DAWN PUTNAM SESSION

Putnam 1985 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1985 — Paper A **5**
1985 A1 5
1985 A2 7
1985 A3 10
1985 A4 13
1985 A5 15
1985 A6 17

Putnam 1985 — Paper B **21**
1985 B1 21
1985 B2 23
1985 B3 25
1985 B4 27
1985 B5 30
1985 B6 32

Putnam 1985 — Paper A

1985 A1

1985 A1. Determine, with proof, the number of ordered triples (A_1, A_2, A_3) of sets which have the property that

(i) $A_1 \cup A_2 \cup A_3 = \{1,2,3,4,5,6,7,8,9,10\}$, and

(ii) $A_1 \cap A_2 \cap A_3 = \emptyset$.

Express your answer in the form $2^a 3^b 5^c 7^d$, where a, b, c, d are nonnegative integers.

Fields: combinatorics (counting, product rule) **Position:** A1 — the opener **Answer:** $6^{10} = 2^{10}3^{10}5^0 7^0$, i.e. $a = b = 10$, $c = d = 0$. Each of the ten elements independently chooses which of the three sets it lies in; the conditions forbid exactly the choices "none" and "all three", leaving $2^3 - 2 = 6$ choices per element.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read the two conditions as global statements about three sets, and you reach for the tool that handles union-and-intersection conditions globally: inclusion–exclusion. First, how many ordered triples have $A_1 \cup A_2 \cup A_3 = [10]$, where $[10] = \{1, \ldots, 10\}$? You might count triples of subsets (8^{10} of them, since each A_i is any of 2^{10} subsets) and subtract those missing some element, which is an inclusion–exclusion over the missing elements:

$$\#\{\text{union} = [10]\} = \sum_{k=0}^{10} (-1)^k \binom{10}{k} (2^{10-k})^3 = \sum_{k} (-1)^k \binom{10}{k} 8^{10-k} = (8-1)^{10} = 7^{10}.$$

Next you must remove the triples whose triple intersection is non-empty. Another inclusion–exclusion, this time over the set K of elements forced into all three sets: an element of K has exactly one allowed pattern (in all three), and each other element has 7 patterns keeping the union full, so

$$\#\{\text{union} = [10],\ \text{intersection} = \emptyset\} = \sum_{k=0}^{10} (-1)^k \binom{10}{k} 1^k 7^{10-k} = (7-1)^{10} = 6^{10}.$$

This is correct and it is honest work. But look at what happened: twice, a ten-term alternating sum collapsed by the binomial theorem into a single power. And notice what you had to say inside the second sum — "each other element has 7 patterns". You were already reasoning element by element; the inclusion–exclusion was scaffolding around an argument that did not need it. (The worse natural attempt, casework on $|A_1|, |A_2|, |A_3|$ and their overlaps, does not finish in exam time at all.)

The model chain. Encode an ordered triple (A_1, A_2, A_3) of subsets of $[10]$ by recording, for each $x \in [10]$, its *membership pattern*

$$\pi(x) = \big([x \in A_1], [x \in A_2], [x \in A_3]\big) \in \{0,1\}^3.$$

(Condition (i) forces every $A_i \subseteq [10]$, so no other elements need recording.) The map $(A_1, A_2, A_3) \mapsto (\pi(1), \ldots, \pi(10))$ is a bijection between triples of subsets of $[10]$ and sequences of ten patterns: given the patterns, $A_i = \{x : \pi(x)_i = 1\}$ recovers the sets.

Now translate the conditions. Condition (i) says every x is in at least one set: $\pi(x) \neq (0,0,0)$. Condition (ii) says no x is in all three: $\pi(x) \neq (1,1,1)$. Both conditions are statements about each element separately, with no interaction between different elements. So the admissible triples correspond exactly to sequences in which each $\pi(x)$ is one of the $2^3 - 2 = 6$ remaining patterns, chosen independently. By the product rule the count is

$$6^{10} = 2^{10} \cdot 3^{10} \cdot 5^0 \cdot 7^0.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths begin with the same sentence: "the conditions constrain three sets". They split at the choice of what to count. The natural attempt counts *sets* and corrects for global conditions; the model counts *elements*, asks what each element can do, and finds the conditions are local. The line where they part is the model's "for each x, record which sets contain it".

> *Union and intersection are defined element by element: x is in the union if it is in some A_i, in the intersection if it is in every A_i. So both conditions are really conditions on each element separately. Let me ask what one element is allowed to do.*

> *Each element picks a subset of $\{1,2,3\}$ — the sets it belongs to. Eight options, minus "none" (violates the union) and "all" (violates the intersection): six. Ten elements, independent: 6^{10}.*

Layer 3 THE WHY

Why is thinking element by element natural rather than inspired? Go back to the definitions. $x \in A_1 \cup A_2 \cup A_3$ *means* "$x \in A_1$ or $x \in A_2$ or $x \in A_3$"; $x \in A_1 \cap A_2 \cap A_3$ *means* "x is in all three". A statement of the form $A_1 \cup A_2 \cup A_3 = [10]$ is therefore ten statements, one per element, joined by "and". The same holds for $A_1 \cap A_2 \cap A_3 = \emptyset$. When a global condition unpacks into a conjunction of per-element conditions that never mention two elements at once, the elements do not interact, and the count is a product over elements. That is the product rule, and it is the whole problem.

The second structural fact is the encoding itself. A triple of subsets of $[10]$ is the same data as a function $[10] \to \{0,1\}^3$, or equivalently a 10×3 matrix of 0s and 1s (Kedlaya, Poonen and Vakil phrase it as exactly this matrix, with "no row equal to 000 or 111"). Their solution states the bijection and stops. Here is where it comes from: a subset of $[10]$ *is* a function $[10] \to \{0,1\}$ (its indicator), so three subsets are three such functions, which is one function into $\{0,1\}^3$. Read the matrix by columns and you see the sets; read it by rows and you see the elements. The problem is phrased by columns; it is solved by rows. Switching which index you sum over first is the move.

Why does the problem-setter ask for the form $2^a 3^b 5^c 7^d$? Because 6^{10} is the answer and the form hides it only slightly — but notice that 7 is offered as a possible prime. A student who counted only condition (i) gets 7^{10}, and the form accommodates that wrong answer. The distractor is built into the answer format. A student who counts only (ii) also gets 7^{10}. Only a student who handles both conditions gets rid of the 7.

And why is inclusion–exclusion the wrong first tool here, though it works? Inclusion–exclusion exists to handle conditions that *overlap* across the objects you are counting. When the conditions are already independent per element, inclusion–exclusion recomputes the independence and then the binomial theorem undoes the work: $\sum (-1)^k \binom{n}{k} q^{n-k} = (q-1)^n$ is precisely

the statement "each element has $q-1$ options". Seeing an alternating sum collapse to a clean power is a signal that a direct product argument existed.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pair of conditions written with $\cup$ and $\cap$ over a small ground set $\{1,\dots,10\}$. Write down immediately: "each x lies in a subset of $\{A_1, A_2, A_3\}$; (i) says not none, (ii) says not all".

2. The method is the product rule over elements, after encoding the triple as a membership pattern per element (a 10×3 binary matrix, counted by rows).

3. Hesitation occurs at "are the choices really independent?" Reinforce it by rereading the definitions of $\cup$ and $\cap$: each condition is a conjunction over x of statements involving that x alone. No condition relates two elements, so independence is literal.

4. Two traps. First, counting only one condition and getting 7^{10} — which the answer format $2^a3^b5^c7^d$ happily accepts. Second, casework on the sizes of the sets, which spirals.

5. Whenever set conditions are built from $\cup, \cap$ and complements (Venn-diagram conditions), with no condition on sizes, count by asking which Venn region each element goes in. Forbidden regions are simply deleted from each element's menu.

6. Reusable rule: before counting a configuration, ask whether its constraints are local (each involves one element at a time). If so, the count is the product of the per-element option counts. If an inclusion–exclusion collapses to $(q-1)^n$, you have rediscovered a product rule and should have started there.

▷ **Cue:** counting tuples of sets subject to conditions on unions, intersections or complements, with no condition on cardinalities $\longrightarrow$ **Move:** assign each element its Venn region (membership pattern); forbidden regions are removed; count $(\text{allowed regions})^{\#\text{elements}}$

▷ **Cue:** a triple of subsets of an n-set $\longrightarrow$ **Move:** rewrite as an $n \times 3$ binary matrix and count by rows rather than columns

▷ **Cue:** an alternating binomial sum $\sum(-1)^k\binom{n}{k}q^{n-k}$ appearing in a count $\longrightarrow$ **Move:** collapse to $(q-1)^n$ and look for the direct per-element argument it encodes

▷ **Cue:** an answer format offering a prime that should not appear (here 7) $\longrightarrow$ **Move:** check whether a partial count (one condition only) produces that prime — it signals the distractor

1985 A2

1985 A2. Let T be an acute triangle. Inscribe a rectangle R in T with one side along a side of T. Then inscribe a rectangle S in the triangle formed by the side of R opposite the side on the boundary of T, and the other two sides of T, with one side along the side of R. For any polygon X, let $A(X)$ denote the area of X. Find the maximum value, or show that no maximum exists, of $\frac{A(R)+A(S)}{A(T)}$, where T ranges over all triangles and R, S over all rectangles as above.

Fields: geometry (similar triangles), optimisation (quadratic / QM–AM) **Position:** A2 — early, routine once set up **Answer:** The maximum exists and equals $\frac{2}{3}$. With the heights of R, S and the leftover top triangle equal to x, y, z times the altitude ($x+y+z=1$), the ratio is $1-(x^2+y^2+z^2)$, maximised at $x=y=z=\frac{1}{3}$, for every acute T.

Layer 1 THE WHAT — surface analysis

The natural attempt. You set up coordinates. Put the side of T carrying R on the x-axis: $B=(0,0)$, $C=(b,0)$, $A=(p,h)$ with $0<p<b$ (the triangle is acute, so the foot of the altitude from A lies strictly inside BC). Let R have height s, $0<s<h$. The horizontal line at height s meets AB and AC in points P, Q; triangle APQ is similar to ABC with ratio $(h-s)/h$, so

$PQ = b(h-s)/h$ and

$$A(R) = \frac{b\,s(h-s)}{h}.$$

Now S sits on PQ with height t, $0 < t < h-s$; the same computation in the triangle of altitude $h-s$ gives width $b(h-s-t)/h$ and $A(S) = b\,t(h-s-t)/h$. Dividing by $A(T) = bh/2$ and writing $x = s/h$, $y = t/h$,

$$\frac{A(R)+A(S)}{A(T)} = 2\big[x(1-x)+y(1-x-y)\big] =: F(x,y), \qquad x,y>0,\ x+y<1.$$

Then calculus: $F_x = 2(1-2x-y)$, $F_y = 2(1-x-2y)$, both zero at $x = y = \frac{1}{3}$, where $F = \frac{2}{3}$. You then owe a justification that this critical point is a global maximum on the open triangle — a Hessian check ($\left(\begin{smallmatrix}-4 & -2\\ -2 & -4\end{smallmatrix}\right)$, negative definite, so F is concave) or a boundary analysis. This works. It costs a coordinate set-up, two-variable calculus and a second-order check, and at the end you have a number with no idea why it is $\frac{2}{3}$.

A cheaper-looking natural attempt fails: be greedy. The largest rectangle in a triangle is half its area (at $x = \frac{1}{2}$), so take R of area $\frac{1}{2}A(T)$; the triangle above has area $\frac{1}{4}A(T)$, and the best S in it is half of that. Total $\frac{1}{2}+\frac{1}{8} = \frac{5}{8} < \frac{2}{3}$. Greedy is wrong: making R as big as possible starves S.

The model chain. Let h be the altitude of T onto the side carrying R, with foot D; since T is acute, D lies strictly inside that side and the altitude splits T into two right triangles U (left) and V (right), with $A(U)+A(V) = A(T)$.

Stack R and S as described; let their heights be xh and yh, and let the triangle remaining above S have height zh, so $x,y,z>0$ and $x+y+z = 1$. Conversely any such x,y,z give a valid configuration: the horizontal lines at heights xh and $(x+y)h$ cut the two slanted sides, and dropping perpendiculars from the cut points lands inside the lower segment because the base angles are acute.

Now look at what is *not* covered by R and S. It consists of six right triangles: on the left, the triangle beside R, the triangle beside S, and the left half of the top triangle; symmetrically three on the right. The three left ones each have a leg along a line parallel to AD and a hypotenuse along AB, so each is similar to U, with altitudes xh, yh, zh; hence their areas are $x^2A(U)$, $y^2A(U)$, $z^2A(U)$. Likewise on the right with V. Therefore

$$A(T) - A(R) - A(S) = (x^2+y^2+z^2)\big(A(U)+A(V)\big) = (x^2+y^2+z^2)A(T),$$

$$\frac{A(R)+A(S)}{A(T)} = 1-(x^2+y^2+z^2).$$

Finally, with $x+y+z = 1$,

$$3(x^2+y^2+z^2) = (x+y+z)^2+(x-y)^2+(y-z)^2+(z-x)^2 \ \geq\ 1,$$

with equality iff $x = y = z = \frac{1}{3}$. So the ratio is at most $1-\frac{1}{3} = \frac{2}{3}$, attained when R and S each have height one third of the altitude. The bound and the equality case hold for every acute T, so the maximum over all T, R, S exists and equals $\frac{2}{3}$. (Check against the coordinate formula: $1-x^2-y^2-(1-x-y)^2 = 2[x(1-x)+y(1-x-y)]$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to "let the heights of R and S be xh and yh". The natural attempt then computes the areas of the rectangles; the model computes the area of everything else. The split is the line "look at what is not covered by R and S".

> *Rectangles inside a triangle have awkward areas — width times height, and the width depends on the height. But what is left over is made of triangles similar to the original, and similar figures have area proportional to the square of a length. Let me measure the waste, not the rectangles.*

> *The leftover heights add up to the whole altitude. So I am minimising a sum of squares with a fixed sum: equal parts, one third each.*

> *The greedy choice (biggest R first) is not obviously optimal, because R's size decides how much room S has. Two coupled choices — I must optimise them jointly.*

Layer 3 THE WHY

Why is "count the complement" the natural move? Because of what a triangle is under a horizontal cut. A line parallel to a side of a triangle cuts off a triangle similar to the original; that is the definition-level fact (Thales, equal angles) behind every width computation in the natural attempt. Similarity gives lengths proportional to the ratio and areas proportional to its *square*. The rectangles are not similar to anything, so their areas come out as products $x(1-x)$ that must be expanded and fought with. The corner triangles are all similar to the two halves U, V of T, so their areas are pure squares x^2, y^2, z^2 times fixed constants. Once a figure is carved into "good" pieces (similar copies) and "bad" pieces (the rectangles), you compute whichever family is good and get the other by subtraction. That is the same instinct as counting a complement in combinatorics.

