MathCastle
Counting & Probability

Complete course booklet • 11 topics

Contents

1 Core Ideas in Plain Terms

Counting and chance, in everyday language

1.1 The Counting Principle

Multiply independent choices

When you make several choices that don't affect each other, **multiply** the number of options. When choices are alternatives (this or that), **add** them. This is the **Fundamental Counting Principle**, and it underlies almost all counting.

(diagram in the online version)

Outfits

With 3 shirts and 4 pairs of pants, the Fundamental Counting Principle gives $3 \times 4 = 12$ outfits.

(diagram in the online version)

1.2 Arrangements vs. Selections

Does order matter?

If rearranging counts as different (a race finish, a seating), use **permutations**: ${}_nP_k = \frac{n!}{(n-k)!}$. If order does not matter (a committee, a handful of cards), use **combinations**: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. The only difference is dividing out the rearrangements.

Choose vs. arrange

Arranging 3 of 8 people in a row: ${}_8P_3 = 336$. Choosing a committee of 3: $\binom{8}{3} = 56$.

1.3 Probability

Favorable over total

For equally likely outcomes, the probability of an event is given by the classical probability formula $P(A) = \dfrac{\text{favorable outcomes}}{\text{total outcomes}}$. For an "at least one" question, it is usually easier to find the probability it *doesn't* happen and subtract from 1.

At least one head

In 3 coin flips, the chance of no heads is $\left(\frac{1}{2}\right)^3 = \frac{1}{8}$, so at least one head has probability $1-\frac{1}{8} = \frac{7}{8}$.

1.4 Going Deeper: Complements and Expected Value

The complement shortcut

The **complement rule** says $P(A) = 1 - P(\text{not } A)$. That makes "at least one" problems almost always easier backwards: $P(\text{at least one}) = 1 - P(\text{none})$. Roll a die 4 times — $P(\text{at least one six}) = 1 - \left(\frac{5}{6}\right)^4 = 1 - \frac{625}{1296} = \frac{671}{1296}$, one subtraction instead of four cases.

Expected value is the long-run average

The **expected value** formula is $E = \sum \left(\text{outcome} \times \text{probability}\right)$: multiply each outcome by its probability and add. A fair die averages $\dfrac{1+2+3+4+5+6}{6} = 3.5$ per roll — never an actual roll, but exactly what the average approaches over many rolls.

Pigeonhole in one line

Any 13 people include two born in the same month: 13 pigeons, 12 holes — some hole gets two. No probability needed, just counting.

A committee count (multi-step)

From 5 teachers and 8 students, choose a committee of 2 teachers and 3 students. Step 1 — teachers: $\binom{5}{2} = 10$. Step 2 — students: $\binom{8}{3} = 56$. Step 3 — the choices are independent, so by the **Fundamental Counting Principle**: $10 \times 56 = 560$ committees.

1.5 Problem-Solving Playbook

Direct, cases, or complement?

Three ways to count, in order of preference: **directly** (multiply choices), by **cases** (split cleanly, add), or by **complement** (count what you don't want, subtract). If identical arrangements got counted more than once, divide out the repeats.

Worked: complement counting

Choose 4 of 6 people for a committee, but two of them refuse to serve together. Total: $\binom{6}{4} = 15$. Committees containing BOTH: the other 2 seats from the remaining 4 people, $\binom{4}{2} = 6$. By the **complement rule**: $15 - 6 = 9$ committees.

2 Fundamental Counting: Permutations & Combinations

Competition counting & probability, Unit 1

2.1 Addition & Multiplication Principles

The two rules that power all of counting

Multiplication Principle (AND). If a task splits into a sequence of independent stages, and stage i can be done in n_i ways *regardless* of the earlier choices, then the whole task can be done in

$$n_1 \cdot n_2 \cdot n_3 \cdots n_k$$

ways. Use it when you make choice 1 *and* choice 2 *and* ...

Addition Principle (OR). If the outcomes split into *disjoint* cases (no outcome lands in two cases at once), add the counts of the cases:

$$|A_1| + |A_2| + \cdots + |A_k|.$$

Use it when an outcome comes from case 1 *or* case 2 *or* ...

Worked example: license plates

A plate has 2 letters (A–Z) followed by 3 digits (0–9). How many plates are possible?

Solution. Five independent stages: $26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 676{,}000$. Now suppose the first symbol may be *either* a letter *or* a digit (so $26 + 10 = 36$ choices) while the rest stay as before: the count becomes $36 \cdot 26 \cdot 10^3$. The "either/or" at one stage is an addition *inside* a multiplication.

Worked example: casework by a special digit

How many 3-digit numbers (100–999) contain at least one digit equal to 7?

Solution. Complementary counting is cleanest: total 3-digit numbers $= 900$. Those with *no* 7: first digit has 8 choices (1–9 except 7), the other two have 9 each, giving $8 \cdot 9 \cdot 9 = 648$. So the answer is $900 - 648 = 252$.

Pitfall. The multiplication principle requires the *number* of ways at each stage to be constant, not the specific choices. If choosing A first changes how many options remain at stage 2 (versus choosing B), you must split into cases and add.

2.2 Factorials & Permutations

Ordered arrangements

$n! = n(n-1)(n-2)\cdots 2\cdot 1$ counts the orderings of n distinct objects, with $0! = 1$.

A **permutation** of k objects chosen from n distinct objects (order matters, no repetition) is

$$P(n,k) = \frac{n!}{(n-k)!} = \underbrace{n(n-1)\cdots(n-k+1)}_{k \text{ falling factors}}.$$

Read it as: fill k ordered slots, with n choices for the first, $n-1$ for the next, and so on.

Worked example: podium finishes

In a race of 8 runners, how many ways can the gold, silver, and bronze medals be awarded (no ties)?

Solution. Order matters and there is no repetition, so

$$P(8,3) = 8\cdot 7\cdot 6 = 336.$$

Worked example: arrange with a restriction

How many ways can 5 distinct books be lined up on a shelf if the dictionary must be on one of the two ends?

Solution. Place the dictionary first: 2 end choices. Arrange the other 4 books in the remaining slots: $4! = 24$. Total $2\cdot 24 = 48$. *Handle the restricted object first.*

Tip. $P(n,n) = n!$ because $0! = 1$. When a problem says "arrange," "order," "line up," or "rank," you are almost always in permutation territory.

2.3 Combinations

Unordered selections

A **combination** chooses k objects from n distinct objects when order does *not* matter:

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

Dividing by $k!$ removes the $k!$ orderings of each chosen set. Key identities:

$$\binom{n}{k} = \binom{n}{n-k}, \qquad \binom{n}{0} = \binom{n}{n} = 1, \qquad \sum_{k=0}^{n}\binom{n}{k} = 2^n.$$

Worked example: forming a committee

From 10 people, how many ways can we choose a committee of 4?

Solution. Order is irrelevant: $\binom{10}{4} = \frac{10\cdot9\cdot8\cdot7}{4\cdot3\cdot2\cdot1} = 210$.

Worked example: choosing with two groups

A club has 6 freshmen and 5 seniors. How many 4-person committees have exactly 2 freshmen?

Solution. Choose 2 of 6 freshmen *and* 2 of 5 seniors: $\binom{6}{2}\binom{5}{2} = 15 \cdot 10 = 150$. (Multiplication principle joins two independent choices.)

Pitfall. $\binom{n}{k}$ counts *sets*, not sequences. If you ever find yourself later imposing an order on a chosen set, you probably wanted a permutation, or you must multiply by the number of orderings.

2.4 Permutations with Identical Objects

Arrangements when some items repeat

The number of distinct arrangements of n objects, where there are n_1 identical objects of type 1, n_2 of type 2, ..., n_r of type r (with $n_1 + \cdots + n_r = n$), is the **multinomial coefficient**

$$\frac{n!}{n_1!\, n_2! \cdots n_r!}.$$

Each group of identical objects contributes a $k!$ overcount that must be divided out.

Worked example: rearranging a word

How many distinct arrangements of the letters in `MISSISSIPPI` are there?

Solution. There are 11 letters: $M{\times}1,\ I{\times}4,\ S{\times}4,\ P{\times}2$. So

$$\frac{11!}{1!\,4!\,4!\,2!} = \frac{39{,}916{,}800}{1\cdot24\cdot24\cdot2} = 34{,}650.$$

Worked example: lattice paths

How many shortest grid paths go from $(0,0)$ to $(4,3)$ moving only right (R) or up (U)?

Solution. Every path is a sequence of 4 R's and 3 U's, so we arrange 7 steps:

$$\frac{7!}{4!\,3!} = \binom{7}{3} = 35.$$

Lattice-path counting is really "permutations with identical objects" in disguise — a favorite contest trick.

Tip. The multinomial coefficient collapses to $\binom{n}{k}$ when there are only two types (a "k of one kind, $n-k$ of the other" string). Choosing positions and arranging identical objects are the same act.

2.5 Circular Permutations

Seating around a table

Arrangements around a circle count only *relative* position: rotating everyone one seat gives the same arrangement. Fixing one person to kill the n rotations gives

$$\text{circular arrangements of } n \text{ distinct objects} = (n-1)!.$$

If arrangements that are *mirror images* (reflections) are also considered identical — as with a bracelet or necklace of beads — divide by an additional 2:

$$\frac{(n-1)!}{2} \quad (n \geq 3).$$

Worked example: round table

In how many ways can 6 people sit at a round table if only relative order matters?

Solution. $(6-1)! = 5! = 120$.

Worked example: keyed seating

5 couples sit around a round table so that partners are adjacent. How many seatings (rotations equivalent)?

Solution. Treat each couple as a block: arrange 5 blocks around the circle in $(5-1)! = 24$ ways, then order the two partners within each block: $2^5 = 32$. Total $24 \cdot 32 = 768$.

Pitfall. Decide up front whether reflections count as the same. Seats fixed to the table (labeled chairs, a head of table, a clock face) make it *linear* again — use $n!$, not $(n-1)!$.

2.6 Order Matters vs. Order Doesn't (Choosing vs. Arranging)

The master decision

Before counting, ask: *does swapping two chosen items produce a different outcome?*

- **Yes** $\Rightarrow$ order matters $\Rightarrow$ arrangement/permutation: $P(n,k) = \dfrac{n!}{(n-k)!}$.
- **No** $\Rightarrow$ order irrelevant $\Rightarrow$ selection/combination: $\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$.

The bridge between them: $P(n,k) = \binom{n}{k} \cdot k!$. First *choose* the set, then *arrange* it.

Worked example: same numbers, different question

From 9 students:

- How many ways to pick a 3-person study group? $\binom{9}{3} = 84$ (order doesn't matter).
- How many ways to pick a President, VP, and Treasurer? $P(9,3) = 9 \cdot 8 \cdot 7 = 504$ (distinct roles $\Rightarrow$ order matters).

Notice $504 = 84 \cdot 3! = 84 \cdot 6$.

Tip. Titles, ranks, ordered slots, and positions signal "order matters." Committees, teams, hands of cards, and subsets signal "order doesn't."

2.7 Going Deeper: Subtleties, Overcounting & Symmetry

Necklaces vs. circular arrangements

A **circular arrangement** treats only rotations as identical: $(n-1)!$. A **necklace** (or bracelet) treats rotations *and* reflections as identical: $\frac{(n-1)!}{2}$ for $n \geq 3$. Always confirm whether the object can be physically flipped over. A round-table seating cannot be flipped (people keep their left/right neighbors distinct), but a bracelet of beads can.

Worked example: overcounting by symmetry

How many distinct triangles can be drawn using 8 points on a circle as vertices (no three collinear)?

Solution. A triangle is just an *unordered* choice of 3 vertices, so $\binom{8}{3} = 56$. If you had counted ordered triples $8 \cdot 7 \cdot 6 = 336$, you would be overcounting each triangle by its $3! = 6$ vertex orderings: $336/6 = 56$. Recognizing the hidden symmetry factor is the whole game.

Worked example: division-into-groups trap

In how many ways can 6 distinct players be split into 3 *unlabeled* pairs?

Solution. If the pairs were labeled (Court 1, 2, 3), the count would be $\binom{6}{2}\binom{4}{2}\binom{2}{2} = 15\cdot6\cdot1 = 90$. But the 3! orderings of the (identical-status) pairs are the same partition, so divide: $90/3! = 15$. *Dividing by the arrangements of interchangeable groups* is the most common contest overcount.

Worked example: complementary & inclusion–exclusion

How many 4-digit numbers (1000–9999) have at least one repeated digit?

Solution. Total $= 9000$. All-distinct: $9 \cdot 9 \cdot 8 \cdot 7 = 4536$ (first digit 1–9, then any unused digit including 0). At least one repeat $= 9000 - 4536 = 4464$. When "at least one" appears, subtract the "none" case.

Big picture — recognizing when order matters. Every basic counting problem reduces to three questions, asked in order:

1. **Are stages joined by AND (multiply) or OR (add)?** Split disjoint cases; never let one outcome fall into two cases.
2. **Within a stage, does order matter?** Yes $\Rightarrow$ permutation $P(n, k)$; No $\Rightarrow$ combination $\binom{n}{k}$; repeats among items $\Rightarrow$ divide by each group's factorial.
3. **Did symmetry make me overcount?** Rotations ($\div n$), reflections ($\div 2$), and interchangeable groups ($\div k!$) each demand a division. When in doubt, count an easy way with order, then divide out exactly the symmetries you introduced.

2.8 The Bijection Principle

Count a hard set by matching it to an easy one

If there is a **bijection** (a one-to-one, onto pairing) between a finite set A and a finite set B, then $|A| = |B|$. So to count a difficult set A, build a reversible dictionary that turns each object of A into exactly one object of an easier-to-count set B — and back again. The whole method is: *define the map, check it is invertible, count B.*

The lattice-path count you already met is a bijection in disguise: paths $\longleftrightarrow$ words in R's and U's. Most "clever" contest counts are really the discovery of the right bijection.

Worked example: stars and bars via a bijection

How many ways can we write $x_1 + x_2 + x_3 + x_4 = 10$ with each x_i a nonnegative integer?

Solution. Encode a solution as a row of 10 stars split into 4 groups by 3 bars, e.g.

$$\underbrace{\star\star}_{x_1=2} \mid \underbrace{\star\star\star}_{x_2=3} \mid \underbrace{}_{x_3=0} \mid \underbrace{\star\star\star\star\star}_{x_4=5}.$$

Every solution gives exactly one such string of 10 stars and 3 bars, and every string decodes to exactly one solution — a bijection. Counting the strings is a "permutations with identical objects" problem on 13 symbols:

$$\binom{10+3}{3} = \binom{13}{3} = 286.$$

In general $x_1 + \cdots + x_k = n$ (nonnegatives) has $\binom{n+k-1}{k-1}$ solutions.

Worked example: subsets with no two consecutive elements

How many subsets of $\{1, 2, \ldots, 12\}$ contain no two consecutive integers?

Solution. Suppose we pick k elements $a_1 < a_2 < \cdots < a_k$ with each gap $a_{i+1} - a_i \geq 2$. Define $b_i = a_i - (i-1)$. This subtracts off the forced gaps, and the map $\{a_i\} \mapsto \{b_i\}$ is a bijection onto the k-element subsets of $\{1, 2, \ldots, 12 - (k-1)\} = \{1, \ldots, 13-k\}$ with *no* restriction. Hence the count for each k is $\binom{13-k}{k}$, and summing over all k:

$$\sum_{k\geq 0} \binom{13-k}{k} = \binom{13}{0} + \binom{12}{1} + \binom{11}{2} + \binom{10}{3} + \binom{9}{4} + \binom{8}{5} + \binom{7}{6} = 1+12+55+120+126+56+7 = 377.$$

(These are Fibonacci numbers — a bijection can expose hidden structure, not just a number.)

Tip. To trust a bijection, exhibit the inverse: describe how to recover the original object from its image. If you cannot decode uniquely, the map is not a bijection and $|A| = |B|$ may fail. "Shift the indices" ($b_i = a_i - (i-1)$) is the single most useful bijection trick for turning a spacing constraint into a free choice.

2.9 Counting Functions Between Finite Sets

Every function is an assignment; count the assignments

Let $|X| = n$ (the domain) and $|Y| = m$ (the codomain). Think of each element of X as choosing a target in Y.

- **All functions $X \to Y$:** each of the n inputs picks any of m outputs independently, giving

$$m^n.$$

- **Injective (one-to-one) functions:** distinct inputs need distinct outputs, so fill the n

inputs with a falling factorial of choices:

$$m(m-1)(m-2)\cdots(m-n+1) = \frac{m!}{(m-n)!} = P(m,n),$$

which is 0 when $n > m$ (pigeonhole: too many inputs to keep distinct).

- **Surjective (onto) functions:** every output in Y must be hit at least once. There is no bare product formula; inclusion–exclusion (a later unit) gives

$$\mathrm{Surj}(n,m) = \sum_{j=0}^{m} (-1)^j \binom{m}{j} (m-j)^n.$$

Worked example: three formulas, one setup

Let $X = \{1,2,3,4\}$ and $Y = \{a,b,c\}$, so $n = 4$, $m = 3$.

Solution.

- All functions: $3^4 = 81$.
- Injective functions: impossible, since $n = 4 > 3 = m$; count $= 0$.
- Surjective functions: $\sum_{j=0}^{3} (-1)^j \binom{3}{j} (3-j)^4 = 3^4 - 3\cdot 2^4 + 3\cdot 1^4 - 0 = 81 - 48 + 3 = 36.$

Sanity check: a surjection from a 4-set onto a 3-set forces exactly one output to be used twice. Choose the doubled output (3 ways), choose which 2 of the 4 inputs share it ($\binom{4}{2} = 6$), then biject the remaining 2 inputs to the remaining 2 outputs ($2! = 2$): $3 \cdot 6 \cdot 2 = 36$. ✓

Tip. "Number of functions" problems are pure multiplication-principle problems wearing a costume: the domain elements are the independent stages. Watch the direction — functions $X \to Y$ give m^n, *not* n^m. The base is the size of the codomain (the menu); the exponent is the size of the domain (the diners).

2.10 The Multinomial Coefficient as a Distribution

One coefficient, three readings

For nonnegative integers with $k_1 + k_2 + \cdots + k_r = n$, define

$$\binom{n}{k_1,\, k_2,\, \ldots,\, k_r} = \frac{n!}{k_1!\, k_2! \cdots k_r!}.$$

This single number counts three equivalent things:

- **Arrangements:** distinct orderings of n objects where k_i are of type i (the `MISSISSIPPI` idea).
- **Distributions:** ways to deal n distinct objects into r distinct labeled boxes so box i receives exactly k_i objects.
- **Ordered set partitions:** ways to split $\{1, \ldots, n\}$ into an ordered list of blocks of sizes $k_1, \ldots, k_r$.

It also names the coefficients in the **Multinomial Theorem**

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{k_1+\cdots+k_r=n} \binom{n}{k_1, \ldots, k_r} x_1^{k_1} x_2^{k_2} \cdots x_r^{k_r}.$$

Worked example: a specific coefficient

Find the coefficient of $x^3y^2z^2$ in the expansion of $(x + y + z)^7$.

Solution. Here $n = 7$ with exponents $(k_1, k_2, k_3) = (3, 2, 2)$, which sum to 7:

$$\binom{7}{3, 2, 2} = \frac{7!}{3!\,2!\,2!} = \frac{5040}{6 \cdot 2 \cdot 2} = 210.$$

Worked example: dealing cards into hands

In how many ways can a standard 52-card deck be dealt into four labeled hands (North, East, South, West) of 13 cards each?

Solution. This is a distribution of 52 distinct objects into 4 labeled boxes of size 13:

$$\binom{52}{13,\ 13,\ 13,\ 13} = \frac{52!}{(13!)^4}.$$

If instead the four hands were *unlabeled* (just a partition into four piles of 13, no seat names), we would divide by the 4! orderings of the interchangeable piles: $\dfrac{52!}{(13!)^4\,4!}$.

Identity worth knowing. The multinomial factors through binomials:

$$\binom{n}{k_1, k_2, \ldots, k_r} = \binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots\binom{k_r}{k_r}.$$

This is just "fill box 1, then box 2 from what remains, …" — the multiplication principle. Summing all multinomials of order n gives r^n (set $x_1 = \cdots = x_r = 1$ in the theorem), matching the "each of n objects picks one of r boxes" count.

2.11 Orientation: The Twelvefold Way

Distributing balls into boxes — the master chart

Almost every elementary counting problem is "put n balls into m boxes." The answer depends on two switches — are the **balls** distinguishable? are the **boxes** distinguishable? — and on any **restriction** (no limit / at most one per box / at least one per box). The resulting $2 \times 2 \times 3$ grid is the classic *Twelvefold Way*. Reading it now shows how the tools of this unit fit together and previews later units.

Here $S(n,k)$ is a **Stirling number of the second kind** (partitions of a set into k nonempty blocks), $p_m(n)$ counts partitions of the integer n into exactly m positive parts, and $[\cdot]$ is 1 if the condition holds and 0 otherwise.

Worked example: reading four cells of the chart

Distribute balls into 3 boxes; compare the four "any" entries for small cases.

Solution.

- **Distinct balls, distinct boxes**, $n = 4$: functions $\{1,2,3,4\} \to \{\text{box}_1,\text{box}_2,\text{box}_3\}$, so $3^4 = 81$.
- **Identical balls, distinct boxes**, $n = 4$: stars and bars, $\binom{4+3-1}{4} = \binom{6}{4} = 15$.
- **Distinct balls, identical boxes**, $n = 4$ into at most 3 nonempty groups: $S(4,1) + S(4,2) + S(4,3) = 1 + 7 + 6 = 14$.
- **Identical balls, identical boxes**, $n = 4$ into at most 3 parts: partitions of 4 with ≤ 3 parts: 4, 3+1, 2+2, 2+1+1, so 4.

Same phrase, four genuinely different answers — the two "distinguishable?" switches are everything.

Big picture. The top row (distinct/distinct) is this unit's home turf: functions, falling factorials, and surjection counts. The second row (identical balls, distinct boxes) is the stars-and-bars world of the next unit. The bottom two rows introduce set partitions ($S(n,k)$) and integer partitions ($p(n)$), the subject of still later units. When a problem stumps you, first locate its cell: *are the balls alike? are the boxes alike? is there a per-box constraint?*

3 Counting Strategies: Casework, Complementary & Overcounting

Competition counting & probability, Unit 2

3.1 Casework Counting

The idea

When a counting problem has no single clean formula, **split it into cases** that are *mutually exclusive* (no outcome lands in two cases) and *exhaustive* (every outcome lands in some case). Count each case separately, then **add**.

$$\text{Total} = (\text{Case } 1) + (\text{Case } 2) + \cdots + (\text{Case } k).$$

Choose your cases along the feature that is causing the trouble: the value of the largest digit, whether a coin came up heads, how many objects of a certain type were chosen, and so on. Good cases make each sub-count easy.

Worked example: sums on two dice

How many ordered pairs (a, b) with $a, b \in \{1, \dots, 6\}$ have $a + b$ a multiple of 3?

Case $a + b = 3$: $(1, 2), (2, 1)$ — 2 ways. **Case** $a + b = 6$: $(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)$ — 5 ways. **Case** $a + b = 9$: $(3, 6), (4, 5), (5, 4), (6, 3)$ — 4 ways. **Case** $a + b = 12$: $(6, 6)$ — 1 way. These cases are disjoint and cover all multiples of 3 between 2 and 12, so the total is $2 + 5 + 4 + 1 = \boxed{12}$.

Worked example: three-digit numbers with a fixed digit sum

How many three-digit numbers have digits summing to 5? Write the number as $\overline{abc}$ with $a \geq 1$. Case on the hundreds digit a: for each $a \in \{1, 2, 3, 4, 5\}$ the tens and units must sum to $5 - a$, giving $(5 - a) + 1 = 6 - a$ ordered choices. Summing, $5 + 4 + 3 + 2 + 1 = \boxed{15}$ numbers.

Pitfall: cases must not overlap. If an outcome could be counted in two cases, you will double count. Before adding, ask: "Could a single outcome satisfy two of my case descriptions?" If yes, redefine the cases (e.g. "exactly k" instead of "at least k").

3.2 Complementary Counting

The idea

Sometimes counting what you *don't* want is far easier than counting what you do. If there are N total outcomes and B "bad" outcomes, then the number of "good" outcomes is

$$\text{Good} = N - \text{Bad}.$$

This shines when the good outcomes fracture into many messy cases but the bad ones form a tidy family (often a single case).

Worked example: at least one repeated digit

How many three-digit numbers (100 to 999) have at least one repeated digit? Total three-digit numbers: 900. The complement is "all digits distinct": the hundreds digit has 9 choices (1–9), the tens digit 9 choices (any but the hundreds), the units 8 choices, giving $9 \cdot 9 \cdot 8 = 648$. Therefore the answer is $900 - 648 = \boxed{252}$.

Tip: the complement trick pairs perfectly with the phrase **"at least one."** The opposite of "at least one" is "none," which is usually one clean product. See the final section.

3.3 Constructive Counting

The idea

Build a valid configuration **one decision at a time**, and multiply the number of choices at each step (the multiplication principle). The art is choosing an order of decisions so that the number of options at each step does *not* depend on the earlier choices in a way you cannot track. Make the most constrained decision first.

Worked example: seating with a restriction

Five people sit in a row of five chairs, but Alice and Bob refuse to sit in the two end chairs. Build the seating: first place the three "unrestricted" people (call them the others). Actually, seat the restricted pair first, since they are the bottleneck. The two end chairs must be filled by 3 non-restricted people: choose an ordered pair for the two ends in $3 \cdot 2 = 6$ ways. The remaining 3 chairs take the remaining 3 people in $3! = 6$ ways. Total $6 \cdot 6 = \boxed{36}$.

Worked example: strings with no two adjacent A's

How many length-4 strings over $\{A, B\}$ have no two adjacent A's? Build left to right and case on the number of A's. Equivalently, place the A's in gaps around the B's: with k B's there are $k+1$ gaps and we choose positions for the A's. For $0, 1, 2$ A's we get $\binom{5}{0} + \binom{4}{1} + \binom{3}{2} = 1 + 4 + 3 = \boxed{8}$.

Pitfall: the multiplication principle needs the *count* of options at each step to be constant, even if the specific options change. If step 3 sometimes offers 2 choices and sometimes 3 depending on step 1, split into cases first, then multiply within each case.

3.4 Overcounting and Correcting

The idea

Often the easiest count treats some outcomes as different when they should be the same. If every final object is produced **the same number of times** d by your over-count M, then

$$\text{Answer} = \frac{M}{d}.$$

The classic example: arranging n distinct objects in a *circle*. Linear arrangements give $n!$, but each circular arrangement appears n times (one per rotation), so the count is $n!/n = (n-1)!$.

