DAWN PUTNAM SESSION

Putnam 1984 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1984 — Paper A **5**
1984 A1 5
1984 A2 7
1984 A3 10
1984 A4 14
1984 A5 17
1984 A6 19

Putnam 1984 — Paper B **23**
1984 B1 23
1984 B2 25
1984 B3 28
1984 B4 30
1984 B5 33
1984 B6 37

Putnam 1984 — Paper A

1984 A1

1984 A1. Let A be a solid $a \times b \times c$ rectangular brick in three dimensions, where $a, b, c > 0$. Let B be the set of all points which are a distance at most one from some point of A (in particular, B contains A). Express the volume of B as a polynomial in a, b, and c.

Fields: solid geometry (Minkowski sums, offset volumes) **Position:** A1 — the opener **Answer:** $\mathrm{vol}(B) = abc + 2(ab + bc + ca) + \pi(a + b + c) + \dfrac{4\pi}{3}$. Every point of B lies nearest to a unique point of A (by convexity), and that nearest point is either an interior point of a face, a point on an edge, or a vertex; each of the three types of nearest point contributes a recognisable elementary shape.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see "all points within distance 1 of a box" and reach for the obvious first piece: fatten each of the six faces outward by 1. That gives six rectangular slabs — two $a \times b \times 1$, two $a \times c \times 1$, two $b \times c \times 1$ — on top of A itself, for a running total of $abc + 2(ab + bc + ca)$. You check this against a simple case: take a very thin, very long brick and look near one of its long edges. A point sitting just outside that edge, at distance slightly less than 1 from the edge but not directly "above" either adjacent face, is at distance less than 1 from A (it is within 1 of points on the edge) but lies in neither slab. The six slabs leave gaps along every edge.

So you patch: along each edge, add a cylindrical wedge of radius 1. Since two faces of a box meet at a right angle, the wedge needed to fill the gap along an edge is a quarter-cylinder (a full cylinder swept around the edge would double-count the two adjacent slabs, and only a quarter of it lies outside both slabs). Twelve edges, four of each length a, b, c, contribute $4 \cdot \frac{\pi}{4} a \cdot 1^2 \cdot 3 = \pi(a + b + c)$ once you sort them by length. Checking again near a vertex, you find one more gap: three quarter-cylinders meeting at a corner still leave an eighth of a ball uncovered, so you add eight one-eighth-spheres, $8 \cdot \frac{1}{8} \cdot \frac{4}{3}\pi = \frac{4\pi}{3}$, and only now does the boundary close up with no gaps and no overlaps. This is the right answer, reached by trial, patch, and recheck at each stage — workable, but you found the decomposition by hunting for holes rather than by knowing in advance that there could only be three kinds of piece.

The model chain. A is convex, so every point $p \notin A$ has a unique nearest point $q \in A$ (the point minimising $|p - q|$ over the closed convex set A). This lets you classify every point of $B \setminus A$ by the type of q: q lies in the relative interior of a face, on an edge but not at a vertex, or at a vertex. There are only three types because ∂A, the boundary of a rectangular box, is stratified into exactly those three kinds of piece.

- **Face points.** If q is an interior point of a face, the segment pq is perpendicular to that face (the nearest-point map to a hyperplane is orthogonal projection), so p lies directly above the face at height ≤ 1. This traces out, over each face, a rectangular slab of thickness 1; the six slabs contribute $2(ab + bc + ca)$.

- **Edge points.** If q is an interior point of an edge e (say of length a, where the two adjacent faces meet at a right angle), the points p with nearest point on e fill exactly the quarter of the solid cylinder of radius 1 around e that lies outside both adjacent faces' half-spaces — a quarter-cylinder, because the two faces divide the space around the edge into four right-angle sectors and only one of the four is exterior to A. Summed over all 12 edges (four of each length), this is $\pi(a+b+c)$.
- **Vertex points.** If q is a vertex, three mutually perpendicular faces meet there, cutting the ball of radius 1 around q into eight equal octants, exactly one of which lies outside A. Eight vertices contribute eight octants, i.e. one full sphere: $\frac{4}{3}\pi$.

These three types of piece, together with A itself, partition B exactly (they meet only on sets of measure zero, since the nearest point of a generic point is unique). Hence

$$\mathrm{vol}(B) = abc + 2(ab+bc+ca) + \pi(a+b+c) + \frac{4\pi}{3}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths reach $abc + 2(ab+bc+ca)$ by the same first move: fatten each face. They part at how the remaining region is found. The natural attempt discovers the missing edge- and vertex-pieces by testing specific points and patching holes as they turn up. The model instead classifies *every* point of B in advance by the type of its (unique) nearest point of A, which tells you before you draw anything that there are exactly three kinds of leftover piece and no more.

> *A is convex, so every external point has one and only one nearest point on it. If I sort all of B by which kind of point of A is nearest — a face-interior point, an edge point, or a vertex — I get an exhaustive classification for free, instead of hunting for gaps one at a time.*

> *The shape contributed by each type is forced by the local geometry at that point: perpendicular to a flat face gives a prism (slab); a right-angle edge gives a right-angle wedge of a cylinder, i.e. a quarter; three mutually perpendicular faces at a vertex give one octant of a ball, i.e. an eighth.*

Layer 3 THE WHY

Why does convexity make this classification exhaustive and non-overlapping, rather than just a convenient way to organise the answer? The nearest-point map $\pi_A : \mathbb{R}^3 \to A$ (send p to the closest point of A) is well-defined precisely because A is convex — for a non-convex body a point could have two equally close nearest points, and the classification would need extra care at those points. For a convex body, $\pi_A(p) = q$ if and only if $p - q$ lies in the normal cone of A at q: the set of directions along which moving away from q only increases distance to every point of A. So the type of piece contributed near a point p is entirely determined by the normal cone at $\pi_A(p)$, and the normal cone's shape is a purely local fact about the boundary near q — flat, edge, or vertex.

This is also why the box is the cleanest possible instance of the general phenomenon. At a face-interior point the normal cone is a single ray (perpendicular to the face), giving a slab. At an edge, the normal cone is a wedge whose angle is π minus the interior dihedral angle of the box along that edge; because a box's dihedral angles are all exactly $\pi/2$, every edge wedge is exactly a right angle, i.e. a quarter of the full turn, and the resulting offset piece is a quarter-cylinder — not some other fraction that would need computing case by case. At a vertex, the normal cone is the solid-angle complement of the box's interior angle there; because three mutually perpendicular faces meet at each vertex of a box, that solid angle is exactly

one octant of the sphere, $\frac{1}{8}$ of 4π steradians. A polytope with other dihedral or solid angles (a general prism, say) would still decompose the same way in principle, but the edge and vertex fractions would be $(\pi - \theta)/2\pi$ and (solid angle)$/4\pi$ respectively, rather than the clean $1/4$ and $1/8$ a box supplies — which is exactly why the Putnam committee chose the box, and not a general polytope, for the opening problem of the paper: the mechanism (nearest-point stratification) is the whole content, and the box makes every fraction an integer reciprocal so no further computation is needed once the mechanism is seen.

Finally, notice why the eight vertex-octants sum to a single whole sphere independent of a, b, c: the shape of the octant only depends on the three mutually perpendicular directions at the vertex, never on the edge lengths, so all eight octants are congruent regardless of the box's proportions — a fact that is easy to lose sight of while you are still patching holes by trial, but is obvious once the vertex piece is understood as "the normal cone there," which never sees a, b, or c at all.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "all points within distance 1 of" applied to a convex solid. Write immediately: "every such point has a unique nearest point on the solid; classify by what kind of point that is."

2. The method suggested is stratification of ∂A into faces, edges, and vertices, with each stratum's normal cone determining the local shape (prism, wedge, or cone) of its contribution.

3. Hesitation occurs at trusting that these pieces tile B exactly with no double-counting. It is reinforced by convexity: since the nearest point is unique (off a measure-zero exceptional set), the classification is a genuine partition, not an overlapping cover.

4. The trap is stopping too early — declaring victory after the six slabs, or after slabs plus edge-cylinders — because each partial sum looks like a complete polynomial answer and nothing forces you to check it against a picture unless you go looking for the gap yourself.

5. Recognise this family instantly: "offset region of radius r around a convex polytope" always decomposes into (the polytope) + (face prisms, thickness r) + (edge wedges, fraction $(\pi - \text{dihedral angle})/2\pi$ of a cylinder of radius r) + (vertex pieces, fraction (solid angle)$/4\pi$ of a ball of radius r).

6. Reusable rule: for the volume of the r-neighbourhood of a convex body, classify every added point by its unique nearest point on the body, and note that the local shape of the offset there is completely determined by the normal cone at that nearest point — flat boundary gives a prism, an edge gives a cylindrical wedge sized by its exterior dihedral angle, and a vertex gives a conical/spherical piece sized by its exterior solid angle.

▷ **Cue:** "all points within distance r of a convex solid A" $\longrightarrow$ **Move:** classify by the unique nearest point of A (guaranteed by convexity); partition into face/edge/vertex contributions via the normal cone at each nearest point

▷ **Cue:** a right-angle edge of a polytope, offset outward by radius r $\longrightarrow$ **Move:** a quarter-cylinder of radius r and length equal to the edge

▷ **Cue:** a vertex where three mutually perpendicular faces meet, offset outward by radius r $\longrightarrow$ **Move:** one-eighth of a ball of radius r (a spherical octant)

▷ **Cue:** summing fractional-sphere vertex contributions over all vertices of a box $\longrightarrow$ **Move:** $8 \times \frac{1}{8} = 1$ full sphere, independent of the box's dimensions

1984 A2

1984 A2. Express $\displaystyle\sum_{k=1}^{\infty} \frac{6^k}{(3^{k+1} - 2^{k+1})(3^k - 2^k)}$ as a rational number.

Fields: analysis (series, telescoping sums) **Position:** A2 **Answer:** 2. Writing $a_k = 3^k - 2^k$, the summand equals $\frac{3^k}{a_k} - \frac{3^{k+1}}{a_{k+1}}$, so the series telescopes to $\frac{3}{a_1} - \lim_{k\to\infty} \frac{3^{k+1}}{a_{k+1}} = 3 - 1 = 2$.

Caveat. The source is OCR of the printed problem and the expression is garbled on the page (`6*/(3k + 1 - 2k + 1 )(3k - 2*)`). I have reconstructed it as $\dfrac{6^k}{(3^{k+1}-2^{k+1})(3^k-2^k)}$: this is the only reading consistent with the OCR fragments (a numerator that is a power of 6, a denominator built from $3^{k+1}-2^{k+1}$ and 3^k-2^k), and it is confirmed below by an exact algebraic identity, not merely by guesswork.

Layer 1 THE WHAT — surface analysis

The natural attempt. You recognise this as a series to evaluate exactly, so you check convergence first: 3^k dominates, $a_k := 3^k - 2^k \sim 3^k$, so the summand is $\sim 6^k/3^{2k} = (2/3)^k$, and the series converges by comparison. Good, but you still need the exact value, not just convergence.

You try to attack the denominator directly. $(3^{k+1}-2^{k+1})(3^k-2^k)$ is a quadratic-looking expression in the two exponentials 3^k and 2^k, so you try clearing it by writing everything over 3^{2k}: with $t = (2/3)^k$, the summand becomes

$$\frac{6^k}{(3^{k+1}-2^{k+1})(3^k-2^k)} = \frac{2^k 3^k}{3^{k+1}3^k(1-\frac{2}{3}t)(1-t)} = \frac{t}{3(3-2t)(1-t)}\cdot 3 = \frac{t}{(3-2t)(1-t)},$$

a genuine partial-fractions problem in the single variable t. This works — partial fractions in t give $\frac{t}{(3-2t)(1-t)} = \frac{1}{1-t} - \frac{3}{3-2t}$ — but you have introduced a substitution, a partial-fraction decomposition, and then had to translate the result back into k to see that it telescopes at all. It is correct and it does finish, but you did three separate pieces of algebra (substitution, partial fractions, translation back) before the telescoping structure became visible, and it is easy to make a sign or bookkeeping slip in any one of them under time pressure. (Trying instead to numerically estimate $S(1), S(2), S(3), \ldots$ and pattern-match a limit is faster to *suspect* the answer is 2, but it proves nothing and cannot be the final write-up.)

The model chain. Let $a_k = 3^k - 2^k$ for $k \geq 1$, so $a_1 = 1$ and the given denominator is $a_{k+1}a_k$. A sum whose denominator is a product of two *consecutive* terms of an explicit sequence is a standing invitation to look for a telescoping numerator: a sequence $f(k)$ with

$$\frac{f(k)}{a_k} - \frac{f(k+1)}{a_{k+1}} = \frac{6^k}{a_k a_{k+1}}.$$

The natural first guess for f is the dominant term of a_k itself, namely $f(k) = 3^k$ (since $a_{k+1} = 3\cdot 3^k - 2\cdot 2^k$ visibly contains a 3^{k+1} term, matching the 3^k scale of f). Check it directly:

$$\frac{3^k}{a_k} - \frac{3^{k+1}}{a_{k+1}} = \frac{3^k a_{k+1} - 3^{k+1}a_k}{a_k a_{k+1}} = \frac{3^k(3^{k+1}-2^{k+1}) - 3^{k+1}(3^k-2^k)}{a_k a_{k+1}} = \frac{-3^k 2^{k+1} + 3^{k+1}2^k}{a_k a_{k+1}} = \frac{2^k 3^k(3-2)}{a_k a_{k+1}} = \frac{6^k}{a_k a_{k+1}}$$

It matches exactly. So the Nth partial sum telescopes:

$$S(N) = \sum_{k=1}^{N}\left(\frac{3^k}{a_k} - \frac{3^{k+1}}{a_{k+1}}\right) = \frac{3^1}{a_1} - \frac{3^{N+1}}{a_{N+1}} = 3 - \frac{3^{N+1}}{3^{N+1}-2^{N+1}}.$$

As $N \to \infty$, $\dfrac{3^{N+1}}{3^{N+1}-2^{N+1}} = \dfrac{1}{1-(2/3)^{N+1}} \to 1$. Hence the series converges to $3 - 1 = 2$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths notice the same fact about the denominator: it is built from $3^k - 2^k$ at two consecutive indices. They split at what to do with that observation. The natural attempt treats it as an algebraic obstacle to be cleared by substitution and standard partial fractions in a new variable t. The model treats it as a direct cue for the telescoping ansatz $f(k)/a_k - f(k+1)/a_{k+1}$, guesses $f(k) = 3^k$ from the shape of a_{k+1} alone, and verifies the guess by one algebraic identity in k — no substitution required.

> *The denominator is literally $a_k \cdot a_{k+1}$ for one explicit sequence $a_k = 3^k - 2^k$, not an arbitrary quadratic in two independent exponentials. A term of the form $c_k/(a_k a_{k+1})$ is exactly the shape that telescopes if I can find $f(k)$ with $c_k = a_{k+1}f(k) - a_k f(k+1)$. Let me look for that f before doing anything else to the expression.*

> *$a_{k+1} = 3 \cdot 3^k - 2 \cdot 2^k$ already contains a 3^{k+1} term at the same scale as the numerator $6^k = 2^k 3^k$. The cheapest guess for f is $f(k) = 3^k$, matching that scale exactly; if it works, I have avoided introducing any new variable at all.*

Layer 3 THE WHY

Why should the shape "product of consecutive terms of one sequence" be read as a telescoping cue rather than as an invitation to partial fractions in a new variable? Because partial fractions in k (in the calculus sense, over a rational function of k) presupposes a *polynomial* denominator in k; here the denominator is exponential in k, so ordinary partial-fraction machinery does not literally apply to k at all — it only becomes available after the substitution $t = (2/3)^k$ converts the problem into one about a rational function of t. That substitution is not wrong, but it is a detour: it exists to manufacture the polynomial structure that a telescoping ansatz would use directly, in the variable k, without ever leaving the original variable. Recognising "$a_k a_{k+1}$ in the denominator" as the telescoping shape is recognising that the detour is unnecessary — the sum is already in the family $\sum c_k/(a_k a_{k+1})$, for which the identity $c_k = a_{k+1}f(k) - a_k f(k+1)$, when solvable, gives $\sum_{k=1}^{N} c_k/(a_k a_{k+1}) = f(1)/a_1 - f(N+1)/a_{N+1}$ directly by cancellation — this is nothing more than the definition of a telescoping sum, stated for a general sequence a_k rather than the special case (e.g. $a_k = k$) most often drilled.

Why is $f(k) = 3^k$ the right first guess and not a leap? Because $6^k = 2^k \cdot 3^k$ and $a_{k+1} = 3 \cdot 3^k - 2 \cdot 2^k$: the numerator already lives at the scale $2^k 3^k$, which is exactly the cross-term you get when you multiply 3^k (a candidate $f(k)$) into a_{k+1}'s $-2 \cdot 2^k$ part (or 3^{k+1} into a_k's -2^k part). Trying the simplest exponential appearing in a_k itself, before trying anything more elaborate, is the same discipline as trying the lowest-degree polynomial ansatz first in a calculus telescoping problem (e.g. guessing $f(k) = k^2$ before $f(k) = k^3$ for a sum whose terms look quadratic) — match the growth rate of the candidate to the growth rate of the target, and verify by direct substitution rather than by manufacturing more algebra to search blindly.

