

# Essential Combinatorics

*A Working Course in Counting Methods*

Randall J. Swift

California State Polytechnic University, Pomona

# Preface

This manuscript grew from a ten-lecture course given in May 2000 at Western Kentucky University. In that course, I presented what counting methods and combinatorial theory I had found to be useful while studying and working in other areas of mathematics and its applications. Those original lecture notes evolved into this brief text and it provides a path through the broad and beautiful subject that is modern combinatorics.

The text is a self-contained working course in the principal methods of counting. It begins with the sum and product rules, permutations and combinations, stars and bars, and probability by counting, then develops combinatorial proof, the pigeonhole principle, Catalan and Stirling numbers, multinomial methods, inclusion–exclusion, elementary number theory, generating functions, and recurrence relations. The text assumes only basic algebra, knowledge of functions, finite sets, and elementary mathematical notation; no previous course in combinatorics or discrete mathematics is required. A few challenge and extended-exploration problems draw on calculus or other optional background, but they may be omitted without interrupting the logical development of the course.

This is intentionally a compact course in essential combinatorial methods rather than an encyclopedia of the subject. Graph theory, Ramsey theory, and the deeper theory of integer partitions are important areas, but a serious treatment of them would require a different and substantially longer book. Here the emphasis remains on methods that a mathematically literate reader is likely to need elsewhere: constructing a sample space, changing representations, exploiting symmetry, extracting coefficients, recognizing a recurrence, and deciding which of several familiar tools is appropriate to a new problem.

A guiding principle throughout the book is that a formula becomes much more useful when we understand the objects it counts and the decisions that produced it. For that reason, many identities are approached in more than one way. An algebraic derivation may establish the result efficiently, while a combinatorial proof explains why the two sides must agree by identifying a common family of objects. The aim is to make these changes of viewpoint a natural habit.

The manuscript is meant to be read with pencil in hand. A displayed formula usually appears at the end of a counting argument rather than at its beginning: first identify the outcomes, decide when two descriptions represent the same object, determine whether order or repetition matters, and only then choose a formula. Early examples make this modeling step explicit, and later chapters return to the same habit when the objects become paths, set partitions, arithmetic functions, probability distributions, or coefficients of a power series. The exercises are part of the exposition because many of them ask to discover a second representation or to push an idea beyond the formal development in the chapter.

A feature of this book is that several mathematical objects are introduced early and then deliberately carried forward. Rather than treating each chapter as an isolated collection of techniques, the text

returns to a familiar object whenever a new method reveals something that could not be seen as clearly before. Fixed points of permutations progress from elementary counting to derangements, bivariate exponential generating functions, factorial moments, and a Poisson limit. Pairings move from the double factorial to noncrossing Catalan structures and then to involutions. Occupancy problems connect balls and boxes with the pigeonhole principle, Stirling numbers, onto functions, and birthday asymptotics. Circular arrangements develop into questions about periodic words, Möbius inversion, stabilizers, and Burnside's lemma, while compositions evolve from separator arguments to generating functions, tilings, and Fibonacci-like recurrences.

Two longer case studies make the same philosophy visible across a still wider range of mathematics. Dice relabeling begins as a probability question and develops into generating-function factorization, cyclotomic structure, and Pythagorean questions. Set partitions lead from Stirling and Bell numbers to Poisson moments, factorial moments, probability generating functions, and a stochastic population model. These are not detachable side stories; they show how a counting idea can remain useful after the surrounding mathematics changes.

The labels *Continuing exploration* and *Returning exploration* mark these threads when they reappear. Their purpose is to reinforce the habit of combinatorial thinking: keep the underlying object in view, change the representation, and ask what the new representation makes easier to count or understand. Figure 1 summarizes the five principal flow explorations. A reader may proceed chapter by chapter, but may also follow one row of the figure forward and watch a single question acquire progressively stronger tools.

The intent of this text is to develop the skill of viewing a counting problem from different perspectives and to connect these perspectives to provide a solution to the problem.

*The Man in the Wilderness*

The old man of the forest said to me,  
“How many red strawberries grow in the sea?”  
I answered him as I thought good,  
“As many red herrings as swim in the wood.”