Worked example: a necklace

How many distinct necklaces use 5 different colored beads if rotations *and* reflections are considered the same? Line them up: $5! = 120$ arrangements. Each necklace can be rotated 5 ways and flipped 2 ways, so it is counted $5 \cdot 2 = 10$ times. Answer: $120/10 = \boxed{12}$.

Worked example: choosing a committee counted twice

A club counts committees of 2 from 6 people by "pick a first member (6 ways), pick a second (5 ways)" getting 30. But $\{X, Y\}$ and $\{Y, X\}$ are the same committee, counted $d = 2$ times. The true count is $30/2 = 15 = \binom{6}{2}$, showing $\binom{n}{k} = \dfrac{n!/(n-k)!}{k!}$ is exactly correcting for the $k!$ orderings.

Pitfall: division only works when *every* object is overcounted the same number of times. If some beads repeat, or a symmetric arrangement is fixed by a rotation, the copies are not all distinct and simple division fails — you need Burnside-style counting (final section).

3.5 Bijections & Correspondences

The idea

To count a hard set A, find a **bijection** (a one-to-one, onto pairing) with an easier set B; then $|A| = |B|$. A bijection guarantees "no misses, no doubles." Two famous correspondences:

- **Subsets $\leftrightarrow$ binary strings:** subsets of an n-set correspond to length-n strings of `0`/`1`, so there are 2^n subsets.
- **Stars and bars:** the number of ways to write n as an ordered sum of k nonnegative integers equals $\binom{n+k-1}{k-1}$, by pairing each solution with an arrangement of n stars and $k-1$ bars.

Worked example: stars and bars

How many solutions in nonnegative integers does $x_1 + x_2 + x_3 = 7$ have? Represent a solution as 7 stars split by 2 bars, e.g. $\star\star \mid \star\star\star \mid \star\star$. Each arrangement of 7 stars and 2 bars gives exactly one solution, so the count is $\binom{7+2}{2} = \binom{9}{2} = \boxed{36}$.

Worked example: lattice paths

How many shortest grid paths go from $(0,0)$ to $(4,3)$ moving only right/up? Each path is a sequence of 4 R's and 3 U's — a bijection with strings, so $\binom{7}{3} = \boxed{35}$.

Tip: to prove your map is a bijection, describe the **inverse**: given an object in B, recover exactly one object in A. If you can undo the map uniquely, it is a bijection.

3.6 "At Least One" via the Complement

The idea

The event "at least one" is the complement of "none." In counting,

$$\#(\text{at least one } X) = (\text{total}) - \#(\text{no } X),$$

and in probability, $P(\text{at least one}) = 1 - P(\text{none})$. Because "none" usually means every trial avoids X independently, its count/probability is a single product.

Worked example: at least one six

Roll a fair die 4 times. Probability of *at least one* six? "No six" on all four rolls has probability $\left(\frac{5}{6}\right)^4 = \frac{625}{1296}$. So $P(\text{at least one six}) = 1 - \frac{625}{1296} = \boxed{\frac{671}{1296}}$.

Tip: watch for the words *at least*, *at most*, *some*, *none*. These almost always signal complementary counting. "At least one" $\to$ subtract "none"; "at most one" $\to$ subtract "two or more."

3.7 Going Deeper: Combining Strategies

Layering techniques

Contest problems rarely use one idea in isolation. Common combinations:

- **Casework + complementary:** split off the easy-to-count bad outcomes with cases, subtract from the total. E.g. "at least two of a kind" = total − (all distinct).

- **Constructive + overcounting:** build with the multiplication principle, then divide out an ordering or symmetry you introduced (this is exactly how $\binom{n}{k}$ arises).
- **Casework inside a bijection:** map to an easier set, then do casework there.

Correcting for symmetry: a Burnside flavor

When simple division fails because some arrangements are *fixed* by a symmetry, use the **Burnside idea**: the number of distinct arrangements equals the *average* number of arrangements left unchanged by each symmetry:

$$\#(\text{distinct}) = \frac{1}{|G|} \sum_{g \in G} \mathrm{Fix}(g),$$

where G is the group of symmetries (e.g. the n rotations of a circle) and $\mathrm{Fix}(g)$ counts colorings unchanged by g. When no nontrivial symmetry fixes any coloring, every $\mathrm{Fix}(g) = 0$ except the identity, and Burnside collapses to the plain division $M/|G|$.

Worked example: Burnside on a bracelet

Color the 3 beads of a triangular bracelet with 2 colors, rotations only (G has 3 elements: rotate by $0, 1, 2$). Identity fixes all $2^3 = 8$ colorings; each nontrivial rotation fixes only the 2 monochromatic colorings. By Burnside, $\frac{1}{3}(8 + 2 + 2) = \frac{12}{3} = \boxed{4}$ distinct bracelets.

Big picture: every strategy here answers the same question — "how do I count each valid outcome exactly once?" Casework partitions the outcomes; complementary counting flips to the easier side; constructive counting sequences the decisions; overcounting/Burnside repairs double counts; bijections transport the problem somewhere friendlier. When stuck, ask three questions: *Is the opposite easier to count? Can I break this into clean cases? Am I counting anything more than once?*

3.8 Burnside's Lemma (Orbit-Counting Theorem)

The statement

Let a finite group G of symmetries act on a set X of configurations (colorings, labelings, …). Two configurations are "the same" when one can be carried to the other by some $g \in G$; the resulting equivalence classes are called **orbits**. The number of distinct configurations is the number of orbits, and

$$\#(\text{orbits}) \;=\; \frac{1}{|G|} \sum_{g \in G} \bigl|\mathrm{Fix}(g)\bigr|,$$

where $\mathrm{Fix}(g) = \{x \in X : g \cdot x = x\}$ is the set of configurations left *unchanged* by g. In words:

the number of distinct objects equals the average number of colorings fixed by a symmetry. This is the rigorous repair for the “divide by symmetry” shortcut, and it never overcounts even when some configurations have extra symmetry.

The workhorse formula for colorings

When g permutes the n “slots” (beads, faces, cells) and you may color each slot freely with c colors, a coloring is fixed by g *iff* every slot in a cycle of g receives the same color. Hence

$$\bigl|\mathrm{Fix}(g)\bigr| \;=\; c^{\,\mathrm{cyc}(g)},$$

where $\mathrm{cyc}(g)$ is the number of disjoint cycles of g (fixed points count as 1-cycles). So the whole problem reduces to a **cycle census** of the group.

Worked example: two-color bracelets of 6 beads

Count the distinct bracelets of 6 beads, each black or white, where rotations *and* reflections are considered the same. The symmetry group is the dihedral group D_6 with $|G| = 12$. Take a cycle census, using $\bigl|\mathrm{Fix}(g)\bigr| = 2^{\mathrm{cyc}(g)}$.

Rotations by k positions have $\gcd(k, 6)$ cycles:

k	0	1	2	3	4	5
cyc	6	1	2	3	2	1
2^{cyc}	64	2	4	8	4	2

sum $= 84$.

Reflections (six of them, since n is even): 3 axes pass through two opposite beads (2 fixed beads $+2$ swaps $= 4$ cycles $\Rightarrow 2^4 = 16$ each), and 3 axes pass through opposite edges (3 swaps $= 3$ cycles $\Rightarrow 2^3 = 8$ each). Reflection sum $= 3(16) + 3(8) = 48 + 24 = 72$. Therefore

$$\#(\text{bracelets}) = \frac{84 + 72}{12} = \frac{156}{12} = \boxed{13}.$$

(Rotations alone would give $84/6 = 14$ *necklaces*; folding in reflections identifies one mirror pair, leaving 13.)

Worked example: coloring the faces of a cube with 3 colors

How many distinct ways can the 6 faces of a cube be painted using 3 available colors, where two paintings are the same if a rotation of the cube carries one to the other? The rotation group has $|G| = 24$. Census of how each rotation type permutes the 6 faces:

Averaging over the group,

$$\frac{1 \cdot 729 + 6 \cdot 27 + 3 \cdot 81 + 8 \cdot 9 + 6 \cdot 27}{24} = \frac{729 + 162 + 243 + 72 + 162}{24} = \frac{1368}{24} = \boxed{57}.$$

The same recipe with c colors gives $\frac{1}{24}(c^6 + 3c^4 + 12c^3 + 8c^2)$ colorings.

Method: (1) identify the symmetry group G and its size; (2) sort the elements into a few *types* (identity, rotations by angle, reflections through this-or-that axis); (3) for each type find $\mathrm{cyc}(g)$ and multiply $c^{\mathrm{cyc}(g)}$ by how many elements have that type; (4) add and divide by $|G|$. The answer is always an integer — a fast sanity check on your census.

3.9 The "Divide by Symmetry" Trap, Formalized

When plain division is legal

The shortcut Answer $= M/|G|$ (over-count divided by the number of symmetries) is valid **exactly when the group action is free**: no nontrivial symmetry fixes any configuration, so every orbit has the full size $|G|$. This is the *regular/free* case. Equivalently, by the orbit–stabilizer theorem,

$$|\text{orbit of } x| = \frac{|G|}{|\mathrm{Stab}(x)|},$$

so orbits all have size $|G|$ precisely when every stabilizer is trivial. If *some* configuration is fixed by a nontrivial g (a symmetric coloring), its orbit is *smaller* than $|G|$, plain division **undercounts the collapse and gives the wrong answer**, and you must use Burnside.

Worked example: why $2^6/12$ is wrong for bracelets

Naively "2^6 colorings, divide by 12 symmetries" gives $64/12 = 5.33\ldots$ — not even an integer, a red flag. The action is *not* free: the all-black bracelet is fixed by all 12 symmetries (stabilizer size 12, orbit size 1), and a bracelet like BWBWBW is fixed by several rotations and reflections. These small orbits are exactly what Burnside's averaging accounts for, producing the correct 13 from the previous section. Plain division is only correct for the *asymmetric* colorings, whose orbits genuinely have full size 12.

Worked example: seating where division *is* legal

Seat 7 distinct people around a round table; rotations are equivalent (no reflections). No nontrivial rotation can fix an arrangement of 7 *distinct* people, so the action of the 7 rotations is free and every orbit has size 7. Plain division is valid: $\frac{7!}{7} = 6! = \boxed{720}$. The distinctness of the objects is what guarantees freeness.

Decision rule: before dividing by $|G|$, ask "could any nontrivial symmetry leave a configuration *unchanged*?" If the objects being arranged are all distinct and G is just rotations, usually no — divide freely. If repeats are allowed (colorings) or reflections are included, symmetric configurations almost always exist — reach for Burnside. A non-integer after dividing is proof the action was not free.

3.10 Complementary Counting Meets Inclusion–Exclusion

Counting the complement of a union

The strongest complementary arguments subtract a *union* of bad events, and unions are handled by inclusion–exclusion:

$$\left|\overline{A_1 \cup \cdots \cup A_n}\right| = N - \sum_i |A_i| + \sum_{i<j} |A_i \cap A_j| - \cdots + (-1)^n |A_1 \cap \cdots \cap A_n|.$$

Use this when "good" means "avoids every one of several forbidden features." The intersections are usually easy products, even when the individual good count is a casework nightmare.

Worked example: integers coprime-ish to $2, 3, 5$

How many integers in $\{1, 2, \ldots, 1000\}$ are divisible by *none* of $2, 3, 5$? Let A_p be the multiples of p. Then $|A_2| = 500,\ |A_3| = 333,\ |A_5| = 200$; pairwise $|A_6| = 166,\ |A_{10}| = 100,\ |A_{15}| = 66$; and $|A_{30}| = 33$. The count divisible by at least one is

$$500 + 333 + 200 - 166 - 100 - 66 + 33 = 734,$$

so the answer is the complement $1000 - 734 = \boxed{266}$.

Worked example: surjections via inclusion–exclusion

How many functions from a 5-element set *onto* a 3-element set $\{1, 2, 3\}$ are there? Let A_i be the functions that *miss* value i. A surjection avoids every A_i, so

$$\#(\text{onto}) = \sum_{k=0}^{3} (-1)^k \binom{3}{k} (3-k)^5 = 3^5 - 3 \cdot 2^5 + 3 \cdot 1^5 = 243 - 96 + 3 = \boxed{150}.$$

Direct casework on the "shape" of the preimages (3+1+1 vs. 2+2+1) also works but is error-prone; the complementary/IE form is mechanical.

Tip: inclusion–exclusion signs alternate, so a sign slip is the most common mistake. Check the last term's sign: with n forbidden properties it is $(-1)^n$. Also verify a small case by hand (e.g. surjections from a 2-set onto a 2-set should give 2) before trusting the general formula.

3.11 Systematic Casework and Generating Functions

Choosing what to case on

Two cases are not equal in quality. Aim for a parameter that is **most constraining** and **symmetry-friendly**:

- **Case on the extreme:** the largest element, the maximum multiplicity, the position of the first success. Extremes split the problem into independent "above" and "below" pieces.
- **Exploit symmetry to fuse cases:** if the roles of two variables are interchangeable, count one ordering and multiply (remember to halve the tie case). This can cut the case count in half or more.
- **Prefer few fat cases over many thin ones:** a parameter taking 3 values, each an easy product, beats one taking 12 values you must enumerate.

Generating functions as bookkeeping

A generating function replaces casework with algebra: the number of ways to choose objects with a target total is a **coefficient** in a product of polynomials/series. Each factor encodes one independent choice; multiplying performs the "add over all splits" automatically. Bounded choices give finite polynomials; a single die is $x + x^2 + \cdots + x^6 = \dfrac{x(1-x^6)}{1-x}$.

Worked example: three dice summing to 10 — casework vs. GF

In how many ordered ways do three dice (1–6) sum to 10? The generating function is $\left(x + x^2 + \cdots + x^6\right)^3$, and we want the coefficient of x^{10}. Shift each die by 1 (let $a' = a - 1$, etc.), so we need $a' + b' + c' = 7$ with $0 \le a', b', c' \le 5$. Unbounded stars and bars gives $\binom{9}{2} = 36$; subtract the solutions where some variable exceeds 5 (i.e. is ≥ 6). Forcing one variable up by 6 leaves a sum of 1 among three nonnegative variables, $\binom{3}{2} = 3$ ways, and there are 3 choices of which variable overflowed (no two can overflow since $2 \cdot 6 > 7$). Thus

$$36 - 3 \cdot 3 = \boxed{27}.$$

Compare the raw casework: summing $\#\{a + b = 10 - c\}$ over $c = 1, \ldots, 6$ forces you to track the 1–6 bounds on a, b case by case; the GF/IE route handles every bound in one subtraction.

Tip: the substitution "$a' = a -$ (lower bound)" turns any $\ge$ constraint into a nonnegativity constraint so stars and bars applies; each *upper* bound violation is then removed by one inclusion–exclusion term (shift the offending variable past its ceiling). If at most one variable can overflow, that correction is a single clean subtraction, as above.

4 Distributions: Stars & Bars and Integer Solutions

Competition counting & probability, Unit 3

4.1 Combinations with Repetition

The multiset counting formula

Suppose you want to choose n items from k *types*, where each type is available in unlimited supply and **order does not matter**. The number of ways is

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

This counts **multisets** of size n drawn from k types. Unlike ordinary combinations $\binom{k}{n}$, repeats are allowed, so a type may be chosen more than once.

The key contrast to keep straight:

- **Permutations** (n^k or $k!/(k-n)!$): order matters.
- **Combinations** $\binom{k}{n}$: order does not matter, no repeats.
- **Combinations with repetition** $\binom{n+k-1}{n}$: order does not matter, repeats allowed.

Choosing donuts

A shop sells 5 kinds of donuts. You buy a box of 8 donuts. How many different boxes are possible?

Solution. Here $k = 5$ types and $n = 8$ chosen, repeats allowed, order irrelevant. The count is

$$\binom{8+5-1}{8} = \binom{12}{8} = \binom{12}{4} = 495.$$

Nondecreasing sequences

How many nondecreasing sequences $1 \le a_1 \le a_2 \le a_3 \le a_4 \le 6$ of length 4 are there?

Solution. A nondecreasing sequence is the same as a multiset of size 4 chosen from the 6 values $\{1, \ldots, 6\}$: sorting a multiset gives exactly one such sequence. So the count is $\binom{4+6-1}{4} = \binom{9}{4} = 126$.

Tip. "How many ways to pick with repetition allowed, order not mattering" and "how many nondecreasing sequences" are *the same problem*. Spotting one form in disguise is half the battle.

4.2 Stars and Bars (identical items to distinct boxes)

The stars-and-bars picture

To distribute n **identical** items among k **distinct** boxes (any box may get zero), draw n stars and $k-1$ bars in a row. The bars split the stars into k groups; group i becomes the contents of box i. Every arrangement of

$$\underbrace{\star\star\cdots\star}_{n} \quad \text{and} \quad \underbrace{|\,|\,\cdots\,|}_{k-1}$$

gives one distribution, so the number of distributions is

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

For example, $\star\star \mid \star \mid\mid \star\star\star$ with $k=4$ boxes means box 1 gets 2, box 2 gets 1, box 3 gets 0, box 4 gets 3. There are $n+k-1$ symbol positions; choose which $k-1$ are bars.

Identical candies to children

In how many ways can 10 identical candies be given to 4 children (a child may get none)?

Solution. Distribute $n=10$ identical items into $k=4$ distinct boxes:

$$\binom{10+4-1}{4-1} = \binom{13}{3} = 286.$$

Careful. Stars and bars requires the **items to be identical** and the **boxes to be distinct**. If the objects are distinguishable, this formula does *not* apply — see "Distributing Distinct Objects."

4.3 Nonnegative vs. Positive Integer Solutions

Counting integer solutions of an equation

The number of **nonnegative** integer solutions $(x_1, \dots, x_k)$ to

$$x_1 + x_2 + \cdots + x_k = n, \qquad x_i \geq 0$$

is $\binom{n+k-1}{k-1}$. This is stars and bars: each x_i is the number of stars in box i.

For **positive** solutions ($x_i \geq 1$), first give each variable 1 unit. Substitute $y_i = x_i - 1 \geq 0$;

then $y_1 + \cdots + y_k = n - k$, so the count is

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

Two counts, one equation

How many solutions does $x_1 + x_2 + x_3 + x_4 = 12$ have (a) in nonnegative integers, (b) in positive integers?

Solution. (a) Nonnegative: $\binom{12+4-1}{4-1} = \binom{15}{3} = 455$.

(b) Positive: set $y_i = x_i - 1$, giving $y_1 + \cdots + y_4 = 8$, so $\binom{8+4-1}{3} = \binom{11}{3} = 165$.

Substitution is your workhorse. A lower bound $x_i \geq c$ is removed by the substitution $y_i = x_i - c$. This turns any constant floor into the standard nonnegative problem. Just remember to subtract c from the right-hand side.

Mixed lower bounds

Count nonnegative solutions to $a + b + c = 20$ with $a \geq 2$, $b \geq 5$.

Solution. Let $a' = a - 2$, $b' = b - 5$. Then $a' + b' + c = 13$ with all ≥ 0, giving $\binom{13+3-1}{3-1} = \binom{15}{2} = 105$.

4.4 Bounded Variables (upper limits)

Upper bounds via inclusion–exclusion

Lower bounds are easy (substitute). **Upper** bounds like $x_i \leq u_i$ need inclusion–exclusion. To count solutions of $x_1 + \cdots + x_k = n$ with $x_i \geq 0$ and some cap $x_i \leq u_i$: subtract the "bad" cases where a variable exceeds its cap.

If a variable is capped at $x_i \leq u$, the bad cases have $x_i \geq u + 1$; substitute to force that excess and count, then alternate signs over all subsets of violated caps.

One upper bound

How many nonnegative solutions does $x + y + z = 15$ have with $x \leq 6$?

Solution. Total (no cap): $\binom{15+2}{2} = \binom{17}{2} = 136$. Bad cases have $x \geq 7$: set $x' = x - 7$, so $x' + y + z = 8$, giving $\binom{8+2}{2} = \binom{10}{2} = 45$. Answer $= 136 - 45 = 91$.

Two upper bounds (inclusion–exclusion)

Count nonnegative solutions of $x + y + z = 15$ with $x \le 6$ and $y \le 6$.

Solution. Let A be solutions with $x \ge 7$, B with $y \ge 7$.

$$\begin{aligned} \text{Total} &= \binom{17}{2} = 136, \\ |A| = |B| &= \binom{10}{2} = 45, \\ |A \cap B| &= \binom{(15-14)+2}{2} = \binom{3}{2} = 3. \end{aligned}$$

By inclusion–exclusion the good count is $136 - 45 - 45 + 3 = 49$.

Sign pattern. Add the unrestricted total, subtract each single-cap violation, add back each double violation, subtract triples, and so on. Any term whose forced excess exceeds n contributes 0 and can be dropped.

4.5 Distributing Distinct Objects

When the objects are distinguishable

If n **distinct** objects are placed into k **distinct** boxes and each object is chosen independently, there are

$$k^n$$

ways (each object picks its box). This is a different world from stars and bars: here objects carry identity, so labeling matters.

Grouping distinct objects into *unlabeled* nonempty groups is counted by Stirling numbers of the second kind $S(n, k)$; distributing into labeled boxes with all nonempty is $k!\,S(n, k)$ (a surjection count).

Distinct vs. identical

In how many ways can 5 different books be placed on 3 distinct shelves (order on a shelf ignored, empty shelves allowed)?

Solution. Each book independently chooses one of 3 shelves: $3^5 = 243$.

Contrast: 5 *identical* books on 3 shelves would be $\binom{5+2}{2} = 21$.

Decision rule. Ask: are the items identical or distinct? Identical $\Rightarrow$ stars and bars, $\binom{n+k-1}{k-1}$. Distinct with free choice $\Rightarrow$ k^n. Distinct with a fixed size in each box $\Rightarrow$ multinomial coefficient (next section).

4.6 Multinomial Coefficients

Splitting distinct objects into fixed-size groups

The number of ways to partition n distinct objects into labeled groups of sizes $n_1, n_2, \ldots, n_k$ (with $n_1 + \cdots + n_k = n$) is the multinomial coefficient

$$\binom{n}{n_1,\, n_2,\, \ldots,\, n_k} = \frac{n!}{n_1!\, n_2! \cdots n_k!}.$$

These coefficients appear in the multinomial theorem:

$$(x_1 + \cdots + x_k)^n = \sum_{n_1+\cdots+n_k=n} \binom{n}{n_1, \ldots, n_k} x_1^{n_1} \cdots x_k^{n_k}.$$

Dealing into fixed hands

In how many ways can 12 distinct cards be split into three labeled hands of sizes 5, 4, and 3?

Solution. $\binom{12}{5,4,3} = \dfrac{12!}{5!\,4!\,3!} = 27720.$

A multinomial coefficient in an expansion

Find the coefficient of $x^2y^3z^2$ in $(x+y+z)^7$.

Solution. With $n = 7$ and exponents $(2,3,2)$,

$$\binom{7}{2,3,2} = \frac{7!}{2!\,3!\,2!} = \frac{5040}{2 \cdot 6 \cdot 2} = 210.$$

Equal group sizes. If several groups have the *same* size and the groups themselves are **unlabeled**, divide by the factorial of the number of equal-size groups to remove overcounting. Labeled groups need no such correction.

4.7 Going Deeper: Techniques That Combine

Layering the tools

Contest problems rarely use one idea in isolation. The strongest solvers layer them:

- **Lower + upper bounds together.** First substitute away lower bounds ($y_i = x_i - c_i$), then apply inclusion–exclusion on the remaining upper caps. Order matters: clear floors first, then handle ceilings.

- **Bounded variables via inclusion–exclusion.** For $x_1+\cdots+x_k=n$ with each $x_i \le u$, the good count is
$$\sum_{j\ge 0}(-1)^j\binom{k}{j}\binom{n-j(u+1)+k-1}{k-1},$$
dropping any term with a negative top argument.
- **Distributing distinct objects with restrictions.** "No box empty" becomes a surjection count $k!\,S(n,k)=\sum_{j=0}^{k}(-1)^j\binom{k}{j}(k-j)^n$ by inclusion–exclusion on empty boxes.
- **Multinomial expansions.** The number of *terms* in $(x_1+\cdots+x_k)^n$ before combining is k^n; the number of *distinct monomials* after combining is $\binom{n+k-1}{k-1}$ — stars and bars on the exponents.

Bounded caps by the formula

Count nonnegative solutions of $x_1+x_2+x_3+x_4=10$ with each $x_i \le 4$.

Solution. Here $k=4$, $n=10$, $u=4$ so $u+1=5$:
$$\binom{4}{0}\binom{13}{3}-\binom{4}{1}\binom{8}{3}+\binom{4}{2}\binom{3}{3}$$
$$=286-4\cdot 56+6\cdot 1=286-224+6=68.$$
(The $j=3$ term needs $10-15<0$, so it vanishes.)

Surjections onto boxes

In how many ways can 6 distinct prizes be given to 3 students so that every student gets at least one prize?

Solution. Inclusion–exclusion on empty students:
$$\sum_{j=0}^{3}(-1)^j\binom{3}{j}(3-j)^6=3^6-3\cdot 2^6+3\cdot 1^6=729-192+3=540.$$

Big picture. Every distribution problem is answered by three questions: (1) Are the **items** identical or distinct? (2) Are the **boxes** distinct, and can they be empty? (3) Are there **bounds** on how much a box holds? Identical items give binomial (stars-and-bars) counts; distinct items give powers, multinomials, or Stirling/surjection counts; bounds are cleared by substitution (lower) and inclusion–exclusion (upper). Name the three answers before you compute, and the right formula chooses itself.

4.8 Compositions vs. Partitions

Ordered and unordered ways to break up n

A **composition** of n is an ordered sequence of positive parts summing to n; a **partition** is the same but *order is ignored.* Thus $3 = 2 + 1 = 1 + 2$ gives two compositions but only one partition $(2 + 1)$.

Compositions are just positive stars and bars. The number of compositions of n into exactly k positive parts is

$$\binom{n-1}{k-1},$$

since we place $k - 1$ dividers into the $n - 1$ gaps between n dots. Summing over k,

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

compositions in total: each of the $n - 1$ internal gaps is independently a "cut" or "no cut."

Partitions are far subtler — there is *no* simple closed formula. The number of partitions of n is written $p(n)$, and it grows fast: $p(1), \ldots, p(6) = 1, 2, 3, 5, 7, 11$.

Counting both flavors for $n = 5$

List the compositions and partitions of 5 into parts, and reconcile the counts.

Solution. Total compositions of 5: $2^{5-1} = 16$. Compositions into exactly $k = 3$ parts: $\binom{5-1}{3-1} = \binom{4}{2} = 6$, namely

$$3{+}1{+}1,\ 1{+}3{+}1,\ 1{+}1{+}3,\ 2{+}2{+}1,\ 2{+}1{+}2,\ 1{+}2{+}2.$$

By contrast the *partitions* of 5 are 5; 4+1; 3+2; 3+1+1; 2+2+1; 2+1+1+1; 1+1+1+1+1, so $p(5) = 7$ — much smaller than 16, because reorderings collapse.

