DAWN PUTNAM SESSION

Putnam 2017 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2017 — Paper A **5**
2017 A1 5
2017 A2 7
2017 A3 9
2017 A4 12
2017 A5 15
2017 A6 18

Putnam 2017 — Paper B **23**
2017 B1 23
2017 B2 25
2017 B3 27
2017 B4 30
2017 B5 32
2017 B6 35

Putnam 2017 — Paper A

2017 A1

2017 A1. Let S be the smallest set of positive integers such that

(a) 2 is in S,

(b) n is in S whenever n^2 is in S, and

(c) $(n+5)^2$ is in S whenever n is in S.

Which positive integers are not in S?
(The set S is "smallest" in the sense that S is contained in any other such set.)

Fields: Number theory / elementary set closure (residues mod 5, derived recursions) **Position:** A1 – the opener **Answer:** The positive integers not in S are exactly 1 and every multiple of 5.

Layer 1 THE WHAT — surface analysis

The natural attempt. The two rules look asymmetric: (c) squares and shifts, (b) un-squares. A strong first instinct is to just run them and see what falls out. Apply (c) to 2: $(2+5)^2 = 49 \in S$. Apply (c) again to 49: $(49+5)^2 = 54^2 = 2916 \in S$. Since $2916 = 54^2$, rule (b) gives $54 \in S$. Apply (c) to 54: $59^2 = 3481 \in S$; apply (b) if 3481 happens to be a useful square of something smaller – but it is only 59^2, so nothing new falls out that way, and the numbers are only growing. This is the trap: without a plan, (c) doubles the size of the working number roughly every step (squaring), and (b) only occasionally hands back something small, and only when a number you already have happens to be a perfect square. Chased this way, the search sprawls – 49, 2916, 54, 3481, ... – and there is no visible reason it should ever produce a small number again, let alone every number outside a clean residue pattern. A student can spend the whole allotted time generating ever larger elements of S without ever being in a position to say which numbers are *excluded*, because nothing here bounds the search or organizes it.

The model chain. Notice first that (b) and (c), composed in sequence, cancel the squaring: if $n \in S$, then $(n+5)^2 \in S$ by (c), and since $(n+5)^2$ is visibly a perfect square, (b) applied to it gives $n+5 \in S$. So combining the two rules yields a hidden third rule, call it (d): $n \in S \implies n+5 \in S$. This converts a square-root-chasing problem into a residue-class problem mod 5, since adding 5 never changes a number's residue mod 5.

Now hunt, using only (b) and (c) as before, for one small representative in each nonzero residue class mod 5 other than 0. Starting from 2: $2 \xrightarrow{c} 49 \xrightarrow{c} 54^2 = 2916 \xrightarrow{d} 56^2 = 3136 \xrightarrow{b} 56 \xrightarrow{d} 121 = 11^2 \xrightarrow{b} 11 \xrightarrow{d} 16 = 4^2 \xrightarrow{b} 4 \xrightarrow{d} 9 = 3^2 \xrightarrow{b} 3$; and separately $16 \xrightarrow{d} 36 = 6^2 \xrightarrow{b} 6$. (Each arithmetic step is elementary and checks directly: $54 + 220 = 56$ with $220 = 5 \cdot 44$; $56 + 65 = 121$ with $65 = 5 \cdot 13$; $11 + 5 = 16$; $16 + 20 = 36$ with $20 = 5 \cdot 4$.) So $2, 3, 4, 6 \in S$.

By rule (d), S therefore contains every $n + 5k$ ($k \geq 0$) for $n \in \{2, 3, 4, 6\}$: all of residue 2 mod 5 from 2 up, all of residue 3 from 3 up, all of residue 4 from 4 up, and all of residue 1 from 6 up (since $6 \equiv 1 \bmod 5$). That is every positive integer with residue $1, 2, 3, 4$ mod 5 *except* the

number 1 itself, which is residue 1 but below the starting point 6 of that chain. No multiple of 5 is ever produced, since $2 \in S$ and (c),(b) applied to any $n \not\equiv 0 \pmod 5$ never yield a multiple of 5: $(n+5)^2 \equiv n^2 \pmod 5$, and $n^2 \equiv 0$ only if $n \equiv 0$, so the property "not a multiple of 5" is preserved by (c), and taking a square root of a non-multiple of 5 cannot produce a multiple of 5 either. One checks directly that the complementary set – every positive integer except 1 and the multiples of 5 – itself satisfies (a), (b), (c): it certainly contains 2; if n^2 is in it then $n^2 \neq 1$ and $5 \nmid n^2$, so $n \neq 1$ (as $n = 1 \Rightarrow n^2 = 1$, excluded) and $5 \nmid n$, so n is in it; and if n is in it then $(n+5)^2 \geq 49 > 1$ and is not a multiple of 5 since n is not, so $(n+5)^2$ is in it. Hence this complementary set is itself a valid S, and since S is the *smallest* such set, S is contained in it – combined with the chain above showing S contains everything in it, S equals it exactly. The integers not in S are 1 and every multiple of 5.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains run the identical low-level moves – apply (c), check for a perfect square, apply (b). They part company at the very first step of organizing the search: the natural attempt applies (b) and (c) directly to the running numbers with no auxiliary structure, so the search has no stopping condition and no way to characterize an infinite exclusion set. The model chain instead first composes (b) after (c) as a single operation and asks what that composition *means* before doing any more arithmetic.

> *If $n \in S$ forces $(n+5)^2 \in S$, and $(n+5)^2$ is automatically a perfect square, then I get to apply rule (b) to it for free – so composing (c) then (b) is not two separate moves, it is one derived rule "$n \mapsto n+5$" that I should extract and name before I do anything else, because a set closed under $n \mapsto n+5$ is a set I can describe completely just by knowing its residues mod 5 and its smallest element in each residue class.*

