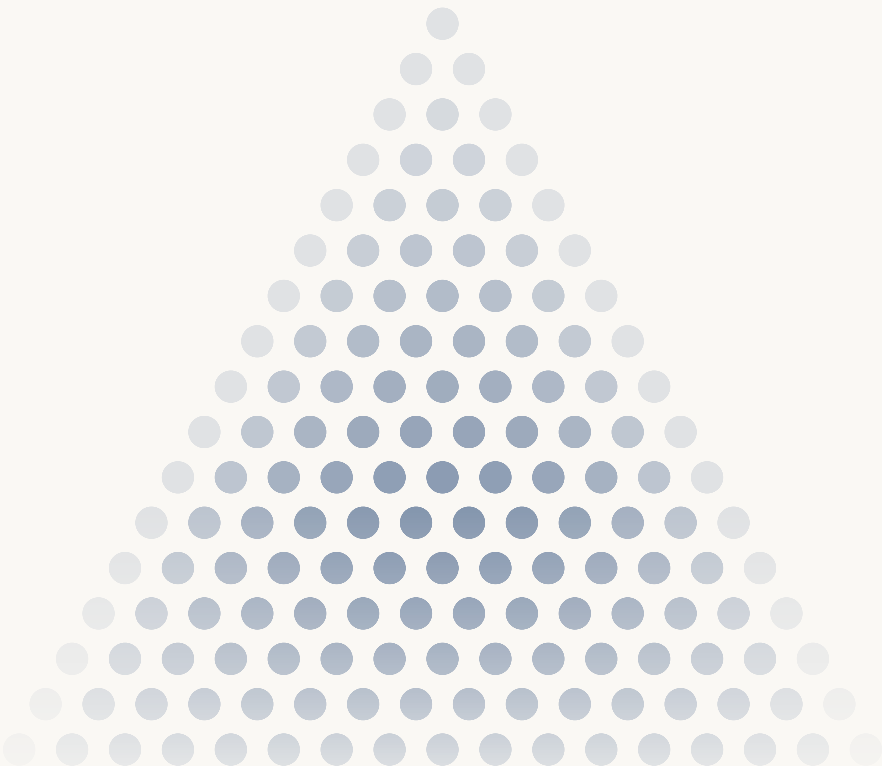


A VISUAL MINI-TEXTBOOK

Counting & Probability

From the product rule to Poisson: permutations, combinations, the binomial theorem, and the probability they generate.

I. Counting

Rules of counting, $P(n,k)$ and $\binom{n}{k}$, Pascal identities, the binomial and multinomial theorems, stars and bars, inclusion–exclusion, bijections.

II. Probability

Axioms, conditional probability and Bayes, expectation by indicators, variance, the named discrete distributions, classic problems.

III. Reference

A formula spread, a distribution table, a "which formula do I use?" decision guide, and problems with full solutions.

Prepared for Ben • September 3, 2026

Contents

How to read this

Everything in the first half of this book is one idea wearing different hats: *if you can describe an object uniquely by a sequence of independent choices, you can count the objects by multiplying the number of choices.* Permutations, combinations, $\binom{n}{k}$, the binomial theorem, stars and bars: all of them are that sentence, plus a correction for overcounting.

Everything in the second half is one further idea: *probability on a finite set where every outcome is equally likely is nothing but counting, divided.* Once outcomes stop being equally likely you need the axioms, and once you want averages you need expectation, but the counting never goes away.

The two-line summary of the whole book

Counting. Build the object by a sequence of choices; multiply; then divide by the number of times you built the same object twice.

Probability. $\mathbb{P}(A) = \dfrac{|A|}{|S|}$ when outcomes are equally likely; otherwise use the axioms; and when you want an average, use $\mathbb{E}[\sum_i X_i] = \sum_i \mathbb{E}[X_i]$, which never asks whether the X_i are independent.

Conventions used throughout

Notation

$n!$	$n \cdot (n-1) \cdots 2 \cdot 1$, with $0! = 1$.
$P(n,k)$, $n^{\underline{k}}$	the number of ordered k-selections from n distinct objects; also written ${}^nP_k$ or $n^{\underline{k}}$ (the *falling factorial*).
$\binom{n}{k}$	the number of unordered k-subsets of an n-set; also written ${}^nC_k$, $C(n,k)$, or "n choose k".
$n^{\overline{k}}$	the *rising factorial* $n(n+1)\cdots(n+k-1)$.
$[n]$	the set $\{1, 2, \ldots, n\}$.
$\lvert A \rvert$	the number of elements of the finite set A.
S	the sample space; A, B, C denote events (subsets of S).
$\mathbf{1}_A$	the indicator random variable of the event A: 1 if A happens, 0 if not.
$X \sim \mathcal{D}$	the random variable X has distribution $\mathcal{D}$.

The colour code

- **Blue boxes** are definitions: what a symbol means.
- **Green boxes** are theorems and identities: what is true.
- **Amber boxes** are worked examples.
- **Purple boxes** are the idea behind the algebra, in words.
- **Red boxes** are the mistakes that actually get made.

If you read nothing else, read the purple boxes and the reference spread at the end.

PART I

Counting

1 The two rules everything is built from

There are exactly two counting rules. Every formula later in this book is a consequence of applying them carefully and then correcting for overcounting.

The sum rule

If a set A can be split into pieces $A_1, A_2, \ldots, A_m$ that are *pairwise disjoint* (no object lies in two pieces) and together make up all of A, then

$$|A| = |A_1| + |A_2| + \cdots + |A_m|.$$

In words: **case-split**, then add. The word *or* usually signals a sum.

The product rule

If an object is built by making a sequence of k choices, where the ith choice can be made in n_i ways *no matter how the earlier choices went*, then the number of objects is

$$n_1 n_2 \cdots n_k.$$

In words: **choose in stages**, then multiply. The word *and* usually signals a product.

The clause "no matter how the earlier choices went" is the whole game. The number of options at stage i may *depend* on the earlier choices; what it may not do is *change in count*. Arranging three of five books on a shelf gives $5 \times 4 \times 3$: which three books remain available depends entirely on what you picked first, but *how many* remain does not.

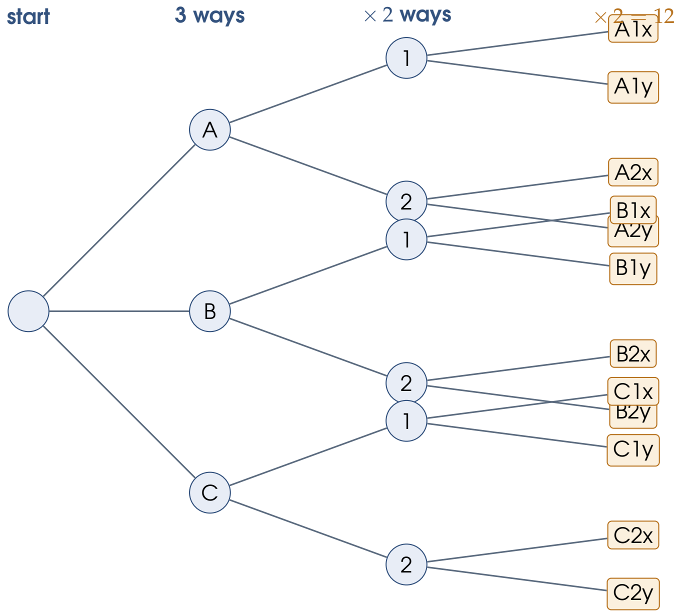


Figure 1: *The product rule is a tree. Each leaf is one object; the number of leaves is the product of the branching factors, provided every node at a given depth has the same number of children.*

1.1 Three corollaries you will use constantly

Immediate consequences

Let A and B be finite sets, $|A| = m$, $|B| = n$.

1. **Strings.** The number of functions $A \to B$, equivalently the number of length-m strings over an alphabet of size n, is n^m. (Product rule: n choices for each of the m inputs.)
2. **Subsets.** The number of subsets of A is 2^m. (Each element is in or out: a function $A \to \{\text{in}, \text{out}\}$.)
3. **Complement.** If $A \subseteq S$, then $|A| = |S| - |S \setminus A|$. Counting the objects you do *not* want is often far easier.

At least one

How many strings of length 8 over $\{a, b, c\}$ contain at least one a?
Counting directly means splitting on how many a's appear: eight cases. Counting the complement means counting strings with *no* a, which is just strings over $\{b, c\}$:

$$3^8 - 2^8 = 6561 - 256 \boxed{6305}.$$

Rule of thumb. "At least one" almost always wants the complement.

1.2 The bijection principle

Bijection principle

If there is a one-to-one correspondence between the elements of A and the elements of B, then $|A| = |B|$.

This sounds like a triviality and is in fact the single most powerful technique in combinatorics. You do not count the hard set; you find an easy set it is secretly the same as. Section 7 is devoted to it, but here is the flavour: the number of subsets of $[n]$ equals the number of binary strings of length n, because "is element i in the subset?" is exactly "is bit i a one?". That observation converts a set problem into a string problem, and the string problem is just the product rule.

The overcounting trap

The product rule counts *procedures*, not objects. If two different sequences of choices produce the same object, the product rule counts that object twice. Almost every combinatorics error is this error. The fix is always the same: if every object arises from exactly d procedures, divide by d. That division is where $\binom{n}{k}$ comes from.

2 Permutations: when order matters

Permutation, and the k-permutation $P(n,k)$

A **permutation** of a set is an ordering of all its elements. A **k-permutation** of an n-set is an ordered selection of k of its elements, without repetition. Their counts are

$$n! = n(n-1)\cdots 2\cdot 1, \qquad P(n,k) = \frac{n!}{(n-k)!} = \underbrace{n(n-1)\cdots(n-k+1)}_{k\text{ factors}} = n^{\underline{k}}.$$

Conventions: $0! = 1$, $P(n,0) = 1$, $P(n,n) = n!$, and $P(n,k) = 0$ if $k > n$.

The second form of $P(n,k)$ is the one to memorise, because it is the one the product rule actually produces: n choices for the first slot, $n-1$ for the second, and so on down to $n-k+1$ for the kth. The factorial-quotient form is a *simplification* of that product, not its origin. Writing $P(9,3)$ as $9\cdot 8\cdot 7 = 504$ is faster and safer than writing $9!/6!$.

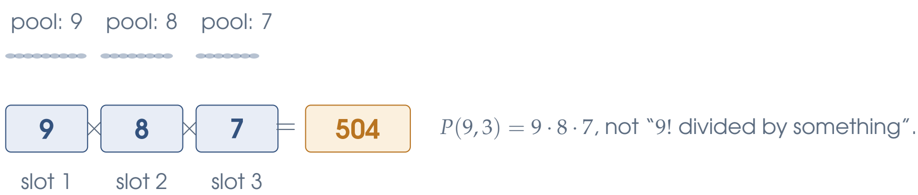


Figure 2: *A k-permutation is the product rule run on a shrinking pool.*

2.1 The four flavours of arrangement

Arranging things in a row

Situation	Description	Count
All n distinct	order all of them	$n!$
k of n distinct	order a selection	$\frac{n!}{(n-k)!} = P(n,k)$
k from n, repeats allowed	each slot independently	n^k
n objects, repeated types	n_1 of type $1, \ldots, n_r$ of type r	$\frac{n!}{n_1!\,n_2!\cdots n_r!}$

The last line is the **multinomial coefficient**, written $\binom{n}{n_1,\, n_2,\, \ldots,\, n_r}$. Its derivation is the overcounting correction in its purest form: pretend the identical objects are distinguishable (giving $n!$ arrangements), then notice that permuting the n_1 copies of type 1 among themselves changes nothing, and likewise for every other type. Each genuine arrangement was therefore counted $n_1!\,n_2!\cdots n_r!$ times.

MISSISSIPPI

The word MISSISSIPPI has 11 letters: one M, four I, four S, two P. The number of distinguishable arrangements is

$$\frac{11!}{1!\;4!\;4!\;2!} = \frac{39\,916\,800}{1\cdot 24\cdot 24\cdot 2} = \boxed{34\,650}.$$

2.2 Circular arrangements

Round a table

The number of ways to seat n distinct people around a circular table, where only the *relative* order matters, is $(n-1)!$. If the arrangement can also be flipped over (a necklace, or a table you may view from either side), the count is $\dfrac{(n-1)!}{2}$ for $n \geq 3$.

Why $(n-1)!$: each circular arrangement corresponds to n linear arrangements, one for each choice of who is written down first, so $n!/n = (n-1)!$. Equivalently, fix one person as a reference point and arrange the remaining $n-1$ people relative to them. The second view is usually the faster one in practice, and it generalises: to handle a restriction, fix the restricted object first.