Order test. If rearranging the parts gives a *different* outcome, you are counting compositions — use stars and bars, $\binom{n-1}{k-1}$ or 2^{n-1}. If rearrangements are the *same* outcome, you are counting partitions, which need generating functions or Ferrers-diagram arguments, not a single binomial.

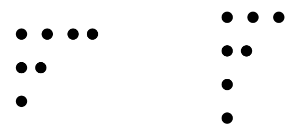
4.9 Integer Partitions and Ferrers Diagrams

Drawing a partition; the conjugate

A partition $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 1$ of n is drawn as a **Ferrers diagram**: left-justified rows of dots, row i having λ_i dots. Reflecting the diagram across its main diagonal (swap rows and columns) gives the **conjugate partition** λ^*, another partition of the same n.

Conjugation is a bijection on partitions of n, so it proves facts by symmetry. For instance: *the number of partitions of n into at most k parts equals the number of partitions of n into parts of size at most k* (conjugation swaps "number of rows" with "length of longest row").

For $\lambda = (4, 2, 1)$ of $n = 7$, the diagram and its conjugate $\lambda^* = (3, 2, 1, 1)$ are



$\lambda = (4, 2, 1)$ $\lambda^* = (3, 2, 1, 1)$

Reading the column heights of the left diagram $(3, 2, 1, 1)$ gives exactly the rows of the right one.

Distinct parts = odd parts (Euler)

Show, for $n = 6$, that the number of partitions into **distinct** parts equals the number into **odd** parts.

Solution. Distinct-part partitions of 6: 6; 5+1; 4+2; 3+2+1 — there are 4. Odd-part partitions of 6: 5+1; 3+3; 3+1+1+1; 1+1+1+1+1+1 — also 4.

Why (the bijection). Write each odd-part partition by grouping equal parts: a part d appearing m times contributes $d \cdot m$. Expand the multiplicity m in binary, $m = \sum 2^j$, and send it to the *distinct* parts $d \cdot 2^j$. Every distinct part factors uniquely as (odd)×(power of 2), so the map reverses — proving the two counts are always equal, matching Euler's generating-function identity

$$\prod_{i \geq 1} (1 + x^i) = \prod_{i \geq 1} \frac{1}{1 - x^{2i-1}}.$$

Bijections beat formulas. Since $p(n)$ has no elementary closed form, partition identities are proved by *matching* objects: conjugate diagrams (rows ↔ columns) or the odd/binary trick above. When a problem claims two partition counts are equal, look for a reversible construction rather than a formula.

4.10 The Generating-Function View of Stars and Bars

Distributions as coefficients

Each box contributes a **factor** recording what it can hold; multiplying the factors and reading off the coefficient of x^n counts the distributions summing to n. An unrestricted box (any count $0, 1, 2, \ldots$) contributes $1 + x + x^2 + \cdots = \dfrac{1}{1 - x}$, so k free boxes give

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

— the stars-and-bars formula, now as a coefficient. **Restrictions just change the factor:**

- box must be nonempty: $x + x^2 + \cdots = \dfrac{x}{1-x}$;
- box holds at most u: $1 + x + \cdots + x^u = \dfrac{1-x^{u+1}}{1-x}$;
- box holds only even amounts: $1 + x^2 + x^4 + \cdots = \dfrac{1}{1-x^2}$;
- coin of value v, any number used: $\dfrac{1}{1-x^v}$.

The whole distribution problem becomes: *extract* $[x^n]$ *from a product of such factors.*

Coin change by coefficient extraction

In how many ways can you make 6 cents using pennies (1), nickels (5), and unlimited 2-cent pieces?

Solution. The generating function is

$$\frac{1}{1-x} \cdot \frac{1}{1-x^2} \cdot \frac{1}{1-x^5},$$

and we want $[x^6]$. Since we need only up to x^6, truncate each factor. A nickel is used 0 or 1 time (≤ 6). With no nickel we need $[x^6]$ of $\frac{1}{(1-x)(1-x^2)}$, i.e. pennies+twos summing to 6: the twos number $0, 1, 2, 3$, giving 4 ways. With one nickel we need $[x^1]$ of the same, i.e. sum 1 from pennies and twos: only 1 way. Total $= 4 + 1 = 5$.

Recovering a bounded-cap count

Use generating functions to count nonnegative solutions of $x_1 + x_2 + x_3 = 8$ with each $x_i \leq 4$.

Solution. Each capped box contributes $\dfrac{1-x^5}{1-x}$, so we need

$$[x^8]\,\frac{(1-x^5)^3}{(1-x)^3} = [x^8]\,(1 - 3x^5 + \cdots)\sum_n \binom{n+2}{2} x^n.$$

Only the 1 and $-3x^5$ terms reach degree 8:

$$\binom{10}{2} - 3\binom{5}{2} = 45 - 3 \cdot 10 = 15.$$

This is exactly the inclusion–exclusion answer — the $(1-x^5)^3$ factor *is* the alternating sum.

One factor per box. Build the generating function box by box: unrestricted $\to \frac{1}{1-x}$, capped $\to \frac{1-x^{u+1}}{1-x}$, step-$v$ $\to \frac{1}{1-x^v}$. The binomial series $\frac{1}{(1-x)^k} = \sum \binom{n+k-1}{k-1} x^n$ turns coefficient extraction back into stars and bars, and the numerator polynomials reproduce inclusion–exclusion automatically.

4.11 The Twelvefold Way

One table for every distribution problem

Placing n **balls** into k **boxes** has $2 \times 2 \times 3 = 12$ variants: balls identical or distinct, boxes identical or distinct, and the map unrestricted / injective ("at most one per box") / surjective ("no box empty"). Every counting formula on this sheet is one cell of the table. Below, $S(n,k)$ is a Stirling number of the second kind and $p_k(n)$ is the number of partitions of n into *at most* k parts.

balls\boxes	any (distinct)	injective	surjective (distinct)
distinct	k^n	$k^{\underline{n}}$	$k!\,S(n,k)$
identical	$\binom{n+k-1}{k-1}$	$\binom{k}{n}$	$\binom{n-1}{k-1}$
dist., boxes ident.	$\sum_{j=1}^{k} S(n,j)$	$[n \le k]$	$S(n,k)$
ident., boxes ident.	$\sum_{j=1}^{k} p_j(n)$	$[n \le k]$	$p_k(n)$ (exactly k)

Here $k^{\underline{n}} = k(k-1)\cdots(k-n+1)$ is the falling factorial and $[\,\cdot\,]$ is 1 if true, else 0.

The four rows are exactly the four big ideas of this unit: distinct balls into distinct boxes (powers, falling factorials, surjections), identical balls into distinct boxes (stars and bars, positive stars and bars), distinct balls into identical boxes (Stirling numbers), and identical balls into identical boxes (integer partitions).

Reading four cells off one scenario

Distribute balls into 3 boxes; compare the counts for $n = 4$ balls across the four ball/box types (unrestricted maps).

Solution.

$$\begin{aligned}
\text{distinct balls, distinct boxes} &: 3^4 = 81,\\
\text{identical balls, distinct boxes} &: \binom{4+2}{2} = \binom{6}{2} = 15,\\
\text{distinct balls, identical boxes} &: S(4,1)+S(4,2)+S(4,3) = 1+7+6 = 14,\\
\text{identical balls, identical boxes} &: p_3(4) = 4 \quad (4;\ 3{+}1;\ 2{+}2;\ 2{+}1{+}1).
\end{aligned}$$

The same physical setup gives four wildly different answers — which is why naming the ball/box types *first* is essential.

AIME-flavored classification

How many ways can 10 identical marbles be split among 4 **identical** bags with each bag nonempty?

Solution. Identical balls, identical boxes, surjective = partitions of 10 into *exactly* 4 positive parts. Subtract 1 from each part: partitions of $10-4=6$ into at most 4 parts, i.e. $p_4(6)$. The partitions of 6 into ≤ 4 parts are

$$6;\ 5{+}1;\ 4{+}2;\ 3{+}3;\ 4{+}1{+}1;\ 3{+}2{+}1;\ 2{+}2{+}2;\ 3{+}1{+}1{+}1;\ 2{+}2{+}1{+}1,$$

which is 9. (Contrast *distinct* bags: that would be the positive stars-and-bars count $\binom{9}{3}=84$.)

Master decision procedure. Fix three switches before computing: **(1)** balls identical or distinct? **(2)** boxes distinct or identical? **(3)** map unrestricted, injective, or surjective? Distinct boxes $\Rightarrow$ closed-form binomials/powers; identical boxes $\Rightarrow$ Stirling numbers (distinct balls) or integer partitions (identical balls), which have *no* simple closed form. The Twelvefold Way is not twelve formulas to memorize — it is one triple of questions that routes you to the right tool on this sheet.

5 Binomial Theorem, Pascal & Combinatorial Identities

Competition counting & probability, Unit 4

5.1 The Binomial Theorem & Specific Coefficients

The Binomial Theorem

For any nonnegative integer n,

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

The coefficient $\binom{n}{k}$ is called a **binomial coefficient**. The term containing y^k (and hence x^{n-k}) is

$$\binom{n}{k} x^{\,n-k} y^{\,k}.$$

Why it is true (combinatorial view): expanding $(x+y)^n = \underbrace{(x+y)(x+y)\cdots(x+y)}_{n}$ means choosing, from each of the n factors, either x or y. A term $x^{n-k}y^k$ arises from every way of choosing y in exactly k of the factors, and there are $\binom{n}{k}$ such choices.

Worked example: a specific coefficient

Find the coefficient of x^5 in $(2x-3)^8$.

The general term is $\binom{8}{k}(2x)^{8-k}(-3)^k = \binom{8}{k}2^{8-k}(-3)^k x^{8-k}$. We need $8-k=5$, so $k=3$:

$$\binom{8}{3}2^5(-3)^3 = 56 \cdot 32 \cdot (-27) = -48384.$$

So the coefficient of x^5 is $\boxed{-48384}$.

Tips. (1) Match the exponent *first*, then plug in k. (2) Track signs carefully: a term like $(-3)^k$ alternates. (3) For $\left(ax+\frac{b}{x}\right)^n$, the general term is $\binom{n}{k}a^{n-k}b^k x^{n-2k}$; set $n-2k$ equal to the desired power.

5.2 Pascal's Triangle & Pascal's Rule

Pascal's Rule

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

Every interior entry of Pascal's Triangle is the sum of the two entries directly above it. Row n (starting at $n=0$) lists $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$.

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

Combinatorial proof: to form a k-element committee from n people, fix one person P. Either P is on the committee (choose the other $k-1$ from the remaining $n-1$: $\binom{n-1}{k-1}$ ways) or P is not (choose all k from the remaining $n-1$: $\binom{n-1}{k}$ ways).

Worked example: using Pascal's Rule

Simplify $\binom{9}{4}+\binom{9}{5}$.

By Pascal's Rule with $n-1=9$, $k=5$: $\binom{9}{4}+\binom{9}{5}=\binom{10}{5}=252$.

Tips. Symmetry $\binom{n}{k} = \binom{n}{n-k}$ mirrors each row. The first nonzero entries are $\binom{n}{0} = 1$ and $\binom{n}{1} = n$. Memorize rows up to $n = 6$ for speed on contests.

5.3 Row Sums & Alternating Sums

Sum and alternating sum of a row

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n, \qquad \sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0 \quad (n \geq 1).$$

Set $x = y = 1$ in the Binomial Theorem for the first; set $x = 1$, $y = -1$ for the second. **Combinatorial view of 2^n:** the total number of subsets of an n-element set, counted by size, is $\sum_k \binom{n}{k}$; but each element is independently in or out, giving 2^n subsets.

Worked example: even-index sum

Evaluate $\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots$.

Add the two identities: adding $\sum \binom{n}{k} = 2^n$ and $\sum(-1)^k \binom{n}{k} = 0$ cancels the odd-index terms and doubles the even ones, so

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

The odd-index sum equals 2^{n-1} as well.

Tips. Plugging clever values ($x = 1, -1, 2, \frac{1}{2}, \ldots$) into $(x+y)^n$ is the master trick for summing binomial coefficients. To split even/odd indices, average $f(1)$ and $f(-1)$.

5.4 The Hockey Stick Identity

Hockey Stick Identity

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

Summing a diagonal of Pascal's Triangle (the "stick") gives the entry just below and to the side of its end (the "blade"). **Combinatorial proof:** to choose $r+1$ numbers from $\{1, 2, \ldots, n+1\}$, condition on the largest chosen number $i+1$. The remaining r numbers come from $\{1, \ldots, i\}$ in $\binom{i}{r}$ ways; summing over $i = r, \ldots, n$ gives the total $\binom{n+1}{r+1}$.

Worked example

Compute $\binom{2}{2}+\binom{3}{2}+\binom{4}{2}+\binom{5}{2}+\binom{6}{2}$.

Here $r=2$, $n=6$, so the sum is $\binom{7}{3}=35$. (Check: $1+3+6+10+15=35$.)

Tips. The lower index r is fixed while the upper index climbs. A common variant, $\sum_{i=0}^{m}\binom{r+i}{i}=\binom{r+m+1}{m}$, follows by symmetry.

5.5 Vandermonde's Identity

Vandermonde's Identity

$$\sum_{k=0}^{p}\binom{m}{k}\binom{n}{p-k}=\binom{m+n}{p}.$$

Combinatorial proof: to choose a committee of p people from a group of m women and n men, choose k women and $p-k$ men; summing over all valid k counts every committee exactly once, giving $\binom{m+n}{p}$. The special case $m=n=p$ gives $\sum_k \binom{n}{k}^2=\binom{2n}{n}$ (using symmetry $\binom{n}{n-k}=\binom{n}{k}$).

Worked example

Evaluate $\sum_{k=0}^{3}\binom{3}{k}\binom{4}{3-k}$.

By Vandermonde with $m=3$, $n=4$, $p=3$: the sum is $\binom{7}{3}=35$.

Tips. Recognize Vandermonde whenever a sum has *two* binomial factors whose *upper* indices are constant and whose *lower* indices add to a constant. The sum-of-squares case $\sum\binom{n}{k}^2=\binom{2n}{n}$ appears constantly.

5.6 Combinatorial (Committee–Chair) Identities

The absorption / committee–chair identity

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

Combinatorial proof (choose a committee and a chair): count pairs (committee of size k, a distinguished chair on it). Left side: pick the committee ($\binom{n}{k}$ ways), then the chair from its k members. Right side: pick the chair first (n ways), then the remaining $k-1$ members from the other $n-1$ people. Two useful consequences:

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

Worked example

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

Replace $k\binom{n}{k} = n\binom{n-1}{k-1}$:

$$\sum_{k=0}^{n} k\binom{n}{k} = n\sum_{k=1}^{n}\binom{n-1}{k-1} = n\sum_{j=0}^{n-1}\binom{n-1}{j} = n\,2^{\,n-1}.$$

Tips. The "double counting" strategy—count one set two ways—proves nearly all binomial identities. The absorption identity is also how you turn a weighted sum $\sum k\binom{n}{k}$ into a plain row sum.

5.7 Lattice Path Counting

Counting monotonic lattice paths

The number of lattice paths from the origin $(0,0)$ to (m,n) using only unit steps **right** (R) and **up** (U) is

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

Each path is a sequence of $m+n$ steps of which exactly m are R; choosing *which* steps are R determines the path.

Worked example

How many shortest paths go from $(0,0)$ to $(4,3)$ along grid lines?

Each path uses 4 rights and 3 ups, 7 steps total: $\binom{7}{4} = \binom{7}{3} = 35$ paths. (One such path is drawn above.)

Tips. Paths through a required point P multiply: (paths $O \to P$)$\times$(paths $P \to$ end). To *avoid* a forbidden point, subtract the paths that pass through it. Pascal's Rule is exactly the lattice-path recursion: paths to (m,n) = paths to $(m-1,n)$ + paths to $(m,n-1)$.

5.8 Going Deeper: Proofs, Generating Functions & Extracting Coefficients

Three lenses on the same identities

1. Combinatorial (double counting). Count a set of objects two different ways; the two counts must be equal. This proved Pascal's Rule, Hockey Stick, Vandermonde, and the committee–chair identity above.

2. Generating functions. Treat $(1+x)^n = \sum_k \binom{n}{k} x^k$ as a polynomial whose coefficients *store* the binomial coefficients. Multiplying series adds exponents, so identities about sums become statements about products of polynomials. For instance, Vandermonde falls out of

$$(1+x)^m(1+x)^n = (1+x)^{m+n}$$

by comparing the coefficient of x^p on both sides: the left side gives $\sum_k \binom{m}{k}\binom{n}{p-k}$, the right gives $\binom{m+n}{p}$.

3. Extracting coefficients. The symbol $[x^p]F(x)$ means "the coefficient of x^p in F." Many contest problems reduce to computing $[x^p]$ of a product of simple factors—exactly the specific-coefficient technique from Section 1.

Worked example: coefficients via a product

Find $\sum_{k=0}^{n} \binom{n}{k}^2$ using generating functions.

Since $\binom{n}{k} = \binom{n}{n-k}$, the sum is $\sum_k \binom{n}{k}\binom{n}{n-k} = [x^n](1+x)^n(1+x)^n = [x^n](1+x)^{2n} = \binom{2n}{n}$.

Big picture. Binomial coefficients live at the crossroads of algebra and counting. Every identity in this unit has a story (*count something two ways*) and a formula (*compare coefficients in a product*). When you see a sum of products of binomial coefficients, ask three questions: *Is it a row sum?* ($x = \pm 1$) *Is it a diagonal?* (Hockey Stick) *Are two groups being merged?* (Vandermonde). And when you need a single coefficient, match the exponent first, then evaluate. Master both the counting story and the algebra, and these problems become routine.

5.9 The Roots of Unity Filter

Extracting an arithmetic slice of a row

Let $\omega = e^{2\pi i/m}$ be a primitive mth root of unity. The key fact is the **orthogonality relation**

$$\frac{1}{m}\sum_{j=0}^{m-1} \omega^{jt} = \begin{cases} 1 & \text{if } m \mid t, \\ 0 & \text{otherwise.} \end{cases}$$

This average acts as an on/off switch that survives only when the exponent is a multiple of m. Apply it to the generating polynomial $f(x) = (1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$. Substituting $x = \omega^j$, weighting by ω^{-jr}, and averaging over j isolates exactly the terms with $k \equiv r \pmod{m}$:

$$\boxed{\sum_{k \equiv r \pmod{m}} \binom{n}{k} = \frac{1}{m} \sum_{j=0}^{m-1} \omega^{-jr} (1+\omega^j)^n.}$$

The $m = 2$, $r = 0$ case recovers the even-index sum 2^{n-1} from Section 3; the filter is its far-reaching generalization.

Worked example (AIME flavor): every third entry

Evaluate $S = \binom{9}{0} + \binom{9}{3} + \binom{9}{6} + \binom{9}{9} = \sum_k \binom{9}{3k}$.

Take $m = 3$, $r = 0$, $n = 9$, and $\omega = e^{2\pi i/3} = -\frac{1}{2} + \frac{\sqrt{3}}{2} i$. The filter gives

$$S = \frac{1}{3}\Big[(1+1)^9 + (1+\omega)^9 + (1+\omega^2)^9\Big].$$

Now simplify the two complex factors. Since $1 + \omega + \omega^2 = 0$,

$$1 + \omega = \tfrac{1}{2} + \tfrac{\sqrt{3}}{2} i = e^{i\pi/3}, \qquad 1 + \omega^2 = \tfrac{1}{2} - \tfrac{\sqrt{3}}{2} i = e^{-i\pi/3}.$$

Both have modulus 1, so raising to the 9th power rotates the angle to $\pm 3\pi$:

$$(1+\omega)^9 = e^{3\pi i} = -1, \qquad (1+\omega^2)^9 = e^{-3\pi i} = -1.$$

Therefore

$$S = \frac{1}{3}\big[512 + (-1) + (-1)\big] = \frac{510}{3} = \boxed{170}.$$

(Check directly: $1 + 84 + 84 + 1 = 170$.)

Tips. (1) The complex factors always pair as conjugates, so their sum is real—write $(1+\omega^j)^n$ in polar form $R^n e^{in\theta}$ and add. (2) The general clean formula for $m = 3$, $r = 0$ is $\sum_k \binom{n}{3k} = \frac{1}{3}\big(2^n + 2\cos\frac{n\pi}{3}\big)$. (3) The same machine filters coefficients of *any* polynomial or power series, not just $(1+x)^n$—it is the standard tool for "count the objects whose size is $\equiv r \pmod{m}$" problems.

5.10 The Generalized Binomial Series

Binomial coefficients with any upper index

Define, for *any* real (or complex) α and integer $k \geq 0$,

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!}.$$

Newton's generalized binomial theorem then extends the expansion to non-integer exponents: for $|x| < 1$,

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

When $\alpha = n$ is a nonnegative integer the numerator eventually hits the factor $(\alpha - \alpha) = 0$, killing all terms past $k = n$ and recovering the ordinary finite Binomial Theorem. For other α the series is genuinely infinite.

Negative integer upper index (ties back to stars and bars). Setting $\alpha = -n$ and simplifying the falling product gives

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

so that

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

The coefficient $\binom{n+k-1}{k}$ is exactly the stars-and-bars count of the ways to write k as an ordered sum of n nonnegative integers—so the negative binomial series *is* the generating function for distributing k identical items into n distinct boxes.

Worked example (olympiad flavor): a square-root series

Show that the central binomial coefficients appear as the coefficients of $(1-4x)^{-1/2}$; that is, prove $[x^k](1-4x)^{-1/2} = \binom{2k}{k}$.

Apply Newton's theorem with $\alpha = -\frac{1}{2}$ and the variable $-4x$:

$$(1-4x)^{-1/2} = \sum_{k=0}^{\infty} \binom{-1/2}{k} (-4)^k x^k.$$

Now evaluate $\binom{-1/2}{k}$. The numerator is

$$\left(-\tfrac{1}{2}\right)\left(-\tfrac{3}{2}\right)\cdots\left(-\tfrac{2k-1}{2}\right) = (-1)^k \frac{1\cdot 3\cdot 5\cdots(2k-1)}{2^k} = (-1)^k \frac{(2k)!}{2^k\, 2^k\, k!},$$

using $1\cdot 3\cdots(2k-1) = \dfrac{(2k)!}{2^k k!}$. Hence

$$\binom{-1/2}{k} = (-1)^k \frac{(2k)!}{4^k\, k!\, k!} = (-1)^k \frac{1}{4^k}\binom{2k}{k}.$$

Multiplying by $(-4)^k$ cancels both the sign and the 4^k:

$$[x^k](1-4x)^{-1/2} = (-1)^k \frac{1}{4^k}\binom{2k}{k}\cdot(-4)^k = \boxed{\binom{2k}{k}}.$$

This is the standard generating function $\sum_{k\ge 0}\binom{2k}{k}x^k = (1-4x)^{-1/2}$.

Tips. (1) The identity $\binom{-n}{k} = (-1)^k\binom{n+k-1}{k}$ ("negating the upper index") converts messy negative-index expressions into ordinary positive binomial coefficients. (2) Whenever a problem produces $\dfrac{1}{(1-x)^n}$, read off $\binom{n+k-1}{k}$—no separate stars-and-bars argument needed. (3) Fractional exponents ($\alpha = \pm\frac{1}{2}$) generate central binomial and Catalan-number series, which recur throughout combinatorics.

5.11 Coefficient Extraction: The Unifying Method

Turning identities into products of series

Every summation identity in this unit is a shadow of a single equation between generating functions, read off by comparing one coefficient. The recipe: encode each side as a power series, multiply, and equate $[x^N]$. Because multiplying series *convolves* their coefficients,

$$[x^N]\, A(x)B(x) = \sum_k \big([x^k]A\big)\big([x^{N-k}]B\big),$$

a product of two known series manufactures a sum of products of coefficients—precisely the shape of Vandermonde, the subcommittee identity, and their kin. This "snake-oil" method turns the *search* for a closed form into the *routine* task of recognizing a product.

Worked example (olympiad flavor): the subcommittee identity

Evaluate $T = \displaystyle\sum_{k=0}^{n}\binom{n}{k}\binom{k}{m}$ (a committee of size k from n people, with a distinguished sub-board of size m).

Combinatorial read. Choosing a committee and then an m-member sub-board is the same as choosing the m-member board *first* ($\binom{n}{m}$ ways) and then freely including or excluding each of the remaining $n-m$ people (2^{n-m} ways). So we expect $T = \binom{n}{m}2^{n-m}$.

Coefficient-extraction proof. Write $\binom{k}{m} = [y^m](1+y)^k$ and sum a geometric-style series in x using $\sum_k \binom{n}{k}\big((1+y)x\big)^k = (1+(1+y)x)^n$. Setting $x=1$,

$$\sum_k \binom{n}{k}(1+y)^k = (2+y)^n, \qquad \text{so} \qquad T = [y^m](2+y)^n.$$

By the ordinary Binomial Theorem $(2+y)^n = \sum_j \binom{n}{j} 2^{n-j} y^j$, whose y^m coefficient is

$$T = \binom{n}{m} 2^{n-m}.$$

Both stories agree, and the algebra needed nothing but "compare the coefficient of y^m."

Master strategy. Faced with an unfamiliar binomial sum, encode it as $[x^N]$ of a product you recognize: use $(1+x)^a(1+x)^b = (1+x)^{a+b}$ for Vandermonde-type merges, $\dfrac{1}{1-x}$ to *accumulate* a running sum (this is Hockey Stick in disguise), and Newton's series from the previous section for negative or fractional exponents. Then the identity is proved the instant you name the coefficient. Combinatorial double-counting supplies the *meaning*; coefficient extraction supplies a mechanical *proof*—and together they make even AIME/olympiad-level binomial sums tractable.

6 Inclusion–Exclusion, Derangements & Pigeonhole

Competition counting & probability, Unit 5

6.1 Two– and Three–Set Inclusion–Exclusion

Counting a Union Without Double–Counting

When two sets overlap, adding their sizes counts the overlap *twice*, so subtract it once:

$$\boxed{|A \cup B| = |A| + |B| - |A \cap B|}$$

For three sets, add the singles, subtract the pairs, add back the triple:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

To count elements in **none** of the sets, subtract the union from the universe $|U|$:

$$|\overline{A} \cap \overline{B}| = |U| - |A \cup B|.$$

Two overlapping sets: the middle region is counted once in $|A|$ and once in $|B|$, so we subtract $|A \cap B|$.

