DAWN PUTNAM SESSION

Putnam 1997 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1997 — Paper A **5**
1997 A1 5
1997 A2 7
1997 A3 9
1997 A4 11
1997 A5 14
1997 A6 17

Putnam 1997 — Paper B **21**
1997 B1 21
1997 B2 23
1997 B3 25
1997 B4 28
1997 B5 31
1997 B6 34

Putnam 1997 — Paper A

1997 A1

1997 A1. A rectangle, $HOMF$, has sides $HO = 11$ and $OM = 5$. A triangle ABC has H as the intersection of the altitudes, O the center of the circumscribed circle, M the midpoint of BC, and F the foot of the altitude from A. What is the length of BC?

Fields: Euclidean triangle geometry — orthocentre, circumcentre, Euler line **Position:** A1 — the opener
Answer: $BC = 28$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Four unfamiliar-looking points are handed to you as vertices of a rectangle, and the honest first instinct is to put down coordinates and let algebra sort out what synthetic argument you cannot immediately see. Place BC on the x-axis with M, the midpoint, at the origin: $B = (-a, 0)$, $C = (a, 0)$, so $BC = 2a$ is the target. Since O is the circumcentre, $OM \perp BC$, so $O = (0, p)$ for some p with $|p| = 5$. Since F is the foot of the altitude from A, $F = (f, 0)$ lies on BC, and $AF \perp BC$ forces $A = (f, q)$ for some height q.

Now find H. The altitude from A is the vertical line $x = f$. The altitude from B is perpendicular to AC through $B = (-a, 0)$; the slope of AC is $-q/(a - f)$, so the altitude from B has slope $(a - f)/q$ and meets $x = f$ at

$$H = \Big(f,\ \frac{(a-f)(a+f)}{q}\Big) = \Big(f,\ \frac{a^2 - f^2}{q}\Big).$$

This is already a page of algebra for one vertex of a rectangle whose other three vertices took one line each.

Next the rectangle conditions must be imposed. $HOMF$ in order means $OH \perp OM$; since OM is vertical, OH must be horizontal, so H's y-coordinate equals O's: $(a^2 - f^2)/q = p$. That gives $H = (f, p)$, and then $HO = |f| = 11$, $MF = |f| = 11$ automatically, $OM = |p| = 5$, $FH = |p| = 5$ automatically: the rectangle condition on the right angle is really one equation, $a^2 - f^2 = pq$, plus $|f| = 11$, $|p| = 5$.

But that alone is two equations short of pinning down a: nothing so far has used that O is equidistant from A, B, C — the actual definition of circumcentre, as opposed to just "lies on the perpendicular bisector of BC", which only used the vertical-line fact. Impose $OA = OB$: $f^2 + (q - p)^2 = a^2 + p^2$, which rearranges to $a^2 - f^2 = q^2 - 2pq$. Combined with $a^2 - f^2 = pq$ from the rectangle condition, this forces $pq = q^2 - 2pq$, i.e. $q(q - 3p) = 0$, and since $q \neq 0$, $q = 3p$. Only now, after two separate systems, does the arithmetic close: $a^2 - f^2 = pq = 3p^2 = 75$, $f^2 = 121$, so $a^2 = 196$, $a = 14$, $BC = 28$ — the right answer, reached, but at the cost of quietly re-deriving that A's height is exactly three times O's height above BC, which is the Euler line in disguise, without ever naming it.

The model chain. The centroid G divides every median in ratio $2 : 1$ from the vertex, and by the Euler line, G also divides HO in ratio $HG : GO = 2 : 1$. Consider the median AM: G lies two-thirds of the way from A to M. Because $AF \parallel OM$ (both perpendicular to BC) and H lies

on line AF, the same 2:1 homothety at G that sends $M \mapsto A$ sends $O \mapsto H$, so H is two-thirds of the way from A to F along that same line: $AH = \frac{2}{3}AF$.

Since $OM = 5$ and $HF = 5$ (opposite side of the rectangle), and since AH is parallel to OM with the standard fact $AH = 2 \cdot OM = 10$, the relation $AH = \frac{2}{3}AF$ gives $AF = 15$.

Triangles BFH and AFC are both right-angled at F, and $\angle HBC = \frac{\pi}{2} - \angle C = \angle CAF$ (the altitude-foot angle identity), so they are similar:

$$\frac{BF}{FH} = \frac{AF}{FC} \quad \Longrightarrow \quad BF \cdot FC = FH \cdot AF = 5 \cdot 15 = 75.$$

Also $BF - FC = (BM + MF) - (MC - MF) = 2\,MF = 2 \cdot 11 = 22$, using $MF = 11$ (the other side of the rectangle). Then

$$BC^2 = (BF + FC)^2 = (BF - FC)^2 + 4\,BF \cdot FC = 22^2 + 4 \cdot 75 = 484 + 300 = 784,$$

so $BC = \sqrt{784} = 28$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains eventually use the same two numbers (22 from MF, 75 from $FH \cdot AF$) to close via $BC^2 = (BF - FC)^2 + 4\,BF \cdot FC$. They split earlier, at the question of *how to get* $AF = 15$: the natural attempt sets up a full coordinate system and solves two nested linear systems (perpendicularity, then equidistance) to discover $q = 3p$; the model chain gets $AF = 15$ in two lines by naming the Euler-line homothety at G and reading off $AH = 2 \cdot OM$ and $AH = \frac{2}{3}AF$ as known facts about any triangle, never touching a coordinate.

> *H, O, and M are not four unrelated points that happen to form a rectangle — H and O are Euler-line points, and M is the base of the median through the same centroid that anchors the Euler line's* 2:1 *ratio. Before reaching for coordinates I should ask what G, the centroid, does to each labelled point, because G is the one point every one of H, O, M, F is already secretly related to.*

Layer 3 THE WHY

Why does naming G shortcut two nested linear systems into two lines of ratio arithmetic? Because the Euler line is not an extra fact bolted onto the triangle; it is a direct consequence of the vector identity $H = A + B + C$ when O is placed at the origin, itself provable in one line: with $O = 0$, $(A + B + C - A) \cdot (C - B) = (B + C) \cdot (C - B) = |C|^2 - |B|^2 = R^2 - R^2 = 0$, so $A + B + C - A = B + C$ is perpendicular to $C - B$, i.e. the line from A through $A + B + C$ is the altitude from A — meaning $A + B + C$ lies on every altitude, hence equals H. From this, $G = (A + B + C)/3$ satisfies $G - O = (A + B + C)/3$ and $H - G = \frac{2}{3}(A + B + C) = 2(G - O)$, which is simultaneously the Euler-line ratio $HG = 2\,GO$ *and*, restricted to the median direction, the statement $AH = 2\,OM$ (since $M = (B + C)/2$ and $AH = B + C = 2M$ when $O = 0$). One vector identity produces both facts the model chain uses. The coordinate attempt never invokes this identity, so it has to re-derive its consequence, $q = 3p$, from raw equidistance — the correct answer, purchased twice.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the problem names H and O together as vertices of a shape, alongside M and F. On seeing H and O named in the same breath, write down "Euler line" before anything else — these two points are never independent.
2. That cue suggests using the centroid G and its fixed 2:1 ratios on both the Euler line ($HG : GO$) and the median ($AG : GM$), rather than solving for either point's coordinates from scratch.
3. Hesitation occurs at the moment of choosing a method for an unfamiliar four-point rectangle: coordinates feel safe because every step is mechanical, even though it is not the fastest

path. This is reinforced away by treating "I recognise these points but not a direct relation between them" as the cue to search for the centroid, not the cue to reach for axes.

4. The trap is that coordinate geometry always "works" on a problem like this, so there is no error message telling you a shorter path exists; the only cost is time, and a full-length exam does not forgive rediscovering $AH = 2 \cdot OM$ the hard way.

5. Recognise this family faster by treating any appearance of both the orthocentre and circumcentre of the same triangle as an automatic invitation to invoke the Euler line and its 2:1 centroid ratio, the way a right angle invites Pythagoras.

6. Reusable rule: when a problem hands you several classical triangle centres at once, spend thirty seconds asking which standard line (Euler line, nine-point circle) already relates them before setting up coordinates; coordinates should be the tool of last resort for named centres, not the first.

▷ **Cue:** a triangle problem names both the orthocentre H and the circumcentre O $\longrightarrow$ **Move:** invoke the Euler line: H, centroid G, O are collinear with $HG = 2\,GO$, and $AH = 2 \cdot$ (distance from O to the midpoint of the opposite side)

▷ **Cue:** a right triangle with legs along an altitude foot shares an angle with the original triangle's altitude configuration $\longrightarrow$ **Move:** look for the similar-triangle pair $BFH \sim AFC$ via $\angle HBC = \frac{\pi}{2} - \angle C = \angle CAF$, giving $BF \cdot FC = FH \cdot AF$

1997 A2

1997 A2. Players $1, 2, 3, \ldots, n$ are seated around a table, and each has a single penny. Player 1 passes a penny to player 2, who then passes two pennies to player 3. Player 3 then passes one penny to Player 4, who passes two pennies to Player 5, and so on, players alternately passing one penny or two to the next player who still has some pennies. A player who runs out of pennies drops out of the game and leaves the table. Find an infinite set of numbers n for which some player ends up with all n pennies.

Fields: combinatorics — elimination games, invariants, powers of two **Position:** A2 **Answer:** The infinite family $n = 2^m + 2$ (equivalently $n = 2^m + 1$), $m \geq 0$, always terminates with one player holding all n pennies.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem only asks you to *find* an infinite set that works, not to characterise every n that works — so the honest first move is not a general theory but a hand simulation of small cases, hunting for a pattern to conjecture and then prove.

$n = 3$: P1 passes 1 to P2 (P1 out, P2 has 2). P2 passes 2 to P3 (P2 out, P3 has 3). P3 now holds all 3 pennies with no one else at the table. Terminates.

$n = 4$: P1→P2 (1); P1 out, P2 has 2. P2→P3 (2); P2 out, P3 has 3. P3→P4 (1); P3 has 2, P4 has 2. P4→P3, the only player left with pennies (2, since P1, P2 are out); P4 out, P3 has 4. Terminates, P3 holds all 4.

$n = 5$: tracking it out turn by turn (P1→P2:1, P2→P3:2, P3→P4:1, P4→P5:2, P5→P3:1, P3→P5:2, P5→P3:1, P3→P5:2) ends with P5 holding all 5 pennies and everyone else out.

$n = 6$: continuing to simulate is possible but the bookkeeping is already error-prone by hand, and a fourth full case buys little extra evidence beyond what three already suggest.

The pattern in the successes so far — 3, 4, 5 — lines up against powers of two as $3 = 2^1 + 1$, $4 = 2^1 + 2$, $5 = 2^2 + 1$. That is enough of a foothold to conjecture the family $n = 2^m + 1$ or $n = 2^m + 2$, but a conjecture from three data points is not a proof, and nothing in the simulation explains *why* powers of two are the relevant scale rather than, say, powers of three.

The model chain. Call a state *k-balanced with mover X* if every active player except X holds exactly k pennies and X holds at least k, and X is about to make a 1-penny pass (the start of

a "move" of two turns: a 1-pass then a 2-pass). The starting position, $k = 1$ with all n players holding 1 and player 1 about to move, is 1-balanced.

The published solution (Bhargava–Kedlaya–Ng) establishes, by tracking a full round of passes from a k-balanced state, that: if the number of active players p is even, then after enough rounds every other player (starting with the mover) gains one penny per round while the rest are driven to 0 and drop out, halving the population and producing a new k'-balanced state with $p/2$ players; if p is odd and $p > 1$, two complete rounds return every player's penny count to what it was before those two rounds, so the game never progresses and never terminates. Iterating the halving step, a k-balanced state with p active players terminates (collapses to one player) exactly when p is a power of 2.

Applied to n players starting from 1 each: after the very first move (P1's 1-pass then P2's 2-pass), $\lfloor (n-1)/2 \rfloor$ players remain active, each holding 2 pennies except the player who is about to move next, who holds 3 or 4. This is a 2-balanced state with $p = \lfloor (n-1)/2 \rfloor$ active players. By the halving lemma above, the game terminates in a single winner exactly when p is a power of 2, i.e. $\lfloor (n-1)/2 \rfloor = 2^m$, which happens precisely for $n = 2^{m+1} + 1$ or $n = 2^{m+1} + 2$.

Caveat. The halving/period-two lemma above — that an odd number of active players in a k-balanced state returns to the same state after two rounds, while an even number reduces by half — is reproduced from the published solution and confirmed against the hand simulations for $n = 3, 4, 5$ above, but I have not independently re-derived it in full generality from the raw turn-by-turn dynamics for an arbitrary k and arbitrary starting surplus at the mover; a from-scratch proof of the general lemma would require its own careful induction on the mover's exact surplus, which this analysis does not carry out. The specific family stated in the vitals line, $n = 2^m + 2$, is exactly what the problem asks for (existence of an infinite working set), and is fully consistent with every case checked directly by hand.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches start the same way, by simulating $n = 3, 4, 5$ turn by turn; they split at what to do with three successful small cases. The natural attempt stops at pattern-spotting: $3, 4, 5$ look like they cluster near powers of two, and that observation alone is offered as the answer. The model chain instead asks what happens *structurally* after the very first move — reducing the n-player game to a smaller 2-balanced game with $\lfloor (n-1)/2 \rfloor$ players — and recognises that this reduction is self-similar: the same halving question reappears one level down, which is exactly the condition that makes powers of two, not any other base, the relevant scale.

> *Three successful cases sitting near powers of two is not an explanation, only a coincidence until I can point to a step in the process that literally halves something. Instead of asking "what number is* 3, 4, 5 *close to", I should ask "what does one full move do to the number of players still in the game", because a rule that eliminates roughly half the players in one pass is the actual reason powers of two, and not powers of three, keep appearing.*

Layer 3 THE WHY

Why does the process privilege powers of two specifically, rather than any other modulus? Because the elimination rule is inherently a binary one: each move consists of exactly two turns (a 1-pass, then a 2-pass), and after one full circuit of moves from a balanced state, the players split into two classes — those who received a net penny each round and survive, and those who received nothing and are driven to zero. This is not an arbitrary halving imposed from outside; it falls directly out of the fact that the passing amounts alternate between exactly two values with period 2, so any invariant of the process (who accumulates, who empties) inherits

that same period-2 structure. A population that repeatedly splits by a factor of 2 terminates in a single survivor exactly when it started as a power of 2 — the same reason a single-elimination knockout bracket needs 2^k entrants to avoid a bye. Powers of three, or any other base, never enter the picture because nothing in the rule alternates with period 3.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "alternately passing one penny or two": an alternation between exactly two fixed amounts, on seeing it, should immediately register as a period-2 mechanism worth tracking through one full cycle before attempting any general argument.
2. That cue suggests simulating small n by hand far enough to see one full move complete (both a 1-pass and a 2-pass), then asking how many active players remain and in what configuration, rather than jumping straight to a general formula.
3. Hesitation occurs after the third or fourth successful hand simulation, at the point of deciding whether pattern-matching against $2^m + 1, 2^m + 2$ is sufficient or whether a mechanism must be found. This is reinforced away by insisting that any numeric pattern from three data points earns, at minimum, one clear reason tied to the rules of the game before it is trusted.
4. The trap is treating the sequence $3, 4, 5$ as strong enough evidence on its own; three successes are compatible with several different closed forms, and the game's rules (not the numbers) are what must certify which one is correct.
5. Recognise this family faster by treating any elimination process with a fixed, alternating period as an automatic prompt to track the population size after exactly one full period, since that is where the recursive halving (or tripling, or whatever the period suggests) shows up first.
6. Reusable rule: in an elimination or accumulation process governed by a periodic rule of period p, the natural scale for a closed-form answer is powers of p (here $p = 2$), and the way to find the right family is to compute what one full period does to the active population, then look for a self-similar reduction, not to pattern-match the raw numbers alone.