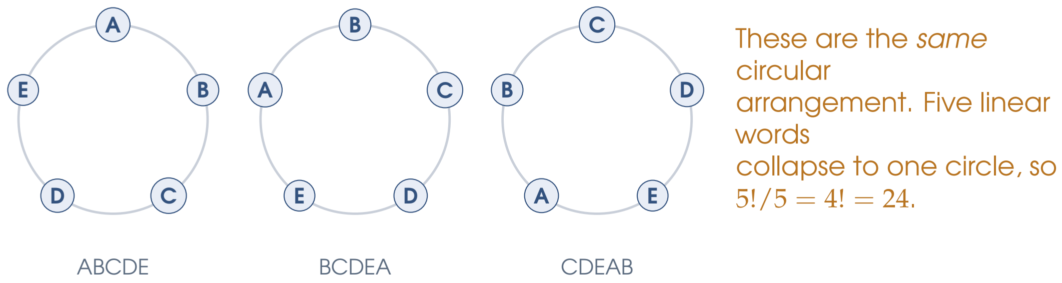


Figure 3: *Rotations of a circular seating are indistinguishable, so divide by n.*

2.3 Two standard restriction techniques

Glue and gap

Glue (objects must be together). Treat the block that must stay together as a single super-object, arrange everything, then multiply by the internal arrangements of the block. Arranging 7 people with a specific 3 insisting on adjacency: $5! \times 3! = 720$.

Gap (objects must be apart). Arrange the unrestricted objects first, then slot the restricted ones into the gaps between them, at most one per gap. Arranging 5 boys and 3 girls so that no two girls are adjacent: place the boys (5! ways), creating 6 gaps, then choose 3 gaps and order the girls into them ($P(6,3) = 120$). Total $5! \times 120 = 14\,400$.

$P(n,k)$ versus n^k

$P(n,k)$ is *without* replacement; n^k is *with* replacement. A four-digit PIN is 10^4, because digits may repeat. A four-digit PIN with distinct digits is $P(10,4) = 5040$. Read the problem for the word "repeat", "distinct", "replaced", or "different" before writing anything down.

3 Combinations: when order does not matter

The binomial coefficient

For integers $0 \le k \le n$, the number of k-element subsets of an n-element set is

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

Set $\binom{n}{k} = 0$ when $k < 0$ or $k > n$. Read it as "n choose k".

Why you divide by $k!$

Count ordered selections: $P(n,k)$. Now ask how many times each *unordered* k-set got counted. Once for every ordering of its k elements, which is $k!$ times, the same number for every set. So dividing by $k!$ is exact, not an approximation:

$$\underbrace{P(n,k)}_{\text{ordered}} = \underbrace{\binom{n}{k}}_{\text{unordered}} \times k!.$$

Order matters $\Rightarrow P(n,k)$. Order does not $\Rightarrow \binom{n}{k}$. That is the entire distinction.

Compute it the short way

$$\binom{10}{3} = \frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1} = \frac{720}{6} = 120.$$
$$\binom{52}{5} = \frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{120} = 2\,598\,960 \text{ (the number of five-card poker hands).}$$

Never expand the factorials. Cancel as you go: $\binom{100}{2} = \frac{100\cdot 99}{2} = 4950$.

3.1 Pascal's triangle

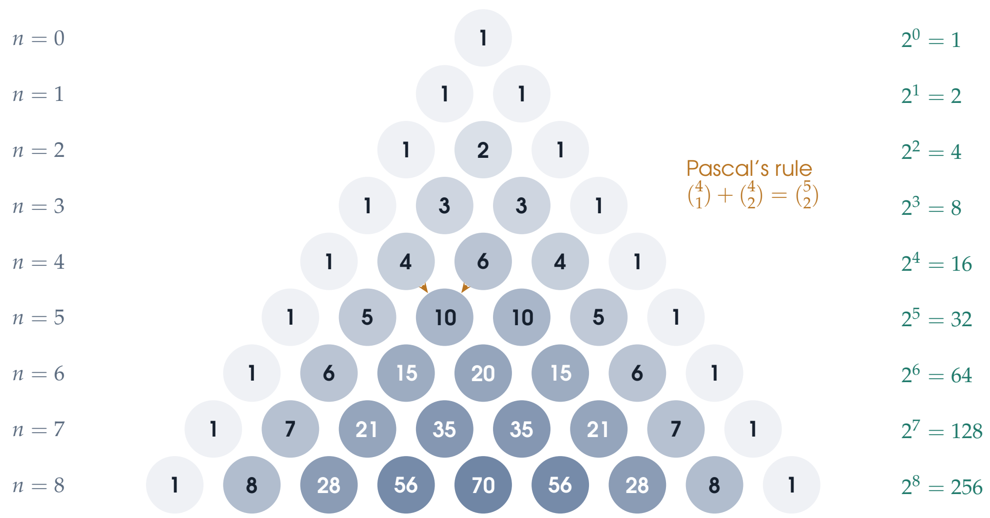


Figure 4: *Pascal's triangle, shaded by magnitude. Each entry is the sum of the two above it (Pascal's rule); each row sums to 2^n; each row is symmetric about its centre.*

Pascal's rule

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

Combinatorial proof. Fix a distinguished element x of the n-set. Every k-subset either contains x or does not. Those containing x are determined by the other $k-1$ elements chosen from the remaining $n-1$: that is $\binom{n-1}{k-1}$. Those avoiding x choose all k elements from the remaining $n-1$: that is $\binom{n-1}{k}$. The two cases are disjoint and exhaustive, so add them.

This proof is the template for the whole subject. To prove a combinatorial identity, find a set, count it two ways, and conclude the counts are equal. It is called **double counting**, and it is nearly always shorter and more illuminating than manipulating factorials.

3.2 The identities worth knowing by heart

Core identities

(a) **Symmetry.** $\binom{n}{k} = \binom{n}{n-k}$.

Choosing the k you keep is the same as choosing the $n-k$ you discard.

(b) **Row sum.** $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Both sides count all subsets of an n-set, one by size, one by deciding in-or-out for each element.

(c) **Alternating sum.** $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$ for $n \ge 1$.

An n-set has as many even-sized subsets as odd-sized ones.

(d) **Absorption / committee-with-chair.** $k\binom{n}{k} = n\binom{n-1}{k-1}$.

Both sides count (committee of size k, with a chair): pick the committee then the chair, or pick the chair then the rest.

(e) **Hockey stick.** $\sum_{i=k}^{n} \binom{i}{k} = \binom{n+1}{k+1}$.

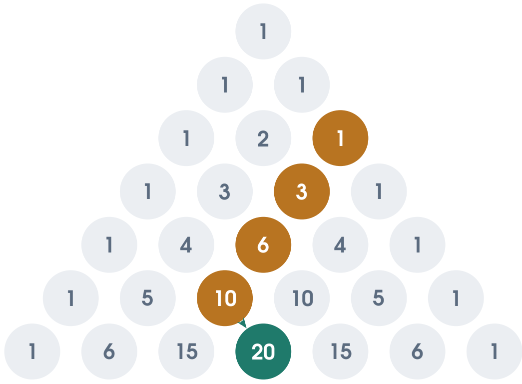
Classify $(k+1)$-subsets of $[n+1]$ by their largest element.

(f) **Vandermonde.** $\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} = \binom{m+n}{k}$.

Choose k people from m women and n men; split on how many women.

(g) **Sum of squares.** $\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$.

Vandermonde with $m = n$, $k = n$, plus symmetry.



$$\binom{2}{2}+\binom{3}{2}+\binom{4}{2}+\binom{5}{2} \;=\; 1+3+6+10 \;=\; 20 \;=\; \binom{6}{3}$$

Figure 5: *The hockey stick: sum a diagonal, land one step down and across.*

Double counting in action: the sum of squares

Choose n people from a group of n women and n men. Directly this is $\binom{2n}{n}$. Alternatively split on the number k of women chosen: $\binom{n}{k}$ ways to pick the women, $\binom{n}{n-k}$ ways to pick the men, and $\binom{n}{n-k} = \binom{n}{k}$ by symmetry. Summing over k:

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

Check at $n = 3$: $1+9+9+1 = 20 = \binom{6}{3}$. ✓

Three habitual errors

- **Dividing twice.** If you already used $\binom{n}{k}$, the order is gone. Do not divide by $k!$ again.
- **Ordering identical groups.** Splitting 12 people into three *unlabelled* teams of 4 is $\frac{1}{3!}\binom{12}{4}\binom{8}{4}\binom{4}{4}$, not $\binom{12}{4}\binom{8}{4}\binom{4}{4}$. If the teams have names (Red, Blue, Green), do *not* divide.
- **Double-counting overlapping cases.** The sum rule requires *disjoint* cases. "Choose a committee containing Alice, or containing Bob" double-counts committees containing both. Use inclusion–exclusion (Section 6).

4 The binomial theorem

Binomial theorem

For any commuting x, y and any integer $n \geq 0$,

$$(x+y)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \;=\; x^n + nx^{n-1}y + \binom{n}{2} x^{n-2}y^2 + \cdots + nxy^{n-1} + y^n.$$

The general term is $T_{k+1} = \binom{n}{k} x^{n-k} y^k$, indexed so that T_1 is the x^n term.

The theorem is a counting statement

Write $(x+y)^n = \underbrace{(x+y)(x+y)\cdots(x+y)}_{n \text{ factors}}$ and expand by choosing, from each factor, either the x or the y. Every choice-sequence produces one monomial. A monomial is $x^{n-k}y^k$ exactly when you chose y from k of the n factors, and the number of ways to do that is $\binom{n}{k}$. So $\binom{n}{k}$ is the coefficient. *The binomial coefficient is called that because of this theorem.*

Extracting one coefficient

Find the coefficient of x^7 in $\left(2x^2 - \frac{3}{x}\right)^8$.
The general term is

$$T_{k+1} = \binom{8}{k}(2x^2)^{8-k}\left(-\frac{3}{x}\right)^k = \binom{8}{k} 2^{8-k}(-3)^k x^{16-2k-k} = \binom{8}{k} 2^{8-k}(-3)^k x^{16-3k}.$$

Set $16 - 3k = 7 \Rightarrow k = 3$. The coefficient is

$$\binom{8}{3} 2^5 (-3)^3 = 56 \cdot 32 \cdot (-27) \boxed{-48\,384}.$$

Method. Always write the general term, collect the exponent of x as a linear function of k, and solve. If the equation has no non-negative integer solution $0 \leq k \leq n$, the coefficient is zero.

4.1 Identities for free: substitute, then differentiate

Because the theorem is an identity in x and y, every substitution gives a new identity, and so does every derivative.

Harvesting identities

Operation	Result	Value
$x = y = 1$	$\sum_k \binom{n}{k}$	2^n
$x = 1,\ y = -1$	$\sum_k (-1)^k \binom{n}{k}$	$0 \ (n \geq 1)$
add the two	$\sum_{k \text{ even}} \binom{n}{k}$	$2^{n-1} \ (n \geq 1)$
$x = 1,\ y = 2$	$\sum_k \binom{n}{k} 2^k$	3^n
$\frac{d}{dy}$ at $x = y = 1$	$\sum_k k\binom{n}{k}$	$n\,2^{n-1}$
$\frac{d^2}{dy^2}$ at $x = y = 1$	$\sum_k k(k-1)\binom{n}{k}$	$n(n-1)2^{n-2}$

The last two are exactly the computations that will give you the mean and variance of the binomial distribution in Section 12. It is worth noticing now that they are the same algebra.

4.2 The shape of a row

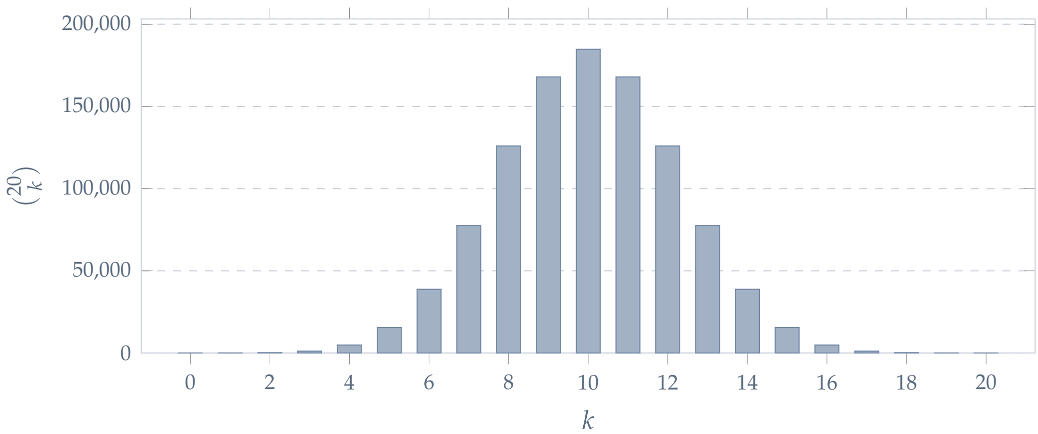


Figure 6: *Row $n = 20$ of Pascal's triangle. The row is symmetric, peaks at $k = \lfloor n/2 \rfloor$, and is already visibly bell-shaped. This is not a coincidence: after dividing by 2^n it is the $\mathrm{Bin}(20, \frac{1}{2})$ distribution, and the de Moivre–Laplace theorem says it converges to a normal curve.*

Where the row is largest

$\binom{n}{k}$ increases in k while $\dfrac{\binom{n}{k+1}}{\binom{n}{k}} = \dfrac{n-k}{k+1} > 1$, that is while $k < \dfrac{n-1}{2}$. So the maximum is at $k = \lfloor n/2 \rfloor$ (two equal maxima when n is odd). The central coefficient satisfies $\binom{2m}{m} \sim \dfrac{4^m}{\sqrt{\pi m}}$ by Stirling's formula, $n! \sim \sqrt{2\pi n}\,(n/e)^n$.