Layer 3 THE WHY

Why is mod 5 the right lens, rather than parity or some other modulus? Because it is forced by the problem's own phrasing, not chosen: rule (c) literally adds 5 before squaring, so the derived rule (d) – the one hidden inside the composition of (b) and (c) – adds exactly 5, and nothing else in the problem statement introduces any other modulus. Once (d) is found, the question "which positive integers are not in S" stops being a question about individual numbers and becomes a question with exactly two parts, one per structural feature of the problem: (i) which residue classes mod 5 does the base case 2, propagated through (b) and (c), eventually touch, and (ii) within a touched class, is the smallest element reachable also the smallest positive integer in that class? The answer to (i) is "all except 0 mod 5", because (c) preserves non-multiples of 5 exactly ($(n+5)^2 \equiv n^2 \not\equiv 0$ whenever $n \not\equiv 0$) and $2 \not\equiv 0$. The answer to (ii) is "yes for residues $2, 3, 4$, no for residue 1", because the explicit chain above finds representatives $2, 3, 4$ that are already the smallest positive members of their classes, but its representative for residue 1 is 6, not 1 – and this gap is not a failure of the search, it is the entire content of the answer: S is smallest, so it contains no more than these chains force, and nothing forces $1 \in S$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that rule (b) reverses exactly the operation rule (c) performs (squaring), so the very first thing to write down on reading the problem is: "what happens if I apply (c) then (b) to the same n?"
2. That cue suggests deriving the hidden linear rule $n \in S \Rightarrow n+5 \in S$ before doing any exploratory arithmetic, which turns the problem into one about residues mod 5 and smallest representatives, rather than a search through squares.
3. Hesitation occurs right after the first couple of applications of (c) and (b) produce large numbers ($49, 2916, \ldots$) with no small numbers yet in sight; it is reinforced away by trusting

that repeated square-rooting must eventually shrink the numbers back down, since each application of (b) roughly halves the number of digits, so a finite search from 2 must terminate.

4. The trap is treating (b) and (c) as independent tools to apply ad hoc, rather than noticing they compose into a single, much simpler rule. A second trap is claiming "$1 \notin S$ and that is the whole answer," forgetting to check the entire residue class 0 mod 5, which the problem's phrasing (adding 5 throughout) should already flag as special.

5. Recognition would be faster next time by treating any generative-set definition that pairs a squaring rule with a rooting rule as an automatic trigger to compose the two and look for a simpler, division/root-free rule hiding inside – here, additive closure mod the constant that appears inside the squaring rule.

6. The reusable decision rule: when a set is defined by closure under an operation and its (partial) inverse, first compose the two to see whether a simpler, structure-revealing rule falls out; if an additive shift by a constant c appears, reduce the whole problem to residues mod c, then separately determine, for each residue class the base case can reach, the smallest element actually forced into the set – the answer to "which are excluded" is almost always a mix of "one full residue class" and "one anomalously small element missed by the search."

2017 A2

2017 A2. Let $Q_0(x) = 1$, $Q_1(x) = x$, and

$$Q_n(x) = \frac{(Q_{n-1}(x))^2 - 1}{Q_{n-2}(x)}$$

for all $n \geq 2$. Show that, whenever n is a positive integer, $Q_n(x)$ is equal to a polynomial with integer coefficients.

Fields: Algebra / recursive polynomial sequences (auxiliary recursion, uniqueness from an invariant) **Position:** A2 – second problem, still an opener **Answer:** $Q_n(x)$ equals $P_n(x)$, where $P_0 = 1$, $P_1 = x$, $P_n = xP_{n-1} - P_{n-2}$ for $n \geq 2$; this is manifestly an integer polynomial by direct induction on the defining recursion.

Layer 1 THE WHAT — surface analysis

The natural attempt. The definition itself divides, so the direct route is strong induction: assume $Q_{n-1}(x)$ and $Q_{n-2}(x)$ are integer polynomials, and try to show $Q_{n-2}(x)$ divides $Q_{n-1}(x)^2 - 1$ *as polynomials*, with the quotient having integer coefficients. This is correct in spirit, but the induction hypothesis gives no leverage: knowing Q_{n-1}, Q_{n-2} are integer polynomials says nothing about whether one divides a combination of the other. A student might try to compute a few terms by hand to find the pattern: $Q_2 = (x^2 - 1)/1 = x^2 - 1$; $Q_3 = ((x^2 - 1)^2 - 1)/x = (x^4 - 2x^2)/x = x^3 - 2x$ (dividing out cleanly, but why should this division *always* be exact for every n, and why should the quotient always be a polynomial rather than a polynomial plus a remainder?); $Q_4 = ((x^3 - 2x)^2 - 1)/(x^2 - 1)$, which requires actually carrying out polynomial long division to confirm it comes out even. Doing this case by case is a dead end: each new n requires a fresh division to check, and nothing from the $n = 2, 3$ computations transfers to justify the $n = 4$ case in general. This is the wall – the recursion, taken at face value, gives no reason the division should ever terminate cleanly, and checking finitely many cases proves nothing about all n.

The model chain. Rewrite Q_n's defining relation with the denominator cleared:

$$Q_{n-1}(x)^2 - Q_n(x)Q_{n-2}(x) = 1 \qquad (n \geq 2).$$

This is a purely multiplicative identity – no division appears in it at all. Now, instead of trying to prove this identity *about* Q directly (which is exactly as hard as the original problem), construct an entirely different, division-free sequence and show *it* satisfies the identical identity. Define $P_0 = 1$, $P_1 = x$, $P_{-1} = 0$, and $P_n = xP_{n-1} - P_{n-2}$ for $n \geq 1$. Every P_n is visibly an integer-coefficient polynomial by direct induction on this recursion, with no denominators anywhere.

Claim: $P_{n-1}(x)^2 - P_n(x)P_{n-2}(x) = 1$ for every $n \geq 1$. Proof by induction on n. Base case $n = 1$: $P_0^2 - P_1P_{-1} = 1 - x \cdot 0 = 1$. Inductive step: assume $P_{n-1}^2 - P_nP_{n-2} = 1$; show $P_n^2 - P_{n+1}P_{n-1} = 1$. Using $P_{n+1} = xP_n - P_{n-1}$,

$$P_n^2 - P_{n+1}P_{n-1} = P_n^2 - (xP_n - P_{n-1})P_{n-1} = P_n^2 - xP_nP_{n-1} + P_{n-1}^2.$$

By the induction hypothesis $P_{n-1}^2 = 1 + P_nP_{n-2}$, so this equals

$$1 + P_nP_{n-2} + P_n^2 - xP_nP_{n-1} = 1 + P_n\big(P_n + P_{n-2} - xP_{n-1}\big).$$

But $P_n = xP_{n-1} - P_{n-2}$ by the defining recursion, so $P_n + P_{n-2} - xP_{n-1} = (xP_{n-1} - P_{n-2}) + P_{n-2} - xP_{n-1} = 0$. Hence the whole bracket vanishes and the expression equals 1, completing the induction.