Example: A Survey

Of 50 people, 30 like coffee, 25 like tea, and 10 like both. How many like at least one drink? How many like neither?

$$|C \cup T| = 30 + 25 - 10 = \mathbf{45}. \qquad \text{Neither} = 50 - 45 = \mathbf{5}.$$

Three sets: + singles − pairs + triple. Each atomic region ends up counted exactly once.

Example: Divisibility

How many integers from 1 to 30 are divisible by 2, 3, or 5?

$$\left\lfloor \tfrac{30}{2} \right\rfloor = 15,\ \left\lfloor \tfrac{30}{3} \right\rfloor = 10,\ \left\lfloor \tfrac{30}{5} \right\rfloor = 6; \quad \left\lfloor \tfrac{30}{6} \right\rfloor = 5,\ \left\lfloor \tfrac{30}{10} \right\rfloor = 3,\ \left\lfloor \tfrac{30}{15} \right\rfloor = 2; \quad \left\lfloor \tfrac{30}{30} \right\rfloor = 1.$$

$$15 + 10 + 6 - 5 - 3 - 2 + 1 = \mathbf{22}.$$

Tip: The count divisible by both a and b is the count divisible by $\operatorname{lcm}(a, b)$ — not ab, unless a and b are coprime. For 6 and 10, use $\operatorname{lcm} = 30$, not 60.

6.2 General Inclusion–Exclusion

The Full Formula

For sets $A_1, \ldots, A_n$, alternate sums over single sets, pairs, triples, and so on:

$$\left| \bigcup_{i=1}^{n} A_i \right| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n+1} \big| A_1 \cap \cdots \cap A_n \big|.$$

Equivalently, the number of elements in **none** of the sets is

$$\big| \overline{A_1} \cap \cdots \cap \overline{A_n} \big| = \sum_{S \subseteq \{1,\ldots,n\}} (-1)^{|S|} \Big| \bigcap_{i \in S} A_i \Big|,$$

where the empty intersection is the whole universe $|U|$.

Example: Coprime–Style Count

How many integers from 1 to 100 are divisible by *none* of 2, 3, 5?

$$|U| = 100; \quad 50 + 33 + 20 \text{ (singles)}, \quad 16 + 10 + 6 \text{ (pairs: } 6, 10, 15), \quad 3 \text{ (triple: } 30).$$

$$100 - (50 + 33 + 20) + (16 + 10 + 6) - 3 = 100 - 103 + 32 - 3 = \mathbf{26}.$$

Sign pattern: the term for an intersection of m sets carries sign $(-1)^{m+1}$ in the union formula (add odd, subtract even). In the "none" formula it is $(-1)^m$. Miscounting signs is the #1 error — write the layers out explicitly.

6.3 Derangements

Permutations With No Fixed Point

A **derangement** is a permutation in which *no* element stays in its original position. Inclusion–exclusion over the "i is fixed" events gives

$$\boxed{D_n = n!\sum_{k=0}^{n}\frac{(-1)^k}{k!} = n!\left(1-\frac{1}{1!}+\frac{1}{2!}-\cdots+\frac{(-1)^n}{n!}\right).}$$

Two handy recurrences: $D_n = (n-1)\big(D_{n-1}+D_{n-2}\big)$ and $D_n = nD_{n-1}+(-1)^n$.

n	1	2	3	4	5	6
D_n	0	1	2	9	44	265

As n grows, $D_n/n! \to 1/e \approx 0.3679$, so a random permutation is a derangement about 37% of the time.

Example: The Hat–Check Problem

Four people check identical-looking hats; on the way out the hats are returned at random. In how many ways does *nobody* get their own hat?

$$D_4 = 4!\left(1-1+\tfrac{1}{2}-\tfrac{1}{6}+\tfrac{1}{24}\right) = 24\cdot\tfrac{9}{24} = \mathbf{9}.$$

Check with the recurrence: $D_4 = 3(D_3+D_2) = 3(2+1) = 9$.

Tip: "Exactly m people get their own item" $= \binom{n}{m}D_{n-m}$: choose who is fixed, then derange the rest. Summing over m recovers $n!$, a useful sanity check.

6.4 Counting Surjections (Onto Functions)

Onto Functions via Inclusion–Exclusion

The number of **surjections** (onto functions) from an n–element set onto a k–element set is found by excluding the maps that miss at least one target value:

$$\boxed{\mathrm{Surj}(n,k) = \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n.}$$

This also counts the ways to distribute n *distinct* objects into k *distinct* boxes with no box empty. It relates to Stirling numbers of the second kind by

$$\mathrm{Surj}(n,k) = k!\, S(n,k),$$

where $S(n,k)$ counts partitions of the objects into k unlabeled nonempty groups.

Example: Onto a 3–Set

How many functions from $\{1,2,3,4\}$ onto $\{a,b,c\}$ are there?

$$\mathrm{Surj}(4,3) = 3^4 - \binom{3}{1}2^4 + \binom{3}{2}1^4 - \binom{3}{3}0^4 = 81 - 48 + 3 - 0 = \mathbf{36}.$$

Equivalently $3!\, S(4,3) = 6 \cdot 6 = 36$.

Trap: k^n counts *all* functions; only $\mathrm{Surj}(n,k)$ are onto. If $n < k$ there are 0 surjections (too few inputs to hit every output), and the alternating sum indeed collapses to 0.

6.5 The Pigeonhole Principle

The Basic Principle

If n objects are placed into k boxes and $n > k$, then some box contains **at least two** objects. More sharply: you cannot have every box with at most one object once the objects outnumber the boxes.

The art is choosing the right "boxes." Common box constructions: remainders modulo m, sign/parity classes, sub-regions of a shape, or pairs that sum to a target.

Five items in four boxes force at least one box to hold two.

Example: Same Remainder

Show that among any 5 integers, two leave the same remainder when divided by 4.

The possible remainders are $\{0,1,2,3\}$ — four boxes. With 5 integers (five objects), two must share a remainder. Their *difference* is then divisible by 4.

Tip: "Two share a remainder mod m" is the same as "their difference is a multiple of m." This turns many divisibility claims into a pigeonhole count with m boxes.

6.6 Generalized Pigeonhole & Extremal Arguments

The Sharpened Count

If n objects go into k boxes, then some box contains at least

$$\boxed{\left\lceil \frac{n}{k} \right\rceil}$$

objects. (If every box had at most $\lceil n/k \rceil - 1$, the total would be under n.) Turned around, to *force* some box to hold at least r objects you need $n > k(r-1)$, i.e. $n = k(r-1)+1$ objects.

Extremal principle: to find the largest "bad-free" set, group elements into forbidden pairs/blocks; you may take at most one representative per block, and the pigeonhole bound is tight when you take exactly one.

Example: Guaranteeing a Color

A drawer holds socks in 5 colors. How many must you draw to be sure of 4 of one color?

$$\text{Worst case: 3 of each color} = 3 \cdot 5 = 15 \text{ socks, still no quartet.} \quad 15 + 1 = \mathbf{16}.$$

Here $k = 5$, $r = 4$, so $n = k(r-1)+1 = 5 \cdot 3 + 1 = 16$.

Example: Extremal Subset

What is the largest subset of $\{1, 2, \ldots, 20\}$ containing no two elements that sum to 21?
Pair them: $(1, 20), (2, 19), \ldots, (10, 11)$ — 10 pairs. From each pair you may keep at most one number, so the largest safe subset has size **10** (e.g. $\{1, 2, \ldots, 10\}$).

Strategy: "How many to guarantee r?" means *build the worst case* ($r-1$ per box) and add one. "Largest set avoiding a condition?" means *group into forbidden blocks* and take one per block. These are the same pigeonhole idea read in two directions.

6.7 Going Deeper: Connections & Advanced Tools

Inclusion–Exclusion and Euler's Totient

Euler's totient $\varphi(n)$ counts integers in $\{1, \ldots, n\}$ coprime to n. Let $p_1, \ldots, p_r$ be the distinct primes dividing n; applying the "none of the divisibility events" form of PIE gives the product formula

$$\varphi(n) = n \prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right).$$

For example $\varphi(12) = 12\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right) = 12 \cdot \frac{1}{2} \cdot \frac{2}{3} = 4$ (namely $1, 5, 7, 11$). This is the surjection/derangement machinery in disguise: a signed sum over subsets of "forbidden" primes.

Example: Surjections and Set Partitions

The number of ways to split 5 labeled tasks among 3 identical days with each day used at least once is a *set partition* count $S(5,3) = 25$. Labeling the days (making them distinct) multiplies by $3!$:

$$\mathrm{Surj}(5,3) = 3!\, S(5,3) = 6 \cdot 25 = 150 = 3^5 - 3 \cdot 2^5 + 3 \cdot 1^5.$$

The PIE formula and the Stirling–number viewpoint always agree.

Pigeonhole in Number Theory & Geometry

Number theory. Among any $n+1$ integers, two are congruent modulo n, so their difference is a multiple of n. Among any $n+1$ numbers chosen from $\{1, \dots, 2n\}$, two are consecutive-in-a-pair and one divides the other (write each as $2^a \cdot (\text{odd})$; there are only n odd parts — the boxes).

Geometry. Partition a region into smaller cells; if you place more points than cells, two points share a cell and lie within its diameter. E.g. 5 points in a unit square: cut it into four $\frac{1}{2} \times \frac{1}{2}$ squares, so two points lie within distance $\frac{\sqrt{2}}{2}$.

Example: Divisibility by a Difference

Prove that among any 12 integers, two have a difference divisible by 11.

The residues modulo 11 form 11 boxes $\{0, 1, \dots, 10\}$. Twelve integers into eleven residue classes force two into the same class; their difference is then divisible by 11. (This is exactly $\lceil 12/11 \rceil = 2$.)

The Extremal Principle

Many existence problems yield to: *consider the largest, smallest, or otherwise most extreme object and argue about it.* Choose the configuration maximizing (or minimizing) some quantity; if the desired property failed, you could push the extreme further — a contradiction. Paired with pigeonhole, the extremal principle proves that a structure must exist without constructing it explicitly.

Big picture. Inclusion–exclusion, derangements, surjection counts, and the totient are all one idea: *count everything, then correct for over-counting with alternating signs over subsets of "bad" conditions.* Pigeonhole and the extremal principle are the reverse temperament — they *prove existence* by counting boxes rather than configurations. Together they let you both **count exactly** and **guarantee** structure. When a direct count stalls, ask: *what is being over-counted* (PIE), or *what are the boxes* (pigeonhole)?

6.8 Stirling Numbers of the Second Kind & Bell Numbers

Partitioning a Set into Blocks

The **Stirling number of the second kind** $S(n,k)$ counts the ways to partition an n–element set into exactly k *nonempty, unlabeled* blocks. Building an element one at a time gives the recurrence

$$\boxed{S(n,k) = k\,S(n-1,k) + S(n-1,k-1)}$$

(the new element either joins one of the k existing blocks, or starts a fresh block with the previous $n-1$ elements split into $k-1$ blocks). Boundary values: $S(n,0) = [n=0]$, $S(n,1) = S(n,n) = 1$, $S(n,2) = 2^{n-1} - 1$.

Labeling the blocks turns each partition into a surjection, recovering the count from the previous section:

$$k!\,S(n,k) = \mathrm{Surj}(n,k) = \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n.$$

Summing over all block-counts gives the **Bell number** $B_n = \sum_{k=0}^{n} S(n,k)$, the total number of partitions of an n–set.

n	0	1	2	3	4	5	6
B_n	1	1	2	5	15	52	203

Each entry is k times the entry directly above plus the entry up-and-left.

Example: Distinct Prizes, Identical Teams (AIME flavor)

In how many ways can 6 distinct prizes be split among 3 *indistinguishable* nonempty groups? This is $S(6,3)$. Build it from the recurrence, using $S(5,3) = 25$ and $S(5,2) = 15$:

$$S(6,3) = 3\,S(5,3) + S(5,2) = 3 \cdot 25 + 15 = \mathbf{90}.$$

If the groups were instead labeled (say, Room 1, Room 2, Room 3), multiply by 3!: $\mathrm{Surj}(6,3) = 6 \cdot 90 = 540 = 3^6 - 3 \cdot 2^6 + 3 \cdot 1^6$.

Labeled vs. unlabeled: identical boxes/groups ⇒ use $S(n,k)$; distinct boxes ⇒ use $k!\,S(n,k) = \mathrm{Surj}(n,k)$. If empty boxes are allowed, sum over the number of nonempty boxes: distinct boxes give k^n, identical boxes give $\sum_{j=1}^{k} S(n,j)$.

6.9 Derangement Asymptotics: The Nearest–Integer Rule

D_n Is the Nearest Integer to $n!/e$

The Maclaurin series $e^{-1} = \sum_{k\geq 0} \frac{(-1)^k}{k!}$ truncated at $k=n$ is exactly $D_n/n!$. The leftover tail is tiny, so

$$\boxed{D_n = \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor \quad (n \geq 1), \qquad D_n = \text{nearest integer to } \frac{n!}{e}.}$$

Bounding the alternating tail gives $\left| D_n - \frac{n!}{e} \right| = \frac{n!}{e} \left| \sum_{k>n} \frac{(-1)^k}{k!} \right| < \dfrac{1}{n+1} \leq \frac{1}{2}$ for $n \geq 1$, which is why rounding is always correct. Consequently the probability that a random permutation of n objects is a derangement,

$$\frac{D_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \longrightarrow \frac{1}{e} \approx 0.367879,$$

converges extraordinarily fast — already at $n=6$ we have $D_6/6! = 265/720 = 0.368055\ldots$, agreeing with $1/e$ to three decimals.

Example: A Fast Derangement Count

Compute D_7 without the full alternating sum. Since $7! = 5040$ and $e \approx 2.718281828$,

$$\frac{7!}{e} = \frac{5040}{2.718281828} \approx 1854.11, \qquad D_7 = \lfloor 1854.11 + 0.5 \rfloor = \mathbf{1854}.$$

Cross-check with $D_n = nD_{n-1} + (-1)^n$: $D_7 = 7\,D_6 - 1 = 7 \cdot 265 - 1 = 1855 - 1 = 1854$.

Contest shortcut: to get D_n for larger n, apply $D_n = nD_{n-1} + (-1)^n$ starting from $D_1 = 0$; the "± 1" alternates. The nearest-integer fact is the fast estimate, the recurrence is the exact check.

6.10 Exactly m Sets: The Bonferroni Refinement

Counting Elements in Exactly (or At Least) m Sets

Let $S_j = \sum_{|T|=j} \left| \bigcap_{i \in T} A_i \right|$ be the j-th symmetric sum of intersection sizes ($S_0 = |U|$). The number of elements lying in **exactly** m of the sets $A_1, \ldots, A_n$ is

$$\boxed{E_m = \sum_{j=m}^{n} (-1)^{j-m} \binom{j}{m} S_j,}$$

and the number in **at least** m of the sets is

$$L_m = \sum_{j=m}^{n} (-1)^{j-m} \binom{j-1}{m-1} S_j.$$

Setting $m = 0$ in E_m gives the "none" formula; $m = 1$ in L_m gives the ordinary union. **Bonferroni inequalities:** truncating the alternating series after an *odd* number of terms over-estimates $|A_1 \cup \cdots \cup A_n|$ and after an *even* number under-estimates it, so partial sums bracket the true count.

Example: Exactly One Own Hat

For the hat–check problem with $n = 5$ people, how many return arrangements give *exactly* 2 people their own hat? Using the "exactly m" idea for fixed points (choose the fixed pair, derange the rest):

$$\binom{5}{2} D_3 = 10 \cdot 2 = \mathbf{20}.$$

As a check that the counts partition $5! = 120$: $\sum_{m=0}^{5} \binom{5}{m} D_{5-m} = 44 + 5 \cdot 9 + 10 \cdot 2 + 10 \cdot 1 + 5 \cdot 0 + 1 = 44 + 45 + 20 + 10 + 0 + 1 = 120$.

Example: In Exactly Two of Three (AIME flavor)

Of integers 1 to 100, how many are divisible by exactly two of $2, 3, 5$? Here $S_2 = \lfloor 100/6 \rfloor + \lfloor 100/10 \rfloor + \lfloor 100/15 \rfloor = 16 + 10 + 6 = 32$ and $S_3 = \lfloor 100/30 \rfloor = 3$. Then

$$E_2 = \binom{2}{2} S_2 - \binom{3}{2} S_3 = 32 - 3 \cdot 3 = 32 - 9 = \mathbf{23}.$$

Why the binomials appear: an element in exactly j sets is counted $\binom{j}{m}$ times inside S_j; the alternating $\binom{j}{m}$-weighted sum cancels all $j \neq m$ and leaves each "exactly-m" element counted once. Memorize the two coefficients: $\binom{j}{m}$ for *exactly*, $\binom{j-1}{m-1}$ for *at least*.

6.11 Ramsey Numbers & Probabilistic Pigeonhole

Order Forced by Size: $R(3,3) = 6$

Color the edges of a complete graph on n vertices with two colors (say, *acquainted* / *stranger*).

A **monochromatic triangle** is three mutually-acquainted or three mutual strangers. The **Ramsey number** $R(3,3)$ is the least n forcing one to exist.

$R(3,3) \le 6$**:** Fix a vertex v among 6. Its 5 edges get 2 colors, so by pigeonhole ($\lceil 5/2 \rceil = 3$) at least 3 share a color — say v is acquainted with a, b, c. If any edge among a, b, c is "acquainted," that edge with v forms an acquainted triangle; otherwise a, b, c are mutual strangers — a stranger triangle. Either way a monochromatic triangle appears.

$R(3,3) > 5$**:** color the 5-cycle's edges red and its "diagonals" blue; neither color contains a triangle. Hence $\boxed{R(3,3) = 6}$: any party of 6 has 3 mutual acquaintances or 3 mutual strangers.

The pentagon (solid) and pentagram (dashed) each avoid a triangle, proving 5 vertices are not enough.

Example: An Averaging (Probabilistic) Pigeonhole

100 people each rate a film from 1 to 10; the ratings sum to **620**. Show some rating value was given by at least **62** people is *false* in general, but the averaging principle still bites: the *mean* rating is $620/100 = 6.2$, so **some** person rated it at least $\lceil 6.2 \rceil = 7$ and some rated it at most 6. More usefully, distributing 620 "rating-points" among 10 possible values, some value carries at least $\lceil 620/10 \rceil$? No — points are not people. The correct averaging statement: with 100 people and 10 scores, some score was chosen by at least $\lceil 100/10 \rceil = \mathbf{10}$ people.

Example: Existence Beats the Average (Olympiad flavor)

Each of 17 scientists corresponds with every other on one of 3 topics. Prove some three of them discuss a single common topic (i.e. $R(3,3,3) \le 17$). Fix a scientist v; its 16 edges use 3 topics, so by pigeonhole $\lceil 16/3 \rceil = 6$ edges share a topic T — say to the group G of ≥ 6 people. If any pair inside G also uses topic T, they and v form a T–triangle. Otherwise G uses only the other 2 topics, and a 6-vertex 2-colored graph contains a monochromatic triangle by $R(3,3) = 6$. Either way a monochromatic triangle exists.

The averaging principle: in any finite collection, *some element is at least the average and some is at most the average.* Combined with pigeonhole ($\lceil n/k \rceil$ per box) and Ramsey-style edge-coloring, this proves a desired structure **must exist** without exhibiting it. Ramsey's slogan: *complete disorder is impossible* — enough size forces order.

7 Recursion, Catalan Numbers & Generating Functions

Competition counting & probability, Unit 6

7.1 Setting Up Recurrences

The core idea

A **recurrence** counts a large problem in terms of smaller versions of itself. To build one:

1. **Define the sequence precisely.** Let a_n be the number of ways to do the task of "size" n. Being sloppy here is the #1 source of errors — write a one-sentence definition.
2. **Condition on a first (or last) choice.** Look at the object and split into cases based on one feature: the first tile, the last step, whether an element is used, the height of the last column, etc.
3. **Express each case using smaller a_k.** Each case should leave a strictly smaller subproblem of the *same* type, so it is counted by a_k for some $k < n$.
4. **Nail the base cases.** Determine a_0 (or a_1) directly by hand. Count the "empty" object carefully: usually $a_0 = 1$.

The recurrence plus base cases determines every term.

Worked example — ways to make a total of n with 1s and 2s

Let a_n be the number of ordered sequences of 1s and 2s summing to n (order matters). Condition on the *first* number in the sequence:

- If it is a 1, the rest sums to $n-1$: that is a_{n-1} ways.
- If it is a 2, the rest sums to $n-2$: that is a_{n-2} ways.

So $a_n = a_{n-1} + a_{n-2}$. Base cases: $a_0 = 1$ (the empty sequence) and $a_1 = 1$. This gives $1, 1, 2, 3, 5, 8, \ldots$ — the Fibonacci numbers.

Tips. (1) Prefer conditioning on the *first* or *last* element — it usually leaves a clean subproblem. (2) Always sanity-check by hand-counting a_1, a_2, a_3 against your formula. (3) Decide up front whether order matters and whether a_0 should be 1 or 0.

7.2 Fibonacci & Tiling / Domino Problems

Fibonacci from tilings

The **Fibonacci numbers** satisfy

$$f_n = f_{n-1} + f_{n-2}, \qquad f_1 = f_2 = 1,$$

giving $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$ Many tiling problems reduce to this.

Square–domino tilings. Let t_n be the number of ways to tile a $1 \times n$ strip using 1×1 squares and 1×2 dominoes. Look at the tile covering the *last* cell: a square (leaving a $1 \times (n-1)$

strip) or a domino (leaving a $1 \times (n-2)$ strip). So $t_n = t_{n-1} + t_{n-2}$ with $t_0 = 1,\ t_1 = 1$, i.e. $t_n = f_{n+1}$.

Worked example — tiling a $2 \times n$ board with dominoes

Let D_n be the number of ways to tile a $2 \times n$ board with 1×2 dominoes. Consider the leftmost column:

- One *vertical* domino fills the column, leaving a $2 \times (n-1)$ board: D_{n-1} ways.
- Two *horizontal* dominoes fill the first two columns, leaving a $2 \times (n-2)$ board: D_{n-2} ways.

Hence $D_n = D_{n-1} + D_{n-2}$, with $D_0 = 1$, $D_1 = 1$. So $D_n = f_{n+1}$; e.g. a 2×4 board has $D_4 = 5$ tilings.

Tips. The recurrence comes from "what covers the boundary cell?" Different tile sets change the recurrence: squares + dominoes + trominoes on a strip give $t_n = t_{n-1} + t_{n-2} + t_{n-3}$. Watch indexing: confirm which Fibonacci index your base cases land on.

7.3 State-Based Counting Recursions

Counting with states

When the next choice depends on the *current situation*, define one sequence per **state** and let them recurse together. Steps:

1. Identify the finite information (the "state") you must remember to extend a valid object by one step.
2. Define $a_n, b_n, \ldots$ = number of length-n objects ending in each state.
3. Write transition equations: each state's next value is a sum over states that may legally precede it.
4. The answer is usually the sum of all states at length n.

This is the "transfer matrix" / finite-automaton viewpoint.

Worked example — binary strings with no two consecutive 1s

Count length-n binary strings with no "11". Track the last bit. Let A_n = number ending in 0, B_n = number ending in 1.

- A string ending in 0 can be preceded by anything: $A_n = A_{n-1} + B_{n-1}$.
- A string ending in 1 must have 0 before it: $B_n = A_{n-1}$.

With $A_1 = 1, B_1 = 1$, the total $T_n = A_n + B_n$ satisfies $T_n = T_{n-1} + T_{n-2}$ (Fibonacci again!), giving $T_1 = 2, T_2 = 3, T_3 = 5, \ldots$

Tips. Keep the state as small as possible — only remember what the next transition needs. If a constraint says "no three in a row," your state is the length of the current run, so you may need two or three coupled sequences.

7.4 Catalan Numbers

The Catalan numbers

The **Catalan numbers** are

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \binom{2n}{n} - \binom{2n}{n+1}, \qquad C_{n+1} = \sum_{i=0}^{n} C_i\, C_{n-i}.$$

First values:

n	0	1	2	3	4	5	6
C_n	1	1	2	5	14	42	132

They count an astonishing number of things. Three canonical interpretations:

- **Lattice paths / ballots:** monotonic paths from $(0,0)$ to (n,n) using steps right/up that never cross above the diagonal $y = x$. Equivalently, sequences of n $+1$s and n -1s with every partial sum ≥ 0.
- **Balanced parentheses:** the number of ways to correctly match n pairs of parentheses (every prefix has $\#(\geq \#)$).
- **Triangulations:** the number of ways to cut a convex $(n+2)$-gon into triangles using non-crossing diagonals.

Other appearances: binary trees with n internal nodes, non-crossing handshakes among $2n$ people, mountain ranges, and stack-sortable permutations.

Worked example — parenthesizations and the convolution

How many ways to fully parenthesize a product of 4 factors, e.g. $a \cdot b \cdot c \cdot d$? This is $C_3 = 5$:

$$((ab)c)d, \quad (a(bc))d, \quad (ab)(cd), \quad a((bc)d), \quad a(b(cd)).$$

Why $C_{n+1} = \sum C_i C_{n-i}$? In a balanced string, the first "(" closes at some point, splitting the string into an inner balanced block (size i) and a following balanced block (size $n-i$). Summing over the split point gives the convolution. This same split proves the triangulation and binary-tree counts.

Tips. If a counting problem involves "never falls behind," "non-crossing," or "a valid bracket/matching structure," suspect Catalan. Memorize C_0 through C_5: $1, 1, 2, 5, 14, 42$. The ratio $C_{n+1}/C_n = \frac{2(2n+1)}{n+2}$ makes them fast to extend.

7.5 Generating Functions (introduction)

Series as bookkeeping devices

A **(ordinary) generating function** packages a sequence $a_0, a_1, a_2, \ldots$ into one formal power series

$$A(x) = \sum_{n\geq 0} a_n x^n.$$

The exponent of x is a "label" tracking a total (size, weight, sum), and the coefficient counts objects with that total. The magic:

$$\textbf{choosing independently} \longleftrightarrow \textbf{multiplying generating functions.}$$

If $A(x)$ counts choices for part 1 (weighted by size) and $B(x)$ counts part 2, then the product $A(x)B(x)$ has x^n-coefficient $\sum_k a_k b_{n-k}$ — exactly the number of ways to split total n between the two parts. Useful building blocks:

$$\frac{1}{1-x} = \sum_{n\geq 0} x^n, \qquad \frac{1}{1-x^k} = \sum_{j\geq 0} x^{jk}, \qquad (1+x)^m = \sum_n \binom{m}{n} x^n.$$

Worked example — coins as a product of GFs

In how many ways can you make n cents using pennies (1c), nickels (5c), and dimes (10c), unlimited supply, order irrelevant? Each coin type contributes a geometric factor tracking how much value it supplies:

$$G(x) = \underbrace{\frac{1}{1-x}}_{\text{pennies}} \cdot \underbrace{\frac{1}{1-x^5}}_{\text{nickels}} \cdot \underbrace{\frac{1}{1-x^{10}}}_{\text{dimes}}.$$

The coefficient of x^n in $G(x)$ is the answer. For instance $[x^{10}]G(x) = 4$: 10 pennies; 5+5 (nickels); a nickel +5 pennies; one dime. Products of GFs automatically handle "distribute the total among independent parts."