4.3 The multinomial theorem

Multinomial theorem

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{\substack{k_1+k_2+\cdots+k_r=n \\ k_i \geq 0}} \binom{n}{k_1,\, k_2,\, \ldots,\, k_r} x_1^{k_1} x_2^{k_2} \cdots x_r^{k_r},$$

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

The number of terms in the expansion is $\dbinom{n+r-1}{r-1}$ (see Section 5), and setting every $x_i = 1$ gives $\sum \dbinom{n}{k_1, \ldots, k_r} = r^n$.

A multinomial coefficient

The coefficient of $x^3y^2z^2$ in $(x+y+z)^7$ is

$$\binom{7}{3,2,2} = \frac{7!}{3!\,2!\,2!} = \frac{5040}{24}\ \boxed{210}.$$

Same number as the arrangements of the multiset $\{x, x, x, y, y, z, z\}$, and for the same reason: each term of the expansion *is* such an arrangement.

4.4 Beyond positive integer exponents

Newton's generalised binomial series

For any real α and $|t| < 1$,

$$(1+t)^{\alpha} = \sum_{k=0}^{\infty} \binom{\alpha}{k} t^k, \qquad \binom{\alpha}{k} := \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!} = \frac{\alpha^{\underline{k}}}{k!}.$$

The falling-factorial definition is the one that survives; the factorial-quotient form does not, because $\alpha!$ is meaningless for non-integers.

Two consequences you will meet again:

$$\frac{1}{1-t} = \sum_{k\geq 0} t^k, \qquad \frac{1}{(1-t)^r} = \sum_{k\geq 0} \binom{k+r-1}{r-1} t^k.$$

The second is the generating function for stars and bars, which is the subject of the next section, and for the negative binomial distribution later on.

Sign and index slips

- In $(x-y)^n$ the terms alternate: $\binom{n}{k}x^{n-k}(-y)^k$ carries $(-1)^k$. Fold the minus sign into y *before* expanding.
- T_{k+1}, not T_k, is the term with y^k. If a question asks for "the 6th term", that is $k=5$.
- $\binom{n}{k}$ counts the y's, so the exponent of x is $n-k$. Check by confirming the exponents sum to n.

5 Stars and bars, and the master table

Everything so far has been about *selecting* without repetition. The remaining case, unordered selection *with* repetition, is the one people find hardest, and it has a picture that makes it easy.

The problem

Count the solutions in non-negative integers of

$$x_1 + x_2 + \cdots + x_r = n.$$

Equivalently: the number of ways to distribute n identical objects into r distinguishable boxes; equivalently, the number of multisets of size n drawn from r types.

The picture

Draw the n objects as **stars**. Insert $r-1$ **bars** to cut the row into r compartments; the number of stars in compartment i is x_i. Every arrangement of n stars and $r-1$ bars gives exactly one solution, and every solution gives exactly one arrangement. So the two sets are in bijection, and we only have to count arrangements of a row of $n+r-1$ symbols, of which $r-1$ are bars.

Figure 7: *One arrangement of 9 stars and 4 bars, and the solution of $x_1 + \cdots + x_5 = 9$ that it encodes. Choosing the solution is the same as choosing which 4 of the 13 positions hold bars.*

Stars and bars

The number of non-negative integer solutions of $x_1 + \cdots + x_r = n$ is

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

The number of *positive* integer solutions ($x_i \geq 1$) is

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

obtained by substituting $y_i = x_i - 1 \geq 0$, which turns the equation into $y_1 + \cdots + y_r = n - r$.

Two quick applications

(a) How many ways to buy 12 doughnuts from 5 varieties, with unlimited supply of each? This is $x_1 + \cdots + x_5 = 12$, so $\binom{16}{4} = 1820$.

(b) How many terms does $(a+b+c+d)^{10}$ have after collecting? One term for each exponent vector with $k_1 + k_2 + k_3 + k_4 = 10$, so $\binom{13}{3} = 286$.

Lower bounds: shift. Upper bounds: inclusion–exclusion.

A constraint $x_i \geq c$ costs nothing: substitute $x_i' = x_i - c$ and reduce n by c. A constraint $x_i \leq c$ is genuinely harder, and is handled by subtracting the bad cases (Section 6). For example, the solutions of $x_1 + x_2 + x_3 = 15$ with every $x_i \leq 8$:

$$\binom{17}{2} - \binom{3}{1}\binom{8}{2} + 0 = 136 - 3 \cdot 28 + 0 \boxed{52},$$

where the middle term subtracts, for each i, the solutions with $x_i \geq 9$ (shift by 9, leaving $x_1 + x_2 + x_3 = 6$, so $\binom{8}{2} = 28$), and no two variables can both exceed 8 when the total is only 15.

5.1 The master table

Almost every elementary counting question is one of four questions. Choose k items from n types; ask whether order matters, and whether repetition is allowed.

	Order matters	Order does not matter
No repetition	$P(n,k) = \dfrac{n!}{(n-k)!}$ sequences of distinct items	$\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$ subsets
Repetition allowed	n^k strings / functions	$\dbinom{n+k-1}{k}$ multisets / stars and bars

Figure 8: *The fourfold way. Identify the row and the column before you write a formula.*

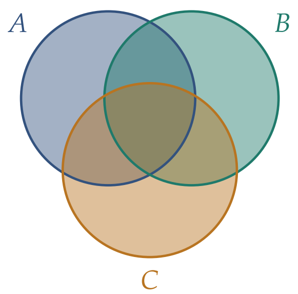
A fifth case that catches people out: partitions into groups

Splitting a set is not the same as selecting from it.

- **Ordered set partition into labelled groups of given sizes** $n_1, \dots, n_r$ summing to n: $\binom{n}{n_1, \dots, n_r} = \dfrac{n!}{n_1! \cdots n_r!}$.
- **Unlabelled groups**, all of the same size m, with $r = n/m$ groups: divide by $r!$ $\dfrac{n!}{(m!)^r\, r!}$.
- **Any number of unlabelled non-empty blocks**: the Bell number B_n; exactly j blocks gives the Stirling number of the second kind $S(n, j)$, with $S(n, j) = j\,S(n-1, j) + S(n-1, j-1)$. These have no closed form as simple as $\binom{n}{k}$, which is why exam questions specify sizes.

6 Inclusion–exclusion

The sum rule needs disjoint cases. When the cases overlap, you have to add the pieces, subtract the double-counted overlaps, add back the triple-counted triple overlaps, and so on. That alternating correction is inclusion–exclusion.



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

Figure 9: *Each point of the union must be counted exactly once. A point in two discs is added twice and subtracted once; a point in three discs is added three times, subtracted three times, and added back once.*

Inclusion–exclusion

For finite sets $A_1, \dots, A_n$,

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{\emptyset \neq J \subseteq [n]} (-1)^{|J|+1} \left|\bigcap_{i \in J} A_i\right| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \dots$$

Equivalently, for the number of elements of a universe S lying in *none* of the A_i:

$$\left|\bigcap_i \overline{A_i}\right| = \sum_{J \subseteq [n]} (-1)^{|J|} \left|\bigcap_{i \in J} A_i\right|, \qquad \text{with } \bigcap_{i \in \emptyset} A_i = S.$$

The identical statement for probabilities holds with $|\cdot|$ replaced by $\mathbb{P}(\cdot)$.

Why the signs alternate

Take an element lying in exactly $m \geq 1$ of the sets. On the right-hand side it is counted once for every non-empty subset of those m sets, with sign $(-1)^{|J|+1}$, contributing

$$\sum_{j=1}^{m} (-1)^{j+1} \binom{m}{j} = 1 - \sum_{j=0}^{m} (-1)^j \binom{m}{j} = 1 - 0 = 1.$$

The alternating row sum of Pascal's triangle is exactly what makes the bookkeeping come out right. Elements in none of the sets contribute 0. So the right-hand side counts each element of the union once.

Surjections

How many functions from a 6-set onto a 3-set (that is, hitting every target)? Let A_i be the set of functions missing target i. Then

$$\#\text{surjections} = 3^6 - \binom{3}{1}2^6 + \binom{3}{2}1^6 - \binom{3}{3}0^6 = 729 - 192 + 3 - 0 \boxed{540}.$$

In general the number of surjections $[n] \twoheadrightarrow [r]$ is $\sum_{j=0}^{r}(-1)^j\binom{r}{j}(r-j)^n = r!\,S(n,r)$.

6.1 Derangements

Derangement

A **derangement** of $[n]$ is a permutation with no fixed point: $\sigma(i) \neq i$ for every i. Write D_n (also $!n$) for the number of them.

Count and asymptotics

Let A_i be the permutations fixing i, so $|\bigcap_{i\in J} A_i| = (n-|J|)!$. Inclusion–exclusion gives

$$D_n \;=\; \sum_{j=0}^{n}(-1)^j\binom{n}{j}(n-j)! \;=\; n!\sum_{j=0}^{n}\frac{(-1)^j}{j!} \;=\; \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor \quad (n \geq 1),$$

with the recurrences $D_n = (n-1)(D_{n-1} + D_{n-2})$ and $D_n = nD_{n-1} + (-1)^n$. Hence

$$\frac{D_n}{n!} \longrightarrow e^{-1} \approx 0.3679.$$

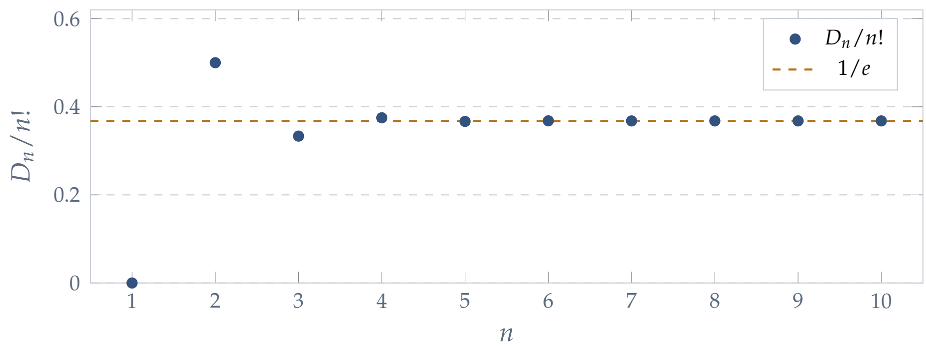


Figure 10: *The proportion of derangements settles onto $1/e$ almost immediately: the error is smaller than $1/(n+1)!$, so by $n = 7$ it is already correct to five decimals.*

The hat-check problem

Ten people check their hats and receive them back at random. The probability that *nobody* gets their own hat is $D_{10}/10! \approx 0.3679$. The probability that *exactly* k people get their own hat is

$$\mathbb{P}(\text{exactly } k) = \frac{\binom{n}{k}D_{n-k}}{n!} \approx \frac{e^{-1}}{k!},$$

which is Poisson(1) in the limit. The expected number of fixed points is exactly 1 for every $n \geq 1$, by linearity of expectation (Section 10).

When to reach for inclusion–exclusion

Look for the words *at least one*, *none*, *no two*, or a list of forbidden patterns. Define A_i = "constraint i is violated", count intersections (which are usually easy, because forcing several constraints to fail typically just shrinks the problem), and alternate. If the intersections are hard to count, the method is the wrong tool.

7 Bijections: counting by disguise

The best counting arguments do not compute anything. They show that the set you care about is secretly a set you already know how to count.