—an anonymous traditional English nursery rhyme

Randall J. Swift  
Pomona, CA

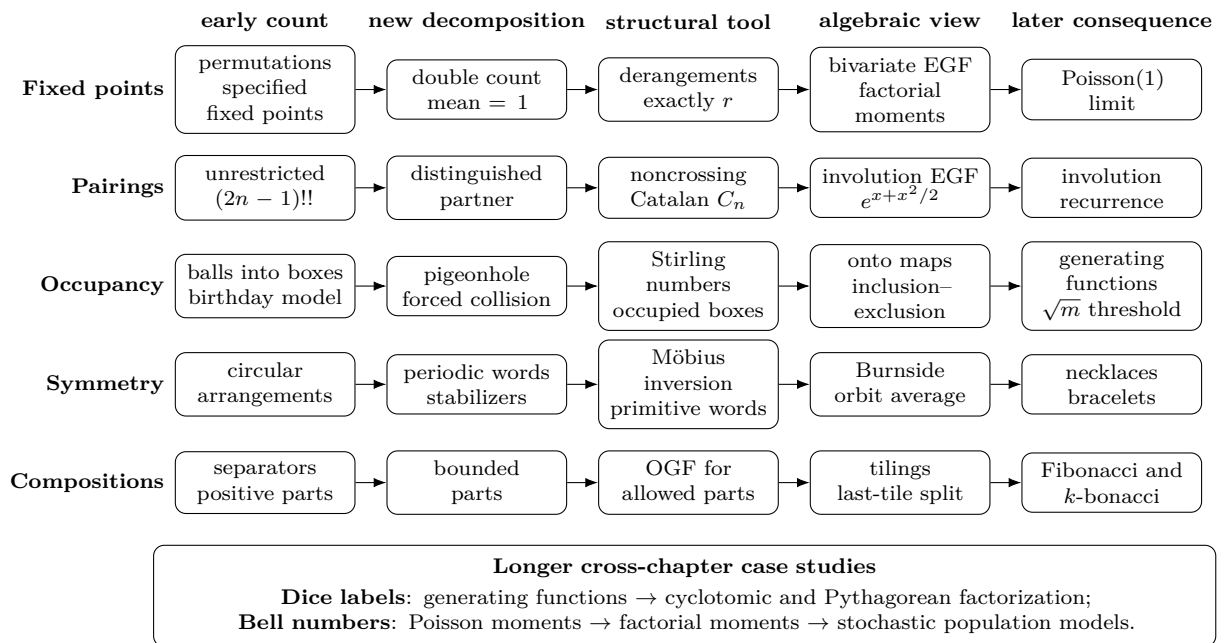


Figure 1: Five recurring flow explorations. The arrows indicate changes of representation rather than prerequisites: later chapters return to familiar objects with stronger tools.

# Contents

|                                                           |            |
|-----------------------------------------------------------|------------|
| <b>Preface</b>                                            | <b>iii</b> |
| <b>1 Fundamentals</b>                                     | <b>1</b>   |
| 1.1 The sum and product rules . . . . .                   | 1          |
| 1.2 Permutations and combinations . . . . .               | 3          |
| 1.2.1 Circular arrangements . . . . .                     | 5          |
| 1.3 Counting functions . . . . .                          | 6          |
| 1.4 The binomial theorem . . . . .                        | 7          |
| 1.5 Multinomial coefficients . . . . .                    | 9          |
| 1.6 Selections with repetition: stars and bars . . . . .  | 11         |
| 1.7 Probability by counting . . . . .                     | 12         |
| 1.8 Estimating large counts: Stirling's formula . . . . . | 14         |
| Exercises . . . . .                                       | 15         |
| <b>2 Combinatorial Proof and Elementary Distributions</b> | <b>18</b>  |
| 2.1 Combinatorial proof methods . . . . .                 | 18         |
| 2.1.1 Bijections and complementary counting . . . . .     | 18         |
| 2.1.2 Counting the same set twice . . . . .               | 19         |
| 2.2 Lattice paths . . . . .                               | 21         |
| 2.3 Sampling without replacement . . . . .                | 23         |
| 2.3.1 Capture–recapture . . . . .                         | 24         |
| 2.4 Sampling with replacement . . . . .                   | 25         |
| 2.5 Waiting for the first success . . . . .               | 27         |
| Exercises . . . . .                                       | 27         |
| <b>3 The Pigeonhole Principle and Latin Squares</b>       | <b>31</b>  |
| 3.1 Basic and general forms . . . . .                     | 31         |
| 3.2 Applications . . . . .                                | 33         |

---

|          |                                                        |           |
|----------|--------------------------------------------------------|-----------|
| 3.3      | Latin squares . . . . .                                | 34        |
|          | Exercises . . . . .                                    | 36        |
| <b>4</b> | <b>Catalan Numbers</b>                                 | <b>38</b> |
| 4.1      | The reflection principle . . . . .                     | 38        |
| 4.2      | Definition and first values . . . . .                  | 40        |
| 4.3      | The ballot problem . . . . .                           | 41        |
| 4.4      | Ticket queues . . . . .                                | 42        |
| 4.5      | Parenthesizations . . . . .                            | 43        |
| 4.6      | Triangulations of convex polygons . . . . .            | 44        |
| 4.7      | The Catalan recurrence . . . . .                       | 45        |
|          | Exercises . . . . .                                    | 46        |
| <b>5</b> | <b>Stirling Numbers</b>                                | <b>49</b> |
| 5.1      | Factorial polynomials . . . . .                        | 50        |
| 5.2      | Stirling numbers of the first kind . . . . .           | 51        |
| 5.3      | Stirling numbers of the second kind . . . . .          | 52        |
| 5.4      | Change of basis and inversion . . . . .                | 53        |
| 5.5      | Distributing objects . . . . .                         | 54        |
| 5.6      | Bell numbers . . . . .                                 | 56        |
|          | Exercises . . . . .                                    | 58        |
| <b>6</b> | <b>Multinomial Counting</b>                            | <b>63</b> |
| 6.1      | Permutations of multisets . . . . .                    | 63        |
| 6.2      | Ordered and unordered blocks . . . . .                 | 64        |
| 6.3      | The multinomial theorem . . . . .                      | 66        |
| 6.4      | Returning walks on the square lattice . . . . .        | 67        |
| 6.5      | The multinomial distribution . . . . .                 | 68        |
| 6.6      | The multivariate hypergeometric distribution . . . . . | 69        |
|          | Exercises . . . . .                                    | 70        |
| <b>7</b> | <b>Inclusion–Exclusion and Derangements</b>            | <b>73</b> |
| 7.1      | A distribution table . . . . .                         | 73        |
| 7.2      | The inclusion–exclusion principle . . . . .            | 75        |
| 7.3      | Bounded integer solutions . . . . .                    | 76        |
| 7.4      | Derangements . . . . .                                 | 78        |