So P satisfies the identical cleared-denominator identity that defines Q. Since $P_0 = 1 = Q_0$ and $P_1 = x = Q_1$, and since $\deg P_n = n$ (immediate by induction from $P_n = xP_{n-1} - P_{n-2}$, so $P_{n-2}(x) \neq 0$ as a polynomial for all $n \geq 2$), the identity $P_{n-1}^2 - P_nP_{n-2} = 1$ can be solved uniquely for P_n given P_{n-1}, P_{n-2} with $P_{n-2} \neq 0$: $P_n = (P_{n-1}^2 - 1)/P_{n-2}$, which is exactly Q_n's defining recursion. By induction on n, since P and Q agree at $n = 0, 1$ and both satisfy the same deterministic recursion at every later step (using that $P_{n-2} = Q_{n-2}$ is a nonzero polynomial, so the division is well defined as rational functions and both sides equal the same ratio), $P_n(x) = Q_n(x)$ as rational functions, hence as polynomials, for every $n \geq 0$. Since every P_n is an integer-coefficient polynomial, so is every Q_n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that the raw recursion, taken step by step, is hard to certify. They split at the point where a proof strategy must be chosen: the natural attempt keeps trying to prove divisibility *inside* the given recursion, computing Q_2, Q_3, Q_4 explicitly and hoping a pattern justifies the general step. The model chain abandons the given recursion as an object of direct proof and instead searches for a *different* recursion, with no division built in, that can be shown to satisfy the same cleared-denominator identity.

> *Proving that a division always comes out exact, from inside the very recursion that performs the division, is circular in difficulty – I am trying to prove integrality using the same machinery whose integrality is in question. What I actually need is a completely different, manifestly-integer sequence that I can independently verify satisfies the identity $a_{n-1}^2 - a_na_{n-2} = 1$ that Q's recursion is equivalent to; if it does, and it starts the same way Q does, uniqueness of the recursion (given nonzero denominators) forces the two sequences to coincide, and the integrality question is then answered by the new sequence's own, division-free recursion instead of Q's.*

Layer 3 THE WHY

Why does guessing an auxiliary linear recursion actually settle the question, rather than merely relocating it? Because the identity $Q_{n-1}(x)^2 - Q_n(x)Q_{n-2}(x) = 1$, once $Q_{n-2}(x) \neq 0$, determines $Q_n(x)$ *uniquely*: that is simply what it means for Q_n to be defined by the division formula. Determinism is the whole content of a well-posed recursion – given the same first two terms and the same rule at every step, two sequences that both obey it must be identical wherever the rule is well defined. This means the problem has quietly split into two separable

questions that the raw recursion tangles together: (i) *existence and non-vanishing* – does the division in Q_n's formula ever fail, i.e., can $Q_{n-2}(x)$ ever be the zero polynomial or fail to divide $Q_{n-1}(x)^2 - 1$ evenly? – and (ii) *integrality* – granting the division works, are the resulting coefficients integers? Trying to answer both at once by staring at Q's own recursion is what stalls the natural attempt. The auxiliary sequence P_n answers (ii) for free, by construction (its recursion has no division in it at all), and the identity proved by induction – which needs nothing but the defining relation $P_n = xP_{n-1} - P_{n-2}$ and simple algebraic substitution – simultaneously answers (i): it shows the division Q's recursion performs is always exact, *because* it coincides at every step with the always-exact, division-free recursion for P. The auxiliary polynomial is not a lucky guess pulled from nowhere: its recursion $xP_{n-1} - P_{n-2}$ is the natural degree-matching linear recursion once $Q_2 = x^2 - 1$ is computed and compared with the Chebyshev-style pattern $x \cdot x - 1$, i.e. it is the same shape one expects from any two-term linear recursion with a quadratic conserved quantity (the same identity underlies the Fibonacci identity $F_{n-1}F_{n+1} - F_n^2 = (-1)^n$).

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that Q_n's definition is a *quotient* of polynomials asserted to be a polynomial; division is the obstruction to "obviously a polynomial," so the first thing to write down is the cleared-denominator identity $Q_{n-1}^2 - Q_nQ_{n-2} = 1$, which contains no division.
2. That identity suggests searching for a second, division-free recursion satisfying the exact same identity and the same initial values, rather than trying to certify the original division directly.
3. Hesitation occurs at the moment of proposing the auxiliary recursion $P_n = xP_{n-1} - P_{n-2}$: it is not obvious why this particular guess should be the right one. The hesitation is reinforced away by remembering that uniqueness, once the identity and matching initial conditions are in hand, means *any* correctly-guessed division-free recursion satisfying the identity will do – the guess only needs to be checked, not justified in advance, and computing $Q_2 = x^2 - 1$ by hand already hints at the linear, degree-matching shape to try.
4. The trap is trying to prove non-vanishing of $Q_{n-2}(x)$ and exactness of the division simultaneously, by strong induction directly on Q's own definition – tangling existence and integrality into a single, much harder induction, when separating them via an independently verified auxiliary sequence makes each part routine.
5. Recognition would be faster next time by treating any recursively-defined sequence given via a rational (division-based) formula, asked to be shown integral or polynomial, as an automatic trigger to clear denominators into an identity first, then search for an equivalent, division-free recursion satisfying that same identity.
6. The reusable decision rule: to prove a division-based recursion secretly always produces integers or integer-coefficient polynomials, do not attempt to prove the division works directly from inside the recursion; instead clear denominators to expose a polynomial identity, then construct an auxiliary, division-free recursion, verify by induction that it satisfies the identical identity and initial conditions, and conclude the two sequences coincide wherever the original division is meant to be defined – integrality then follows for free from the auxiliary sequence's manifestly polynomial recursion.

2017 A3

2017 A3. Let a and b be real numbers with $a < b$, and let f and g be continuous functions from $[a, b]$ to $(0, \infty)$ such that $\int_a^b f(x)\,dx = \int_a^b g(x)\,dx$ but $f \neq g$. For every positive integer

n, define

$$I_n = \int_a^b \frac{(f(x))^{n+1}}{(g(x))^n}\,dx.$$

Show that $I_1, I_2, I_3, \ldots$ is an increasing sequence with $\lim_{n\to\infty} I_n = \infty$.