7.1 Lattice paths

Monotone lattice paths

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

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

Proof. A path is a word of $m+n$ letters, m of them R and n of them U. Choosing the path is choosing which n of the $m+n$ positions are U. □

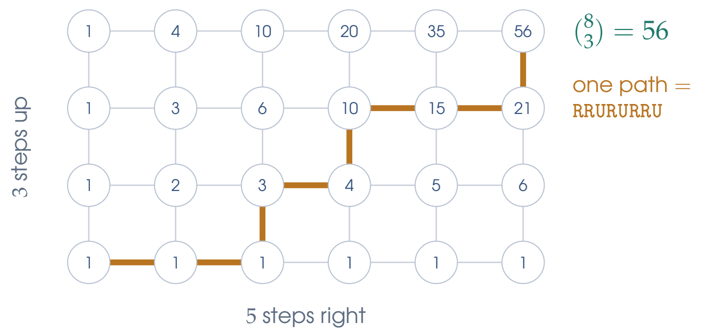


Figure 11: *Every vertex is labelled with the number of paths reaching it from the origin. The labels satisfy Pascal's rule, because a path arrives at a vertex either from the left or from below: Pascal's triangle* is the *lattice-path count, tilted.*

This one bijection explains several formulas at once. A path is a binary string; a binary string is a subset; a subset is a 0/1 vector; a 0/1 vector with prescribed weight is a combination. Any question about one of them can be moved to whichever representation makes it easiest.

7.2 The reflection principle and the ballot problem

Reflect the bad paths

To count paths that *avoid* a forbidden line, count all paths and subtract the bad ones. The bad ones are counted by reflecting the portion of the path before its first touch of the line: this maps bad paths bijectively onto *all* paths from the reflected starting point, and those are unconstrained, so easy to count.

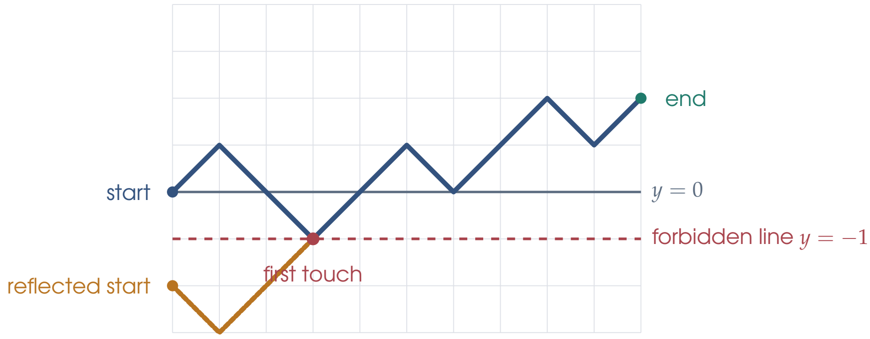


Figure 12: *Reflection. The path segment before the first touch of the forbidden line is mirrored in that line. Bad paths from the true start correspond one-to-one with* all *paths from the mirrored start.*

Bertrand's ballot problem

Candidate A receives a votes and candidate B receives b votes, with $a > b$. If the ballots are counted in a uniformly random order, the probability that A is strictly ahead throughout the count is

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

7.3 Catalan numbers

Catalan number

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

$$C_0, C_1, C_2, \ldots = 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, \ldots$$

with the recurrence $C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}$ and growth $C_n \sim \dfrac{4^n}{n^{3/2}\sqrt{\pi}}$.

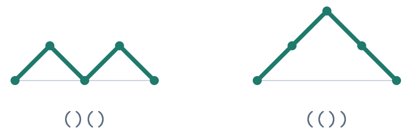
C_n counts, among many hundreds of other things:

- balanced strings of n pairs of parentheses;
- monotone lattice paths from $(0,0)$ to (n,n) that never rise above the diagonal (**Dyck paths**);
- binary trees with n internal nodes;
- triangulations of a convex $(n+2)$-gon;
- ways to stack coins in a row of n touching coins;
- sequences of n pushes and n pops of a stack that never pop an empty stack.

Why $\frac{1}{n+1}\binom{2n}{n}$

Of the $\binom{2n}{n}$ paths from $(0,0)$ to (n,n), the bad ones cross the diagonal. Reflect a bad path in the line $y = x+1$ after its first crossing: the endpoint (n,n) becomes $(n-1,n+1)$, and this is a bijection between bad paths and *all* paths to $(n-1,n+1)$, of which there are $\binom{2n}{n+1}$. Hence

$$C_n = \binom{2n}{n} - \binom{2n}{n+1} = \binom{2n}{n}\left(1 - \frac{n}{n+1}\right) = \frac{1}{n+1}\binom{2n}{n}.$$



$C_2 = 2$: the two Dyck paths of length 4.
Up-steps open a bracket, down-steps close one; staying above the axis means never closing a bracket that was never opened.

Figure 13: *Dyck paths never dip below the axis. Up-steps are open brackets, down-steps are close brackets.*

A checklist for building a bijection

1. Describe an element of your set as a *string of decisions*.
2. Check that different elements give different strings (injective).
3. Check that every legal string comes from an element (surjective).
4. Count the strings.

Steps 2 and 3 are the proof; skipping them is how overcounting sneaks back in.

PART II

Probability

8 From counting to probability

Sample space, event, probability measure

A **sample space** S is the set of possible outcomes of an experiment. An **event** is a subset $A \subseteq S$. A **probability measure** assigns to each event a number $\mathbb{P}(A)$ satisfying **Kolmogorov's axioms**:

A1. $\mathbb{P}(A) \geq 0$ for every event A;

A2. $\mathbb{P}(S) = 1$;

A3. if $A_1, A_2, \ldots$ are pairwise disjoint then $\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i)$.

Axiom A3 is the sum rule again. Everything below follows from these three lines alone.

Consequences

Complement	$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$, and $\mathbb{P}(\emptyset) = 0$
Monotonicity	$A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$, so $\mathbb{P}(A) \leq 1$
Two events	$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
Union bound	$\mathbb{P}\left(\bigcup_i A_i\right) \leq \sum_i \mathbb{P}(A_i)$
Inclusion–exclusion	$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{\emptyset \neq J \subseteq [n]} (-1)^{\lvert J\rvert+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right)$
Partition	if $B_1, \ldots, B_m$ partition S then $\mathbb{P}(A) = \sum_j \mathbb{P}(A \cap B_j)$

The equally likely (classical) model

If S is finite and every outcome is equally likely, then for every event A

$$\mathbb{P}(A) = \frac{|A|}{|S|}.$$

The model is a choice, not a fact

"Equally likely" is an assumption you impose on the sample space, and you can make it true or false by choosing the sample space badly. If you model two coin flips by $S = \{0 \text{ heads}, 1 \text{ head}, 2 \text{ heads}\}$ and assume equal likelihood, you get $\mathbb{P}(2 \text{ heads}) = 1/3$, which is wrong. Model with $S = \{HH, HT, TH, TT\}$ instead. **Rule.** Build the sample space out of the *physical symmetries* of the experiment (individual coins, individual cards, individual people), never out of summary statistics.

8.1 A worked catalogue: five-card poker hands

There are $\binom{52}{5} = 2\,598\,960$ equally likely hands. Each row below is a pure counting exercise, and together they are a good test of whether the fourfold way has sunk in.

Hand	Count, as a product	Number	Probability	Odds
Straight flush	$10 \cdot 4$	40	0.000015	1 in 64 974
Four of a kind	$13 \cdot 48$	624	0.000240	1 in 4 165
Full house	$13\binom{4}{3} \cdot 12\binom{4}{2}$	3 744	0.001441	1 in 694
Flush (not straight)	$4\binom{13}{5} - 40$	5 108	0.001965	1 in 509
Straight (not flush)	$10 \cdot 4^5 - 40$	10 200	0.003925	1 in 255
Three of a kind	$13\binom{4}{3}\binom{12}{2}4^2$	54 912	0.021128	1 in 47
Two pair	$\binom{13}{2}\binom{4}{2}^2 \cdot 44$	123 552	0.047539	1 in 21
One pair	$13\binom{4}{2}\binom{12}{3}4^3$	1 098 240	0.422569	1 in 2
High card (nothing)	remainder	1 302 540	0.501177	1 in 2
Total		2 598 960	1.000000	

Table 1: *The counts sum exactly to $\binom{52}{5}$, which is the check that the case split was disjoint and exhaustive.*

How to read the products

Full house. Choose the rank that appears three times (13), choose which three of its four suits ($\binom{4}{3}$), choose the rank appearing twice (12, not 13: it must differ), choose its two suits ($\binom{4}{2}$).
Two pair. Choose the two pair-ranks together ($\binom{13}{2}$, unordered, because naming them in the other order gives the same hand), suits for each ($\binom{4}{2}^2$), then the fifth card from the 44 cards of the remaining 11 ranks.
The -40. Straights and flushes overlap in the 40 straight flushes, so each category must exclude them to keep the cases disjoint.

Complement, again

What is the probability that a five-card hand contains at least one ace?

$$1 - \frac{\binom{48}{5}}{\binom{52}{5}} = 1 - \frac{1\,712\,304}{2\,598\,960} \approx \boxed{0.3412}.$$

Note that $\binom{4}{1}\binom{48}{4}/\binom{52}{5} \approx 0.2995$ is the probability of *exactly* one ace, which is a different question. And the tempting shortcut "choose an ace, then any four other cards", $\binom{4}{1}\binom{51}{4}/\binom{52}{5} \approx 0.3846$, is simply wrong: a hand with two aces is produced twice, once for each ace that could have been the designated one.

9 Conditional probability and Bayes

Conditional probability

For events A, B with $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Conditioning on B is a change of universe: B becomes the new sample space, and all probabilities are renormalised by $\mathbb{P}(B)$ so they sum to 1 again. In the equally-likely model this is just $\mathbb{P}(A \mid B) = |A \cap B|/|B|$: counting inside B.

The rules that follow

$$
\begin{aligned}
\text{Multiplication rule:}\quad & \mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A \mid B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A) \\
\text{Chain rule:}\quad & \mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\cdots\mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \\
\text{Total probability:}\quad & \mathbb{P}(A) = \sum_j \mathbb{P}(B_j)\,\mathbb{P}(A \mid B_j) \quad \text{for any partition } \{B_j\} \text{ of } S \\
\text{Bayes:}\quad & \mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(B_i)\,\mathbb{P}(A \mid B_i)}{\sum_j \mathbb{P}(B_j)\,\mathbb{P}(A \mid B_j)}
\end{aligned}
$$

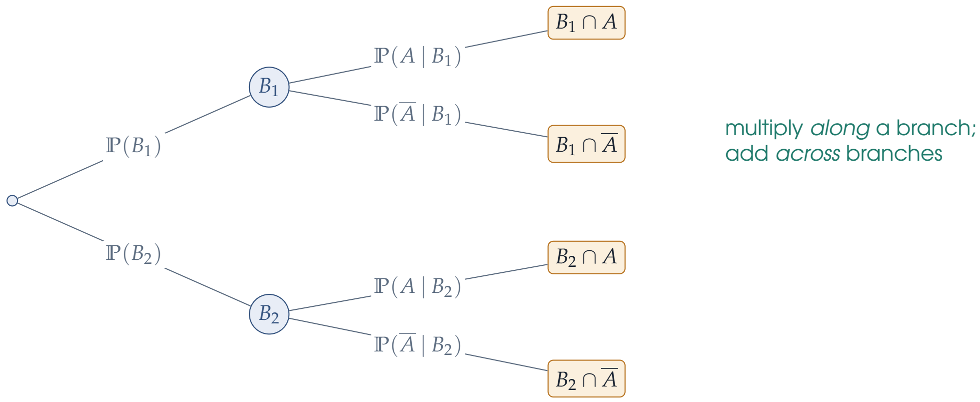


Figure 14: *The probability tree. $\mathbb{P}(A)$ is the sum of the two amber leaves containing A; $\mathbb{P}(B_1 \mid A)$ is the top leaf divided by that sum. Bayes' theorem is nothing more than this picture read backwards.*

Independence

A and B are **independent** iff $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, equivalently $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ when $\mathbb{P}(B) > 0$. Events $A_1, \ldots, A_n$ are **mutually independent** iff for *every* subset J, $\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$.

Independence is not pairwise independence, and is not disjointness

Pairwise is weaker. Toss two fair coins; let A = "first is heads", B = "second is heads", C = "the two agree". Any two of these are independent, but $\mathbb{P}(A \cap B \cap C) = 1/4 \neq 1/8$. Mutual independence requires *all* $2^n - n - 1$ conditions.

Disjoint is nearly the opposite. If A and B are disjoint and both have positive probability, they are strongly *dependent*: knowing A occurred tells you B did not, so $\mathbb{P}(B \mid A) = 0 \neq \mathbb{P}(B)$.