Finally, why does the limit come out to a clean integer rather than an irrational number? Because $3^{k+1}/a_{k+1} = 1/(1 - (2/3)^{k+1}) \to 1$ exactly, since $(2/3)^{k+1} \to 0$: the two exponential bases 2 and 3 enter only through their ratio, and that ratio vanishing in the limit is what collapses the tail of the telescoped sum to a rational number (1) rather than leaving an irrational residue — this is the mechanism the Putnam committee is testing when it asks for the answer "as a rational number": a student who reroutes to the wrong tool (say, trying to force this into a geometric series formula directly) will not obviously get a rational number out, whereas the correct telescoping expression manifestly is one at every finite stage.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the denominator is $a_{k+1}a_k$ for the single explicit sequence $a_k =$

$3^k - 2^k$. Write down immediately: "look for $f(k)$ with $f(k)/a_k - f(k+1)/a_{k+1}$ equal to the summand."

2. The method suggested is the telescoping-ansatz for sums of the shape $\sum c_k/(a_k a_{k+1})$, verified by a direct algebraic identity in k rather than a substitution.

3. Hesitation occurs at choosing f. It is resolved by matching scales: the numerator $6^k = 2^k 3^k$ and the term $3 \cdot 3^k$ inside a_{k+1} are at the same order, so try $f(k) = 3^k$ (the dominant term of a_k) before trying anything more complicated.

4. The trap is treating the exponential denominator as a job for ordinary rational-function partial fractions in k, which forces an unnecessary substitution $t = (2/3)^k$ before the telescoping structure becomes visible — extra steps, extra chances for a sign error, no extra insight.

5. Recognise this shape instantly: whenever a series' denominator is a product of two consecutive terms of an explicit sequence a_k, default to hunting a telescoping numerator $f(k)/a_k$ before reaching for any other technique.

6. Reusable rule: $\sum_{k=1}^{\infty} c_k/(a_k a_{k+1})$ telescopes to $f(1)/a_1 - \lim_{k\to\infty} f(k+1)/a_{k+1}$ whenever $c_k = a_{k+1}f(k) - a_k f(k+1)$; try $f(k)$ equal to the dominant term of a_k first, and confirm by direct substitution in k, not by introducing a new variable.

▷ **Cue:** a series whose denominator is a product of two consecutive terms $a_k a_{k+1}$ of an explicit sequence $\longrightarrow$ **Move:** search for a telescoping numerator $f(k)$ with $f(k)/a_k - f(k+1)/a_{k+1}$ equal to the summand; verify by direct algebra in k

▷ **Cue:** a_k is a difference of two exponentials $p^k - q^k$ with $p > q > 0$, appearing in a series denominator $\longrightarrow$ **Move:** try $f(k) = p^k$ (the dominant term) as the first telescoping candidate

▷ **Cue:** after telescoping, a partial sum reduces to $f(1)/a_1 - f(N+1)/a_{N+1}$ $\longrightarrow$ **Move:** evaluate the limit of the tail term by dividing through by the dominant exponential, turning it into $1/(1-(q/p)^{N+1}) \to 1$

1984 A3

1984 A3. Let n be a positive integer. Let a, b, x be real numbers, with $a \neq b$, and let M_x denote the $2n \times 2n$ matrix whose (i, j) entry m_{ij} is given by

$$m_{ij} = \begin{cases} x & \text{if } i = j, \\ a & \text{if } i \neq j \text{ and } i + j \text{ is even}, \\ b & \text{if } i \neq j \text{ and } i + j \text{ is odd}. \end{cases}$$

Thus, for example,

$$M_2 = \begin{pmatrix} x & b & a & b \\ b & x & b & a \\ a & b & x & b \\ b & a & b & x \end{pmatrix}.$$

Express $\lim_{x \to a} \dfrac{\det M_x}{(x-a)^{2n-2}}$ as a polynomial in a, b, and n, where $\det M_x$ denotes the determinant of M_x.

Fields: linear algebra (eigenvalues, determinants) **Position:** A3 **Answer:** $n^2(a^2 - b^2) = n^2(a-b)(a+b)$. The off-diagonal entries never involve x, so $M_x = N + (x-a)I$ where $N := M_a$; the eigenvalues of N are $n(a+b)$, $n(a-b)$, and 0 with multiplicity $2n-2$, so $\det M_x = (x-a)^{2n-2}\big(n(a+b) + (x-a)\big)\big(n(a-b) + (x-a)\big)$, and the limit is the product $n(a+b) \cdot n(a-b)$.

Caveat. The OCR of the piecewise definition of m_{ij} drops the three values "x", "a", "b" entirely, leaving only the conditions. I have reconstructed the definition above from the given 4×4 example M_2 (checking, entry by entry, that $m_{ii} = x$, and $m_{ij} = a$ exactly when $i \neq j$ and $i + j$ is even, $m_{ij} = b$ exactly when $i \neq j$ and $i + j$ is odd, reproduces every entry of M_2 as printed) and confirm it is the intended reading below by an independent verification against the small case $n = 1$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You first check $n = 1$: $M_x = \begin{pmatrix} x & b \\ b & x \end{pmatrix}$ (since for 2×2, $i + j = 3$ is always odd off the diagonal), so $\det M_x = x^2 - b^2$, and $\lim_{x \to a} (x^2 - b^2)/(x - a)^0 = a^2 - b^2$. That matches $n^2(a^2 - b^2)$ at $n = 1$, which is reassuring, but a single case does not prove the general formula, and you need a method that works for all n.

For general n, you notice something structural: the off-diagonal pattern only depends on the *parity* of the index, not on the index itself, so you sort the $2n$ indices into odds and evens and reorder rows and columns accordingly (an operation that does not change the determinant, since it is the same permutation on rows and columns). In the new order, M_x becomes the 2×2 block matrix

$$M'_x = \begin{pmatrix} D & B \\ B & D \end{pmatrix}, \qquad D = (x - a) I_n + a J_n, \qquad B = b J_n,$$

where J_n is the $n \times n$ all-ones matrix (within a class, off-diagonal entries are a; across classes, all entries are b; the diagonal correction $(x - a)$ absorbs $m_{ii} = x$ into a matrix that is otherwise aJ_n). Since I_n and J_n commute, so do D and B, and the identity $\det \begin{pmatrix} D & B \\ B & D \end{pmatrix} = \det(D + B) \det(D - B)$ applies. Using the eigenvalues of J_n (namely n, once, and 0, with multiplicity $n - 1$), you get $\det(cI_n + dJ_n) = c^{n-1}(c + dn)$, and after two applications and a fair amount of bookkeeping you arrive at

$$\det M_x = (x - a)^{2n-2} \Big[(x - a + an)^2 - (bn)^2 \Big],$$

which does give the right limit. This is a complete and correct proof, but it computes the *entire polynomial* $\det M_x$ in x for general n — far more than the problem actually asks for, since only the limiting behaviour at $x = a$ is wanted, and every step (the reordering, the block-determinant identity, the eigenvalues of J_n, two separate expansions) is machinery you have to set up and then discard once you take $x \to a$.

The model chain. Look again at what the problem is actually varying: x appears *only* on the diagonal. Every off-diagonal entry is a or b, fixed, independent of x. So, writing $N := M_a$ (the matrix you get by setting $x = a$),

$$M_x = N + (x - a)\, I_{2n}.$$

This is not an arbitrary one-parameter family of matrices — it is a uniform shift of the diagonal of a single fixed matrix N. For such a family, $\det M_x = \det(N + tI)$ with $t = x - a$ is exactly the characteristic polynomial of $-N$ evaluated at $-t$, so if $\lambda_1, \ldots, \lambda_{2n}$ are the eigenvalues of N (with multiplicity),

$$\det M_x = \prod_{i=1}^{2n} (\lambda_i + t).$$

The exponent $2n - 2$ in the problem is now a direct instruction: it says $t = 0$ (i.e. $x = a$) is a root of this product of multiplicity (at least) $2n - 2$, i.e. N has eigenvalue 0 with multiplicity $2n - 2$, i.e. $\operatorname{rank}(N) \leq 2$. Confirm this directly: N has $m_{ii} = a$, $m_{ij} = a$ for $i \neq j$ same parity, $m_{ij} = b$ for $i \neq j$ different parity — so, in the same odd/even reordering as before, every row

of N restricted to the first n (odd-index) columns is $(a,\dots,a)$ or $(b,\dots,b)$, and likewise for the last n; concretely, every row of N is one of only two distinct row vectors, $(a,\dots,a,b,\dots,b)$ or $(b,\dots,b,a,\dots,a)$ (n entries each). A matrix with only two distinct rows (and $a \neq b$, so they are not proportional) has rank exactly 2. So N has eigenvalue 0 with multiplicity exactly $2n-2$, and two more eigenvalues to find.

Find them without the characteristic polynomial. Let e be the all-ones column vector. Every row of N sums to the same value: for any index i, among the other $2n-1$ indices exactly $n-1$ share its parity (contributing a each) and n have the opposite parity (contributing b each), plus the diagonal entry a itself — giving row sum $a+(n-1)a+nb = na+nb = n(a+b)$, regardless of whether i is odd or even. So $Ne = n(a+b)e$: e is an eigenvector with eigenvalue $n(a+b)$. Since N has rank 2, exactly one more eigenvalue is nonzero, call it μ; the trace of N is $2n \cdot a$ (its diagonal is a everywhere), and the trace equals the sum of all eigenvalues, which is $n(a+b)+\mu$ (the remaining $2n-2$ eigenvalues are 0). So

$$\mu = 2na - n(a+b) = n(a-b).$$

Hence $\det M_x = t^{2n-2}\big(n(a+b)+t\big)\big(n(a-b)+t\big)$ with $t = x-a$, and

$$\lim_{x\to a} \frac{\det M_x}{(x-a)^{2n-2}} = n(a+b)\cdot n(a-b) = n^2(a^2-b^2).$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths notice that off-diagonal entries never depend on x. They split at what to do with that fact. The natural attempt uses it to reorganise M_x into a 2×2 block form and computes the entire determinant polynomial in x, for general n, before specialising to $x = a$. The model uses the same fact to write $M_x = N + (x-a)I$ outright, recognises this as a spectral-shift problem, and reduces the whole question to the eigenvalues of the single fixed matrix $N = M_a$ — never writing down $\det M_x$ as a polynomial in x at all.

> *x only ever multiplies the diagonal; every off-diagonal entry is fixed. That is not an arbitrary parametrised family — it is $M_x = M_a + (x-a)I$, a uniform shift of a fixed matrix's diagonal. Whenever a matrix is fixed-part-plus-scalar-times-identity, its determinant is the characteristic polynomial of the fixed part; I should go straight to the eigenvalues of $N = M_a$, not to a general expansion in x.*

> *The problem hands me the exponent $2n-2$ inside $\lim_{x\to a} \det M_x/(x-a)^{2n-2}$. That number is not decoration — it is telling me $\det M_x$ vanishes to order (at least) $2n-2$ at $x=a$, which for $\det(N+tI) = \prod(\lambda_i + t)$ means exactly that N has eigenvalue 0 with multiplicity $2n-2$, i.e. $\mathrm{rank}(N) \leq 2$. I should verify the rank directly rather than deriving it as a byproduct of a full computation.*

> *To find the nonzero eigenvalues of a low-rank matrix, I do not need the characteristic polynomial: an obvious symmetry vector (here, the all-ones vector, since N has constant row sums) gives one eigenvalue directly, and the trace — the sum of all eigenvalues, most of which are already known to be zero — gives the last one for free.*

Layer 3 THE WHY

Why is "$M_x = N + (x-a)I$" the observation to reach for, rather than an incidental simplification? Because a determinant's behaviour near a specific value of a parameter is governed by how the matrix depends on that parameter, and the simplest possible dependence a matrix

can have on a scalar is an additive shift of the whole diagonal — which is, by definition, the situation $\det(N + tI)$ that the characteristic polynomial exists to describe. Once you see that x enters only this way, the entire machinery of eigenvalues (which are defined exactly so that $\det(N + tI) = \prod(\lambda_i + t)$) becomes available for free, and the specific matrix N is reduced to two numbers to find (its nonzero eigenvalues) rather than a polynomial in two variables to expand. This is a fact about *any* matrix family of the form (fixed matrix) $+x \cdot I$, independent of what the fixed matrix happens to look like — it is the transferable content of this problem, not an accident of this particular a, b pattern.

Why does the exponent $2n - 2$, printed in the problem itself, deserve to be read as information about rank rather than treated as just the correct power to divide by at the end? Because the problem is handing you, for free, the answer to "what is the order of vanishing of $\det M_x$ at $x = a$?" — and for a matrix of the shifted form $N + tI$, order of vanishing at $t = 0$ *is* the algebraic multiplicity of the eigenvalue 0 of N, which *is* $2n - \operatorname{rank}(N)$ (since N, being symmetric — real, and equal to its own transpose by inspection of the definition — is diagonalisable, so algebraic and geometric multiplicity coincide, and the rank is the number of nonzero eigenvalues counted with multiplicity). So the stated exponent is simultaneously the answer's justification (why rank 2 is even plausible) and a checkable prediction (verify $\operatorname{rank}(N) = 2$ directly, as done above, rather than trusting the problem's exponent blindly) — using the given data to aim the proof is not laziness, it is reading the problem correctly.

Why is the all-ones vector automatically an eigenvector of N, and not a lucky guess? N is built from a two-class partition of the indices (odd/even), and its entries depend only on which classes i and j fall into. For *any* such matrix, the row sum over index i depends only on how many other indices share i's class and how many do not — and because there are exactly n odd and n even indices among $2n$ total, that count ($n - 1$ same-class, n different-class) is the same no matter which class i itself belongs to. A matrix whose entries depend only on class membership, over equally-sized classes, therefore always has constant row sums, and a matrix with constant row sums always has the all-ones vector as an eigenvector, by the definition of matrix-vector multiplication (Ne just re-sums each row). This is the general mechanism — sometimes called an *equitable partition* in graph and matrix theory — and it is why guessing e costs nothing: it follows from the symmetry of the construction, not from noticing this particular problem's numbers.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that x appears only on the diagonal. Write immediately: "$M_x = M_a + (x - a)I$", a scalar shift of a fixed matrix.

2. The method suggested is spectral: $\det(N + tI) = \prod(\lambda_i + t)$, so the limit reduces to finding the eigenvalues of $N = M_a$.

3. Hesitation occurs at finding those eigenvalues without grinding through a characteristic polynomial. It is reinforced by reading the problem's own exponent $2n - 2$ as telling you N has rank 2, and then finding the two nonzero eigenvalues cheaply: one from a symmetry vector (constant row sums $\Rightarrow$ the all-ones vector is an eigenvector), the other from the trace.

4. The trap is doing the reordering-into-blocks computation for general n in full: correct, but far more algebra than the problem requires, since it computes the whole polynomial $\det M_x$ when only its leading behaviour at $x = a$ is needed. A second trap is checking only $n = 1$ (or $n = 1, 2$) and pattern-matching the formula without proof.

5. Recognise this shape instantly: whenever a parameter appears in a matrix only as an additive shift of the entire diagonal, reduce immediately to the eigenvalues of the matrix at the base value, rather than expanding the determinant as a polynomial in that parameter.

6. Reusable rule: if $M_x = N + (x - c)I$, then $\det M_x/(x - c)^{2n-k} \to$ (product of the k nonzero eigenvalues of N) as $x \to c$, where $2n - k$ is the multiplicity of 0 as an eigenvalue of N; find those nonzero eigenvalues cheaply via rank, an obvious symmetry eigenvector (e.g. constant row/column sums $\Rightarrow$ the all-ones vector), and the trace, rather than via the full

characteristic polynomial.

- ▷ **Cue:** a matrix family M_x in which the parameter x appears only on the diagonal $\longrightarrow$ **Move:** write $M_x = N + (x-c)I$ for the fixed matrix $N = M_c$, and use $\det(N + tI) = \prod(\lambda_i + t)$
- ▷ **Cue:** a limit of the form $\det M_x/(x-c)^{2n-k}$ as $x \to c$ $\longrightarrow$ **Move:** read the exponent as the multiplicity of the eigenvalue 0 of $N = M_c$, i.e. $\operatorname{rank}(N) = k$; the limit is the product of the k nonzero eigenvalues
- ▷ **Cue:** a matrix whose entries depend only on which class (of an equally-sized partition of the indices) each index belongs to $\longrightarrow$ **Move:** it automatically has constant row sums, so the all-ones vector is an eigenvector with eigenvalue equal to the common row sum
- ▷ **Cue:** needing the remaining eigenvalue(s) of a matrix once its rank and one eigenvalue are known $\longrightarrow$ **Move:** use trace = sum of eigenvalues, since every eigenvalue outside the known rank is 0

1984 A4

1984 A4. A convex pentagon $P = ABCDE$, with vertices labeled consecutively, is inscribed in a circle of radius 1. Find the maximum area of P subject to the condition that the chords AC and BD be perpendicular.