|                                                                           |            |
|---------------------------------------------------------------------------|------------|
| Exercises . . . . .                                                       | 79         |
| <b>8 Combinatorial Number Theory</b>                                      | <b>82</b>  |
| 8.1 Counting integers with forbidden divisors . . . . .                   | 82         |
| 8.2 A sieve formula for primes . . . . .                                  | 83         |
| 8.3 Euler's totient function . . . . .                                    | 84         |
| 8.4 The Möbius function and inversion . . . . .                           | 86         |
| 8.5 Circular words and necklaces . . . . .                                | 87         |
| 8.6 Burnside's lemma: counting up to symmetry . . . . .                   | 89         |
| Exercises . . . . .                                                       | 89         |
| <b>9 Generating Functions</b>                                             | <b>93</b>  |
| 9.1 Ordinary, exponential, and probability generating functions . . . . . | 93         |
| 9.2 Bell numbers, Poisson moments, and Touchard polynomials . . . . .     | 96         |
| 9.3 Factorial moments and probability generating functions . . . . .      | 98         |
| 9.4 Polynomial-rate population processes . . . . .                        | 98         |
| 9.5 Basic operations . . . . .                                            | 100        |
| 9.6 Selection problems . . . . .                                          | 102        |
| 9.7 Relabeling dice . . . . .                                             | 104        |
| 9.8 Uniqueness of base representation . . . . .                           | 106        |
| Exercises . . . . .                                                       | 107        |
| <b>10 Recurrence Relations</b>                                            | <b>116</b> |
| 10.1 First examples . . . . .                                             | 116        |
| 10.2 Linear recurrences with constant coefficients . . . . .              | 118        |
| 10.3 Nonhomogeneous recurrences . . . . .                                 | 119        |
| 10.4 Solving recurrences with generating functions . . . . .              | 120        |
| 10.5 The Fibonacci sequence . . . . .                                     | 121        |
| 10.6 A model for annual plants . . . . .                                  | 123        |
| Exercises . . . . .                                                       | 125        |
| <b>Extended Problems and Explorations</b>                                 | <b>131</b> |
| <b>Answers and Hints</b>                                                  | <b>137</b> |
| Chapter 1: Fundamentals . . . . .                                         | 137        |
| Chapter 2: Combinatorial Proof and Elementary Distributions . . . . .     | 141        |
| Chapter 3: The Pigeonhole Principle and Latin Squares . . . . .           | 145        |

---

|                                                           |            |
|-----------------------------------------------------------|------------|
| Chapter 4: Catalan Numbers . . . . .                      | 147        |
| Chapter 5: Stirling Numbers . . . . .                     | 150        |
| Chapter 6: Multinomial Counting . . . . .                 | 155        |
| Chapter 7: Inclusion–Exclusion and Derangements . . . . . | 158        |
| Chapter 8: Combinatorial Number Theory . . . . .          | 161        |
| Chapter 9: Generating Functions . . . . .                 | 164        |
| Chapter 10: Recurrence Relations . . . . .                | 174        |
| Extended Problems and Explorations . . . . .              | 180        |
| <b>Closing Remark</b>                                     | <b>185</b> |
| <b>References</b>                                         | <b>187</b> |
| <b>Subject Index</b>                                      | <b>189</b> |

## Chapter 2

# Combinatorial Proof and Elementary Distributions

Chapter 1 developed formulas for constructing and counting finite sets; we now use those formulas as the raw material for proofs. A numerical identity can always be checked algebraically once the correct expression is known, but a combinatorial proof explains where the expression comes from by describing one family of objects and organizing it in two different ways. Because the underlying objects do not change, the two counts must agree, and the identity becomes a statement about structure rather than symbolic simplification.

The same viewpoint carries naturally into probability. A finite probability distribution is obtained by counting all admissible samples and then counting those with a prescribed statistic, so the crucial step remains the construction of the sample space. Sampling with and without replacement lead to different spaces and different distributions; recognizing that distinction prepares the way for Chapter 3, where counting gives way to a different question about what structure must occur once a configuration becomes large enough.

### 2.1 Combinatorial proof methods

Two of the most useful proof patterns are to construct an explicit correspondence and to count one family in two different ways. The first gives a bijection; the second gives a double count. Both methods replace symbolic manipulation by a description of the objects being counted.

#### 2.1.1 Bijections and complementary counting

A bijection is a one-to-one correspondence between two finite sets. If every object in one class can be paired with exactly one object in another class, then the two classes have the same size. This is often the cleanest kind of combinatorial proof because it does not calculate either size directly; it changes the representation until the answer is familiar.

Complementary counting is a closely related change of viewpoint. Instead of constructing the desired objects, count all objects and subtract those that violate the condition. The method is especially effective when “at least one” is awkward but “none” has a simple product form.

**Example 2.1** (Continuing example: subsets, words, and paths). A subset  $S \subseteq \{1, \dots, n\}$  corresponds to a binary word  $b_1 \cdots b_n$  by setting  $b_i = 1$  exactly when  $i \in S$ . The same binary word corresponds to a lattice path with an east step for 0 and a north step for 1. Thus

$$\{\text{subsets of } [n]\} \longleftrightarrow \{\text{binary words of length } n\} \longleftrightarrow \{n\text{-step E/N paths}\}.$$

Restricting to subsets of size  $r$  gives words with  $r$  ones and paths with  $r$  north steps, all counted by  $\binom{n}{r}$ .