Bayes with urns

Urn 1 holds 3 red and 7 blue balls; urn 2 holds 6 red and 4 blue. An urn is chosen by a fair coin, then one ball is drawn from it, and it is red. What is the probability the ball came from urn 1?

$$\mathbb{P}(U_1 \mid R) = \frac{\frac{1}{2} \cdot \frac{3}{10}}{\frac{1}{2} \cdot \frac{3}{10} + \frac{1}{2} \cdot \frac{6}{10}} = \frac{3}{3+6} \boxed{\tfrac{1}{3}}.$$

Notice the equal priors cancelled entirely: with a fair coin, the posterior is just the red counts in proportion, 3 : 6. Priors only matter when they differ.

9.1 The odds form, which is the one to use

Bayes in odds form

$$\underbrace{\frac{\mathbb{P}(H_1 \mid E)}{\mathbb{P}(H_2 \mid E)}}_{\text{posterior odds}} = \underbrace{\frac{\mathbb{P}(H_1)}{\mathbb{P}(H_2)}}_{\text{prior odds}} \times \underbrace{\frac{\mathbb{P}(E \mid H_1)}{\mathbb{P}(E \mid H_2)}}_{\text{likelihood ratio}} .$$

No normalising constant, no denominator to compute, and evidence multiplies: two independent pieces of evidence multiply their likelihood ratios.

Rare events and the base rate

A property is held by 1 in 1000 objects. A test detects it with probability 0.99, and gives a false positive with probability 0.05. An object tests positive. The posterior odds are

$$\frac{1}{999} \times \frac{0.99}{0.05} = \frac{0.99}{49.95} \approx \frac{1}{50.45},$$

so $\mathbb{P}(\text{property} \mid +) \approx 1/51.45 \approx 0.0194$.
Fewer than 2%, despite a 99%-accurate test. The reason is visible in the odds form: the likelihood ratio is only about 20, while the prior odds are 1:999. Evidence has to overcome the prior, and a ratio of 20 does not overcome a ratio of 999.

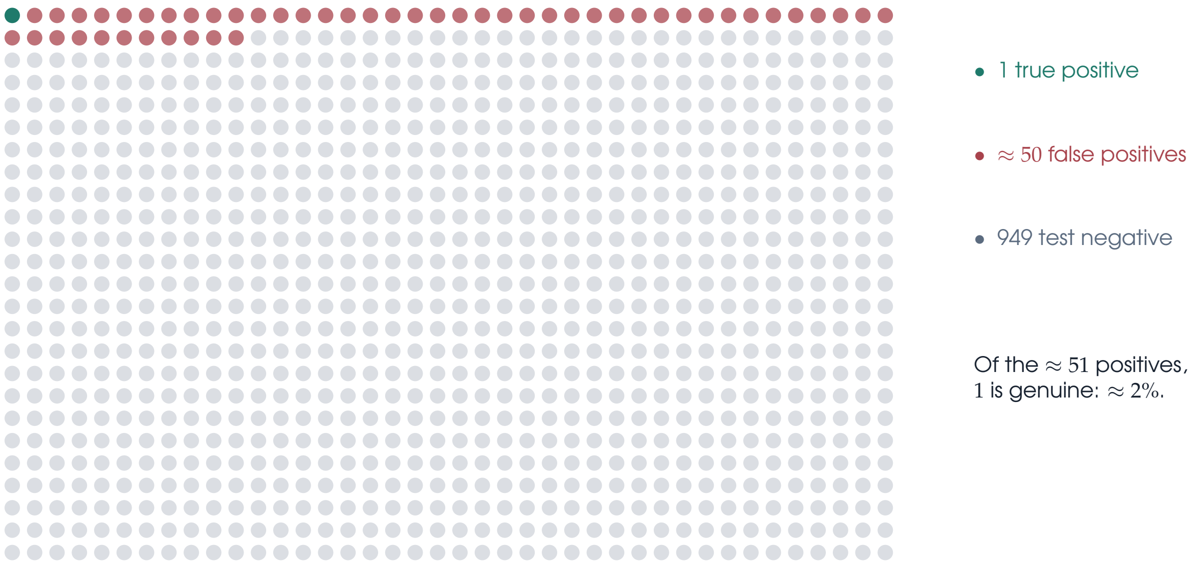


Figure 15: *Natural frequencies. Conditional probability becomes obvious the moment you count objects instead of manipulating fractions: restrict to the coloured dots and ask what fraction is green.*

The two most common conditional-probability errors

- **Transposing the conditional.** $\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$ in general. They are related by Bayes and differ by the ratio $\mathbb{P}(A)/\mathbb{P}(B)$.
- **Forgetting to condition on the information-generating process.** What you learn is not "a fact" but "a fact was reported to me by some mechanism". The two-child problem turns on exactly this: "at least one is a boy" gives $\mathbb{P}(\text{both boys}) = 1/3$, while "the elder is a boy" gives $1/2$, and "I met one of them at random and he is a boy" gives $1/2$. Same words, different sample spaces.

10 Expectation, and the one trick that solves everything

Random variable, PMF, expectation

A **discrete random variable** is a function $X : S \to \mathbb{R}$ taking countably many values. Its **probability mass function** is $p_X(x) = \mathbb{P}(X = x)$, and its **expectation** (mean, expected value) is

$$\mathbb{E}[X] = \sum_x x\,\mathbb{P}(X = x),$$

provided $\sum_x |x|\mathbb{P}(X = x) < \infty$. Interpretation: the long-run average value of X over many independent repetitions, and the centre of mass of the PMF.

Linearity of expectation

For *any* random variables $X_1, \ldots, X_n$ on the same sample space and any constants a_i, b,

$$\mathbb{E}\left[\sum_{i=1}^n a_i X_i + b\right] = \sum_{i=1}^n a_i\,\mathbb{E}[X_i] + b.$$

No independence is required. The X_i may be wildly dependent; linearity still holds, because it is just a rearrangement of a finite (or absolutely convergent) sum over the sample space.

The indicator method

To find $\mathbb{E}[X]$ where X counts how many things happen:

1. Write $X = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n}$, one indicator per thing that could happen.
2. Note $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$, since $\mathbf{1}_A$ is 1 with probability $\mathbb{P}(A)$ and 0 otherwise.
3. Conclude $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{P}(A_i)$.

This converts a hard distributional question (what is the distribution of X?) into n easy questions (what is the chance of this one event?). It is by far the most useful technique in elementary probability.

Four problems, one method

(a) Fixed points of a random permutation. $X = \sum_{i=1}^n \mathbf{1}_{\{\sigma(i)=i\}}$, and $\mathbb{P}(\sigma(i) = i) = 1/n$, so $\mathbb{E}[X] = n \cdot \frac{1}{n} = 1$, for every n. The distribution of X is complicated (Section 6); its mean is not.

(b) Matching cards. Two shuffled decks are dealt in parallel; the expected number of positions where the cards agree is $52 \cdot \frac{1}{52} = 1$.

(c) Ascents. In a uniformly random permutation of $[n]$, the expected number of i with $\sigma(i) < \sigma(i+1)$ is $(n-1) \cdot \frac{1}{2} = \frac{n-1}{2}$, since each adjacent pair is increasing with probability $1/2$.

(d) Records. Reading a random permutation left to right, position i is a *record* (larger than everything before it) with probability $1/i$, because the maximum of the first i entries is equally likely to be in any of the i positions. So the expected number of records is the harmonic number

$$H_n = \sum_{i=1}^n \frac{1}{i} \approx \ln n + \gamma, \qquad \gamma \approx 0.5772.$$

Other tools for computing $\mathbb{E}$

LOTUS	$\mathbb{E}[g(X)] = \sum_x g(x)\mathbb{P}(X = x)$: you do *not* need the distribution of $g(X)$.
Tail sum	for X taking values in $\{0, 1, 2, \dots\}$, $\mathbb{E}[X] = \sum_{k \geq 1} \mathbb{P}(X \geq k)$.
Tower property	$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X \mid Y]\big]$: condition on whatever makes the problem easy, then average.
First-step analysis	set up an equation for $\mathbb{E}[X]$ by conditioning on the first step, then solve algebraically.
Symmetry	if $X_1, \dots, X_n$ are exchangeable and sum to T, then $\mathbb{E}[X_i] = \mathbb{E}[T]/n$ without any further work.

First-step analysis: waiting for a run

How many fair coin tosses on average until the first head? Let $\mu = \mathbb{E}[N]$. Condition on the first toss: with probability $\frac{1}{2}$ we are done in one toss; with probability $\frac{1}{2}$ we have used one toss and are back where we started. So

$$\mu = \tfrac{1}{2}(1) + \tfrac{1}{2}(1 + \mu) \implies \mu = 2.$$

For the first occurrence of `HH`, the same method with two states gives 6; for `HT` it gives 4. *Different patterns of the same length have different waiting times*, which surprises almost everyone the first time. The reason is that `HH` can overlap itself and `HT` cannot.

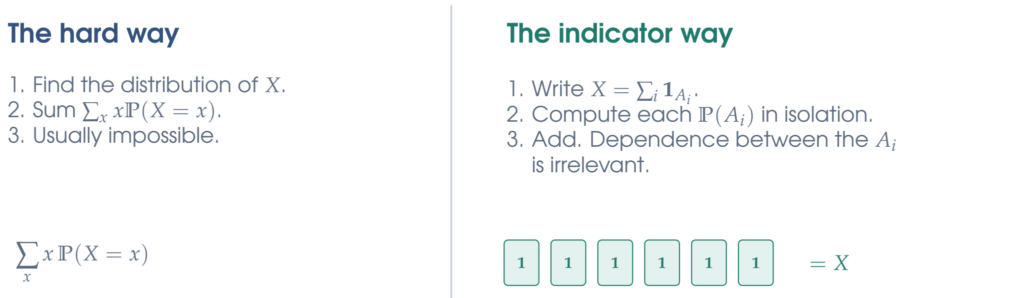


Figure 16: *Linearity lets you decompose before you compute, which is the opposite of the usual order of operations.*

Linearity is for sums, not products

$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ requires independence (or at least zero covariance). $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ requires nothing. Likewise $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$ in general: for convex g, Jensen's inequality gives $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$. In particular $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$, with the gap being exactly the variance.

11 Variance, covariance, and how far from the mean

Variance and standard deviation

$$\operatorname{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}X)^2\big] = \mathbb{E}[X^2] - (\mathbb{E}X)^2, \qquad \operatorname{SD}(X) = \sqrt{\operatorname{Var}(X)}.$$

The second form is the one you compute with. Variance is in squared units; the standard deviation is in the same units as X, which is why it is the one that gets reported.

Rules

$$\operatorname{Var}(aX+b) = a^2 \operatorname{Var}(X)$$ shifts do not change spread; scaling squares

$$\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$ with $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$

$$\operatorname{Var}\Big(\sum_i X_i\Big) = \sum_i \operatorname{Var}(X_i) + 2\sum_{i<j} \operatorname{Cov}(X_i, X_j)$$

$$\operatorname{Var}\Big(\sum_i X_i\Big) = \sum_i \operatorname{Var}(X_i)$$ if the X_i are pairwise uncorrelated

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\operatorname{SD}(X)\ \operatorname{SD}(Y)} \in [-1,1]$$

Variance is not linear

The identity $\mathbb{E}[\sum X_i] = \sum \mathbb{E}[X_i]$ is unconditional. The analogous statement for variance is *false* without an assumption: dependent variables have a covariance term. Independence implies zero covariance; the converse is false (X uniform on $\{-1,0,1\}$ and $Y = X^2$ have $\operatorname{Cov} = 0$ but are utterly dependent).

Variance of a binomial, by indicators

Let $X \sim \operatorname{Bin}(n,p)$, written as $X = \sum_{i=1}^n \mathbf{1}_{A_i}$ with independent A_i of probability p. For a single indicator,

$$\mathbb{E}[\mathbf{1}_A] = p, \qquad \mathbb{E}[\mathbf{1}_A^2] = \mathbb{E}[\mathbf{1}_A] = p, \qquad \operatorname{Var}(\mathbf{1}_A) = p - p^2 = p(1-p).$$

Independence kills the covariances, so $\operatorname{Var}(X) = np(1-p)$.
Now compare the **hypergeometric** case: drawing n balls *without* replacement from N balls of which K are red. The indicators still each have mean $p = K/N$, so $\mathbb{E}[X] = np$ exactly as before, by linearity. But now they are negatively correlated (drawing a red one makes the next draw less likely to be red), and the covariance term reduces the variance to

$$\operatorname{Var}(X) = np(1-p) \cdot \underbrace{\frac{N-n}{N-1}}_{\text{finite population correction}}.$$

Same mean, smaller spread: sampling without replacement is more informative.

11.1 How far from the mean can you get?

Markov and Chebyshev

For $X \geq 0$ and $a > 0$: $\quad \mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}[X]}{a} \quad$ (Markov).