Fields: geometry (circle geometry, arcs and central angles); optimisation (concavity / Jensen) **Position:** A4 — mid-paper; a five-line problem once the pentagon is described by its arcs rather than by its vertices, and a multivariable-calculus swamp otherwise **Answer:** $1 + \dfrac{3\sqrt{3}}{4}$, attained when arcs AB and CD are each a quarter of the circle and arcs BC, DE, EA are each a sixth of it.

Layer 1 THE WHAT — surface analysis

The natural attempt. You put the circle at the origin and the vertices at $A = (\cos\alpha_1, \sin\alpha_1)$, $\dots$, $E = (\cos\alpha_5, \sin\alpha_5)$ with $\alpha_1 < \alpha_2 < \cdots < \alpha_5 < \alpha_1 + 2\pi$. The area comes from the shoelace formula,

$$[P] = \tfrac{1}{2}\sum_{i=1}^{5}(x_i y_{i+1} - x_{i+1}y_i) = \tfrac{1}{2}\sum_{i=1}^{5}\sin(\alpha_{i+1} - \alpha_i),$$

which is already a pleasant surprise. Then the constraint: $\overrightarrow{AC} \cdot \overrightarrow{BD} = 0$, that is

$$(\cos\alpha_3 - \cos\alpha_1)(\cos\alpha_4 - \cos\alpha_2) + (\sin\alpha_3 - \sin\alpha_1)(\sin\alpha_4 - \sin\alpha_2) = 0.$$

You expand it: $\cos(\alpha_3 - \alpha_4) - \cos(\alpha_3 - \alpha_2) - \cos(\alpha_1 - \alpha_4) + \cos(\alpha_1 - \alpha_2) = 0$. Now you have five angles (four, after fixing $\alpha_1 = 0$ by rotation), a trigonometric objective and a trigonometric constraint, and the reflex is Lagrange multipliers. Five partial derivatives, a multiplier, a nonlinear system in cosines. It can be pushed through, but the system does not visibly decouple, you have to argue separately that the critical point you find is a maximum rather than a saddle, and the boundary (degenerate pentagons, where some arc shrinks to 0) has to be checked by hand. Most people who start here either run out of time or produce a critical point without a proof that it is the maximum.

The model chain. *Coordinates.* Let O be the centre. Going round the circle, write the central angles (arc measures) as

$$\theta = \overset{\frown}{AB}, \quad u = \overset{\frown}{BC}, \quad \varphi = \overset{\frown}{CD}, \quad v = \overset{\frown}{DE}, \quad w = \overset{\frown}{EA}, \qquad \theta + u + \varphi + v + w = 2\pi,$$

all positive, since the five vertices are distinct and in order.

Area. Split P into the five triangles $OAB, \dots, OEA$. Triangle OXY with $OX = OY = 1$ and angle t at O has area $\frac{1}{2}\sin t$. This holds as a signed area even if some central angle exceeds π (then O lies outside P, the triangle is counted negatively, and $\sin t < 0$ does exactly that). So

$$[P] = \tfrac{1}{2}\big(\sin\theta + \sin u + \sin\varphi + \sin v + \sin w\big).$$

Constraint. Since A, B, C, D are in order round the circle, the chords AC and BD cross at an interior point X. The angle between two chords meeting inside a circle is half the sum of the arcs they cut off: $\angle AXB = \frac{1}{2}(\overset{\frown}{AB} + \overset{\frown}{CD})$. (Proof: $\angle AXB$ is an exterior angle of triangle XBC, so $\angle AXB = \angle XCB + \angle XBC = \angle ACB + \angle DBC$, and each inscribed angle is half its arc: $\frac{1}{2}\theta + \frac{1}{2}\varphi$.) Hence

$$AC \perp BD \iff \theta + \varphi = \pi.$$

Separate. Then $\sin\varphi = \sin(\pi - \theta) = \sin\theta$, and $u + v + w = 2\pi - \pi = \pi$. So

$$[P] = \sin\theta + \tfrac{1}{2}\big(\sin u + \sin v + \sin w\big), \qquad 0 < \theta < \pi, \quad u, v, w > 0,\ u + v + w = \pi.$$

The two pieces involve disjoint variables, so maximise each alone.

First piece. $\sin\theta \le 1$, with equality at $\theta = \pi/2$.

Second piece. sin is concave on $[0, \pi]$ (its second derivative $-\sin t \le 0$ there), and $u, v, w \in (0, \pi)$. By Jensen,

$$\sin u + \sin v + \sin w \le 3\sin\frac{u + v + w}{3} = 3\sin\frac{\pi}{3} = \frac{3\sqrt{3}}{2},$$

with equality at $u = v = w = \pi/3$.

Attained. $\theta = \varphi = \pi/2$, $u = v = w = \pi/3$ gives five positive arcs summing to 2π, hence a genuine convex inscribed pentagon satisfying the constraint. Its area is

$$1 + \tfrac{1}{2}\cdot\tfrac{3\sqrt{3}}{2} = 1 + \frac{3\sqrt{3}}{4} \approx 2.299.$$

(For comparison, the regular pentagon has area $\frac{5}{2}\sin 72^\circ \approx 2.378$; it is larger, and it fails the constraint, since there $\theta + \varphi = \frac{4\pi}{5} \ne \pi$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the area formula $\frac{1}{2}\sum\sin(\text{central angle})$ — even the shoelace computation reaches it. They split at the constraint. The natural attempt writes perpendicularity as a dot product in coordinates and gets a four-cosine equation. The model writes it as a statement about arcs, $\overset{\frown}{AB} + \overset{\frown}{CD} = \pi$, and at that moment the problem falls apart into two independent one-line maximisations.

> *The area is already a sum over arcs. So I want the constraint in terms of arcs too, not in terms of coordinates. What does "two chords are perpendicular" say about arcs? The angle between crossing chords is half the sum of the intercepted arcs — so perpendicular means those two arcs add to a semicircle.*

> *Now* $\sin\varphi = \sin\theta$*, and the other three arcs share the other semicircle with no further condition. That is two separate problems: make one sine as big as possible, and share a fixed total among three sines — which is concavity, so share it equally.*

Layer 3 THE WHY

Why arcs? Go back to what the problem fixes. Every vertex lies on the same circle of radius 1, so a vertex has exactly one degree of freedom: where it sits round the circle. Differences of those positions are the arcs, and a rotation of the whole figure changes nothing, so the arcs (four free, since they sum to 2π) are the true coordinates of the problem. The x- and y-coordinates are two numbers per vertex tied by $x^2 + y^2 = 1$ — a redundant description, and every redundancy in the description turns into algebra you have to undo later. That is why the natural attempt's constraint came out as a sum of four cosines: it is the arc condition, written in a language that hides it.

Why does the area come out in arcs? Because the centre is the one point you know, and a triangle with two sides equal to the radius has area $\frac{1}{2}\sin(\text{apex angle})$. The circle is the set of points at distance 1 from O; therefore O is the object of interest; therefore coning the pentagon from O is not a trick. This is the circle example from the Blueprint in its purest form.

Why does the constraint come out in arcs? Because every angle formed by chords of a circle is measured by arcs: an inscribed angle is half its arc (again because of isosceles triangles with apex O), and the angle between two chords crossing inside is an exterior angle of a triangle whose other two angles are inscribed. So once both the objective and the constraint are "about arcs", the question is purely: which arcs appear in which? The constraint mentions $\overset{\frown}{AB}$ and $\overset{\frown}{CD}$ only; the objective is a sum of one function of each arc. That is the structural reason for the decoupling: the constraint locks two arcs to each other and fences off a semicircle for the other three.

Why equal sharing? sin on $[0, \pi]$ bends downwards, so moving two unequal arcs towards each other, keeping their sum, raises $\sin u + \sin v$ (the mean of a concave function at two points lies below its value at the mean). Any unequal triple can be improved, so the maximum is at the equal one. That is the whole content of Jensen here, and it is the same reason the regular n-gon is the largest n-gon in a circle.

The official solution in the MAA volume begins "Let $\theta = \overset{\frown}{AB}$, $\alpha = \overset{\frown}{DE}$, $\beta = \overset{\frown}{EA}$; then $\overset{\frown}{CD} = \pi - \theta$" and states the maximiser without the chord-angle theorem or the concavity argument. Both omitted steps are above. An essay by a former Questions Committee member in the 1985–2000 volume recalls this very problem as made by taking a combination of simple objects and staring at them until a problem appeared — which is the tell: the problem was built to separate once you use the circle's own coordinates.

Layer $+\alpha$ BACKTESTING

1. The cue is "inscribed in a circle of radius 1". Write down: the centre O, the five arcs θ, u, φ, v, w summing to 2π, and $[P] = \frac{1}{2}\sum\sin(\text{arc})$.

2. Parametrise by central angles and translate every condition into arcs; then optimise one-variable pieces, using concavity for a fixed-sum share-out.

3. Hesitation comes at "$AC \perp BD$", where the dot product is the reflex. Reinforce it away with: every angle made by chords is measured by arcs — inscribed angle $= \frac{1}{2}$ arc, crossing chords $= \frac{1}{2}$ sum of arcs, secants from outside $= \frac{1}{2}$ difference.

4. The trap is Lagrange multipliers on a coordinate constraint: correct in principle, uncoupled in fact, but you cannot see the uncoupling and you inherit a second-order check and a boundary analysis. A smaller trap: assuming the regular pentagon is the answer; it violates the constraint.

5. Any polygon inscribed in a fixed circle, any constraint on angles between chords: the arcs are the coordinates. When the objective is a sum of one function per arc, look for constraints that touch only some of the arcs — they split the problem.

6. Reusable rule: describe the configuration in the variables in which the objective is additive, then rewrite the constraint in the same variables. If the constraint only links some variables,

the optimisation separates; a fixed sum under a concave function is shared equally.

- ▷ **Cue:** a polygon inscribed in a circle of known radius, area asked ⟶ **Move:** cone from the centre: area $= \frac{1}{2}R^2 \sum \sin(\text{central angles})$
- ▷ **Cue:** two chords of a circle are perpendicular (or meet at a given angle) ⟶ **Move:** intercepted arcs: the angle is half the sum of the two arcs cut off
- ▷ **Cue:** maximise $\sum g(x_i)$ with $\sum x_i$ fixed and g concave on the range ⟶ **Move:** Jensen: all x_i equal
- ▷ **Cue:** a constrained optimisation whose constraint involves only some of the variables ⟶ **Move:** maximise the untouched block and the constrained block separately
- ▷ **Cue:** points on a fixed circle described by (x, y) ⟶ **Move:** replace by angles; the redundancy $x^2 + y^2 = 1$ is where the algebra comes from

1984 A5

1984 A5. Let R be the region consisting of all triples (x, y, z) of nonnegative real numbers satisfying $x + y + z \le 1$. Let $w = 1 - x - y - z$. Express the value of the triple integral

$$\iiint_R x^1 y^9 z^8 w^4 \, dx \, dy \, dz$$

in the form $a!\, b!\, c!\, d!/n!$, where a, b, c, d, and n are positive integers.

Fields: analysis (multiple integrals, Beta and Gamma integrals, Dirichlet integral) **Position:** A5 — late in the paper but far gentler than its position suggests; a ten-line computation if you integrate one variable at a time and rescale, a page of multinomial coefficients if you expand w^4 **Answer:** $\dfrac{1!\,9!\,8!\,4!}{25!}$, i.e. $a = 1$, $b = 9$, $c = 8$, $d = 4$, $n = 25$. (Numerically $1/44\,172\,388\,260\,000$.)

Layer 1 THE WHAT — surface analysis

The natural attempt. You set up the iterated integral in the standard order,

$$\int_0^1 \int_0^{1-x} \int_0^{1-x-y} x\, y^9 z^8 (1 - x - y - z)^4 \, dz \, dy \, dx,$$

and look at the inner integral. The factor $(1 - x - y - z)^4$ is a polynomial, so you expand it — binomially in z against $(1 - x - y)$, say:

$$(1 - x - y - z)^4 = \sum_{j=0}^{4} \binom{4}{j} (1 - x - y)^{4-j} (-z)^j.$$

Integrating z^{8+j} from 0 to $1 - x - y$ gives $(1 - x - y)^{9+j}/(9 + j)$, so the inner integral is $(1 - x - y)^{13} \sum_j \binom{4}{j}(-1)^j/(9 + j)$. That alternating sum is a specific fraction, and you compute it $(\frac{1}{9} - \frac{4}{10} + \frac{6}{11} - \frac{4}{12} + \frac{1}{13})$. Then the middle integral has $y^9(1 - x - y)^{13}$, and expanding $(1 - x - y)^{13}$ gives fourteen terms with alternating signs. At this point the arithmetic is out of hand, and — worse — the form $a!b!c!d!/n!$ that the problem demands is invisible: you are producing a fraction with a denominator like $9 \cdot 10 \cdot 11 \cdot 12 \cdot 13$ and no reason to expect it to reorganise into factorials. Some people notice that the first alternating sum equals $8!\,4!/13!$ and guess the pattern; the ones who do not stop here.

The model chain. *The one-variable lemma.* For integers $p, q \ge 0$ and $L > 0$,

$$\int_0^L t^p (L - t)^q \, dt = \frac{p!\, q!}{(p + q + 1)!} L^{p+q+1}.$$

Proof. Substitute $t = Ls$: the integral is $L^{p+q+1}J(p,q)$ with $J(p,q) = \int_0^1 s^p(1-s)^q\,ds$. Integrating by parts (differentiate $(1-s)^q$, integrate s^p; the boundary terms vanish for $q \geq 1$) gives $J(p,q) = \frac{q}{p+1}J(p+1,q-1)$. Repeating q times, $J(p,q) = \frac{q!\,p!}{(p+q)!}J(p+q,0) = \frac{p!\,q!}{(p+q)!}\cdot\frac{1}{p+q+1}$.

Inner integral (z). Fix x, y with $x + y \leq 1$ and put $L = 1 - x - y$. Then $w = L - z$, and

$$\int_0^L z^8(L-z)^4\,dz = \frac{8!\,4!}{13!}\,(1-x-y)^{13}.$$

Middle integral (y). Fix x and put $L = 1 - x$. The integrand is now $\frac{8!4!}{13!}\,y^9(L-y)^{13}$, so

$$\int_0^L y^9(L-y)^{13}\,dy = \frac{9!\,13!}{23!}\,(1-x)^{23}.$$

Outer integral (x).

$$\int_0^1 x^1(1-x)^{23}\,dx = \frac{1!\,23!}{25!}.$$

Multiply.

$$\iiint_R xy^9z^8w^4 = \frac{8!\,4!}{13!}\cdot\frac{9!\,13!}{23!}\cdot\frac{1!\,23!}{25!} = \frac{1!\,9!\,8!\,4!}{25!}.$$

The intermediate factorials 13! and 23! cancel, as they had to.

General form (Dirichlet). The same three steps prove, for nonnegative integers a, b, c, d,

$$\iiint_R x^a y^b z^c w^d\,dx\,dy\,dz = \frac{a!\,b!\,c!\,d!}{(a+b+c+d+3)!},$$

and here $1 + 9 + 8 + 4 + 3 = 25$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write the same iterated integral and both begin with the inner z-integral. They split on the first line of that integral: the natural attempt sees $(1-x-y-z)^4$ and expands it; the model sees $z^8(L-z)^4$ on $[0, L]$ — a power of the distance to the left end times a power of the distance to the right end — and rescales to $[0, 1]$.

> *The inner integrand is z^8 times $(L-z)^4$, where $L = 1 - x - y$ is the length of the interval. That is the same shape at every stage: a power of t times a power of (length minus t). If I rescale $t = Ls$, the answer is a constant times a power of L, and the power of L is the only thing that goes forward into the next integral.*

> *The answer is demanded as $a!b!c!d!/n!$. So the one-variable integral had better be a ratio of factorials too. $\int_0^1 s^p(1-s)^q\,ds$ is the one to know: $p!q!/(p+q+1)!$.*

Layer 3 THE WHY

Why is rescaling natural rather than a trick? Look at the region. At fixed x, the slice of R is the triangle $y, z \geq 0$, $y + z \leq 1 - x$ — a copy of the two-dimensional version of R, shrunk by the factor $1 - x$. At fixed x, y, the slice is the interval $[0, 1 - x - y]$, a shrunk copy of $[0, 1]$. The simplex is built out of smaller simplices, and the integrand $x^a y^b z^c w^d$ is a product of monomials in the coordinates *including* w, the distance to the far face. So every slice sees the same problem at a different scale, and homogeneity says a degree-m integrand integrated over a region scaled by L in k dimensions picks up L^{m+k}. The official solution in the MAA volume makes precisely this move in three dimensions at once: it defines $I(t)$ over the simplex $x + y + z \leq t$, observes $I(t) = I(1)t^{25}$ by the substitution $x = tu$, $y = tv$, $z = tr$, and then evaluates $\int_0^\infty I(t)e^{-t}\,dt$ in two

ways — once as $I(1)\Gamma(26) = 25!\, I(1)$, once by the change of variables $s = t - x - y - z$, which factors the four-fold integral into $\Gamma(2)\Gamma(10)\Gamma(9)\Gamma(5)$. That argument is the Gamma function's product formula; the iterated version above is the same scaling done one dimension at a time, and asks for nothing beyond integration by parts.