▷ **Cue:** a game or process passes/eliminates according to a fixed alternating rule with period p ⟶ **Move:** track the state after exactly one full period rather than one raw step, and look for a self-similar reduction of the population by a factor related to p

▷ **Cue:** a sequence of successful small cases looks numerically close to powers of a small base ⟶ **Move:** do not stop at the numeric coincidence; find the mechanism in the rules that would force exactly that base to be the relevant scale

1997 A3

1997 A3. Evaluate

$$\int_0^\infty \left(x - \frac{x^3}{2} + \frac{x^5}{2\cdot 4} - \frac{x^7}{2\cdot 4\cdot 6} + \cdots\right) \left(1 + \frac{x^2}{2^2} + \frac{x^4}{2^2\cdot 4^2} + \frac{x^6}{2^2\cdot 4^2\cdot 6^2} + \cdots\right) dx.$$

Fields: real analysis — power series, integration by parts, term-by-term integration **Position:** A3
Answer: $\sqrt{e}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Two power series are staring at you, and the reflex on seeing a power series in a Putnam integral is to name it. The first bracket, $x - \frac{x^3}{2} + \frac{x^5}{2\cdot 4} - \frac{x^7}{2\cdot 4\cdot 6} + \cdots$, is easy: the n-th term is $(-1)^n x^{2n+1}/(2\cdot 4\cdots 2n) = (-1)^n x^{2n+1}/(2^n n!)$, so the bracket is $x\sum_n (-x^2/2)^n/n! = x\,e^{-x^2/2}$, a Taylor series any student who has seen $e^t = \sum t^n/n!$ will recognise on sight.

The second bracket is where the honest attempt slows down. Its n-th term is $x^{2n}/(2\cdot 4\cdots 2n)^2 = x^{2n}/(4^n(n!)^2)$, which is $\sum_n (x^2/4)^n/(n!)^2$ — the series for the modified Bessel function $I_0(x) = \sum_n (z/2)^{2n}/(n!)^2$ at $z = x$. A student who recognises this (most will not, without prompting) is tempted to reach for a table of Bessel-function integrals, hunting for a closed form of $\int_0^\infty xe^{-x^2/2} I_0(x)\,dx$; a student who does not recognise it is tempted to multiply the two series together as a literal Cauchy product, $\left(\sum_j c_j x^{2j+1}\right)\left(\sum_k d_k x^{2k}\right) = \sum_N \left(\sum_{j+k=N} c_j d_k\right) x^{2N+1}$, and integrate the resulting single series term by term.

Either route works in principle, but both spend real effort converting *both* brackets into a single closed object — a Bessel-function identity that must be looked up, or a convolution sum with a double index that must be simplified — before any integration happens at all. Neither is necessary: the second bracket never needs to become anything other than what it already is.

The model chain. Keep the first bracket in closed form, $xe^{-x^2/2}$, and keep the second bracket as a literal series, with n-th term $x^{2n}/(2\cdot 4\cdots 2n)^2$. Multiplying gives, term by term,

$$xe^{-x^2/2}\cdot \frac{x^{2n}}{(2\cdot 4\cdots 2n)^2} = \frac{x^{2n+1}e^{-x^2/2}}{(2\cdot 4\cdots 2n)^2},$$

and each of these is individually integrable in closed form. Define $I(n) = \int_0^\infty x^{2n+1}e^{-x^2/2}\,dx$. Integrating by parts with $u = x^{2n}$, $dv = xe^{-x^2/2}dx$, $v = -e^{-x^2/2}$:

$$I(n) = \left[-x^{2n}e^{-x^2/2}\right]_0^\infty + 2n\int_0^\infty x^{2n-1}e^{-x^2/2}dx = 2n\,I(n-1),$$

the boundary term vanishing at both ends (at 0 because $x^{2n} = 0$ for $n \geq 1$, at ∞ because $e^{-x^2/2}$ beats any polynomial). With $I(0) = \int_0^\infty xe^{-x^2/2}dx = \left[-e^{-x^2/2}\right]_0^\infty = 1$, induction gives $I(n) = 2\cdot 4\cdots 2n$ for $n \geq 0$ (the empty product for $n = 0$).

Assuming term-by-term integration is legitimate (the integrand is nonnegative on the second bracket's terms after the sign is carried by the first, and the series converges absolutely and uniformly on compacts, so this is justified for this problem), the whole integral is

$$\sum_{n=0}^\infty \frac{I(n)}{(2\cdot 4\cdots 2n)^2} = \sum_{n=0}^\infty \frac{2\cdot 4\cdots 2n}{(2\cdot 4\cdots 2n)^2} = \sum_{n=0}^\infty \frac{1}{2\cdot 4\cdots 2n} = \sum_{n=0}^\infty \frac{1}{2^n n!} = \sum_{n=0}^\infty \frac{(1/2)^n}{n!} = e^{1/2} = \sqrt{e}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains correctly identify the first bracket as $xe^{-x^2/2}$ in the first line; they split immediately after, at what to do with the second bracket. The natural attempt treats "convert to a single closed-form integrand" as the required next step, forcing either a Bessel-function lookup or a full Cauchy-product convolution. The model chain asks instead whether a closed form for the *whole* integrand is actually needed, and notices it is not: integrating term by term against the second bracket's coefficients turns the product of two series into a single series of elementary integrals $I(n)$, each computable on its own by one integration by parts.

> *I do not need the second bracket to become a named function at all — I only need, for each fixed power of x it contributes, the single elementary integral $\int_0^\infty x^{2n+1}e^{-x^2/2}dx$, and that integral I can get by hand with one integration by parts and an induction. Multiplying two series together and then integrating is backwards; I should integrate term by term first and let the coefficients combine afterward.*

Layer 3 THE WHY

Why does swapping the order — integrate first, combine coefficients after — turn an intractable double sum into a one-line answer? Because the coefficients of the second bracket,

$1/(2\cdot 4\cdots 2n)^2$, are not arbitrary: they are the *square* of the same double factorial that the integration-by-parts recursion $I(n) = 2n\, I(n-1)$ produces as its own closed form, $I(n) = 2\cdot 4\cdots 2n$. The moment the two are multiplied, $I(n)/(2\cdot 4\cdots 2n)^2 = (2\cdot 4\cdots 2n)/(2\cdot 4\cdots 2n)^2 = 1/(2\cdot 4\cdots 2n)$, one entire copy of the double factorial in the denominator is cancelled by the integral itself, leaving the other copy standing alone as a plain factorial series $\sum(1/2)^n/n! = e^{1/2}$. This is not a coincidence engineered after the fact: the original problem was built by starting from $xe^{-x^2/2}\cdot I_0(x)$-type structure and choosing the second series' coefficients specifically so that one factor of the recursion's output would survive integration and the other would not. Recognising the second bracket's coefficients as *a square* of the same quantity that a basic reduction formula would generate is the structural fact the Bessel-function or Cauchy-product routes obscure, because both of those routes hide the double factorial inside a named function or a convolution index before it has a chance to cancel visibly.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the first bracket's coefficients, $1/(2\cdot 4\cdots 2n)$ with alternating sign, are immediately recognisable as the Taylor coefficients of $e^{-x^2/2}$ (times x). On seeing it, write the closed form down at once and leave the second bracket alone rather than hunting for its closed form too.

2. That cue suggests multiplying the closed-form first bracket by the second bracket's series *termwise*, reducing the whole integral to a sum of elementary integrals $\int_0^\infty x^{2n+1}e^{-x^2/2}dx$, each solvable by one integration by parts.

3. Hesitation occurs at the second bracket: it does not match a memorised elementary Taylor series, and the instinct is to either look it up (Bessel function) or brute-force multiply the two series together. This is reinforced away by asking whether a name for the second bracket is actually needed for the next step, or whether its raw coefficients are already usable as they stand.

4. The trap is assuming that because one bracket became a clean closed form, the other must too, before the integral can be evaluated; in fact only one bracket needs to become closed-form, since the other's coefficients can be carried straight through term-by-term integration.

5. Recognise this family faster by treating the product of an obviously-nameable series with an unfamiliar one as a signal to integrate term by term against the unfamiliar series' raw coefficients, rather than searching for its name first.

6. Reusable rule: when an integral is a product of two power series and only one has an obvious closed form, do not force the other into closed form as well; integrate the closed-form factor against the unfamiliar series term by term, compute the resulting family of elementary integrals by a reduction formula, and look for its output to cancel structurally against the unfamiliar series' own coefficients.

▷ **Cue:** an integral is the product of two power series, one recognisable as a named function and one not $\longrightarrow$ **Move:** keep the recognisable factor in closed form, leave the other as a raw series, and integrate term by term against its coefficients rather than searching for a second closed form

▷ **Cue:** a term-by-term integral of $x^{2n+1}e^{-x^2/2}$ (or similar) appears against a coefficient built from the same double-factorial-type quantity $\longrightarrow$ **Move:** compute the integral by a reduction formula via integration by parts and look for the resulting expression to cancel against the given coefficient

1997 A4

1997 A4. Let G be a group with identity e and $\phi : G \to G$ a function such that

$$\phi(g_1)\phi(g_2)\phi(g_3) = \phi(h_1)\phi(h_2)\phi(h_3)$$

whenever $g_1g_2g_3 = e = h_1h_2h_3$. Prove that there exists an element $a \in G$ such that $\psi(x) = a\phi(x)$ is a homomorphism (i.e. $\psi(xy) = \psi(x)\psi(y)$ for all $x, y \in G$).

Fields: algebra (group theory; functional equations on groups) **Position:** A4 — mid-paper, but short once the element a is pinned down; the difficulty is entirely in not knowing where to start. **Answer:** Take $a = \phi(e)^{-1}$; then $\psi(x) = \phi(e)^{-1}\phi(x)$ is a homomorphism.

Layer 1 THE WHAT — surface analysis

The natural attempt. You want $\psi(xy) = \psi(x)\psi(y)$, so you look for a way to make the hypothesis produce $\phi(x)\phi(y)$ and $\phi(xy)$ in one equation. The hypothesis compares two triples multiplying to e, so the triple containing x and y has to be $(x, y, (xy)^{-1})$, and the triple containing xy has to have $(xy)^{-1}$ in it as well. The first thing that comes to hand is $(xy, (xy)^{-1}, e)$. The hypothesis gives

$$\phi(x)\phi(y)\phi\big((xy)^{-1}\big) = \phi(xy)\,\phi\big((xy)^{-1}\big)\,\phi(e).$$

Now you want to cancel $\phi((xy)^{-1})$, and you can't: on the left it sits on the far right, on the right it sits in the middle, with $\phi(e)$ stuck behind it. In a non-abelian group you cannot slide it past. Writing $u = \phi((xy)^{-1})$ and $c = \phi(e)$, you get

$$\phi(x)\phi(y) = \phi(xy)\,u\,c\,u^{-1},$$

a conjugate of c where you wanted something tidy. You then try to guess a: perhaps $a = \phi(e)$, perhaps $a = \phi(e)^{-1}$, perhaps something built from $\phi(g)\phi(g^{-1})$. Without a reason to prefer one, you test each against $\psi(xy) = \psi(x)\psi(y)$ and each test runs into the same commutation problem. This is the wall: the algebra is fine, but non-commutativity stops every cancellation, and you have no principled choice of a.

The model chain. *Step 1: a is forced.* Any homomorphism satisfies $\psi(e) = \psi(e)\psi(e)$, so $\psi(e) = e$. If $\psi(x) = a\phi(x)$ is to be a homomorphism, then $a\phi(e) = e$, so $a = \phi(e)^{-1}$. There is no choice to make. Write $c = \phi(e)$ and $\psi(x) = c^{-1}\phi(x)$.

Step 2: read the hypothesis as one equation. The triple (e, e, e) multiplies to e, so for every triple with $g_1g_2g_3 = e$,

$$\phi(g_1)\phi(g_2)\phi(g_3) = c^3. \qquad (*)$$

Step 3: c commutes with every value of ϕ. For any $g \in G$, both (g, e, g^{-1}) and (e, g, g^{-1}) multiply to e, so by $(*)$

$$\phi(g)\,c\,\phi(g^{-1}) = c\,\phi(g)\,\phi(g^{-1}).$$

Multiply on the right by $\phi(g^{-1})^{-1}$: $\phi(g)c = c\phi(g)$. Consequently c^{-1} also commutes with $\phi(g)$ (multiply $\phi(g)c = c\phi(g)$ by c^{-1} on both sides).

Step 4: the homomorphism property. For $x, y \in G$, both $(x, y, (xy)^{-1})$ and $(xy, e, (xy)^{-1})$ multiply to e, so

$$\phi(x)\phi(y)\phi\big((xy)^{-1}\big) = \phi(xy)\,c\,\phi\big((xy)^{-1}\big).$$

Cancel $\phi((xy)^{-1})$ on the right: $\phi(x)\phi(y) = \phi(xy)c = c\,\phi(xy)$, using Step 3. Multiply on the left by c^{-2}:

$$c^{-2}\phi(x)\phi(y) = c^{-1}\phi(xy).$$

Since c^{-1} commutes with $\phi(x)$, the left side equals $c^{-1}\phi(x)\,c^{-1}\phi(y)$. Hence $\psi(x)\psi(y) = \psi(xy)$ for all x, y. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains use the triple $(x, y, (xy)^{-1})$. The split is the *partner* triple and the order of work. The natural attempt pairs it with $(xy, (xy)^{-1}, e)$, which strands $\phi(e)$ behind the element to be cancelled, and it looks for a last. The model decides a first — by evaluating the desired conclusion at $x = e$ — and then pairs with $(xy, e, (xy)^{-1})$, putting the common factor $\phi((xy)^{-1})$ in the *same* position on both sides so it cancels.