Kedlaya, Poonen and Vakil's solution opens by splitting T along its altitude into U and V. Here is where that comes from: the altitude is exactly the line along which every corner triangle has its vertical leg, so it is the line that makes every corner triangle a scaled copy of U or of V. Without it, the left and right corner pieces of a single rectangle form a triangle similar to T only after you slide them together; with it, the similarity is immediate. It also explains the hypothesis "acute": the foot of the altitude must lie inside the base for U and V to be honest pieces of T, and for the rectangles to sit on the base at all.

Why is $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ the answer? The constraint $x+y+z=1$ is forced by the geometry — the three stacked heights exhaust the altitude — and $x^2+y^2+z^2$ under a fixed sum is smallest at equal parts. This is the most basic inequality there is: the identity $3\sum x^2 = (\sum x)^2 + \sum(x-y)^2$ says "the sum of squares exceeds its minimum by exactly the spread". The same computation with $n-1$ stacked rectangles gives $1-\frac{1}{n}$, which is how the book generalises it; with one rectangle, $n=2$, it recovers the classical fact that the largest inscribed rectangle is half the triangle. So $\frac{2}{3}$ is not a coincidence: it is $1-\frac{1}{3}$, three strips of equal height.

Finally, why does greedy fail? Because the objective is a function of the *partition* (x,y,z), and greedy chooses $x=\frac{1}{2}$ without regard to the other two parts. Greedy's partition is $(x,y,z) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, with sum of squares $\frac{1}{4}+\frac{1}{16}+\frac{1}{16}=\frac{3}{8}$, which is exactly $1-\frac{5}{8}$. The complement formula makes the failure of greedy visible at a glance: greedy makes one part large, and squares punish large parts.

Layer $+\alpha$ BACKTESTING

1. The first cue is "inscribe a rectangle ... with one side along a side of T" repeated in a smaller triangle. Write down: the triangle above each rectangle is similar to T; areas scale as squares of heights.

2. It suggests parametrising by heights as fractions of the altitude and either (a) computing the rectangle areas and optimising a quadratic, or better (b) computing the complementary similar triangles and minimising a sum of squares under a fixed sum.

3. Hesitation occurs after reaching $F(x,y) = 2[x(1-x)+y(1-x-y)]$: is the critical point really the global maximum? Reinforce it away by introducing $z = 1-x-y$ and seeing $F = 1-(x^2+y^2+z^2)$, where the minimum is a one-line identity.
4. The trap is greedy: maximising R first gives $\frac{5}{8}$. A second trap is thinking the shape of T matters; it does not, which you only see once the ratio is written without b, h, p.
5. Whenever a region is cut into pieces some of which are similar copies of a fixed shape, compute the similar pieces (areas are squares of scale factors) and subtract; do not compute the awkward pieces directly.
6. Reusable rule: an objective equal to "whole minus a sum of squares of parts whose lengths add to a constant" is maximised at equal parts. More generally, when a quantity looks like a messy product, check whether its complement is a clean sum of squares.

▷ **Cue:** a line parallel to a side of a triangle, or rectangles stacked on a side of a triangle $\longrightarrow$ **Move:** the cut-off pieces are similar to the triangle (or its altitude halves); write their areas as (scale factor)^2 times a fixed area
▷ **Cue:** area of irregular pieces (rectangles) inside a figure made of similar triangles $\longrightarrow$ **Move:** compute the complement — the similar triangles — and subtract
▷ **Cue:** minimise $x_1^2+\cdots+x_n^2$ subject to $x_1+\cdots+x_n$ fixed $\longrightarrow$ **Move:** equal parts, via $n\sum x_i^2 = (\sum x_i)^2 + \sum_{i<j}(x_i-x_j)^2$
▷ **Cue:** two sequential choices where the first constrains the second $\longrightarrow$ **Move:** do not optimise greedily; parametrise both and optimise jointly (or optimise the second for fixed first, then optimise the first)
▷ **Cue:** "T ranges over all triangles" in an extremal geometry problem $\longrightarrow$ **Move:** expect the ratio to be scale- and shape-invariant; normalise lengths by the altitude and look for the answer independent of T

1985 A3

1985 A3. Let d be a real number. For each integer $m \geq 0$, define a sequence $\{a_m(j)\}$, $j = 0, 1, 2, \ldots$ by the condition

$$a_m(0) = d/2^m,$$
$$a_m(j+1) = (a_m(j))^2 + 2a_m(j), \qquad j \geq 0.$$

Evaluate $\lim_{n\to\infty} a_n(n)$.

Fields: analysis (limits), recurrences (conjugation to squaring) **Position:** A3 — mid-paper, short once the substitution is seen **Answer:** $e^d - 1$. Since $a_m(j+1)+1 = (a_m(j)+1)^2$, we get $a_n(n) = \left(1+\frac{d}{2^n}\right)^{2^n} - 1 \to e^d - 1$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. Write $\varepsilon = d/2^n$, so $a_n(0) = \varepsilon$, and iterate $a \mapsto a^2 + 2a$:

$$a_n(1) = \varepsilon^2 + 2\varepsilon, \qquad a_n(2) = (\varepsilon^2+2\varepsilon)^2 + 2(\varepsilon^2+2\varepsilon) = \varepsilon^4 + 4\varepsilon^3 + 6\varepsilon^2 + 4\varepsilon, \ \ldots$$

The polynomials grow in degree (2^j after j steps) and you have to iterate n times, with $n \to \infty$. So you try to approximate instead. ε is tiny, so linearise: $a^2 + 2a \approx 2a$ near 0. Then $a_n(j) \approx 2^j\varepsilon$ and

$$a_n(n) \approx 2^n \cdot \frac{d}{2^n} = d.$$

The guess $\lim a_n(n) = d$ is wrong. The linearisation throws away a^2 at every step, and although each discarded term is tiny, there are n steps, and by the late steps $a_n(j)$ is no longer tiny — it

is of order d. The errors accumulate to a finite amount. (Test $d = 1$ numerically with $n = 10$: $a_{10}(10) = (1 + 2^{-10})^{1024} - 1 \approx 1.7169$, not 1.) A student who pushes on to second order finds corrections of size $d^2/2$, then $d^3/6$, and may guess the exponential series from them — but that is guessing, not proof.

Had you looked harder at $a_n(2) = \varepsilon^4 + 4\varepsilon^3 + 6\varepsilon^2 + 4\varepsilon$, you would have seen the coefficients $1, 4, 6, 4$: row four of Pascal's triangle with the final 1 missing. That is $(1 + \varepsilon)^4 - 1$. The computation was trying to tell you the answer.

The model chain. Add 1 to both sides of the recurrence:

$$a_m(j+1) + 1 = a_m(j)^2 + 2a_m(j) + 1 = \big(a_m(j) + 1\big)^2.$$

So $b_m(j) := a_m(j) + 1$ satisfies $b_m(j+1) = b_m(j)^2$, and by induction on j,

$$b_m(j) = b_m(0)^{2^j} = \Big(1 + \frac{d}{2^m}\Big)^{2^j}.$$

Taking $m = j = n$ and writing $N = 2^n$,

$$a_n(n) = \Big(1 + \frac{d}{N}\Big)^N - 1.$$

It remains to show $(1 + d/N)^N \to e^d$ as $n \to \infty$ (so $N \to \infty$ through powers of 2). If $d = 0$ every term is 0 and the limit is $0 = e^0 - 1$. Otherwise, for n large enough that $N > |d|$, we have $1 + d/N > 0$ and may take logarithms:

$$\ln\Big(1 + \frac{d}{N}\Big)^N = N \ln\Big(1 + \frac{d}{N}\Big) = d \cdot \frac{\ln(1+u)}{u}\Big|_{u=d/N}.$$

As $n \to \infty$, $u = d/N \to 0$ with $u \neq 0$, and $\ln(1+u)/u \to 1$ (this is the derivative of $\ln(1+u)$ at $u = 0$). Hence the logarithm tends to d, and since exp is continuous, $(1 + d/N)^N \to e^d$. Therefore

$$\lim_{n\to\infty} a_n(n) = e^d - 1.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths begin by writing out the first iterate $a^2 + 2a$. The natural attempt then either expands further or linearises; the model completes the square. The split is the single line $a^2 + 2a = (a+1)^2 - 1$.

> *$a^2 + 2a$ is one short of a perfect square. If I shift by 1, the map becomes pure squaring, and I know exactly what n squarings do: raise to the power 2^n.*

> *$\big(1 + \frac{d}{2^n}\big)^{2^n}$ — one plus something over N, all to the N. That is the compound-interest limit; the answer is e^d, and I only need to justify it for real d of either sign.*

Layer 3 THE WHY

Why should you look for a shift that turns the map into squaring? Because iterating a map is the one operation where a change of variables pays off every single step. If $f = h^{-1} \circ g \circ h$ for some invertible h, then

$$f^{\circ n} = h^{-1} \circ g^{\circ n} \circ h,$$

since the inner $h \circ h^{-1}$ pairs cancel. So the whole difficulty of an iteration question is finding a coordinate in which the map is one you can iterate by hand. The maps you can iterate by hand

are few: $x \mapsto cx$ (giving $c^n x$), $x \mapsto x + c$ (giving $x + nc$), $x \mapsto x^k$ (giving x^{k^n}), and trigonometric ones like $x \mapsto 2x^2 - 1$ (which is $\cos\theta \mapsto \cos 2\theta$). Every quadratic map is affinely conjugate to one of the form $y \mapsto y^2 + c$, and the standard values of c for which the iterates have an elementary closed form are $c = 0$ (pure squaring) and $c = -2$ (the Chebyshev map, via $y = 2\cos\theta$). So for $f(x) = x^2 + 2x$ the first question is whether it is conjugate to pure squaring, and completing the square is exactly the computation that answers it and finds h. Here $h(x) = x + 1$. It is not a trick; it is the first thing to try on any quadratic recurrence $x_{j+1} = x_j^2 + px_j + q$, and it succeeds precisely when $q = p^2/4 - p/2$ — which the problem-setter arranged.

Kedlaya, Poonen and Vakil's solution begins with $a_m(j+1) + 1 = (a_m(j) + 1)^2$ and gives no reason. The reason is above: the only quadratic you can iterate in closed form is a pure square, so you complete the square and read off the shift.

Why does e appear? Look at how the problem couples the two indices. The start value is divided by 2^m; the number of squarings is j; and the limit sets $m = j = n$. Squaring n times multiplies exponents by 2^n; dividing by 2^n shrinks the increment by 2^n. The two cancel to leave $(1 + x/N)^N$, the defining limit of e^x (Bernoulli's compound interest: split a growth rate d into N equal compounding steps). The problem is a disguised definition of the exponential.

That also explains precisely why the linearised guess gives d: replacing $(1 + d/N)^N$ by $1 + N \cdot d/N = 1 + d$ keeps only the first two terms of the binomial expansion. The true limit is $1 + d + d^2/2! + \cdots$, and the terms you dropped add up to $e^d - 1 - d$. Linearisation is legitimate for a fixed number of steps; it fails when the number of steps grows while the perturbation shrinks, because the product "steps × perturbation" stays finite and higher-order terms survive.

On rigour: the logarithm argument needs $1 + d/N > 0$, which fails for small n when $d \le -1$. That is harmless because a limit ignores finitely many terms, but you must say so. Note also that the problem asks only for $N \to \infty$ through powers of 2; a sequential limit along a subsequence follows from the limit $\lim_{u\to 0} \ln(1+u)/u = 1$.

Layer $+\alpha$ BACKTESTING

1. The first cue is the recurrence $a_{j+1} = a_j^2 + 2a_j$: a quadratic map to be iterated many times. Write down $a_{j+1} + 1 = (a_j + 1)^2$ at once.

2. It suggests conjugating to the pure squaring map, obtaining $a_m(j) = (1 + d/2^m)^{2^j} - 1$, and then recognising the limit $(1 + x/N)^N \to e^x$.

3. Hesitation occurs at the limit: is it legitimate for negative d, and what about terms where $1 + d/2^n \le 0$? Reinforce by noting that only the tail matters, and for $2^n > |d|$ everything is positive and the logarithm argument goes through.

4. The trap is linearisation: $a^2 + 2a \approx 2a$ gives the answer d. It fails because the number of steps grows as the starting value shrinks, so neglected higher-order terms accumulate to a finite amount.

5. Whenever a recurrence is iterated a growing number of times from a shrinking starting value, expect a $(1 + x/N)^N$-type limit and look for the exact closed form before approximating. And whenever a quadratic recurrence appears, complete the square first.

6. Reusable rule: to iterate a map, find a change of variable h that turns it into multiplication, addition or powering; then $f^{\circ n} = h^{-1} \circ g^{\circ n} \circ h$. Never linearise across a number of steps that tends to infinity without checking the accumulated error.

▷ **Cue:** a quadratic recurrence $x_{j+1} = x_j^2 + px_j + q$ ⟶ **Move:** complete the square; if it becomes $(x_j + c)^2 - c$, shift by c to get pure squaring $y_{j+1} = y_j^2$

▷ **Cue:** a map to be iterated n times ⟶ **Move:** look for a conjugacy $f = h^{-1}gh$ with g one of cx, $x + c$, x^k, $2x^2 - 1$; then $f^{\circ n} = h^{-1}g^{\circ n}h$

▷ **Cue:** an expression $(1 + x/N)^N$, or a start value divided by 2^n combined with n squarings ⟶ **Move:** the limit is e^x; prove it via $N\ln(1 + x/N) \to x$ and continuity of exp

▷ **Cue:** expanded iterates whose coefficients look like $1, 4, 6, 4$ ⟶ **Move:** recognise a binomial row: the iterate is $(1 + \varepsilon)^k - 1$

▷ **Cue:** a tempting linearisation over a growing number of steps ⟶ **Move:** suspect it: the neglected terms accumulate; find the exact form instead

1985 A4

1985 A4. Define a sequence $\{a_i\}$ by $a_1 = 3$ and $a_{i+1} = 3^{a_i}$ for $i \geq 1$. Which integers between 00 and 99 inclusive occur as the last two digits in the decimal expansion of infinitely many a_i?