Why is $\int_0^1 s^p(1-s)^q\,ds = p!q!/(p+q+1)!$ "of course" and not only an identity? Take $p+q+1$ independent uniform random numbers in $[0,1]$ and single one of them out, X. The probability that a named set of p of the others falls below X and the remaining q above is $\int_0^1 s^p(1-s)^q\,ds$ (condition on $X = s$). On the other hand all $(p+q+1)!$ orderings are equally likely, and the event holds for exactly $p!\,q!$ of them (order the p below among themselves, order the q above). So the integral is $p!q!/(p+q+1)!$. The factorials in the answer are counting orderings; the demanded form $a!b!c!d!/n!$ is the problem telling you so. The triple integral itself has the same reading: with 25 uniform points in $[0,1]$ split into consecutive gaps, it is the probability of a prescribed arrangement of blocks of sizes $1, 9, 8, 4$ separated by three marked points — $1!9!8!4!$ favourable orderings out of $25!$.

Why is expanding the enemy? Because it destroys the structure you need. $(L-z)^4$ is a single factor with a meaning (a distance to the boundary); the binomial expansion turns it into five terms of alternating sign whose sum is small by cancellation. Alternating sums that cancel are a warning that you have broken something symmetric. This is zombie maths in its clearest form: the expansion is always available, always correct, and never informative.

Layer $+\alpha$ BACKTESTING

1. The cue is $w = 1 - x - y - z$ defined for you, and the answer demanded as $a!b!c!d!/n!$. Write down: "$\int_0^L t^p(L-t)^q\,dt = \frac{p!q!}{(p+q+1)!}L^{p+q+1}$".

2. Iterated integration with rescaling at each stage (a Beta integral per variable), or the Gamma-function (Dirichlet) formula directly.

3. Hesitation comes at the first inner integral, where $(1-x-y-z)^4$ looks like it needs expanding. Reinforce it away by renaming $1-x-y$ as L: the integrand becomes $z^8(L-z)^4$ and the Beta shape is visible.

4. The trap is expanding the power of w: correct, lengthy, and it hides the factorial structure the question asks for. A second trap is the answer's form: $n = 25$, not 22 $(= 1+9+8+4)$ and not 26: n is the sum of the exponents plus the number of integration variables, $22 + 3$ (w is not a fourth variable of integration).

5. Any integrand that is a product of powers of distances to the faces of a simplex (or ends of an interval), any answer requested as a ratio of factorials or Gamma values: rescale to the unit simplex and use the Beta/Dirichlet formula.

6. Reusable rule: when a region is self-similar under slicing and the integrand is homogeneous in the coordinates that describe it, integrate one variable, rescale, and carry only the power of the scale forward. Never expand a factor whose meaning is a distance to the boundary.

▷ **Cue:** $\int_0^L t^p(L-t)^q\,dt$, or $(1-x-y-z)^d$ integrated over a simplex $\longrightarrow$ **Move:** rescale $t = Ls$; Beta integral $p!q!/(p+q+1)!$; carry L^{p+q+1} forward

▷ **Cue:** an answer demanded as a ratio of factorials $\longrightarrow$ **Move:** expect Beta/Gamma integrals or a count of orderings

▷ **Cue:** a region whose slices are scaled copies of a smaller version of itself $\longrightarrow$ **Move:** homogeneity: the slice integral is a constant times a power of the scale

▷ **Cue:** the temptation to expand $(A-t)^n$ inside an integral $\longrightarrow$ **Move:** keep it as a factor and substitute so that A becomes 1

▷ **Cue:** $\int_0^1 s^p(1-s)^q\,ds$ with integer p, q $\longrightarrow$ **Move:** probability that a chosen uniform point has a named p of $p+q+1$ others below it: $p!q!/(p+q+1)!$

1984 A6

1984 A6. Let n be a positive integer, and let $f(n)$ denote the last nonzero digit in the decimal expansion of $n!$. For instance, $f(5) = 2$.

(a) Show that if $a_1, a_2, \ldots, a_k$ are distinct nonnegative integers, then $f(5^{a_1} + 5^{a_2} + \cdots + 5^{a_k})$ depends only on the sum $a_1 + a_2 + \cdots + a_k$.

(b) Assuming part (a), we can define

$$g(s) = f(5^{a_1} + 5^{a_2} + \cdots + 5^{a_k}),$$

where $s = a_1 + a_2 + \cdots + a_k$. Find the least positive integer p for which

$$g(s) = g(s + p), \qquad \text{for all } s \geq 1,$$

or else show that no such p exists.

Fields: number theory (factorials, last nonzero digit, base-5 expansions, induction) **Position:** A6 — last on the morning paper; hard until you see that the hypothesis is a base-5 numeral, then a two-case induction on one lemma **Answer:** (a) $f(5^{a_1} + \cdots + 5^{a_k}) \equiv 2^{a_1+\cdots+a_k} \pmod{10}$; (b) $g(s)$ runs $2, 4, 8, 6, 2, 4, 8, 6, \ldots$ for $s \geq 1$, so $p = 4$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. $f(1) = 1$, $f(5) = 2$ (from 120), $f(6) = 2$ (from 720), $f(25) = 4$, $f(26) = 4$, $f(30) = 8$, $f(125) = 8$, $f(130) = 6$, $f(150) = 2$. Against s: $6 = 5 + 1$ has $s = 1$, value 2; 25 and 26 have $s = 2$, value 4; $30 = 25 + 5$ and 125 have $s = 3$, value 8; $130 = 125 + 5$ has $s = 4$, value 6; $150 = 125 + 25$ has $s = 5$, value 2. The pattern $2, 4, 8, 6$ is unmistakable, and you conjecture $g(s) \equiv 2^s \pmod{10}$ and $p = 4$. That is part (b), *given* part (a) and the conjecture.

For (a) you reach for a formula for $n!$. Legendre gives the number of trailing zeros, $e = \sum_j \lfloor n/5^j \rfloor$, so $f(n)$ is $n!/10^e$ mod 10. You try to compute $n!/10^e$ mod 10 directly: strip every factor of 5 from $1, 2, \ldots, n$, strip the same number of factors of 2, multiply what is left. Mod 2 the answer is 0, so you need it mod 5, and you start pairing residues. For a general $n = 5^{a_1} + \cdots + 5^{a_k}$ this becomes a bookkeeping exercise over every level $5, 25, 125, \ldots$ at once, and there is no point at which the sum $a_1 + \cdots + a_k$ appears. The attempt stalls with a correct conjecture and no proof.

The model chain. *Notation.* For a positive integer N write $N = 10^e M$ with $10 \nmid M$, and let $L(N)$ be the last digit of M, so $f(n) = L(n!)$. Two facts used throughout:

(i) $L(10N) = L(N)$.

(ii) If $5 \nmid M$ and $5 \nmid K$, then $10 \nmid MK$, so $L(NK) \equiv L(N)\,K \pmod{10}$.

For $n \geq 2$, $n!$ contains more factors 2 than 5, so M is even and $f(n) \in \{2, 4, 6, 8\}$; in particular $5 \nmid M$. For $n = 0, 1$, $f(n) = 1$.

Lemma. For $m \geq 1$, $f(5m) \equiv 2^m f(m) \pmod{10}$.

Proof. Split $1, \ldots, 5m$ into multiples of 5 and the rest:

$$(5m)! = \underbrace{(5 \cdot 10 \cdots 5m)}_{= 5^m\, m!} \prod_{i=0}^{m-1} B_i, \qquad B_i = (5i+1)(5i+2)(5i+3)(5i+4).$$

Each B_i contains two consecutive even numbers, one of them divisible by 4, so $8 \mid B_i$; and $B_i \equiv 1 \cdot 2 \cdot 3 \cdot 4 = 24 \equiv 4 \pmod{5}$. So $B_i \equiv 0 \pmod{4}$ and $B_i \equiv 4 \pmod{5}$, which forces $B_i \equiv 4 \pmod{20}$. Write $B_i = 2c_i$; then $c_i \equiv 2 \pmod{10}$ and $5 \nmid c_i$. Now

$$(5m)! = 5^m\, m!\, 2^m \prod c_i = 10^m \cdot m! \cdot \prod c_i.$$

By (i) and (ii) (with $K = \prod c_i$, coprime to 5), $f(5m) \equiv f(m) \prod c_i \equiv 2^m f(m) \pmod{10}$. □

Two facts about 2^u *mod* 10. $6 \cdot x \equiv x \pmod{10}$ for every even x (since $5x \equiv 0$). As $2^4 = 16$, it follows that $2^{u+4} \equiv 2^u$ for $u \geq 1$; so for $u, v \geq 1$, $u \equiv v \pmod 4$ implies $2^u \equiv 2^v \pmod{10}$.

Claim. If $a_1, \ldots, a_k$ are distinct nonnegative integers and $N = \sum 5^{a_i}$, then $f(N) \equiv 2^s \pmod{10}$, where $s = \sum a_i$. This proves (a).

Proof, by strong induction on N. If $N = 1$ then $k = 1$, $a_1 = 0$, $s = 0$, and $f(1) = 1 = 2^0$.

Case 1: every $a_i \geq 1$. Then $N = 5m$ with $m = \sum 5^{a_i - 1}$, a sum of distinct powers of 5 with exponent sum $s - k$, and $m < N$. By the lemma and induction, $f(N) \equiv 2^m \cdot 2^{s-k}$. Since $5 \equiv 1 \pmod 4$, $m \equiv k \pmod 4$, and $m, k \geq 1$, so $2^m \equiv 2^k$ and $f(N) \equiv 2^s$.

Case 2: some $a_i = 0$. By distinctness exactly one, so $N = 5m + 1$ where $5m$ is the sum of the other powers, whose exponent sum is still s. If $m = 0$ we are in the base case. Otherwise $N! = N \cdot (5m)!$; by (ii), $f(N) \equiv N f(5m)$; and $N f(5m) - f(5m) = 5m\, f(5m) \equiv 0 \pmod{10}$ because $f(5m)$ is even. So $f(N) = f(5m) \equiv 2^s$ by Case 1. □

Part (b). For $s \geq 1$, $g(s) \equiv 2^s$: $g(1), g(2), g(3), g(4) = 2, 4, 8, 6$. The period $p = 4$ works by the fact above. No smaller p works: $p = 1$ fails at $g(2) \neq g(1)$, $p = 2$ at $g(3) \neq g(1)$, $p = 3$ at $g(4) \neq g(1)$. So $p = 4$. (The condition $s \geq 1$ matters: $g(0) = f(1) = 1$, which is not $g(4) = 6$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the data and the conjecture $g(s) \equiv 2^s$. They split at how to use the hypothesis. The natural attempt treats $5^{a_1} + \cdots + 5^{a_k}$ as "some number" and attacks $n!$ for general n. The model reads it as a base-5 numeral whose digits are all 0 or 1, and asks what happens to f when you append a digit: $m \mapsto 5m$ (append 0) and $5m \mapsto 5m + 1$ (change the last 0 to 1).

> *A sum of distinct powers of* 5 *is a number whose base-*5 *digits are only* 0*s and* 1*s. So I should induct on the base-*5 *digits: I need* $f(5m)$ *in terms of* $f(m)$*, and* $f(5m + 1)$ *in terms of* $f(5m)$*.*

> *In* $(5m)!$ *the multiples of* 5 *are* $5 \cdot 1, 5 \cdot 2, \ldots, 5 \cdot m$ *— that is* $5^m\, m!$*, the smaller factorial reappearing. Each of those* m *fives wants a* 2 *to become a* 10*, and there are exactly* m *blocks of four non-multiples, each with plenty of* 2*s. Take one* 2 *from each block and see what is left mod* 10*.*

Layer 3 THE WHY

Why base 5? Go back to what "last nonzero digit" means. Trailing zeros are factors of $10 = 2 \cdot 5$, and in $n!$ the 5s are scarcer, so the 5s decide everything. Where are the 5s? In the multiples of 5 — and the multiples of 5 up to $5m$ are $5 \cdot 1, \ldots, 5 \cdot m$, whose product is $5^m\, m!$. So $n!$ contains a copy of $\lfloor n/5 \rfloor!$, and the question about $n!$ becomes a question about $\lfloor n/5 \rfloor!$ plus a local correction. Passing from n to $\lfloor n/5 \rfloor$ is deleting the last base-5 digit. That is why the problem's hypothesis is phrased in powers of 5: it is handing you a number already written in the only base in which the recursion is visible. Once you see that, induction on digits is not a choice; it is the definition of the object you were given.

Why does each block contribute exactly a factor 2 mod 10? Two facts at axiom level. Mod 5, the block $(5i + 1)(5i + 2)(5i + 3)(5i + 4)$ is $1 \cdot 2 \cdot 3 \cdot 4 = -1$: the product of all nonzero residues mod a prime is -1 (Wilson's theorem — every residue pairs with its inverse except 1 and -1). Mod 4, the block is 0. So the block is $4 \pmod{20}$, and halving it — which is exactly the act of lending a 2 to one 5 — gives $2 \pmod{10}$. Nothing about i survives, which is why the lemma has no case split on i. The official solution in the MAA volume states "$\frac{1}{2}(5i + 1)(5i + 2)(5i + 3)(5i + 4) \equiv 2$", checks i even and i odd numerically, and moves on; the mod-20 argument is where that line comes from, and it is what makes the halving legitimate (reading a number mod 10 after dividing by 2 requires knowing it mod 20).

Why does the sum of exponents appear? Each time you peel a digit 0 you pick up 2^m, and 2^m mod 10 depends only on m mod 4. Since $5 \equiv 1 \pmod 4$, $m = \sum 5^{a_i - 1} \equiv k$, the number of

1-digits. So peeling the last digit off a numeral with k ones contributes 2^k; peeling repeatedly, the digit 1 in position a is counted once for each of the a peels that pass over it, contributing 2^a in total. That is the whole reason the answer is $2^{\sum a_i}$: each 1-digit collects a factor 2 per level it sits above the units place. And the digit 1 in the units place costs nothing, because multiplying an even last digit by a number $\equiv 1 \pmod 5$ changes nothing mod 10 — which is also why the problem allows only digits 0 and 1. (With general digits $r \in \{0, \ldots, 4\}$ the same argument gives $f(5m + r) \equiv 2^m f(m)\, r! \pmod{10}$ for $m \geq 1$, and the digit factorials $r!$ spoil the clean dependence on the exponent sum; I have checked this general recursion numerically for $n < 3000$, and it follows from the lemma and the evenness step exactly as in Case 2.)

Why $p = 4$ and why "$s \geq 1$"? The multiplicative order of 2 modulo 5 is 4, and the factor of 2 in 10 is absorbed because every $g(s)$ with $s \geq 1$ is even. At $s = 0$ the value is the odd number 1, outside the cycle — the problem's "for all $s \geq 1$" is fencing off exactly that.

Layer $+\alpha$ BACKTESTING

1. The cue is "$5^{a_1} + \cdots + 5^{a_k}$ with distinct a_i". Write down: this is a base-5 numeral with digits in $\{0, 1\}$; I will induct on digits, so I need $f(5m)$ versus $f(m)$.

2. Recursion $n! \supset 5^{\lfloor n/5 \rfloor} \lfloor n/5 \rfloor !$, a lemma $f(5m) \equiv 2^m f(m)$, and strong induction on the numeral.

3. Hesitation comes at $(5m)!/10^m$: which 2s do you remove, and is the leftover determined mod 10? Reinforce it away by removing one 2 per block of four, and computing the block mod 20 (mod 4 and mod 5 separately).

4. Two traps. Proving (b) from the numerical pattern without (a) — the problem lets you assume (a) but you still need the formula $g(s) \equiv 2^s$, which the proof of (a) supplies. And forgetting that "$s \geq 1$": $g(0) = 1$ breaks the cycle. A smaller trap is dividing a congruence mod 10 by 2, which is illegitimate without working mod 20.

5. Any question about trailing zeros or the last nonzero digit of $n!$, any sum of distinct powers of a prime p: pass from n to $\lfloor n/p \rfloor$, i.e. induct on base-p digits, because the multiples of p in $n!$ form $p^{\lfloor n/p \rfloor} \lfloor n/p \rfloor !$.

6. Reusable rule: when an object contains a scaled copy of itself, find the recursion that peels off the copy and read the hypothesis as a description of the recursion's input (here, a base-5 numeral). To read a quotient mod 10, compute the dividend modulo a larger modulus (mod 20 to halve), and use CRT.