**Example 2.2** (Complementary counting). How many length-8 decimal strings contain at least one zero? There are  $10^8$  strings in all and  $9^8$  containing no zero. Hence the answer is

$$10^8 - 9^8.$$

Trying to count strings with exactly one zero, exactly two zeros, and so on would be correct but needlessly elaborate.

Bijections compare two classes by changing the objects' representation, whereas double counting keeps one class fixed and reorganizes the order in which its features are chosen. The two methods share the same goal: expose the structure of the objects first, and let the resulting algebra record that structure afterward.

### 2.1.2 Counting the same set twice

Double counting begins by identifying one set of objects and then organizing it in two ways. One count may choose all parts at once, while another count first selects a distinguished element or an intermediate structure. Because the underlying objects do not change, the two expressions must agree.

Pascal's identity is the basic example. A fixed element divides all  $r$ -subsets into those that contain it and those that do not. Newton's identity adds an intermediate selection: one may choose a large committee and then a subcommittee, or choose the subcommittee first and complete the larger group. The proof records the order of decisions rather than manipulating factorials.

**Example 2.3.** Suppose a group contains  $m$  mathematics majors and  $n$  physical-education majors. The number of mixed pairs is  $mn$ . It is also the number of all pairs minus the same-major pairs, so

$$\binom{m+n}{2} - \binom{m}{2} - \binom{n}{2} = mn.$$

**Theorem 2.4** (Pascal's identity). For  $1 \leq r \leq n$ ,

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

*Proof.* Fix an element  $a$  of an  $n$ -element set. An  $r$ -subset either omits  $a$ , giving  $\binom{n-1}{r}$  choices, or contains  $a$ , in which case the remaining  $r-1$  elements can be chosen in  $\binom{n-1}{r-1}$  ways.  $\square$

Iterating Pascal's identity gives the hockey-stick identity

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

**Theorem 2.5** (Vandermonde's identity). *For nonnegative integers  $m, n, r$ ,*

$$\sum_k \binom{m}{k} \binom{n}{r-k} = \binom{m+n}{r},$$

*where impossible binomial coefficients are interpreted as zero.*

*Proof.* Choose an  $r$ -element committee from two disjoint groups of sizes  $m$  and  $n$ . If exactly  $k$  members come from the first group, the choice can be made in  $\binom{m}{k} \binom{n}{r-k}$  ways; summing over all admissible values of  $k$  gives the identity.  $\square$

A useful nested-selection identity is

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

Both sides count a pair  $(K, R)$  of subsets with  $K \subseteq R$ ,  $|K| = k$ , and  $|R| = r$ .

**Example 2.6** (Worked example: subsets with two distinguished elements). We prove the identity

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

by counting pairs consisting of a subset  $S \subseteq \{1, \dots, n\}$  and an ordered pair  $(a, b)$  of distinct distinguished elements of  $S$ .

If  $|S| = k$ , there are  $\binom{n}{k}$  choices for  $S$  and  $k(k-1)$  choices for  $(a, b)$ , which gives the left-hand side after summing over  $k$ . Alternatively, choose the ordered pair first in  $n(n-1)$  ways. Each of the other  $n-2$  elements may then be included in or excluded from  $S$  independently, giving  $2^{n-2}$  possibilities. This gives the right-hand side.

The identity is therefore not an accidental algebraic simplification: both expressions enumerate exactly the same marked subsets.

Double counting is not restricted to committees and subsets. Lattice paths provide a geometric model in which binomial coefficients become visible as routes through a grid.

Marked objects are a particularly useful source of double counts. One count groups objects by their total size; another chooses the mark first. This simple pattern produces many weighted binomial identities.

**Example 2.7** (Continuing example: counting all 1's in binary words). Consider all  $2^n$  binary words of length  $n$  and mark one occurrence of 1. If a word contains  $k$  ones, it contributes  $k$  marked words, so the total is

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

Alternatively choose the marked position first in  $n$  ways and fill the remaining  $n-1$  positions arbitrarily, giving  $n2^{n-1}$ . Hence

$$\sum_{k=0}^n k \binom{n}{k} = n2^{n-1}.$$

The identity is now a statement about the same family of marked words viewed from two directions.

**Example 2.8** (Continuing exploration: fixed points on average). Consider pairs  $(\pi, i)$  in which  $\pi$  is a permutation of  $[n]$  and  $i$  is a fixed point of  $\pi$ . Count the pairs by choosing  $i$  first. There are  $n$  choices for  $i$ , and then  $(n-1)!$  permutations fixing it, for a total of

$$n(n-1)! = n!.$$

There are also  $n!$  permutations altogether. Therefore the average number of fixed points over all permutations is exactly

$$\frac{n!}{n!} = 1.$$

The value 1 will later reappear in a less obvious form: the number of fixed points in a large random permutation approaches a Poisson random variable with parameter 1.

**Example 2.9** (Continuing exploration: pairings by a distinguished partner). Let  $M_{2n}$  be the number of pairings of  $2n$  labeled objects. Object 1 has  $2n-1$  possible partners. Once that pair is chosen, the remaining  $2n-2$  objects may be paired independently. Hence

$$M_{2n} = (2n-1)M_{2n-2}, \quad M_0 = 1.$$

Iterating gives  $M_{2n} = (2n-1)!!$ , recovering the Chapter 1 formula by a decomposition rather than by division by symmetry. The same “distinguished component” idea will later produce Catalan and involution recurrences.