> *Before I hunt for a, what must a be? A homomorphism sends e to e, so $a\phi(e) = e$ and $a = \phi(e)^{-1}$. There's nothing to search for.*

> *In a non-abelian group I can only cancel a factor if it sits at the same end on both sides. So I choose the second triple to put $\phi((xy)^{-1})$ at the right-hand end, and I let e sit in the middle.*

> *Now $\phi(e)$ is in the way, sitting between things. The hypothesis lets me move e around inside a triple for free — (g, e, g^{-1}) versus (e, g, g^{-1}) — so I can find out whether $\phi(e)$ commutes with everything.*

Layer 3 THE WHY

Start with what a homomorphism is: a map respecting products, and therefore sending the identity to the identity. The problem asks for an existence proof, and the existential "there exists a" is the thing that makes it feel open-ended. But the conclusion is so rigid that it evaluates itself at one point: $\psi(e) = e$. The moment you write the conclusion at $x = e$, the unknown a is determined. This is the general principle behind the first critical thought: *when asked to show some object exists with a property, first ask what the property forces the object to be*. Kedlaya's solution opens "we must in particular have this for $x = e$, and so we take $a = \phi(e)^{-1}$". That sentence is the whole strategic content of the problem, and it is not a guess; it is necessity run backwards into sufficiency.

Next, why does commutation matter, and why is it natural to expect it? Look at what the hypothesis says once you plug in (e, e, e): every product $\phi(g_1)\phi(g_2)\phi(g_3)$ along a triple with $g_1g_2g_3 = e$ equals the single element c^3. A condition of the form "this product doesn't depend on the triple" is a statement about *positions* inside a product, and in a group the positions only matter because of non-commutativity. The hypothesis is invariant under inserting e anywhere in the triple (the triples (g, e, g^{-1}) and (e, g, g^{-1}) both multiply to e), so it must be telling you that $\phi(e)$ can be moved around — that is, $\phi(e)$ commutes with the values of ϕ. You do not need to be told this; you need to ask "what does the freedom to move e in the triple buy me?", and the answer is one line.

To see that this commuting property is not an accident of the proof but the actual shape of every such ϕ, run the converse. Take any homomorphism χ and any c commuting with every $\chi(g)$, and set $\phi = c\chi$. Then along any triple with $g_1g_2g_3 = e$,

$$\phi(g_1)\phi(g_2)\phi(g_3) = c\chi(g_1)c\chi(g_2)c\chi(g_3) = c^3\chi(g_1g_2g_3) = c^3,$$

so the hypothesis holds. The problem is therefore the statement that the hypothesis *characterises* "a homomorphism times a constant commuting with its image". Once you see the family, the proof writes itself: peel the constant off (Step 1), confirm it commutes (Step 3), and what remains is a homomorphism (Step 4).

Finally, the choice of triples is dictated by the one axiom of cancellation available in a group: from $pr = qr$ you get $p = q$ by multiplying by r^{-1} *on the right*. So both triples must end in the same element, and the triples $(x, y, (xy)^{-1})$ and $(xy, e, (xy)^{-1})$ are the only natural pair whose last entries agree and whose first entries produce $\phi(x)\phi(y)$ and $\phi(xy)$. The natural attempt

chose $(xy, (xy)^{-1}, e)$ because it looked equally valid — and it is, but it puts the common factor in the wrong slot. That is the whole trap: in a non-abelian setting, *where* a factor sits is part of the information.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "there exists an element a such that $\psi(x) = a\phi(x)$ is a homomorphism". Write down: $\psi(e) = e$, so $a = \phi(e)^{-1}$.

2. Necessity first: determine the object from the conclusion evaluated at a special point, then prove sufficiency with that object fixed. Combined with choosing substitutions into the hypothesis that align the factors you want to cancel.

3. Hesitation comes when the cancellation fails because of $\phi(e)$ sitting in the middle. Reinforce it away by asking directly whether $\phi(e)$ commutes with $\phi(g)$, and testing that with two triples differing only by where e sits.

4. The trap is pairing $(x, y, (xy)^{-1})$ with $(xy, (xy)^{-1}, e)$: legal, but the common factor is not in the same position on both sides, so it cannot be cancelled in a non-abelian group. A secondary trap is treating "there exists a" as a search.

5. Slogan: *an existence claim with a rigid conclusion names its own witness — evaluate the conclusion at the identity.* And: *in a non-abelian computation, arrange the substitution so the factor to cancel sits at the same end.*

6. Decision rule: **when asked to show that some constant or parameter can be chosen to make a property hold, first derive what the property forces the constant to be at the simplest input; then prove the property with that value.**

▷ **Cue:** "prove there exists a such that $a\phi$ is a homomorphism" $\longrightarrow$ **Move:** evaluate at e: a homomorphism fixes e, so $a = \phi(e)^{-1}$

▷ **Cue:** a hypothesis "the product along any triple with $g_1g_2g_3 = e$ is the same" $\longrightarrow$ **Move:** plug in (e, e, e): every such product equals $\phi(e)^3$

▷ **Cue:** a functional equation on a group where some factor $\phi(e)$ blocks a cancellation $\longrightarrow$ **Move:** compare triples differing only by the position of e to show $\phi(e)$ commutes with every $\phi(g)$

▷ **Cue:** need to cancel a group element from both sides of an equation in a non-abelian group $\longrightarrow$ **Move:** choose substitutions so that element is at the same end (right or left) on both sides

▷ **Cue:** a functional condition on a group that looks like "homomorphism up to something" $\longrightarrow$ **Move:** guess the family (homomorphism times a central-ish constant), check it satisfies the condition, then prove the converse

1997 A5

1997 A5. Let N_n denote the number of ordered n-tuples of positive integers $(a_1, a_2, \ldots, a_n)$ such that $1/a_1 + 1/a_2 + \ldots + 1/a_n = 1$. Determine whether N_{10} is even or odd.

Fields: combinatorics (parity via involutions); number theory (Egyptian fractions) **Position:** A5 — late in the paper, but a short problem once the right tool is named; the cost is entirely in not attempting a count. **Answer:** N_{10} is odd.

Layer 1 THE WHAT — surface analysis

The natural attempt. You read "the number of ordered n-tuples" and start counting. Ten unit fractions summing to 1 is far too many to enumerate, so you look for a recursion: fix a_1, and count 9-tuples summing to $1 - 1/a_1$. That needs a formula for the number of k-tuples summing to an arbitrary rational, and no such formula exists. You compute small cases to look for a pattern: $N_1 = 1$, $N_2 = 1$ (only $(2, 2)$), $N_3 = 10$ (the arrangements of $\{2, 3, 6\}$, $\{2, 4, 4\}$, $\{3, 3, 3\}$,

giving $6+3+1$). By hand N_4 is already a serious job (it is 215), and the parities $1,1,0,1$ show no obvious rhythm you would trust to extrapolate to $n=10$. You might switch to counting unordered solutions and multiplying by multinomial coefficients, which is how the 10 above was obtained — but for $n=10$ the unordered solutions are themselves astronomically many. The wall: the question asks for one bit of information, and every method you have computes the whole number.

The model chain. *Step 1: pairing kills most tuples.* Let S be the set of solutions. The map σ_{12} that swaps a_1 and a_2 sends solutions to solutions and is its own inverse. Its non-fixed points split into pairs $\{t, \sigma_{12}(t)\}$, so

$$|S| \equiv |\{t \in S : a_1 = a_2\}| \pmod 2.$$

Apply the same reasoning to σ_{34} acting on the set $\{a_1 = a_2\}$ (it preserves that condition), then σ_{56}, σ_{78}, $\sigma_{9,10}$. So N_{10} has the parity of the number of solutions with

$$a_1 = a_2,\ a_3 = a_4,\ a_5 = a_6,\ a_7 = a_8,\ a_9 = a_{10}.$$

Writing $b_1, \ldots, b_5$ for these common values, this is the number of ordered 5-tuples of positive integers with

$$\frac{2}{b_1} + \frac{2}{b_2} + \frac{2}{b_3} + \frac{2}{b_4} + \frac{2}{b_5} = 1.$$

Step 2: repeat. Swapping $b_1 \leftrightarrow b_2$ and then $b_3 \leftrightarrow b_4$ reduces to $b_1 = b_2$, $b_3 = b_4$:

$$\frac{4}{b_1} + \frac{4}{b_3} + \frac{2}{b_5} = 1,$$

counted over ordered triples (b_1, b_3, b_5). Swapping $b_1 \leftrightarrow b_3$ reduces to $b_1 = b_3$:

$$\frac{8}{p} + \frac{2}{q} = 1, \qquad p, q \text{ positive integers.}$$

Step 3: count exactly. Clearing denominators, $pq - 2p - 8q = 0$, i.e. $(p-8)(q-2) = 16$. Since $8/p < 1$ and $2/q < 1$, we need $p > 8$ and $q > 2$, so both factors are positive divisors of 16. (If both factors were negative we would need $q - 2 \in \{-1\}$ and $p - 8 = -16$, impossible for positive p.) The divisors $1, 2, 4, 8, 16$ give exactly five solutions:

$$(p,q) = (9,18),\ (10,10),\ (12,6),\ (16,4),\ (24,3).$$

Five is odd, so N_{10} is odd. □

Check. The same reduction predicts the parity of N_4 from the number of pairs with $2/b_1 + 2/b_2 = 1$, namely $(3,6), (4,4), (6,3)$: three, odd. Indeed $N_4 = 215$. For N_6 it predicts the parity of the count for $4/p + 2/q = 1$, i.e. $(p-4)(q-2) = 8$: four solutions, even; and $N_6 = 2\,025\,462$. A direct machine count of ordered 5-tuples with $\sum 1/b_i = \frac{1}{2}$ gives 307 831, odd, agreeing with Step 3.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains read the same statement. The natural attempt treats "determine whether N_{10} is even or odd" as "compute N_{10}, then look at the last digit". The model treats "even or odd" as the question itself and reaches for the tool that answers parity questions without computing the number: a pairing of the objects, where only unpaired objects contribute.

> *I'm asked for one bit, not the number. Parity questions are answered by pairing things off. What pairs up ordered tuples for free? Swapping two entries — a tuple and its swap are different unless the two entries are equal.*

> *After the first swap, the survivors still have lots of symmetry left: swap the next two entries. Keep going until what's left is small enough to count by hand.*

Layer 3 THE WHY

Why is the pairing the natural move and not a trick? Go to the definition of parity. A finite set has even size exactly when it can be split into disjoint pairs. So the most basic statement about parity is: *if you can pair up some of the elements, the parity of the whole equals the parity of what's left unpaired.* Formally, if $\sigma : S \to S$ satisfies $\sigma(\sigma(t)) = t$ (an *involution*), then the orbits of σ have size 1 or 2, and

$$|S| \equiv |\mathrm{Fix}(\sigma)| \pmod 2.$$

This is not a special technique; it is what "even" means.

So the real question is: where does the involution come from? From the word "ordered" in the statement. The equation $\sum 1/a_i = 1$ is symmetric in the a_i — any permutation of a solution is a solution. Counting *ordered* tuples of a symmetric equation means that every solution comes with all its rearrangements, and a transposition of two coordinates is the simplest involution on that set. The problem-setter chose "ordered" deliberately: for unordered solutions no such free pairing exists. Whenever a counting problem hands you a symmetric condition and asks for a count of ordered objects modulo 2, the symmetry group is there to be used.

Why five transpositions and then more? Each round halves the number of free variables and doubles the coefficients. The fixed points of the first round are the solutions of $\sum 2/b_i = 1$ in five variables. This equation is still symmetric, so the same trick applies. You stop when you reach an equation you can solve outright, and the two-variable equation $8/p + 2/q = 1$ is exactly the kind you can: multiply out, complete the product, and read solutions off the divisors of a fixed integer. That last move — turning $pq - 2p - 8q = 0$ into $(p-8)(q-2) = 16$ — is the standard "Simon's favourite factoring" step: add the product of the coefficients to both sides so the left factorises. It is forced by wanting a product equal to a constant, since then solutions correspond to divisor pairs.

One rigour point the official solution skims: each later swap must be an involution on the *current* set of survivors, not on all solutions. Swapping a_3 and a_4 preserves the condition $a_1 = a_2$, so it does map the set $\{a_1 = a_2\}$ to itself. Similarly, the last swap acts on ordered triples (b_1, b_3, b_5) solving $4/b_1 + 4/b_3 + 2/b_5 = 1$; read back in the five-variable form it exchanges the block (b_1, b_2) with the block (b_3, b_4), which preserves both equalities $b_1 = b_2$ and $b_3 = b_4$. State this once; a grader will look for it.

A second rigour point: the grouping is arbitrary, and you are free to choose it. With ten variables and ten being $2 \cdot 5$, the first round leaves five, which is odd, so one variable (b_5) is left unpaired in round two. That is why the final equation has coefficients 8 and 2 rather than being fully symmetric. Any grouping works; the answer (a parity) does not depend on it, because each step is an exact congruence.

Kedlaya's solution says "we may discard any solutions for which $a_1 \neq a_2$, since those come in pairs". That sentence is the involution principle stated in eight words. The archive's job is to say why a solver reaches for it: because the question asks for a parity, and parity is pairing.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "determine whether N_{10} is even or odd", together with "ordered n-tuples" of a symmetric equation. Write down: *swap a_1, a_2; non-fixed solutions pair off;* $N_{10} \equiv \#\{a_1 = a_2\} \pmod 2$.
2. Involution counting: $|S| \equiv |\mathrm{Fix}(\sigma)| \pmod 2$ for any involution σ of S, iterated until the fixed-point set is small, then a direct count by factoring.
3. Hesitation occurs after the first round, when the equation $\sum 2/b_i = 1$ still looks uncountable. Reinforce it away by noticing that the new equation is again symmetric, so the same move applies again.

4. The trap is trying to compute N_{10}, or to find a pattern in the parities of $N_1, \ldots, N_6$. Both are hopeless: the numbers grow super-exponentially and the parities show no rhythm. A secondary trap is forgetting that $p > 8$ and $q > 2$ when counting divisor pairs, or counting negative divisor pairs.

5. Slogan: *a parity question about a set with a symmetry is a question about the fixed points of that symmetry.*

6. Decision rule: **when asked whether a count is even or odd, do not compute it — find an involution on the set, discard the pairs it creates, and count only the fixed points; iterate if the fixed-point set is still too large.**

▷ **Cue:** "determine whether the number of … is even or odd" $\longrightarrow$ **Move:** find an involution on the set; $|S| \equiv |\mathrm{Fix}| \pmod 2$

▷ **Cue:** counting ordered tuples satisfying a symmetric equation $\longrightarrow$ **Move:** transpose two coordinates: the non-fixed tuples pair off

▷ **Cue:** after one round of pairing, the surviving equation is still symmetric $\longrightarrow$ **Move:** apply another transposition to the survivors, checking it preserves the earlier equalities

▷ **Cue:** an equation $pq - \alpha p - \beta q = 0$ in positive integers $\longrightarrow$ **Move:** rewrite as $(p-\beta)(q-\alpha) = \alpha\beta$ and count divisor pairs, checking sign constraints

1997 A6

1997 A6. For a positive integer n and any real number c, define x_k recursively by $x_0 = 0$, $x_1 = 1$, and for $k \geq 0$,

$$x_{k+2} = \frac{c x_{k+1} - (n-k) x_k}{k+1}.$$

Fix n and then take c to be the largest value for which $x_{n+1} = 0$. Find x_k in terms of n and k, $1 \leq k \leq n$.