Fields: Real analysis (integral inequalities, using a linear hypothesis as a free zero-insertion device) **Position:** A3 – third problem, the first with real bite **Answer:** The sequence is strictly increasing and diverges to ∞: increase follows from writing $I_{n+1} - I_n$ in terms of the ratio $r = f/g$ and factoring it as a manifestly nonnegative same-sign product, and divergence follows from the existence of a subinterval where $f \geq cg$ for some $c > 1$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The most natural first move on seeing $I_1, I_2, I_3, \ldots$ is to compute $I_1 - I_0$ (extending the definition to $I_0 = \int_a^b f$), since that is the simplest case and the hypothesis $\int f = \int g$ ought to appear there directly. Indeed

$$I_1 - I_0 = \int_a^b \frac{f^2}{g}\,dx - \int_a^b f\,dx = \int_a^b \frac{f}{g}(f-g)\,dx = \int_a^b \frac{(f-g)^2}{g}\,dx > 0,$$

using $\int_a^b (f-g)\,dx = 0$ to rewrite $\int (f/g)(f-g) = \int \left(\frac{f}{g} - 1\right)(f-g) + \int (f-g) = \int \frac{(f-g)^2}{g}$. This is clean and correct, and a student can be forgiven for feeling the problem is basically solved – just repeat this for $I_2 - I_1$, $I_3 - I_2$, and so on. But trying to redo it for $I_2 - I_1$ by direct expansion,

$$I_2 - I_1 = \int_a^b \left(\frac{f^3}{g^2} - \frac{f^2}{g}\right) dx = \int_a^b \frac{f^2}{g^2}(f-g)\,dx,$$

now has an f^2/g^2 in front instead of f/g, and the same trick of subtracting $\int (f-g) = 0$ gives $\int \left(\frac{f^2}{g^2} - 1\right)(f-g)\,dx$ – a genuinely different-looking expression from the $n = 1$ case, with $f^2/g^2 - 1$ in place of $f/g - 1$. A student working case by case has no reason yet to expect these to factor the same way for every n; each new n looks like it needs its own algebra, and nothing forces the realization that $r^{n+1} - 1$ and $r - 1$ always share a sign for *every* n at once, for the same structural reason. Worked case by case, this is an infinite family of separate small proofs, not one proof – and divergence is not addressed by this approach at all, since "each term is bigger than the last" does not by itself imply the terms grow without bound.

The model chain. Write $r(x) = f(x)/g(x) > 0$. For general n,

$$I_{n+1} - I_n = \int_a^b \left(\frac{f^{n+2}}{g^{n+1}} - \frac{f^{n+1}}{g^n}\right) dx = \int_a^b \frac{f^{n+1}}{g^{n+1}}(f-g)\,dx = \int_a^b r^{n+1}(f-g)\,dx.$$

Subtract the free zero term $\int_a^b (f-g)\,dx = 0$ (licensed by the hypothesis $\int f = \int g$):

$$I_{n+1} - I_n = \int_a^b \left(r^{n+1} - 1\right)(f-g)\,dx.$$

Now factor $r^{n+1} - 1 = (r-1)(r^n + r^{n-1} + \cdots + 1)$ and note $f - g = g(r-1)$, so

$$I_{n+1} - I_n = \int_a^b g(x)\,(r(x)-1)^2\left(r(x)^n + r(x)^{n-1} + \cdots + 1\right) dx.$$

Every factor in the integrand is nonnegative – $g > 0$, $(r-1)^2 \geq 0$, and $r^n + \cdots + 1 > 0$ since $r > 0$ – so $I_{n+1} \geq I_n$ for every $n \geq 0$, uniformly, by one identity that did not need to be re-derived for each n. Equality would force $(r(x)-1)^2 = 0$ for (almost) every x, i.e. $f = g$ on a

dense subset of $[a,b]$, hence everywhere by continuity – contradicting $f \neq g$. So the sequence is strictly increasing.

For divergence, use the hypothesis $\int f = \int g$ a second time, differently: if $f(x) \leq g(x)$ for every $x \in [a,b]$, then $\int f \leq \int g$ with equality forcing $f = g$ everywhere (both continuous), contradicting $f \neq g$; so there is some point $t \in [a,b]$ with $f(t) > g(t)$. By continuity of $f - g$, there is a closed subinterval $[t_0, t_1] \subset [a,b]$ (with $t_0 < t_1$) and a constant $c > 1$ such that $f(x) \geq c\, g(x)$ for all $x \in [t_0, t_1]$. Then, since the integrand of I_n is positive everywhere and can only be enlarged by keeping just this subinterval,

$$I_n \;\geq\; \int_{t_0}^{t_1} \frac{f^{n+1}}{g^n}\, dx \;\geq\; \int_{t_0}^{t_1} c^n f(x)\, dx \;=\; c^n \int_{t_0}^{t_1} f(x)\, dx.$$

Since $f > 0$ is continuous on $[t_0, t_1]$ with $t_0 < t_1$, $\int_{t_0}^{t_1} f\, dx > 0$ is a fixed positive constant, and $c^n \to \infty$ as $n \to \infty$ because $c > 1$. Hence $I_n \to \infty$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin with the identical computation of $I_{n+1} - I_n$ as $\int (f^{n+1}/g^{n+1})(f - g)\, dx$, and both use $\int (f - g) = 0$ once, at $n = 0$, to turn that into a manifest square. They split exactly at the decision of *how* to use the zero-integral hypothesis for general n: the natural attempt applies it once, gets a clean $n = 0$ result, and then re-derives the algebra by hand for each subsequent n, never separating out the part of the argument that is n-independent. The model chain instead abstracts to the ratio $r = f/g$ *before* fixing n, so that the zero-integral subtraction and the resulting factorization are performed symbolically, once, for a generic exponent $n + 1$.

> *The hypothesis $\int_a^b (f - g)\, dx = 0$ is not a fact to be used once at $n = 0$ and then forgotten – it is a standing license to insert or remove $\lambda \int (f - g)\, dx$ from any expression I am integrating, for absolutely any constant or function λ, without changing the value. Rather than expanding $I_{n+1} - I_n$ in terms of f and g separately for each n and hoping a pattern emerges, I should rewrite the whole integrand purely in terms of the single ratio $r = f/g$ first, since that is the one quantity every I_n actually depends on structurally, and then apply the zero-insertion trick symbolically in n rather than numerically.*

Layer 3 THE WHY

Why does collapsing everything to the single ratio $r = f/g$ work, and why is $(r^{n+1} - 1)(r - 1) \geq 0$ the natural target rather than something else? Because for $r > 0$, the function $x \mapsto x^{n+1}$ is strictly increasing through the fixed point $x = 1$: it sends $r > 1$ to something > 1 and $r < 1$ to something < 1, for every positive integer n, simultaneously and for the same reason (monotonicity of a power function on the positive reals). So $r^{n+1} - 1$ and $r - 1$ are automatically the same sign, for every n at once – this is not a computation to redo per n, it is a single fact about power functions through 1 that the algebra only needs to invoke, not reprove. The reduction to r is forced by the problem's own structure: I_n's integrand f^{n+1}/g^n can be written $g \cdot r^{n+1}$, so the only place n actually enters is as an exponent on r, and the only hypothesis linking f and g globally – $\int f = \int g$, i.e. $\int g(r - 1) = 0$ – is exactly a statement about r weighted by g. Once both the definition and the hypothesis are re-expressed in terms of r, the n-dependence of the whole problem is confined to a single scalar exponent inside a monotone function, and monotonicity through a fixed point is the correct tool precisely because it is the one property of $x \mapsto x^{n+1}$ that holds uniformly in n. The same reasoning explains why the divergence argument needs a genuinely separate device (a subinterval where $r \geq c > 1$): an increasing sequence need not diverge, so boundedness must be ruled out independently, and it is ruled out by the same ratio r exceeding 1 somewhere – which must happen, again by the

zero-mean property of $r-1$ weighted by g, since $r \le 1$ everywhere would force $\int g(r-1) \le 0$ with equality only when $r \equiv 1$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that $\int_a^b f = \int_a^b g$ is equivalent to $\int_a^b (f-g)\,dx = 0$; this should be written down immediately as a standing license to add or subtract $\lambda \int (f-g)$ from any integral being computed, for any λ, without changing its value – not as a one-time substitution to be used only where it is first needed.