Tips. Treat x as a formal symbol — convergence does not matter. "Choose one of several" becomes a *sum* of terms; "do several things independently" becomes a *product* of series. To read off a coefficient, use $\frac{1}{(1-x)^k} = \sum_n \binom{n+k-1}{k-1} x^n$ or partial fractions.

7.6 Solving Simple Linear Recurrences

The characteristic-equation method

For a linear recurrence with constant coefficients

$$a_n = c_1 a_{n-1} + c_2 a_{n-2},$$

guess $a_n = r^n$. Substituting and dividing by r^{n-2} gives the **characteristic equation**

$$r^2 = c_1 r + c_2.$$

- **Distinct roots** $r_1 \neq r_2$: the general solution is $a_n = \alpha\, r_1^n + \beta\, r_2^n$.
- **Repeated root** r: use $a_n = (\alpha + \beta n)\, r^{\,n}$.

Fix α, β from two initial values. The same idea extends to order k: a degree-k characteristic polynomial with k roots.

Worked example — solving $a_n = 5a_{n-1} - 6a_{n-2}$

Given $a_0 = 0$, $a_1 = 1$. Characteristic equation: $r^2 = 5r - 6$, i.e. $r^2 - 5r + 6 = 0$, so $(r-2)(r-3) = 0$ and $r = 2, 3$. Thus

$$a_n = \alpha\, 2^n + \beta\, 3^n.$$

Initial conditions: $a_0 = \alpha + \beta = 0$ and $a_1 = 2\alpha + 3\beta = 1$. Solving, $\beta = 1$, $\alpha = -1$, so

$$a_n = 3^n - 2^n.$$

Check: $a_2 = 9 - 4 = 5 = 5(1) - 6(0)$. ✓

Tips. Bring everything to one side so the characteristic polynomial is clear. Complex roots are fine — they produce oscillating solutions. For a *nonhomogeneous* recurrence ($a_n = c_1 a_{n-1} + \cdots + g(n)$), add a particular solution (try a constant for constant g, a linear guess for linear g) to the homogeneous solution.

7.7 Going Deeper: Closed Forms and Proofs

Three derivations worth knowing

(A) The Catalan closed form by reflection. Count lattice paths from $(0,0)$ to (n,n) (steps R and U) that stay weakly below the diagonal. Total paths: $\binom{2n}{n}$. A *bad* path touches the line $y = x + 1$. Reflect the portion of a bad path *after* its first touch across $y = x + 1$: this bijects bad paths with *all* paths from $(0,0)$ to $(n-1,n+1)$, of which there are $\binom{2n}{n+1}$. Hence

$$C_n = \binom{2n}{n} - \binom{2n}{n+1} = \frac{1}{n+1}\binom{2n}{n}.$$

(B) Solving a recurrence with a generating function. Let $C(x) = \sum_{n\geq 0} C_n x^n$. The convolution $C_{n+1} = \sum_i C_i C_{n-i}$ says $C(x) = 1 + xC(x)^2$. Solving the quadratic,

$$C(x) = \frac{1-\sqrt{1-4x}}{2x},$$

and expanding $\sqrt{1-4x}$ by the binomial series recovers $C_n = \frac{1}{n+1}\binom{2n}{n}$. This "turn the recurrence into an equation for $A(x)$, solve, expand" pipeline is the general method.

(C) Binet's formula for Fibonacci. The characteristic equation $r^2 = r+1$ has roots $\varphi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$. Then

$$f_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}.$$

Since $|\psi| < 1$, f_n is the nearest integer to $\varphi^n/\sqrt{5}$, so Fibonacci numbers grow like φ^n.

Worked example — coefficients from $\dfrac{1}{1-x-x^2}$

The Fibonacci GF is $F(x) = \sum f_n x^n = \dfrac{x}{1-x-x^2}$. Factor the denominator using φ, ψ and split by partial fractions:

$$F(x) = \frac{1}{\sqrt{5}}\left(\frac{1}{1-\varphi x} - \frac{1}{1-\psi x}\right) = \frac{1}{\sqrt{5}}\sum_{n\geq 0}(\varphi^n - \psi^n)x^n,$$

which reads off Binet's formula directly. Partial fractions on $1/(1-x-x^2)$-type denominators is the bridge from recurrence to closed form.

Big picture. Recursion, Catalan numbers, and generating functions are one connected toolkit. *Recursion* is how you *discover* a count by conditioning on a first/last choice or a state. *Generating functions* are how you *package and solve* those recurrences — products encode independent choices, and solving an algebraic equation for $A(x)$ yields closed forms. *Catalan numbers* are the flagship example where all three views (a convolution recurrence, a reflection/bijection argument, and a GF quadratic) meet in one answer. When you see "count in terms of smaller cases," set up a_n; when you see "combine independent totals," multiply GFs; when you see "non-crossing / balanced / never-behind," reach for Catalan.

7.8 Nonhomogeneous & Higher-Order Linear Recurrences

Particular + homogeneous, with resonance

A linear recurrence with a forcing term,

$$a_n = c_1 a_{n-1} + \cdots + c_k a_{n-k} + g(n),$$

is solved by

$$a_n \;=\; a_n^{(h)} \;+\; a_n^{(p)},$$

where $a_n^{(h)}$ is the general solution of the *homogeneous* part (characteristic roots, as before) and $a_n^{(p)}$ is *any one* particular solution. To find $a_n^{(p)}$, guess a form matching $g(n)$ and solve for its constants:

$g(n)$	try $a_n^{(p)} =$
constant	A
polynomial of degree d	$A_d n^d + \cdots + A_1 n + A_0$
exponential b^n	$A\, b^n$

Resonance (the crucial trap). If your guess already solves the homogeneous equation — e.g. $g(n) = b^n$ but b is a characteristic root, or $g(n)$ is constant but $r = 1$ is a root — multiply the guess by n (by n^m if b is a root of multiplicity m). This is the same "multiply by n for a repeated root" rule seen in the homogeneous case.

Worked example — a nonhomogeneous recurrence with resonance

Solve $a_n = 3a_{n-1} - 2a_{n-2} + 2^n$ with $a_0 = 0,\ a_1 = 0$.

Homogeneous part: $r^2 - 3r + 2 = (r-1)(r-2) = 0$, so $r = 1, 2$ and $a_n^{(h)} = \alpha \cdot 1^n + \beta \cdot 2^n$.

Particular part: the natural guess $A\,2^n$ is **resonant** because $r = 2$ is already a root. Multiply by n: try $a_n^{(p)} = A\,n\,2^n$. Substituting,

$$An2^n = 3A(n-1)2^{n-1} - 2A(n-2)2^{n-2} + 2^n.$$

Divide by 2^{n-2}: $4An = 6A(n-1) - 2A(n-2) + 4$, i.e. $4An = 4An - 2A + 4$, giving $A = 2$. So $a_n^{(p)} = 2n\,2^n = n\,2^{n+1}$.

Combine and fit: $a_n = \alpha + \beta\,2^n + n\,2^{n+1}$. From $a_0 = \alpha + \beta = 0$ and $a_1 = \alpha + 2\beta + 4 = 0$ we get $\beta = -4,\ \alpha = 4$. Thus

$$a_n = 4 - 4 \cdot 2^n + n\,2^{n+1} = 4 - 2^{n+2} + n\,2^{n+1}.$$

Check: $a_2 = 4 - 16 + 16 = 4$, and $3a_1 - 2a_0 + 2^2 = 0 - 0 + 4 = 4$. ✓

Tips. (1) Always solve the homogeneous roots *first* — you cannot detect resonance until you know them. (2) A constant forcing term with a root $r = 1$ needs the guess An (not A). (3) If $g(n)$ is a sum of pieces, find a particular solution for each piece separately and add. (4) Fit the constants α, β using the *full* solution $a_n^{(h)} + a_n^{(p)}$, never the homogeneous part alone.

7.9 The Transfer-Matrix Method

Recurrences as matrix powers

A coupled linear recurrence among finitely many states is one **matrix** T acting on a state vector. If $\mathbf{v}_n$ collects the state values at step n and $\mathbf{v}_n = T\mathbf{v}_{n-1}$, then

$$\mathbf{v}_n = T^n \mathbf{v}_0.$$

For the Fibonacci recurrence, taking $\mathbf{v}_n = \begin{bmatrix} f_{n+1} \\ f_n \end{bmatrix}$,

$$\begin{bmatrix} f_{n+1} \\ f_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_n \\ f_{n-1} \end{bmatrix}, \qquad \begin{bmatrix} f_{n+1} & f_n \\ f_n & f_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n.$$

The matrix's **eigenvalues are exactly the characteristic roots** φ, ψ, which is why T^n (hence every entry) is a combination of φ^n and ψ^n — another route to Binet's formula. Taking determinants of the boxed identity gives $f_{n+1}f_{n-1} - f_n^2 = (-1)^n$ instantly (Cassini's identity).

Counting walks. If M is the adjacency matrix of a (multi)graph, then $(M^n)_{ij}$ counts walks of length n from vertex i to j, and $\mathrm{tr}(M^n) = \sum_k \lambda_k^n$ counts closed walks. This packages "count paths through states" as matrix powers.

Worked example — walks on a triangle

Three vertices $\{1, 2, 3\}$ of a triangle are all mutually adjacent. How many closed walks of length n start and end at vertex 1? The adjacency matrix is $M = J - I$ where J is all-ones:

$$M = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Its eigenvalues: J has eigenvalues $3, 0, 0$, so $M = J - I$ has eigenvalues $2, -1, -1$. By symmetry every diagonal entry of M^n is equal, and $\mathrm{tr}(M^n) = 2^n + 2(-1)^n$, so

$$(M^n)_{11} = \frac{2^n + 2(-1)^n}{3}.$$

Check $n = 2$: closed walks $1 \to k \to 1$ number 2, and $\frac{4+2}{3} = 2$. ✓ For $n = 3$ we get $\frac{8-2}{3} = 2$, the two directed triangles $1 \to 2 \to 3 \to 1$ and $1 \to 3 \to 2 \to 1$. This is the standard AIME-style "ant walks on a solid/graph and returns home" setup.

Tips. (1) Order-k scalar recurrences become $k \times k$ *companion* matrices whose characteristic polynomial is the recurrence's. (2) For symmetric transition structures, diagonalizing via eigenvalues turns T^n into a clean closed form — look for an all-ones or circulant pattern. (3) $\mathrm{tr}(M^n) = \sum \lambda^n$ and det identities (Cassini) fall out for free. (4) On a highly symmetric graph, "lump" equivalent vertices into one state to shrink the matrix.

7.10 The Cycle Lemma & the Ballot Problem

Rotation counting and the general ballot theorem

The reflection principle is one proof of the Catalan/ballot counts; the **cycle lemma** is a slicker, purely combinatorial one.

Cycle Lemma (Dvoretzky–Motzkin). Take any sequence of n steps that are $+1$'s and -1's summing to $+1$ (so $\frac{n+1}{2}$ ups and $\frac{n-1}{2}$ downs). Among its n cyclic rotations, *exactly one* has all partial sums strictly positive.

Ballot Theorem. In an election where A gets a votes and B gets $b < a$ votes, the number of orderings of the ballots in which A is *strictly* ahead throughout is

$$\frac{a-b}{a+b}\binom{a+b}{a}.$$

Catalan is the boundary case: paths from $(0,0)$ to (n,n) staying strictly below the diagonal except at the ends correspond to $a = n+1,\ b = n$ shifted, and the ballot formula collapses to $C_n = \frac{1}{n+1}\binom{2n}{n}$.

Worked example — counting good arrangements by rotation

In how many ways can $n+1$ people holding a \$5 bill and n people holding a \$10 bill line up at a box office charging \$5, starting with an empty till, so the cashier can *always* give change?

Model each \$5 as $+1$ and each \$10 as -1. There are $2n+1$ people; the partial sums must stay ≥ 0 (in fact > 0 after the first person once we require never getting stuck), and the total is $(n+1) - n = +1$. Any arrangement of the $2n+1$ steps summing to $+1$ has, by the **cycle lemma**, exactly one of its $2n+1$ rotations with all partial sums positive. Since the total number of step-sequences is $\binom{2n+1}{n}$ and they group into rotation-classes each contributing exactly one good sequence,

$$\#\{\text{good lines}\} = \frac{1}{2n+1}\binom{2n+1}{n} = C_n.$$

For $n = 2$ (\$5,\$5,\$5 and \$10,\$10): $\frac{1}{5}\binom{5}{2} = 2 = C_2$. ✓ The one-good-rotation-per-class argument avoids any reflection bijection entirely.

Tips. (1) The cycle lemma also proves the **Fuss–Catalan** numbers $\frac{1}{kn+1}\binom{kn+1}{n}$ (paths with up-steps $+1$ and down-steps $-(k-1)$) — same one-good-rotation idea. (2) "Strictly ahead" vs. "never behind" shifts an index by one; recount a tiny case to pin it down. (3) When a problem says "running total never goes negative" *and* you can cyclically shift without changing the count, reach for the cycle lemma instead of reflection.

7.11 Exponential Generating Functions

EGFs for labeled structures

When objects are built on **labeled** elements (people, positions $1..n$ that are distinguishable), the right bookkeeping tool is the **exponential generating function**

$$\widehat{A}(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}.$$

The key rule mirrors the ordinary case but with a binomial-flavored product:

$$\widehat{A}(x)\,\widehat{B}(x) = \sum_{n\geq 0} \left(\sum_k \binom{n}{k} a_k\, b_{n-k} \right) \frac{x^n}{n!}.$$

The $\binom{n}{k}$ appears because you first *choose which labels* go to the A-part. Building blocks:

$$e^x = \sum_{n\geq 0} \frac{x^n}{n!} \text{ (one labeled set)}, \qquad \frac{1}{1-x} = \sum_{n\geq 0} n!\,\frac{x^n}{n!} \text{ (a linear order)}.$$

The **Exponential Formula** $\widehat{A}(x) = e^{\widehat{C}(x)}$ passes from “connected” pieces $\widehat{C}$ to arbitrary disjoint unions of them (set partitions into blocks, permutations into cycles, graphs into components).

Worked example — derangements via an EGF

A **derangement** is a permutation with no fixed point; let D_n be their count. Every permutation splits its n labels into a set of fixed points and a derangement of the rest:

$$n! = \sum_{k=0}^{n} \binom{n}{k} D_{n-k} \cdot 1.$$

In EGF language this is a product: (all permutations) = (choose fixed points) × (derange the rest), i.e.

$$\frac{1}{1-x} = e^x \cdot \widehat{D}(x) \quad \Longrightarrow \quad \widehat{D}(x) = \frac{e^{-x}}{1-x}.$$

Reading off the coefficient of $x^n/n!$,

$$D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!},$$

the familiar inclusion–exclusion formula, and D_n is the nearest integer to $n!/e$. From $\widehat{D}(x) = e^{-x}/(1-x)$ one also reads the recurrences $D_n = nD_{n-1} + (-1)^n$ and $D_n = (n-1)(D_{n-1} + D_{n-2})$.

Big picture. Use an **ordinary** GF $\sum a_n x^n$ when parts combine by "split the total" (unlabeled: coins, compositions, tilings); use an **exponential** GF $\sum a_n x^n/n!$ when parts combine by "distribute distinct labels" (labeled: permutations, set partitions, surjections). The dictionary — product = independent combination, $e^{\widehat{C}}$ = disjoint components — lets you write the answer's generating function directly from a structural description, then extract a_n by expanding.

8 Probability Fundamentals & Geometric Probability

Competition counting & probability, Unit 7

8.1 Sample Spaces & Equally Likely Outcomes

What we are counting

An **experiment** is any process with an uncertain result. The **sample space** S is the set of *all* possible outcomes, and an **event** E is any subset of S (a collection of outcomes we care about).

When every outcome in S is **equally likely**, probability becomes pure counting:

$$P(E) = \frac{\text{number of favorable outcomes}}{\text{number of total outcomes}} = \frac{|E|}{|S|}.$$

The single most important habit is to build a sample space in which the outcomes really *are* equally likely — otherwise the formula does not apply.

Worked example: two dice, equally likely outcomes

Two fair dice are rolled. What is the sample space, and how likely is the sum 7?

Solution. List *ordered* pairs (a, b) so each is equally likely: $|S| = 6 \cdot 6 = 36$. The sum is 7 for $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$, so $|E| = 6$ and

$$P(\text{sum} = 7) = \frac{6}{36} = \frac{1}{6}.$$

If instead we had used the 11 possible sums $2, 3, \ldots, 12$ as our sample space, those outcomes are *not* equally likely, and the counting formula would give a wrong answer.

Pitfall. "Number of things that can happen" is only a valid denominator when those things are equally likely. Prefer ordered, atomic outcomes (each coin, each die, each draw distinguished) to force equal likelihood.

8.2 Probability as Favorable over Total (Counting-Based)

Probability is a counting ratio

Once outcomes are equally likely, every probability question is two counting questions: count $|E|$ (favorable) and count $|S|$ (total), then divide. Basic properties follow immediately:

$$0 \leq P(E) \leq 1, \qquad P(\varnothing) = 0, \qquad P(S) = 1.$$

For **disjoint** (mutually exclusive) events, probabilities add: $P(A \cup B) = P(A) + P(B)$. In general, inclusion–exclusion gives $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Worked example: drawing a card

One card is drawn from a standard 52-card deck. Find P(a face card or a heart).

Solution. Face cards: 12. Hearts: 13. Overlap (face-card hearts): 3. By inclusion–exclusion the favorable count is $12 + 13 - 3 = 22$, so

$$P = \frac{22}{52} = \frac{11}{26}.$$

Worked example: reduce every answer

A jar has 4 red and 6 blue marbles. One marble is drawn. Find P(red).

Solution. $P(\text{red}) = \dfrac{4}{10} = \dfrac{2}{5}$. Contest answers are almost always expected as *reduced* fractions.

Tip. Keep the same "kind" of object in numerator and denominator. If you count the total as ordered outcomes, count the favorable as ordered too; if unordered, unordered. Mixing the two is the classic source of wrong probabilities.

8.3 Complementary Probability

Count the opposite instead

The **complement** E^c is the event "E does not happen." Since E and E^c partition S,

$$P(E) + P(E^c) = 1 \qquad \Longrightarrow \qquad P(E) = 1 - P(E^c).$$

Reach for the complement whenever E is described by "at least one," "at least," or "not all" — the opposite event ("none," "exactly zero") is usually far easier to count.

Worked example: at least one head

A fair coin is flipped 5 times. Find $P(\text{at least one head})$.

Solution. The complement is "no heads," i.e. all 5 tails, with probability $\left(\frac{1}{2}\right)^5 = \frac{1}{32}$. Hence

$$P(\text{at least one head}) = 1 - \frac{1}{32} = \frac{31}{32}.$$

Counting the 5 separate "exactly k heads" cases would be far more work.

Worked example: shared birthday flavor

Three people each pick a day of the week at random. Find $P(\text{at least two pick the same day})$.

Solution. Complement: all three days distinct. Total $= 7^3 = 343$; all-distinct $= 7 \cdot 6 \cdot 5 = 210$. So

$$P(\text{some match}) = 1 - \frac{210}{343} = 1 - \frac{30}{49} = \frac{19}{49}.$$

Tip. "At least one" $\Rightarrow$ subtract "none." The complement trick converts an OR of many cases into a single, clean count.

8.4 Probability with Permutations & Combinations

Counting tools power the ratio

For selection problems, $|S|$ and $|E|$ are themselves permutation or combination counts. Decide once whether order matters and use the *same* convention on top and bottom:

$$P(E) = \frac{\text{favorable arrangements/selections}}{\text{total arrangements/selections}}, \qquad \binom{n}{k} = \frac{n!}{k!\,(n-k)!}, \qquad P(n,k) = \frac{n!}{(n-k)!}.$$

For committees, hands, and unordered draws, $\binom{}{}$ appears in both numerator and denominator.

Worked example: a committee count

A club of 6 boys and 4 girls chooses a 3-person committee at random. Find $P(\text{exactly 2 girls})$.

Solution. Total committees: $\binom{10}{3} = 120$. Favorable: choose 2 of 4 girls *and* 1 of 6 boys, $\binom{4}{2}\binom{6}{1} = 6 \cdot 6 = 36$. So

$$P = \frac{36}{120} = \frac{3}{10}.$$

Worked example: a poker-style hand

Five cards are dealt from a standard deck. Find $P(\text{all five are hearts})$.

Solution. Total 5-card hands: $\binom{52}{5}$. All-heart hands: $\binom{13}{5}$. So

$$P = \frac{\binom{13}{5}}{\binom{52}{5}} = \frac{1287}{2{,}598{,}960} = \frac{33}{66640}.$$

Order is irrelevant to a hand, so combinations appear top and bottom.

Pitfall. Never mix ordered and unordered counts. Using $\binom{}{}$ on top but $P(n, k)$ (or $n!$) on the bottom silently multiplies your answer by a stray factorial.

8.5 Independent vs. Dependent Events (Basic)

Does one event change the other?

Events A and B are **independent** if knowing one occurred does not change the probability of the other. Then

$$P(A \cap B) = P(A)\,P(B).$$

If the events are **dependent**, use the multiplication rule with a conditional probability:

$$P(A \cap B) = P(A) \cdot P(B \mid A),$$

where $P(B \mid A)$ is computed *after* A has happened. Drawing **with replacement** keeps events independent; drawing **without replacement** makes them dependent.

Worked example: independent (with replacement)

A bag has 3 red and 5 green marbles. Two are drawn *with* replacement. Find $P(\text{both red})$.

Solution. Replacement restores the bag, so the draws are independent, each with $P(\text{red}) = \frac{3}{8}$:

$$P(\text{both red}) = \frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64}.$$

Worked example: dependent (without replacement)

Same bag (3 red, 5 green), but the two marbles are drawn *without* replacement. Find $P(\text{both red})$.

Solution. The second draw depends on the first:

$$P(\text{both red}) = \frac{3}{8} \cdot \frac{2}{7} = \frac{6}{56} = \frac{3}{28}.$$

Fewer red marbles and fewer total marbles remain, so the second factor shrinks.

Tip. "With replacement" $\Rightarrow$ independent $\Rightarrow$ multiply unchanged probabilities. "Without replacement" $\Rightarrow$ dependent $\Rightarrow$ update the counts after each draw. When in doubt, count the whole thing with combinations instead.

8.6 Geometric Probability (Length & Area)

Probability as a ratio of measures

When an outcome is a *uniformly random* point in a region — a spot on a segment, an instant in a time interval, a dart in a target — the sample space is continuous and we replace counting with **measure**:

$$P(E) = \frac{\text{length of favorable interval}}{\text{length of whole interval}} \quad \text{or} \quad \frac{\text{area of favorable region}}{\text{area of whole region}}.$$

The idea is identical to $\frac{|E|}{|S|}$: favorable "size" over total "size," now measured by length or area instead of by count.

Worked example: dart in a square

A dart lands at a uniformly random point in a 4×4 square. Find the probability it lands inside the inscribed circle of radius 2 (figure above).

Solution. Favorable area $= \pi(2)^2 = 4\pi$; total area $= 16$. So

$$P = \frac{4\pi}{16} = \frac{\pi}{4} \approx 0.785.$$

Geometric probabilities need not be rational — they inherit whatever constants (π, radicals) the geometry contains.

Worked example: a random point on a segment

A point X is chosen uniformly on the segment from 0 to 12. Find $P(X > 9)$.

Solution. Favorable length $= 12 - 9 = 3$; total length $= 12$. So $P = \dfrac{3}{12} = \dfrac{1}{4}$.

Tip. Match dimensions: a one-variable (single random point) problem uses *length*; a two-variable (two independent random points, or a point in the plane) problem uses *area*. Sketch the favorable region before computing.

8.7 Going Deeper: Why Probability Is Counting, and the Meeting Problem

Equally likely outcomes make probability into counting

The formula $P(E) = |E|/|S|$ is not a definition of probability in general — it is a *theorem* about the special case of equally likely outcomes. Each of the $|S|$ atomic outcomes carries probability $\frac{1}{|S|}$, and an event's probability is the sum of the probabilities of the outcomes it contains, which is exactly $|E| \cdot \frac{1}{|S|}$. This is why all of Units 1–6 (permutations, combinations, casework, complementary counting, inclusion–exclusion) are really *probability* tools in disguise: master the count and the probability is one division away.

Geometric probability is the same statement pushed to a continuum. When outcomes are a uniformly random *point*, no single point can carry positive probability, so instead of counting points we measure events by their *size* — length, area, or volume. "Favorable size over total size" plays exactly the role that "favorable count over total count" played, and uniform randomness is precisely what makes that swap legitimate.

Worked example: the meeting problem (area method)

Two friends each arrive at the café at a uniformly random time between 12:00 and 1:00, independently, and each waits 15 minutes. Find the probability that they meet.

Solution. Let $x, y \in [0, 60]$ be their arrival minutes. The sample space is the 60×60 square, area 3600. They meet exactly when $|x - y| \le 15$, the band between the lines $y = x + 15$ and $y = x - 15$. The two *failure* corners are right triangles with legs $60 - 15 = 45$, so the failure area is $2 \cdot \frac{1}{2} \cdot 45^2 = 45^2 = 2025$. Thus

$$P(\text{meet}) = 1 - \frac{2025}{3600} = 1 - \frac{9}{16} = \frac{7}{16}.$$

Two independent random times ⇒ a point in a square ⇒ *area* ratio. This "two random reals" setup is the signature of area-based geometric probability.

Worked example: uniform points and a broken stick

A stick of length 1 is broken at a uniformly random point. What is the probability the longer piece is at least twice the shorter?

Solution. Let the break be at $X \in [0, 1]$. The longer piece is at least twice the shorter exactly when the shorter piece is at most $\frac{1}{3}$, i.e. $X \le \frac{1}{3}$ or $X \ge \frac{2}{3}$. Favorable length $= \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$, so $P = \dfrac{2}{3}$. One random point ⇒ *length* ratio.