Fields: algebra (recurrences, generating functions); analysis (first-order linear ODEs); linear algebra (eigenvalues) **Position:** A6 — last problem of the morning, and hard: the answer is guessable from small cases, but proving that $c = n-1$ is the *largest* root needs a real idea. **Answer:** $c = n-1$ and $x_k = \binom{n-1}{k-1}$ for $1 \leq k \leq n$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You compute. With $x_0 = 0$, $x_1 = 1$:

$$x_2 = c, \qquad x_3 = \frac{c^2 - (n-1)}{2}, \qquad x_4 = \frac{c\,x_3 - (n-2)c}{3}.$$

For $n = 2$: $x_3 = \frac{c^2-1}{2}$, roots ± 1, largest $c = 1$, giving $x_1, x_2 = 1, 1$. For $n = 3$: $x_4 = \frac{c(c^2-2)/2 - c}{3} = \frac{c(c^2-4)}{6}$, roots $0, \pm 2$, largest $c = 2$, giving $1, 2, 1$. For $n = 4$ you find roots $\pm 1, \pm 3$ and the values $1, 3, 3, 1$. The pattern is loud: $c = n-1$ and $x_k = \binom{n-1}{k-1}$, the roots are $n-1, n-3, \ldots, -(n-1)$.

Checking that $c = n-1$ works is a binomial identity. With $x_k = \binom{n-1}{k-1}$ you need

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

Using $(k+1)\binom{n-1}{k+1} = (n-1-k)\binom{n-1}{k}$ and $(n-k)\binom{n-1}{k-1} = k\binom{n-1}{k}$, both sides equal $(n-1-k)\binom{n-1}{k}$. And $x_{n+1} = \binom{n-1}{n} = 0$. So $c = n-1$ is a root.

Now the wall. The problem says *largest*. You have one root and a verified formula, but nothing stops $x_{n+1}(c)$ from having a root larger than $n-1$. You try to show $x_{n+1}(c) > 0$ for

$c > n-1$ by induction on k — the terms x_k stay positive and grow? — but the subtraction $-(n-k)x_k$ spoils any easy monotonicity, and the induction does not close. Many strong candidates stop here with a correct answer and an incomplete proof.

The model chain. *Step 1: count the roots.* By induction x_{k+1} is a polynomial in c of degree exactly k with leading coefficient $1/k!$: true for $k = 0, 1$ ($x_1 = 1$, $x_2 = c$), and the recurrence multiplies the leading term of x_{k+1} by $c/(k+1)$ while x_k has lower degree. So $x_{n+1}(c)$ has degree n and at most n real roots. If we exhibit n distinct roots, we have all of them and can read off the largest.

Step 2: package the sequence. Let $f(t) = \sum_{k\geq 1} x_k t^{k-1}$ as a formal power series. Multiply the recurrence by $(k+1)$ and read it as a coefficient identity. The coefficient of t^k in $f'(t)$ is $(k+1)x_{k+2}$; in $t^2 f'(t)$ it is $(k-1)x_k$; in $tf(t)$ it is x_k; in $f(t)$ it is x_{k+1}. The recurrence $(k+1)x_{k+2} = cx_{k+1} - (n-k)x_k = cx_{k+1} - (n-1)x_k + (k-1)x_k$ is therefore, for every $k \geq 0$ (at $k = 0$ all x_0-terms vanish),

$$(1-t^2)f'(t) = \big(c-(n-1)t\big)f(t), \qquad f(0) = 1. \tag{†}$$

Conversely, (†) with $f(0) = 1$ determines every coefficient of f through the recurrence, so any power series satisfying (†) with constant term 1 is our f.

Step 3: solve. Divide by $(1-t^2)f$ and split into partial fractions:

$$\frac{f'}{f} = \frac{c-(n-1)t}{(1-t)(1+t)} = \frac{A}{1+t} - \frac{r}{1-t}, \qquad A = \frac{n-1+c}{2},\ r = \frac{n-1-c}{2}.$$

Note $A + r = n-1$. So $f(t) = (1+t)^A(1-t)^r$. When $r \in \{0, 1, \ldots, n-1\}$, both exponents are non-negative integers and f is a polynomial of degree $A + r = n-1$. To avoid any appeal to real powers, verify directly: for $f = (1+t)^{n-1-r}(1-t)^r$, logarithmic differentiation gives

$$(1-t^2)f' = f\big[(n-1-r)(1-t) - r(1+t)\big] = f\big[(n-1-2r) - (n-1)t\big],$$

which is (†) with $c = n-1-2r$. So for $c = n-1-2r$ the sequence x_k is the coefficient sequence of a polynomial of degree $n-1$, and x_{n+1}, the coefficient of t^n, is 0.

Step 4: conclude. The values $c = n-1-2r$, $r = 0, 1, \ldots, n-1$, are n distinct roots of the degree-n polynomial $x_{n+1}(c)$, so they are all its roots. The largest is $c = n-1$ ($r = 0$), where $f(t) = (1+t)^{n-1}$ and

$$x_k = \binom{n-1}{k-1}, \qquad 1 \leq k \leq n. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains find $c = n-1$ and $x_k = \binom{n-1}{k-1}$. The split comes at "largest". The natural attempt tries to show no root exceeds $n-1$ by bounding $x_{n+1}(c)$ for $c > n-1$. The model never compares roots at all: it counts them. A degree-n polynomial with n known roots has no others.

> *I can't bound this polynomial for large c, but I know its degree. If I can find all n roots, "largest" is a matter of reading the list. So the pattern I should be chasing is the full list of roots — $n-1, n-3, \ldots, 1-n$ — not the single top one.*

> *The binomial answer $(1+t)^{n-1}$ is a guess about the generating function. The other roots should also be nice polynomials. The recurrence has coefficients linear in k, so the generating function obeys a first-order ODE; let me write it down and solve it for general c.*

Layer 3 THE WHY

Why is counting roots the natural route to "largest", and why does a generating function appear? Take them in turn.

First, "largest root" is a comparison, and comparisons among roots of a polynomial are hard in general. But there is one situation where they are free: when you know every root. A real polynomial of degree n has at most n roots — that is the factor theorem, applied n times. So the degree bound converts "find the largest root" into "find n roots". Your small cases already told you the roots for $n = 2, 3, 4$: ± 1; $0, \pm 2$; $\pm 1, \pm 3$. The pattern is the arithmetic progression $n-1, n-3, \ldots, 1-n$, symmetric about 0 (as it must be: replacing c by $-c$ and x_k by $(-1)^{k-1}x_k$ preserves the recurrence, so the roots come in $\pm$ pairs). The natural attempt looked only at the top root because the problem only named the top root. The problem was hiding the easier target behind the harder one.

Second, why a generating function? Look at the recurrence with $(k+1)$ multiplied through:

$$(k+1)x_{k+2} - cx_{k+1} + (n-k)x_k = 0.$$

The coefficients are polynomials of degree 1 in the index k. In a power series, multiplying the k-th coefficient by k corresponds to the operator $t\frac{d}{dt}$. So a recurrence with coefficients linear in k becomes a differential equation of first order for the generating function, with polynomial coefficients. A first-order linear ODE is solvable by separation, $f'/f =$ rational function, and partial fractions finish it. This is the "of course" of the problem: constant-coefficient recurrences become rational generating functions, and index-linear recurrences become first-order ODEs. Kedlaya's solution announces "x_k is the coefficient of t^{k-1} in $(1-t)^r(1+t)^{n-1-r}$" as a claim to be verified. Here is where it comes from: it is the solution of (†) with the exponent $r = \frac{n-1-c}{2}$ written in terms of c.

Third, why exactly those c? The problem's condition $x_{n+1} = 0$ has a structural meaning you should check before anything else. Put $k = n$ into the recurrence: $x_{n+2} = \frac{cx_{n+1} - 0 \cdot x_n}{n+1}$. The coefficient $(n-k)$ *vanishes* at $k = n$. So if $x_{n+1} = 0$ then $x_{n+2} = 0$, and then every later term is 0 too. In other words, $x_{n+1} = 0$ is precisely the condition that the generating function is a polynomial of degree at most $n-1$. And $(1+t)^A(1-t)^r$ with $A + r = n-1$ is visibly a polynomial when A and r are non-negative integers — which is $c = n-1-2r$, $r = 0, \ldots, n-1$. You do not need to prove that no other c gives a polynomial: the degree count in Step 4 does that for you. The design of the recurrence (the factor $n-k$) is what makes truncation possible, and the ODE tells you which truncations occur.

There is a second route to "largest" worth knowing, due to Kuperberg and quoted in the official solution. Since $x_0 = 0$ and (at a root) $x_{n+1} = 0$, the equations $i\,x_{i+1} + (n-i+1)x_{i-1} = c\,x_i$ for $i = 1, \ldots, n$ say that $(x_1, \ldots, x_n)$ is an eigenvector with eigenvalue c of a tridiagonal matrix with non-negative off-diagonal entries. By Perron–Frobenius the eigenvector with all entries positive belongs to the eigenvalue of largest modulus. The binomial vector is positive, so $n-1$ is the largest. This uses a heavier theorem; the root-counting route above is self-contained, and I would write that one in the exam.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the largest value for which $x_{n+1} = 0$". Write down: *x_{n+1} is a polynomial in c of degree n; find all n roots.*

2. Compute small cases to guess all roots; then encode the recurrence as a first-order ODE for the generating function $f(t) = \sum x_k t^{k-1}$, solve by partial fractions, and verify the polynomial solutions.

3. Hesitation occurs after verifying $c = n-1$, when "largest" seems to require an inequality. Reinforce it away with the degree count: n distinct roots of a degree-n polynomial are all of them.

4. The trap is proving only that $c = n-1$ is *a* root and asserting it is the largest from small cases. A second trap is trying to show $x_{n+1}(c) \neq 0$ for $c > n-1$ by a positivity induction; the minus sign in the recurrence defeats it.

5. Slogan: *to find the extreme root of a polynomial of known degree, find all the roots.* And: *a recurrence whose coefficients are linear in the index is a first-order ODE in disguise.*

6. Decision rule: **when asked for an extreme member of a finite set defined by an equation, bound the size of the set first; if you can exhibit that many members, the extreme one is read off, with no inequality needed.**

- ▷ **Cue:** "the largest c for which" a polynomial-in-c expression vanishes $\longrightarrow$ **Move:** compute its degree; exhibit that many distinct roots; read off the largest
- ▷ **Cue:** a recurrence $(k+1)x_{k+2} = \alpha x_{k+1} + (\beta + \gamma k)x_k$ with coefficients linear in k $\longrightarrow$ **Move:** form $f(t) = \sum x_k t^k$; the recurrence becomes a first-order linear ODE, solved by separation and partial fractions
- ▷ **Cue:** a coefficient like $(n-k)$ in a recurrence that vanishes at $k = n$ $\longrightarrow$ **Move:** a single zero term $x_{n+1} = 0$ forces all later terms to vanish: the generating function is a polynomial
- ▷ **Cue:** small cases produce binomial coefficients $\longrightarrow$ **Move:** guess a generating function of the form $(1+t)^a(1-t)^b$ and verify by logarithmic differentiation
- ▷ **Cue:** a three-term recurrence with boundary conditions $x_0 = 0 = x_{n+1}$ $\longrightarrow$ **Move:** rewrite as an eigenvector equation for a tridiagonal matrix; consider Perron–Frobenius if entries are non-negative

Putnam 1997 — Paper B

1997 B1

1997 B1. Let $\{x\}$ denote the distance between the real number x and the nearest integer. For each positive integer n, evaluate

$$F_n = \sum_{m=1}^{6n-1} \min(\{\frac{m}{6n}\}, \{\frac{m}{3n}\}).$$

(Here $\min(a, b)$ denotes the minimum of a and b.)

Fields: number theory, fractional parts, piecewise-linear summation **Position:** B1 — the opener **Answer:** $F_n = n$ for every positive integer n.

Layer 1 THE WHAT — surface analysis

The natural attempt. Compute F_1 by hand: for $6n = 6$, $m = 1, \ldots, 5$ gives $\{m/6\} = \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{1}{3}, \frac{1}{6}$ and $\{m/3\} = \frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}$, so the minima are $\frac{1}{6}, \frac{1}{3}, 0, \frac{1}{3}, \frac{1}{6}$, summing to 1. Compute F_2 the same way and it comes to 2. The pattern $F_n = n$ is now visible, and the honest next move is to try to prove it by relating F_{n+1} to F_n — an induction on n.

This is where the natural attempt stalls. Going from n to $n + 1$ changes *both* the number of terms in the sum ($6n - 1$ to $6n + 5$) and the mesh of every fraction in it ($m/6n$ to $m/(6n + 6)$): there is no reindexing $m \mapsto m'$ that turns the n-th sum into a sub-sum or a shift of the $(n + 1)$-th, because $6n$ and $6(n + 1)$ are not related by a shift, only by a change of denominator that moves every single term. Writing out $F_2 - F_1$ term-by-term to look for a telescoping $+1$ produces two sums with no common index set at all, and the induction dies before it starts. The pattern is correct but the proof strategy suggested by "prove a formula in n" — relate consecutive n — is the wrong lever for a sum whose *shape*, not just its length, depends on n.

The model chain. Substitute $\theta = m/(6n)$. As m runs over $1, \ldots, 6n - 1$, θ runs over every point of the form $k/(6n)$ strictly between 0 and 1, and

$$\min\left(\left\{\frac{m}{6n}\right\}, \left\{\frac{m}{3n}\right\}\right) = \varphi(\theta), \qquad \varphi(\theta) := \min(\{\theta\}, \{2\theta\}),$$

where φ is a single function of $\theta \in [0, 1)$ that does not depend on n at all. So F_n is not "a new sum for each n"; it is the same fixed function φ, sampled at the $6n - 1$ equally spaced points of mesh $1/(6n)$.

Find where φ changes its linear formula. $\{\theta\} = \{2\theta\}$ has solutions at $\theta = 0, \frac{1}{3}, \frac{2}{3}$ (check: $\{\frac{1}{3}\} = \frac{1}{3} = \{\frac{2}{3}\}$; $\{\frac{2}{3}\} = \frac{1}{3} = \{\frac{4}{3}\}$), and $\{\theta\}$ itself has a corner at $\theta = \frac{1}{2}$. These points — $0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1$ — cut $[0, 1)$ into four intervals on which φ is exactly one linear function: on $[0, \frac{1}{3})$, $\{\theta\} = \theta$ wins; on $[\frac{1}{3}, \frac{1}{2})$, $\{2\theta\} = 1 - 2\theta$ wins; on $[\frac{1}{2}, \frac{2}{3})$, $\{2\theta\} = 2\theta - 1$ wins; on $[\frac{2}{3}, 1)$, $\{\theta\} = 1 - \theta$ wins.