2. That cue suggests rewriting the entire family I_n, and its consecutive difference, in terms of the single ratio $r = f/g$ before fixing any particular n, so that the zero-insertion trick and the resulting sign argument are done symbolically in n rather than recomputed by hand for each value.

3. Hesitation occurs exactly after the $n = 0$ (or $n = 1$) case comes out cleanly: the natural next move is to redo the same expansion for $n = 2$ and see if it still works, rather than asking whether the $n = 0$ computation used anything specific to $n = 0$ at all. It is reinforced away by explicitly checking, before moving to a new n, which parts of the previous computation depended on the exponent and which did not.

4. The trap is treating "the sequence is increasing" and "the sequence diverges" as two consequences of the same estimate. They are not: an increasing sequence can converge to a finite limit, so divergence needs its own, separate argument (here, the existence of a subinterval with $f \ge cg$, $c > 1$), and a solution that proves monotonicity and stops has answered only half the question.

5. Recognition would be faster next time by treating any problem with two positive functions sharing equal total integral, compared via a ratio raised to a growing power, as an automatic trigger to substitute $r = f/g$ immediately and re-express both the objects being compared and the zero-mean hypothesis purely in terms of r.

6. The reusable decision rule: whenever a hypothesis gives $\int (f-g) = 0$ (or any linear identity summing or integrating to zero), treat it as a standing license to insert a matching zero term into any difference under study, and prefer doing so *after* reducing the objects being compared to a single ratio or normalized quantity, so that the resulting sign argument is proved once, uniformly in whatever index or exponent varies, rather than re-derived case by case.

2017 A4

2017 A4. A class with $2N$ students took a quiz, on which the possible scores were $0, 1, \ldots, 10$. Each of these scores occurred at least once, and the average score was exactly 7.4. Show that the class can be divided into two groups of N students in such a way that the average score for each group was exactly 7.4.

Fields: combinatorics: discrete intermediate value argument, local improvement (swapping), parity invariant **Position:** A4 — mid-paper. One idea, but it is an idea about *method* (stop constructing, start improving), and the number 7.4 is placed to lure you into arithmetic that does not matter. **Answer:** Always possible: start from any split into two groups of N and repeatedly swap a score $v+1$ in the heavier group for a score v in the lighter one; each swap cuts the difference of the sums by exactly 2, and the difference is always even, so it reaches 0.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start with the number, because the number is the most conspicuous thing on the page. The total score is $7.4 \cdot 2N = \frac{74N}{5}$, an integer, so $5 \mid N$; write $N = 5m$. Then the total is $74m$ and each group must total $37m$. So far so good, and all of it is correct.

Now you try to *build* a group. Perhaps: sort the scores and deal them out alternately, first to group 1, second to group 2, and so on. The sums come out close, but "close" is all you can say — the difference is at most the largest gap between consecutive sorted scores, which can be anything up to 10. Perhaps greedy: hand each student, largest score first, to whichever group currently has the smaller total. Greedy guarantees the final difference is at most 10, and again gives no control over whether it is 0. Perhaps blocks of ten students averaging 7.4 (total 74)? The distribution of scores is arbitrary — it could be $2N - 11$ students on 8 and one each on $0, \ldots, 10$, adjusted to fit — and there is no reason the class decomposes into such blocks.

This is the wall. Every construction you try gives you a near-miss, the near-miss is off by an unknown even-ish amount, and you start inventing case-by-case repairs. The honest state of the attempt after twenty minutes is: "I can get within 10; I cannot get to 0." Notice also what you have *not* used: the hypothesis that every score from 0 to 10 occurs. A construction that does not use a hypothesis cannot be the proof, because without that hypothesis the claim is false: ten students scoring $10, 10, 10, 10, 10, 10, 10, 4, 0, 0$ average exactly 7.4, but no five of them total 37, since the tens contribute a multiple of 10 and the rest contribute 0 or 4.

The model chain. *Step 1: parity.* As above, $5 \mid N$ and the total score is $\Sigma = 74m$, which is even. For any split of the class into groups A and B of N students each, write ΣA and ΣB for the group totals and $D = \Sigma A - \Sigma B$. Since $D = 2\Sigma A - \Sigma$, the difference D is always even.

Step 2: the goal restated. The groups have equal averages exactly when $D = 0$. So it suffices to show: if $D \neq 0$, there is a swap of one student from each group that changes $|D|$ to $|D| - 2$. Then, starting from any split and repeating, $|D|$ runs through $|D|, |D| - 2, |D| - 4, \ldots$ and, being a nonnegative even integer, reaches 0 after finitely many swaps.

Step 3: a useful pair exists. Suppose $D > 0$ (otherwise rename the groups), so $D \geq 2$. If every score in A were at most every score in B, then, because $|A| = |B| = N$, pairing them off would give $\Sigma A \leq \Sigma B$. So some student in A has score a and some student in B has score b with $a > b$. I claim there is a value v with $b \leq v < a$ such that some student in B scored v and some student in A scored $v + 1$. Suppose not. Then for each v in $b, b + 1, \ldots, a - 1$: if v occurs in B, then $v + 1$ does not occur in A; but every value $0, \ldots, 10$ occurs somewhere, so $v + 1$ occurs in B. Since b occurs in B, induction up the values gives that a occurs only in B — contradicting a occurring in A.

Step 4: the swap. Exchange that $(v + 1)$-student from A with that v-student from B. Group sizes are unchanged, ΣA drops by 1, ΣB rises by 1, and D drops by exactly 2. Since $D \geq 2$ beforehand, $0 \leq D - 2 < D$. Combined with Step 2, this finishes the proof.

Note what was used: equal group sizes (Step 3), an even total (Step 1), and the fact that the occurring scores form a block of consecutive integers (Step 3). The specific value 7.4 was used only to get the parity.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree up to "each group must total $37m$". At that line the natural attempt asks *which* N students to put together, and tries to exhibit them. The model asks *how far* an arbitrary split is from working, and whether that distance can always be reduced by a fixed step. One is a construction problem; the other is a monotone-descent problem, and only the second has a finite, checkable answer.