Big picture — one idea, three costumes. Every probability in this unit is "favorable size ÷ total size," and the whole skill is choosing the right notion of *size*:

1. **Finite, equally likely outcomes** ⇒ size = count. Use permutations/combinations for $|E|$ and $|S|$; keep ordered-vs-unordered consistent top and bottom.
2. **"At least one" / messy OR** ⇒ count the complement and subtract from 1.
3. **Sequential draws** ⇒ ask independent (multiply, unchanged) or dependent (multiply with updated counts).
4. **Uniform random point(s)** ⇒ size = length (one variable) or area (two variables). Sketch the favorable region; use the complement when the failure region is the simpler shape.

Get the sample space right — equally likely, or uniform — and every problem collapses to a single ratio.

8.8 Symmetry & Bijection Arguments

Let the problem's symmetry do the counting

Often the cleanest probability argument computes *nothing*: instead you observe that several outcomes must, by symmetry, be equally likely, and read the answer off directly.

- **Random permutations.** If n distinct objects are placed in a uniformly random order, every one of the $n!$ orderings is equally likely. Consequently any statement of the form "object A comes before object B" holds in exactly half of the orderings, so $P(A \text{ before } B) = \frac{1}{2}$ — no matter how many other objects there are.
- **Symmetric roles.** If k specific items are equally likely to occupy any of n positions, then the item in a *fixed* position ("the first card," "the top of the deck") is a uniformly random one of the n items.
- **Bijections.** To show two events are equally likely, exhibit an explicit one-to-one correspondence between their favorable outcomes. Equal-size favorable sets over the same sample space ⇒ equal probability.

The payoff: a symmetry or bijection argument replaces a hard count with a one-line observation.

Worked example: relative order in a shuffle (AIME flavor)

A standard 52-card deck is shuffled uniformly at random. Find the probability that the ace of spades appears somewhere above the king of hearts, and that both appear above the two of clubs.

Solution. Ignore all 49 other cards: by symmetry the three named cards appear among themselves in each of the $3! = 6$ relative orders equally likely. Exactly one of those orders is "ace of spades, then king of hearts, then two of clubs." Hence

$$P = \frac{1}{3!} = \frac{1}{6}.$$

The 49 irrelevant cards never enter the computation — symmetry collapses the whole shuffle to the relative order of the three cards we care about.

Worked example: a bijection between draws

From a bag of chips numbered 1 through n, two distinct chips are drawn. Show that the smaller number is equally likely to be any value that leaves room for a larger partner, and use symmetry to find P(the two numbers differ by exactly 1) for $n = 10$.

Solution. Total unordered pairs: $\binom{10}{2} = 45$, all equally likely. Consecutive pairs are $(1,2), (2,3), \ldots, (9,10)$: there are 9 of them. So $P = \frac{9}{45} = \frac{1}{5}$. The bijection "pair $(k, k+1) \leftrightarrow k$" shows there are exactly $n-1$ consecutive pairs, giving the general answer $\frac{n-1}{\binom{n}{2}} = \frac{2}{n}$.

Tip. Before counting, ask "are these outcomes interchangeable?" If swapping two objects (or reversing an order) maps favorable outcomes to favorable outcomes and unfavorable to unfavorable, you have a symmetry — use it. "A before B" in a random order is the workhorse: its probability is always $\frac{1}{2}$.

8.9 Higher-Dimensional Geometric Probability

Pick the right measure: length, area, or volume

Uniform random points still give $P(E) = \dfrac{\text{measure of favorable region}}{\text{measure of whole region}}$, but in 2 or 3 variables the "measure" is an area or a volume.

- **Use a ratio of measures** when the favorable region is a recognizable shape (triangle, disk, polygon, box) whose area or volume you can find by geometry. This is the first thing to try.
- **Cut, complement, or exploit symmetry** when the region is awkward: split it into triangles and rectangles, measure the *failure* region instead, or find a rigid motion (like swapping coordinates) that pairs the region with one you already know.

Three independent uniform reals $\Rightarrow$ a point in a cube $\Rightarrow$ *volume* ratio; a curved boundary $\Rightarrow$ reach for a known circle or sphere formula.

Worked example: a probability is a ratio, an average comes from symmetry

A point (x, y) is chosen uniformly in the unit square $[0, 1]^2$.

(a) Find $P\big(x + y \le \frac{1}{2}\big)$. *(b)* Find the expected value of $x + y$.

Solution. *(a)* The whole square has area 1. The favorable set $x + y \le \frac{1}{2}$ is a right triangle with legs $\frac{1}{2}$, area $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$. A clean shape, so use a ratio:

$$P\big(x + y \le \tfrac{1}{2}\big) = \frac{1/8}{1} = \frac{1}{8}.$$

(b) An average, so use symmetry and linearity of expectation (previewed below). The mirror $x \mapsto 1 - x$ sends the uniform point to another uniform point, so $\mathbb{E}[x] = \mathbb{E}[1 - x] = 1 - \mathbb{E}[x]$, forcing $\mathbb{E}[x] = \frac{1}{2}$; the same holds for y. Adding the two averages,

$$\mathbb{E}[x + y] = \mathbb{E}[x] + \mathbb{E}[y] = \tfrac{1}{2} + \tfrac{1}{2} = 1.$$

Part (a) wanted the size of a region (ratio); part (b) wanted an average (symmetry plus linearity).

Worked example: three reals and a cube (olympiad flavor)

Real numbers x, y, z are chosen independently and uniformly from $[0, 1]$. Find the probability that they can be the side lengths of the three edges meeting at a corner of a box whose space diagonal is at most 1, i.e. $x^2 + y^2 + z^2 \le 1$.

Solution. The sample space is the unit cube, volume 1. The favorable region is the part of the cube inside the sphere of radius 1 centered at the origin — exactly one octant of that ball:

$$P\big(x^2 + y^2 + z^2 \le 1\big) = \frac{\frac{1}{8} \cdot \frac{4}{3}\pi(1)^3}{1} = \frac{\pi}{6}.$$

Three independent uniform reals put us in a cube (volume ratio); the curved boundary is a sphere, so we lean on the known ball-volume formula — no heavier machinery needed.

Tip. Count the number of independent random reals: $1 \Rightarrow$ length, $2 \Rightarrow$ area, $3 \Rightarrow$ volume. If the favorable boundary is straight or circular/spherical, reach for a known area/volume formula; if you are asked for an *expected value* (an average over the region), reach for symmetry and linearity of expectation (Unit 9) before anything fancier.

8.10 Inclusion–Exclusion for Probability & the Union Bound

Probability of a union of many events

For any events $A_1, \ldots, A_n$, the probability that *at least one* occurs is

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right).$$

This is inclusion–exclusion applied to probabilities: add the singles, subtract the pairwise overlaps, add back the triples, and so on. When exact overlaps are hard, the **union bound** (Boole's inequality) gives a quick one-sided estimate by keeping only the first sum:

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

The union bound is loose but never requires knowing any intersection — ideal for showing a bad event is *unlikely*.

Worked example: derangement-style matching (AIME flavor)

Four letters are placed at random into four pre-addressed envelopes, one per envelope. Find the probability that *at least one* letter lands in its correct envelope.

Solution. Let A_i be the event that letter i is correct. By symmetry $P(A_i) = \frac{3!}{4!} = \frac{1}{4}$; each pairwise $P(A_i \cap A_j) = \frac{2!}{4!} = \frac{1}{12}$; each triple $\frac{1!}{4!} = \frac{1}{24}$; the quadruple $\frac{1}{24}$. Inclusion–exclusion:

$$P\left(\bigcup A_i\right) = \binom{4}{1}\tfrac{1}{4} - \binom{4}{2}\tfrac{1}{12} + \binom{4}{3}\tfrac{1}{24} - \binom{4}{4}\tfrac{1}{24} = 1 - \tfrac{1}{2} + \tfrac{1}{6} - \tfrac{1}{24} = \frac{15}{24} = \frac{5}{8}.$$

Equivalently, 1 minus the derangement probability $\frac{9}{24} = \frac{3}{8}$. As $n \to \infty$ this "at least one fixed point" probability tends to $1 - e^{-1}$.

Worked example: the union bound as a safety net

Each of 10 independent components fails on a given day with probability 0.002. Bound the probability that the system, which fails if *any* component fails, fails today.

Solution. Exact inclusion–exclusion is messy, but the union bound is instant:

$$P(\text{system fails}) = P\left(\bigcup_{i=1}^{10} A_i\right) \leq \sum_{i=1}^{10} P(A_i) = 10(0.002) = 0.02.$$

The true value $1 - (0.998)^{10} \approx 0.0198$ is just under the bound — when the events are rare and barely overlap, the union bound is both quick and tight.

Tip. For an *exact* "at least one of many" probability, use full inclusion–exclusion (and exploit symmetry so each k-fold term is $\binom{n}{k}$ copies of a single value). For a fast *upper* bound —

especially to prove something is unlikely — drop to the union bound $P(\bigcup A_i) \le \sum P(A_i)$. Complementary counting $(1 - P(\text{none}))$ is a third route when the A_i are independent.

8.11 A Catalog of Discrete Distributions & a Preview of Expected Value

The standard families and their key probabilities

Most contest setups reduce to one of a few named **distributions**. Recognizing which one you are in gives the probability formula immediately.

- **Discrete uniform** on $\{1, 2, \dots, n\}$: each value has probability $\frac{1}{n}$; mean $\dfrac{n+1}{2}$. (One fair die, a random card value.)
- **Binomial** $B(n, p)$: number of successes in n independent trials, each succeeding with probability p. The chance of exactly k successes is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n.$$

- **Geometric**: the trial number of the *first* success in independent trials with success probability p. The chance the first success is on trial k is

$$P(X = k) = (1-p)^{k-1}\, p, \qquad k = 1, 2, 3, \dots.$$

Each is just the equally-likely ratio bookkeeping done once and packaged as a formula.

Worked example: binomial and geometric together

A biased coin shows heads with probability $p = \frac{1}{3}$. It is flipped repeatedly and independently.

(a) In 5 flips, find $P(\text{exactly 2 heads})$. *(b)* Find $P(\text{the first head is on flip 4})$.

Solution. *(a)* Binomial with $n = 5$, $k = 2$:

$$P(X = 2) = \binom{5}{2} \left(\tfrac{1}{3}\right)^2 \left(\tfrac{2}{3}\right)^3 = 10 \cdot \tfrac{1}{9} \cdot \tfrac{8}{27} = \frac{80}{243}.$$

(b) Geometric with $p = \frac{1}{3}$: three failures then a success,

$$P(X = 4) = \left(\tfrac{2}{3}\right)^3 \left(\tfrac{1}{3}\right) = \frac{8}{81}.$$

Identify the family (fixed number of trials ⇒ binomial; wait for the first success ⇒ geometric) and the formula writes itself.

Worked example: expected value preview (linking to Unit 9)

A fair die is rolled once. Find the expected value of the number shown. Then find the expected number of heads in 5 flips of the coin with $p = \frac{1}{3}$.

Solution. *Definition.* The **expected value** of a discrete random variable is the probability-weighted average of its values, $\mathbb{E}[X] = \sum_x x\, P(X = x)$. For the die (discrete uniform on 1–6):

$$\mathbb{E}[X] = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}.$$

For the coin, rather than sum the binomial terms, use **linearity of expectation**: each flip contributes an expected $p = \frac{1}{3}$ head, so

$$\mathbb{E}[\text{heads}] = 5 \cdot \tfrac{1}{3} = \frac{5}{3}.$$

In general $\mathbb{E}[B(n,p)] = np$ and $\mathbb{E}[\text{geometric}(p)] = \frac{1}{p}$. Unit 9 develops linearity of expectation into a power tool; for now, note it lets you average without ever computing the full distribution.

Tip. Classify before you compute: *fixed number of independent trials, count successes* $\Rightarrow$ binomial $\binom{n}{k}p^k(1-p)^{n-k}$; *repeat until the first success* $\Rightarrow$ geometric $(1-p)^{k-1}p$; *one draw from equally likely values* $\Rightarrow$ uniform. For averages, prefer $\mathbb{E}[X] = \sum x\, P(X = x)$, and remember $\mathbb{E}[B(n,p)] = np$ and $\mathbb{E}[\text{geometric}] = \frac{1}{p}$ as instant shortcuts (full treatment in Unit 9).

9 Conditional Probability, Independence & Bayes

Competition counting & probability, Unit 8

9.1 Conditional Probability $P(A \mid B)$

Probability once you already know something

The **conditional probability** of A given that B has occurred is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0.$$

Conditioning on B shrinks the sample space to just the outcomes in B: we ask what fraction of B's probability also lies in A. Rearranging gives the form you will use constantly,

$$P(A \cap B) = P(A \mid B)\, P(B) = P(B \mid A)\, P(A).$$

In a finite equally-likely setting this is simply $P(A \mid B) = \dfrac{|A \cap B|}{|B|}$: recount inside the restricted world B.

Worked example: a die given "even"

A fair die is rolled. Given that the result is even, what is the probability it is greater than 3?

Solution. Let $B = \{2, 4, 6\}$ and $A = \{4, 5, 6\}$. Then $A \cap B = \{4, 6\}$, so

$$P(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{2}{3}.$$

Knowing "even" throws away the odd outcomes; we only compare within $\{2, 4, 6\}$.

Worked example: two cards without replacement

Two cards are drawn from a standard 52-card deck without replacement. What is the probability the second is a heart given the first was a heart?

Solution. After removing one heart, 12 hearts remain among 51 cards, so

$$P(\text{2nd heart} \mid \text{1st heart}) = \frac{12}{51} = \frac{4}{17}.$$

Conditioning literally updates the deck.

Pitfall. $P(A \mid B)$ and $P(B \mid A)$ are *different* numbers. "The probability of a rash given measles" is near 1; "the probability of measles given a rash" is tiny. Swapping the two is the single most common conditional-probability error.

9.2 The Multiplication Rule & Probability Trees

Chaining conditional probabilities

Solving $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ for the intersection gives the **multiplication rule**:

$$P(A \cap B) = P(B)\, P(A \mid B).$$

For a sequence of stages it chains (the *chain rule*):

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2).$$

A **probability tree** draws this: each branch carries a conditional probability, and the probability of a full path is the *product* of the branch probabilities along it. Probabilities on branches leaving a node sum to 1; probabilities of the leaves sum to 1.

Multiply along a path for a leaf; add leaves for an event.

Worked example: drawing two aces

Two cards are drawn without replacement. Probability both are aces?

Solution. Multiply along the path:

$$P(\text{ace}_1 \cap \text{ace}_2) = \frac{4}{52} \cdot \frac{3}{51} = \frac{12}{2652} = \frac{1}{221}.$$

Worked example: reading a two-stage tree

A box has 2 red and 3 blue balls. Draw two without replacement. Probability of exactly one red?

Solution. Two paths give "one red": RB and BR.

$$P = \underbrace{\frac{2}{5} \cdot \frac{3}{4}}_{RB} + \underbrace{\frac{3}{5} \cdot \frac{2}{4}}_{BR} = \frac{6}{20} + \frac{6}{20} = \frac{12}{20} = \frac{3}{5}.$$

Multiply *along* each path, then add the paths that satisfy the event.

Tip. Trees turn "and" into multiplication and "or" into addition automatically. Whenever a problem unfolds in stages (draw, then draw; test, then confirm), sketch the tree before writing any algebra.

9.3 Independence

When knowing B tells you nothing about A

Events A and B are **independent** exactly when

$$P(A \cap B) = P(A)\,P(B),$$

equivalently (when the conditionals are defined) $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$. Learning that B happened does not change the probability of A.

For several events, **mutual independence** requires the product rule to hold for *every* sub-collection, not just pairs. Pairwise independence does not imply mutual independence.

Worked example: independent vs. dependent draws

Draw a card, replace it, reshuffle, draw again. Are "first is a heart" and "second is a heart" independent?

Solution. With replacement the deck resets: $P(\text{2nd heart} \mid \text{1st heart}) = \frac{1}{4} = P(\text{2nd heart})$, so **yes**. Without replacement it was $\frac{12}{51} \neq \frac{1}{4}$, so **no**. Replacement is the usual switch between independent and dependent.

Worked example: testing the definition

Roll two fair dice. Let $A =$ "first die is ≤ 2" and $B =$ "sum is 7". Independent?

Solution. $P(A) = \frac{2}{6} = \frac{1}{3}$, $P(B) = \frac{6}{36} = \frac{1}{6}$. The pairs with first die ≤ 2 and sum 7 are $(1,6), (2,5)$, so $P(A \cap B) = \frac{2}{36} = \frac{1}{18}$. Check: $P(A)P(B) = \frac{1}{3} \cdot \frac{1}{6} = \frac{1}{18}$. Equal, so **independent**.

Pitfall. Independent is not the same as mutually exclusive. Disjoint events with positive probability are strongly *dependent*: if $A \cap B = \varnothing$ then $P(A \mid B) = 0 \neq P(A)$. Always test independence with the product rule, never by intuition.

9.4 The Law of Total Probability

Break an event across a partition of causes

Let $B_1, B_2, \ldots, B_n$ **partition** the sample space (disjoint, together all outcomes, each $P(B_i) > 0$). For any event A,

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)\, P(B_i).$$

Each term is one path of a tree: choose the scenario B_i, then let A happen inside it. This is the workhorse for computing an overall probability when the world splits into cases.

Worked example: two urns

Urn 1 has 3 red, 1 white; Urn 2 has 1 red, 3 white. Flip a fair coin to pick an urn, then draw a ball. Probability the ball is red?

Solution. Partition by the urn:

$$P(R) = P(R \mid U_1)P(U_1) + P(R \mid U_2)P(U_2) = \frac{3}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}.$$

Worked example: defective parts from three machines

Machines A, B, C make $50\%, 30\%, 20\%$ of output with defect rates $1\%, 2\%, 3\%$. Probability a random item is defective?

Solution.

$$P(D) = (0.01)(0.50) + (0.02)(0.30) + (0.03)(0.20) = 0.005 + 0.006 + 0.006 = 0.017.$$

A weighted average of the defect rates, weighted by production share.

Tip. The weights $P(B_i)$ must sum to 1 and the cases must not overlap. When you see "there are several types/sources, each with its own rate," reach for total probability.

9.5 Bayes' Theorem

Reversing the conditioning

Combining $P(B_i \cap A) = P(A \mid B_i)P(B_i)$ with the law of total probability gives **Bayes' Theorem**:

$$P(B_i \mid A) = \frac{P(A \mid B_i)\, P(B_i)}{\sum_j P(A \mid B_j)\, P(B_j)} = \frac{P(A \mid B_i)\, P(B_i)}{P(A)}.$$

It converts the *prior* $P(B_i)$ (belief before the evidence) into the *posterior* $P(B_i \mid A)$ (belief after seeing A). The numerator is one path; the denominator is all paths reaching A.

Worked example: which urn produced the red ball?

Using the two-urn setup above, a red ball was drawn. Probability it came from Urn 1?

Solution.

$$P(U_1 \mid R) = \frac{P(R \mid U_1)P(U_1)}{P(R)} = \frac{\frac{3}{4}\cdot\frac{1}{2}}{\frac{1}{2}} = \frac{3/8}{1/2} = \frac{3}{4}.$$

The red ball is evidence that shifts belief toward the red-heavy urn.

Worked example: the defective came from machine C?

For the three machines above, a defective item is found. Probability it was machine C's?

Solution.

$$P(C \mid D) = \frac{(0.03)(0.20)}{0.017} = \frac{0.006}{0.017} = \frac{6}{17} \approx 0.353.$$

Even though C makes only 20% of output, its high defect rate makes it responsible for over a third of defects.

Tip. A clean recipe: (1) list the scenarios B_i and their priors, (2) attach the likelihood $P(A \mid B_i)$ to each, (3) multiply for each path, (4) divide the target path by the sum of all paths. Steps 1–3 are just a tree; Bayes is only the final division.

9.6 Classic Traps: Monty Hall, Two-Child, False Positives

Why these fool everyone

Each trap hides a conditioning step. The intuition "symmetry means 50/50" fails because the evidence is not symmetric across the hidden scenarios. Draw the tree, weight by priors, and let Bayes overrule the gut.

Monty Hall, fully explained

Three doors hide one car and two goats. You pick a door; the host — who *knows* where the car is — opens a different door revealing a goat, then offers a switch. Should you switch?

Solution. Your first pick is right with probability $\frac{1}{3}$ and wrong with probability $\frac{2}{3}$. The host's guaranteed goat-reveal gives no new information about your original door, so:

- If you initially picked the car ($\frac{1}{3}$), switching *loses*.
- If you initially picked a goat ($\frac{2}{3}$), the host is forced to reveal the other goat, so the remaining door hides the car — switching *wins*.

Therefore $P(\text{win by switching}) = \frac{2}{3}$, double the $\frac{1}{3}$ from staying. The key is that the host's choice is *constrained* by knowledge; a host opening a random door would change the analysis.

The false-positive (medical-test) paradox

A disease affects 1 in 1000 people. A test is 99% accurate both ways (sensitivity 99%, specificity 99%). You test positive. What is the probability you actually have the disease?

Solution. Let D be disease, $+$ a positive test. Priors $P(D) = 0.001$, $P(D^c) = 0.999$.

$$P(+ \mid D) = 0.99, \qquad P(+ \mid D^c) = 0.01.$$

By Bayes,

$$P(D \mid +) = \frac{(0.99)(0.001)}{(0.99)(0.001) + (0.01)(0.999)} = \frac{0.00099}{0.00099 + 0.00999} = \frac{0.00099}{0.01098} \approx 0.090.$$

Only about 9%! Because the disease is rare, the many false positives from the huge healthy population swamp the few true positives. The base rate dominates.

The two-child problem and its ambiguity

A family has two children.

- "At least one is a boy." Probability both are boys? Sample space $\{BB, BG, GB, GG\}$ is equally likely; condition on at least one boy: $\{BB, BG, GB\}$, of which BB is one. Answer $\frac{1}{3}$.
- "The older is a boy." Probability both boys? Now condition on $\{BB, BG\}$: answer $\frac{1}{2}$.

Same family, different information, different answers — see Going Deeper for why the phrasing matters.

Pitfall. In every trap, *how* the information was obtained changes the conditioning set. "I met one of the children and he's a boy" is not the same evidence as "at least one child is a boy," even though both mention a boy. Model the process, not just the sentence.

9.7 Going Deeper: Bayesian Updating, Conditioning, and Ambiguity

Belief as something you update

Bayesian updating treats probability as a state of belief refined by evidence. The posterior after one observation becomes the prior for the next:

$$P(H \mid E_1, E_2) \;\propto\; P(E_2 \mid H)\,\underbrace{P(E_1 \mid H)\,P(H)}_{\text{posterior after } E_1}.$$

Independent pieces of evidence multiply their likelihood ratios. A rare hypothesis can still win if the evidence is strong enough, and a common one can survive weak evidence — exactly the tension in the medical-test example.

Worked example: updating twice

The rare disease ($P(D) = 0.001$) test above comes back positive; a second *independent* test (same accuracy) is also positive. Now find $P(D \mid {+}{+})$.

Solution. Use the first posterior $P(D \mid +) \approx 0.0902$ as the new prior:

$$P(D \mid {+}{+}) = \frac{(0.99)(0.0902)}{(0.99)(0.0902) + (0.01)(0.9098)} \approx \frac{0.0893}{0.0893 + 0.0091} \approx 0.907.$$

Two positives push belief from 9% to about 91%. Evidence compounds.

Worked example: why conditioning defies intuition

Why does "at least one boy" give $\frac{1}{3}$ but "the older is a boy" give $\frac{1}{2}$?

Solution. Conditioning removes outcomes unevenly. "Older is a boy" eliminates GG *and* GB, leaving two equally likely worlds BB, BG. "At least one boy" eliminates only GG, leaving three worlds BB, BG, GB, and BB is just one of them. The information "at least one" is weaker — it does not pin down *which* child — so it cuts the space less. Intuition fails because we imagine the evidence naming a specific child when it does not.

Worked example: the ambiguity made precise

"A family has two children; I tell you at least one is a boy born on a Tuesday. Both boys?" The Tuesday detail, seemingly irrelevant, shifts the answer to $\frac{13}{27}$, not $\frac{1}{3}$.

Solution sketch. Each child is one of 14 equally likely (gender, weekday) types. Counting the ordered pairs containing at least one "boy-Tuesday" gives 27; those with two boys give 13; hence $\frac{13}{27} \approx 0.481$. The lesson: the more specifically the evidence identifies a child, the closer the answer creeps to $\frac{1}{2}$. Ambiguous wording has no single "right" answer until the sampling process is fixed.

Big picture — conditioning is re-weighting your world. Every result in this unit is one identity, $P(A \cap B) = P(A \mid B)P(B)$, read in different directions:

1. **Forward (trees, multiplication, total probability):** split into cases, multiply along paths, add paths to get $P(A)$.
2. **Backward (Bayes):** once A is observed, divide the target path by all paths to get $P(B_i \mid A)$ — prior times likelihood, normalized.
3. **Independence** is the special case where conditioning changes nothing: $P(A \mid B) = P(A)$.
4. **The traps** all come from misreading *which* outcomes the evidence keeps. Always model the process that produced the information, weight by the priors (mind the base rate), and trust the arithmetic over the gut.

9.8 The Odds Form of Bayes & Sequential Updating

Bayes as a multiplier on odds

Write Bayes' theorem for two competing hypotheses H and its complement H^c and divide the two versions. The awkward normalizing denominator $P(E)$ cancels, leaving the **odds form**:

$$\underbrace{\frac{P(H \mid E)}{P(H^c \mid E)}}_{\text{posterior odds}} = \underbrace{\frac{P(E \mid H)}{P(E \mid H^c)}}_{\text{likelihood ratio (LR)}} \cdot \underbrace{\frac{P(H)}{P(H^c)}}_{\text{prior odds}}.$$

Updating is now pure multiplication: *posterior odds* $=$ *LR* $\times$ *prior odds*. For a stream of *conditionally independent* pieces of evidence $E_1, \dots, E_n$ the likelihood ratios simply multiply,

$$\frac{P(H \mid E_1, \dots, E_n)}{P(H^c \mid E_1, \dots, E_n)} = \left(\prod_{k=1}^{n} \frac{P(E_k \mid H)}{P(E_k \mid H^c)} \right) \frac{P(H)}{P(H^c)}.$$

This is **iterated Bayesian updating**: each posterior is the next prior, and in odds language the whole computation collapses to one product. Recover a probability at the end via $p =$

$$\frac{\text{odds}}{1+\text{odds}}.$$

Worked example: three positive tests, done in odds

A disease has prevalence $P(D) = \frac{1}{1000}$. A test has sensitivity $P(+ \mid D) = 0.99$ and false-positive rate $P(+ \mid D^c) = 0.01$, giving likelihood ratio $\mathrm{LR} = \frac{0.99}{0.01} = 99$. Three *independent* tests all read positive. Find $P(D \mid {+}{+}{+})$.