The sample points $\theta = m/(6n)$ hit $\frac{1}{3}, \frac{1}{2}, \frac{2}{3}$ exactly at $m = 2n, 3n, 4n$ — always integers, so the sample mesh respects the corners for every n. Summing φ at $6n - 1$ equally spaced points over one period is therefore four arithmetic series, and each one scales linearly in n because

doubling the mesh density of a fixed-shape piecewise-linear function summed on a fixed set of sub-intervals exactly doubles the sum. Carrying out the four arithmetic series (as in the official solution) gives $F_n = n$ for every n: the four pieces sum to $\frac{(2n-1)n}{6n} \cdot (\text{const}) + \cdots$, and collecting them (a routine but not illuminating computation) returns exactly n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin by computing $F_1 = 1$, $F_2 = 2$ and noticing the pattern $F_n = n$. They split at the choice of what to induct on. Quoting the natural attempt: "relate F_{n+1} to F_n." The model chain instead asks: "what does $\min(\{\theta\}, \{2\theta\})$ look like as a fixed function of $\theta \in [0, 1)$?" — discarding n as the variable to induct on and treating it instead as a sampling density.

> *The upper limit $6n$ and the factor of 2 between the two denominators $6n$ and $3n$ are not incidental to this particular n — if I substitute $\theta = m/(6n)$, both fractional-part arguments become $\{\theta\}$ and $\{2\theta\}$, a function of θ alone with n controlling only how finely I sample it. So the right question is not "how does the sum change from n to $n+1$" but "where does this fixed function change formula, and does my sample mesh always land on those changeover points" — and it does, because $\frac{1}{3} \cdot 6n$, $\frac{1}{2} \cdot 6n$, $\frac{2}{3} \cdot 6n$ are always integers.*

Layer 3 THE WHY

Why does recognising $\theta = m/(6n)$ turn an apparently n-dependent problem into an n-independent one? Because the summand $\min(\{m/6n\}, \{m/3n\})$ is homogeneous under the rescaling $m \mapsto m/(6n)$: both fractional-part arguments are ratios with $6n$ (or a fixed multiple of it) in the denominator, so writing $\theta = m/(6n)$ is not a substitution invented for convenience, it is the substitution that makes the two arguments $\{\theta\}$ and $\{2\theta\}$ literally stop mentioning n. Once n has been pushed entirely into "how many equally spaced samples of φ I take, and at what mesh," the only way n can still affect the answer is through the *number of complete linear pieces* it samples — and since φ is periodic-shaped with corners at rational points independent of n, and the mesh $1/(6n)$ always lands exactly on those corners (because $\frac{1}{3}, \frac{1}{2}, \frac{2}{3}$ have denominators dividing 6), every value of n samples the *same* four-piece linear shape, just with n times as many points per piece. Summing a linear function at k equally spaced points is a fixed arithmetic-series formula linear in k; that is the entire reason F_n comes out proportional to n, and it is invisible if n is treated as the thing to induct on rather than as a magnification factor on a fixed picture.

An induction on n was never going to work cleanly here because induction wants a recursive relationship between consecutive instances of the *same* object, and F_n and F_{n+1} are sums over genuinely different, non-nested index sets. The fix is not a cleverer induction; it is recognising that n does not belong in the summand at all once you look at it correctly.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the two fractional-part arguments $m/(6n)$ and $m/(3n)$ differ by exactly a factor of 2, and both have n appearing only through the same denominator $6n$. On seeing it, write $\theta = m/6n$ and rewrite the summand as $\min(\{\theta\}, \{2\theta\})$ before doing anything else.
2. That cue suggests treating F_n as a fixed piecewise-linear function of $\theta \in [0, 1)$ sampled at $6n - 1$ equally spaced points, rather than as a sequence in n to be inducted on.
3. Hesitation occurs right after computing $F_1 = 1, F_2 = 2$ and seeing the pattern: the instinct is to reach for induction on n immediately, since that is the default proof technique for "prove a formula holds for all n." This is reinforced away by checking, before committing to induction, whether consecutive instances of the sum are actually related by a shift or sub-sum relationship — here they are not, which is the signal to look for a substitution that removes n from the summand instead.

4. The trap is assuming that any formula indexed by n should be proved by relating instance n to instance $n+1$. Here the sum's underlying shape does not change with n at all — only the sampling density does — so induction on n attacks a relationship that does not exist, and a lot of algebra can be spent trying to force it before noticing the index sets never align.

5. Recognise this family faster: whenever a sum's bound and every denominator inside it share the same multiple of n (here, everything is $m/(6n)$ or a fixed rational multiple of it), substitute to make that shared factor explicit and check whether the summand becomes a function of a single rescaled variable with n pushed entirely into the sample count.

6. Reusable rule: before inducting on n in a sum $\sum_{m=1}^{f(n)} h(m,n)$, check whether $h(m,n)$ is actually a fixed function of $\theta = m/f(n)$ (or another natural rescaling). If it is, the problem is "sum a fixed piecewise-linear (or otherwise simple) function at an n-dependent mesh," which is solved by locating the function's own breakpoints once, not by relating consecutive values of n.

▷ **Cue:** a sum $\sum_{m=1}^{f(n)} h(m,n)$ where every argument of h is a ratio with the same multiple of n in its denominator $\longrightarrow$ **Move:** substitute $\theta = m/f(n)$ to rewrite h as a fixed function of θ alone, turning the problem into summing a fixed shape at an n-dependent mesh, not inducting on n

▷ **Cue:** a piecewise-linear function built from $\{\theta\}$ and $\{k\theta\}$ (a fractional part and an integer multiple of it) $\longrightarrow$ **Move:** find its corners by solving $\{\theta\} = \{k\theta\}$ (the crossover points) together with the corners of $\{\theta\}$ itself, not by taking the corners of $\{k\theta\}$ in isolation — a corner of the non-winning function does not create a new region for the minimum

1997 B2

1997 B2. Let f be a twice-differentiable real-valued function satisfying

$$f(x) + f''(x) = -xg(x)f'(x),$$

where $g(x) \geq 0$ for all real x. Prove that $|f(x)|$ is bounded.

Fields: ordinary differential equations, energy/Lyapunov methods **Position:** B2 **Answer:** Proof (not a numeral): with $E(x) = f(x)^2 + f'(x)^2$, a single computation of $E'(x)$ shows $|f(x)| \leq \sqrt{f(0)^2 + f'(0)^2}$ for every real x.

Layer 1 THE WHAT — surface analysis

The natural attempt. The equation $f'' + f = -xg(x)f'$ is $f'' + f = 0$ (bounded, oscillatory solutions) with an extra term on the right. The reflex is to try to control that extra term: multiply through by f', since $f' \cdot f'' = \frac{1}{2}\frac{d}{dx}(f')^2$ and $f' \cdot f = \frac{1}{2}\frac{d}{dx}f^2$ are both exact derivatives, and this is exactly what makes energy methods work for $f'' + f = 0$. Define $E(x) = f(x)^2 + f'(x)^2$ and compute

$$E'(x) = 2f(x)f'(x) + 2f'(x)f''(x) = 2f'(x)\big(f(x) + f''(x)\big) = 2f'(x) \cdot \big(-xg(x)f'(x)\big) = -2xg(x)\, f'(x)^2.$$

For $x \geq 0$ this is ≤ 0 (since $g \geq 0$), so E is non-increasing on $[0, \infty)$ and $E(x) \leq E(0)$ there — boundedness for $x \geq 0$ is secured cleanly.

But the same line, read at $x < 0$, gives $E'(x) = -2xg(x)f'(x)^2 \geq 0$: E is now shown *increasing* as x increases through negative values, which sounds like the wrong direction to conclude anything about $x < 0$ directly, and the natural reaction is to abandon this computation for $x < 0$ and reach for a separate device. The device that suggests itself is symmetry: define $h(x) = f(-x)$, check that h satisfies the same type of equation with $g(-x)$ in place of $g(x)$ (still ≥ 0, since $g \geq 0$ everywhere by hypothesis), and re-run the entire $x \geq 0$ argument on h to get

$|h(x)| = |f(-x)|$ bounded for $x \geq 0$, i.e. $|f(x)|$ bounded for $x \leq 0$. This works, and it is exactly the route the official solution takes ("it suffices to show $|f(x)|$ is bounded for $x \geq 0$, since $f(-x)$ satisfies the same equation as $f(x)$"). It costs a second full argument, however: setting up h, re-deriving that its governing g is still non-negative, and re-running the energy computation from scratch.

The model chain. Do not abandon the single formula $E'(x) = -2xg(x)f'(x)^2$ at $x < 0$ — read it correctly instead. It holds for *every* real x, because the governing equation $f + f'' = -xg(x)f'$ itself holds for every real x; nothing about its derivation used $x \geq 0$. Its sign is controlled entirely by $-2x$, since $g(x) \geq 0$ and $f'(x)^2 \geq 0$ are non-negative factors regardless of the sign of x. So:

- on $[0, \infty)$: $-2x \leq 0$, so $E'(x) \leq 0$, E is non-increasing, and for any $x \geq 0$, $E(x) \leq E(0)$;
- on $(-\infty, 0]$: $-2x \geq 0$, so $E'(x) \geq 0$, E is non-decreasing, and for any $x \leq 0$ (so $x \leq 0$), non-decreasingness gives $E(x) \leq E(0)$ directly — the value at an earlier point of a non-decreasing function is at most the value at a later point, and 0 is later than any $x \leq 0$.

Both bullets conclude $E(x) \leq E(0)$, from the one formula, on the two halves of the real line separately and without introducing $f(-x)$ at all. Hence $f(x)^2 \leq E(x) \leq E(0) = f(0)^2 + f'(0)^2$ for every real x, so

$$|f(x)| \leq \sqrt{f(0)^2 + f'(0)^2} \qquad \text{for all } x \in \mathbb{R},$$

and $|f|$ is bounded.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same displayed line, $E'(x) = -2xg(x)f'(x)^2$, and both correctly read off E non-increasing on $[0, \infty)$. They split immediately after, at the question of what to do for $x < 0$. The natural attempt treats the sign flip of E' there as a dead end for this computation and reaches for the reflection $h(x) = f(-x)$ to manufacture a second, structurally identical, argument. The model chain instead asks whether "E non-decreasing on $(-\infty, 0]$" — which is exactly what the same formula already says — is itself enough, and checks that it is.

> *This one formula for $E'(x)$ already has $-2x$ built into it, so its sign for $x < 0$ is not something I need to go and establish by symmetry — I already have it, on the same line I just wrote down. A non-decreasing function on $(-\infty, 0]$ is bounded above by its value at the right endpoint, 0, exactly as a non-increasing function on $[0, \infty)$ is bounded above by its value at the left endpoint, 0 — these are the same fact, monotone-bounded-by-the-nearer-endpoint, applied on the two sides of a single point, and I have both monotonicities for free from one derivative computation. Introducing $f(-x)$ solves a problem I have already solved.*

Layer 3 THE WHY

Why is the reflection unnecessary here, when it is exactly the move the official solution reaches for? Because the only place the sign of x enters the whole argument is as an explicit, isolated multiplicative factor, $-2x$, sitting in front of two other factors, $g(x)$ and $f'(x)^2$, whose *signs* are fixed (non-negative) independently of the sign of x by hypothesis. That is the structural fact the official solution's own hypothesis, "$g(x) \geq 0$ for all real x" — stated for all x, not just $x \geq 0$ — is built to supply. Once a derivative's sign decomposes as (a fixed-sign function of x alone) $\times$ (a fixed-sign quantity), the derivative's sign on the two sides of 0 is already determined by inspection of the one factor that depends on x; there is no need to reconstruct the argument on the mirror image of the domain, because the domain was never actually asymmetric — only the factor $-2x$ was, and it changes sign exactly where you need it to for a two-sided bound by the same base point.

The reflection trick $h(x) = f(-x)$ is not wrong; it is a general-purpose device for turning "prove something for $x \geq 0$" into "prove it for all x" when you cannot see, or have not checked, that the argument already covers $x < 0$. It earns its keep when the governing inequality's sign genuinely does depend on more than just x's sign in a way that isn't already isolated in the formula. Here it does not: the isolation was already done by the algebra that produced $E'(x) = -2xg(x)f'(x)^2$, and STYLE's own instinct about cheapness applies directly — rerouting to the shorter argument, once you notice the formula already contains the answer for $x < 0$, is the actual skill being tested, not the energy-function idea itself (which both chains find identically).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that $f' \cdot (f + f'')$ is $\frac{1}{2}\frac{d}{dx}(f^2 + (f')^2)$, an exact derivative, and the right-hand side of the given equation, $-xg(x)f'(x)$, becomes $-xg(x)f'(x)^2$ once multiplied by f'. On seeing an ODE of the shape $f'' + f = (\text{something}) \cdot f'$, write down $E = f^2 + (f')^2$ and compute E' immediately.

2. That cue suggests an energy/Lyapunov argument: if $E'(x)$ has one sign on an interval, E is monotone there and bounded by its value at an endpoint of that interval.

3. Hesitation occurs at $x < 0$, right after seeing $E'(x) \geq 0$ there and instinctively reading "increasing" as "unbounded" or "the wrong direction." This is reinforced away by asking, before reaching for any new device, which endpoint of the current interval the monotonicity actually bounds the function by — a non-decreasing function on $(-\infty, 0]$ is bounded by its value at 0, the right endpoint, just as reliably as a non-increasing function on $[0, \infty)$ is bounded by its value at 0, the left endpoint.

4. The trap is treating "the sign of E' flipped" as "the argument has failed," rather than "the argument is still working, just bounding from the other direction." That reflex is exactly what sends a solver looking for machinery (a reflection, a second differential inequality) to redo work a single monotonicity statement, read correctly, already did.

5. Recognise this family faster: whenever a derivative's sign is a product of an x-dependent sign-controlling factor (here $-2x$) and fixed-sign factors, write out explicitly which endpoint each half of the domain is bounded by, before deciding whether a symmetry argument is needed at all.

6. Reusable rule: if a monotone quantity's sign on $(-\infty, a]$ and $[a, \infty)$ both trace back to a single formula whose x-dependence is isolated in one explicit factor, both halves are already bounded by the value at a — check this directly before introducing a reflection or comparison device to re-derive the side you already have.

▷ **Cue:** an ODE of the form $f'' + f = h(x)f'$ (or similarly, a first-order-in-derivative term on the right) $\longrightarrow$ **Move:** multiply by f' and recognise $f'(f + f'') = \frac{1}{2}\frac{d}{dx}(f^2 + (f')^2)$; compute $E' = h(x)f'(x)^2$ (up to the constant factor) and read its sign directly from the sign of $h(x)$

▷ **Cue:** a monotonicity statement $E'(x) \leq 0$ on one half of the domain flips sign on the other half $\longrightarrow$ **Move:** before reaching for a symmetry or reflection argument, check which endpoint the flipped-sign half is bounded by — non-decreasing on $(-\infty, a]$ is bounded by $E(a)$ exactly as non-increasing on $[a, \infty)$ is bounded by $E(a)$

1997 B3

1997 B3. For each positive integer n, write the sum $\sum_{m=1}^{n} 1/m$ in the form p_n/q_n, where p_n and q_n are relatively prime positive integers. Determine all n such that 5 does not divide q_n.