> *I keep getting splits that are nearly right. So I should stop constructing and take any split at all, measure the error D, and ask what the smallest possible change to a split does to D. The smallest change is swapping one student for one student.*

> *A swap changes D by twice the difference of the two swapped scores. If I can always find two scores that differ by exactly 1 and point the right way, then D moves in steps of 2. It starts even, so it*

> *cannot step over* 0.

> *Every score* 0 *to* 10 *occurs. That is a hypothesis about* gaps*: there are no gaps in the set of values. Steps of exactly* 1 *are what that hypothesis is for.*

Layer 3 THE WHY

Why is local improvement the natural move and not a trick? Because of what the problem asks you to prove. It is an existence statement about a very large finite set (splits of $2N$ students) with no structure that singles out any particular split as "the" answer. When nothing picks out the answer, trying to write it down is fighting the problem. The only structure present is a single integer-valued quantity, D, and an existence claim "$D = 0$ somewhere". That is the shape of the intermediate value theorem: a quantity that moves in small steps and takes values of both signs must pass through zero. Here D takes both signs automatically (swap the names of the groups), so what remains is to show the steps are small enough not to jump over zero.

That is exactly where the hypotheses bite, and it explains each of them. A swap of a student scoring x in A for one scoring y in B changes D by $2(y - x)$. For the descent to be unable to overshoot, the step must be the smallest possible nonzero change, namely 2, which needs $y - x = -1$: adjacent values. The "every score occurs" hypothesis guarantees adjacent values are available, and parity guarantees that a walk in steps of 2 from an even number lands on 0 rather than on ± 1. Remove either and the general claim fails: a class scoring only 7s and 8s has no gaps but can have an odd total, and then no equal split exists; a class scoring only 0s and 10s, with an odd number of 10s, has an even total, but its swaps move D by 20 and D is always an odd multiple of 10.

Kedlaya's first solution sorts the scores $a_1 \leq \cdots \leq a_{2N}$ and slides a window $s_i = a_{i+1} + \cdots + a_{i+N}$. Here is where that comes from. It is the same intermediate value idea along a different path through the set of splits: $s_0 \leq 7.4N \leq s_N$ because $s_0 + s_N$ is the total, and as i increases the window sum moves monotonically. The difficulty is that one window move can jump by up to 10, so the window may straddle the target without hitting it; the official solution repairs the jump by taking $N + 1$ consecutive scores and omitting one whose value is the overshoot, which exists because every value between a_i and a_{i+N+1} occurs. The swap version builds the "no gaps" hypothesis into the path itself, so no repair is needed. Both are discrete intermediate value arguments; the lesson is to recognise the family.

And the 7.4? It is there to be computed with. Its only contribution to the proof is "the total is even". The problem punishes the student who treats the specific number as the heart of the matter — this is a small instance of zombie maths, where the most visible datum gets manipulated before anyone asks what role it plays.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the pair "show that the class *can be divided*" together with "each of these scores occurred at least once". Write down: *any split; error* $D = \Sigma A - \Sigma B$*; D is even; want* $D = 0$.

2. A discrete intermediate value / descent argument: find a move on splits that changes D by the smallest possible amount, and show such a move exists whenever $D \neq 0$.

3. Hesitation occurs after the arithmetic ($5 \mid N$, target $37m$), when you feel you have "set up" and must now construct. It is reinforced away by asking what each hypothesis is for: the "every score occurs" condition has not been used, and a condition about no gaps in the values is a condition about step sizes.

4. The trap is the number 7.4, which invites arithmetic and block constructions; its only role is parity. A second, quieter trap is claiming "some score in A exceeds some score in B" without the equal-size argument that justifies it.

5. The slogan: *existence of a balanced configuration among many, with no canonical candidate, is an intermediate value problem — measure the imbalance and show it can always be reduced by the minimal step.* It applies to splitting sets, balancing weights, fair division, and any claim that some quantity "can be made exactly" a target.

6. The reusable decision rule: **when asked to show that a configuration with an exact property exists and no configuration is canonical, do not construct it; define an integer-valued error, find the smallest local move, and show that move can always shrink the error without overshooting.** Check each hypothesis against the two needs: step size (no overshoot) and parity or divisibility (the target is reachable in those steps).

▷ **Cue:** show that a finite set can be split so that two sums are exactly equal, and every value in a range occurs $\longrightarrow$ **Move:** take any split, let D be the difference, and swap an adjacent pair of values v, $v+1$ across the groups to change D by exactly 2

▷ **Cue:** a hypothesis says every value in a consecutive range $0, 1, \ldots, k$ occurs $\longrightarrow$ **Move:** it is a no-gaps condition; use it to guarantee moves of the smallest possible size, typically in an intermediate value argument

▷ **Cue:** a specific decimal such as 7.4 appears in a structural existence problem $\longrightarrow$ **Move:** extract its arithmetic consequences (here: $5 \mid N$, total even) and then stop using it

▷ **Cue:** an integer quantity must be shown to hit 0 and it moves in steps of fixed size s $\longrightarrow$ **Move:** check that it starts at a multiple of s (a parity or divisibility invariant) and that a step toward 0 always exists

2017 A5

2017 A5. Each of the integers from 1 to n is written on a separate card, and then the cards are combined into a deck and shuffled. Three players, A, B, and C, take turns in the order $A, B, C, A, \ldots$ choosing one card at random from the deck. (Each card in the deck is equally likely to be chosen.) After a card is chosen, that card and all higher-numbered cards are removed from the deck, and the remaining cards are reshuffled before the next turn. Play continues until one of the three players wins the game by drawing the card numbered 1.

Show that for each of the three players, there are arbitrarily large values of n for which that player has the highest probability among the three players of winning the game.

Fields: probability: first-step analysis, probability generating functions; algebra: roots-of-unity filter; analysis: divergent series of small angles **Position:** A5 — hard. The recurrences are easy to write and useless to stare at; the problem is decided by whether you think to encode "who wins" as a residue mod 3. **Answer:** With T_n the number of turns, $\mathbb{E}[X^{T_n}] = X(X+1)\cdots(X+n-1)/n!$; at $X = \omega = e^{2\pi i/3}$ this is a complex number of argument $\frac{2\pi}{3} + \sum_{j=1}^{n-1} \arctan \frac{\sqrt{3}}{2j-1}$, which increases to ∞ in steps of at most $\pi/3$, and the leader is read off from which third of the circle that argument lies in.