▷ **Cue:** a sum of distinct powers of b (or "n written with digits 0 and 1") $\longrightarrow$ **Move:** it is a base-b numeral; induct by deleting the last digit

▷ **Cue:** trailing zeros or last nonzero digit of $n!$ $\longrightarrow$ **Move:** split off the multiples of 5: $(5m)! = 5^m\, m! \prod(\text{blocks of four})$

▷ **Cue:** a product of a full set of nonzero residues mod a prime p $\longrightarrow$ **Move:** Wilson: it is $-1 \pmod p$

▷ **Cue:** you need $x/2$ mod 10 (or x/d mod m) $\longrightarrow$ **Move:** compute x mod 20 (mod dm), via CRT

▷ **Cue:** an even number multiplied by something $\equiv 1 \pmod 5$ $\longrightarrow$ **Move:** the last digit is unchanged: $x(5t + 1) \equiv x \pmod{10}$ for even x

▷ **Cue:** periodicity of 2^s mod 10 $\longrightarrow$ **Move:** order of 2 mod 5 is 4; valid only for $s \geq 1$ because of the factor 2

Putnam 1984 — Paper B

1984 B1

1984 B1. Let n be a positive integer, and define

$$f(n) = 1! + 2! + \cdots + n!.$$

Find polynomials $P(x)$ and $Q(x)$ such that

$$f(n+2) = P(n)f(n+1) + Q(n)f(n)$$

for all $n \geq 1$.

Fields: algebra (factorial sums, linear recurrences) **Position:** B1 – opening problem of the second session **Answer:** $P(x) = x + 3$, $Q(x) = -(x+2)$, i.e. $f(n+2) = (n+3)f(n+1) - (n+2)f(n)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute a few values to get a feel for the sequence: $f(1) = 1$, $f(2) = 3$, $f(3) = 9$, $f(4) = 33$, $f(5) = 153$. The problem asks for polynomials P, Q, so you guess they are low degree – linear, say – and try to pin down $P(x) = ax + b$, $Q(x) = cx + d$ by substitution. From $n = 1$: $f(3) = P(1)f(2) + Q(1)f(1)$, i.e. $9 = 3(a+b) + (c+d)$. From $n = 2$: $f(4) = P(2)f(3) + Q(2)f(2)$, i.e. $33 = 9(2a+b) + 3(2c+d)$. From $n = 3$: $153 = 33(3a+b) + 9(3c+d)$. That is three equations in four unknowns a, b, c, d – you need a fourth data point from $n = 4$ to even solve for the coefficients, and once you have solved the linear system you still only know that P, Q *fit these four values of* n. Nothing in the arithmetic tells you the fit continues at $n = 5, 6, 7, \ldots$; that would need a separate induction on top of the curve-fit. You have not used the definition of f as a sum of factorials at all – you have only used its numerical output – so the proof, if you finish it, is doing much more work than the problem requires.

The model chain. Do not treat $f(n+2)$, $f(n+1)$, $f(n)$ as three independent numbers to be matched by a template. f is defined as a running sum, so consecutive values differ by exactly the next factorial:

$$f(n+1) - f(n) = (n+1)!, \qquad f(n+2) - f(n+1) = (n+2)!.$$

Now use the one fact about factorials that lets you go from $(n+2)!$ to $(n+1)!$: $(n+2)! = (n+2) \cdot (n+1)!$. Substitute the difference expression for $(n+1)!$:

$$f(n+2) - f(n+1) = (n+2)! = (n+2)(n+1)! = (n+2)\big(f(n+1) - f(n)\big).$$

Expand and solve for $f(n+2)$:

$$f(n+2) = f(n+1) + (n+2)f(n+1) - (n+2)f(n) = (n+3)f(n+1) - (n+2)f(n).$$

This is an identity for every $n \geq 1$, derived directly from the definitions of f and of the factorial – no induction is needed because both ingredient facts ($f(n+1) - f(n) = (n+1)!$ and

$(n+2)! = (n+2)(n+1)!$) are already true for every n individually. Hence $P(x) = x+3$ and $Q(x) = -(x+2)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start by noticing f grows through factorials. They split at how they treat the target identity $f(n+2) = P(n)f(n+1) + Q(n)f(n)$. The natural attempt reads it as a template to be matched by plugging in numbers and solving for unknown coefficients – an interpolation problem. The model reads it as an algebraic consequence to be derived by working one level down, with the differences $f(n+1) - f(n)$ and $f(n+2) - f(n+1)$, which are exactly the factorial terms that define f in the first place.

> *$f(n+2)$, $f(n+1)$, and $f(n)$ are not free unknowns to be matched against a guessed template – each is a partial sum of factorials, so their pairwise differences are just the next factorial term. I should work with those differences, not the raw values.*

> *Since $(n+2)! = (n+2)\cdot(n+1)!$, and $(n+1)!$ is itself the difference $f(n+1) - f(n)$, I can rewrite the jump from $f(n+1)$ to $f(n+2)$ entirely in terms of $f(n+1)$ and $f(n)$ – which is exactly the shape $P(n)f(n+1) + Q(n)f(n)$ the problem is asking for, so I should build the identity by substitution rather than solve for coefficients numerically.*

Layer 3 THE WHY

Why is curve-fitting the wrong tool here, and not merely a slower one? Because finitely many numerical values pin down a polynomial only once its degree is already known or bounded, and nothing in the problem statement guarantees P, Q are linear – you would need at least $\deg P + \deg Q + 2$ data points to even determine the coefficients, and then a further argument (an induction) that the fitted formula, having been checked at finitely many n, holds for every n. Curve-fitting supplies a candidate, not a proof; the "find polynomials P, Q such that … for all $n \geq 1$" phrasing is a universally quantified claim, and a universally quantified claim about a sequence defined by a recursive/summed construction is answered most directly by working algebraically inside that construction, not by sampling its outputs.

The deeper reason the substitution route works so cleanly is that f is a partial sum, $f(n) = \sum_{k=1}^{n} k!$, and the summand $k!$ obeys its own multiplicative recursion, $(k+1)! = (k+1)\cdot k!$. Any time a sequence is defined as $s_n = \sum a_k$, the identity $s_{n+1} - s_n = a_{n+1}$ is automatic from the definition of a partial sum – it costs nothing to write down. If the summand a_n itself satisfies a recursion linking a_{n+1} to a_n (here, the factorial recursion), that recursion transports onto s by direct substitution of the difference identity, and produces a linear recursion for s with coefficients that are polynomials in n (here, degree 1, because the factorial recursion has a coefficient $(n+2)$ linear in n). This is the general mechanism – sometimes called holonomic transport – behind almost every "sum of a recursively-defined sequence satisfies a nice recursion" problem: the recursion is inherited from the summand, term by term, and the inheritance is a substitution, never a search.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that f is defined as $1! + 2! + \cdots + n!$, a partial sum. On seeing it, write down immediately: $f(n+1) - f(n) = (n+1)!$.

2. It suggests the difference/telescoping method: express the recursion you want for f by working one level down, in terms of the summand's own recursion, rather than treating $f(n), f(n+1), f(n+2)$ as free values.

3. Hesitation occurs right after writing $f(n+2) - f(n+1) = (n+2)!$, when it is unclear how to bring $f(n)$ into the picture. It is reinforced away by recalling $(n+2)! = (n+2)(n+1)!$ and that $(n+1)!$ is itself $f(n+1) - f(n)$.

4. The trap is treating "find polynomials P, Q" as free license to guess a low degree and solve a numerical system from small cases; this produces a formula that fits the sampled data but is not proved for all n without a further induction.

5. Any time a sequence is defined as a partial sum (or product) and a problem asks for a linear recursion linking several consecutive terms, write the recursion for the summand (or factor) first, then substitute the difference (or ratio) identity to eliminate it.

6. Reusable rule: to find a recursion for a partial sum $s_n = \sum a_k$, do not fit coefficients numerically. Use $s_{n+1} - s_n = a_{n+1}$ together with any known recursion for a_n itself, and substitute; the coefficients of the resulting recursion for s fall out as polynomials in n automatically, with no separate induction required.

▷ **Cue:** a sequence is defined as $f(n) =$ a sum up to n, and you are asked for a recursion linking $f(n+2), f(n+1), f(n)$ $\longrightarrow$ **Move:** write $f(n+1) - f(n) =$ (the summand at $n+1$), and use the summand's own recursion to eliminate it in favour of $f(n+1) - f(n)$ and $f(n) - f(n-1)$

▷ **Cue:** $(n+2)!$ appears next to $(n+1)!$ $\longrightarrow$ **Move:** use $(n+2)! = (n+2)(n+1)!$ to convert between consecutive factorials

▷ **Cue:** asked to "find polynomials $P(x), Q(x)$" satisfying an identity for all n $\longrightarrow$ **Move:** derive the identity algebraically from the definitions in play; do not fit coefficients from finitely many numerical values, since that only produces a candidate, not a proof for all n

1984 B2

1984 B2. Find the minimum value of

$$(u-v)^2 + \left(\sqrt{2-u^2} - \frac{9}{v}\right)^2$$

for $0 < u < \sqrt{2}$ and $v > 0$.

Fields: algebra/geometry (optimisation via a geometric reformulation; AM–GM, Cauchy–Schwarz) **Position:** B2 – second problem of the second session **Answer:** The minimum value is 8, attained at $u = v =$ no single value – rather at $u = 1$, $v = 3$ (giving points $(1,1)$ on the circle and $(3,3)$ on the hyperbola).

Caveat. The book's OCR of this problem is badly garbled (symbols dropped, "$\sqrt{2}$" rendered as "fi", the fraction unreadable). The statement above is reconstructed from the book's own solution text, which explicitly describes the problem as finding the minimum squared distance between the arc of $x^2 + y^2 = 2$ in the open first quadrant and the branch of $xy = 9$ in that quadrant, and states the answer is 8 – this pins the expression down uniquely, since $(u, \sqrt{2-u^2})$ for $0 < u < \sqrt{2}$ parametrises exactly that circular arc and $(v, 9/v)$ for $v > 0$ parametrises exactly that hyperbola branch, and the squared distance between them is precisely the displayed expression. The reconstruction has been verified independently below, both by a direct calculus attempt and by a self-contained non-calculus proof, and both agree with the value 8 that the archive's own (otherwise unreadable) solution states.

Layer 1 THE WHAT — surface analysis

The natural attempt. The expression is a function of two independent variables u, v, so the reflex is calculus: set both partial derivatives to zero. Write $D(u,v) = (u-v)^2 + (\sqrt{2-u^2} - 9/v)^2$. Then

$$\frac{\partial D}{\partial u} = 2(u-v) - \frac{2u}{\sqrt{2-u^2}}\left(\sqrt{2-u^2} - \frac{9}{v}\right) = 0,$$

$$\frac{\partial D}{\partial v} = -2(u-v) + \frac{18}{v^2}\left(\sqrt{2-u^2} - \frac{9}{v}\right) = 0.$$

Adding the two equations, the $(u-v)$ terms cancel, leaving

$$\left(\sqrt{2-u^2} - \frac{9}{v}\right)\left(\frac{9}{v^2} - \frac{u}{\sqrt{2-u^2}}\right) = 0.$$

If the first factor vanishes, $\sqrt{2-u^2} = 9/v$; substituting back into the original $\partial D/\partial u = 0$ forces $u = v$ as well, and then $u = v$ together with $\sqrt{2-u^2} = 9/u$ has no solution in the required range (the left side is at most $\sqrt{2}$, the right side is at least $9/\sqrt{2} \approx 6.4$ for $u < \sqrt{2}$) – so this branch is empty. That forces the second factor to vanish: $9/v^2 = u/\sqrt{2-u^2}$, a genuinely transcendental-looking relation between u and v once combined with $\partial D/\partial u = 0$ itself. You now have two nonlinear equations in u, v involving a square root and a reciprocal, with no obvious substitution – solving it is possible in principle but this is exactly the kind of brute-force manipulation ("zombie maths") that the problem does not reward: it costs many lines and several sign cases to grind through, purely because the calculus route never looked at what the expression *means*.

The model chain. Read $(u, \sqrt{2-u^2})$ as a point on the arc of the circle $x^2 + y^2 = 2$ lying in the open first quadrant (as u ranges over $(0, \sqrt{2})$, this traces exactly that arc), and $(v, 9/v)$ as a point on the branch of the hyperbola $xy = 9$ in the first quadrant (as v ranges over $(0, \infty)$). The given expression is then exactly the squared Euclidean distance between a point on the circle and a point on the hyperbola, and the problem is: minimise the squared distance between these two curves.

Bound each curve's points using the diagonal direction $x + y$. For a point (x, y) on the circle, Cauchy–Schwarz gives $x + y \le \sqrt{2(x^2+y^2)} = \sqrt{2 \cdot 2} = 2$, with equality iff $x = y$, i.e. at $(1,1)$. For a point (x, y) on the hyperbola with $x, y > 0$, AM–GM gives $x + y \ge 2\sqrt{xy} = 2\sqrt{9} = 6$, with equality iff $x = y$, i.e. at $(3,3)$.

So every point $P = (u, \sqrt{2-u^2})$ on the circular arc has $x_P + y_P \le 2$, and every point $Q = (v, 9/v)$ on the hyperbola branch has $x_Q + y_Q \ge 6$. Project the vector $\overrightarrow{PQ}$ onto the unit vector $\frac{1}{\sqrt{2}}(1,1)$: this projection has length at most $|PQ|$ (a projection never exceeds the vector's own length), and it equals

$$\frac{(x_Q - x_P) + (y_Q - y_P)}{\sqrt{2}} = \frac{(x_Q + y_Q) - (x_P + y_P)}{\sqrt{2}} \ge \frac{6-2}{\sqrt{2}} = 2\sqrt{2}.$$

Hence $|PQ| \ge 2\sqrt{2}$ for every choice of P on the arc and Q on the branch, so the squared distance is at least 8. Equality in both the Cauchy–Schwarz and AM–GM steps forces $P = (1,1)$ and $Q = (3,3)$, and equality in the projection bound forces $\overrightarrow{PQ}$ to be parallel to $(1,1)$, which it is: $\overrightarrow{PQ} = (2,2)$. So the bound 8 is attained exactly at $u = 1$, $v = 3$, and the minimum value is $\boxed{8}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths treat $D(u, v)$ as a function to be minimised over two variables. They split at the very first move. The natural attempt takes "minimise a function of u, v" at face value and reaches for partial derivatives, without pausing to ask what D *is*. The model chain first asks what geometric object $(u, \sqrt{2-u^2})$ and $(v, 9/v)$ describe, recognises D as a squared distance between two named curves, and only then chooses a method – one suited to bounding a distance between curves, not one suited to an arbitrary two-variable function.

> *Before differentiating anything, I should ask what $(u, \sqrt{2-u^2})$ and $(v, 9/v)$ actually parametrise. The first satisfies $x^2 + y^2 = 2$ and the second satisfies $xy = 9$; the whole expression is a squared*

> *distance between a point on one curve and a point on the other, so this is a distance-between-curves problem in disguise, not a bare calculus-in-two-variables problem.*

> *To bound a distance between two curves without calculus, I want a single linear functional (a projection) that is small on one curve and large on the other; here $x + y$ is that functional, because Cauchy–Schwarz bounds it above on the circle and AM–GM bounds it below on the hyperbola, and both bounds are tight at points related by the symmetry $x = y$ that both curves share.*

Layer 3 THE WHY

Why does reaching for partial derivatives fail to be the efficient move here, even though it is not wrong? Because $D(u, v)$ is not an arbitrary smooth function – it carries geometric content (a squared distance between two specific, highly symmetric curves) that a blind calculus attack discards. Setting $\partial D/\partial u = \partial D/\partial v = 0$ recovers only the *local* first-order condition for a critical point; it says nothing about which curves are involved or why the critical point should be easy to find, so the resulting system mixes a square root and a reciprocal with no structural handle. The lesson generalises: whenever an expression $(u - v)^2 + (g(u) - h(v))^2$ appears, first ask whether $(u, g(u))$ and $(v, h(v))$ are two named curves – if so, the problem is distance-between-curves, and distance-between-curves problems are almost always solved geometrically (via a separating line, a common normal, or a monotone functional), because a distance is a global, coordinate-free quantity and calculus produces only local information.