Fields: number theory, p-adic valuation, harmonic numbers **Position:** B3 **Answer:** $n \in \{1, 2, 3, 4\} \cup \{20, 21, 22, 23, 24\} \cup \{100, 101, 102, 103, 104\} \cup \{120, 121, 122, 123, 124\}$ — nineteen values in total.

Layer 1 THE WHAT — surface analysis

The natural attempt. Compute the harmonic numbers directly: $H_1 = 1$, $H_2 = 3/2$, $H_3 = 11/6$, $H_4 = 25/12$ (denominators $1, 2, 6, 12$ — none divisible by 5, so $n = 1, 2, 3, 4$ qualify), then $H_5 = 137/60$, and $60 = 5 \cdot 12$ kills it. The natural next move is to keep computing: $H_6, H_7, \ldots$ by hand or by tracking, at each step, whether adding $1/(n+1)$ to p_n/q_n reintroduces a factor of 5 into the reduced denominator. This is exactly correct as far as it goes, but it gives no way to predict the answer without doing it — and the true answer includes $n = 20$ through 24, and then 100–104 and 120–124, values far beyond where hand computation of $137/60, \ldots$ can reach.

Worse, the naive expectation built from the first failure at $n = 5$ — "avoid n near a multiple of 5" — is actively misleading: every n from 6 to 19 fails (fourteen consecutive values), and then $20, 21, 22, 23, 24$ — five consecutive values, all multiples-of-5-adjacent or exactly a multiple of 5 — all succeed. A rule built from watching $1/(n+1)$ get added term-by-term has no mechanism for predicting that a whole run of five values straddling 20 (itself a multiple of 5) would come back into the "safe" set after fourteen unsafe ones; the divisibility behaviour of q_n is not monotone or locally determined by the last term added, it is a global property of the sum up to n, and no amount of watching individual additions of $1/(n+1)$ reveals why the safe windows recur near powers of 5 and near numbers like $20, 24, 100, 104, 120, 124$ specifically.

The model chain. Split H_n by divisibility by 5. Let $I_n = \sum_{1 \le m \le n,\, 5 \nmid m} 1/m$ (the part with denominators coprime to 5). Every term of H_n is either counted in I_n or is $1/(5j)$ for some $j \le n/5$, so

$$H_n = I_n + \frac{1}{5} \sum_{j=1}^{\lfloor n/5 \rfloor} \frac{1}{j} = I_n + \frac{1}{5} H_{\lfloor n/5 \rfloor}.$$

(Verified independently: this is exactly the statement "split off the multiples of 5*, and the leftover is* $\frac{1}{5}$ *of a smaller harmonic sum," and it is immediate from partitioning* $\{1, \ldots, n\}$ *into multiples of* 5 *and non-multiples.)* Applying it again to $H_{\lfloor n/5 \rfloor}$ and iterating gives the unrolled identity

$$H_n = \sum_{i=0}^{k} \frac{1}{5^i} I_{\lfloor n/5^i \rfloor}, \qquad k = k(n) = \text{largest integer with } 5^k \le n.$$

Since every term of I_r has a denominator coprime to 5, I_r itself is a 5-adic integer (its denominator is invertible mod any power of 5), and its value mod 5 is periodic in r with period 5: direct computation ($1/1 \equiv 1, 1/2 \equiv 3, 1/3 \equiv 2, 1/4 \equiv 4 \pmod 5$, since $2 \cdot 3 \equiv 1, 3 \cdot 2 \equiv 1, 4 \cdot 4 \equiv 1 \pmod 5$) gives

$$I_r \equiv 1,\ -1,\ 1,\ 0,\ 0 \pmod 5 \quad \text{for } r \equiv 1, 2, 3, 4, 0 \pmod 5 \text{ respectively}$$

(both congruences checked directly above: $I_1 \equiv 1$, $I_2 \equiv 1 + 3 = 4 \equiv -1$, $I_3 \equiv -1 + 2 = 1$, $I_4 \equiv 1 + 4 = 5 \equiv 0 \pmod 5$*, and* $I_5 = I_4$ *since the term* $1/5$ *is excluded, so* $I_5 \equiv 0$ *too, matching* $r \equiv 0$*).*

The 5-adic valuation of H_n — whether $5 \mid q_n$ or not is exactly whether this valuation is ≥ 0 — is governed by the *last* term of the unrolled sum whose $I_{\lfloor n/5^i \rfloor}$ is not divisible by 5 (i.e. $\lfloor n/5^i \rfloor \equiv 1, 2, 3 \pmod 5$), since that term, divided by 5^i, then carries valuation exactly $-i$ unless it is cancelled by another term of the same valuation. Values $n = 1, 2, 3$ ($k = 0$, no division needed) are always safe; $n = 4$ is the boundary case where $I_4 \equiv 0$ but the division stops at $k = 0$ so no negative valuation is created either. Once $n \ge 5$, some prefix $\lfloor n/5^i \rfloor$ generically lands on residue $1, 2, 3 \pmod 5$ and forces q_n to acquire a factor of 5 — *unless* every relevant $\lfloor n/5^i \rfloor$ lands on residue 4 or 0 (the two residues with $I_r \equiv 0$), which is precisely the mechanism that reopens the safe windows at $n = 20$–24 (where $\lfloor n/5 \rfloor = 4$), and recursively at 100–104 and 120–124 (where the second level also lands on residue 4, i.e. $\lfloor n/25 \rfloor = 4$, giving $100 \le n \le 124$ before intersecting with the first-level condition again).

Caveat. The mechanism above — the recursive decomposition $H_n = I_n + \frac{1}{5}H_{\lfloor n/5 \rfloor}$ and the period-5 residues of $I_r \pmod 5$ — I have verified independently by direct computation for small r and by re-deriving the decomposition from first principles. The finer claim the official solution needs to pin the exact boundaries of the safe windows at 100–104 versus 120–124 (in particular, that $I_{20} \equiv I_{24} \equiv 0 \pmod{25}$, a mod-25 statement about sums of roughly sixteen unit fractions' 5-adic inverses) is a genuine second-order computation that the official solution carries out and I have not independently re-verified digit by digit; I am reporting it as sourced from Bhargava–Kedlaya–Ng rather than re-derived. The architecture — recurse on the prime-p decomposition of a harmonic sum, track residues mod increasing powers of p — is the part I have confirmed and is the reusable content.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains correctly identify $n = 1, 2, 3, 4$ as safe and $n = 5$ as the first failure. They split at the point of deciding how to search for the *next* safe value. Quoting the natural attempt's implicit strategy: keep adding $1/(n+1)$ and watch the denominator. The model chain instead asks: "what controls whether $5 \mid q_n$ at all, structurally, rather than term by term?" and splits H_n by divisibility by 5 before doing any further arithmetic.

> *Watching q_n term by term as $1/(n+1)$ is added tells me only about local, one-step changes, but the property "$5 \nmid q_n$" is a statement about the entire sum $1 + \frac{1}{2} + \cdots + \frac{1}{n}$ at once. If I instead split every term of H_n into "denominator coprime to 5" and "denominator a multiple of 5," the second group is literally $\frac{1}{5}$ of a smaller harmonic sum over $\lfloor n/5 \rfloor$ terms — so I have written H_n in terms of a strictly smaller instance of exactly the same object, which is a recursion I can unroll, rather than a term-by-term walk I can only extend one step at a time.*

Layer 3 THE WHY

Why does splitting off the multiples of 5 turn an unpredictable term-by-term walk into a structure with recognisable periodic windows? Because $5 \mid q_n$ is fundamentally a statement about the 5-adic valuation of H_n (a rational number is 5-integral, i.e. has $5 \nmid$ denominator, exactly when its 5-adic valuation is ≥ 0), and valuation is not something that changes tamely under "add one more term" — adding $1/(n+1)$ can shift the valuation by any amount depending on cancellation with the existing sum. What valuation *does* respect is exactly this kind of prime-indexed recursive decomposition: separating the terms whose denominators already carry a factor of the prime in question from those that do not is the standard move for computing p-adic valuations of sums (the same idea underlies, for instance, Legendre's formula for the valuation of $n!$, which splits off multiples of p recursively in an entirely analogous way). Once H_n is written as $\sum_i I_{\lfloor n/5^i \rfloor}/5^i$, each summand's own valuation is transparent (it is $-i$ plus the valuation of $I_{\lfloor n/5^i \rfloor}$, and the latter is 5-integral by construction, so its valuation is ≥ 0, with equality controlled by its residue mod 5), and the total valuation is governed by whichever summand has the most negative individual valuation that is not exactly cancelled — which is precisely why the answer comes in *windows*: a window opens whenever every level of the base-5-style expansion of n happens to land on the residues (0 or 4 mod 5) for which I_r vanishes mod 5, and it closes again the moment any level lands on $1, 2, 3$.

The naive term-by-term view cannot see this because it has no notion of "level" at all — it only ever looks at the innermost term, whereas the actual obstruction lives in how n's own digits, read in base 5, interact with the fixed residue pattern of I_r at every level of the recursion simultaneously.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the question "does 5 divide q_n" is a statement about 5-adic valuation, not about the numerator or denominator as computed integers. On seeing a divisibility-of-

denominator question about a sum of unit fractions, immediately reframe it as a valuation question.

2. That cue suggests the standard move for computing the p-adic valuation of a sum indexed by $1, \ldots, n$: split off the terms already divisible by p (here, denominators that are multiples of 5) and recurse on the remainder, exactly as Legendre's formula splits off multiples of p to compute the valuation of $n!$.

3. Hesitation occurs right after the first failure at $n = 5$: the instinct is to keep computing $H_6, H_7, \ldots$ by hand, hoping a pattern in the raw denominators will announce itself. This is reinforced away by refusing to trust any pattern read off consecutive raw denominators, and instead writing the exact recursive decomposition $H_n = I_n + \frac{1}{5}H_{\lfloor n/5 \rfloor}$ before searching further.

4. The trap is expecting the safe set to be "small n, plus n far from multiples of 5" — a locally-plausible guess from the first failure at $n = 5$ that is falsified by the fourteen consecutive failures from $n = 6$ to 19 and the return of a safe window exactly straddling the multiple of 5 at $n = 20$–24. Valuation-based windows do not track "distance from a multiple of 5" in any simple way; they track residues at every level of a base-5-style recursion at once.

5. Recognise this family faster: whenever a problem asks "for which n does p fail to divide the denominator of $\sum_{m=1}^{n} 1/m$ (or a similarly indexed sum)," write the recursive split by divisibility by p immediately, and compute the residues of the coprime-to-p partial sum mod p for a full period before attempting to characterise the answer set.

6. Reusable rule: a p-adic-valuation question about a sum indexed by $1, \ldots, n$ is answered by recursing on the prime-p digits of n (split off multiples of p, recurse on the quotient), not by tracking the raw sum term by term. The answer set typically comes in windows determined by which residues of the recursion's "coprime part" vanish mod p, and those windows can recur, unpredictably from a term-by-term view, near powers of p.

▷ **Cue:** a question about whether a prime p divides the denominator of $\sum_{m=1}^{n} 1/m$ (or an analogous indexed sum) in lowest terms ⟶ **Move:** reframe as a p-adic valuation question, then split the sum into terms with denominator divisible by p and terms coprime to p; the divisible part is $\frac{1}{p}$ times a smaller instance of the same sum, giving a recursion to unroll

▷ **Cue:** a pattern guessed from the first few failures of a divisibility condition on a growing sum ⟶ **Move:** do not extrapolate "avoid values near the first failure" without checking a full period of the underlying residue — valuation-type conditions can reopen in windows far from where they first failed

1997 B4

1997 B4. Let $a_{m,n}$ denote the coefficient of x^n in the expansion of $(1 + x + x^2)^m$. Prove that for all [integers] $k \geq 0$,

$$0 \leq \sum_{i=0}^{\lfloor \frac{2k}{3} \rfloor} (-1)^i a_{k-i,i} \leq 1.$$

Fields: combinatorics; generating functions; polynomial coefficients **Position:** B4 — mid-paper; a two-line proof once the sum is recognised as a diagonal of a double generating function, and a bookkeeping swamp otherwise **Answer:** Proof: the sum s_k is the coefficient of y^k in $1/((1-y)(1+y^2))$, so $s_k = 1, 1, 0, 0, 1, 1, 0, 0, \ldots$ with period 4; every value is 0 or 1.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student computes. Write s_k for the sum. The rows of $(1 + x + x^2)^m$ are 1; 1, 1, 1; 1, 2, 3, 2, 1; 1, 3, 6, 7, 6, 3, 1; 1, 4, 10, 16, 19, 16, 10, 4, 1. Then

$$s_0 = 1, \quad s_1 = a_{1,0} = 1, \quad s_2 = a_{2,0} - a_{1,1} = 0, \quad s_3 = 1 - 2 + 1 = 0,$$

$$s_4 = a_{4,0} - a_{3,1} + a_{2,2} = 1 - 3 + 3 = 1, \qquad s_5 = 1 - 4 + 6 - 2 = 1.$$

The pattern $1, 1, 0, 0, 1, 1$ appears and the student conjectures period 4. Now a proof is needed, and the natural instinct is to use the Pascal-type recurrence $a_{m+1,n} = a_{m,n} + a_{m,n-1} + a_{m,n-2}$ (multiply $(1 + x + x^2)^m$ by one more factor) and push it through the alternating sum. This is where the time goes. The sum runs along an *anti-diagonal* $m + i = k$, while the recurrence steps m by one and n by up to two, so substituting it into s_k produces three sums along three different lines, and one has to see that they recombine as $s_k = s_{k-1} - s_{k-2} + s_{k-3}$ for $k \geq 3$. It can be done — the official solution does it — but the index shifts and the upper limits $\lfloor 2k/3 \rfloor$ make every line an opportunity to be off by one, and at the end you still need initial values and a four-term induction.

The model chain. **Step 1: remove the awkward upper limit.** The coefficient $a_{m,i}$ is zero for $i > 2m$, since $(1 + x + x^2)^m$ has degree $2m$. With $m = k - i$, the term $a_{k-i,i}$ is nonzero only when $i \leq 2(k - i)$, that is $3i \leq 2k$, that is $i \leq \lfloor 2k/3 \rfloor$. So

$$s_k = \sum_{i=0}^{\lfloor 2k/3 \rfloor} (-1)^i a_{k-i,i} = \sum_{\substack{m,i \geq 0 \\ m+i=k}} (-1)^i a_{m,i}.$$

The floor was only ever the natural end of the support.