For any X with finite variance and $k > 0$: $\quad \mathbb{P}\big(|X - \mathbb{E}X| \geq k\ \operatorname{SD}(X)\big) \leq \dfrac{1}{k^2} \quad$ (Chebyshev).

Chebyshev is Markov applied to $(X - \mathbb{E}X)^2$. Both are weak, and that weakness is the point: they hold for *every* distribution with the stated moments, so they are the right tools when you know nothing else. When you do know more (independent bounded summands, say), sharper exponential bounds are available.

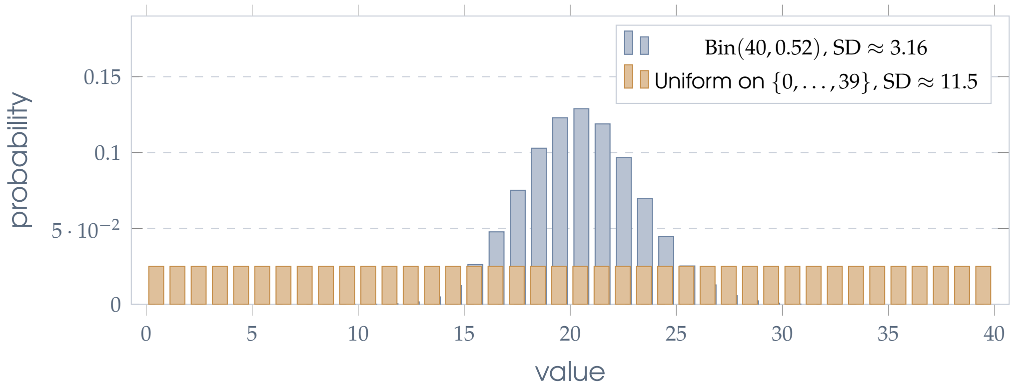


Figure 17: *Two distributions with almost the same mean (20.8 and 19.5) and completely different variance. The mean tells you where; the variance tells you how much you should trust it.*

Moments in one place

For a discrete X:

$$\mathbb{E}[X] = \sum_x x\,p(x), \qquad \mathbb{E}[X^2] = \sum_x x^2 p(x), \qquad \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

The **probability generating function** $G_X(t) = \mathbb{E}[t^X] = \sum_k \mathbb{P}(X = k)t^k$ packages all of this: $G_X(1) = 1$, $G_X'(1) = \mathbb{E}[X]$, $G_X''(1) = \mathbb{E}[X(X-1)]$, and hence $\mathrm{Var}(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2$. For independent X, Y, $G_{X+Y} = G_X G_Y$, which is why sums of independent binomials with the same p are binomial: $((1-p)+pt)^m((1-p)+pt)^n = ((1-p)+pt)^{m+n}$.

12 The named discrete distributions

Six distributions cover almost everything. They are not six unrelated facts: they are six different questions asked about the same repeated experiment.

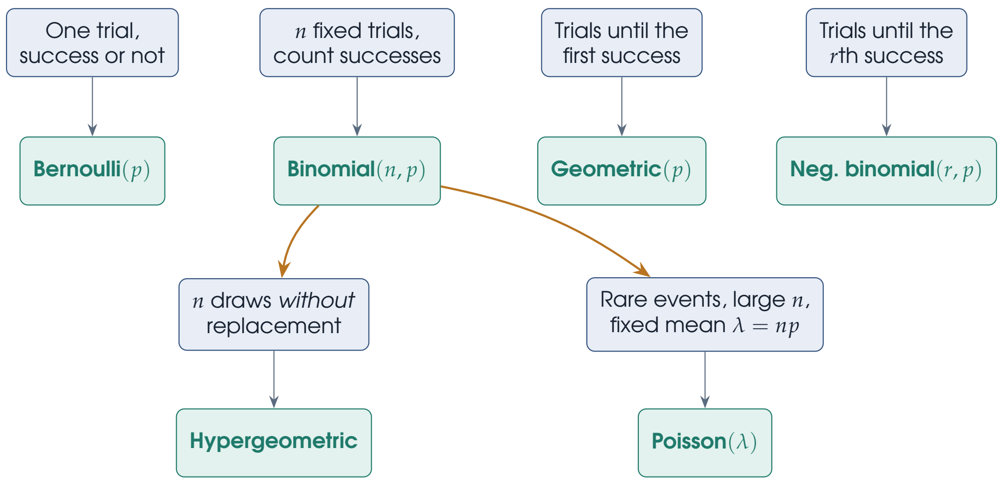


Figure 18: *Which distribution is which question.*

12.1 The reference table

Distribution	Support	PMF $\mathbb{P}(X=k)$	Mean	Variance
Bernoulli(p)	$\{0,1\}$	$p^k(1-p)^{1-k}$	p	$p(1-p)$
Binomial(n,p)	$0,1,\ldots,n$	$\binom{n}{k}\, p^k(1-p)^{n-k}$	np	$np(1-p)$
Geometric(p)	$1,2,3,\ldots$	$(1-p)^{k-1}p$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
Neg. binomial(r,p)	$r, r+1,\ldots$	$\binom{k-1}{r-1}p^r(1-p)^{k-r}$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$
Hypergeom.(N,K,n)	$\max(0,n-N+K)\ldots\min(n,K)$	$\dfrac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$n\pi$	$n\pi(1-\pi)\dfrac{N-n}{N-1}$
Poisson(λ)	$0,1,2,\ldots$	$e^{-\lambda}\dfrac{\lambda^k}{k!}$	λ	λ
Uniform on $[n]$	$1,2,\ldots,n$	$\dfrac{1}{n}$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$

For the hypergeometric, $\pi = K/N$ is the proportion of successes in the population.

Table 2: *Every PMF here sums to 1 by an identity from Part I: binomial by the binomial theorem, negative binomial by Newton's series, hypergeometric by Vandermonde, Poisson by the exponential series.*

Read the table as counting

$\binom{n}{k}p^k(1-p)^{n-k}$: choose *which* k of the n trials succeed ($\binom{n}{k}$), then the probability of that specific pattern ($p^k(1-p)^{n-k}$, by independence). The binomial coefficient is doing exactly the job it did in Part I.
$\binom{k-1}{r-1}p^r(1-p)^{k-r}$: the kth trial must be the rth success, so choose which $r-1$ of the first $k-1$ trials were the earlier successes.
$\binom{K}{k}\binom{N-K}{n-k}/\binom{N}{n}$: this is the equally-likely model, $|A|/|S|$, with $|S| = \binom{N}{n}$ subsets of size n.

12.2 How the shape depends on the parameters

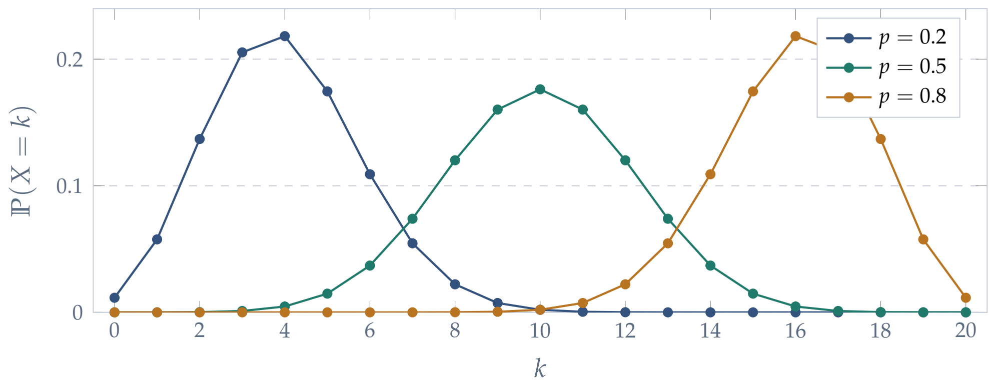


Figure 19: $\mathrm{Bin}(20,p)$. *The mean sits at np; the distribution is symmetric only at $p=\frac{1}{2}$, right-skewed for small p, left-skewed for large p. Variance $np(1-p)$ is largest at $p=\frac{1}{2}$ and vanishes at the extremes.*

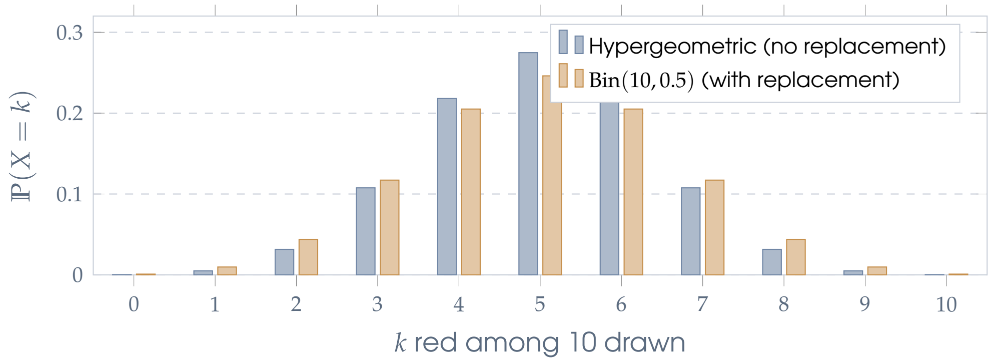


Figure 20: *Hypergeometric versus binomial with the same mean. Sampling without replacement is more concentrated: the finite-population correction* $\frac{N-n}{N-1}$ *is less than 1 whenever* $n > 1$. *As* $N \to \infty$ *with* K/N *fixed, the two coincide.*

12.3 The Poisson limit

Poisson limit theorem

If $n \to \infty$ and $p \to 0$ with $np \to \lambda$ fixed, then for each k

$$\binom{n}{k} p^k (1-p)^{n-k} \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

Sketch. With $p = \lambda/n$,

$$\binom{n}{k} p^k (1-p)^{n-k} = \underbrace{\frac{n^{\underline{k}}}{n^k}}_{\to 1} \cdot \frac{\lambda^k}{k!} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1}.$$

The falling factorial from Section 2 is what makes the first factor tend to 1.

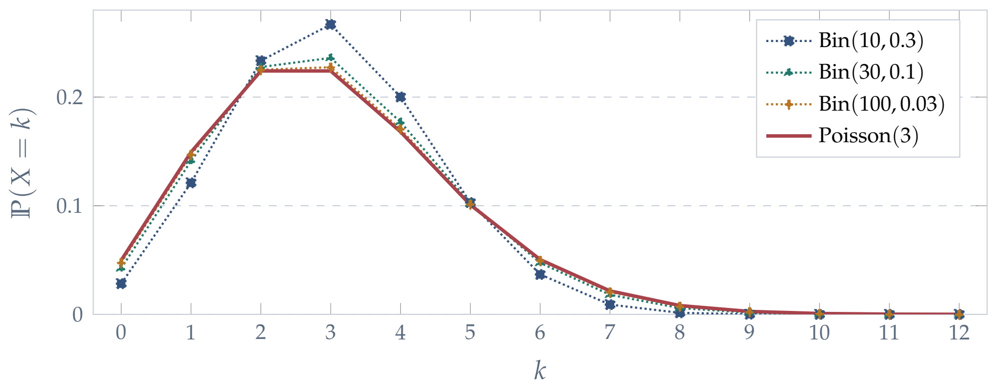


Figure 21: *Three binomials with mean 3 converging on* Poisson(3). *By* $n = 100$, $p = 0.03$ *the two are indistinguishable at this scale. Rule of thumb: the approximation is good when* $n \geq 20$ *and* $p \leq 0.05$, *or* $n \geq 100$ *and* $np \leq 10$.

Poisson facts worth carrying

- Mean = variance = λ. If your data has variance far from its mean, it is not Poisson.
- Sums: $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$ independent $\Rightarrow X + Y \sim \text{Poisson}(\lambda + \mu)$.
- Thinning: keep each Poisson(λ) event independently with probability q; the survivors are Poisson(λq).

- $\mathbb{P}(X = 0) = e^{-\lambda}$, which is the single most-used value.

12.4 The normal approximation

de Moivre–Laplace

For $X \sim \text{Bin}(n, p)$ with $np(1-p)$ large,

$$\mathbb{P}(a \le X \le b) \;\approx\; \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right),$$

where Φ is the standard normal CDF. The $\pm\frac{1}{2}$ is the **continuity correction**: you are approximating a bar of width 1 by an area under a curve. Rule of thumb: usable when $np \ge 10$ and $n(1-p) \ge 10$.