Fields: number theory (modular arithmetic, orders, power towers) **Position:** A4 — mid-paper; a two-minute problem once you let the modulus shrink at each level of the tower, a swamp if you keep working mod 100 **Answer:** Only 87. The terms are $a_1 = 3$, $a_2 = 27$, and $a_i \equiv 87 \pmod{100}$ for every $i \geq 3$; so 03 and 27 each occur once and 87 is the only pair of last digits that occurs infinitely often.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. $a_1 = 3$, $a_2 = 3^3 = 27$, $a_3 = 3^{27}$. That last one is a 13-digit number, but you only want it mod 100, so you reduce as you go: $3^5 = 243 \equiv 43$, $3^{10} \equiv 43^2 = 1849 \equiv 49$, $3^{20} \equiv 49^2 = 2401 \equiv 1$. So $3^{27} = 3^{20} \cdot 3^7 \equiv 3^7 = 2187 \equiv 87$. Good: a_3 ends in 87.

Now $a_4 = 3^{a_3} = 3^{3^{27}}$. You cannot write the exponent down, so you look for structure. You have found $3^{20} \equiv 1 \pmod{100}$, so 3^k mod 100 repeats with period 20 in k. You write out the whole cycle: $03, 09, 27, 81, 43, 29, 87, 61, 83, 49, 47, 41, 23, 69, 07, 21, 63, 89, 67, 01$. To place a_4 in this cycle you need a_3 mod 20. And here the attempt stalls in one of two ways. Either you try to get a_3 mod 20 from the fact that $a_3 \equiv 87 \pmod{100}$ (which gives $a_3 \equiv 7 \pmod{20}$ — lucky, and it works for this step), then worry that for a_5 you need a_4 mod 20, and that you are not sure your last-two-digits information is enough at every level; or you use Euler's theorem with $\varphi(100) = 40$, and then need a_3 mod 40, which needs a_2 mod $\varphi(40) = 16$, which needs a_1 mod 8, and you are running a chain of four reductions with larger numbers than necessary. Both routes can be pushed through. Neither tells you in advance that the sequence *must* settle, so you finish unsure whether 87 is a fixed point or a stop on a longer cycle you have not yet seen.

The model chain. *Principle.* If $3^d \equiv 1 \pmod{m}$, then 3^a mod m depends only on a mod d, because $3^{a+d} = 3^a \cdot 3^d \equiv 3^a$.

Orders needed.

- mod 4: $3 \equiv -1$, so $3^a \equiv (-1)^a$; the order is 2.
- mod 20: $3^4 = 81 \equiv 1$; the order is 4 ($3^2 = 9 \not\equiv 1$).
- mod 100: $3^{20} \equiv 1$ (computed above); the order is 20, since $3^4 = 81$ and $3^{10} \equiv 49$ are not 1.

So the moduli descend $100 \to 20 \to 4 \to 2$.

Bottom level. Every a_i is a power of 3, hence odd.

Climb. For $i \geq 2$: $a_i = 3^{a_{i-1}}$ with a_{i-1} odd, so $a_i \equiv (-1)^{\text{odd}} = -1 \equiv 3 \pmod{4}$. (Also $a_1 = 3 \equiv 3 \pmod{4}$.)

For $i \geq 2$: $a_i = 3^{a_{i-1}}$ with $a_{i-1} \equiv 3 \pmod{4}$, and the order of 3 mod 20 is 4, so $a_i \equiv 3^3 = 27 \equiv 7 \pmod{20}$.

For $i \geq 3$: $a_i = 3^{a_{i-1}}$ with $a_{i-1} \equiv 7 \pmod{20}$ (valid since $i - 1 \geq 2$), and the order of 3 mod 100 is 20, so $a_i \equiv 3^7 = 2187 \equiv 87 \pmod{100}$.

Conclusion. $a_1 = 03$, $a_2 = 27$, and every later term ends in 87. The only two-digit ending that occurs for infinitely many i is 87.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree up to $3^{20} \equiv 1 \pmod{100}$. They split at the next sentence. The natural attempt says "so I need a_3 mod 20" and then tries to get it from what it already knows mod 100, level by level, forwards. The model says "so the question about a_i mod 100 *is* a question about a_{i-1} mod 20, which is a question about a_{i-2} mod 4, which is a question about parity" — and solves from the bottom up.

> *The exponent only matters modulo the order of* 3. *So I never need a_{i-1} mod* 100*; I need it mod* 20*, and for that I need the level below mod* 4*, and below that only parity. The modulus shrinks every time I go down a level, so this must bottom out.*

> *Once the bottom level is constant (every term is odd), each level above is constant too, from some index on. So the sequence is eventually constant mod* 100*, and I only have to compute the constant once.*

Layer 3 THE WHY

Why must the modulus change as you descend the tower? Go back to what an exponent is. 3^a means a factors of 3. If d factors of 3 multiply to 1 mod m, then any block of d factors can be deleted without changing the residue. So the exponent is only ever read modulo d. That is not a theorem about towers; it is the definition of multiplication applied to a repeating block. Therefore, for a tower $3^{3^{3^{\cdots}}} \bmod m$, the top of the tower is irrelevant beyond its residue mod d, and the problem literally *becomes* a problem about the next term at modulus d. The recursion $a_{i+1} = 3^{a_i}$ is shifting you down a level, and the modulus has to shift with it.

Why must it terminate? The order d of 3 mod m is at most the number of residues coprime to m, which is less than m for $m > 1$. A strictly decreasing sequence of positive integers ends, here at 2 (and below that, at 1, where everything is 0). Kedlaya–Poonen–Vakil state the general result: for any c, the tower $c, c^c, c^{c^c}, \ldots$ is eventually constant mod any n. The proof is this descent. So the answer to "which endings occur infinitely often" was always going to be a single number; the problem's phrasing ("which integers … occur … infinitely many") invites you to expect a set, and the structure says the set has one element.

Why choose the order and not φ? Kedlaya's main solution uses Euler's theorem, $\varphi(100) = 40$, and descends $100 \to 40 \to 16 \to 8 \to 2$. That is correct and requires five residues. Their remark replaces φ by Carmichael's λ, the least universal exponent, giving $100 \to 20 \to 4 \to 2$. For a single base you can do better still by using the actual order of 3, which you have already computed on your way to a_3. The cheapest invariant is the one you already found: when the computation hands you $3^{20} \equiv 1$, use 20, not a textbook 40.

Why the natural attempt felt unsafe. Working forwards ("I know a_3 mod 100, now get a_4") needs, at every step, the previous term modulo the order — and whether mod-100 information suffices for mod 20 is a divisibility question ($20 \mid 100$, so yes) that you are answering by luck rather than design. Working backwards from parity makes the dependency explicit, so you know which information each level needs before you compute it.

Layer $+\alpha$ BACKTESTING

1. The cue is $a_{i+1} = 3^{a_i}$ with a question about last digits: a power tower modulo 100. Write down immediately "3^a mod 100 depends on a mod $\operatorname{ord}_{100}(3)$" and compute that order (20).
2. Descent through orders: $100 \to 20 \to 4 \to 2$. Solve the bottom level (parity) and climb.
3. Hesitation comes at $a_4 = 3^{3^{27}}$, where the exponent cannot be written. Reinforce it away with the principle "I never need the exponent, only its residue mod the order".
4. Two traps. The phrasing asks for a set, which suggests a cycle of several endings; the descent shows there is only one. And 03 and 27 occur, so a careless reader lists them; they occur once each, not infinitely often.

5. Any iterated exponential, any $x^{x^{\cdots}}$, any "last digits of a huge power": the modulus of the exponent is the order of the base, and each level down uses a smaller modulus.

6. Reusable rule: when an operation is periodic, the input matters only modulo the period. Nested operations therefore reduce modulo nested periods, which shrink, so nested periodic processes are eventually constant. Solve them from the innermost level outwards.

▷ **Cue:** a power tower $a_{i+1} = c^{a_i}$ asked modulo m ⟶ **Move:** reduce the exponent modulo the order of c mod m; recurse on the smaller modulus

▷ **Cue:** the problem asks which values occur infinitely often in a sequence mod m ⟶ **Move:** look for eventual constancy or periodicity; discard the initial terms explicitly

▷ **Cue:** you have found $c^d \equiv 1 \pmod m$ during a computation ⟶ **Move:** use that d as the exponent modulus instead of $\varphi(m)$

▷ **Cue:** a chain of reductions whose bottom level is "every term is odd" ⟶ **Move:** the chain is constant from the start at the bottom, so every level above is eventually constant

▷ **Cue:** an exponent too large to write ⟶ **Move:** you never need it; you need its residue modulo a period

1985 A5

1985 A5. Let $I_m = \int_0^{2\pi} \cos(x)\cos(2x)\cdots\cos(mx)\,dx$. For which integers m, $1 \le m \le 10$ is $I_m \neq 0$?

Fields: analysis (Fourier / orthogonality), combinatorics (signed sums) **Position:** A5 — late-paper; routine once the cosines are written as exponentials, genuinely hard to finish if you only use trigonometric identities, because proving *non*-vanishing needs a positivity argument **Answer:** $I_m \neq 0$ exactly for $m = 3, 4, 7, 8$, i.e. $m \equiv 0$ or $3 \pmod 4$. In fact $I_m = 2\pi N_m / 2^m$, where N_m is the number of sign choices with $\pm 1 \pm 2 \pm \cdots \pm m = 0$; for instance $I_3 = \pi/2$, $I_4 = \pi/4$, $I_7 = \pi/8$, $I_8 = 7\pi/64$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You know product-to-sum: $\cos A \cos B = \frac{1}{2}[\cos(A-B) + \cos(A+B)]$. So you start multiplying. $\cos x \cos 2x = \frac{1}{2}(\cos x + \cos 3x)$. Multiply by $\cos 3x$: $\frac{1}{4}(\cos 2x + \cos 4x + 1 + \cos 6x)$. The constant $\frac{1}{4}$ integrates to $\frac{2\pi}{4} = \frac{\pi}{2}$ and the rest integrate to 0, so $I_3 = \pi/2 \neq 0$. For $m = 1, 2$ there is no constant term, so $I_1 = I_2 = 0$. Then $m = 4$: multiply by $\cos 4x$ and you get eight terms. By $m = 10$ you would be tracking 512 terms. This is zombie maths, and you know it.

A stronger student looks for symmetry. Substitute $x \mapsto \pi - x$. Since $\cos(k(\pi - x)) = (-1)^k \cos kx$, the integrand becomes $(-1)^{1+2+\cdots+m}$ times itself, and the substitution maps one full period to another, so

$$I_m = (-1)^{m(m+1)/2} I_m.$$

When $m(m+1)/2$ is odd — that is, $m \equiv 1, 2 \pmod 4$ — this forces $I_m = 0$. That settles $m = 1, 2, 5, 6, 9, 10$ in two lines, and it is a genuinely good move.

And here is the wall. For $m = 3, 4, 7, 8$ the symmetry gives $I_m = I_m$, which is no information. You suspect $I_m \neq 0$, but to *prove* an integral of an oscillating function is non-zero you need either to compute it or to show it is a sum of non-negative things, not all zero. The product-to-sum route does compute it, at a cost of 2^{m-1} terms for $m = 8$. The symmetry route does not reach it at all.

The model chain. Write each cosine as $\cos kx = \frac{1}{2}(e^{ikx} + e^{-ikx})$ and expand the product:

$$\cos x \cos 2x \cdots \cos mx = \frac{1}{2^m} \sum_{\epsilon \in \{\pm 1\}^m} e^{i(\epsilon_1 + 2\epsilon_2 + \cdots + m\epsilon_m)x}.$$

For an integer t, $\int_0^{2\pi} e^{itx}\,dx$ equals 2π if $t = 0$ and $\left[e^{itx}/(it)\right]_0^{2\pi} = 0$ otherwise. Therefore

$$I_m = \frac{2\pi}{2^m} N_m, \qquad N_m = \#\{\epsilon \in \{\pm 1\}^m : \epsilon_1 + 2\epsilon_2 + \cdots + m\epsilon_m = 0\}.$$

Every term is non-negative, so $I_m \neq 0 \iff N_m \geq 1 \iff$ some choice of signs makes $\pm 1 \pm 2 \pm \cdots \pm m = 0$.

Necessity. Changing a sign changes the sum by an even number, so every signed sum has the parity of $1 + 2 + \cdots + m = m(m+1)/2$. A zero sum needs $m(m+1)/2$ even, i.e. $4 \mid m(m+1)$. Since m and $m+1$ are coprime and one of them is odd, this means $4 \mid m$ or $4 \mid m+1$: $m \equiv 0$ or $3 \pmod 4$.

Sufficiency. Four consecutive integers satisfy $n - (n+1) - (n+2) + (n+3) = 0$. If $m \equiv 0 \pmod 4$, split $1, \ldots, m$ into blocks of four and sign each block $+\,-\,-\,+$. If $m \equiv 3 \pmod 4$, use $1 + 2 - 3 = 0$ and then blocks of four from 4 onwards. Explicitly: $1 - 2 - 3 + 4 = 0$, $1 + 2 - 3 = 0$, $1 + 2 - 3 + 4 - 5 - 6 + 7 = 0$, $1 - 2 - 3 + 4 + 5 - 6 - 7 + 8 = 0$.

Answer. In $1 \leq m \leq 10$: $I_m \neq 0$ exactly for $m = 3, 4, 7, 8$. (Counting sign patterns gives $N_3 = 2$, $N_4 = 2$, $N_7 = 8$, $N_8 = 14$, hence the values in the vitals; they agree with numerical integration.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the integrand is a product of cosines whose integral over a period is wanted. They split at the first line of work. The natural attempt reaches for a real identity (product-to-sum, or a symmetry $x \mapsto \pi - x$). The model rewrites every cosine as a sum of two exponentials, so that the product becomes a sum of pure frequencies with *positive* coefficients.