Step 2: package all s_k at once. Work with formal power series in y. Since $\sum_i a_{m,i} x^i = (1 + x + x^2)^m$,

$$\sum_{k \geq 0} s_k y^k = \sum_{m \geq 0} \sum_{i \geq 0} a_{m,i} (-1)^i y^{m+i} = \sum_{m \geq 0} y^m \sum_i a_{m,i} (-y)^i = \sum_{m \geq 0} y^m (1 - y + y^2)^m.$$

Each coefficient of y^k on the right is a finite sum (only $m \leq k$ contribute), so the rearrangement is legitimate for formal series. The right side is a geometric series in $u = y(1 - y + y^2)$, which has zero constant term, so

$$\sum_{k \geq 0} s_k y^k = \frac{1}{1 - y + y^2 - y^3} = \frac{1}{(1 - y)(1 + y^2)}.$$

Step 3: read off the coefficients.

$$\frac{1}{(1 - y)(1 + y^2)} = \Big(\sum_{j \geq 0} y^j \Big) \Big(\sum_{l \geq 0} (-1)^l y^{2l} \Big), \qquad \text{so} \qquad s_k = \sum_{l=0}^{\lfloor k/2 \rfloor} (-1)^l.$$

A partial sum of $1 - 1 + 1 - \cdots$ is 1 if it has an odd number of terms and 0 if even. Here there are $\lfloor k/2 \rfloor + 1$ terms, so $s_k = 1$ when $\lfloor k/2 \rfloor$ is even ($k \equiv 0, 1 \pmod 4$) and $s_k = 0$ otherwise. In every case $0 \leq s_k \leq 1$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains compute the same small cases and agree on the period-4 conjecture. They split at the choice of proof tool. (N) reaches for the recurrence $a_{m+1,n} = a_{m,n} + a_{m,n-1} + a_{m,n-2}$ and tries to transport it along anti-diagonals. (M) notices that the sum is over $m + i = k$ with a weight depending only on i, which is exactly the shape of a coefficient in a *product-in-disguise*: attach y^m to the row index and $(-y)^i$ to the column index, and the anti-diagonal sum becomes "the coefficient of y^k". After that, the row generating function $(1 + x + x^2)^m$ is given in the problem, so the whole double sum collapses to a geometric series.

> *The sum runs over $m + i = k$. A sum over a fixed total is a coefficient of a product. If I give the row m the weight y^m and the column i the weight $(-y)^i$, then s_k is the coefficient of y^k in something I can write down.*

> *I already know the generating function of each row: it is $(1 + x + x^2)^m$. So putting $x = -y$ and summing over m with y^m gives a geometric series. I never need the individual $a_{m,n}$.*

> *The upper limit $\lfloor 2k/3 \rfloor$ looks like a constraint, but it is where $a_{k-i,i}$ becomes zero. I can drop it and sum over all i.*

- ▷ **Cue:** a sum $\sum_i c_i\, a_{k-i,i}$ over an anti-diagonal of a table whose rows have a known generating function $\longrightarrow$ **Move:** form $\sum_k s_k y^k = \sum_m y^m R_m(\text{weight} \cdot y)$ where R_m is the row generating function
- ▷ **Cue:** an upper summation limit such as $\lfloor 2k/3 \rfloor$ that looks arbitrary $\longrightarrow$ **Move:** check whether it is the support boundary of the summand; if so, sum over everything
- ▷ **Cue:** an integer sequence claimed to take only a few values, e.g. $0 \le s_k \le 1$ $\longrightarrow$ **Move:** compute ten terms, spot the period, then find the rational generating function whose denominator is a product of cyclotomic factors
- ▷ **Cue:** a rational function $1/((1-y)(1+y^2))$ $\longrightarrow$ **Move:** expand as a product of two geometric series; the coefficient is a partial sum of a periodic sequence

Layer 3 THE WHY

Why is the generating function the natural move rather than a trick? Go to what $a_{m,n}$ *is*. The problem defines it as a coefficient: "the coefficient of x^n in the expansion of $(1 + x + x^2)^m$". The table is presented to you already packaged, row by row, as polynomials. When an object is defined as a coefficient, the cheapest way to handle sums of it is to stay inside the algebra of polynomials that produced it, not to unpack it into numbers and repack. The recurrence approach of (N) unpacks: it treats the $a_{m,n}$ as a table of integers and tries to discover how the anti-diagonal sums interact. The generating-function approach never leaves the polynomial world.

Why the particular weights y^m and $(-y)^i$? The sum has two features: the indices satisfy $m + i = k$, and the sign depends on i. "Indices adding to k" is the defining property of multiplication of power series: the coefficient of y^k in $\sum_m \sum_i b_{m,i} y^m y^i$ is exactly $\sum_{m+i=k} b_{m,i}$. So the exponent of y must be $m + i$, which forces weight y^m on the row and y^i on the column. The sign $(-1)^i$ rides along with the column, giving $(-y)^i$. That is not a choice; it is the unique way to make "coefficient of y^k" equal s_k. Once you have written it, the inner sum is the row polynomial evaluated at $x = -y$, by definition.

Now the official solution. Bhargava, Kedlaya and Ng take the recurrence route: they derive a four-term linear recurrence for s_k from the three-term Pascal rule and finish by induction from s_0, s_1, s_2. (As printed, their recurrence reads $s_{k+3} = s_k - s_{k-1} + s_{k+2}$ and their definition of s_k uses $a_{k-1,i}$, with a stray n in the intermediate line; taken literally with $s_{4j} = s_{4j+1} = 1$, $s_{4j+2} = s_{4j+3} = 0$, it fails at $k = 1$, since $s_4 = 1$ but $s_1 - s_0 + s_3 = 0$. The correct recurrence, with s_k as defined in the problem, is $s_{k+3} = s_{k+2} - s_{k+1} + s_k$, and I checked it against the first forty values by computer. The typo is harmless to the idea but would cost marks if copied.) Where does that recurrence come from? It is the denominator: $1 - y + y^2 - y^3$ means precisely $s_k - s_{k-1} + s_{k-2} - s_{k-3} = 0$ for $k \ge 3$, which is the corrected linear recurrence. The official route discovers the denominator of the generating function one coefficient at a time. The parenthetical alternative they credit to John Rickert — s_k is the coefficient of y^k in $(1 + y)/(1 - y^4)$ — is Step 2 above, with the fraction multiplied top and bottom by $1 + y$ so that the period 4 is visible in the denominator.

Why is the answer periodic at all? Because the denominator factors as $(1-y)(1+y^2)$, and every root $(1, i, -i)$ is a root of unity. A rational generating function whose denominator has only roots of unity as zeros, all simple, has a periodic coefficient sequence (here period 4, because $1-y^4=(1-y)(1+y)(1+y^2)$ is a multiple of the denominator). So the bounded, two-valued behaviour that the problem asks you to prove is forced by where the roots of $1-y+y^2-y^3$ sit. The problem was set to reward reaching the denominator.

Finally, why was (N) not wrong, only expensive? It proved the same recurrence by hand. Any linear recurrence you find by manipulating sums is the denominator of a generating function you did not write down. Writing it down first is the cheaper route — and it is the one that also tells you the answer, rather than needing the answer to be guessed first. That is the zombie-maths lesson here.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the index pattern $a_{k-i,i}$: the two subscripts add to k. On seeing it, write $\sum_k s_k y^k = \sum_{m,i}(-1)^i a_{m,i} y^{m+i}$.

2. A bivariate-to-univariate generating function: sum over rows using the given row polynomial $(1+x+x^2)^m$ at $x=-y$, then the geometric series.

3. Hesitation comes at the upper limit $\lfloor 2k/3 \rfloor$, which makes the sum look non-standard. It is reinforced away by asking when $a_{k-i,i}$ vanishes: degree $2(k-i) < i$ exactly when $i > 2k/3$.

4. The trap is the recurrence $a_{m+1,n} = a_{m,n} + a_{m,n-1} + a_{m,n-2}$: it is true and it does lead to a proof, but pushing it along anti-diagonals is error-prone, and it only confirms a pattern you had to guess. A second trap is stopping after the numerical pattern $1, 1, 0, 0$ and calling it proved.

5. Recognise faster by reading the definition: whenever a table entry is defined as "the coefficient of x^n in $P(x)^m$", sums of entries along any line $\alpha m + \beta n = k$ are coefficients of $\sum_m (y^\alpha P(y^\beta))^m$.

6. Reusable decision rule: *when a sum runs over indices with a fixed total, turn it into a coefficient of a product by weighting each index with a power of the same variable; if the pieces have known generating functions, the sum becomes a rational function and its denominator gives the recurrence and the long-run behaviour.*

1997 B5

1997 B5. Prove that for $n \geq 2$,

$$\underbrace{2^{2^{\cdot^{\cdot^{\cdot^2}}}}}_{n \text{ terms}} \equiv \underbrace{2^{2^{\cdot^{\cdot^{\cdot^2}}}}}_{n-1 \text{ terms}} \pmod{n}.$$

Fields: number theory; Euler's theorem; strong induction; Chinese remainder theorem **Position:** B5 — late paper; the proof is short, but only after the statement is strengthened into one that can carry its own induction **Answer:** Proof: with $t_1 = 2$, $t_{k+1} = 2^{t_k}$, prove the stronger claim that for every $m \geq 1$, $t_k \equiv t_{k+1} \pmod{m}$ for all $k \geq m-1$, by strong induction on m, splitting $m = 2^a b$ and using Euler's theorem modulo b with the hypothesis at $\varphi(b) < m$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Name the towers: $t_1 = 2$ and $t_{k+1} = 2^{t_k}$, so the claim is $t_n \equiv t_{n-1} \pmod{n}$. The strong student checks small cases: $n=2$ ($4 \equiv 2$, both even), $n=3$ ($16 \equiv 4 \equiv 1$), $n=4$ ($2^{16} \equiv 16 \equiv 0$), $n=5$ ($t_5 = 2^{65536}$; $2^4 \equiv 1$ and $65536 \equiv 0 \pmod 4$, so $t_5 \equiv 1$, and $t_4 = 65536 \equiv 1$). The $n=5$ check contains the whole mechanism: to reduce 2^{huge} mod 5 you reduce the exponent mod 4. So the student tries induction on n: "assume $t_{n-1} \equiv t_{n-2} \pmod{n-1}$ and prove $t_n \equiv t_{n-1} \pmod n$". The wall appears at once. To compare $t_n = 2^{t_{n-1}}$ with $t_{n-1} = 2^{t_{n-2}}$ modulo n, you need the exponents t_{n-1} and t_{n-2} to agree modulo the order of 2 mod n — something

like $\varphi(n)$, not $n-1$. The hypothesis about $n-1$ is about the wrong modulus. And for even n, 2 is not invertible mod n, so Euler's theorem does not apply at all. The induction as stated does not close.

The model chain. Let $t_1 = 2$, $t_{k+1} = 2^{t_k}$. Two facts about sizes: t_k is even for all k, and $t_k \geq k+1$ (true for $k=1$; and $t_{k+1} = 2^{t_k} \geq 2^{k+1} \geq k+2$).

Claim. For every integer $m \geq 1$: $t_k \equiv t_{k+1} \pmod{m}$ for every $k \geq \max(1, m-1)$.

The problem is the case $m = n$, $k = n-1 \geq 1$.

Proof by strong induction on m. For $m = 1$ there is nothing to prove; for $m = 2$ all t_k are even. Let $m \geq 3$ and assume the claim for all smaller moduli. Fix $k \geq m-1 \geq 2$. Write $m = 2^a b$ with b odd. By the Chinese remainder theorem it suffices to prove $t_k \equiv t_{k+1}$ modulo 2^a and modulo b.

Modulo 2^a. Since $k \geq 2$, $t_k = 2^{t_{k-1}}$ and $t_{k+1} = 2^{t_k}$. Now $a < 2^a \leq m$, so $a \leq m-1$, while $t_{k-1} \geq k \geq m-1 \geq a$ and $t_k \geq t_{k-1}$. So 2^a divides both t_k and t_{k+1}; both are $\equiv 0 \pmod{2^a}$.

Modulo b. If $b = 1$ there is nothing to do. Otherwise $b \geq 3$ is odd, $\gcd(2, b) = 1$, and Euler's theorem gives $2^{\varphi(b)} \equiv 1 \pmod{b}$. Hence for non-negative integers $u \equiv v \pmod{\varphi(b)}$ we have $2^u \equiv 2^v \pmod{b}$. Now $\varphi(b) \leq b-1 \leq m-1 < m$, so the induction hypothesis applies to the modulus $\varphi(b)$: $t_j \equiv t_{j+1} \pmod{\varphi(b)}$ for all $j \geq \max(1, \varphi(b)-1)$. Take $j = k-1$; this is allowed because $k-1 \geq m-2 \geq b-2 \geq \varphi(b)-1$ and $k-1 \geq 1$. So $t_{k-1} \equiv t_k \pmod{\varphi(b)}$, and therefore

$$t_k = 2^{t_{k-1}} \equiv 2^{t_k} = t_{k+1} \pmod{b}.$$

Combining the two moduli, $t_k \equiv t_{k+1} \pmod{m}$. This completes the induction, and with $m = n$, $k = n-1$ it gives the problem. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains see that reducing 2^{tower} modulo n means reducing the upper tower modulo $\varphi(n)$ (or the order of 2). They split at what they choose to induct on. (N) inducts on n with the statement exactly as given, so the hypothesis is about the modulus $n-1$, which is useless. (M) inducts on the *modulus*, with a statement that is uniform in the height: "modulo m, the tower sequence has stopped changing by index $m-1$". That statement is strong enough to be fed to a smaller modulus $\varphi(b)$ with a different index, which is exactly what the Euler step requires.

> *To compute 2^x mod n I need x mod $\varphi(n)$, not mod $n-1$. So my induction hypothesis has to talk about* every *smaller modulus, not the previous one. I'll induct on the modulus.*

> *The hypothesis at $\varphi(b)$ must hold at an index I choose, not a fixed one, so I prove more: the sequence t_k mod m is constant from some point on, and that point is no later than $m-1$.*

> *Two is not invertible mod an even n, so split $n = 2^a b$: modulo 2^a the towers are both 0 because they are enormous powers of 2; modulo b Euler applies.*

- ▷ **Cue:** a power tower, or any iterated exponential, reduced modulo n $\longrightarrow$ **Move:** reduce the top modulo $\varphi(n)$ (or the order of the base), and set up induction on the modulus
- ▷ **Cue:** an induction on n where the step needs the hypothesis at an unrelated smaller value such as $\varphi(n)$ $\longrightarrow$ **Move:** use strong induction and strengthen the statement so it holds for every smaller value and every sufficiently large index
- ▷ **Cue:** a congruence $x \equiv y \pmod{n}$ with a base not coprime to n $\longrightarrow$ **Move:** split $n = 2^a b$ by the Chinese remainder theorem; treat the prime-power part by size (divisibility), the coprime part by Euler

▷ **Cue:** a problem statement that ties the index to the modulus, like "n terms mod n" $\longrightarrow$ **Move:** treat it as a statement about when a sequence mod m stabilises, and prove the stabilisation bound

Layer 3 THE WHY

Why must the induction be on the modulus, with a strengthened statement? Go to what exponentiation modulo n depends on. For $\gcd(2, b) = 1$, the residue of 2^u mod b depends only on u modulo the order of 2, which divides $\varphi(b)$. That is the definition of order together with Euler's theorem, nothing more. So the question "is $2^u \equiv 2^v \pmod{b}$?" is, by definition, a question about u and v modulo a *different, smaller* number. In a tower, u and v are themselves towers. Therefore any proof about towers modulo b must invoke a fact about towers modulo $\varphi(b)$, and then modulo $\varphi(\varphi(b))$, and so on down. The induction is on the modulus because the mathematics itself descends through moduli. That is the "of course": the structure of 2^u mod b tells you which variable to induct on.