Solution. Prior odds $= \dfrac{1/1000}{999/1000} = \dfrac{1}{999}$. Multiply by LR once per positive:

$$\text{posterior odds} = 99^3 \cdot \frac{1}{999} = \frac{970299}{999} \approx 971.3.$$

Convert back:

$$P(D \mid {+}{+}{+}) = \frac{971.3}{1 + 971.3} \approx 0.99897.$$

One positive left belief near 9% (the base rate dominated); the odds form shows why each extra positive multiplies the odds by 99, so three positives overwhelm even a 1-in-1000 prior.

Worked example: mixed evidence (a positive then a negative)

Same test. A patient tests positive, then negative on an independent retest. What is $P(D)$ now?

Solution. The negative test has its own likelihood ratio $\dfrac{P(- \mid D)}{P(- \mid D^c)} = \dfrac{0.01}{0.99} = \dfrac{1}{99}$. Chain both:

$$\text{posterior odds} = \underbrace{99}_{+} \cdot \underbrace{\tfrac{1}{99}}_{-} \cdot \underbrace{\tfrac{1}{999}}_{\text{prior}} = \frac{1}{999}.$$

The two tests exactly cancel, returning belief to the prior $\frac{1}{1000}$. Contradictory evidence multiplies to a net ratio of 1 — a fact almost invisible in the fraction form but obvious in odds.

Tip. Whenever a problem chains several independent clues toward one yes/no hypothesis, switch to odds: multiply the prior odds by each likelihood ratio, then convert once at the end. It sidesteps the denominator entirely and makes "how much does this clue matter?" literally a multiplier.

9.9 Markov Chains & Transition Matrices

Memoryless step-by-step randomness

A **Markov chain** moves among states $1, \ldots, m$ so that the next state depends only on the current one, not the past history:

$$P(X_{n+1} = j \mid X_n = i,\ X_{n-1}, \dots) = P(X_{n+1} = j \mid X_n = i) =: p_{ij}.$$

Collect these into the **transition matrix** $P = [p_{ij}]$; each row sums to 1. If the current distribution over states is a row vector $\mathbf{v}$, then after one step it is $\mathbf{v}P$, and after n steps $\mathbf{v}P^n$. The (i, j) entry of P^n is exactly $P(X_n = j \mid X_0 = i)$ — multi-step probabilities are just matrix powers.

A **stationary distribution** $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi}P = \boldsymbol{\pi}$ with $\sum_i \pi_i = 1$: the chain, once in $\boldsymbol{\pi}$, stays distributed as $\boldsymbol{\pi}$ forever. An **absorbing state** has $p_{ii} = 1$ (once entered, never left).

Worked example: sunny/rainy two-step and long-run

Weather follows a chain on states (Sunny, Rainy) with

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}.$$

(a) If today is sunny, probability it is rainy the day after tomorrow? (b) What fraction of days are sunny in the long run?

Solution. (a) Compute the two-step matrix:

$$P^2 = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.72 & 0.28 \\ 0.56 & 0.44 \end{bmatrix}.$$

The (Sunny $\to$ Rainy) entry is 0.28.

(b) Solve $\boldsymbol{\pi}P = \boldsymbol{\pi}$ with $\boldsymbol{\pi} = (\pi_S, \pi_R)$, $\pi_S + \pi_R = 1$. The first component gives $0.8\pi_S + 0.4\pi_R = \pi_S$, i.e. $0.4\pi_R = 0.2\pi_S$, so $\pi_S = 2\pi_R$. With $\pi_S + \pi_R = 1$ we get $\pi_S = \frac{2}{3}$, $\pi_R = \frac{1}{3}$. Sunny two-thirds of the time.

Worked example: a three-state absorbing chain

A token sits on states A, B, C with C absorbing:

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}.$$

Starting at A, what is the probability of *eventual* absorption at C? (Here it is forced, so instead find the probability that the token visits B before being absorbed.)

Solution. From A the first move is to B or C with probability $\frac{1}{2}$ each. It visits B before absorption iff the very first step lands on B (from B the only escape is to C eventually, but the "before absorption" visit is already secured). Hence $P(\text{visit } B \mid X_0 = A) = \frac{1}{2}$. To confirm absorption is certain, note C is reachable from both A and B and is absorbing, so with probability 1 the token is eventually trapped at C: powers P^n push all mass into column C.

Tip. Rows of P must each sum to 1 — always sanity-check this first. Multi-step questions (n days later, n moves later) are P^n; "in the long run / eventually / steady state" questions are the stationary equation $\boldsymbol{\pi} P = \boldsymbol{\pi}$. Never confuse a single large power with the stationary vector unless the chain has actually settled.

9.10 Gambler's Ruin

Reaching N before 0

A gambler holding \$$k$ bets \$1 per round, winning with probability p and losing with probability $q = 1 - p$, until reaching a target \$$N$ (win) or \$0 (ruin). Let r_k be the probability of reaching N before 0 starting from k. Conditioning on the first bet (this is **first-step analysis**) gives the recurrence

$$r_k = p\, r_{k+1} + q\, r_{k-1}, \qquad r_0 = 0,\ r_N = 1.$$

Solving this linear recurrence yields the classic formulas

$$r_k = \begin{cases} \dfrac{k}{N}, & p = q = \frac{1}{2} \text{ (fair game)}, \\[2ex] \dfrac{1 - (q/p)^k}{1 - (q/p)^N}, & p \neq q \text{ (biased game)}. \end{cases}$$

The fair case is a straight line in k; the biased case is governed by the ratio q/p.

Worked example: fair game, doubling your money

You start with \$50 and play a fair game ($p = \frac{1}{2}$), quitting at \$0 or \$100. Probability you reach \$100?

Solution. With $k = 50$, $N = 100$, $p = q = \frac{1}{2}$:

$$r_{50} = \frac{k}{N} = \frac{50}{100} = \frac{1}{2}.$$

In a fair game the chance of hitting the target before ruin is just your current fortune as a fraction of the target — no arithmetic beyond a ratio.

Worked example: a slightly unfair game

A gambler starts at \$1, wants to reach \$5, and wins each \$1 bet with probability $p = 0.4$ (so $q = 0.6$, ratio $q/p = \frac{3}{2}$). Probability of success?

Solution. Use the biased formula with $k = 1,\ N = 5,\ q/p = \frac{3}{2}$:

$$r_1 = \frac{1 - (3/2)^1}{1 - (3/2)^5} = \frac{1 - 1.5}{1 - 7.59375} = \frac{-0.5}{-6.59375} \approx 0.0758.$$

Only about 7.6%. A small per-bet disadvantage compounds brutally over the distance to the target: fighting the house edge across many rounds is nearly hopeless. (Contrast the fair game, where the same $k/N = \frac{1}{5} = 20\%$.)

Big picture. Gambler's ruin is the template for every "random walk with two barriers" problem: set r_k for the boundary event, condition on the first step to get $r_k = p\,r_{k+1} + q\,r_{k-1}$, and impose the boundary values. The same recurrence machinery reappears in fair-coin streak problems, tennis deuce, and best-of-N series.

9.11 First-Step Analysis & Conditional Expectation (Unit 9 Preview)

Condition on the first move to set up equations

First-step analysis solves "how long / how likely" questions on a chain by conditioning on the very first transition and using the memoryless property: after one step the problem *looks like the same problem from the new state.* This turns a hard infinite process into a small linear system.

For **expected times** it uses the tower property of **conditional expectation**,

$$E[T] = \sum_i E[T \mid \text{first step to } i]\, P(\text{first step to } i),$$

where reaching a new state i contributes the 1 step already taken plus the expected remaining time $E_i[T]$ from i. Letting $e_k = E[\text{steps to absorption from } k]$ typically yields a recurrence such as $e_k = 1 + p\,e_{k+1} + q\,e_{k-1}$ with $e = 0$ at absorbing states. This conditional-expectation viewpoint is the gateway to Unit 9.

Worked example: expected flips to get two heads in a row

Let e be the expected number of fair-coin flips to first see `HH`. Track states by current progress: S_0 (no useful streak), S_1 (one H so far). Let $a = E[\text{flips from } S_0]$, $b = E[\text{flips from } S_1]$.

Solution. First-step analysis on each state:

$$a = 1 + \tfrac{1}{2}\,b + \tfrac{1}{2}\,a, \qquad b = 1 + \tfrac{1}{2}\cdot 0 + \tfrac{1}{2}\,a.$$

From the first equation $\frac{1}{2}a = 1 + \frac{1}{2}b \Rightarrow a = 2 + b$. Substitute into the second: $b = 1 + \frac{1}{2}a = 1 + \frac{1}{2}(2 + b) = 2 + \frac{1}{2}b$, so $\frac{1}{2}b = 2 \Rightarrow b = 4$, and $a = 6$. The expected number of flips is $\boxed{6}$. (The general pattern for a streak of n heads is $2^{n+1} - 2$.)

Worked example: expected duration of a fair gambler's game

In the fair gambler's-ruin game ($p = \frac{1}{2}$), let e_k be the expected number of bets until hitting 0 or N from \$$k$. Find a formula.

Solution. Condition on the first bet:

$$e_k = 1 + \tfrac{1}{2}\, e_{k+1} + \tfrac{1}{2}\, e_{k-1}, \qquad e_0 = e_N = 0.$$

Rewrite as $e_{k+1} - 2e_k + e_{k-1} = -2$: the second difference is constant, so e_k is quadratic. Trying $e_k = k(N-k)$ gives $(k+1)(N-k-1) - 2k(N-k) + (k-1)(N-k+1) = -2$, which checks, and $e_0 = e_N = 0$. Hence

$$e_k = k(N-k).$$

Starting at the midpoint $k = N/2$ the game lasts an expected $N^2/4$ bets — surprisingly long. First-step analysis converts the recurrence directly into the answer.

Bridge to Unit 9. Everything here is one idea: *condition on the first step, then reuse the structure of the problem.* For **probabilities** it gives $r_k = p\, r_{k+1} + q\, r_{k-1}$; for **expected times** it gives $e_k = 1 + p\, e_{k+1} + q\, e_{k-1}$. The tool is conditional expectation and its tower property, $E[X] = E\big[E[X \mid Y]\big]$ — the central engine of Unit 9. Master first-step analysis now and expectation will feel like the same move applied to averages instead of probabilities.

10 Expected Value, Linearity & Advanced Probability

Competition counting & probability, Unit 9

10.1 Expected Value (Definition)

The expected value of a random variable

A **random variable** X assigns a number to each outcome of a random experiment. Its **expected value** (or *mean*) is the probability-weighted average of the values X can take:

$$E[X] = \sum_x x\, P(X = x),$$

summing over every value x in the range of X. Think of $E[X]$ as the long-run average of X over many independent repetitions.

Expectation is **linear** under scaling and shifting: for constants a, b,

$$E[aX + b] = a\, E[X] + b.$$

The weights $P(X = x)$ are nonnegative and sum to 1, so $E[X]$ is a genuine weighted average and always lies between the smallest and largest possible values of X.

A weighted average of payoffs

A game pays \$3 if a fair die shows a 6, pays \$1 if it shows a 4 or 5, and pays nothing otherwise. What is the expected payoff?

Solution. The payoff X takes value 3 with probability $\frac{1}{6}$, value 1 with probability $\frac{2}{6}$, and value 0 with probability $\frac{3}{6}$:

$$E[X] = 3 \cdot \tfrac{1}{6} + 1 \cdot \tfrac{2}{6} + 0 \cdot \tfrac{3}{6} = \tfrac{3}{6} + \tfrac{2}{6} = \tfrac{5}{6} \text{ dollars.}$$

Expected value of a single die

A fair six-sided die is rolled. Then

$$E[X] = \tfrac{1}{6}(1+2+3+4+5+6) = \tfrac{21}{6} = \tfrac{7}{2} = 3.5.$$

Note 3.5 is not a possible outcome — the mean need not be attainable.

Tip. To build $E[X]$ from scratch, list the distinct values, attach a probability to each, check the probabilities sum to 1, then take the dot product "value $\times$ probability." A missing or extra probability is the most common error.

10.2 Linearity of Expectation

Expectation adds — always

For *any* random variables X and Y defined on the same experiment,

$$E[X+Y] = E[X] + E[Y],$$

and more generally $E[\sum_i X_i] = \sum_i E[X_i]$. This holds **even when the X_i are dependent** — no independence is required. Combined with scaling,

$$E\left[\sum_i a_i X_i\right] = \sum_i a_i \, E[X_i].$$

Linearity is the single most powerful tool in this unit. It lets you compute the mean of a complicated total by breaking it into simple pieces, computing each piece's mean separately, and adding — *without ever finding the distribution of the total.*

Sum of two dice

Roll two fair dice; let S be their sum. Working out the full distribution of S (values 2 through 12) is tedious. Instead let $S = X_1 + X_2$ where X_i is die i. Each die has mean 3.5, so

$$E[S] = E[X_1] + E[X_2] = 3.5 + 3.5 = 7.$$

The dice need not even be independent for this to work.

Ten coins

Flip 10 fair coins and let H count the heads. Write $H = X_1 + \cdots + X_{10}$ where $X_i = 1$ if flip i is heads and 0 otherwise. Each $E[X_i] = \frac{1}{2}$, so $E[H] = 10 \cdot \frac{1}{2} = 5$.

Key insight. When you must find the expected value of a *count* or a *total*, do not chase its distribution. Split it into a sum of simpler variables and add their means. Dependence between the pieces is irrelevant to the *sum* of the expectations.

10.3 Indicator Random Variables

Turning "how many" into a sum of $0/1$ variables

An **indicator $\mathbf{1}_A$** equals 1 if event A happens and 0 otherwise. Its expectation is just the probability of the event:

$$E[\mathbf{1}_A] = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A).$$

If X counts how many of the events $A_1, \ldots, A_n$ occur, then $X = \sum_{i=1}^{n} \mathbf{1}_{A_i}$, and by linearity

$$E[X] = \sum_{i=1}^{n} P(A_i).$$

This is the workhorse pattern: whenever a quantity counts occurrences (heads, matches, fixed points, adjacent pairs, successes), define one indicator per potential occurrence and add up the individual probabilities. The events may overlap or be highly dependent — it does not matter.

Expected number of aces in a hand

Deal 5 cards from a standard 52-card deck. Let A be the number of aces. For position i, let $\mathbf{1}_i$ indicate that the i-th card dealt is an ace; by symmetry $P(\mathbf{1}_i = 1) = \frac{4}{52} = \frac{1}{13}$. Then

$$E[A] = \sum_{i=1}^{5} P(\text{card } i \text{ is an ace}) = 5 \cdot \tfrac{1}{13} = \tfrac{5}{13}.$$

The cards are dependent (drawing without replacement), yet linearity still applies.

Expected number of fixed points

Randomly permute $1, 2, \ldots, n$. Let F count the *fixed points* (positions i with $a_i = i$). Let $\mathbf{1}_i$

indicate $a_i = i$; then $P(a_i = i) = \frac{1}{n}$, so

$$E[F] = \sum_{i=1}^{n} \tfrac{1}{n} = n \cdot \tfrac{1}{n} = 1.$$

On average a random shuffle leaves exactly one element in place, for every n.

Recipe. (1) Identify the count. (2) Write it as $\sum \mathbf{1}_{A_i}$, one indicator per possible occurrence. (3) Find $P(A_i)$ for a single i (symmetry usually makes them all equal). (4) Add. This four-step recipe cracks a huge fraction of contest expectation problems.

10.4 Expected Value via States & Recursion

Conditioning on the first step

When a process repeats until something happens, set up an equation for the expected value by **conditioning on the first step**. Let E be the expected number of steps to finish. After one step you either finish or land back in a known state; averaging over the possibilities gives a linear equation in E that you solve.

For a process that, each step, succeeds with probability p (and otherwise restarts identically):

$$E = 1 + (1-p)\,E \implies E = \tfrac{1}{p}.$$

With several states, write one equation per state and solve the system.

The template is always "$E =$ (cost of one step)$+\sum$(probability of each next state)$\times$(expected cost from that state
Finished/absorbing states contribute expected cost 0.

First heads

Flip a fair coin until the first heads. Let E be the expected number of flips. One flip always happens; with probability $\frac{1}{2}$ we are done, and with probability $\frac{1}{2}$ we are back where we started:

$$E = 1 + \tfrac{1}{2} \cdot 0 + \tfrac{1}{2} \cdot E \implies \tfrac{1}{2}E = 1 \implies E = 2.$$

Two states: waiting for HH

Flip a fair coin until two heads occur in a row. Let a be the expected additional flips from the start (or just after a tails) and b the expected additional flips right after a single heads. Then

$$\begin{aligned} a &= 1 + \tfrac{1}{2}a + \tfrac{1}{2}b, \\ b &= 1 + \tfrac{1}{2}a + \tfrac{1}{2} \cdot 0. \end{aligned}$$

Substituting the second into the first gives $a = 1 + \frac{1}{2}a + \frac{1}{2}(1 + \frac{1}{2}a)$, so $a = \frac{3}{2} + \frac{3}{4}a$, giving

$\frac{1}{4}a = \frac{3}{2}$ and $a = 6$.

Tip. Choose states so that the future depends only on the current state, not the full history. Then each state gets one equation. Absorbing ("done") states have expected remaining cost 0 — they anchor the system.

10.5 Symmetry Arguments

Let symmetry do the counting

When outcomes are interchangeable, symmetry pins down probabilities and expectations with almost no computation. Two common forms:

- **Uniform position.** In a random permutation, each element is equally likely to occupy any given position, so its expected position is the average position $\frac{n+1}{2}$.
- **Equal-likelihood ordering.** Among k specific items in random order, each of the $k!$ relative orders is equally likely; e.g. any one of them is first with probability $\frac{1}{k}$.

Symmetry pairs beautifully with linearity: use symmetry to argue every indicator has the *same* probability, then multiply by the number of indicators.

Position of the first ace

Shuffle a 52-card deck. The 4 aces split the remaining 48 cards into 5 gaps (before the first ace, between consecutive aces, after the last). By symmetry each of the 48 non-aces is equally likely to fall in any gap, so each gap holds $\frac{48}{5}$ non-aces on average. The first ace sits just after the first gap, so its expected position is

$$\frac{48}{5} + 1 = \frac{53}{5} = 10.6.$$

More generally the first of k special cards among n has expected position $\frac{n+1}{k+1}$.

Who comes first?

Three runners A, B, C finish in a uniformly random order. By symmetry each is equally likely to be the fastest, so $P(A \text{ finishes first}) = \frac{1}{3}$. No casework needed.

Tip. Before computing, ask: "Are these objects interchangeable?" If relabeling them leaves the setup unchanged, they share the same probability or expected value. That observation often replaces pages of casework with a single fraction.

10.6 Waiting Times & the Geometric Distribution

Expected wait for a first success

Repeat independent trials, each succeeding with probability $p > 0$. Let N be the number of trials up to and including the first success. Then N has a **geometric distribution**:

$$P(N = k) = (1-p)^{k-1}p \quad (k = 1, 2, 3, \dots), \qquad E[N] = \frac{1}{p}.$$

The mean wait is the reciprocal of the success probability: rarer events take proportionally longer. (The variance is $\frac{1-p}{p^2}$, rarely needed on contests.)

The formula $E[N] = 1/p$ follows instantly from the recursion $E = 1 + (1-p)E$ of the previous section, so "waiting time" problems and "restart" recursions are two views of the same idea.

Rolling for a six

Roll a fair die until a 6 appears. Each roll succeeds with $p = \frac{1}{6}$, so the expected number of rolls is

$$E[N] = \frac{1}{1/6} = 6.$$

Waiting for a 1 *or* a 2 instead has $p = \frac{1}{3}$, giving $E[N] = 3$.

Free throws

A player makes each free throw with probability 0.8. The expected number of attempts to record the first make is $\frac{1}{0.8} = \frac{5}{4} = 1.25$.

Careful. $E[N] = 1/p$ counts the trial *on which* success occurs. If a problem asks for the expected number of *failures* before the first success, that is $\frac{1}{p} - 1 = \frac{1-p}{p}$. Read carefully which count is wanted.

10.7 Going Deeper: Where These Ideas Combine

Signature techniques of advanced expectation

The four moves below appear again and again in olympiad and AMC/AIME problems:

- **Linearity magic.** To find the expected value of a total, never compute its distribution — decompose into indicators or simple pieces and add. This defeats problems whose totals are hopelessly complicated to distribute directly.
- **Expected fixed points** $= 1$**.** In any random permutation of n objects, the expected number of elements left in place is exactly 1, independent of n (each of the n positions

is a fixed point with probability $\frac{1}{n}$).

- **Absorbing-chain expectations.** Model a repeating process as a Markov chain with "done" (absorbing) states; write one linear equation per transient state by conditioning on the next step, then solve. Symmetric gambler's-ruin on $\{0, 1, \ldots, N\}$ started at k lasts $k(N-k)$ steps in expectation.
- **Coupon collector.** To collect all n distinct coupons, once you hold j of them the chance a new draw is fresh is $\frac{n-j}{n}$, a geometric wait of $\frac{n}{n-j}$. Summing gives $E = n\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) = n\,H_n$.

Coupon collector for a die

Roll a fair die repeatedly until all six faces have appeared. Waiting for the first new face costs 1 roll; from j distinct faces seen, the next new face has probability $\frac{6-j}{6}$, so its expected wait is $\frac{6}{6-j}$. Summing over $j = 0, 1, \ldots, 5$:

$$E = 6\left(\tfrac{1}{6} + \tfrac{1}{5} + \tfrac{1}{4} + \tfrac{1}{3} + \tfrac{1}{2} + 1\right) = 6 \cdot \tfrac{49}{20} = \tfrac{147}{10} = 14.7 \text{ rolls.}$$

Absorbing chain on a triangle

A token starts at vertex A of triangle ABC and each step moves to one of the other two vertices with equal probability. How many steps, on average, to first reach C? Let a, b be the expected steps from A and B. By the transition rules,

$$a = 1 + \tfrac{1}{2}b, \qquad b = 1 + \tfrac{1}{2}a.$$

By symmetry $a = b$, so $a = 1 + \frac{1}{2}a$, giving $a = 2$ steps.

Big picture. Expectation problems reward the solver who *refuses to find the whole distribution*. Ask three questions: (1) Is this a **count or total**? Then split into indicators and add ($E = \sum P(A_i)$) — linearity ignores all dependence. (2) Does the process **repeat or wander among states**? Then condition on the first step and solve a linear system; a pure "restart" collapses to the geometric mean $1/p$. (3) Are the objects **interchangeable**? Then symmetry hands you the per-piece probability for free. Almost every advanced expectation problem is one of these three shapes in disguise — name the shape, and the computation shrinks to a few lines.

10.8 Linearity Masterclass: Records, Inversions & Cycles

One permutation, many statistics — all by indicators

A uniformly random permutation $a_1, a_2, \ldots, a_n$ of $\{1, \ldots, n\}$ hides a zoo of natural statistics. Linearity computes the mean of each one *without* independence:

- **Records** (left-to-right maxima): position i is a record if a_i exceeds all earlier entries. Among the first i values, each is equally likely to be the largest, so $P(\text{record at } i) = \frac{1}{i}$.
- **Inversions**: a pair $i < j$ is inverted if $a_i > a_j$. By symmetry each unordered pair is inverted with probability $\frac{1}{2}$.
- **Fixed points**: $P(a_i = i) = \frac{1}{n}$ (already seen: mean $= 1$).
- **Cycles**: the expected number of cycles equals the harmonic number $H_n = \sum_{k=1}^{n} \frac{1}{k}$.

In every case, sum the per-object probabilities. The events are wildly dependent; linearity does not care.

The cycle count is the subtle one. Build the permutation by the *Chinese-restaurant* process: insert elements one at a time, each either sitting after some existing element or starting a new cycle. Element k (the k-th inserted) closes a new cycle exactly when it is the one that "points back to the start," which happens with probability $\frac{1}{k}$ among its k equally likely choices. So the number of cycles is $\sum_{k=1}^{n} \mathbf{1}_k$ with $E[\mathbf{1}_k] = \frac{1}{k}$, giving $E[\#\text{cycles}] = H_n$.

Expected records and inversions (AIME flavor)

Let R count the left-to-right maxima and I the inversions of a random permutation of $1, \ldots, n$.

Records. $E[R] = \sum_{i=1}^{n} \frac{1}{i} = H_n$. For $n = 10$, $E[R] = 1 + \frac{1}{2} + \cdots + \frac{1}{10} = \frac{7381}{2520} \approx 2.93$.

Inversions. There are $\binom{n}{2}$ pairs, each inverted with probability $\frac{1}{2}$:

$$E[I] = \binom{n}{2} \cdot \tfrac{1}{2} = \frac{n(n-1)}{4}.$$

For $n = 10$ this is $\frac{10 \cdot 9}{4} = \frac{90}{4} = 22.5$. Notice records and cycles share the same mean H_n — not a coincidence, but a bijective fact (Foata's correspondence) that linearity reveals for free.

Key insight. A permutation statistic that is a *sum over positions or pairs* yields to indicators. The whole art is choosing the right "atom": positions for records and fixed points, unordered pairs for inversions, insertion steps for cycles. Find the atom, find one probability, multiply or sum. The joint distribution never appears.

10.9 Variance & the Second Moment

Measuring spread with $\mathrm{Var}(X)$

The **variance** of X measures how far X typically strays from its mean $\mu = E[X]$:

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

The second form ("mean of the square minus square of the mean") is almost always the easier one to compute. Variance scales quadratically: $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$ — shifting by b changes nothing, stretching by a multiplies spread by a^2.

Unlike expectation, **variance does NOT add in general**. It adds only when the pieces are *independent* (or merely uncorrelated):

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_i \mathrm{Var}(X_i) \quad \text{if the } X_i \text{ are independent.}$$

For a single indicator $\mathbf{1}_A$ with $p = P(A)$, we have $E[\mathbf{1}_A^2] = E[\mathbf{1}_A] = p$ (since $0^2 = 0$, $1^2 = 1$), so

$$\mathrm{Var}(\mathbf{1}_A) = p - p^2 = p(1-p).$$

This one fact powers the variance of every count built from independent indicators.

Variance of a binomial count

Flip n independent coins, each heads with probability p; let $H = \sum_{i=1}^n \mathbf{1}_i$ be the number of heads. The flips are independent, so variances add:

$$\mathrm{Var}(H) = \sum_{i=1}^n \mathrm{Var}(\mathbf{1}_i) = \sum_{i=1}^n p(1-p) = np(1-p).$$

For $n = 100$ fair coins ($p = \frac{1}{2}$): $E[H] = 50$ and $\mathrm{Var}(H) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25$, so the standard deviation is $\sqrt{25} = 5$. Typical outcomes cluster near 50 ± 5.