> *Integrating over a full period kills every frequency except zero. So the integral is the constant term of the product. Products of exponentials add their frequencies, so I want the product in exponential form: the constant term counts the sign choices with $\pm 1 \pm 2 \cdots \pm m = 0$.*

> *Every coefficient in that expansion is $+2^{-m}$. So nothing can cancel: I_m is non-zero as soon as one signed sum is zero. The non-vanishing half of the problem is a construction, not a computation.*

Layer 3 THE WHY

Why exponentials, and why is this not a trick? Start from what the integral does. $\int_0^{2\pi} \cos(kx)\,dx = 0$ for every integer $k \neq 0$, because the curve spends as much of the period below the axis as above it. So the integral of any trigonometric polynomial over a full period reads off its constant term and throws away everything else. The whole problem is therefore: *what is the constant term of $\prod \cos kx$?* Constant terms of products are computed by adding frequencies, and the object in which frequencies add is the exponential: $e^{iax}e^{ibx} = e^{i(a+b)x}$. A cosine is the average of the two frequencies $\pm k$, which is the definition $\cos kx = \frac{1}{2}(e^{ikx} + e^{-ikx})$ read as a statement about frequencies. So the product of m cosines is the average over all 2^m ways of choosing a sign for each frequency, and the constant term is the proportion of choices summing to 0. Product-to-sum is the same expansion done two factors at a time; the exponential form does all m at once and, crucially, displays every coefficient as positive.

Why positivity matters. The symmetry argument can only ever prove $I_m = 0$, because a symmetry equation $I = \pm I$ says nothing when the sign is $+$. To prove $I \neq 0$ you need a different kind of fact: a sum of non-negative terms with one positive term. The exponential expansion supplies exactly that, and the moment you see it the problem splits into a parity obstruction (necessity) and a block construction (sufficiency) — the standard shape of every "for which n does a signed sum vanish" question.

Why parity is the obstruction. Flipping the sign of k changes the total by $2k$, an even number. So the parity of $\pm 1 \pm \cdots \pm m$ is an invariant, equal to the parity of $1 + \cdots + m$. An invariant that differs from the target's value is the cheapest impossibility proof there is. Note that this parity condition and the $x \mapsto \pi - x$ symmetry give the same condition $m(m+1)/2$ even; they are the same fact seen from the two sides of the Fourier transform, since $x \mapsto \pi - x$ sends e^{itx} to $(-1)^t e^{-itx}$, and every frequency t in the expansion has the parity of $m(m+1)/2$.

Why blocks of four. You want to cancel consecutive integers with signs. Pairs fail ($n - (n+1) = -1$). But two such pairs with opposite orientation cancel: $(n - (n+1)) - ((n+2) - (n+3)) = -1 + 1 = 0$. So four is the shortest self-cancelling block, which is why the answer is periodic mod 4.

Kedlaya–Poonen–Vakil's solution begins "By de Moivre's theorem". Here is where that comes from: the integral is a constant-term extractor, and exponentials are the basis in which constant terms of products are computable.

Layer $+\alpha$ BACKTESTING

1. The cue is $\int_0^{2\pi}$ of a product of $\cos(kx)$ with integer k. Write down $\cos kx = \frac{1}{2}(e^{ikx} + e^{-ikx})$ and "the integral over a period picks out the constant term".

2. Expand into exponentials; convert the integral into a count of sign patterns with $\sum k\epsilon_k = 0$; then parity for necessity and a periodic block construction for sufficiency.

3. Hesitation comes after the symmetry argument has disposed of $m \equiv 1, 2 \pmod 4$ and the remaining cases seem to need a computation. Reinforce it away by asking "are all coefficients positive?" — if yes, non-vanishing reduces to exhibiting one term.

4. The trap is product-to-sum carried to $m = 8$ or $m = 10$: correct, exhausting, and error-prone. A subtler trap is believing the symmetry argument has finished the problem; it only proves the zero cases.

5. Whenever an integral over a full period, or an average over roots of unity, is applied to a product, think "constant term of a product of Laurent polynomials" and expand in exponentials.

6. Reusable rule: to prove a quantity non-zero, express it as a sum of non-negative terms and exhibit one positive term; symmetry arguments can prove vanishing but never non-vanishing. For signed-sum existence questions, parity (or another invariant) gives necessity, and a short self-cancelling block repeated periodically gives sufficiency.

▷ **Cue:** $\int_0^{2\pi}$ of a product of $\cos(kx)$ or $\sin(kx)$ with integer k $\longrightarrow$ **Move:** write each factor in exponentials; the integral is 2π times the constant term

▷ **Cue:** need to prove an integral or sum is non-zero $\longrightarrow$ **Move:** rewrite it as a sum of non-negative terms and find one positive term

▷ **Cue:** does $\pm 1 \pm 2 \pm \cdots \pm m = 0$ have a solution $\longrightarrow$ **Move:** parity of $m(m+1)/2$ for necessity; blocks $n - (n+1) - (n+2) + (n+3) = 0$ for sufficiency

▷ **Cue:** a substitution sends the integrand to $\pm$ itself $\longrightarrow$ **Move:** use it to prove vanishing when the sign is $-$; expect no information when it is $+$

▷ **Cue:** an answer that should be periodic in m $\longrightarrow$ **Move:** find the shortest self-cancelling block; its length is the period

1985 A6

1985 A6. If $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ is a polynomial with real coefficients a_i, then set

$$\Gamma(p(x)) = a_0^2 + a_1^2 + \cdots + a_m^2.$$

Let $F(x) = 3x^2 + 7x + 2$. Find, with proof, a polynomial $g(x)$ with real coefficients such that

(i) $g(0) = 1$, and

(ii) $\Gamma(f(x)^n) = \Gamma(g(x)^n)$

for every integer $n \geq 1$.

Caveat. The archive's source `.tex` writes $F(x)$ in the definition and $f(x)$ in condition (ii); the statement is reproduced verbatim. They denote the same polynomial — the Kedlaya–Poonen–Vakil book prints $f(x) = 3x^2 + 7x + 2$ throughout. The analysis below writes f.

Fields: algebra (polynomials, Laurent polynomials), analysis (Parseval on the unit circle) **Position:** A6 — last problem of the morning; the answer takes one line once you see what Γ really measures, and no amount of coefficient-matching will reach "for every n" **Answer:** $g(x) = 6x^2 + 5x + 1 = (3x + 1)(2x + 1)$. Reason: $\Gamma(p)$ is the constant term of $p(x)p(1/x)$, this is multiplicative in p, and $(x + 2)(x^{-1} + 2) = (2x + 1)(2x^{-1} + 1)$. Other valid answers exist, e.g. $(3x - 1)(2x - 1)$ and $(3x^k + 1)(2x^k + 1)$; uniqueness is not claimed.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute $\Gamma(f) = 2^2 + 7^2 + 3^2 = 62$. You guess g has degree 2 with $g(0) = 1$, say $g = 1 + bx + ax^2$, so condition (ii) at $n = 1$ is $1 + b^2 + a^2 = 62$, i.e. $a^2 + b^2 = 61 = 25 + 36$. Integer candidates: $(a, b) = (\pm 6, \pm 5)$ or $(\pm 5, \pm 6)$.

Now $n = 2$. $f^2 = 9x^4 + 42x^3 + 61x^2 + 28x + 4$, so $\Gamma(f^2) = 81 + 1764 + 3721 + 784 + 16 = 6366$. For $g = 1 + bx + ax^2$, $g^2 = 1 + 2bx + (b^2 + 2a)x^2 + 2abx^3 + a^2x^4$, so

$$\Gamma(g^2) = 1 + 4b^2 + (b^2 + 2a)^2 + 4a^2b^2 + a^4.$$

With $(a, b) = (6, 5)$: $1 + 100 + 1369 + 3600 + 1296 = 6366$. It matches. With $(5, 6)$: $1 + 144 + 2116 + 3600 + 625 = 6486$. It does not. So $g = 6x^2 + 5x + 1$ survives two tests.

And here is the wall. The problem says *for every* $n \geq 1$. You can check $n = 3$ (751520 both sides, if you have the stamina), but the coefficients of f^n are not given by any formula you can square and sum in closed form, and Γ is not multiplicative: $\Gamma(f^2) = 6366 \neq 62^2$. So there is no way to lift the $n = 1$ check to all n by algebra on Γ itself. You have the right answer and no proof, which on a "find, with proof" problem is close to nothing.

The model chain. *Step 1: rewrite* Γ. For $p(x) = \sum_{i=0}^{m} a_i x^i$ with real a_i, define the Laurent polynomial $\gamma(p)(x) = p(x)\, p(x^{-1})$. Then

$$p(x)p(x^{-1}) = \sum_{i,j} a_i a_j x^{i-j},$$

and the terms with $i = j$ are exactly the x^0 terms, so $\Gamma(p) =$ the coefficient of x^0 in $\gamma(p)$.

Step 2: γ *is multiplicative.* $\gamma(pq)(x) = p(x)q(x)p(x^{-1})q(x^{-1}) = \gamma(p)(x)\gamma(q)(x)$. In particular $\gamma(p^n) = \gamma(p)^n$.

Step 3: find a factor with the same γ. Factor $f(x) = (3x + 1)(x + 2)$. Then

$$\gamma(x + 2) = (x + 2)(x^{-1} + 2) = 5 + 2x + 2x^{-1} = (2x + 1)(2x^{-1} + 1) = \gamma(2x + 1).$$

Set $g(x) = (3x + 1)(2x + 1) = 6x^2 + 5x + 1$. By multiplicativity, $\gamma(g) = \gamma(3x + 1)\gamma(2x + 1) = \gamma(3x + 1)\gamma(x + 2) = \gamma(f)$.

Step 4: conclude. For every $n \geq 1$, $\gamma(g^n) = \gamma(g)^n = \gamma(f)^n = \gamma(f^n)$, and taking the x^0 coefficient gives $\Gamma(g^n) = \Gamma(f^n)$. Finally $g(0) = 1$.

(Checked numerically: $\Gamma(f^n) = \Gamma(g^n) = 62,\ 6366,\ 751520,\ 93950470, \ldots$ for $n = 1, \ldots, 7$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from $\Gamma(f) = 62$. They split at the phrase "for every integer $n \geq 1$". The natural attempt treats it as a list of equations to satisfy one at a time. The model reads it as a demand for an *invariant*: some object attached to p that behaves well under powers and determines Γ. Since Γ itself is not multiplicative, the model looks for something multiplicative underneath it.

> *"For every n" means I will never verify this term by term. I need a quantity attached to p that respects multiplication and from which Γ can be read off. A sum of squares of coefficients is $\sum a_i a_j$ over $i = j$ — that is the constant term of $p(x)p(1/x)$, and* that *is multiplicative.*

> *Now I only need g with $g(x)g(1/x) = f(x)f(1/x)$. That product does not change if I replace a factor $x + 2$ by its reversal $2x + 1$. The condition $g(0) = 1$ tells me which factor to reverse: the one whose constant term is* 2.

Layer 3 THE WHY

Why is $p(x)p(1/x)$ the natural object? Go back to the definition. $\Gamma(p) = \sum a_i^2$ is a sum of products $a_i \cdot a_j$ in which the indices agree. To make a product of two polynomials produce $a_i a_j$ with a tag saying whether $i = j$, give a_i the tag x^i and a_j the tag x^{-j}: then $a_i a_j$ carries x^{i-j}, and the diagonal $i = j$ is the constant term. That is how you encode "indices agree" in algebra — the same device as $\sum_k \binom{n}{k}^2 = [x^n](1+x)^n(1+x)^n$, the Vandermonde trick. Nothing is being pulled from a hat; the definition of Γ is a diagonal sum, and diagonal sums are constant terms of $p(x) \cdot p(x^{-1})$.

There is a second, geometric reading, and it explains the answer. Put $x = e^{i\theta}$. For real coefficients, $p(e^{-i\theta}) = \overline{p(e^{i\theta})}$, so $p(x)p(1/x) = |p(e^{i\theta})|^2$, and integrating over θ extracts the constant term:

$$\Gamma(p) = \frac{1}{2\pi} \int_0^{2\pi} \left|p(e^{i\theta})\right|^2 d\theta.$$

This is Parseval's identity (Kedlaya–Poonen–Vakil mention the contour-integral form). So $\Gamma(p)$ depends only on the modulus $|p|$ on the unit circle, and $\Gamma(p^n)$ depends only on $|p|^n$ there. The problem is therefore asking: find g with $|g| = |f|$ on the unit circle. Now the choice of g is forced by geometry. On $|x| = 1$, $|x + 2| = |x|\,|1 + 2\bar{x}| = |1 + 2\bar{x}| = |\overline{1 + 2x}| = |2x + 1|$. In words: replacing a root r by its reflection $1/\bar{r}$ in the unit circle, and rescaling, does not change the modulus on the circle. The roots of f are $-1/3$ and -2; reflect -2 to $-1/2$ and you have g.

Why the condition $g(0) = 1$. Without it, $g = f$ is an answer, and so is $\pm x^k f$. The condition exists to rule out the trivial answer and to force you to change f in a structural way. It also tells you which way to reflect: $f(0) = 2$ comes from the factor $x + 2$, and reversing that factor to $2x + 1$ turns its constant term into 1, while $3x + 1$ already has constant term 1. So the normalisation is a signpost, not a nuisance.

Why coefficient matching could not finish. Γ is a quadratic function of the coefficients and does not respect multiplication, so information about $\Gamma(g)$ does not propagate to $\Gamma(g^2)$. The natural attempt found the right g because two equations happened to pin it down among integer candidates, but a proof "for every n" must pass through a multiplicative object. The general principle: when a problem imposes infinitely many conditions indexed by powers, look for a homomorphism.

A note on what is not proved. The book remarks that $(3x^k + 1)(2x^k + 1)$ and $(3x^k - 1)(2x^k - 1)$ also work and that uniqueness beyond these is open (it would follow from a conjecture of Kuperberg). The problem asks only for one g, so nothing here depends on it.

Layer $+\alpha$ BACKTESTING

1. The cue is the definition $\Gamma(p) = \sum a_i^2$ together with "for every $n \geq 1$". Write down $\Gamma(p) = [x^0]\, p(x)p(x^{-1})$ at once.

2. Multiplicativity of $p \mapsto p(x)p(x^{-1})$ (equivalently $|p|^2$ on the unit circle). Then factor f and look for factor-by-factor substitutes with the same γ: reverse a linear factor.