Why strengthen? The problem's statement pins the modulus to the height: "n terms versus $n-1$ terms, mod n". When you descend from b to $\varphi(b)$, the height does not descend with it: you still need information about t_{k-1} and t_k, but now modulo $\varphi(b)$, a modulus much smaller than k. The original statement at modulus $\varphi(b)$ would only tell you about $t_{\varphi(b)-1}$ and $t_{\varphi(b)}$, which are the wrong terms. So you need "from index $\varphi(b) - 1$ onwards, all terms agree". The strengthened claim is the smallest statement that survives the descent. A statement that does not survive its own inductive step must be strengthened until it does; this is the standard behaviour of induction, and here the problem makes it unavoidable.

Why the bound $m - 1$ on the index? Tally the losses. Each Euler step costs one level of the tower (going from t_{k+1} to t_k) and at least one unit of modulus ($\varphi(b) \leq b - 1$). The 2-power part needs the tower to be at least $a \leq \log_2 m$, which is small. So the index at which the sequence mod m freezes grows at most as fast as the modulus shrinks, and "index $m - 1$" has room to spare. The problem's "n terms, mod n" is a comfortably true stabilisation bound, not a tight one (computation shows $t_{m-2} \equiv t_{m-1} \pmod{m}$ for every m from 4 to 39; only $m = 3$ needs the full index).

Now the official solution. Bhargava, Kedlaya and Ng write: "It suffices to show that for every n, $x_m \equiv x_{m+1} \equiv \cdots \pmod{n}$ for some $m < n$." Here is where that comes from: it is precisely the strengthening forced by the descent. "For some $m < n$" is the stabilisation formulation; the claim above makes the index explicit, which is what makes the induction step checkable line by line. Their mod-2^a step uses "$x_{n-1} \geq n$"; the precise requirement is that the exponent t_{k-1} is at least a, which the size bound $t_j \geq j + 1$ supplies. Their mod-b step invokes the hypothesis at $\varphi(b)$ and needs $\varphi(b) + 1 \leq n$, which is where oddness of b (so $b \geq 3$ and $\varphi(b) \leq b - 1$) enters. They also thank a reader for catching a "lethal typo" — a warning that the index bookkeeping here is exactly where proofs of this problem die.

What does zombie maths look like here? Computing t_5, t_6 modulo 5, 6 by hand and hoping to see a pattern in n. The pattern that matters is not in n; it is in the chain $b \to \varphi(b) \to \varphi(\varphi(b)) \to \cdots \to 1$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a tower of exponents compared modulo n. On seeing it, write: 2^u mod b depends only on u mod $\varphi(b)$ when b is odd; so the proof descends from b to $\varphi(b)$.

2. Strong induction on the modulus, with the claim strengthened to "t_k mod m is constant for $k \geq m - 1$"; split $m = 2^a b$ with the Chinese remainder theorem; Euler's theorem on the odd part, divisibility on the 2-part.

3. Hesitation comes when the induction hypothesis at $n - 1$ turns out to be useless. It is reinforced away by asking: *which smaller modulus does my step actually need?* The answer, $\varphi(b)$, tells you to use strong induction over all moduli.

4. The trap is inducting on n with the problem's exact statement. A second trap is applying

Euler's theorem modulo an even n, where 2 is not a unit. A third is getting the index conditions wrong in the step (check $k-1 \geq 1$ and $k-1 \geq \varphi(b)-1$ explicitly).

5. Recognise faster: any iterated exponential modulo n is a descent along $n \to \varphi(n) \to \cdots$; set up the induction on the modulus before computing anything.

6. Reusable decision rule: *when the inductive step needs the hypothesis at a value unrelated to $n-1$, switch to strong induction and strengthen the statement until it is uniform in every parameter the step changes; here, uniform in the index once it passes a threshold.*

1997 B6

1997 B6. The dissection of the 3–4–5 triangle shown below (into four congruent right triangles similar to the original) has diameter $5/2$. Find the least diameter of a dissection of this triangle into four parts. (The diameter of a dissection is the least upper bound of the distances between pairs of points belonging to the same part.)

Fields: geometry; extremal combinatorics; pigeonhole principle; coordinate geometry **Position:** B6 — last problem; the answer $25/13$ is not guessable, and both a lower bound and a matching construction are needed **Answer:** $25/13$. Lower bound: five points of the triangle pairwise at distance $\geq 25/13$, so two share a part. Upper bound: an explicit dissection into four convex pieces of diameter exactly $25/13$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student starts from the given dissection (the medial-triangle cut, diameter $5/2$) and tries to improve it by moving cuts. Place $C=(0,0)$, $A=(0,3)$, $B=(4,0)$. The obvious pieces are "corners": a piece at A, a piece at B, a piece at C, and a middle piece. Try cutting a disc sector of radius r at each vertex and see what is left. It is an honest search, and it gets somewhere — one can reach diameters near 2 by trial — but it has no stopping rule. Every configuration gives an upper bound; nothing tells you that you have reached the minimum. The student also tries a lower bound by area (four parts, total area 6, so some part has area $\geq 3/2$; a set of diameter δ has area at most $\pi\delta^2/4$ by the isodiametric inequality, so $\delta \geq 2\sqrt{(3/2)/\pi} \approx 1.38$), which is far too weak. The two bounds never meet, and the answer $25/13 \approx 1.923$ is not a number anyone guesses.

The model chain. **Step 1: the lower bound comes from five points.** If a set S of five points of the triangle has all pairwise distances $\geq d$, then in any dissection into four parts two of the points lie in a common part (pigeonhole: every point lies in some part), so the dissection has diameter $\geq d$. The vertices A, B, C are far apart, so choose two more points $D \in AB$ and $E \in CB$ with

$$AD = DE = EB = d.$$

Then $D = A + \frac{d}{5}(4,-3) = \left(\frac{4d}{5}, 3-\frac{3d}{5}\right)$ and $E=(4-d,0)$, and $DE^2 = d^2$ becomes

$$\left(4-\tfrac{9d}{5}\right)^2 + \left(3-\tfrac{3d}{5}\right)^2 = d^2 \iff 13d^2 - 90d + 125 = 0 \iff d = 5 \text{ or } d = \tfrac{25}{13}.$$

Take $d = 25/13$: $D=(20/13, 24/13)$, $E=(27/13, 0)$. In units of $1/13$, the ten distances are

$$AB = 65,\ AC = 39,\ BC = 52,\ AD = DE = BE = 25,\ BD = 40,\ CE = 27,\ CD = \sqrt{976} \approx 31.2,\ AE = \sqrt{2250} \approx$$

All are ≥ 25, so every dissection into four parts has diameter $\geq 25/13$.

Step 2: the matching dissection. Let

$$F = \left(\tfrac{15}{13}, \tfrac{19}{13}\right),\ G = \left(\tfrac{15}{13}, 0\right),\ H = \left(0, \tfrac{19}{13}\right),\ J = \left(\tfrac{32}{13}, \tfrac{15}{13}\right),$$

so J lies on AB ($\frac{32}{52} + \frac{15}{39} = 1$) with $BJ = 25/13$. Cut the triangle into the four convex polygons $ADFH$, BEJ, $CGFH$, $DFGEJ$. They tile the triangle: the boundary of ABC is shared out as CH, HA on CA; AD, DJ, JB on AB; CG, GE, EB on CB; the interior edges HF, FD, FG, EJ are each shared by exactly two pieces; and the areas (in units of $1/169$) are $237.5 + 187.5 + 285 + 304 = 1014 = 6 \cdot 169$.

The diameter of a convex polygon is its largest vertex-to-vertex distance (the squared distance is convex in each argument, so its maximum over the polygon is attained at vertices). In units of $1/13$:

- $ADFH$: $AD = AF = 25$, $AH = 20$, $DH = \sqrt{425}$, $FH = 15$, $DF = \sqrt{50}$. Max 25.
- BEJ: $BE = BJ = 25$, $EJ = \sqrt{250}$. Max 25.
- $CGFH$: a 15×19 rectangle, diagonal $\sqrt{586} < 25$.
- $DFGEJ$: $DE = 25$, $DG = \sqrt{601}$, $GJ = \sqrt{514}$, $EF = \sqrt{505}$, $FG = 19$, $FJ = \sqrt{305}$, $DJ = 15$, $GE = 12$, $EJ = \sqrt{250}$, $DF = \sqrt{50}$. Max 25.

So this dissection has diameter $25/13$, and the least diameter is $25/13$. □

Caveat. The official solution prints $J = (32/15, 15/13)$. That point is not on AB ($\frac{32/15}{4} + \frac{15/13}{3} = \frac{179}{195} \neq 1$), and with it the four pieces would not cover the triangle. The intended point is $J = (32/13, 15/13)$, which is on AB at distance $25/13$ from B. With this correction every claim above — the ten distances, the tiling, convexity of each piece, and all the vertex distances — was checked in exact rational arithmetic by computer, and the tiling was also confirmed by sampling 10^5 random points of the triangle, each of which lay in exactly one piece.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start by drawing pieces around the corners. They split at the question asked of those corners. (N) asks "how small can I make each corner piece?" and searches the space of dissections, which only ever produces upper bounds. (M) asks "what forces a large piece?" and answers with pigeonhole: four parts, so find *five* points that are pairwise far apart. That reframes the problem as a finite optimisation — spread five points in the triangle to maximise the minimum distance — and the optimal configuration then *tells you* the dissection: each far-apart point wants its own piece, and the pieces at A and B are sectors of radius d.

> *Four parts. If I find five points pairwise at distance at least d, two of them share a part and the diameter is at least d. So the lower bound is a packing problem: five points in the triangle, maximise the smallest distance.*

> *The vertices are the three points furthest apart; I take all three. The other two go on the long sides, and at the optimum the tight distances should be equal: $AD = DE = EB$. That is one equation in one unknown.*

▷ **Cue:** "least diameter of a dissection into k parts" ⟶ **Move:** lower bound: find $k+1$ points pairwise at distance $\geq d$; pigeonhole puts two in one part

▷ **Cue:** an extremal point configuration with several distances that must all be large ⟶ **Move:** guess that the binding distances are equal at the optimum; write that as equations and solve

▷ **Cue:** a piece around a vertex whose angle is less than 60° ⟶ **Move:** take a circular sector of radius d at that vertex; its diameter is exactly d

▷ **Cue:** a convex polygon whose diameter must be checked $\longrightarrow$ **Move:** check only vertex-to-vertex distances

Layer 3 THE WHY

Why is the five-point pigeonhole the natural move? Go to the definition. The diameter of a dissection is a supremum over pairs of points *in the same part*. To force it to be large you must force two far-apart points into the same part. You do not control the dissection — the adversary chooses it — so the only way to force two points together is to have more points than parts. Four parts, so five points. That is not a trick; it is the only lever the definition offers. For a "min over partitions of max within a part" quantity, the standard lower bound is exactly this pigeonhole on $k+1$ well-separated points.

Why those five points? Maximising the minimum distance among five points of a triangle pushes points outwards to the boundary, and the three vertices are the most isolated points of the region. So A, B, C go in. The remaining two points must be at distance $\geq d$ from every chosen point, and the most room is on the two longest sides, $AB = 5$ and $CB = 4$. At an optimum you cannot move any point to increase the smallest distance, which in practice means the smallest distances are all equal and "locked". Here the lock is the chain $A \to D \to E \to B$ running round from A to B, and $AD = DE = EB = d$ gives a quadratic with roots 5 and $25/13$. The root $d = 5$ puts D at B and E at $(-1, 0)$, outside the triangle, so it is spurious and $d = 25/13$. The strange number is not luck: the 13 is the leading coefficient of the quadratic, and $13d^2 - 90d + 125$ has discriminant $8100 - 6500 = 1600 = 40^2$, which is why the answer is rational at all. I have not proved that this five-point configuration is optimal among all five-point configurations; the proof does not need it, because the matching dissection proves the bound is attained.

Why does the construction look the way it does? Once you know d, each of the five points wants a piece that contains no other of the five points. The pieces at A and B have a natural shape: the angle at A is $\arctan(4/3) \approx 53.1°$ and at B it is $\approx 36.9°$, both below $60°$, so a circular sector of radius d at each has diameter exactly d (two points on the arc are at most a chord apart, and the chord subtending an angle below $60°$ is shorter than the radius). So think of sectors of radius $25/13$ at A and B: the arc at A runs from D on AB to $(0, 14/13)$ on AC, and the arc at B runs from J on AB to E on CB. The piece BEJ is the triangle inscribed in the sector at B; the piece $ADFH$ is a polygon whose vertices are all within $25/13$ of A ($AF = AD = 25/13$), so it sits inside the sector at A. Then split the remainder into a piece around C (the rectangle $CGFH$) and a middle piece containing the segment DE, which has length exactly d. The middle piece's diameter being DE is forced: D and E belong to it, and they are at distance d.

Now the official solution. Bhargava, Kedlaya and Ng state the points D and E and list the distances without saying where they come from. Here is where they come from: they are the solution of $AD = DE = EB$, the equal-tension condition on the chain of close pairs. Their remark that the pieces at A and B lie in sectors of angle less than $\pi/3$ is the design principle of the construction, placed at the end as a verification shortcut; it should be read as the reason the construction exists. And their $J = (32/15, 15/13)$ is a typo for $(32/13, 15/13)$; see the caveat.

The natural attempt failed for a structural reason: it searched only on the upper-bound side. On an extremal problem, the lower-bound certificate is what tells you what the extremal object must look like.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "least diameter of a dissection into four parts". On seeing it, write: *five points, pairwise $\geq d$, forces diameter $\geq d$.*

2. Pigeonhole for the lower bound; an equal-distance equation $AD = DE = EB$ to find d; then sectors at the acute vertices and vertex-distance checks for the construction.

3. Hesitation comes after $d = 25/13$ appears: it looks too strange to be right. It is reinforced away by building the dissection and checking vertex distances; if the construction reaches

d, the number is the answer.

4. The trap is searching dissections without a lower bound, which never terminates, and the area/isodiametric bound, which is far too weak. A second trap is copying the official point $J = (32/15, 15/13)$, which is off the hypotenuse.

5. Recognise faster: any "minimise the largest within-part quantity over partitions into k parts" is paired with "find $k+1$ objects pairwise bad". Look for both at once, and let the extremal $k+1$ objects dictate the partition.

6. Reusable decision rule: *for a min–max over partitions into k parts, get the lower bound from $k+1$ points pairwise far apart, find those points by equating the binding distances, and then build each part around one of the points.*