**Example 2.10** (Continuing exploration: the average number of collision pairs). Place  $n$  labeled balls independently and uniformly into  $m$  labeled boxes. For each unordered pair of balls, the probability that the two land in the same box is  $1/m$ . Summing over the  $\binom{n}{2}$  pairs shows that the average number  $C$  of colliding pairs is

$$\mathbb{E}[C] = \frac{\binom{n}{2}}{m}.$$

This estimate already predicts the birthday scale: collisions should become common when  $\binom{n}{2}/m$  is of order one, that is, when  $n$  is of order  $\sqrt{m}$ . The end-of-book exploration will make this heuristic quantitative.

**Linearity of expectation.** The preceding calculation uses one elementary probabilistic fact that is useful in counting arguments. If  $X = X_1 + \cdots + X_r$ , then

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_r],$$

whether or not the variables are independent. For an indicator  $I_A$ ,  $\mathbb{E}[I_A] = \mathbb{P}(A)$ . Thus the collision count is the sum of  $\binom{n}{2}$  indicators, each having mean  $1/m$ . This short device will reappear in exercises involving random subsets and occupancy.

## 2.2 Lattice paths

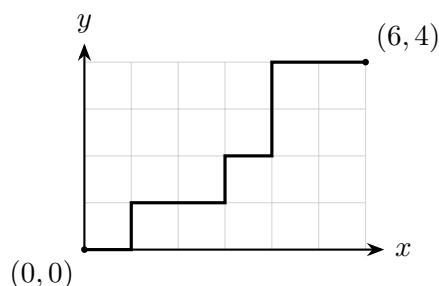
A monotone lattice path can be encoded as a word. Write  $E$  for an east step and  $N$  for a north step. A path from  $(0,0)$  to  $(a,b)$  is then a word containing  $a$  copies of  $E$  and  $b$  copies of  $N$ .

Counting paths is therefore the same as choosing the positions of one kind of step among all  $a + b$  positions.

This encoding is valuable because geometric restrictions become restrictions on prefixes of the word. Passing through a point splits a path into two independent pieces, while staying below a diagonal imposes an inequality on every initial segment. The unconstrained count in this section will be refined by the reflection principle in the Catalan chapter.

A path from  $(0, 0)$  to  $(a, b)$  using only unit east and north steps has  $a + b$  steps, of which  $b$  are north. Therefore

$$\#\text{paths} = \binom{a+b}{b} = \binom{a+b}{a}.$$



This interpretation turns Pascal's identity into the observation that the last step into  $(a, b)$  comes either from  $(a - 1, b)$  or from  $(a, b - 1)$ .

**Example 2.11** (Worked example: forcing a path through a point). How many monotone lattice paths from  $(0, 0)$  to  $(8, 6)$  pass through  $(3, 2)$ ? Passing through the specified point divides the path into two independent pieces.

From  $(0, 0)$  to  $(3, 2)$  a path uses three east steps and two north steps, so there are

$$\binom{5}{2} = 10$$

possibilities. From  $(3, 2)$  to  $(8, 6)$  it uses five east steps and four north steps, giving

$$\binom{9}{4} = 126$$

possibilities. By the product rule, the number of complete paths is

$$\binom{5}{2} \binom{9}{4} = 10 \cdot 126 = 1260.$$

This factorization is a general technique: a condition that a path pass through a fixed lattice point converts one global path into two ordinary path-counting problems.

The same binomial coefficients that count paths also count samples. We now use them to distinguish two basic probability models, beginning with sampling without replacement.

The word encoding also makes local path restrictions easy to translate. Specifying an initial or terminal step deletes one symbol; specifying an intermediate point breaks the word into two independent blocks. More global prefix restrictions require the reflection principle developed in Chapter 4.

**Example 2.12** (Continuing example: first-step restriction). A monotone path from  $(0, 0)$  to  $(5, 5)$  has five east and five north steps. If the first step must be east, the remaining nine positions contain four east steps and five north steps, so there are

$$\binom{9}{4} = 126$$

paths. In binary-word language, we fixed the first symbol and chose the positions of the remaining four symbols of the same type.

## 2.3 Sampling without replacement

Without replacement, the selected objects form a subset rather than an independent sequence of trials. Once an object has been drawn, the composition of the remaining population changes. The hypergeometric distribution accounts for this dependence by counting samples directly: choose the required number from the first type, choose the remainder from the second type, and divide by the number of all samples of the same size.

The allowable range of  $x$  is part of the model. We cannot select more objects of a type than the population contains, and we may be forced to select some objects of that type if the other type is too small to fill the sample. Writing these bounds prevents formally meaningful but combinatorially impossible terms.

**Definition 2.13.** A random variable  $X$  has a *hypergeometric distribution* with parameters  $N, M, n$  if a sample of size  $n$  is drawn without replacement from a population containing  $N$  objects of one type and  $M$  of another, and  $X$  counts the first type.

Then

$$\mathbb{P}(X = x) = \frac{\binom{N}{x} \binom{M}{n-x}}{\binom{N+M}{n}},$$

for

$$\max(0, n - M) \leq x \leq \min(N, n).$$

**Example 2.14.** A box contains ten red and five blue marbles. If six are selected without replacement, the probability of four red marbles is

$$\frac{\binom{10}{4} \binom{5}{2}}{\binom{15}{6}} \approx 0.41958.$$

**Example 2.15** (Worked example: inspecting a finite shipment). A shipment contains 20 components, of which 5 are defective. Six components are selected uniformly without replacement. What is the probability that exactly two of the selected components are defective?

