adv. Math. **111** (1995), no. 1, 166–194, DOI 10.1006/aima.1995.1020. MR1317387

[92] Richard P. Stanley, *Catalan numbers*, Cambridge University Press, New York, 2015. MR3467982

[93] Richard P. Stanley and John R. Stembridge, *On immanants of Jacobi-Trudi matrices and permutations with restricted position*, J. Combin. Theory Ser. A **62** (1993), no. 2, 261–279, DOI 10.1016/0097-3165(93)90048-D. MR1207737

[94] N. Trudi, *Intorno un determinante piu generale di quello che suol dirsi determinante delle radici di una equazione, ed alle funzioni simmetriche complete di queste radici*, Rend. Accad. Sci. Fis. Mat. Napoli **3** (1864), 121–134.

[95] J. H. van Lint, *Introduction to coding theory*, 3rd ed., Graduate Texts in Mathematics, vol. 86, Springer-Verlag, Berlin, 1999. MR1664228

[96] Michelle L. Wachs, *Poset topology: tools and applications*, Geometric combinatorics, IAS/Park City Math. Ser., vol. 13, Amer. Math. Soc., Providence, RI, 2007, pp. 497–615. MR2383132

[97] Edward Waring, *Meditationes algebraicæ*, translated from the Latin, edited and with a foreword by Dennis Weeks, with an appendix by Franz X. Mayer, translated from the German by Weeks, American Mathematical Society, Providence, RI, 1991. MR1146921

[98] Louis Weisner, *Abstract theory of inversion of finite series*, Trans. Amer. Math. Soc. **38** (1935), no. 3, 474–484, DOI 10.2307/1989808. MR1501822

[99] Hassler Whitney, *A logical expansion in mathematics*, Bull. Amer. Math. Soc. **38** (1932), no. 8, 572–579, DOI 10.1090/S0002-9904-1932-05460-X. MR1562461

[100] Herbert S. Wilf, *Sieve equivalence in generalized partition theory*, J. Combin. Theory Ser. A **34** (1983), no. 1, 80–89, DOI 10.1016/0097-3165(83)90042-0. MR685214

[101] Herbert S. Wilf, *generatingfunctionology*, 2nd ed., Academic Press, Inc., Boston, MA, 1994. MR1277813

Index

Selected Published Titles in This Series

210 **Bruce E. Sagan,** Combinatorics: The Art of Counting, 2020
207 **Dmitry N. Kozlov,** Organized Collapse: An Introduction to Discrete Morse Theory, 2020
206 **Ben Andrews, Bennett Chow, Christine Guenther, and Mat Langford,** Extrinsic Geometric Flows, 2020
205 **Mikhail Shubin,** Invitation to Partial Differential Equations, 2020
204 **Sarah J. Witherspoon,** Hochschild Cohomology for Algebras, 2019
203 **Dimitris Koukoulopoulos,** The Distribution of Prime Numbers, 2019
202 **Michael E. Taylor,** Introduction to Complex Analysis, 2019
201 **Dan A. Lee,** Geometric Relativity, 2019
200 **Semyon Dyatlov and Maciej Zworski,** Mathematical Theory of Scattering Resonances, 2019
199 **Weinan E, Tiejun Li, and Eric Vanden-Eijnden,** Applied Stochastic Analysis, 2019
198 **Robert L. Benedetto,** Dynamics in One Non-Archimedean Variable, 2019
197 **Walter Craig,** A Course on Partial Differential Equations, 2018
196 **Martin Stynes and David Stynes,** Convection-Diffusion Problems, 2018
195 **Matthias Beck and Raman Sanyal,** Combinatorial Reciprocity Theorems, 2018
194 **Seth Sullivant,** Algebraic Statistics, 2018
193 **Martin Lorenz,** A Tour of Representation Theory, 2018
192 **Tai-Peng Tsai,** Lectures on Navier-Stokes Equations, 2018
191 **Theo Bühler and Dietmar A. Salamon,** Functional Analysis, 2018
190 **Xiang-dong Hou,** Lectures on Finite Fields, 2018
189 **I. Martin Isaacs,** Characters of Solvable Groups, 2018
188 **Steven Dale Cutkosky,** Introduction to Algebraic Geometry, 2018
187 **John Douglas Moore,** Introduction to Global Analysis, 2017
186 **Bjorn Poonen,** Rational Points on Varieties, 2017
185 **Douglas J. LaFountain and William W. Menasco,** Braid Foliations in Low-Dimensional Topology, 2017
184 **Harm Derksen and Jerzy Weyman,** An Introduction to Quiver Representations, 2017
183 **Timothy J. Ford,** Separable Algebras, 2017
182 **Guido Schneider and Hannes Uecker,** Nonlinear PDEs, 2017
181 **Giovanni Leoni,** A First Course in Sobolev Spaces, Second Edition, 2017
180 **Joseph J. Rotman,** Advanced Modern Algebra: Third Edition, Part 2, 2017
179 **Henri Cohen and Fredrik Strömberg,** Modular Forms, 2017
178 **Jeanne N. Clelland,** From Frenet to Cartan: The Method of Moving Frames, 2017
177 **Jacques Sauloy,** Differential Galois Theory through Riemann-Hilbert Correspondence, 2016
176 **Adam Clay and Dale Rolfsen,** Ordered Groups and Topology, 2016
175 **Thomas A. Ivey and Joseph M. Landsberg,** Cartan for Beginners: Differential Geometry via Moving Frames and Exterior Differential Systems, Second Edition, 2016
174 **Alexander Kirillov Jr.,** Quiver Representations and Quiver Varieties, 2016
173 **Lan Wen,** Differentiable Dynamical Systems, 2016
172 **Jinho Baik, Percy Deift, and Toufic Suidan,** Combinatorics and Random Matrix Theory, 2016
171 **Qing Han,** Nonlinear Elliptic Equations of the Second Order, 2016
170 **Donald Yau,** Colored Operads, 2016

For a complete list of titles in this series, visit the
AMS Bookstore at **www.ams.org/bookstore/gsmseries/**.