Why is $x + y$ exactly the right functional to bound, and not, say, x or xy or $x^2 + y^2$? Because both curves are symmetric under swapping $x \leftrightarrow y$ (the circle $x^2 + y^2 = 2$ and the hyperbola $xy = 9$ are both invariant under reflection across the line $y = x$), and $x + y$ is precisely the linear functional whose level sets are the lines perpendicular to that axis of symmetry. A functional that respects the shared symmetry of both curves is the one whose extreme values on each curve occur exactly on the symmetry axis itself ($x = y$) – that is why Cauchy–Schwarz's equality case $x = y$ on the circle and AM–GM's equality case $x = y$ on the hyperbola coincide, and why the nearest points $(1, 1)$ and $(3, 3)$ both lie on $y = x$. Put differently: the axis of symmetry common to two curves is always a natural place to look for the closest pair of points between them, because the tangent lines there are forced (by symmetry) to be perpendicular to the axis, which is exactly the geometric condition (parallel tangents, common normal) that characterises a nearest pair on convex curves. The archive's own solution states this via the tangent-line version of the same fact ("the tangents at $(1, 1)$ and $(3, 3)$ separate the curves and are both perpendicular to $x = y$"); the Cauchy–Schwarz/AM–GM argument above is a self-contained, calculus-free way of proving the same separation and is fully verified without appeal to that tangent-line claim.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape $(u - v)^2 + (g(u) - h(v))^2$. On seeing it, write down: this is the squared distance between the point $(u, g(u))$ and the point $(v, h(v))$ – identify the two curves before doing anything else.
2. It suggests the method of bounding a distance between two curves geometrically: find a linear functional (or a pair of parallel supporting lines) that separates the curves, using whatever symmetry the curves share.
3. Hesitation occurs right after setting up the partial derivatives and reaching the messy simultaneous equations; it is reinforced away by stepping back and asking what curves $u \mapsto (u, \sqrt{2 - u^2})$ and $v \mapsto (v, 9/v)$ trace, which reframes the whole problem.
4. The trap is committing to calculus immediately because two independent variables and a minimisation are present; the messy system that results is solvable but wastes time that a geometric read-off avoids entirely.

5. Any time an expression has the shape (difference of first coordinates)^2 + (difference of second coordinates)^2, recognise it as a squared distance between two parametrised curves, and look for a symmetry the two curves share before differentiating.

6. Reusable rule: to minimise the distance between two curves, look for their common symmetry axis; bound a linear functional respecting that symmetry above on one curve and below on the other (via Cauchy–Schwarz, AM–GM, or a supporting line), and check that both bounds are tight at symmetric points.

▷ **Cue:** an expression of the form $(u-v)^2+(g(u)-h(v))^2 \longrightarrow$ **Move:** recognise it as the squared distance between the point $(u, g(u))$ and the point $(v, h(v))$; identify the two curves these parametrise before choosing a method

▷ **Cue:** two curves both symmetric under $x \leftrightarrow y$ (e.g. a circle centred at the origin and a hyperbola $xy = \text{const}$) $\longrightarrow$ **Move:** look for the nearest pair of points on the line $y = x$, and bound the functional $x + y$ above on one curve and below on the other via Cauchy–Schwarz and AM–GM respectively

▷ **Cue:** minimising a two-variable expression that turns out to be geometric $\longrightarrow$ **Move:** before differentiating, ask what curves the coordinates parametrise; a distance-between-curves problem is solved by separation/symmetry arguments, not by blind partial derivatives

1984 B3

1984 B3. Prove or disprove the following statement: If F is a finite set with two or more elements, then there exists a binary operation $*$ on F such that for all x, y, z in F, (i) $x * z = y * z$ implies $x = y$ (right cancellation holds), and (ii) $x * (y * z) \neq (x * y) * z$ (no case of associativity holds).

Fields: abstract algebra (binary operations, bijections/permutations) **Position:** B3 – third problem of the second session **Answer:** The statement is true: for any finite F with $|F| \geq 2$, define $x * y = \varphi(x)$ for a fixed-point-free bijection $\varphi : F \to F$.

Layer 1 THE WHAT — surface analysis

The natural attempt. "Prove or disprove" with a two-part condition on a finite set naturally invites you to test the smallest case first: $F = \{a, b\}$. You start building a Cayley table. Right cancellation with two elements means: for each fixed z, the column $x \mapsto x * z$ must be injective, i.e. a permutation of $\{a, b\}$ – so each of the two columns of the table is either the identity permutation or the swap. That already narrows the table to 4 possibilities. Now you check each for non-associativity on all $2^3 = 8$ triples (x, y, z). This is doable by hand for $|F| = 2$, and with some trial and error you find a table that works. But you have only answered the problem for $F = \{a, b\}$: nothing about a 2×2 Cayley table search generalises to F with, say, 17 elements, where the number of Cayley tables satisfying the column-permutation condition is astronomically larger and hand-checking every triple is impossible. The "prove or disprove" is a claim about *every* finite F with ≥ 2 elements, and a case-by-case search, however successful at $n = 2$, is not a proof of the general statement – it would need to be redone, or generalised, for every size.

The model chain. Look for a single formula for $*$ that works uniformly for every finite F with $|F| \geq 2$, rather than searching a table for one size. Condition (i) says: for every fixed z, the map $x \mapsto x * z$ is injective. The cheapest way to guarantee an injective map "for every z" is to make the map not depend on z at all. So define

$$x * y := \varphi(x) \quad \text{for all } x, y \in F,$$

where $\varphi : F \to F$ is a bijection to be chosen. Condition (i) is then immediate: $x * z = y * z$ means $\varphi(x) = \varphi(y)$, and φ is injective (being a bijection), so $x = y$. This holds for every z simultaneously, since $*$ never looks at its second argument.

For condition (ii), compute both sides of the associativity comparison under this definition:

$$x * (y * z) = \varphi(x), \qquad (x * y) * z = \varphi(x * y) = \varphi(\varphi(x)).$$

(The first equality holds because $*$ applied to *any* pair returns φ of the first argument; the second because $x * y = \varphi(x)$ is itself the "first argument" fed into the outer $*$.) So condition (ii) becomes: $\varphi(x) \neq \varphi(\varphi(x))$ for every $x \in F$. Since φ is injective, $\varphi(x) = \varphi(\varphi(x))$ is equivalent to $x = \varphi(x)$; so condition (ii) holds for every x exactly when φ has no fixed point.

A fixed-point-free bijection (a derangement) of a finite set with $|F| \geq 2$ always exists: list the elements $a_1, \ldots, a_n$ ($n \geq 2$) and let φ be the single n-cycle $\varphi(a_i) = a_{i+1}$ (indices mod n); no element maps to itself. Hence $x * y = \varphi(x)$ satisfies both (i) and (ii) for every finite F with $|F| \geq 2$, and the statement is **true**.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths recognise that condition (i) is an injectivity requirement and condition (ii) is a non-associativity requirement holding everywhere. They split at how to search for an operation satisfying both. The natural attempt fixes a specific small F and tries to fill in a full Cayley table by trial, checking (i) and (ii) after the fact on that table. The model chain instead restricts the *form* of the operation before checking anything – it lets $x * y$ ignore y entirely – so that condition (i) is satisfied automatically by any bijection, and condition (ii) collapses from a three-variable check into a one-variable check on that bijection's fixed points.

> *If I make $x * y$ depend only on x, then condition (i) – for every fixed z, the map $x \mapsto x * z$ is injective – stops depending on z at all: it becomes a single injectivity requirement on one function, and any bijection satisfies it for free, regardless of which z is used.*

> *For non-associativity everywhere, I need $\varphi(x) \neq \varphi(\varphi(x))$ for every x; since φ is injective, this is the same as needing $x \neq \varphi(x)$ for every x, i.e. I want φ to be a permutation of F with no fixed point – and such a permutation always exists on a finite set with two or more elements (a single cycle through all of them, for instance).*

Layer 3 THE WHY

Why is "make $*$ ignore one argument" the natural move, and not a trick pulled from nowhere? Because condition (i) is quantified "for all z", and it asks that a certain map – one map per value of z – be injective. The cheapest way to satisfy a condition quantified over a parameter is to remove the object being checked from depending on that parameter: if there is only one map to check (because $*$ doesn't look at z), then a single choice of bijection settles the "for all z" at once, instead of needing to verify injectivity separately for every z. This is a general principle, not specific to this problem: whenever a construction must satisfy a property "for all" some variable, first ask whether the object can be built so that it simply does not depend on that variable, collapsing the universal quantifier into a single instance.

Why does this same restriction make condition (ii) easy rather than merely convenient? Because once $*$ ignores its second argument, the two sides of the associativity comparison, $x * (y * z)$ and $(x * y) * z$, stop being genuinely three-variable expressions: y and z contribute nothing to either side except by being fed, indirectly, back into φ. Concretely, $x * (y * z)$ applies φ to x exactly once (since $*$ always returns φ of its first argument, and here that first argument is x itself), while $(x * y) * z$ applies φ to x exactly *twice* (because $x * y = \varphi(x)$ becomes the new first argument fed into $*$ a second time). So the entire "for all x, y, z" condition (ii) reduces to a

one-variable question about φ and its iterate $\varphi \circ \varphi$: does $\varphi(x)$ ever equal $\varphi^2(x)$? This is exactly the fixed-point condition on φ, a fact about permutations of a finite set that is elementary once the reduction is seen – and every finite set of size ≥ 2 admits a permutation moving every point (an n-cycle is the simplest example). The problem's apparent difficulty – constructing an operation that simultaneously satisfies a positive universal condition and a negative universal condition – dissolves once the operation is deliberately made "thin" (dependent on as little as possible), because a thin operation turns both universal conditions on the operation into simple, checkable conditions on a single auxiliary function.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that condition (i) is quantified "for all z" and already looks like an injectivity statement. On seeing it, write down: "(i) says $x \mapsto x * z$ is injective, for every fixed z."

2. It suggests trying to make the object being checked (the map $x \mapsto x * z$) independent of the quantified variable z, i.e. defining $*$ so it ignores its second argument.

3. Hesitation occurs when checking condition (ii) after making that restriction: it is not immediately obvious how removing the y-dependence affects the associativity comparison. It is reinforced away by writing out both sides, $x * (y * z)$ and $(x * y) * z$, symbolically under the restricted definition and seeing they only differ in how many times φ has been applied.

4. The trap is treating the problem as "find one example for one F" via Cayley-table trial and error, which answers a specific size but not the general "for every finite F with ≥ 2 elements" claim the problem actually makes.

5. Any time a construction must satisfy a condition "for all" some parameter, ask whether the object can be defined so that it simply does not depend on that parameter – this turns a universally quantified condition into a single check.

6. Reusable rule: to build an algebraic structure satisfying both a positive universal condition (like a cancellation/injectivity law) and a negative universal condition (like failure of associativity), try defining the operation to depend on as few of its arguments as possible; this typically reduces both conditions to elementary properties (bijectivity, absence of fixed points) of a single auxiliary function, which are then supplied by standard finite combinatorics (e.g. an n-cycle).

▷ **Cue:** "prove or disprove: there exists an operation on a finite set F with property P holding for all z (or all triples)" $\longrightarrow$ **Move:** try defining the operation to ignore one or more of its arguments, so that the "for all" condition on the ignored variable collapses into a single condition on the resulting function

▷ **Cue:** right cancellation: $x * z = y * z \implies x = y$ $\longrightarrow$ **Move:** this says $x \mapsto x * z$ is injective; if $*$ does not depend on its second argument, this is exactly the statement that a single fixed function is a bijection

▷ **Cue:** want $x * (y * z) \neq (x * y) * z$ for all x, y, z, with $*$ depending only on the first argument $\longrightarrow$ **Move:** compute both sides as iterates of the underlying function ($\varphi(x)$ versus $\varphi(\varphi(x))$), and require that function to have no fixed points

▷ **Cue:** need a bijection on a finite set with ≥ 2 elements that has no fixed point (a derangement) $\longrightarrow$ **Move:** use a single cycle through all the elements of the set

1984 B4

1984 B4. Find, with proof, all real-valued functions $y = g(x)$ defined and continuous on $[0, \infty)$, positive on $(0, \infty)$, such that for all $x > 0$ the y-coordinate of the centroid of the region

$$R_x = \{(s, t) \mid 0 \leq s \leq x,\ 0 \leq t \leq g(s)\}$$

is the same as the average value of g on $[0, x]$.

Fields: real analysis; integral equations; first-order ODEs; centroids **Position:** B4 — middle of the paper; a functional equation in integrals that becomes a homogeneous ODE once the integral you cannot control is eliminated **Answer:** $g(x) = Cx^{1+\sqrt{2}}$ with $C > 0$, and nothing else. The condition is $x\int_0^x g^2 = 2\left(\int_0^x g\right)^2$; with $G = \int_0^x g$ it forces $xg/G \in \{2+\sqrt{2},\, 2-\sqrt{2}\}$, continuity makes the ratio constant, and the root $2-\sqrt{2}$ gives $g \propto x^{1-\sqrt{2}}$, which blows up at 0.

Layer 1 THE WHAT — surface analysis

The natural attempt. The centroid formula first. The region under a graph has area $\int_0^x g$, and the moment about the s-axis is $\int_0^x \int_0^{g(s)} t\,dt\,ds = \frac{1}{2}\int_0^x g^2$. So the condition reads

$$\frac{\frac{1}{2}\int_0^x g(s)^2\,ds}{\int_0^x g(s)\,ds} = \frac{1}{x}\int_0^x g(s)\,ds \quad\Longleftrightarrow\quad x\int_0^x g^2 = 2\Big(\int_0^x g\Big)^2. \tag{$*$}$$

Now the strong student tries the functions that make integrals of $[0, x]$ easy: powers. For $g = Cx^a$ with $a > -1$,

$$x\cdot\frac{C^2x^{2a+1}}{2a+1} = 2\cdot\frac{C^2x^{2a+2}}{(a+1)^2} \iff (a+1)^2 = 2(2a+1) \iff a^2-2a-1=0 \iff a = 1\pm\sqrt{2}.$$

$a = 1-\sqrt{2} \approx -0.414$ gives a function that is not continuous at 0, so $g = Cx^{1+\sqrt{2}}$ works. Many scripts stopped here.

That is a verification, not a solution. The problem says "find *all*". The student who tries to close the gap differentiates $(*)$:

$$\int_0^x g^2 + xg(x)^2 = 4g(x)\int_0^x g.$$

Now there are two unknown integrals, $\int g$ and $\int g^2$, and a pointwise g. Differentiating again brings in g', which need not exist — g is only continuous. The attempt stalls on an equation that looks second order and non-linear.

The model chain. Write $G(x) = \int_0^x g(s)\,ds$ and $H(x) = \int_0^x g(s)^2\,ds$. Since g is continuous, G and H are C^1 on $[0,\infty)$ with $G' = g$, $H' = g^2$ (fundamental theorem of calculus). Since $g > 0$ on $(0,\infty)$, $G(x) > 0$ for $x > 0$.

The condition. As computed in the natural attempt, the centroid height of R_x is $H/(2G)$ and the average of g is G/x, so for all $x > 0$

$$xH(x) = 2G(x)^2. \tag{1}$$

Differentiate once, then eliminate H. Both sides of (1) are differentiable, so $H + xg^2 = 4Gg$. Use (1) again to replace H by $2G^2/x$:

$$x^2g^2 - 4xGg + 2G^2 = 0 \qquad (x > 0).$$

Divide by $G^2 > 0$ and set $u(x) = xg(x)/G(x)$, which is continuous on $(0,\infty)$:

$$u^2 - 4u + 2 = 0, \qquad u \in \{r_1, r_2\}, \quad r_1 = 2+\sqrt{2},\ r_2 = 2-\sqrt{2}.$$

No switching. u is continuous on the connected set $(0,\infty)$ and takes values in a two-point set, so it is constant (intermediate value theorem): $u \equiv r$ with $r \in \{r_1, r_2\}$.

Solve. $G'/G = r/x$ on $(0,\infty)$, so $(\ln G)' = (r\ln x)'$ and $G = C_1x^r$ with $C_1 > 0$. Hence $g = G' = Cx^{r-1}$ with $C = rC_1 > 0$.

Continuity at 0. If $r = r_2$ then $r - 1 = 1 - \sqrt{2} < 0$ and $g(x) \to \infty$ as $x \to 0^+$, contradicting continuity on $[0, \infty)$. So $r = r_1$ and $g(x) = Cx^{1+\sqrt{2}}$.

Check. For $g = Cx^a$, $a = 1 + \sqrt{2}$: g is continuous on $[0, \infty)$, positive on $(0, \infty)$, and (1) holds because $(a+1)^2 = 2(2a+1)$, which is $a^2 - 2a - 1 = 0$. Every step above was an implication from (1), and the check closes the converse. The answer is $\boxed{g(x) = Cx^{1+\sqrt{2}},\ C > 0}$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach (1) and both differentiate it. The split is the line after: $H + xg^2 = 4Gg$. The natural attempt treats this as a new equation with three unknowns and thinks about differentiating again. The model notices that (1) is still true and still available, and uses it to delete H. What is left is an algebraic relation between g, G and x alone — a first-order ODE for G, and a homogeneous one.