Layer 1 THE WHAT — surface analysis

The natural attempt. Let a_n, b_n, c_n be the winning probabilities of A, B, C with n cards. If A draws card k, the game restarts on cards $1, \ldots, k-1$ with B to move, so A now plays the role of the third player. With the conventions $a_0 = b_0 = 0$, $c_0 = 1$ (an empty deck means the previous player, in the C seat, has won),

$$a_n = \frac{1}{n}\sum_{k=1}^{n} c_{k-1}, \qquad b_n = \frac{1}{n}\sum_{k=1}^{n} a_{k-1}, \qquad c_n = \frac{1}{n}\sum_{k=1}^{n} b_{k-1}.$$

Compute: (a_n, b_n, c_n) is $(1,0,0)$, $(\frac{1}{2}, \frac{1}{2}, 0)$, $(\frac{1}{3}, \frac{1}{2}, \frac{1}{6})$, $(\frac{7}{24}, \frac{11}{24}, \frac{1}{4})$, … and by $n = 10$ it is about $(0.302, 0.357, 0.341)$. Everything is drifting to $\frac{1}{3}$. The leader is B for a long time; you compute on and C takes over at $n = 14$, and then nothing changes for as far as hand computation goes.

This is the wall. The question is about the *order* of three numbers that all converge to $\frac{1}{3}$, so it is a question about the tiny differences, and the recurrences give no handle on them. You can subtract the recurrences to get a linear system for $x_n = a_n - b_n$, $y_n = b_n - c_n$, $z_n = c_n - a_n$ (this is what Kedlaya's first solution does, turning it into a point rotating on a circle), but on first contact you do not know that rotation is the picture, and a 3×3 non-autonomous linear recurrence is not something one solves by staring.

The model chain. *Step 1: reduce to one random variable.* Let T_n be the number of turns the game lasts. The winner is determined by T_n mod 3: A wins if $T_n \equiv 1$, B if $T_n \equiv 2$, C if $T_n \equiv 0 \pmod 3$. The players' identities have dropped out.

Step 2: the distribution of T_n. Put $T_0 = 0$. With $n \geq 1$ cards the first draw is k, uniform on $\{1, \dots, n\}$, and then the game continues on $\{1, \dots, k-1\}$, so $T_n = 1 + T_{k-1}$. Let $G_n(X) = \mathbb{E}[X^{T_n}]$, a polynomial. Conditioning on k,

$$n\, G_n(X) = X \sum_{m=0}^{n-1} G_m(X).$$

Write the same identity for $n-1$ (valid for $n \geq 2$) and subtract: $nG_n - (n-1)G_{n-1} = XG_{n-1}$, so $G_n = \frac{X+n-1}{n} G_{n-1}$. This also holds at $n = 1$, since $G_1 = X$. Hence

$$G_n(X) = \frac{X(X+1)(X+2)\cdots(X+n-1)}{n!}.$$

Step 3: filter by residue. Let $\omega = e^{2\pi i/3}$, so $\omega^3 = 1$ and $1 + \omega + \omega^2 = 0$. Then

$$z_n := G_n(\omega) = c_n + a_n\, \omega + b_n\, \omega^2.$$

Multiply by $\bar{\omega} = \omega^2$: $z_n\bar{\omega} = a_n + b_n\omega + c_n\omega^2$, whose real part is $a_n - \frac{1}{2}(b_n + c_n) = \frac{1}{2}(3a_n - 1)$, using $a_n + b_n + c_n = 1$. The same computation for the other two players gives

$$a_n = \tfrac{1}{3}\big(1 + 2\,\mathrm{Re}(z_n\omega^{-1})\big), \quad b_n = \tfrac{1}{3}\big(1 + 2\,\mathrm{Re}(z_n\omega^{-2})\big), \quad c_n = \tfrac{1}{3}\big(1 + 2\,\mathrm{Re}(z_n)\big).$$

Since $z_n \neq 0$ (each factor $\omega + j$ is non-real), write $z_n = |z_n|e^{i\phi_n}$; then $\mathrm{Re}(z_n\omega^{-k}) = |z_n|\cos(\phi_n - 2\pi k/3)$. So the player A, B or C has the strictly highest probability exactly when ϕ_n lies, modulo 2π, in the open arc $(\frac{\pi}{3}, \pi)$, $(\pi, \frac{5\pi}{3})$ or $(-\frac{\pi}{3}, \frac{\pi}{3})$ respectively: the arc of width $2\pi/3$ centred on ω, ω^2 or 1.

Step 4: the argument. Since $n! > 0$, $\phi_n = \sum_{j=0}^{n-1} \arg(\omega + j)$. For $j = 0$ this is $2\pi/3$. For $j \geq 1$, $\omega + j = (j - \frac{1}{2}) + \frac{\sqrt{3}}{2}i$ has argument

$$\theta_j = \arctan\frac{\sqrt{3}}{2j-1} \in \big(0, \tfrac{\pi}{3}\big].$$

So $\phi_n = \frac{2\pi}{3} + \theta_1 + \cdots + \theta_{n-1}$, taken as a real number (not reduced mod 2π), is increasing in n with every step at most $\pi/3$.

Step 5: the argument diverges. arctan is concave on $[0, \infty)$, so on $[0, \sqrt{3}]$ it lies above its chord: $\arctan x \geq \frac{\pi/3}{\sqrt{3}}\, x$. Hence $\theta_j \geq \frac{\pi}{3(2j-1)} \geq \frac{\pi}{6j}$, and the harmonic series gives $\phi_n \to \infty$.

Step 6: no arc can be skipped. Let $I = (L, L + \frac{2\pi}{3})$ be any open interval with $L \geq \phi_1$. Let n be the least index with $\phi_n > L$ (it exists by Step 5, and $n \geq 2$). Then $\phi_n \leq \phi_{n-1} + \frac{\pi}{3} \leq L + \frac{\pi}{3}$, so $\phi_n \in I$. Taking $L = \frac{\pi}{3} + 2\pi t$, $\pi + 2\pi t$, $\frac{5\pi}{3} + 2\pi t$ for $t = 1, 2, 3, \dots$ produces, for each player, infinitely many n at which that player strictly leads. They are arbitrarily large because each ϕ_n is finite, so $n \to \infty$ as $L \to \infty$.