3. Hesitation comes after $n = 1$ and $n = 2$ have both checked out and there seems to be no way to reach general n. Reinforce it away with: "Γ is not multiplicative, so I must find what is."

4. The trap is the successful numerical check. Solving $a^2 + b^2 = 61$ and testing $n = 2$ produces the right g and invites you to stop without a proof. A secondary trap is trying to make Γ multiplicative, which it is not.

5. Any sum of squares of coefficients, any $\sum_k c_k^2$ for coefficients of a product or power, any "norm of a polynomial": rewrite as the constant term of $p(x)p(1/x)$, or as the mean of $|p|^2$ on the unit circle.

6. Reusable rule: when a condition must hold for every power n, find a multiplicative object from which the condition can be read off, and make the two objects equal before taking powers. For polynomials measured on the unit circle, reflecting a root $r \mapsto 1/\bar{r}$ (reversing a linear factor) is the move that changes the polynomial without changing the measurement.

▷ **Cue:** sum of squares of the coefficients of a polynomial $\longrightarrow$ **Move:** it is the constant term of $p(x)p(1/x)$, equivalently $\frac{1}{2\pi}\int |p(e^{i\theta})|^2\, d\theta$

▷ **Cue:** a condition required for every power p^n $\longrightarrow$ **Move:** find a multiplicative invariant that determines the condition; equate invariants before taking powers

▷ **Cue:** need a different polynomial with the same modulus on $|x| = 1$ $\longrightarrow$ **Move:** reverse a linear factor: $x + c \mapsto cx + 1$ (reflect the root in the unit circle)

▷ **Cue:** a normalisation like $g(0) = 1$ in a "find g" problem $\longrightarrow$ **Move:** it excludes the trivial answer; use it to decide which factor to change

▷ **Cue:** a candidate answer passes the first two numerical checks $\longrightarrow$ **Move:** do not stop; find the structural reason, or you have a guess and not a proof

Putnam 1985 — Paper B

1985 B1

1985 B1. Let k be the smallest positive integer for which there exist distinct integers m_1, m_2, m_3, m_4, m_5 such that the polynomial

$$p(x) = (x - m_1)(x - m_2)(x - m_3)(x - m_4)(x - m_5)$$

has exactly k nonzero coefficients. Find, with proof, a set of integers m_1, m_2, m_3, m_4, m_5 for which this minimum k is achieved.

Fields: algebra (polynomials, real roots), construction + lower bound **Position:** B1 – opening problem of the second session **Answer:** $k = 3$, attained by $\{m_1, \dots, m_5\} = \{-2, -1, 0, 1, 2\}$, since $x(x^2 - 1)(x^2 - 4) = x^5 - 5x^3 + 4x$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start by expanding. With five roots, $p(x) = x^5 - e_1x^4 + e_2x^3 - e_3x^2 + e_4x - e_5$, where e_j are the elementary symmetric functions of the m_i. "Few nonzero coefficients" means "many e_j vanish", so you try to kill them one at a time. $e_5 = m_1m_2m_3m_4m_5 = 0$ forces a root at 0 – fine, put $m_5 = 0$. Now $e_1 = m_1 + m_2 + m_3 + m_4 = 0$, and $e_3 = 0 \dots$ and you are solving a small system of symmetric Diophantine equations. You might stumble on $\{-2, -1, 0, 1, 2\}$ by trial, and then you have an example with three terms. But the problem says "with proof": you now owe the reader an argument that two terms are impossible, and the e_j formulation gives you nothing to hold – it says which coefficients are zero, not why at most three can vanish. So you are left with an example and a guess, and "find, with proof" is half done.

The model chain. Split the problem into its two halves before touching any coefficient: a lower bound ($k \geq 3$) and a construction ($k = 3$ is attained).

Lower bound. The polynomial is monic of degree 5, so the coefficient of x^5 is $1 \neq 0$, and $k \geq 1$.

- If $k = 1$, then $p(x) = x^5$, whose only root is 0 (with multiplicity 5): the m_i are not distinct.
- If $k = 2$, then $p(x) = x^5 + ax^r$ with $a \neq 0$ and $0 \leq r \leq 4$. If $r \geq 2$, then $x^2 \mid p(x)$, so 0 is a repeated root: not distinct. If $r = 1$, then $p(x) = x(x^4 + a)$; the equation $x^4 = -a$ has no real solution if $a > 0$ and exactly two ($\pm|a|^{1/4}$) if $a < 0$, so p has at most three real roots, not five. If $r = 0$, then $p(x) = x^5 + a$; since $x \mapsto x^5$ is strictly increasing on $\mathbb{R}$, $x^5 = -a$ has exactly one real solution.

In every case p fails to have five distinct real (let alone integer) roots, so $k \geq 3$.

Construction. Pair the roots symmetrically about 0:

$$x(x - 1)(x + 1)(x - 2)(x + 2) = x(x^2 - 1)(x^2 - 4) = x^5 - 5x^3 + 4x,$$

which has exactly three nonzero coefficients. Hence $k = 3$, attained by $\{-2, -1, 0, 1, 2\}$. (A computer search over all 5-subsets of $\{-6, \dots, 6\}$ confirms the minimum 3 there, and finds it only at sets of the form $\{-b, -a, 0, a, b\}$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree that p is monic of degree 5. They split at the first move. The natural attempt starts by expanding and trying to force coefficients to vanish – it treats the problem as a construction and discovers only afterwards that it owes a lower bound. The model reads "smallest" and "with proof" as a two-part contract: first rule out $k = 1, 2$ by writing down what such a polynomial must *look like* (x^5 or $x^5 + ax^r$) and checking its roots; then build a three-term example by symmetry.

> *"Smallest k" means I owe two things: an example with k terms and a proof that $k-1$ terms is impossible. The impossibility is cheap if I describe the few-term polynomial directly: with two terms it is $x^5 + ax^r$, and I can read its roots off.*

> *To make coefficients vanish, I should not solve for them – I should use a symmetry that kills them wholesale. Roots in pairs $\pm m$ give factors $x^2 - m^2$, which are even; an even polynomial times x is odd, and an odd polynomial has no even-degree terms.*

Layer 3 THE WHY

Why is "describe the sparse polynomial directly" the right move for the lower bound? Because the hypothesis "exactly k nonzero coefficients" is a statement about the *shape* of the polynomial, and a two-term shape is so rigid that its roots can be computed by hand. The conclusion we want to contradict – five distinct real roots – is about roots. So the cheapest bridge is to take the shape, compute the roots, and compare. Working through $e_1, \ldots, e_5$ goes the other way (roots to coefficients) and is the direction where information is lost: many root sets give the same sparsity pattern, and a system of vanishing symmetric functions has no obvious obstruction to solvability. The rule is to reason in the representation where the hypothesis is simplest; here the hypothesis is about coefficients, so write the polynomial in coefficients.

There is a deeper reason why sparsity limits real roots, and it is worth knowing because it turns this problem into a family. Descartes' rule of signs says that a real polynomial whose nonzero coefficients, listed in order of degree, have s sign changes has at most s positive roots. A polynomial with k nonzero terms has at most $k-1$ sign changes, so at most $k-1$ positive roots; applying the same rule to $p(-x)$ gives at most $k-1$ negative roots; and 0 adds at most one more. So k terms support at most $2k-1$ distinct real roots. For five roots, $2k-1 \geq 5$ gives $k \geq 3$ immediately. Kedlaya, Poonen and Vakil record exactly this in their remark, and conclude that n distinct integer roots need $\lceil (n+1)/2 \rceil$ terms, attained by $x\prod_{j=1}^{k-1}(x^2 - j^2)$ or $\prod_{j=1}^{k-1}(x^2 - j^2)$. The case analysis in the model chain is Descartes' rule done by hand for $k = 2$: the only shapes are $x^5 + ax^r$, and each loses roots for a visible reason (a repeated root at 0, or x^4 and x^5 hitting a constant too few times).

Why is symmetric pairing natural for the construction, and not a trick? Because the Descartes count is tight only when roots are balanced: at most $k-1$ positive, at most $k-1$ negative, possibly 0. With $k = 3$ and five roots that forces two positive, two negative and zero. So the lower-bound argument itself tells you where the roots of an extremal example must sit: two on each side of 0, and 0 itself. It does not by itself force mirror images, but mirror images are the placement that makes the coefficients collapse – $x \mapsto -x$ fixes the root set, so $p(-x) = -p(x)$ and every even-degree coefficient dies. Once you have $\pm a, \pm b, 0$, the smallest integer choice $a = 1, b = 2$ and the identity $(x-m)(x+m) = x^2 - m^2$ finish it. The point is that the extremal example is dictated by the proof of the bound; you do not search for it.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "smallest k … find, with proof". On seeing it, write down two headings: "$k \geq$?" and "example with $k =$?".

2. It suggests the extremal template: bound plus construction. For the bound, enumerate the possible shapes with 1 and 2 terms (x^5, $x^5 + ax^r$) and count real roots.
3. Hesitation occurs after finding $x^5 - 5x^3 + 4x$, when the example feels like the answer. It is reinforced away by asking what would stop two terms, and noticing that two-term polynomials are few enough to list.
4. The trap is treating the problem as pure construction: expanding in elementary symmetric functions, finding an example, and submitting no lower bound. A smaller trap is forgetting the case $r = 1$, where $x(x^4 + a)$ does have nonzero real roots, only too few.
5. Any time a problem links the number of terms (or sign changes) of a polynomial to its number of real roots, think Descartes' rule of signs, and expect extremal examples with roots placed symmetrically.
6. Reusable rule: for a minimum, split into bound and construction before calculating; prove the bound in whichever representation makes the hypothesis simplest; and read off the shape of the extremal example from where the bound is tight.

▷ **Cue:** "find the smallest k such that there exist ..., with proof" $\longrightarrow$ **Move:** write two headings, a lower bound and a construction, and do not stop after the construction
▷ **Cue:** a polynomial is constrained to have few nonzero coefficients $\longrightarrow$ **Move:** list the possible shapes ($x^n + ax^r, \dots$) and compute their real roots directly, or apply Descartes' rule of signs to $p(x)$ and $p(-x)$
▷ **Cue:** you want many coefficients of a polynomial to vanish $\longrightarrow$ **Move:** use symmetry: pair roots as $\pm m$ to make an even or odd polynomial, which kills every other coefficient
▷ **Cue:** a monic polynomial with a term ax^r, $r \geq 2$, and no lower terms $\longrightarrow$ **Move:** x^r divides it, so 0 is a repeated root

1985 B2

1985 B2. Define polynomials $f_n(x)$ for $n \geq 0$ by $f_0(x) = 1$, $f_n(0) = 0$ for $n \geq 1$, and

$$\frac{d}{dx} f_{n+1}(x) = (n+1) f_n(x+1)$$

for $n \geq 0$. Find, with proof, the explicit factorization of $f_{100}(1)$ into powers of distinct primes.

Fields: algebra (polynomial recursions), induction; combinatorics in the background (Abel polynomials, Cayley's formula) **Position:** B2 – second problem of the second session, routine once the pattern is found **Answer:** $f_{100}(1) = 101^{99}$, because $f_n(x) = x(x+n)^{n-1}$ for $n \geq 1$ and 101 is prime.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. $f_1'(x) = f_0(x+1) = 1$ and $f_1(0) = 0$ give $f_1(x) = x$. Then $f_2'(x) = 2f_1(x+1) = 2x + 2$, so $f_2(x) = x^2 + 2x$. Then $f_3'(x) = 3f_2(x+1) = 3\big((x+1)^2 + 2(x+1)\big) = 3x^2 + 12x + 9$, so $f_3(x) = x^3 + 6x^2 + 9x$. You carry on in expanded form: $f_4'(x) = 4f_3(x+1)$, which needs $(x+1)^3 + 6(x+1)^2 + 9(x+1)$, and so on. The coefficients grow fast ($f_4 = x^4 + 12x^3 + 48x^2 + 64x$) and no pattern in them jumps out. Alternatively, you tabulate only the numbers the problem asks about: $f_1(1) = 1$, $f_2(1) = 3$, $f_3(1) = 16$, $f_4(1) = 125$. Now a pattern does jump out – $1, 3, 16, 125$ are $2^0, 3^1, 4^2, 5^3$ – and you guess $f_n(1) = (n+1)^{n-1}$, hence $f_{100}(1) = 101^{99}$. The answer is right. But you cannot prove it: the recursion relates f_{n+1} to $f_n(x+1)$ on a whole interval, not to the single number $f_n(1)$, so a statement about $f_n(1)$ alone does not feed itself into an induction. The number-level guess is a dead end for "with proof".

The model chain. Uniqueness. For $n \geq 0$, f_{n+1} is determined by its derivative $(n+1)f_n(x+1)$ and its value $f_{n+1}(0) = 0$, namely $f_{n+1}(x) = (n+1)\int_0^x f_n(t+1)\,dt$. So by induction there is

exactly one sequence of polynomials satisfying the conditions, and any sequence we exhibit that satisfies them is it.

Guess. Factor the first few: $f_1 = x$, $f_2 = x(x+2)$, $f_3 = x(x^2+6x+9) = x(x+3)^2$, $f_4 = x(x^3+12x^2+48x+64) = x(x+4)^3$. Conjecture

$$f_n(x) = x(x+n)^{n-1} \qquad (n \geq 1).$$

Proof by induction on $n \geq 1$. For $n = 1$, $x(x+1)^0 = x = f_1(x)$. Suppose $f_n(x) = x(x+n)^{n-1}$ for some $n \geq 1$, and let $g(x) = x(x+n+1)^n$. Then $g(0) = 0$, and by the product rule

$$g'(x) = (x+n+1)^n + nx(x+n+1)^{n-1} = (x+n+1)^{n-1}\big(x+n+1+nx\big) = (n+1)(x+1)(x+n+1)^{n-1}.$$

On the other side, $(n+1)f_n(x+1) = (n+1)(x+1)\big((x+1)+n\big)^{n-1} = (n+1)(x+1)(x+n+1)^{n-1}$. So g satisfies the same derivative condition and initial value as f_{n+1}; by uniqueness, $f_{n+1} = g$.