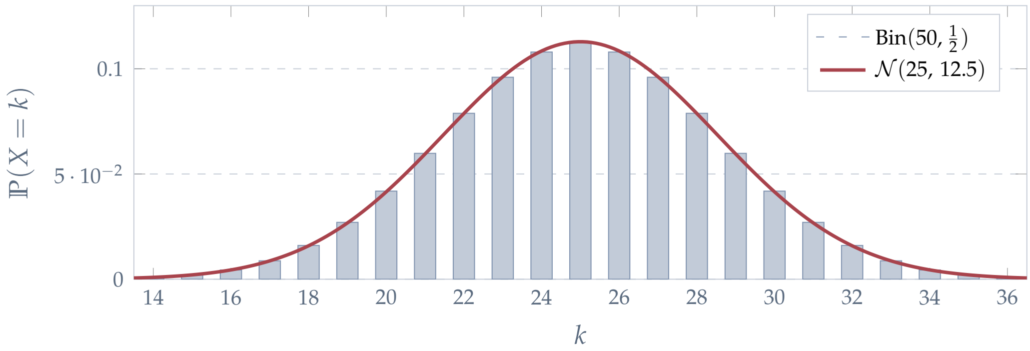


Figure 22: *Row 50 of Pascal's triangle, normalised, against the normal density with the same mean and variance. The bell shape you noticed in Section 5 was the central limit theorem all along.*

Which one is it?

1. *Twelve fair dice are rolled; how many show a six?* Fixed number of independent trials, count successes: $\text{Bin}(12, \frac{1}{6})$, mean 2, variance $\frac{5}{3}$.
2. *How many rolls until the first six?* $\text{Geometric}(\frac{1}{6})$, mean 6, variance 30.
3. *Five cards are dealt from a deck; how many are hearts?* No replacement: $\text{Hypergeometric}(52, 13, 5)$, mean $5 \cdot \frac{13}{52} = 1.25$.
4. *A page has 2000 characters, each mistyped with probability 0.001; how many typos?* Rare, many trials: $\text{Poisson}(2)$, so $\mathbb{P}(\text{no typos}) = e^{-2} \approx 0.135$.

13 Classic problems

These are the problems that everyone meets, gets wrong, and then never forgets. Each one isolates a single idea.

13.1 The birthday problem

How many people before a shared birthday is likely?

With n people and 365 equally likely birthdays, use the complement:

$$\mathbb{P}(\text{all distinct}) = \frac{P(365,n)}{365^n} = \prod_{i=0}^{n-1}\left(1 - \frac{i}{365}\right), \qquad \mathbb{P}(\text{some match}) = 1 - \frac{P(365,n)}{365^n}.$$

At $n = 23$ this exceeds $\frac{1}{2}$; at $n = 57$ it exceeds 0.99.

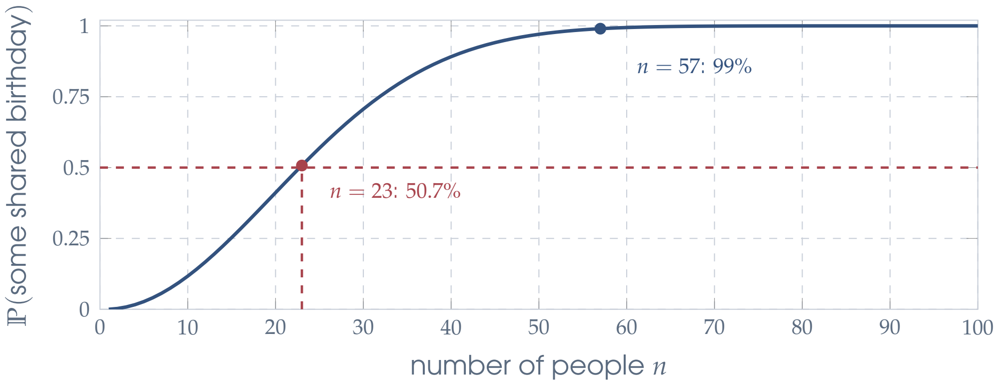


Figure 23: *The birthday curve. It rises far faster than intuition suggests.*

Why 23 and not 183

Intuition counts *people*; the problem counts *pairs*. With n people there are $\binom{n}{2}$ pairs, each matching with probability $1/365$, so the expected number of matching pairs is $\binom{n}{2}/365$. This reaches 1 at $\binom{n}{2} \approx 365$, that is $n \approx 27$. Since matches are roughly Poisson, $\mathbb{P}(\text{at least one}) \approx 1 - \exp\left(-\binom{n}{2}/365\right)$, which crosses $\frac{1}{2}$ at $\binom{n}{2} \approx 365 \ln 2 \approx 253$, giving $n \approx 23$. *Quadratic growth in n is the whole phenomenon.* The same computation is why hash collisions appear after about $\sqrt{N}$ insertions into N buckets.

13.2 The coupon collector

How long to collect all n coupons?

Once you hold i distinct coupons, each new one is new with probability $(n-i)/n$, so the wait for the next new coupon is Geometric$\left(\frac{n-i}{n}\right)$ with mean $\frac{n}{n-i}$. The total time T is the sum of these waits, and linearity gives

$$\mathbb{E}[T] = \sum_{i=0}^{n-1} \frac{n}{n-i} = n \sum_{j=1}^{n} \frac{1}{j} = nH_n \approx n \ln n + \gamma n + \tfrac{1}{2}.$$

The waits are independent, so the variances add too: $\mathrm{Var}(T) = \sum_{i=0}^{n-1} \frac{n\,i}{(n-i)^2} < \frac{\pi^2 n^2}{6}$.
For $n = 50$: $\mathbb{E}[T] = 50\,H_{50} \approx 224.96$, so about 225 purchases for 50 coupons. Doubling n slightly more than doubles the wait; the $\ln n$ factor is gentle.

13.3 Monty Hall

Switch

Three doors, one prize. You pick a door; the host, who knows where the prize is and who *always* opens a different door revealing no prize, does so. Should you switch?
Condition on where the prize is. Your initial pick is correct with probability $\frac{1}{3}$. Switching wins exactly when your initial pick was *wrong*, which has probability $\frac{2}{3}$. So switching wins with probability $\frac{2}{3}$, and staying with probability $\frac{1}{3}$.

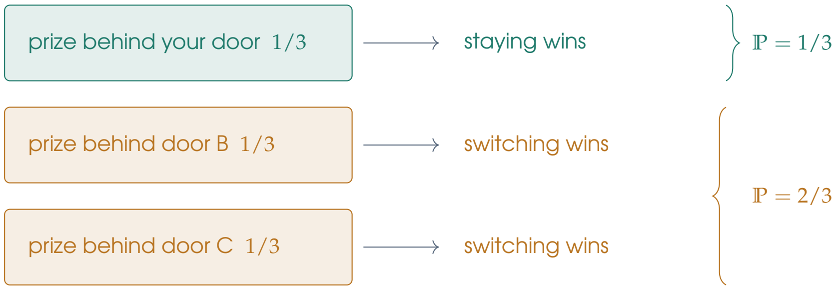


Figure 24: *The host's action gives you no information about your own door, because he would have opened* some *empty door whatever you chose. All the information lands on the remaining door.*

The protocol is the problem

Change the host's rule and the answer changes. If the host opens a door *at random* and it happens to be empty, switching wins with probability $\frac{1}{2}$. The words spoken are almost identical; the sample spaces are not. Whenever a problem involves someone revealing information, write down explicitly what they would have done in every case.

13.4 Two more worth knowing

Quick classics

The matching / hat-check problem. Expected matches $= 1$; probability of no match $\to e^{-1}$; number of matches is asymptotically $\text{Poisson}(1)$ (Section 6).

The secretary problem. To maximise the chance of selecting the best of n candidates interviewed in random order, reject the first $\approx n/e$ and then take the first candidate better than all seen so far. Success probability $\to 1/e \approx 0.368$, independent of n.

Banach's matchbox. Two boxes of n matches, one drawn at random each time; when one box is first found empty, the expected number remaining in the other is $(2n+1)\frac{\binom{2n}{n}}{4^n} - 1 \approx 2\sqrt{n/\pi}$. The central binomial coefficient from Section 5 appears again, and again with $\sqrt{\pi}$ in tow.

PART III

Reference & Practice

14 The one-page decision guide

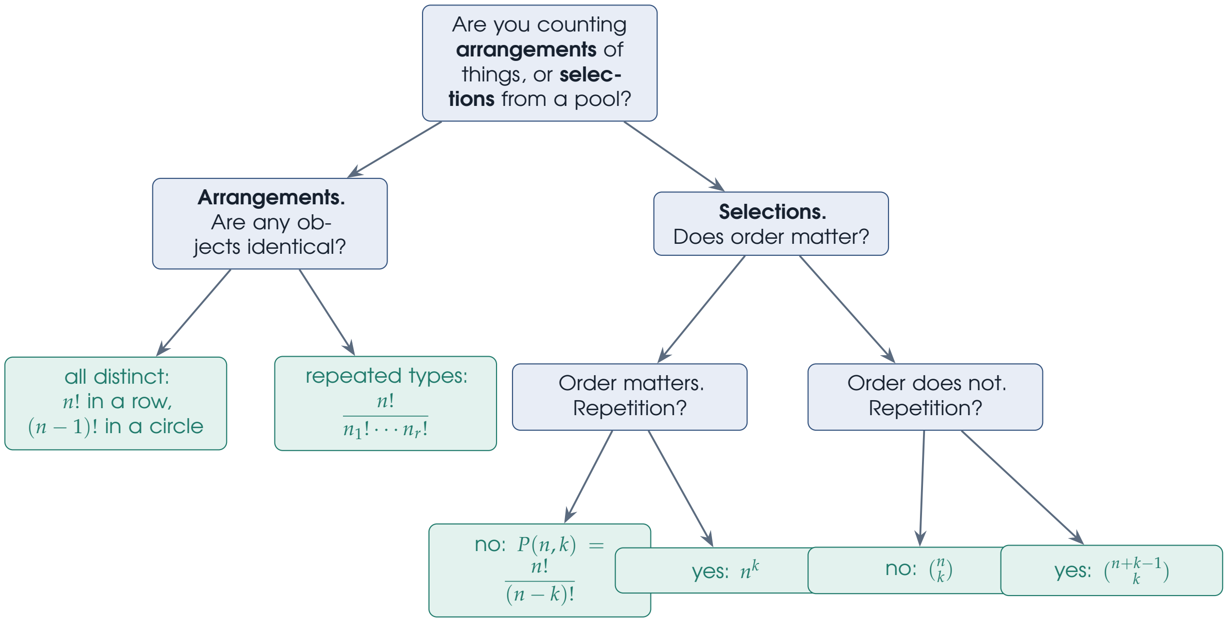


Figure 25: *Four questions, four formulas. If none of the leaves fits, the object is probably a* partition *rather than a selection: see the panel in Section 6.*

Signal words

and, "then", stages	multiply (product rule)
or, "either", disjoint cases	add (sum rule)
at least one, *none*	take the complement, or inclusion–exclusion
at most, upper bounds on variables	inclusion–exclusion
arrangement, *sequence*, *ranking*, *password*	order matters
committee, *subset*, *hand*, *team*	order does not matter
identical, *indistinguishable* objects into boxes	stars and bars
how many on average, *expected*	indicators plus linearity
given that, *knowing that*	conditional probability, Bayes

15 Formula spread

Small values worth recognising

$n!$	1, 1, 2, 6, 24, 120, 720, 5040, 40320, 362880, 3628800
2^n	1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096
$\binom{2n}{n}$	1, 2, 6, 20, 70, 252, 924, 3432, 12870
Catalan C_n	1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862
Derangements D_n	1, 0, 1, 2, 9, 44, 265, 1854, 14833
Bell B_n	1, 1, 2, 5, 15, 52, 203, 877, 4140

Counting

$$n! = n(n-1)\cdots 1, \quad 0! = 1$$
$$P(n,k) = \frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1)$$
$$\binom{n}{k} = \frac{P(n,k)}{k!} = \frac{n!}{k!(n-k)!}$$
$$\binom{n}{n_1,\ldots,n_r} = \frac{n!}{n_1!\cdots n_r!}$$
$$\text{strings} : n^k \qquad \text{multisets} : \binom{n+k-1}{k}$$
$$\text{circular} : (n-1)! \qquad \text{necklace} : \tfrac{(n-1)!}{2}$$

Identities

$$\binom{n}{k} = \binom{n}{n-k}$$
$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$
$$k\binom{n}{k} = n\binom{n-1}{k-1}$$
$$\sum_k \binom{n}{k} = 2^n, \qquad \sum_k (-1)^k \binom{n}{k} = 0$$
$$\sum_k k\binom{n}{k} = n2^{n-1}$$
$$\sum_{i=k}^{n} \binom{i}{k} = \binom{n+1}{k+1}$$
$$\sum_j \binom{m}{j}\binom{n}{k-j} = \binom{m+n}{k}$$
$$\sum_k \binom{n}{k}^2 = \binom{2n}{n}$$