Second moment of a single die

Let X be a fair die roll. We found $E[X] = \frac{7}{2}$. For the variance, compute the second moment:

$$E[X^2] = \tfrac{1}{6}(1^2 + 2^2 + \cdots + 6^2) = \tfrac{1}{6} \cdot \tfrac{6\cdot 7\cdot 13}{6} = \tfrac{91}{6}.$$

Then

$$\mathrm{Var}(X) = E[X^2] - \big(E[X]\big)^2 = \tfrac{91}{6} - \tfrac{49}{4} = \tfrac{182-147}{12} = \tfrac{35}{12}.$$

The sum of two *independent* dice therefore has variance $2 \cdot \frac{35}{12} = \frac{35}{6}$, even though its *mean* 7 came from linearity without any independence.

Careful. Expectation always adds; variance adds *only under independence*. When indicators are dependent (drawing without replacement, fixed points of a permutation), you must include covariance terms: $\mathrm{Var}(\sum X_i) = \sum \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)$. If a problem asks for a variance and the pieces overlap, do not just add — the cross terms are where the points hide.

10.10 Conditional Expectation & the Tower Property

Averaging in stages

The **conditional expectation** $E[X \mid Y]$ is the average of X once the value of Y is known; it is itself a random variable (a function of Y). The **tower property** (law of total expectation) says you may average in two stages and recover the plain mean:

$$E[X] = E\big[E[X \mid Y]\big] = \sum_y E[X \mid Y = y]\, P(Y = y).$$

Read it as: "condition on whatever you wish you knew, solve the easy conditional problem, then average over the thing you conditioned on." This is exactly the engine behind first-step analysis — there Y is the outcome of the first step.

The tower property turns a hard unconditional expectation into a menu of easy conditional ones. Choosing the right Y (the first coin flip, the first move, the number of successes so far) is the whole game, and it directly generalizes the state-recursion method of Unit 8's Markov chains.

A random number of dice (compound expectation)

Roll a fair die to get a number N; then roll N fair dice and let S be their total. Find $E[S]$.

Solution. Condition on N. Given $N = n$, the inner sum has mean $E[S \mid N = n] = n \cdot \frac{7}{2}$, so as a random variable $E[S \mid N] = \frac{7}{2}N$. By the tower property,

$$E[S] = E\big[\tfrac{7}{2}N\big] = \tfrac{7}{2}\, E[N] = \tfrac{7}{2} \cdot \tfrac{7}{2} = \tfrac{49}{4} = 12.25.$$

This is **Wald's identity** in miniature: the mean of a random-length sum of i.i.d. terms is (expected length)$\times$(term mean).

Expected hitting time on a path (first-step analysis)

A frog sits at position k on $\{0, 1, 2, 3, 4\}$ and each second hops left or right with equal probability; positions 0 and 4 are absorbing. Let h_k be the expected time to get absorbed. Conditioning on the first hop (the tower property with $Y =$ first move),

$$h_k = 1 + \tfrac{1}{2}h_{k-1} + \tfrac{1}{2}h_{k+1}, \qquad h_0 = h_4 = 0.$$

Solving the symmetric system gives $h_k = k(4-k)$: $h_1 = h_3 = 3$ and $h_2 = 4$. This is the general gambler's-ruin duration $k(N-k)$ on $\{0, \ldots, N\}$, derived here purely by conditioning on the first step.

Recipe. To use the tower property: (1) pick a variable Y that, once known, makes X easy to average; (2) write $E[X \mid Y = y]$ for each y; (3) average against $P(Y = y)$. When Y is "the first step," this reproduces the one-equation-per-state method — conditional expectation and first-step analysis are the same tool wearing different hats.

10.11 Probability Generating Functions & the Tail-Sum Formula

Encoding a distribution in a single function

For a nonnegative integer-valued X, its **probability generating function** (PGF) packs all the probabilities into one power series:

$$G_X(s) = E\left[s^X\right] = \sum_{k \geq 0} P(X = k)\, s^k.$$

Three facts make PGFs powerful on contests:

- **Normalization:** $G_X(1) = \sum_k P(X = k) = 1$.
- **Coefficients are probabilities:** the coefficient of s^k in $G_X(s)$ is exactly $P(X = k)$, so extracting one coefficient answers a "what is the chance the total is k" question with pure polynomial bookkeeping.
- **Sums multiply:** if X, Y are *independent*, then $G_{X+Y}(s) = G_X(s)\, G_Y(s)$. Convolving distributions becomes multiplying polynomials.

For the *mean*, the workhorse is the **tail-sum formula**: for a nonnegative integer variable,

$$E[X] = \sum_{k \geq 1} P(X \geq k).$$

It re-derives many waiting-time means in one line and is often the fastest route when the tail probabilities $P(X \geq k)$ are simpler than the point probabilities $P(X = k)$.

PGF of a sum, then a tail-sum shortcut

PGF part. A single die has PGF $G(s) = \frac{1}{6}(s + s^2 + \cdots + s^6)$. The sum S of two independent dice therefore has $G_S(s) = G(s)^2$; the coefficient of s^7 in $G(s)^2$ is $\frac{6}{36} = \frac{1}{6}$, recovering $P(S = 7) = \frac{1}{6}$ by polynomial multiplication rather than casework. And by linearity of expectation, $E[S] = E[\text{die } 1] + E[\text{die } 2] = \frac{7}{2} + \frac{7}{2} = 7$ — no new machinery needed for the mean.

Tail-sum part. Roll a die until the first 6; let N be the number of rolls. Then $P(N \geq k) = \left(\frac{5}{6}\right)^{k-1}$ (the first $k - 1$ rolls all miss). So

$$E[N] = \sum_{k \geq 1} P(N \geq k) = \sum_{k \geq 1} \left(\frac{5}{6}\right)^{k-1} = \frac{1}{1 - \frac{5}{6}} = 6,$$

recovering the geometric mean $1/p$ with a single geometric series and no distribution-by-distribution bookkeeping.

Key insight. Reach for a PGF when a problem involves a **sum of independent variables**: multiply their PGFs and read the answer off a coefficient. Reach for **linearity of expectation** for means of sums. Reach for the **tail-sum formula** $E[X] = \sum_{k \geq 1} P(X \geq k)$ when the "at least k" probabilities are clean — it turns a mean into a single series and is the slickest derivation of most waiting-time expectations.

11 Course Review

Every formula, identity, and technique for competition math

11.1 Fundamental Counting

Addition & Multiplication Principles

Multiplication (AND). A task done in independent stages with $n_1, n_2, \ldots, n_k$ ways per stage has $n_1 n_2 \cdots n_k$ total outcomes. **Addition (OR).** Disjoint cases add: $|A_1| + \cdots + |A_k|$.
Ex. 2 letters then 3 digits: $26^2 \cdot 10^3 = 676{,}000$ plates.

Permutations (order matters, no repeats)

Arrange k of n distinct objects in ordered slots:

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

Ex. Gold/silver/bronze among 8 runners: $P(8, 3) = 8 \cdot 7 \cdot 6 = 336$.

Combinations (order doesn't matter)

Choose k of n distinct objects, ignoring order:

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

Ex. A 4-person committee from 10: $\binom{10}{4} = 210$.

Permutations with Identical Objects (multinomial)

Arrange n objects with n_i copies of type i ($\sum n_i = n$):

$$\frac{n!}{n_1!\,n_2!\cdots n_r!}.$$

Ex. `MISSISSIPPI`: $\dfrac{11!}{1!\,4!\,4!\,2!} = 34{,}650$.

Circular Permutations

Arrangements around a circle count only relative position (fix one seat to remove n rotations):

$$(n-1)!; \qquad \text{if reflections also identical (bracelet): } \frac{(n-1)!}{2} \ (n \geq 3).$$

Ex. 6 people at a round table: $(6-1)! = 120$.

Tip. Titles, ranks, and ordered slots signal permutations; committees, subsets, and hands signal combinations. Labeled chairs make a "circular" problem linear again ($n!$, not $(n-1)!$).

11.2 Counting Strategies

Casework

Split into disjoint, exhaustive cases; count each; add. Choose the split that makes each case easy and keeps cases from overlapping.
Ex. Rolls of two dice summing to a multiple of 4: cases sum$= 4, 8, 12$ give $3+5+1=9$.

Complementary Counting

Count the opposite and subtract from the total: $|E| = |S| - |E^c|$. Ideal for "at least one."
Ex. 3-digit numbers with a digit 7: $900 - 8 \cdot 9 \cdot 9 = 252$.

Constructive Counting

Build a valid object step by step, multiplying the choices available at each step (keep the count per step constant, or split into cases).
Ex. 5 books, dictionary on an end: $2 \cdot 4! = 48$.

Correcting Overcounting (divide by symmetry)

Count an easy ordered/labeled way, then divide by the symmetries you introduced: rotations ($\div n$), reflections ($\div 2$), interchangeable groups ($\div k!$).
Ex. Split 6 players into 3 unlabeled pairs: $\dfrac{\binom{6}{2}\binom{4}{2}\binom{2}{2}}{3!} = \dfrac{90}{6} = 15$.

Bijections

If a set is hard to count, find a one-to-one correspondence with a set you can count; equal sizes.

Ex. Grid paths $(0,0) \to (4,3) \leftrightarrow$ strings of 4 R's, 3 U's: $\binom{7}{3} = 35$.

Tip. "At least one" $\Rightarrow$ complement. When a direct build overcounts by a fixed factor, count with order and divide. A clever bijection often turns a messy object into a simple string or lattice path.

11.3 Distributions: Stars & Bars

Nonnegative Integer Solutions

The number of solutions to $x_1+x_2+\cdots+x_k = n$ with each $x_i \geq 0$ (identical items into distinct boxes) is

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

Ex. 10 identical candies among 4 kids: $\binom{13}{3} = 286$.

Positive Integer Solutions

For $x_i \geq 1$, give each variable one unit first ($y_i = x_i - 1 \geq 0$):

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

Ex. 10 candies, each kid at least one: $\binom{9}{3} = 84$.

Bounded Variables (inclusion–exclusion)

For upper bounds like $x_i \leq c_i$, subtract the "overflow" cases with inclusion–exclusion (give a violating variable $c_i + 1$ units, then recount).
Ex. $x_1 + x_2 + x_3 = 10$, each ≤ 4: $\binom{12}{2} - 3\binom{7}{2} + 3\binom{2}{2} = 66 - 63 + 3 = 6$.

Multinomial Coefficients (distinct items)

Split n *distinct* items into labeled groups of sizes $n_1, \ldots, n_r$:

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!}.$$

Ex. Deal 9 distinct cards into piles of $4, 3, 2$: $\dfrac{9!}{4!\,3!\,2!} = 1260$.

Tip. Stars and bars needs *identical* items and *distinct* boxes. "Each at least one" subtracts k from n first; upper bounds require inclusion–exclusion; distinct items use the multinomial instead.

11.4 Binomial Theorem & Identities

Binomial Theorem

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

Ex. Coefficient of x^2 in $(1+x)^5$ is $\dbinom{5}{2} = 10$.

Pascal's Rule

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

Ex. $\dbinom{5}{2} = \dbinom{4}{1} + \dbinom{4}{2} = 4 + 6 = 10$.

Row Sums

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

Ex. Subsets of a 5-set: $2^5 = 32$; equal numbers of even- and odd-sized subsets.

Hockey Stick

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

Ex. $\dbinom{2}{2} + \dbinom{3}{2} + \dbinom{4}{2} = 1 + 3 + 6 = 10 = \dbinom{5}{3}$.

Vandermonde's Identity

$$\sum_{k=0}^{p}\binom{m}{k}\binom{n}{p-k}=\binom{m+n}{p}.$$

Ex. Pick 2 from 3 boys + 4 girls: $\sum_k \binom{3}{k}\binom{4}{2-k}=\binom{7}{2}=21$.

Absorption & Lattice Paths

$$k\binom{n}{k}=n\binom{n-1}{k-1}; \qquad \text{paths } (0,0)\to(m,n) \text{ (R/U steps)} = \binom{m+n}{m}.$$

Ex. $3\binom{6}{3}=6\binom{5}{2}=60$; paths to $(4,3)$: $\binom{7}{4}=35$.

Tip. Prove identities combinatorially (count one set two ways) when algebra stalls. To evaluate a weird sum, match it to the binomial theorem by choosing clever x, y (e.g. $x=y=1$ gives 2^n; $x=1, y=-1$ gives 0).

11.5 Inclusion–Exclusion, Derangements & Pigeonhole

Inclusion–Exclusion (PIE)

$$\left|\bigcup_i A_i\right|=\sum|A_i|-\sum|A_i\cap A_j|+\sum|A_i\cap A_j\cap A_k|-\cdots$$

Ex. ≤ 100 divisible by 2 or 3: $50+33-16=67$.

Derangements (no fixed points)

$$D_n=n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}=(n-1)(D_{n-1}+D_{n-2}), \qquad D_n\approx\frac{n!}{e}.$$

Ex. 4 letters, wrong envelopes: $D_4=9$.

Surjections (onto functions)

Functions from an n-set onto a k-set, by PIE:

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

Ex. Onto maps $\{1, 2, 3\} \to \{a, b\}$: $2^3 - 2 = 6$.

Pigeonhole Principle

Place n items into k boxes: some box holds at least $\left\lceil \frac{n}{k} \right\rceil$ items (basic case: $n > k \Rightarrow$ a box has ≥ 2).
Ex. 13 people share a birth month: some month has $\lceil 13/12 \rceil = 2$.

Tip. PIE shines for "at least one property" with overlapping conditions. For pigeonhole, the art is choosing the pigeons and holes; look for a quantity forced to repeat.

11.6 Recursion, Catalan & Generating Functions

Recursion & Fibonacci Tilings

Set up a_n from smaller cases by conditioning on the last choice. Tiling a $1 \times n$ strip with 1×1 and 1×2 pieces:

$$a_n = a_{n-1} + a_{n-2} \text{ (Fibonacci)}, \quad a_1 = 1,\ a_2 = 2.$$

Ex. 1×4 strip: $a_4 = 5$ tilings.

Catalan Numbers

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \binom{2n}{n} - \binom{2n}{n+1}, \qquad C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}.$$

Counts balanced parentheses, non-crossing paths, triangulations, binary trees.
Ex. Valid sequences of 3 pairs of parentheses: $C_3 = 5$.

Generating Functions

Encode a sequence a_n as coefficients of $\sum a_n x^n$; products of series combine choices. Key:

$$\frac{1}{1-x} = \sum_{n \geq 0} x^n, \qquad (1+x)^n = \sum_k \binom{n}{k} x^k.$$

Ex. Ways to make 6 from parts $\{1, 2\}$: coefficient of x^6 in $\frac{1}{(1-x)(1-x^2)}$ is 4.

Tip. When a count depends on a process built up one step at a time, define a recurrence by conditioning on the first or last move. Recognize Catalan whenever "balanced/non-crossing/neverdips-below" appears.

11.7 Probability Fundamentals

Equally Likely Outcomes

$$P(E) = \frac{|E|}{|S|}, \qquad 0 \le P(E) \le 1, \qquad P(E^c) = 1 - P(E).$$

Ex. Two dice sum to 7: $\dfrac{6}{36} = \dfrac{1}{6}$.

Counting-Based Probability

Compute favorable and total outcomes with the same counting rules (order-consistent on both).

Ex. 2 aces in a 5-card hand: $\dfrac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}$.

Independence

Events are independent iff

$$P(A \cap B) = P(A)\,P(B).$$

Ex. Two coins both heads: $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Geometric Probability

When outcomes fill a continuous region, probability is a ratio of measures (length, area, volume):

$$P = \frac{\text{measure of favorable region}}{\text{measure of sample region}}.$$

Ex. Random point in $[0,1]^2$ with $x + y < 1$: area $\frac{1}{2}$, so $P = \frac{1}{2}$.

Tip. Keep numerator and denominator counted the same way (both ordered or both unordered). "At least one" in probability also loves the complement: $P(\ge 1) = 1 - P(\text{none})$.

11.8 Conditional Probability & Bayes

Conditional Probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0.$$

Ex. Die shows even, chance it's 2: $\dfrac{1/6}{1/2} = \dfrac{1}{3}$.

Multiplication Rule

$$P(A \cap B) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B).$$

Ex. Two cards, both hearts (no replacement): $\dfrac{13}{52} \cdot \dfrac{12}{51} = \dfrac{1}{17}$.

Law of Total Probability

For a partition $B_1, \dots, B_n$ of the sample space:

$$P(A) = \sum_i P(A \mid B_i)\,P(B_i).$$

Ex. Urn A (bag $\frac{1}{2}$): $P(\text{red}) = \frac{1}{2} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{1}{3} = \frac{11}{30}$.

Bayes' Theorem

$$P(B \mid A) = \frac{P(A \mid B)\,P(B)}{P(A)} = \frac{P(A \mid B)P(B)}{\sum_i P(A \mid B_i)P(B_i)}.$$

Ex. Test 99% accurate, disease rate 1%: $P(\text{sick} \mid +) = \dfrac{.99(.01)}{.99(.01) + .01(.99)} = \frac{1}{2}$.

Tip. "Given that" $\Rightarrow$ conditional; restrict the sample space to the given event. When you know $P(\text{evidence} \mid \text{cause})$ but want $P(\text{cause} \mid \text{evidence})$, use Bayes. Rare conditions make "accurate" tests deceptively unreliable.

11.9 Expected Value

Definition

$$E[X] = \sum_x x\,P(X = x).$$

Ex. One fair die: $E[X] = \dfrac{1 + 2 + \cdots + 6}{6} = 3.5$.

Linearity of Expectation

Always true, even for dependent variables:

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n].$$

Ex. Sum of 10 dice: $10 \cdot 3.5 = 35$.

Indicator Variables

Let $X_i = 1$ if event i occurs, else 0; then $E[X_i] = P(\text{event } i)$ and $E[\#\text{events}] = \sum P(\text{event } i)$.
Ex. Expected fixed points of a random permutation of n: $n \cdot \frac{1}{n} = 1$.

Expectation via States / Recursion

Let E be the expected value from a state; condition on the next step and solve the resulting equation(s).
Ex. Flips until first head: $E = 1 + \frac{1}{2}E \Rightarrow E = 2$.

Waiting Time & Coupon Collector

Geometric waiting time for a prob-p event: $E = \dfrac{1}{p}$. Collecting all n coupons:

$$E = n\Big(1 + \tfrac{1}{2} + \cdots + \tfrac{1}{n}\Big) = nH_n.$$

Ex. All 6 die faces: $6H_6 = 6 \cdot \frac{49}{20} = 14.7$.

Tip. "Expected number of . . . " almost always means indicators + linearity — never fight the dependence. For a repeating process with memory, set up a states equation and solve for E.

11.10 Problem-Solving Playbook

Cue $\rightarrow$ Technique decision guide

- order matters, no repeat $\to P(n,k)$
- order doesn't matter $\to \binom{n}{k}$
- repeated letters/objects $\to$ multinomial $\frac{n!}{n_1!\cdots n_r!}$
- around a circle $\to (n-1)!$
- "at least one" $\to$ complement $1 - P(\text{none})$
- distribute identical items $\to$ stars and bars
- distinct items into groups $\to$ multinomial
- overlapping conditions / "or" $\to$ inclusion–exclusion
- nothing in its place $\to$ derangement D_n
- something forced to repeat $\to$ pigeonhole
- balanced / non-crossing $\to$ Catalan C_n
- process built step by step $\to$ recursion
- "given that" / updated info $\to$ conditional / Bayes
- continuous / random point $\to$ geometric probability
- expected count of something $\to$ linearity + indicators
- repeated states with memory $\to$ expectation via states
- time until first success $\to$ waiting time $1/p$
- weird binomial sum $\to$ known identity / count two ways

Master flow. (1) AND $\Rightarrow$ multiply, OR $\Rightarrow$ add. (2) Does order matter? (3) Did I overcount by symmetry? (4) Is the complement easier? (5) For probability, count favorable and total the same way; for expectation, reach for linearity before anything clever.

11.11 Reference Tables

Factorials $0!$ through $10!$

Catalan numbers C_0–C_8 and Derangements D_0–D_6

Pascal's Triangle (rows $n = 0$ to 8)

$$
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1 \\
1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1
\end{array}
$$

Row n sums to 2^n; entry k is $\binom{n}{k}$; each entry is the sum of the two above it (Pascal's rule).

11.12 Advanced Reference

Burnside's Lemma (counting up to symmetry)

The number of distinct configurations under a symmetry group G equals the average number of colorings fixed by each group element:

$$\#\text{orbits} = \frac{1}{|G|} \sum_{g \in G} \big|\mathrm{Fix}(g)\big|.$$

Ex. 2-color a necklace of 4 beads under 4 rotations: $\frac{1}{4}(2^4 + 2^1 + 2^2 + 2^1) = 6$.

Stirling Numbers (2nd kind) & Bell Numbers

$S(n,k)$ partitions an n-set into k nonempty unlabeled blocks; Bell number $B_n = \sum_k S(n,k)$ counts all partitions:

$$S(n,k) = S(n-1,k-1) + k\,S(n-1,k), \qquad B_n = \sum_{k=0}^{n} S(n,k).$$

Surjections from n-set onto k-set $= k!\,S(n,k)$.
Ex. $S(4,2) = 7$, so onto maps $\{1,2,3,4\} \to \{a,b\}$ number $2! \cdot 7 = 14$.

Derangement Nearest-Integer Formula

Since $D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}$ converges to $n!/e$ fast,

$$D_n = \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor \qquad (n \geq 1).$$

Ex. $D_5 = \lfloor 120/e + 0.5 \rfloor = \lfloor 44.65 \rfloor = 44$.

Roots of Unity Filter

To sum binomial coefficients over k in one residue class mod m, average $(1+\omega^j)^n$ over the m-th roots of unity $\omega = e^{2\pi i/m}$:

$$\sum_{k\equiv r\ (m)} \binom{n}{k} = \frac{1}{m}\sum_{j=0}^{m-1} \omega^{-jr}(1+\omega^j)^n.$$

Ex. $\sum_{k\equiv 0\ (2)} \binom{n}{k} = \frac{1}{2}\big(2^n + 0^n\big) = 2^{n-1}$ for $n \geq 1$.

Generalized Binomial Series

For any real (or complex) α with $|x|<1$, and the negative-power expansion:

$$(1+x)^\alpha = \sum_{k\geq 0}\binom{\alpha}{k}x^k, \qquad \frac{1}{(1-x)^k} = \sum_{n\geq 0}\binom{n+k-1}{k-1}x^n.$$

Here $\binom{\alpha}{k} = \dfrac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}$.

Ex. $\dfrac{1}{(1-x)^3}$ has x^n coefficient $\dbinom{n+2}{2}$ (stars and bars).

Key Weighted & Squared Sum Identities

$$\sum_{k=0}^{n} k\binom{n}{k} = n\,2^{n-1}, \qquad \sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n},$$

and Vandermonde $\displaystyle\sum_k \binom{m}{k}\binom{n}{p-k} = \binom{m+n}{p}$ (the squared sum is its $m=n=p$ case).

Ex. $\sum k\binom{4}{k} = 4\cdot 2^3 = 32$; $\sum \binom{4}{k}^2 = \binom{8}{4} = 70$.

Catalan Interpretations & Closed Form

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \binom{2n}{n} - \binom{2n}{n-1}.$$

Counts: balanced parentheses, Dyck paths never dipping below the axis, triangulations of an $(n{+}2)$-gon, full binary trees with $n{+}1$ leaves, and non-crossing chord matchings.
Ex. $C_4 = \binom{8}{4} - \binom{8}{3} = 70 - 56 = 14$.

Solving Linear Recurrences (characteristic roots)

For $a_n = c_1a_{n-1}+\cdots+c_da_{n-d}$, solve $x^d = c_1x^{d-1}+\cdots+c_d$. Distinct roots r_i give $a_n = \sum A_i r_i^n$;

a root of multiplicity m contributes $(A_0+A_1n+\cdots+A_{m-1}n^{m-1})r^n$. Fix constants from initial values.

Ex. $a_n = 2a_{n-1} - a_{n-2}$ has double root 1: $a_n = A + Bn$ (arithmetic).

Solving Recurrences (generating functions)

Multiply the recurrence by x^n, sum over n, and solve for $A(x) = \sum a_n x^n$ as a rational function; expand via partial fractions and $\frac{1}{(1-rx)^m}$ to read off a_n.

Ex. Fibonacci: $A(x) = \dfrac{x}{1-x-x^2} \Rightarrow a_n = \dfrac{\varphi^n - \psi^n}{\sqrt{5}}$.

11.13 Advanced Expectation & Markov Chains

Expectation Toolkit

Beyond linearity and indicators ($E[\sum X_i] = \sum P(\text{event } i)$), for a nonnegative integer X use the tail sum:

$$E[X] = \sum_{k\geq 1} P(X \geq k), \qquad \mathrm{Var}(X) = E[X^2] - E[X]^2.$$

Geometric wait for a prob-p event: $E = 1/p$. Coupon collector over n types: $E = nH_n$.

Ex. Rolls to see all 6 faces: $6H_6 = 14.7$.

Odds-Form Bayes

Updating hypothesis H against $\bar{H}$ on evidence E multiplies prior odds by the likelihood ratio:

$$\frac{P(H \mid E)}{P(\bar{H} \mid E)} = \frac{P(H)}{P(\bar{H})} \cdot \frac{P(E \mid H)}{P(E \mid \bar{H})}.$$

Ex. Prior odds 1:99, likelihood ratio 99:1 $\Rightarrow$ posterior odds 1:1, i.e. $P = \frac{1}{2}$.

First-Step Analysis & Gambler's Ruin

Define hitting probabilities/times per state and condition on the first step. Fair-coin gambler's ruin from \$$k$ with absorbing barriers 0 and N:

$$P(\text{reach } N) = \frac{k}{N}, \qquad E[\text{duration}] = k(N-k).$$

Ex. Start at \$3 of \$10: reach goal with prob $\frac{3}{10}$, expected 21 steps.

Stationary Distribution

A distribution π is stationary for transition matrix P when it is unchanged by a step:

$$\pi P = \pi, \qquad \sum_i \pi_i = 1.$$

For an irreducible chain the long-run fraction of time in state i is π_i, and the mean return time is $1/\pi_i$.
Ex. Two-state chain with rates a (out of 1), b (out of 2): $\pi = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$.

Tip. "Distinct up to rotation/reflection" $\Rightarrow$ Burnside. "Partition into groups" (unlabeled) $\Rightarrow$ Stirling/Bell. A weird sum over every m-th term $\Rightarrow$ roots of unity filter. For long-run behavior of a random process, solve $\pi P = \pi$; for hitting probabilities, use first-step analysis.

Stirling Numbers $S(n,k)$ (2nd kind), $n \le 6$

Row sums give the Bell numbers.

Bell Numbers B_0–B_6

B_n counts all set partitions of an n-set; $B_n = \sum_k \binom{n-1}{k} B_k$.