Answer. $f_{100}(1) = 1 \cdot (1+100)^{99} = 101^{99}$. Since 101 is prime (it is not divisible by $2, 3, 5, 7$, and $11^2 > 101$), this is the factorization into powers of distinct primes.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths compute f_1, f_2, f_3. They split at what is done with the computed polynomials. The natural attempt either keeps them expanded (no pattern visible) or collapses them to numbers at $x = 1$ (pattern visible but unprovable). The model *factors* them, and guesses at the level of the whole polynomial – the level at which the recursion operates.

> *The question asks for a prime factorization, so the answer is a product. Products come from factored forms, so I should factor f_2 and f_3 rather than expand them.*

> *My guess must be about $f_n(x)$ as a function, not about the number $f_n(1)$, because the recursion needs f_n at $x+1$ for every x. Whatever I induct on must be the thing the recursion eats.*

Layer 3 THE WHY

Why factor? Because the problem tells you to. "The explicit factorization of $f_{100}(1)$ into powers of distinct primes" is the problem-setter announcing that the answer is a clean product, most likely a single prime power. A clean product at $x = 1$ for every n is overwhelmingly likely to come from a clean product in x. So the phrasing is the cue: the form requested for the answer is the form to request of the intermediate objects. Expanded coefficients $1, 12, 48, 64$ are $\binom{3}{j}4^j$ in disguise, but you only see that once you have factored.

Why must the induction hypothesis be at the function level? Look at what the recursion consumes. To compute f_{n+1} anywhere, you integrate $f_n(t+1)$ over $[0, x]$ – you need all values of f_n on an interval. An induction whose hypothesis is "$f_n(1) = (n+1)^{n-1}$" supplies one number, and the step needs a function. This is a general law of inductions: the hypothesis must be strong enough to be the input of the step. When the step involves a shift, derivative or integral, the hypothesis has to be a statement about the whole function. Strengthening the claim is not extra work; it is what makes the step possible.

Why is $x(x+n)^{n-1}$ natural and not a lucky guess? Two structural facts force its shape. First, $f_n(0) = 0$ for $n \geq 1$, so x divides f_n: the factor x is imposed by the initial condition. Second, the recursion shifts the argument by 1 at each step, and n steps of shifting move things by n; so a root drifting to $-n$ is what a repeated shift looks like. Having seen x, $x(x+2)$, $x(x+3)^2$, the remaining degree $n-1$ must be carried somewhere, and the data put it all on $x+n$. Once guessed, the check is a single product-rule computation, and it closes because $x+n+1+nx = (n+1)(x+1)$ – the factor $(x+1)$ that the recursion demands appears on cue.

The polynomials $x(x+n)^{n-1}$ are the *Abel polynomials*, and $f_n(1) = (n+1)^{n-1}$ is Cayley's count of labelled trees on $n+1$ vertices. You do not need either fact here, but they explain why the numbers $1, 3, 16, 125$ looked familiar and why the answer is a perfect prime power: 101 vertices, 101^{99} trees.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "explicit factorization … into powers of distinct primes". On seeing it, write down: the answer is a product, so factor every f_n you compute.

2. It suggests compute–factor–conjecture–induct: find $f_1, \ldots, f_4$, factor them, conjecture a closed form for $f_n(x)$, and prove it by induction using uniqueness of the solution of the recursion.

3. Hesitation occurs when the numbers $1, 3, 16, 125$ reveal $(n+1)^{n-1}$ and it is tempting to stop. It is reinforced away by asking whether that statement can be the hypothesis of an induction; since the step needs $f_n(x+1)$ for all x, it cannot.

4. The trap is proving (or asserting) a pattern in $f_n(1)$ only, which gives the right answer with no valid proof. A second trap is keeping the polynomials expanded and grinding to f_5, f_6 – zombie maths that hides the pattern.

5. When a problem asks for an answer in a specific form (a factorization, a power, a ratio of integers), put the intermediate objects in that form too. When a recursion involves shifts, derivatives or integrals, the induction hypothesis must be about the whole function.

6. Reusable rule: in a recursively defined family, conjecture at the level on which the recursion acts, prove existence by checking the conjectured form satisfies every defining condition, and cite uniqueness to identify it with the family.

▷ **Cue:** answer requested as "a factorization into powers of distinct primes" $\longrightarrow$ **Move:** expect a clean product; factor every intermediate object rather than expanding

▷ **Cue:** recursion of the form $f'_{n+1}(x) = c_n f_n(x+1)$, $f_{n+1}(0) = 0$ $\longrightarrow$ **Move:** uniqueness is automatic; guess a closed form for $f_n(x)$, verify derivative and initial value, conclude by uniqueness

▷ **Cue:** a pattern in the values $f_n(1)$ but the recursion uses f_n at shifted arguments $\longrightarrow$ **Move:** strengthen to a statement about $f_n(x)$ for all x before inducting

▷ **Cue:** $f_n(0) = 0$ given as a condition $\longrightarrow$ **Move:** x is a factor of f_n; divide it out and look at what is left

1985 B3

1985 B3. Let

$$\begin{array}{cccc} a_{1,1} & a_{1,2} & a_{1,3} & \cdots \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

be a doubly infinite array of positive integers, and suppose each positive integer appears exactly eight times in the array. Prove that $a_{m,n} > mn$ for some pair of positive integers (m, n).

Fields: combinatorics (double counting), analysis (divergence of the harmonic series) **Position:** B3 – middle of the second session **Answer:** Proof by contradiction: if $a_{m,n} \le mn$ everywhere, the cells with $mn \le k$ (about $k \ln k$ of them) must all hold values $\le k$, of which there are only $8k$; this fails for large k (first at $k = 2550$).

Layer 1 THE WHAT — surface analysis

The natural attempt. You assume the contrary, $a_{m,n} \le mn$ for all m, n, and look for the contradiction locally. The corner cell has $a_{1,1} \le 1$, so $a_{1,1} = 1$. Row one has $a_{1,n} \le n$, column one $a_{m,1} \le m$.

The cells with $mn \le 2$ are $(1,1),(1,2),(2,1)$, which must hold values in $\{1,2\}$ – but there are sixteen copies of 1 and 2 available, so no contradiction. You try building an array that satisfies the constraint, filling small values near the corner, and it seems to go fine: each value v can be placed in any cell with $mn \ge v$, and there are infinitely many of those. After ten minutes you have a partial array, no contradiction, and a suspicion that the statement is about some subtle clash between rows and columns. There isn't one. The obstruction exists, but it does not appear until k is in the thousands, so no amount of hand-exploration near the corner will find it.

The model chain. Suppose, for contradiction, that $a_{m,n} \le mn$ for all positive integers m, n. Fix a positive integer k and count the set

$$R(k) = \{(m,n) : a_{m,n} \le k\}$$

in two ways.

From the values. Each of the values $1, 2, \dots, k$ appears exactly eight times, and no other value is $\le k$, so $\#R(k) = 8k$.

From the positions. If $mn \le k$, then $a_{m,n} \le mn \le k$, so $(m,n) \in R(k)$. For each $m \in \{1,\dots,k\}$ the number of $n \ge 1$ with $mn \le k$ is $\lfloor k/m \rfloor$, and $m > k$ contributes nothing. Hence

$$\#R(k) \;\ge\; \sum_{m=1}^{k} \left\lfloor \frac{k}{m} \right\rfloor \;>\; \sum_{m=1}^{k} \left(\frac{k}{m} - 1 \right) = k\big(H_k - 1\big), \qquad H_k = 1 + \tfrac{1}{2} + \cdots + \tfrac{1}{k}.$$

(The strict inequality uses $\lfloor y \rfloor > y - 1$.)

Contradiction. Combining, $8k > k(H_k - 1)$, i.e. $H_k < 9$, for every k. But the harmonic series diverges: $H_k > \int_1^{k+1} \frac{dx}{x} = \ln(k+1)$, since $\frac{1}{m} > \int_m^{m+1} \frac{dx}{x}$ for each m. Taking any $k \ge e^9$ gives $H_k > 9$, a contradiction. Hence $a_{m,n} > mn$ for some (m,n).

(For the record: the exact count $\sum_{m \le k} \lfloor k/m \rfloor$ first exceeds $8k$ at $k = 2550$, where it equals $20410 > 20400$, checked by direct computation. So the contradiction lives at values in the thousands – which is why the corner never shows it.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths assume $a_{m,n} \le mn$ everywhere. They split at what gets examined next. The natural attempt examines individual cells and small values, looking for a local clash. The model fixes a threshold k and *counts*: how many cells are forced to hold values $\le k$ (all those under the hyperbola $mn \le k$), versus how many such values exist ($8k$).

> *"Each integer appears exactly eight times" is a statement about how many cells hold small values: exactly $8k$ cells hold a value $\le k$. The inequality $a_{m,n} \le mn$ says certain cells must hold small values. So I should count the cells forced below k and compare with $8k$.*

> *The cells with $mn \le k$ are the lattice points under a hyperbola, and there are about $k \ln k$ of them. That grows faster than $8k$, so the contradiction is at large k, not near the corner. I will stop exploring small cases.*

Layer 3 THE WHY

Why count? Look at the two hypotheses side by side. "Each positive integer appears exactly eight times" is a supply statement: it tells you exactly how many cells can carry values in any range, and nothing about where they sit. "$a_{m,n} \le mn$" is a demand statement: it tells you, cell by cell, an upper limit on the value, and nothing about how many times a value is used. A supply hypothesis and a demand hypothesis can only be made to talk to each other through a count – the number of slots of a given kind. The natural object is therefore a level set, "cells with value $\le k$", because both hypotheses say something about its size: supply says it has

exactly $8k$ elements, demand says it contains every cell with $mn \le k$. That is pigeonhole in its double-counting form, and it is dictated by the shape of the hypotheses, not invented.

Why is the contradiction at large k, and why does that matter for the method? Because the demand side grows like $\sum_{m \le k} k/m \approx k \ln k$ and the supply side like $8k$. The ratio is $\ln k$ against 8, so they cross only when $\ln k$ is around 8, i.e. k in the low thousands. Any exploration of small cases is structurally doomed: the constraint is slack near the corner by design. The moment you see that one side of a comparison is linear and the other is "linear times something unbounded", you know the argument is asymptotic, and you stop drawing grids.

Where does $k \ln k$ come from? Count lattice points under $xy = k$ in the positive quadrant row by row: row m contributes $\lfloor k/m \rfloor$. Summing k/m is k times the harmonic sum, and the harmonic sum is compared with $\int dx/x$ by the rectangle picture (each rectangle of width 1 and height $1/m$ lies above the curve on $[m, m+1]$). So the logarithm is the area under the hyperbola, and the hyperbola is there because the problem's bound is the product mn.

Kedlaya, Poonen and Vakil remark that the explicit rate is not needed: all you need is $H_k \to \infty$. That observation is the real content, and it generalises the problem. Replace "eight" by any fixed number c: the argument gives $H_k < c + 1$ for all k, still impossible. So the eight is a decoy – the problem is true for every finite multiplicity, and fails only if some values may appear infinitely often. What the problem is really testing is whether you recognise $\sum 1/m$ hiding inside a count of lattice points.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "each positive integer appears exactly eight times". On seeing it, write down: for every k, exactly $8k$ cells have value $\le k$.

2. It suggests double counting a level set $R(k) = \{a_{m,n} \le k\}$: once from the multiplicity (supply), once from the positional constraint (demand).

3. Hesitation occurs after checking small cases and finding no clash. It is reinforced away by estimating the growth of both sides: $8k$ against $\sum \lfloor k/m \rfloor \sim k \ln k$, which shows the clash is asymptotic.

4. The trap is local search: building the array near the corner, where the constraint is slack until $k \approx 2550$. A secondary trap is dropping the floor carelessly and writing $\sum \lfloor k/m \rfloor \ge kH_k$ (false); the correct bound is $> k(H_k - 1)$.

5. When one hypothesis limits how often values occur and another limits where they may occur, count a level set both ways. When a count involves pairs with $mn \le k$, expect $k \ln k$ and the harmonic series.

6. Reusable rule: to contradict a pointwise bound, pick a threshold, count the objects forced below it and the objects available below it, and compare growth rates; if one is linear and the other superlinear, the argument is asymptotic and small cases are irrelevant.

▷ **Cue:** "each value appears exactly (or at most) c times" together with an upper bound on each entry $\longrightarrow$ **Move:** fix k and double count the entries $\le k$: at most ck from multiplicity, at least the number of positions forced below k

▷ **Cue:** counting lattice points (m, n) with $mn \le k$ $\longrightarrow$ **Move:** sum $\lfloor k/m \rfloor$ over rows; it exceeds $k(H_k - 1)$ and grows like $k \ln k$

▷ **Cue:** a contradiction that should come from comparing a linear count with a count involving $\sum 1/m$ $\longrightarrow$ **Move:** use divergence of the harmonic series ($H_k > \ln(k+1)$ by comparison with $\int dx/x$) and take k large

▷ **Cue:** small cases show no contradiction in a problem that claims one $\longrightarrow$ **Move:** estimate the growth rates of the competing counts before exploring further; the clash may be at large parameter

1985 B4

1985 B4. Let C be the unit circle $x^2 + y^2 = 1$. A point p is chosen randomly on the circumference C and another point q is chosen randomly from the interior of C (these points are chosen independently and uniformly over their domains). Let R be the rectangle with sides parallel to the x and y-axes with diagonal pq. What is the probability that no point of R lies outside of C?

Fields: geometric probability; coordinate geometry; integration **Position:** B4 — middle of the paper; a two-variable probability that collapses to one integral once you condition on the right variable **Answer:** $4/\pi^2 \approx 0.405$. For $p = (\cos\theta, \sin\theta)$ the good positions of q form the axis-parallel rectangle $|x| \le |\cos\theta|$, $|y| \le |\sin\theta|$, of area $2|\sin 2\theta|$; averaging $2|\sin 2\theta|/\pi$ over θ gives $4/\pi^2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Two random points, so the strong student starts with the one that feels like it carries the geometry: the interior point $q = (a, b)$. Fix q and ask which positions of p on the circle are good.