Expansions

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$
$$(1+t)^\alpha = \sum_{k\ge 0} \frac{\alpha^{\underline{k}}}{k!} t^k, \quad |t| < 1$$
$$\frac{1}{(1-t)^r} = \sum_{k\ge 0} \binom{k+r-1}{r-1} t^k$$
$$e^x = \sum_{k\ge 0} \frac{x^k}{k!}$$

Probability

$$\mathbb{P}(A) = \frac{|A|}{|S|} \quad \text{(equally likely)}$$
$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$$
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$
$$\mathbb{P}(A) = \sum_j \mathbb{P}(B_j)\mathbb{P}(A \mid B_j)$$
$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(B_i)\mathbb{P}(A \mid B_i)}{\mathbb{P}(A)}$$
$$\frac{\mathbb{P}(H_1 \mid E)}{\mathbb{P}(H_2 \mid E)} = \frac{\mathbb{P}(H_1)}{\mathbb{P}(H_2)} \cdot \frac{\mathbb{P}(E \mid H_1)}{\mathbb{P}(E \mid H_2)}$$

Moments

$$\mathbb{E}[X] = \sum_x x\,\mathbb{P}(X = x) = \sum_{k\ge 1} \mathbb{P}(X \ge k)$$
$$\mathbb{E}\Big[\sum_i X_i\Big] = \sum_i \mathbb{E}[X_i] \quad \text{(always)}$$
$$\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$$
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$
$$\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$$
$$\mathrm{Var}\Big(\sum_i X_i\Big) = \sum_i \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$
$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X \mid Y]\big]$$

Approximations

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$
$$\binom{2m}{m} \sim \frac{4^m}{\sqrt{\pi m}}$$
$$H_n = \sum_{j=1}^{n} \tfrac{1}{j} \approx \ln n + \gamma, \ \gamma \approx 0.5772$$
$$\left(1 - \tfrac{\lambda}{n}\right)^n \to e^{-\lambda}$$
$$\mathrm{Bin}(n,p) \approx \mathrm{Poisson}(np) \ \ (p \text{ small})$$
$$\mathrm{Bin}(n,p) \approx \mathcal{N}\big(np,\, np(1-p)\big)$$

16 Problems

Answers and full solutions follow in Section 17. Try them before looking; the value is entirely in the attempt.

Counting

P1. How many five-letter words can be formed from the 26 letters if no letter repeats?

P2. How many distinguishable arrangements does BOOKKEEPER have?

P3. A committee of 5 is chosen from 12 people, and one member is designated chair. In how many ways? Give two derivations.

P4. How many monotone lattice paths run from $(0,0)$ to $(7,5)$? How many of them pass through $(3,2)$?

P5. How many solutions in integers does $a+b+c+d=20$ have with $0 \le a,b,c,d \le 9$?

P6. Find the coefficient of x^5 in $(1+2x)^9$.

P7. How many subsets of $\{1,\ldots,10\}$ contain no two consecutive integers?

P8. Eight people sit at a round table. In how many seatings are two particular people *not* adjacent?

P9. In how many permutations of $\{1,\ldots,5\}$ does no element stay in its own place?

P10. How many six-digit numbers (no leading zero) have digits summing to 10?

P11. In how many ways can 10 identical balls go into 4 distinct boxes with no box empty?

P12. How many functions from a 5-set onto a 3-set are there?

P13. Ten points lie in the plane with no three collinear. How many triangles do they determine?

P14. Prove $\displaystyle\sum_{k=0}^{n} k^2 \binom{n}{k} = n(n+1)2^{n-2}$.

P15. How many binary strings of length 10 have exactly three 1s, no two adjacent?

Probability

P16. Two fair dice are rolled. Given that at least one shows a 5, what is the probability the total is 8?

P17. What is the probability that a five-card hand contains exactly two aces?

P18. A fair die is rolled five times. What is the probability all five results differ?

P19. One of two coins is chosen at random: one fair, one two-headed. It is flipped three times and lands heads every time. What is the probability it is the two-headed coin?

P20. A fair die is rolled six times. What is the expected number of *distinct* faces seen?

P21. Show that the number of fixed points of a uniformly random permutation of $[n]$ has variance 1 for every $n \ge 2$.

P22. On average, how many rolls of a fair die are needed to see all six faces?

P23. A fair die is rolled ten times. What is the probability of at least two sixes?

P24. Five people check hats and receive them back at random. What is the probability that exactly two people get their own hat?

17 Solutions

P1. Ordered, no repetition: $P(26,5) = 26 \cdot 25 \cdot 24 \cdot 23 \cdot 22 \boxed{7\,893\,600}$.

P2. BOOKKEEPER has 10 letters with O twice, K twice, E three times: $\dfrac{10!}{2!\,2!\,3!} = \dfrac{3\,628\,800}{24} \boxed{151\,200}$.

P3. *Chair last:* $\binom{12}{5} \cdot 5 = 792 \cdot 5 = 3960$. *Chair first:* $12 \cdot \binom{11}{4} = 12 \cdot 330 = 3960$. The agreement is the absorption identity $k\binom{n}{k} = n\binom{n-1}{k-1}$ with $n = 12$, $k = 5$. $\boxed{3960}$

P4. A path is a word with 7 Rs and 5 Us: $\binom{12}{5} = 792$. Through $(3,2)$: split the path at that vertex, $\binom{5}{2}\binom{7}{3} = 10 \cdot 35 \boxed{350}$.

P5. Without the upper bounds: $\binom{23}{3} = 1771$. Subtract, for each variable, the solutions where it is ≥ 10 (shift by 10, leaving total 10): $4\binom{13}{3} = 4 \cdot 286 = 1144$. Two variables can both exceed 9 only if the total is at least 20, which happens exactly when both equal 10 and the others are 0: add back $\binom{4}{2}\binom{3}{3} = 6$. Total $1771 - 1144 + 6 \boxed{633}$.

P6. $T_{k+1} = \binom{9}{k}(2x)^k$, so $k = 5$ gives $\binom{9}{5}2^5 = 126 \cdot 32 \boxed{4032}$.

P7. Choose m elements with no two adjacent from $[n]$: $\binom{n-m+1}{m}$ ways (a stars-and-bars argument on the gaps). Summing over $m = 0, \ldots, 5$: $1 + 10 + 36 + 56 + 35 + 6 = 144 = F_{12}$, the Fibonacci number. $\boxed{144}$

P8. Total circular seatings $7! = 5040$. Seatings with the pair adjacent: glue them into a block, giving $6!$ circular arrangements of 7 objects, times 2 for the internal order: $2 \cdot 720 = 1440$. Answer $5040 - 1440 \boxed{3600}$.

P9. $D_5 = 5!\left(1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120}\right) = 120 \cdot \frac{11}{30} \boxed{44}$. (Check: $\lfloor 120/e + \frac{1}{2} \rfloor = 44$.)

P10. Let the digits be $d_1 \geq 1$ and $d_2, \ldots, d_6 \geq 0$, all at most 9, summing to 10. Substitute $e_1 = d_1 - 1 \geq 0$, so $e_1 + d_2 + \cdots + d_6 = 9$: unrestricted, that is $\binom{14}{5} = 2002$ solutions. Because the total is only 9, no variable can reach 10, and the single upper-bound violation is $e_1 = 9$ (that is $d_1 = 10$). Subtracting it gives $2002 - 1 \boxed{2001}$.

P11. Every box non-empty: substitute $y_i = x_i - 1$, so $y_1 + \cdots + y_4 = 6$, giving $\binom{9}{3} \boxed{84}$.

P12. Inclusion–exclusion on which targets are missed: $3^5 - \binom{3}{1}2^5 + \binom{3}{2}1^5 = 243 - 96 + 3 \boxed{150}$.

P13. Each triangle is a 3-subset: $\binom{10}{3} \boxed{120}$.

P14. Use $k\binom{n}{k} = n\binom{n-1}{k-1}$ twice. Write $k^2 = k(k-1) + k$. Then

$$\sum_k k(k-1)\binom{n}{k} = n(n-1)\sum_k \binom{n-2}{k-2} = n(n-1)2^{n-2}, \qquad \sum_k k\binom{n}{k} = n2^{n-1}.$$

Adding: $n(n-1)2^{n-2} + n2^{n-1} = 2^{n-2}\big(n(n-1) + 2n\big) = n(n+1)2^{n-2}$. (Check $n = 3$: $24 = 3 \cdot 4 \cdot 2$.)

P15. Place the seven 0s, creating eight gaps; choose three gaps for the 1s: $\binom{8}{3} \boxed{56}$.

P16. The conditioning event "at least one 5" has 11 outcomes. Of these, the ones totalling 8 are $(5,3)$ and $(3,5)$. So the answer is $\boxed{\frac{2}{11}} \approx 0.1818$. Note this is *not* $\mathbb{P}(\text{total } 8) = 5/36$: conditioning changed the universe.

P17. $\dfrac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} = \dfrac{6 \cdot 17\,296}{2\,598\,960} = \dfrac{103\,776}{2\,598\,960} \boxed{\approx 0.0399}$.

P18. Ordered with no repetition over ordered with repetition: $\dfrac{P(6,5)}{6^5} = \dfrac{720}{7776} = \dfrac{5}{54} \boxed{\approx 0.0926}$.

P19. Odds form: prior odds 1:1, likelihood ratio $1/(\frac{1}{2})^3 = 8$. Posterior odds 8:1, so the probability is $\boxed{\frac{8}{9}} \approx 0.889$.

P20. Let X count the faces that appear. Write $X = \sum_{j=1}^{6} \mathbf{1}_{\{j \text{ appears}\}}$. Each face fails to appear with probability $(5/6)^6$, so

$$\mathbb{E}[X] = 6\left(1 - \left(\tfrac{5}{6}\right)^6\right) = 6 \cdot \frac{31\,031}{46\,656} \boxed{\approx 3.99}.$$

The indicators are dependent; linearity does not care.

P21. $X = \sum_i \mathbf{1}_{\{\sigma(i)=i\}}$ with $\mathbb{P}(\sigma(i) = i) = \frac{1}{n}$, so $\mathbb{E}[X] = 1$. For the second factorial moment, for $i \neq j$, $\mathbb{P}(\sigma(i) = i,\ \sigma(j) = j) = \dfrac{(n-2)!}{n!} = \dfrac{1}{n(n-1)}$, and there are $n(n-1)$ ordered pairs, so $\mathbb{E}[X(X-1)] = 1$. Hence $\mathbb{E}[X^2] = 2$ and $\mathrm{Var}(X) = 2 - 1^2 \boxed{1}$, independent of n. (This is another sign that X is close to Poisson(1), for which mean and variance are both 1.)

P22. Coupon collector with $n = 6$: $\mathbb{E}[T] = 6H_6 = 6\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}\right) = 6 \cdot \frac{49}{20} = \frac{147}{10} \boxed{14.7}$.

P23. $X \sim \mathrm{Bin}(10, \frac{1}{6})$, and use the complement:

$$\mathbb{P}(X \geq 2) = 1 - \left(\tfrac{5}{6}\right)^{10} - 10 \cdot \tfrac{1}{6}\left(\tfrac{5}{6}\right)^9 \boxed{\approx 0.5155}.$$

P24. Choose the two matched people, derange the other three: $\dfrac{\binom{5}{2} D_3}{5!} = \dfrac{10 \cdot 2}{120} = \dfrac{1}{6} \boxed{\approx 0.1667}$. Compare $e^{-1}/2! \approx 0.1839$: the Poisson approximation is already close at $n = 5$.

If you want to keep going

The natural next steps from here, in rough order of distance travelled:

- **Generating functions.** Encode a counting sequence $a_0, a_1, \ldots$ as $\sum a_n x^n$ and let algebra do the combinatorics. Stars and bars, Catalan numbers, Fibonacci, and partitions all collapse into one method. Start with Wilf's *generatingfunctionology*.
- **The probabilistic method.** Prove a combinatorial object exists by showing a random one works with positive probability. Alon and Spencer is the standard reference; Erdős's lower bound for Ramsey numbers is the one-paragraph entry point.
- **Concentration inequalities.** Chebyshev is the weakest useful bound; Chernoff, Hoeffding, and Azuma give exponential tails and are the working tools of randomised algorithms and statistics.
- **Continuous probability.** Densities, the exponential and normal distributions, the central limit theorem in its general form. Everything in Part II survives with sums replaced by integrals.
- **Pólya theory.** Counting objects up to symmetry (necklaces, colourings, chemical isomers) via Burnside's lemma and the cycle index.

Typeset in LaTeX with Palatino and Avant Garde. Diagrams in TikZ and pgfplots.