> *Differentiating got rid of nothing I couldn't handle except H, and I already have a formula for H: the original equation. Substitute it back before doing anything else.*

> *Every term in $x^2g^2 - 4xGg + 2G^2$ has the same weight once I count xg as the same size as G. Divide by G^2 and the whole thing is a quadratic in the single quantity xg/G.*

> *A quadratic gives two roots at each x. Could g use one root on some intervals and the other elsewhere? xg/G is continuous and the two roots are separated, so no.*

- ▷ **Cue:** an identity between integrals $\int_0^x$ of g and g^2, required for all x, with g only continuous ⟶ **Move:** name the integrals G, H; differentiate once; substitute the original identity back to eliminate whichever integral has no derivative relation you want
- ▷ **Cue:** an equation $F(x, G, xG') = 0$ in which every term has the same total degree in G and xG' ⟶ **Move:** set $u = xG'/G$; the equation becomes algebraic in u, and G is a power of x
- ▷ **Cue:** a continuous quantity forced to lie in a finite set at every point of an interval ⟶ **Move:** it is constant; say so, citing the intermediate value theorem

Layer 3 THE WHY

Why is substituting back natural and not a trick? Look at what (1) *is*: an identity in x, holding on the whole half-line. An identity on an interval can be differentiated, and the derivative is a new identity — but the old one does not stop being true. You now have two equations in the three functions g, G, H, and H appears in both without its derivative g^2 doing any work that g cannot do. Two equations, one unwanted unknown: eliminate it. This is the same move as eliminating a parameter between $x = f(t)$, $y = h(t)$, and it is the same move as solving $y = \int_0^x y$ by differentiating to $y' = y$ and using the original equation for the initial condition. The integral equation is never discarded; it is used as data.

Why does the result have to be homogeneous? Because the problem is scale-invariant, and the phrasing tells you so. "Centroid" and "average value" are both unchanged if you stretch the picture: replace $g(s)$ by $\lambda g(s)$ and both sides of the condition multiply by λ; replace $g(s)$ by $g(\mu s)$ and the condition at x becomes the condition at μx. Any condition with both symmetries has power functions as its natural solutions, and any ODE it produces is invariant under $x \mapsto \mu x$ and $G \mapsto \lambda G$. The combination of x, G and G' that both scalings leave unchanged is $u = xG'/G = d \ln G / d \ln x$, the "local exponent" of G, and for $G = Cx^r$ it is the constant r. So the substitution $u = xg/G$ is asking, point by point, "what power of x does G look like here?" — and the equation answers: $2 \pm \sqrt{2}$, everywhere.

That is also why the natural attempt's guess of $g = Cx^a$ was right and yet insufficient. Scale invariance suggests powers; it does not prove that only powers work. The model chain's

content beyond the guess is exactly two lines: the elimination that turns the condition into "the local exponent is a root of $u^2 - 4u + 2$", and the continuity argument that stops the exponent jumping between roots.

Finally, the r_2 branch. It is not discarded because it "looks wrong"; it satisfies (1) on $(0, \infty)$ perfectly. It fails a hypothesis the problem stated deliberately: continuity on the *closed* half-line $[0, \infty)$. The problem's authors put the 0 in that bracket to kill $x^{1-\sqrt{2}}$. Read every endpoint in a hypothesis as a filter someone designed.

The official solution runs the same line — it writes $z = \int_0^x g$, differentiates and "has" the quadratic in xz'/z without saying where $\int g^2$ went. Here is where: (1) was substituted back.

Layer $+\alpha$ BACKTESTING

1. The first cue is "y-coordinate of the centroid". Write down $\bar{y} = \dfrac{\frac{1}{2}\int_0^x g^2}{\int_0^x g}$ at once, and then the condition as the clean identity $x \int_0^x g^2 = 2(\int_0^x g)^2$ with no fractions.
2. Integral identity for all $x \Rightarrow$ differentiate (fundamental theorem of calculus), eliminate using the undifferentiated identity, solve the resulting ODE.
3. Hesitation comes after the first differentiation, when $\int g^2$ is still present and a second differentiation is illegal (g need not be differentiable). It is reinforced away by the habit: before differentiating again, ask whether the original equation already expresses the unwanted term.
4. Three traps. (i) Guessing x^a, finding $a = 1 + \sqrt{2}$, and stopping — that is at most a partial answer to "find all". (ii) Forgetting that a pointwise quadratic allows branch switching; one sentence of continuity is owed. (iii) Keeping $x^{1-\sqrt{2}}$, or discarding it for the wrong reason: it fails continuity at 0, not the identity.
5. Whenever a condition is unchanged by rescaling the input and by rescaling the output, expect powers, and expect the equation to close up in the local exponent xf'/f.
6. Reusable decision rule: *after differentiating an identity, substitute the identity itself back before differentiating again; if every term then has equal weight, divide to get an equation in the scale-free ratio.*

1984 B5

1984 B5. For each nonnegative integer k, let $d(k)$ denote the number of 1's in the binary expansion of k (for example, $d(0) = 0$ and $d(5) = 2$). Let m be a positive integer. Express

$$\sum_{k=0}^{2^m-1} (-1)^{d(k)} k^m$$

in the form $(-1)^m a^{f(m)} (g(m))!$, where a is an integer and f and g are polynomials.

Fields: combinatorics; inclusion–exclusion; generating functions / multilinear algebra **Position:** B5 — the second-hardest problem on the paper, and it is hard for a specific reason: the object $(-1)^{d(k)}$ looks number-theoretic, but the entire difficulty is recognising it as something else. **Answer:** $a = 2$, $f(m) = \binom{m}{2} = \frac{m(m-1)}{2}$, $g(m) = m$, so the sum equals $(-1)^m \, 2^{m(m-1)/2} \, m!$.

Caveat. The book's own solution is present in the OCR but cuts off mid-argument: it defines $D(x) = (1-x)(1-x^2)(1-x^4)\cdots(1-x^{2^{m-1}})$, correctly identifies $D(x) = \sum_{k=0}^{2^m-1} (-1)^{d(k)} x^k$ from uniqueness of binary representations, and states that applying $(x\frac{d}{dx})^m$ to D and substituting $F(x) = D(x+1)$ recovers the sum – then the scan ends

before the evaluation is carried out. That generating-function route is legitimate and reaches the same place I do (a zero of D of order exactly m at $x = 1$ is what makes the m-th application of $x\frac{d}{dx}$ pick out a single coefficient), but I cannot reproduce the book's last steps from what survived the scan, so what follows is solved independently, using a cleaner route through the same fact about D, and checked numerically for $m = 1, \dots, 6$ before being written up.

Layer 1 THE WHAT — surface analysis

The natural attempt. The honest first move is to compute. $m = 1$: $k = 0, 1$, sum $= 0 - 1 = -1$. $m = 2$: $k = 0, 1, 2, 3$ with $d = 0, 1, 1, 2$, so the sum is $0 - 1 - 4 + 9 = 4$. $m = 3$: $d(0..7) = 0, 1, 1, 2, 1, 2, 2, 3$, giving $0 - 1 - 8 + 27 - 64 + 125 + 216 - 343 = -48$.

So the values are $-1, 4, -48, \dots$, and the target shape $(-1)^m a^{f(m)}(g(m))!$ invites guessing $g(m) = m$ immediately, since $1!, 2!, 3! = 1, 2, 6$ divide $1, 4, 48$ cleanly with quotients $1, 2, 8$ – powers of 2 with exponents $0, 1, 3$. Triangular numbers. A well-prepared student can reach the conjecture $a = 2$, $f(m) = \binom{m}{2}$ from three data points in a few minutes and would be right.

The trouble is that guessing the closed form is not solving the problem – the problem says *express*, which the graders read as *prove*, and three data points are not a proof. So the natural attempt turns to proving the conjecture directly, and the direct route is a recursion on m by splitting the range of summation at the leading bit. Every k with $0 \le k \le 2^m - 1$ is either $k = j$ with $0 \le j \le 2^{m-1} - 1$ (leading bit 0, so $d(k) = d(j)$) or $k = 2^{m-1} + j$ with the same range of j (leading bit 1, so $d(k) = d(j) + 1$). Writing S_m for the target sum,

$$S_m = \sum_{j=0}^{2^{m-1}-1} (-1)^{d(j)} j^m - \sum_{j=0}^{2^{m-1}-1} (-1)^{d(j)} \big(2^{m-1} + j\big)^m.$$

Expanding $(2^{m-1} + j)^m$ by the binomial theorem and subtracting cancels the j^m terms, leaving $S_m = -\sum_{r=0}^{m-1} \binom{m}{r} 2^{(m-1)(m-r)} \sum_{j=0}^{2^{m-1}-1} (-1)^{d(j)} j^r$. This is a genuine recursion – it expresses S_m in terms of sums $T_r := \sum_{j<2^{m-1}} (-1)^{d(j)} j^r$ for *every* $r < m$, not just $S_{m-1} = T_{m-1}$. To close the recursion the student now needs all the lower moments $T_0, \dots, T_{m-2}$ as well, which is a second unknown family, and proving they vanish (they do, for $r < m - 1$, but that is itself a fact requiring proof) is a second problem stacked on the first. The approach is not wrong – it can be pushed through with enough patience and a subsidiary induction that $T_r = 0$ for $r < m - 1$ – but it is a page of binomial-coefficient bookkeeping for a competition allowing perhaps 30 minutes on the last B-problem.

The model chain. Re-read $d(k)$: it is not really a digit count, it is the size of a set. Write k's binary expansion using m bits (legitimate since $0 \le k \le 2^m - 1$), so $k = \sum_{i=0}^{m-1} b_i 2^i$ with each $b_i \in \{0, 1\}$. Let $I = \{0, 1, \dots, m - 1\}$ and let $S = \{i \in I : b_i = 1\}$ be the set of positions where the bit is 1. Then $d(k) = |S|$ and $k = \sum_{i \in S} 2^i$, and because binary representations are unique, $k \mapsto S$ is a bijection between $\{0, \dots, 2^m - 1\}$ and the 2^m subsets of I. So

$$S_m = \sum_{k=0}^{2^m-1} (-1)^{d(k)} k^m = \sum_{S \subseteq I} (-1)^{|S|} \Big(\sum_{i \in S} 2^i \Big)^m.$$

This is now an instance of a general identity. For any n indeterminates $x_1, \dots, x_n$,

$$\sum_{T \subseteq [n]} (-1)^{n-|T|} \Big(\sum_{i \in T} x_i \Big)^n = n!\, x_1 x_2 \cdots x_n. \tag{$*$}$$

Proof of $(*)$. Expand each $\big(\sum_{i \in T} x_i\big)^n$ by the multinomial theorem: it is $\sum_a \binom{n}{a_1, \dots, a_n} \prod_i x_i^{a_i}$, summed over tuples $a = (a_i)_{i \in [n]}$ with $a_i = 0$ for $i \notin T$ and $\sum a_i = n$. Fix a monomial $x^a = \prod x_i^{a_i}$ with

$\sum a_i = n$ and support $\operatorname{supp}(a) = \{i : a_i > 0\}$; it contributes to the coefficient of x^a in $\left(\sum_{i\in T} x_i\right)^n$ exactly when $T \supseteq \operatorname{supp}(a)$, with coefficient $\binom{n}{a_1,\dots,a_n}$ every time. So the total coefficient of x^a in the left side of $(*)$ is

$$\binom{n}{a_1,\dots,a_n} \sum_{T\supseteq \operatorname{supp}(a)} (-1)^{n-|T|}.$$

Let $U = [n] \setminus \operatorname{supp}(a)$. Writing $T = \operatorname{supp}(a) \cup V$ for $V \subseteq U$, the inner sum is $(-1)^{n-|\operatorname{supp}(a)|} \sum_{V\subseteq U} (-1)^{-|V|} = (-1)^{n-|\operatorname{supp}(a)|}(1-1)^{|U|}$, which is 0 unless $U = \varnothing$, i.e. $\operatorname{supp}(a) = [n]$. Since $\sum a_i = n$ and there are n indices each with $a_i \geq 1$, this forces $a_i = 1$ for every i. So every monomial coefficient vanishes except that of $x_1 x_2 \cdots x_n$, whose coefficient is $\binom{n}{1,1,\dots,1} = n!$. This proves $(*)$.

Applying $(*)$. Take $n = m$, index set $[n] = I = \{0,\dots,m-1\}$, and $x_i = 2^i$. Since $(-1)^{|S|} = (-1)^m(-1)^{m-|S|}$,

$$S_m = (-1)^m \sum_{S\subseteq I} (-1)^{m-|S|} \Big(\sum_{i\in S} 2^i\Big)^m = (-1)^m \cdot m! \cdot \prod_{i=0}^{m-1} 2^i = (-1)^m\, m!\, 2^{0+1+\cdots+(m-1)} = (-1)^m\, 2^{\binom{m}{2}}\, m!.$$

Numerical check. $m = 1$: $(-1)(2^0)(1) = -1$. $m = 2$: $(1)(2^1)(2) = 4$. $m = 3$: $(-1)(2^3)(6) = -48$. Direct computation (also carried out for $m = 4, 5, 6$, giving $1536, -122880, 23592960$) matches in every case.

Answer. $a = 2$, $f(m) = \binom{m}{2} = \frac{m(m-1)}{2}$, $g(m) = m$: $S_m = (-1)^m 2^{m(m-1)/2}\, m!$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same three numbers $-1, 4, -48$ by direct computation, and both notice the shape $2^? \cdot (\text{something})!$. They split at what to do with the recursion that direct expansion produces. The natural attempt reads $d(k)$ literally, as a digit-count that happens to interact with k^m through the binomial theorem when k's leading bit is peeled off. The model chain rereads $d(k)$ structurally, as $|S|$ for a subset S of an m-element index set, and asks what identity has exactly the shape "alternating sum over subsets of a set-sum raised to the m".

> *$d(k)$ is not really about digits, it's the size of the set of positions where the binary expansion has a 1. If I let S be that set, then k is just $\sum_{i\in S} 2^i$, and the whole sum is $\sum_{S\subseteq I}(-1)^{|S|}(\sum_{i\in S} 2^i)^m$ – an alternating sum, over* all *subsets of an m-element set, of a subset-sum raised to the m-th power. I have seen that shape before: it is what you get when you expand $(\sum x_i)^n$ and use inclusion–exclusion to kill every term except the one that uses every variable.*

> *If I trust that structural read, I don't need to know the individual value of $\sum_{i\in S} 2^i$ until the very last line – I should first prove the identity $(*)$ for general indeterminates $x_1, \dots, x_n$, because in that generality the proof is transparent (inclusion–exclusion on the support of a monomial), and only then substitute the specific numbers $x_i = 2^i$. Proving the identity with the actual powers of 2 already plugged in would obscure exactly the cancellation that makes it work.*

- ▷ **Cue:** an alternating sum, over all subsets of a finite index set, of (subset-sum)$^{(\text{size of the index set})}$ ⟶ **Move:** recognise it as the coefficient-extraction identity $\sum_{T\subseteq[n]}(-1)^{n-|T|}(\sum_{i\in T} x_i)^n = n!\, x_1 \cdots x_n$; prove it in general (support argument on a monomial) before substituting specific numbers
- ▷ **Cue:** a quantity defined as "the number of 1's in a binary expansion", appearing inside a sign $(-1)^{d(k)}$ ⟶ **Move:** stop reading it as a digit count and read it as $|S|$ for the set S of active bit positions; then $k = \sum_{i\in S} 2^i$ turns the whole sum into a sum over subsets
- ▷ **Cue:** a competition answer requested in the form (integer)$^{\text{polynomial}}\times$(polynomial)! ⟶ **Move:** expect the proof to end in a clean product like $n! \prod x_i$ – work backwards from that shape to guess which classical identity is being asked for

Layer 3 THE WHY

Why does inclusion–exclusion over subsets kill every monomial except the fully mixed one? This is worth deriving rather than quoting, because the identity $(*)$ is the entire problem in disguise, and the official solution (so far as the scan preserves it) reaches the same fact by a different door – a generating function $D(x) = \prod_{i=0}^{m-1}(1 - x^{2^i})$ that vanishes to order exactly m at $x = 1$. Both routes are answering the same underlying question: *what mechanism has an alternating sum survive only when every one of m independent conditions is simultaneously "used"*? That mechanism is $(1-1)^{|U|} = 0$ unless $U = \varnothing$ – the same fact used to prove $\sum_{V \subseteq U}(-1)^{|V|} = 0$ for every nonempty U, which is itself just the binomial theorem at $x = -1$.

So the axiom underneath both solutions is the same one: a sum over all subsets of a set, weighted by $(-1)^{|\text{subset}|}$, vanishes unless the object being summed does not distinguish "in" from "out" for at least one element – and if it does distinguish for every element, the sum collects that distinguishing exactly once, with coefficient ± 1. In $(*)$, the monomial x^a "distinguishes" index i (cares whether $i \in T$) exactly when $a_i > 0$; if some $a_i = 0$, the monomial doesn't care whether $i \in T$, so summing over whether $i \in T$ contributes a factor of $(1 - 1) = 0$ to that coefficient. The only monomial that cares about every index is the fully multilinear one, and it survives with coefficient $n!$ because that's the number of orderings of n distinct factors, i.e. the multinomial coefficient $\binom{n}{1,\ldots,1}$.

Why is $2^{m(m-1)/2}$ specifically the right power once $(*)$ is available? Because the x_i are not abstract: they are $2^0, 2^1, \ldots, 2^{m-1}$, forced by the fact that k ranges exactly over $\{0, \ldots, 2^m - 1\}$ and binary digits are the unique way to hit every such k exactly once as k runs over all 2^m subsets. The product $\prod x_i$ is then forced to be $2^{0+1+\cdots+(m-1)} = 2^{\binom{m}{2}}$ – the exponent is a triangular number for the same reason the sum of the first $m - 1$ nonnegative integers is a triangular number, nothing deeper. The $m!$ is forced by $(*)$ regardless of what the x_i are, which is exactly why the problem's answer has an $m!$ that is "clean" (independent of the specific powers of two) sitting next to a $2^{f(m)}$ that is not (it encodes exactly which powers of two were used).