First the condition. The rectangle with diagonal pq has vertices p, q, and the two "cross" corners $(\cos\theta, b)$ and $(a, \sin\theta)$, where $p = (\cos\theta, \sin\theta)$. A rectangle is the convex hull of its corners and the closed disc is convex, so R stays in the disc exactly when all four corners do. p and q already do, so the condition is

$$\cos^2\theta + b^2 \le 1, \qquad a^2 + \sin^2\theta \le 1, \quad \text{i.e.} \quad |b| \le |\sin\theta|, \;\; |a| \le |\cos\theta|.$$

Now fix q and find the good θ. In the first quadrant, $|\sin\theta| \ge |b|$ means $\theta \ge \arcsin|b|$, and $|\cos\theta| \ge |a|$ means $\theta \le \arccos|a|$. Since $a^2 + b^2 < 1$, $\arcsin|b| < \arccos|a|$, so the good arc in that quadrant has length $\arccos|a| - \arcsin|b|$, and the four quadrants contribute equally by the symmetries $\theta \mapsto \pi - \theta$, $\theta \mapsto -\theta$. Hence

$$P = \mathbb{E}_q\left[\frac{4(\arccos|a| - \arcsin|b|)}{2\pi}\right] = \frac{2}{\pi}\,\mathbb{E}\left[\arccos|a| + \arccos|b| - \tfrac{\pi}{2}\right].$$

Now the student must integrate $\arccos|a|$ over the disc. The marginal density of a for a uniform point of the disc is $\frac{2}{\pi}\sqrt{1-a^2}$ on $[-1,1]$, so

$$\mathbb{E}[\arccos|a|] = \frac{4}{\pi}\int_0^1 \arccos(a)\sqrt{1-a^2}\,da = \frac{4}{\pi}\int_0^{\pi/2} \varphi\sin^2\varphi\,d\varphi \quad (a = \cos\varphi).$$

Integration by parts gives $\int_0^{\pi/2}\varphi\sin^2\varphi\,d\varphi = \frac{\pi^2}{16} + \frac{1}{4}$, so $\mathbb{E}[\arccos|a|] = \frac{\pi}{4} + \frac{1}{\pi}$, the same for b by symmetry, and $P = \frac{2}{\pi}\left(\frac{\pi}{2} + \frac{2}{\pi} - \frac{\pi}{2}\right) = \frac{4}{\pi^2}$.

It works. It costs a marginal density, an inverse-trig substitution, an integration by parts and a sign-sensitive bookkeeping of arcs — four places to drop a factor of 2 under time pressure.

The model chain. Write $p = (\cos\theta, \sin\theta)$ with θ uniform on $[0, 2\pi)$, and $q = (a, b)$ uniform on the open disc D of area π, independent of θ.

The condition. R has corners p, q, $(\cos\theta, b)$, $(a, \sin\theta)$. R is the convex hull of its corners and the closed disc $\bar{D}$ is convex, so $R \subseteq \bar{D}$ if and only if the four corners lie in $\bar{D}$. The corners p and q do. Hence $R \subseteq \bar{D}$ iff

$$|a| \le |\cos\theta| \quad \text{and} \quad |b| \le |\sin\theta|. \tag{$*$}$$

Condition on p. For fixed θ, $(*)$ says q lies in the axis-parallel rectangle $Q_\theta = [-|\cos\theta|, |\cos\theta|] \times [-|\sin\theta|, |\sin\theta|]$. Its corners $(\pm\cos\theta, \pm\sin\theta)$ lie on C, so $Q_\theta \subseteq \bar{D}$ (convexity again), and Q_θ minus its boundary lies in D. So the conditional probability is

$$\frac{\text{area } Q_\theta}{\pi} = \frac{4|\cos\theta\sin\theta|}{\pi} = \frac{2|\sin 2\theta|}{\pi}.$$

Average over p.

$$P = \frac{1}{2\pi}\int_0^{2\pi} \frac{2|\sin 2\theta|}{\pi}\,d\theta = \frac{1}{\pi^2}\int_0^{2\pi} |\sin 2\theta|\,d\theta = \frac{1}{\pi^2}\cdot 4\int_0^{\pi/2} \sin 2\theta\,d\theta = \frac{4}{\pi^2},$$

since $|\sin 2\theta|$ has period $\pi/2$ and $\int_0^{\pi/2}\sin 2\theta\,d\theta = 1$. Boundary cases (degenerate rectangles, equality in $(*)$) have probability zero. The answer is $\boxed{4/\pi^2}$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains derive $(*)$ identically. They split on the next line: which variable to hold fixed. (N) fixes the interior point q and asks for the set of good θ — an arc whose endpoints are $\arcsin|b|$ and $\arccos|a|$ — and then integrates inverse-trig functions over a disc. (M) fixes the boundary point p and asks for the set of good q — a rectangle whose area is a single product — and then integrates a sine over an interval.

> *I have a two-dimensional variable and a one-dimensional one. If I fix the one-dimensional variable, the thing left to measure is a region in the plane, and $(*)$ is a pair of bounds on $|a|$ and $|b|$ separately — that region is a box, and a box has an area I can write down. Fix p.*

> *The box $|x| \le |\cos\theta|$, $|y| \le |\sin\theta|$ has its corners on the circle. So it sits inside the disc on its own; I do not need to intersect it with anything.*

- ▷ **Cue:** a probability over two independent variables, one of which appears in the condition only through separate bounds on its coordinates ⟶ **Move:** condition on the *other* variable, so the event becomes a product region whose measure is a product
- ▷ **Cue:** "no point of R lies outside" a convex set, with R a polygon ⟶ **Move:** reduce to the vertices: a convex set contains a polygon iff it contains its corners
- ▷ **Cue:** an axis-parallel rectangle with a corner at $(\cos\theta, \sin\theta)$ ⟶ **Move:** its area is $4|\cos\theta\sin\theta| = 2|\sin 2\theta|$; the double-angle formula is waiting

Layer 3 THE WHY

Why is fixing p the natural move rather than a lucky guess? Look at what $(*)$ is made of. It is two inequalities, one on $|a|$ and one on $|b|$, with no term mixing a and b. An inequality system that separates in the coordinates of q describes a *product set* in q-space — an interval in a times an interval in b — and the measure of a product set is the product of the lengths. That is the definition of area for a rectangle, and it is the only kind of region whose area needs no integration. So the moment $(*)$ is written down, the problem is telling you: for fixed p the good q form a box. The natural attempt reads the same system in the other direction and sees two conditions on a single angle, which intersect as an arc with inverse-trigonometric endpoints; everything that follows is the price of reading $(*)$ against its grain.

A second reason, from the phrasing. The problem hands you p on the circumference and q in the interior. The variable on a one-dimensional set is the cheap one to integrate over last, because its integral is a single-variable integral. Integrating over the disc last means a two-variable integral with a non-rectangular domain — which is exactly where the natural attempt had to invent a marginal density $\frac{2}{\pi}\sqrt{1-a^2}$. The general principle is Fubini's theorem: the probability is $\iint \mathbf{1}_{\text{good}}\,d\mu_q\,d\mu_p$, and you may integrate in either order; choose the order that makes the inner integral a known area.

Why does the box fit inside the disc for free? Because $(\cos\theta, \sin\theta)$ is on the circle, and the box Q_θ is the rectangle inscribed in the circle with a corner at p — the diameter through p is its diagonal. This is also the geometric meaning of $(*)$: R stays inside the disc exactly when q lies in the rectangle inscribed in C with diagonal the diameter through p. Seen this way, the

answer $2|\sin 2\theta|$ is the area of an inscribed rectangle with diagonal 2 and angle 2θ between its diagonals — half the product of the diagonals times the sine of the angle between them.

Kedlaya, Poonen and Vakil fix θ in their first line without comment. Here is where that comes from: $(*)$ separates in a and b, and a separated condition is a box.

Layer $+\alpha$ BACKTESTING

1. The first cue is "rectangle with sides parallel to the x and y-axes". On seeing it write down the four corners, $(\cos\theta, \sin\theta)$, (a, b), $(\cos\theta, b)$, $(a, \sin\theta)$, and the vertex reduction: R is inside iff the two cross corners are.
2. Geometric probability by conditioning (Fubini): fix one variable, compute the measure of the good set of the other, average.
3. Hesitation comes at the choice of which variable to fix. It is reinforced away by looking at the shape of $(*)$ in each variable: in q it is two independent coordinate bounds (a box); in θ it is two trigonometric bounds on one angle (an arc with inverse-trig endpoints). Choose the box.
4. The trap is conditioning on q because q "has more information". It still reaches $4/\pi^2$, but through $\mathbb{E}[\arccos|a|] = \frac{\pi}{4} + \frac{1}{\pi}$, where a dropped factor of 2 in the marginal density or in the quadrant count is very likely. A secondary trap is forgetting that R's interior need not be checked: convexity reduces the problem to corners.
5. Whenever a condition on a point is a list of inequalities on its coordinates separately, the set of good points is a box; condition on everything else first so that box is the inner measure.
6. Reusable decision rule: *in a two-variable probability, integrate first over the variable in which the event is a product set, and last over the variable living on the lower-dimensional domain.*

1985 B5

1985 B5. Evaluate $\int_0^\infty t^{-1/2} e^{-1985(t+t^{-1})}\,dt$. You may assume that $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$.

Fields: real analysis; improper integrals; substitution **Position:** B5 — late in the paper, and most of the top solvers left it blank; the difficulty is entirely in seeing which substitution the symmetry $t \leftrightarrow 1/t$ asks for **Answer:** $\sqrt{\dfrac{\pi}{1985}}\, e^{-3970}$. In general $\int_0^\infty t^{-1/2} e^{-a(t+1/t)}\,dt = \sqrt{\pi/a}\, e^{-2a}$ for $a > 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $a = 1985$ and $I(a) = \int_0^\infty t^{-1/2} e^{-a(t+1/t)}\,dt$. The hint mentions e^{-x^2}, so the first move is to make a square appear: $t = x^2$, $dt = 2x\,dx$, $t^{-1/2} = 1/x$, giving

$$I(a) = 2\int_0^\infty e^{-a(x^2+x^{-2})}\,dx.$$

Good so far. Now the strong student tries to split the exponential, $e^{-ax^2} \cdot e^{-a/x^2}$, and expand the second factor as a power series $\sum_k (-a)^k x^{-2k}/k!$. Term by term this gives $\int_0^\infty x^{-2k} e^{-ax^2}\,dx$, which diverges at 0 for every $k \geq 1$. Wall.

Second try: differentiate with respect to a parameter. Put $F(b) = \int_0^\infty e^{-x^2 - b^2/x^2}\,dx$ for $b \geq 0$. Then

$$F'(b) = \int_0^\infty \left(-\frac{2b}{x^2}\right) e^{-x^2 - b^2/x^2}\,dx,$$

and with $x = b/y$, $dx = -b\,dy/y^2$, this becomes $-2\int_0^\infty e^{-b^2/y^2 - y^2}\,dy = -2F(b)$. So $F(b) = F(0)e^{-2b} = \frac{\sqrt{\pi}}{2} e^{-2b}$. Rescaling $x \mapsto x/\sqrt{a}$ gives $I(a) = \frac{2}{\sqrt{a}} F(a) = \sqrt{\pi/a}\, e^{-2a}$.

This second attempt is correct and is a good piece of work, but it spends its effort in the wrong place. The differentiation under the integral sign needs a domination argument (the

derivative's integrand blows up like b/x^2 near $x = 0$ unless you use the exponential factor), the continuity of F at $b = 0$ needs checking to use $F(0) = \sqrt{\pi}/2$, and inside it you still had to find the substitution $x = b/y$ — which is the $t \leftrightarrow 1/t$ symmetry in disguise.

The model chain. Let $a > 0$ and $I(a) = \int_0^\infty t^{-1/2} e^{-a(t+1/t)}\, dt$. The integral converges: on $(0, 1]$ the integrand is at most $t^{-1/2}$, and on $[1, \infty)$ at most e^{-at}.

Step 1: square root. Put $t = x^2$ ($x > 0$). Then

$$I(a) = 2\int_0^\infty e^{-a(x^2+1/x^2)}\, dx = 2e^{-2a}\int_0^\infty e^{-a(x-1/x)^2}\, dx,$$

using $x^2 + 1/x^2 = (x - 1/x)^2 + 2$. Call the last integral J.

Step 2: use $x \leftrightarrow 1/x$. Substituting $x = 1/y$, $dx = -dy/y^2$, and noting $(1/y - y)^2 = (y - 1/y)^2$,

$$J = \int_0^\infty e^{-a(y-1/y)^2}\, \frac{dy}{y^2}.$$

Adding the two expressions for J,

$$2J = \int_0^\infty e^{-a(x-1/x)^2}\Big(1 + \frac{1}{x^2}\Big)dx.$$

Step 3: the Gaussian. Put $u = x - 1/x$. Then $du = (1 + 1/x^2)\, dx$, and $x \mapsto u$ is a strictly increasing bijection of $(0, \infty)$ onto $(-\infty, \infty)$ (derivative positive; $u \to -\infty$ as $x \to 0^+$, $u \to \infty$ as $x \to \infty$). So

$$2J = \int_{-\infty}^\infty e^{-au^2}\, du = \frac{1}{\sqrt{a}}\int_{-\infty}^\infty e^{-v^2}\, dv = \sqrt{\frac{\pi}{a}} \qquad (v = \sqrt{a}\, u).$$

All integrands are positive, so every rearrangement of these convergent improper integrals is legitimate.

Step 4: assemble. $J = \frac{1}{2}\sqrt{\pi/a}$ and $I(a) = 2e^{-2a}J = \sqrt{\pi/a}\, e^{-2a}$. With $a = 1985$,

$$\int_0^\infty t^{-1/2} e^{-1985(t+t^{-1})}\, dt = \boxed{\sqrt{\frac{\pi}{1985}}\, e^{-3970}}. \qquad \square$$

(Checked numerically at $a = 0.7$: both sides equal $0.522412\ldots$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $I(a) = 2\int_0^\infty e^{-a(x^2+x^{-2})}\, dx$. They split at the next line. (N) treats x^2 and x^{-2} as two separate pieces of the exponent and tries to separate them — by series, then by a parameter. (M) treats $x^2 + x^{-2}$ as one object, completes the square to $(x - 1/x)^2 + 2$, and looks for the du that goes with $u = x - 1/x$.