A sample is an unordered six-element subset of the shipment, so there are  $\binom{20}{6}$  equally likely samples. A favorable sample contains two of the five defective components and four of the fifteen good components. Hence

$$\mathbb{P}(X = 2) = \frac{\binom{5}{2}\binom{15}{4}}{\binom{20}{6}} \approx 0.3522.$$

The factors in the numerator describe the composition of the sample directly. No independence assumption is made: after each physical draw the population changes, but counting complete subsets avoids having to track those changing conditional probabilities one draw at a time.

The hypergeometric model is especially useful when the sampled population is finite and sampling changes the remaining composition. Capture–recapture turns that observation into a method for estimating an unknown population size.

For small samples it is useful to compare the hypergeometric count with a sequential calculation. The two methods agree, but the subset calculation usually exposes the symmetry more clearly and avoids a long chain of changing conditional probabilities.

**Example 2.16** (Continuing example: selecting a codebook committee). A collection contains 12 devices of type A and 8 of type B. Five are selected without replacement. The probability that the sample contains exactly three type-A devices is

$$\frac{\binom{12}{3}\binom{8}{2}}{\binom{20}{5}}.$$

The numerator classifies the selected subset by content, exactly as a multinomial word is classified by symbol counts.

### 2.3.1 Capture–recapture

Capture–recapture turns a counting distribution into an estimation method. The marked animals play the role of one type in a finite population, and the later sample is drawn without replacement. If marked and unmarked animals have mixed thoroughly and are equally likely to be caught, the proportion of marked animals in the sample should resemble the marked proportion in the population.

The estimator obtained by equating these proportions is intuitive, but its reliability depends on the assumptions: marks must remain visible, marking must not change survival or catchability, and the population should remain effectively closed between samples. The hypergeometric likelihood makes the probabilistic content explicit and provides a route from the heuristic ratio to a maximum-likelihood estimate.

Suppose a population has unknown size  $K$ . Mark and release  $M$  individuals, then later draw a sample of size  $n$  and observe  $X$  marked individuals. Equating the marked proportions suggests the estimator

$$\widehat{K} = \frac{Mn}{X},$$

provided  $X > 0$ . Conditional on  $K$ , the count  $X$  is hypergeometric:

$$\mathbb{P}_K(X = x) = \frac{\binom{M}{x}\binom{K-M}{n-x}}{\binom{K}{n}}.$$

The ratio argument  $\mathbb{P}_K(X = x)/\mathbb{P}_{K-1}(X = x)$  leads to a maximum-likelihood estimate close to  $Mn/x$ , with integer and boundary adjustments as necessary.

**Example 2.17** (Worked example: estimating a population). Suppose 40 animals are captured, marked, and released. Later a sample of 30 animals contains 6 marked individuals. The observed marked proportion is  $6/30 = 0.2$ , so the proportion argument gives

$$\frac{40}{K} \approx \frac{6}{30}, \quad \widehat{K} = \frac{40 \cdot 30}{6} = 200.$$

If the true population size were  $K = 200$ , the probability of observing exactly six marked animals in the second sample would be

$$\frac{\binom{40}{6} \binom{160}{24}}{\binom{200}{30}} \approx 0.1946.$$

The numerical estimate is meaningful only under the modeling assumptions stated above. In particular, the calculation treats all 200 animals as equally catchable and assumes that the marked population has mixed back into the full population before the second sample is taken.

Replacing each sampled object restores the same population before the next draw. This small change in the experiment replaces hypergeometric dependence with the simpler binomial model.

The ratio estimator is best viewed as a first approximation rather than a magical formula. The combinatorial model underneath it is the same without-replacement sampling model, and the assumptions determine whether the marked sample is representative.

**Example 2.18** (A second capture–recapture calculation). Suppose 60 animals are marked and, after mixing, a sample of 50 contains 10 marked animals. The proportion estimate is

$$\widehat{K} = \frac{60 \cdot 50}{10} = 300.$$

If  $K = 300$ , the probability of exactly ten recaptures is

$$\frac{\binom{60}{10} \binom{240}{40}}{\binom{300}{50}}.$$

The first expression estimates the unknown population; the second evaluates how compatible the observed recapture count is with one proposed population size.

## 2.4 Sampling with replacement

With replacement, every trial begins with the same population composition. The success probability therefore remains constant, and the trials may be modeled as independent Bernoulli experiments. A particular sequence with  $x$  successes and  $n - x$  failures has probability  $p^x(1 - p)^{n-x}$ ; the binomial coefficient counts how many such sequences have the same total number of successes.

The distribution ignores the order in the final statistic but counts order while deriving its probability. This two-level viewpoint is common in combinatorics: first count detailed outcomes that are easy to assign probabilities to, then group them according to the statistic of interest.

If  $n$  independent trials each succeed with probability  $p$ , then the number  $X$  of successes has the binomial distribution

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \dots, n.$$

The probabilities sum to one by the binomial theorem.

**Example 2.19.** For the same box of ten red and five blue marbles, sampling six times with replacement gives

$$\mathbb{P}(X = 4) = \binom{6}{4} \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^2 = \frac{80}{243} \approx 0.32922.$$

This differs from the without-replacement probability because the composition of the box changes after each unreturned draw.