Why does the problem's phrasing "express... in the form $(-1)^m a^{f(m)}(g(m))!$" all but announce the identity $(*)$? Because that shape – integer power times a factorial – is the signature of "$n!$ times a product", and $n!$ appearing alone (not $\binom{2n}{n}$ or some other combinatorial count) is the signature of a fully multilinear extraction, i.e. of inclusion–exclusion collapsing onto exactly one term out of 2^n. A solver who has internalised $(*)$ as a named tool reads the requested answer-shape as a hint before doing any algebra at all.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $(-1)^{d(k)}$ next to k^m summed over *all* of $0, \ldots, 2^m - 1$ – the range is exactly 2^m, matching the number of subsets of an m-set. On seeing it, write down: "$d(k) = |S|$ for the set S of 1-bits; $k = \sum_{i \in S} 2^i$; the sum is over all subsets of an m-element index set."

2. That cue suggests inclusion–exclusion / the multinomial-extraction identity $(*)$: an alternating sum over all subsets of (subset-sum)^n equals $n!$ times the product of the underlying quantities.

3. Hesitation happens right after the small-case computation, at the point of deciding whether to keep computing (recursion on the leading bit, binomial theorem) or to stop and look for a structural identity that matches the alternating-sum-over-subsets shape. It is reinforced away by noticing that the recursive approach reintroduces new unknown sums T_r for every $r < m$, which is a sign the approach is fighting the problem's structure rather than using it.

4. The trap is trusting the guessed closed form from three data points and writing it up as if pattern-matching were a proof. A second, subtler trap is proving $(*)$ with the specific numbers 2^i already substituted, which hides the cancellation and makes the identity look like a coincidence of binary arithmetic rather than the general combinatorial fact it is.

5. Any time a sign $(-1)^{(\text{popcount or subset-size})}$ multiplies a power of a subset-defined quantity, and the outer sum runs over all 2^n subsets, check the shape $\sum_T (-1)^{n-|T|}(\sum_{i \in T} x_i)^n = n! \prod x_i$

before doing any expansion by hand.

6. Reusable decision rule: *when an alternating sum runs over all subsets of a finite index set (recognisable from a $(-1)^{|S|}$ or $(-1)^{d(k)}$-type sign and a range of size 2^n), first ask what happens to a generic monomial as the subset ranges over all supersets of its support – it either vanishes (support incomplete) or survives with a clean combinatorial coefficient (support complete). This single fact reduces an entire family of "alternating sum over binary/subset data" problems to identifying which monomial is the survivor.*

1984 B6

1984 B6. A sequence of convex polygons $\{P_n\}$, $n \geq 0$, is defined inductively as follows. P_0 is an equilateral triangle with sides of length 1. Once P_n has been determined, its sides are trisected; the vertices of P_{n+1} are the interior trisection points of the sides of P_n. Thus, P_{n+1} is obtained by cutting corners off P_n, and P_n has $3 \cdot 2^n$ sides. (P_1 is a regular hexagon with sides of length $1/3$.)

Express $\lim_{n\to\infty} \text{Area}(P_n)$ in the form $\sqrt{a}/b$, where a and b are positive integers.

Fields: plane geometry; vectors and the cross product; geometric series; iterated constructions **Position:** B6 — last problem of the paper, but its difficulty is entirely in choosing the right bookkeeping; once the corners are measured with edge vectors the computation is four lines **Answer:** $\lim_{n\to\infty} \text{Area}(P_n) = \sqrt{3}/7$, i.e. $a = 3$, $b = 7$. The area cut off at step n is $\frac{1}{18}\sum u \times v$ over consecutive edge vectors, that sum is multiplied by exactly $2/9$ at every step, so the total cut is $\frac{\sqrt{3}}{12} \cdot \frac{9}{7} = \frac{3\sqrt{3}}{28}$ and $\frac{\sqrt{3}}{4} - \frac{3\sqrt{3}}{28} = \frac{\sqrt{3}}{7}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student starts with pictures and exact lengths. Going from the triangle P_0 to the hexagon P_1 cuts off three equilateral triangles of side $\frac{1}{3}$, each of area $\frac{\sqrt{3}}{4} \cdot \frac{1}{9} = \frac{\sqrt{3}}{36}$, so

$$\text{Area}(P_1) = \frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{12} = \frac{\sqrt{3}}{6},$$

which agrees with the regular hexagon of side $\frac{1}{3}$: $\frac{3\sqrt{3}}{2} \cdot \frac{1}{9}$.

Next, $P_1 \to P_2$. Each corner of the hexagon has interior angle 120° and the cut-off triangle has two sides $\frac{1}{9}$ enclosing it, so its area is $\frac{1}{2} \cdot \frac{1}{9} \cdot \frac{1}{9} \cdot \frac{\sqrt{3}}{2} = \frac{\sqrt{3}}{324}$; six corners remove $\frac{\sqrt{3}}{54}$. The ratio of the second cut to the first is $\frac{\sqrt{3}/54}{\sqrt{3}/12} = \frac{2}{9}$.

Now the student guesses: "the cuts form a geometric series with ratio $\frac{2}{9}$". That gives total cut $\frac{\sqrt{3}}{12} \cdot \frac{1}{1-2/9} = \frac{3\sqrt{3}}{28}$ and the right answer, $\frac{\sqrt{3}}{7}$. But two terms do not make a series, and the attempt to check a third term hits the wall. P_2 is a 12-gon whose sides alternate between $\frac{1}{9}$ and $\frac{\sqrt{3}}{9}$ (the middle thirds of the hexagon's sides, and the chords across its corners), all interior angles 150°. It is not regular. One step further, P_3 has 24 sides, and if the hexagon's consecutive edge vectors are u, v, the edges of P_3 near that corner point in the directions of u, $2u + v$, $u + v$, $u + 2v$, v: lengths now involve $|2u + v| = \frac{\sqrt{7}}{3}$, angles involve arctan of surds, and there are several non-congruent kinds of corner. Each further step multiplies the number of corner types. Bookkeeping by side lengths and angles does not close; the student either writes down the unjustified guess or runs out of time drawing 24-gons.

The model chain. *Notation.* For plane vectors $u = (u_1, u_2)$, $v = (v_1, v_2)$ write $u \times v = u_1 v_2 - u_2 v_1$. It is bilinear, antisymmetric ($u \times u = 0$), and the triangle spanned by u and v from a common point has area $\frac{1}{2}|u \times v|$.

Traverse P_n anticlockwise and let $e_1, \ldots, e_m$ ($m = 3 \cdot 2^n$) be its edge vectors in order, indices mod m. Convexity (with the anticlockwise orientation) means every turn is a left turn: $e_i \times$

$e_{i+1} > 0$. Define

$$S_n = \sum_{i=1}^{m} e_i \times e_{i+1}.$$

The cut at one corner. At the vertex where edge $u = e_i$ ends and $v = e_{i+1}$ begins, the two interior trisection points adjacent to it are the vertex minus $u/3$ and the vertex plus $v/3$. The cut-off triangle is spanned by $u/3$ and $v/3$ at that vertex, so its area is $\frac{1}{2} \cdot \frac{1}{9}\, u \times v = \frac{1}{18}\, u \times v$. These corner triangles lie in P_n (convexity), each is bounded by one chord and the piece of ∂P_n between that chord's endpoints, and these boundary pieces are disjoint (they are the outer thirds of the edges, grouped by vertex). So P_n is P_{n+1} together with m corner triangles, interiors disjoint, and

$$\operatorname{Area}(P_n) - \operatorname{Area}(P_{n+1}) = \frac{S_n}{18}. \tag{1}$$

The new edges. Along u, P_{n+1} keeps the middle third, an edge $u/3$. Across the corner it has the chord from the $\frac{2}{3}$-point of u to the $\frac{1}{3}$-point of v, which is $\frac{u}{3} + \frac{v}{3}$. So the edge list of P_{n+1} is obtained from that of P_n by

$$\ldots,\ u,\ v,\ w,\ \ldots \longmapsto \ldots,\ \tfrac{u}{3},\ \tfrac{u+v}{3},\ \tfrac{v}{3},\ \tfrac{v+w}{3},\ \tfrac{w}{3},\ \ldots$$

The recursion. Each old corner (u, v) produces exactly two consecutive pairs in the new list, and

$$\tfrac{u}{3} \times \tfrac{u+v}{3} = \tfrac{1}{9}\big(u \times u + u \times v\big) = \tfrac{1}{9}\, u \times v, \qquad \tfrac{u+v}{3} \times \tfrac{v}{3} = \tfrac{1}{9}\big(u \times v + v \times v\big) = \tfrac{1}{9}\, u \times v.$$

Every consecutive pair of P_{n+1} arises this way, once. Hence

$$S_{n+1} = \tfrac{2}{9}\, S_n, \qquad S_n = \big(\tfrac{2}{9}\big)^n S_0. \tag{2}$$

(Both new cross products are positive, so P_{n+1} again turns left at every vertex; the problem asserts convexity anyway.)

Start value. In the equilateral triangle, consecutive edges have length 1 and meet at exterior angle 120°, so $e_i \times e_{i+1} = \sin 120^\circ = \frac{\sqrt{3}}{2}$ and $S_0 = \frac{3\sqrt{3}}{2}$. Check with (1): the first cut is $\frac{S_0}{18} = \frac{\sqrt{3}}{12}$, as computed directly.

Sum. By (1) and (2), $\operatorname{Area}(P_n)$ decreases and

$$\lim_{n\to\infty} \operatorname{Area}(P_n) = \frac{\sqrt{3}}{4} - \sum_{n\ge 0} \frac{S_0}{18}\Big(\frac{2}{9}\Big)^n = \frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{12} \cdot \frac{9}{7} = \frac{\sqrt{3}}{4} - \frac{3\sqrt{3}}{28} = \frac{\sqrt{3}}{7}.$$

So $a = 3$, $b = 7$ (with a squarefree; $\sqrt{12}/14$ is the same number). A direct numerical iteration of the construction to $n = 14$ gives $0.2474358\ldots$, agreeing with $\sqrt{3}/7 = 0.24743583\ldots$, and the ratio of successive cuts is $0.22222\ldots$ at every step.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains compute the first cut correctly. The split is the moment of writing down the area of a corner. The natural attempt writes $\frac{1}{2} \cdot \frac{|u|}{3} \cdot \frac{|v|}{3} \sin\theta$ — lengths and an angle, quantities that have to be recomputed for every new shape. The model writes $\frac{1}{18}\, u \times v$ — a bilinear function of the edge vectors, which the construction transforms linearly. After that line, the model's third term is not a new computation: it is the same two-line expansion as the second.

> *The new polygon's edges are $u/3$, $(u+v)/3$, $v/3$ — linear combinations of the old edges. So I should stop describing the polygon by lengths and angles and describe it by its list of edge vectors; that is the thing the construction acts on.*

> *A corner triangle's area is $\frac{1}{18}\,u \times v$, and $\times$ is bilinear. If I push the new edges through the cross product, the $u \times u$ terms die, and the new corner sum should come out as a multiple of the old one. Compute it before drawing anything.*

> *I've got the ratio $\frac{2}{9}$ from two terms. That is a conjecture. What quantity would have to be multiplied by exactly $\frac{2}{9}$ at every step for the conjecture to be true? Find that quantity and prove the recursion for it.*

- ▷ **Cue:** a figure built by an iterated rule in which each new edge or point is a fixed linear combination of old edges or points (trisection, midpoints, weighted averages) $\longrightarrow$ **Move:** describe the figure by its vector data (edge vectors or vertex vectors) and write the rule as a linear map on that data
- ▷ **Cue:** an area to track through a linear construction on plane vectors $\longrightarrow$ **Move:** express it with the cross product $u \times v$, which is bilinear, and expand; antisymmetry $u \times u = 0$ kills the diagonal terms
- ▷ **Cue:** "cutting corners" repeatedly and a limiting area asked for $\longrightarrow$ **Move:** area of the limit = original area minus the sum of the cuts; look for a geometric ratio between consecutive total cuts
- ▷ **Cue:** the first two terms of a sequence suggest a constant ratio but later terms get harder to compute directly $\longrightarrow$ **Move:** find a single quantity defined at every stage whose one-step recursion you can prove with the same computation at every stage

Layer 3 THE WHY

Why is the vector description natural and not a trick? Go to what the problem actually defines. It never says anything about angles or lengths. It says "the vertices of P_{n+1} are the interior trisection points of the sides of P_n". A trisection point is $\frac{2}{3}A + \frac{1}{3}B$ — an affine combination of two vertices. The whole construction is therefore a linear operation on vertex coordinates, and on edge vectors (which are differences of vertices) it is the map $u, v \mapsto \frac{u}{3}, \frac{u+v}{3}, \frac{v}{3}$. The circle is defined by its centre, so the centre is the object of interest; this construction is defined by linear combinations, so the vector data it combines is the object of interest. Lengths ($\sqrt{\text{quadratic form}}$) and angles (arccos of a ratio) are non-linear functions of that data, and that non-linearity is exactly why the natural attempt's cases multiply: every new direction $2u + v$, $u + 2v$, $3u + 2v$, ... gives a new surd.

Why the cross product specifically? Area is the quantity asked for, and in the plane area is the bilinear, antisymmetric form $u \times v$ (it is the 2×2 determinant; the shoelace formula is the same object). A bilinear quantity fed a linear recursion comes out as a combination of the old bilinear quantities. That is all the $\frac{2}{9}$ is:

$$\frac{2}{9} = \underbrace{2}_{\text{each corner becomes two corners}} \times \underbrace{\frac{1}{9}}_{\text{edges scaled by } 1/3\text{, area by } 1/9},$$

with the diagonal terms $u \times u$ vanishing because a vector spans no area with itself. Once you see the factorisation, the answer is not surprising at all; what would be surprising is a construction linear in the edges whose corner sum did *not* obey a clean recursion.

Why does the natural attempt get the right answer and still fail? Because P_1 and P_2 are special. P_1 is regular, and P_2 has all angles equal, so "all corners cut are congruent" is true for the first two steps and the student's area-per-corner computation happens to be uniform. From P_3 on, the corners are genuinely different, but the *sum* $\sum u \times v$ still scales by $\frac{2}{9}$. The individual terms are not self-similar; their sum is. When a construction produces irregular shapes but a

clean answer is requested, the clean thing is almost always an aggregate that transforms as a whole.

Why does the requested form $\sqrt{a}/b$ matter? It is a cue. No π is allowed in the answer, so the limit is not a disc or any conic region with a transcendental area, and the answer is reachable by finitely many algebraic operations plus one geometric series. A student who suspects "the limit is a circle" should be stopped by the form alone. (It is not a circle: de Rham, cited in the official comment, studied this curve, which is C^1 but has infinite curvature on a dense set; numerically, it even dips inside the incircle of P_1, reaching distance about 0.2771 from the centre against the incircle radius $\frac{\sqrt{3}}{6} \approx 0.2887$.)

The official solution writes: "Suppose that u and v are consecutive edges in P_n. Then $u/3$, $(u+v)/3$, and $v/3$ are consecutive edges in P_{n+1}." It does not say why anyone would think in edge vectors. Here is where it comes from: the construction is made of trisection points, trisection is linear, and so edge vectors are the coordinates in which the construction is simplest.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the vertices of P_{n+1} are the interior trisection points of the sides of P_n". Write down immediately: if consecutive edges are u, v, the new edges are $u/3$, $(u+v)/3$, $v/3$, and the removed corner has area $\frac{1}{18}u \times v$.

2. Linear rule on vectors plus a bilinear quantity (area) $\Rightarrow$ push the rule through the cross product, get a one-step recursion for the total, sum the geometric series.

3. Hesitation occurs after computing $\frac{2}{9}$ from $P_0 \to P_1 \to P_2$: is it really constant? It is reinforced away by asking which single quantity carries the ratio and proving its recursion in general, which here costs two lines.

4. Two traps. (i) Extrapolating a ratio from two terms computed on shapes that are more symmetric than the general P_n (regular hexagon, equiangular 12-gon). (ii) Believing the limit is a circle and reaching for π; the answer form $\sqrt{a}/b$ rules this out. A third, smaller one: forgetting that corner cuts must be disjoint and inside P_n for "area = original minus cuts" — convexity supplies both.

5. Whenever a construction builds new points as fixed weighted averages of old ones, change coordinates to the vectors it averages. Quadratic quantities (areas, sums of squares, energies) then satisfy linear recursions, even when the figures themselves become irregular.

6. Reusable decision rule: *if an iterated rule is linear in some vector data, measure everything you track by a bilinear or quadratic form in that data and look for the recursion of the total, not of the individual pieces.*