> *The hint is a Gaussian, so I want the exponent to be (something)^2 and the dx to be d(that something). The exponent $x^2 + 1/x^2$ is a square plus a constant: $(x - 1/x)^2 + 2$. So I want $u = x - 1/x$. But $du = (1 + 1/x^2)\, dx$, and I only have dx. Where does a $1/x^2\, dx$ come from? From the substitution $x \to 1/x$, which leaves $(x - 1/x)^2$ alone and turns dx into dx/x^2. Average the two versions.*

> *The integrand is unchanged in form under $t \to 1/t$ except for the measure. When a problem is symmetric under an involution, write the integral both ways and add; the sum is often an exact derivative.*

▷ **Cue:** an exponent of the form $x^2 + 1/x^2$ (or $t + 1/t$ after $t = x^2$) ⟶ **Move:** complete the square: $(x - 1/x)^2 + 2$, then aim for $u = x - 1/x$

▷ **Cue:** you need $(1+1/x^2)\,dx$ but only have dx, and the integrand is invariant under $x \to 1/x$ $\longrightarrow$ **Move:** substitute $x \to 1/x$ to get the dx/x^2 copy, add, and halve

▷ **Cue:** a problem hands you $\int e^{-x^2} = \sqrt{\pi}$ $\longrightarrow$ **Move:** the answer route ends in a Gaussian; work backwards to find the substitution that produces $e^{-u^2}\,du$

Layer 3 THE WHY

Kedlaya, Poonen and Vakil (after Bernau) split the integral at $t = 1$, flip the first half by $t \to 1/t$, and then substitute $u = a^{1/2}(t^{1/2} - t^{-1/2})$. Here is where that comes from.

Start from the hint. The only integral you are allowed to know is $\int e^{-x^2}dx$. So the solution must end in the form $\int e^{-u^2}\,du$, which means two things must happen together: the exponent must become a perfect square in some new variable u, and the measure must become du. The problem is a matching problem between exponent and measure.

The exponent $a(t+1/t)$ is not a square, but after $t = x^2$ it is $a(x^2+x^{-2})$, and x^2+x^{-2} is one step from a square, by the identity $(x-1/x)^2 = x^2 - 2 + x^{-2}$. (The alternative $(x+1/x)^2 - 2$ also works for the exponent, but $x+1/x$ is not monotonic on $(0,\infty)$ — it turns round at $x=1$ — so it cannot serve as a single substitution. $x-1/x$ is increasing and sweeps all of $\mathbb{R}$: exactly the range of the Gaussian.) So $u = x - 1/x$ is forced by the exponent, not guessed.

Now the measure. $du = (1+x^{-2})\,dx$, so you are short a term $x^{-2}dx$. This is where the symmetry of the problem pays. The expression $t+1/t$ is unchanged by $t \to 1/t$, and correspondingly $(x-1/x)^2$ is unchanged by $x \to 1/x$. An involution that fixes the exponent can only change the measure, and here it changes dx into dx/x^2 — the missing piece. Writing J both ways and adding is the averaging principle: if a quantity is invariant under a symmetry, you may replace the integrand by its average over the symmetry, and the average is often better behaved than either copy. Bernau's split at $t = 1$ is the same move, performed on the fixed point of the involution: $[1/B, 1]$ and $[1, B]$ are swapped by $t \to 1/t$.

Why is the parameter-differentiation attempt the expensive route? Because it hides the same symmetry inside the substitution $x = b/y$ and then pays extra for analytic justification. The model chain uses the symmetry once, openly, and needs only positivity of integrands for rigour.

The general fact behind the problem is that $\int_0^\infty t^{\nu-1}e^{-a(t+1/t)}\,dt = 2K_\nu(2a)$, a modified Bessel function; for $\nu = \pm\frac{1}{2}$ (the case here) the Bessel function is elementary, and the reason is exactly that $t^{\mp 1/2}$ is the weight for which the averaging trick produces an exact du.

Layer $+\alpha$ BACKTESTING

1. The first cue is the pair "$t + t^{-1}$" and "you may assume $\int e^{-x^2} = \sqrt{\pi}$". Write down: *I need $e^{-u^2}\,du$; after $t = x^2$ the exponent is $(x-1/x)^2 + 2$.*

2. Substitution driven by the target form, plus symmetrisation under the involution $x \to 1/x$.

3. Hesitation comes at $du = (1+1/x^2)dx$ when only dx is present: students abandon $u = x - 1/x$ here. Reinforce it away by asking what the problem's symmetry does to the measure — the missing term appears.

4. The trap is expanding $e^{-a/t}$ as a series (every term diverges) or differentiating under the integral sign without checking domination. A second trap is choosing $u = x + 1/x$, which fixes the exponent but is not injective on $(0,\infty)$.

5. Any integrand invariant under $x \to 1/x$ (or $x \to c - x$, or $x \to -x$) invites writing the integral twice and adding. Look at what the involution does to dx.

6. Reusable decision rule: *when the answer must be a known integral, fix the target form first; choose u to match the exponent, then use the problem's symmetry to manufacture the missing part of du.*

1985 B6

1985 B6. Let G be a finite set of real $n \times n$ matrices $\{M_i\}$, $1 \le i \le r$, which form a group under matrix multiplication. Suppose that $\sum_{i=1}^{r} \operatorname{tr}(M_i) = 0$, where $\operatorname{tr}(A)$ denotes the trace of the matrix A. Prove that $\sum_{i=1}^{r} M_i$ is the $n \times n$ zero matrix.

Fields: linear algebra; group theory (averaging over a finite group) **Position:** B6 — last problem, and in 1985 almost nobody solved it; the whole proof is five lines once one identity is found **Answer:** Let $S = \sum_i M_i$. Left multiplication by any M_j permutes G, so $M_j S = S$; summing over j gives $S^2 = rS$. So every eigenvalue of S is 0 or r, and $\operatorname{tr} S = 0$ forces all to be 0; then $S - rI$ is invertible and $S(S - rI) = 0$ gives $S = 0$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student tests small cases first, which is right. $G = \{I, -I\}$: traces n and $-n$ sum to 0, and $I + (-I) = 0$. $G = \{I, R, R^2\}$ with R rotation of the plane by $120°$: traces $2, -1, -1$ sum to 0, and $I + R + R^2 = 0$ because $R^3 - I = (R - I)(I + R + R^2) = 0$ and $R - I$ is invertible. So the statement looks like "roots of unity sum to zero", and the first plan is eigenvalues. Each M_i has finite order, so over $\mathbb{C}$ it is diagonalisable with root-of-unity eigenvalues, and $\operatorname{tr} M_i$ is a sum of roots of unity.

The plan then stalls: the M_i need not commute, so they cannot be diagonalised together, and the eigenvalues of $\sum M_i$ are not sums of eigenvalues of the M_i. A trace condition on a sum does not obviously control each summand.

Second plan, vector by vector. Take any v and set $w = Sv$. Then $M_j w = M_j S v$, and $M_j S = \sum_i M_j M_i = S$ because $M_j M_1, \ldots, M_j M_r$ is the list of elements of G in another order. So w is fixed by every M_j. The student has found the key identity. Now they want to show the fixed space $W = \{w : M_j w = w \ \forall j\}$ is zero, from $\operatorname{tr} S = 0$. On W, S acts as rI, contributing $r \dim W$ to the trace. To finish they need S to act with trace 0 on a complement of W — which requires a complement preserved by every M_j. That is Maschke's theorem, and proving it on the spot (average an inner product over G) costs a page. Wall, or at least a very long detour.

The model chain. Let $S = \sum_{i=1}^{r} M_i$ and let I denote the $n \times n$ identity matrix. (The identity *of the group* G need not be I: $G = \{\operatorname{diag}(1,0), \operatorname{diag}(-1,0)\}$ is a group whose identity is $\operatorname{diag}(1,0)$. The argument never uses $I \in G$.)

Step 1: $M_j S = S$. Fix j. The map $X \mapsto M_j X$ sends G to G and is injective, since M_j has an inverse M_j' in G and $M_j'(M_j X) = X$ for $X \in G$. A self-injection of a finite set is a bijection, so $M_j M_1, \ldots, M_j M_r$ is a rearrangement of $M_1, \ldots, M_r$. Summing,

$$M_j S = \sum_{i=1}^{r} M_j M_i = \sum_{i=1}^{r} M_i = S.$$

Step 2: $S^2 = rS$. Sum Step 1 over j: $S^2 = \left(\sum_j M_j\right) S = \sum_j M_j S = rS$. So

$$S(S - rI) = 0.$$

Step 3: eigenvalues. Work over $\mathbb{C}$. If $Sv = \lambda v$ with $v \neq 0$, then $\lambda^2 v = S^2 v = rSv = r\lambda v$, so $\lambda \in \{0, r\}$. The trace is the sum of the eigenvalues with algebraic multiplicity, so $\operatorname{tr} S = r\,m$, where m is the multiplicity of r. But $\operatorname{tr} S = \sum_i \operatorname{tr} M_i = 0$ and $r \ge 1$, so $m = 0$: r is not an eigenvalue of S.

Step 4: conclude. Since r is not an eigenvalue, $S - rI$ is invertible. Multiply $S(S - rI) = 0$ on the right by $(S - rI)^{-1}$: $S = 0$. □

Same proof, second dress. $P = S/r$ satisfies $P^2 = P$: it is a projection. A projection is diagonalisable with eigenvalues in $\{0, 1\}$ (every v splits as $Pv + (v - Pv)$, with $P(Pv) = Pv$ and $P(v - Pv) = 0$), so $\operatorname{tr} P = \operatorname{rank} P$. Here $\operatorname{tr} P = 0$, so $\operatorname{rank} P = 0$ and $S = 0$. As a by-product: in general $\sum_i \operatorname{tr} M_i = r \cdot \operatorname{rank}(S/r)$ is always a non-negative multiple of r.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find $M_jS = S$. The natural attempt then *applies* the identity to vectors, reads it as "the image of S is fixed by G", and needs a G-invariant complement to compute the trace. The model *sums* the identity over j and gets the matrix equation $S^2 = rS$, which controls every eigenvalue of S at once without mentioning any subspace.

> *I have $M_jS = S$ for every j, and S is itself the sum of the M_j. So sum the identity over j — the left side becomes $S \cdot S$. Now I have a polynomial equation in the single matrix S, and a polynomial equation pins down the eigenvalues.*

> *The hypothesis is about a trace, and the conclusion is about a matrix. The only bridge between them is eigenvalues. So whatever I prove about S must be a statement that restricts its eigenvalues to a set where "sum is zero" forces "all are zero".*

- ▷ **Cue:** a sum over all elements of a finite group $\longrightarrow$ **Move:** multiply by one element: the sum is unchanged because multiplication permutes the group
- ▷ **Cue:** an identity $AS = S$ holding for every summand A of S $\longrightarrow$ **Move:** sum it over the summands to get $S^2 = (\text{count})\, S$
- ▷ **Cue:** a matrix satisfying a polynomial $p(S) = 0$ and a condition on its trace $\longrightarrow$ **Move:** the eigenvalues are roots of p; the trace is their sum; decide which roots can occur

Layer 3 THE WHY

Why is summing $M_jS = S$ the natural move? Because of what S is. S is built from the M_j by addition; an identity that holds for each M_j separately and is linear in M_j transfers to the sum. That is all Step 2 is — linearity — and it converts a family of r equations into one equation about S alone. The natural attempt keeps the r equations separate ("w is fixed by each M_j") and so is forced to reason about the common fixed space and its complement. The model collapses the family before reasoning.

Why is a polynomial equation exactly what is wanted? Go to the hypothesis. It gives $\operatorname{tr} S = 0$ and asks for $S = 0$. Trace alone never determines a matrix: $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ have trace 0. So something else must restrict S, and it must do two jobs: (i) stop eigenvalues cancelling in the trace, and (ii) stop nilpotent parts, which are invisible to eigenvalues. $S^2 = rS$ does both. It confines eigenvalues to $\{0, r\}$, all of one sign, so a zero sum means all zero; and it has distinct roots $0 \neq r$, so S is diagonalisable (Step 4 is the concrete form of this) and there is no nilpotent part to hide. Note where $r \neq 0$ is used: if the two roots coincided, $S^2 = 0$ would allow $S = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$.

Why should one expect S/r to be a projection in the first place? Averaging over a group is the canonical way to make something invariant. $P = \frac{1}{r}\sum_g g$ maps every vector to a G-fixed vector (that is $M_jP = P$), and it leaves G-fixed vectors alone (if $M_jw = w$ for all j then $Pw = \frac{1}{r}\sum_j w = w$). A map that lands in a subspace and is the identity on it is, by definition, a projection onto it. So $\frac{1}{r}\sum_i \operatorname{tr} M_i = \operatorname{tr} P = \dim(\text{fixed space})$: the average trace counts the invariant vectors. The problem is the statement "if the average trace is 0, there are no invariant vectors, so the average kills everything". In representation theory this is the multiplicity of the trivial representation, which is Kedlaya–Poonen–Vakil's Solution 2; the matrix proof is the same fact with the vocabulary removed.

Kedlaya, Poonen and Vakil begin by observing that $M_jM_1, \ldots, M_jM_r$ is a permutation of G. Here is where that comes from: the only information you have about G is that it is a group, and the defining property of a group that interacts with a sum over the whole group is that multiplication by a fixed element is a bijection.

Layer $+\alpha$ BACKTESTING

1. The first cue is "form a group" together with a sum over *all* of G. Write down immediately: $M_j S = S$ for every j, because left multiplication permutes G.

2. Averaging over a finite group, then a polynomial identity for the average, then eigenvalues.

3. Hesitation occurs after $M_j S = S$, when it is unclear how a statement about each M_j helps with a trace. Reinforce it away with the question: *what does this identity say about S alone?* Sum over j.

4. Two traps. First, assuming $I \in G$ or that the M_i are invertible; the problem allows singular matrices (the group's identity can be a non-trivial projection), and a proof relying on invertibility is incomplete. Second, concluding $S = 0$ from "all eigenvalues are 0" without excluding a nilpotent S; that step needs $S(S - rI) = 0$ with $S - rI$ invertible.

5. Any time a problem sums over every element of a group (or every residue mod n, or every rotation), expect the sum to be invariant under the group action, and expect the normalised sum to be a projection.

6. Reusable decision rule: *to get from a trace condition to a matrix identity, first find a polynomial the matrix satisfies with distinct roots of one sign; then the trace determines the eigenvalue multiplicities and the polynomial rules out nilpotent parts.*