A hypergeometric distribution is well approximated by a binomial distribution when the population is large relative to the sample. A common practical rule is that the sampling fraction  $n/(N + M)$  should be at most about 0.05; the condition concerns the sampling fraction, not the success probability.

**Example 2.20** (Worked example: a binomial tail). A manufacturing test is passed independently with probability 0.3. Eight items are tested. What is the probability that at least three pass?

It is shorter to count the complementary event, in which zero, one, or two items pass. If  $X$  is the number of passing items, then  $X$  is binomial with parameters  $n = 8$  and  $p = 0.3$ , so

$$\begin{aligned} \mathbb{P}(X \geq 3) &= 1 - \sum_{k=0}^2 \binom{8}{k} (0.3)^k (0.7)^{8-k} \\ &\approx 0.4482. \end{aligned}$$

The complement avoids adding six separate binomial probabilities. The model is binomial because every test has the same success probability and the outcomes are assumed independent.

The binomial distribution fixes the number of trials and counts successes. A complementary question fixes the first success and allows the number of trials to vary, leading to the geometric distribution.

The binomial distribution is also a counting formula for words over a two-symbol alphabet whose symbols have unequal weights. The binomial coefficient counts the locations of the successes, while the powers of  $p$  and  $1 - p$  supply the probability weight of each word.

**Example 2.21** (Continuing example: symbol counts). A communication channel transmits a 1 with probability 0.4 and a 0 with probability 0.6, independently from position to position. The probability that a length-12 transmission contains exactly five 1's is

$$\binom{12}{5} (0.4)^5 (0.6)^7.$$

The combinatorial factor is the same count of fixed-weight binary words seen in Chapter 1; the probabilistic model attaches a common weight to every word of that content.

## 2.5 Waiting for the first success

The geometric distribution records a different statistic of repeated trials. Instead of fixing the number of trials and counting successes, we continue until the first success and record its position. The event  $X = x$  has a rigid form:  $x - 1$  failures followed by one success. No binomial coefficient is needed because there is only one possible order.

For the geometric distribution, tail probabilities have an especially simple form and often shorten a calculation. The event  $X > r$  means that the first  $r$  trials all fail, so it has probability  $(1 - p)^r$ . This observation makes many waiting-time questions easier to answer through complements than by summing individual probabilities.

If independent trials succeed with probability  $p$  and  $X$  is the trial number of the first success, then

$$\mathbb{P}(X = x) = (1 - p)^{x-1}p, \quad x = 1, 2, \dots$$

This is the geometric distribution, and its probabilities sum to one because

$$\sum_{x \geq 1} (1 - p)^{x-1}p = 1.$$

**Example 2.22** (Worked example: success in a specified window). Independent trials succeed with probability  $p = 0.2$ , and  $X$  is the trial number of the first success. Find the probability that the first success occurs after trial 5 but no later than trial 9.

The event is  $6 \leq X \leq 9$ . One approach is to sum four geometric probabilities. A shorter approach uses tails:

$$\mathbb{P}(X > r) = (0.8)^r.$$

Therefore

$$\mathbb{P}(6 \leq X \leq 9) = \mathbb{P}(X > 5) - \mathbb{P}(X > 9) = (0.8)^5 - (0.8)^9 \approx 0.19346.$$

The subtraction works because the event  $X > 9$  is contained in  $X > 5$ ; removing it leaves precisely those outcomes whose first success occurs on trials 6, 7, 8, or 9.

The geometric model is a useful reminder that not every repeated-trial problem requires a binomial coefficient. When the stopping rule determines a unique pattern of failures and success, there is no choice of locations left to count.

**Example 2.23** (Continuing example: waiting for a symbol). Suppose independent symbols are drawn uniformly from an alphabet of size 5. The probability that the first occurrence of a specified symbol is on draw  $r$  is

$$\left(\frac{4}{5}\right)^{r-1} \frac{1}{5}.$$

For example, the probability that it first appears on the sixth draw is  $(4/5)^5(1/5)$ . The word before the first occurrence is forced to avoid the distinguished symbol, and the final position is forced to contain it.

## Exercises

These exercises continue the chapter's emphasis on representation. The early problems practice bijections, double counting, lattice paths, and the standard finite distributions, while the later

problems ask the reader to decide which sampling model is appropriate or to discover a second proof. Several problems also begin threads that will return in later chapters. Answers, substantial hints, or solutions are provided in the back matter.

## Elementary exercises

**Exercise 2.1.** How many shortest routes lead six blocks east and four blocks north?

**Exercise 2.2.** Give algebraic and combinatorial proofs that

$$\sum_{r=0}^n \binom{n}{r} = 2^n.$$

**Exercise 2.3.** Use a combinatorial argument to prove

$$\binom{2n}{2} = 2\binom{n}{2} + n^2.$$

**Exercise 2.4** (The hockey-stick identity). Give a combinatorial proof that, for  $0 \leq r \leq n$ ,

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

**Exercise 2.5** (Vandermonde's identity). Count  $r$ -element subsets of two disjoint sets, of sizes  $m$  and  $n$ , to prove

$$\sum_j \binom{m}{j} \binom{n}{r-j} = \binom{m+n}{r}.$$

State explicitly the range of  $j$  over which the summands are nonzero.

## Developing technique

**Exercise 2.6.** An urn contains twelve red balls and eight blue balls. Five balls are drawn without replacement. Find the probability that exactly three are red.

**Exercise 2.7.** A ten-question multiple-choice examination has four choices for each question. A student guesses independently on every question. Find the probability of (a) exactly six correct answers and (b) at least one correct answer.

**Exercise 2.8** (Quality control: spoiled chicken). A quality-control inspector examines 10 packages from each shipment of 1000 packages of frozen chicken for spoilage. The shipment is approved if and only if all 10 packages pass inspection. Find the probability that a shipment is approved even though 10% of its packages are badly spoiled. This problem is adapted from [27].

**Exercise 2.9.** Independent trials succeed with probability 0.3. Find the probability that the first success occurs (a) on the fourth trial, (b) within the first four trials, and (c) on an odd-numbered trial.

**Exercise 2.10.** A shipment contains 42 good devices and 8 defective devices. Five are selected without replacement. Find the probability that exactly three are good and the probability that at most three are good.

**Exercise 2.11.** A population contains 400 objects of one type and 600 of another. For a sample of size 10, compare the exact hypergeometric probability of obtaining exactly five objects of the first type with the corresponding binomial approximation. Explain why the two values are close.

### Challenge exercises

**Exercise 2.12.** Show that if  $p$  is prime, then  $p$  divides  $\binom{p}{r}$  for  $1 \leq r \leq p-1$ .

**Exercise 2.13** (Capture–recapture). A population of unknown size  $K$  contains  $M$  marked individuals. A later sample of size  $n$  contains exactly  $x$  marked individuals. Write the hypergeometric likelihood as a function of  $K$ , compare consecutive likelihood values, and determine the integer value or values of  $K$  that maximize it. Compare the result with the heuristic estimate  $Mn/x$ .

**Exercise 2.14** (Pepys’s problem). Compare the probabilities of at least one six in six rolls, at least two sixes in twelve rolls, and at least three sixes in eighteen rolls.

**Exercise 2.15** (Banach’s matchbox problem). A mathematician starts with  $n$  matches in each of two boxes and chooses a box at random whenever a match is needed. Find the probability that, when an empty box is first selected, the other box contains exactly  $r$  matches.

**Exercise 2.16.** Use Vandermonde’s identity to verify directly that the hypergeometric probabilities sum to one:

$$\sum_x \frac{\binom{N}{x} \binom{M}{n-x}}{\binom{N+M}{n}} = 1.$$

Give the correct range of summation.

### Further exercise on bijections

**Exercise 2.17.** Give an explicit bijection between  $r$ -element subsets of  $\{1, \dots, n\}$  and monotone lattice paths from  $(0, 0)$  to  $(n-r, r)$ . Use the bijection to explain  $\binom{n}{r} = \binom{n}{n-r}$  geometrically.

### Additional exercises: bijections, double counts, and sampling

The following problems are arranged from direct applications to short explorations.

**Exercise 2.18.** Give an explicit bijection between compositions of  $n$  into  $k$  positive parts and  $(k-1)$ -element subsets of  $\{1, 2, \dots, n-1\}$ . Recover the formula  $\binom{n-1}{k-1}$ .

**Exercise 2.19.** Prove by double counting that

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

Interpret the two terms obtained from  $k^2 = k(k-1) + k$ .

**Exercise 2.20.** How many monotone lattice paths from  $(0, 0)$  to  $(7, 5)$  avoid the point  $(3, 2)$ ?

**Exercise 2.21** (Waiting without replacement). An urn contains five red and seven blue balls. Balls are drawn without replacement until the first red ball appears. Find the probability that the first red ball is drawn fourth.

**Exercise 2.22.** A population contains  $N$  objects of one type and  $M$  of another. A sample of size  $n$  is drawn without replacement, and  $X$  counts the first type. Introduce indicator variables to prove

$$\mathbb{E}[X] = n \frac{N}{N + M}.$$

**Exercise 2.23.** Use Vandermonde's identity and the symmetry of binomial coefficients to prove

$$\sum_k \binom{m}{k} \binom{n}{k} = \binom{m+n}{m}.$$

Give a selection interpretation.

**Exercise 2.24** (The handshake lemma). A finite graph consists of vertices joined by edges. The degree of a vertex is the number of incident edges. Double-count vertex–edge incidences to prove that the sum of all vertex degrees is twice the number of edges. Deduce that every finite graph has an even number of vertices of odd degree.

**Exercise 2.25.** A committee of size  $r$  is chosen uniformly from  $n$  people. Find the probability that two specified people are both chosen, and simplify the result to

$$\frac{r(r-1)}{n(n-1)}.$$

**Exercise 2.26.** Prove combinatorially that

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

Describe the object being counted on both sides.

**Exercise 2.27** (Waiting for the  $r$ th success). Independent Bernoulli trials succeed with probability  $p$ . Let  $T$  be the trial on which the  $r$ th success occurs. Derive

$$\mathbb{P}(T = t) = \binom{t-1}{r-1} p^r (1-p)^{t-r}, \quad t = r, r+1, \dots$$

This distribution is called the negative binomial distribution.

**Exercise 2.28** (Exploration: Prüfer codes and Cayley's formula). A labeled tree has vertex set  $[n] = \{1, \dots, n\}$ . Repeatedly delete the smallest labeled leaf and record its unique neighbor; stop when two vertices remain.

- (a) Show that this produces a sequence of length  $n-2$  with entries in  $[n]$ .
- (b) Describe a reconstruction algorithm that recovers the tree from the sequence.
- (c) Deduce Cayley's formula: the number of labeled trees on  $n$  vertices is  $n^{n-2}$ .

This exercise introduces a classical graph-theoretic idea without requiring a separate chapter on graph theory.