Check: $z_1 = \omega$ gives $(a, b, c) = (1, 0, 0)$; $z_2 = \omega(1 + \omega)/2 = -\frac{1}{2}$ sits on the boundary between A and B, matching the tie $(\frac{1}{2}, \frac{1}{2}, 0)$. Computing ϕ_n numerically, the leader is B for

$3 \le n \le 13$, C for $14 \le n \le 144$, A for $145 \le n \le 1607$, B from 1608, C from 18041, A from 202545 — consecutive ratios near $e^{4\pi/(3\sqrt{3})} \approx 11.2$, because $\phi_n \approx \frac{\sqrt{3}}{2} \ln n$ and each arc has width $2\pi/3$. The values for $n \le 119$ were also checked against the exact rational recurrences.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on the first-step conditioning. At that point the natural attempt conditions on the first card *and keeps track of whose turn it is*, producing three coupled recurrences in three unknowns. The model conditions on the first card *and throws away whose turn it is*, keeping only the length of the game. One recurrence in one unknown, and it has a product solution.

> *The winner is whoever makes the last draw, and the turns go round in threes. So the winner is a function of the number of turns mod* 3*. I should study the number of turns, which does not care who is playing, and filter it mod* 3 *at the end.*

> *Residue mod* 3 *of a random count: evaluate the generating function at a cube root of unity. Real parts at* $\omega^0, \omega^{-1}, \omega^{-2}$ *recover the three probabilities.*

> *My recurrence is* $nG_n = X\sum_{m<n} G_m$*, a running sum. Write it one step down and subtract, and it becomes a ratio* G_n/G_{n-1} *— so* G_n *is a product.*

> *A product of complex numbers: moduli multiply, arguments add. The arguments* $\arctan\frac{\sqrt{3}}{2j-1}$ *shrink like* $1/j$*, so they add up to infinity in small steps. The point spirals round forever and cannot jump over a sector.*

Layer 3 THE WHY

Why is "the number of turns" the right object, and not a lucky reformulation? Because the three players are indistinguishable except for their position in the rotation. The game itself — a deck $\{1, \dots, m\}$, a uniform draw, a smaller deck — has no idea who is drawing. Any description that carries the players' names is carrying information the process does not use, and that redundancy is precisely why the three recurrences are coupled. Strip it out and the state is one integer, the size of the deck, and the outcome is one integer, the number of draws. The player names come back only through T_n mod 3.

Why a generating function? Because the question has become "what is the distribution of T_n modulo 3?", and a residue class of a count is extracted by a root of unity: $\sum_m p(m)\omega^m$ groups the terms by m mod 3 automatically, because ω^m depends only on m mod 3. This is the same device as the familiar identity that $\frac{1}{3}\big((1+1)^N + (1+\omega)^N + (1+\omega^2)^N\big)$ counts subsets whose size is a multiple of 3. The problem says "in the order $A, B, C, A, \dots$": a period-3 rotation. Period 3 is the cue for cube roots of unity.

Why a product? The official second solution states $P_n(X) = X(X+1)\cdots(X+n-1)/n!$ with the note that it "may be guessed by computing". Here is where it comes from. The factor $\frac{X+j-1}{j} = \big(1 - \frac{1}{j}\big) + \frac{1}{j}X$ is the generating function of a coin that shows 1 with probability $1/j$. So the product formula says T_n has the same distribution as a sum of independent indicators with probabilities $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}$. That has a direct reading: card j is ever drawn exactly when, at the first moment a draw lands in $\{1, \dots, j\}$, it lands on j itself, and that first landing is uniform on $\{1, \dots, j\}$. (This is the record/cycle structure of a random permutation: T_n is distributed as the number of cycles, with the unsigned Stirling numbers of the first kind as the coefficients.) The reading explains the product; the shift-and-subtract in Step 2 proves it.

Why must the leader rotate? At $X = \omega$ each factor $\frac{\omega+j}{j+1}$ turns the point by a small angle θ_j and shrinks it. The angles decay like $\frac{\sqrt{3}}{2j}$ — small, so the point cannot leap across a sector, yet not summable, so it never stops turning. The probabilities themselves converge to $\frac{1}{3}$ because $|z_n| \to 0$, which is what made the hand computation so uninformative: the ordering lives entirely in the direction of a vector whose length is vanishing.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "take turns in the order $A, B, C, A, \ldots$" combined with "wins the game by drawing the card numbered 1". Write down: *winner* $= T \bmod 3$, *where T is the number of turns.*

2. Probability generating function of T_n via first-step analysis, then a roots-of-unity filter at $\omega = e^{2\pi i/3}$.

3. Hesitation comes after the three coupled recurrences are written: they are correct and it feels like progress to iterate them. Reinforce it away by asking whether the process knows who is playing. It does not; so the player labels are extra state, and removing them should reduce three unknowns to one.

4. The trap is numerical: every probability tends to $\frac{1}{3}$, B leads for $3 \le n \le 13$ and C from 14 to 144, so hand computation suggests the leadership settles. It does not, but the next change is at $n = 145$. A second trap is claiming "arbitrarily large n" from the divergence of $\sum \theta_j$ alone, without the step bound $\theta_j \le \pi/3$ that prevents the argument from skipping a sector.

5. The slogan: *when an outcome depends on a count modulo k, compute the generating function of the count and evaluate at k-th roots of unity.* This covers alternating turns, colourings by residue, subset sizes in a residue class, and random walks on a cycle.

6. The reusable decision rule: **before setting up one recurrence per player (or per colour, or per state), ask what the underlying process actually distinguishes; if the labels only enter through a residue, track the unlabelled count and filter at the end.** Corollary: a product $\prod(1 + w/j)$ with $|w| = 1$, $w \neq \pm 1$, has argument growing like $\mathrm{Im}(w) \ln n$, so it winds round the circle forever in shrinking steps.

▷ **Cue:** players (or states) take turns cyclically with period k and the outcome is "who made the last move" $\longrightarrow$ **Move:** track only the number of moves T; the winner is T mod k

▷ **Cue:** a probability of a count lying in a residue class mod k is required $\longrightarrow$ **Move:** evaluate the probability generating function at the k-th roots of unity and take real parts (roots-of-unity filter)

▷ **Cue:** first-step analysis gives $n\,G_n = X \sum_{m<n} G_m$ $\longrightarrow$ **Move:** shift and subtract to get G_n / G_{n-1}, so G_n is a product (here a rising factorial over n!)

▷ **Cue:** several probabilities all converge to the same limit and the question asks which is largest $\longrightarrow$ **Move:** subtract the limit and study the direction of the error vector; the ordering is a question about an angle, not a size

▷ **Cue:** a sequence of angles increases to infinity with steps bounded by δ $\longrightarrow$ **Move:** it enters every arc of length greater than δ infinitely often: take the first index past the arc's left end

2017 A6

2017 A6. The 30 edges of a regular icosahedron are distinguished by labeling them $1, 2, \ldots, 30$. How many different ways are there to paint each edge red, white, or blue such that each of the 20 triangular faces of the icosahedron has two edges of the same color and a third edge of a different color? [Note: the top matter on each exam paper included the logo of the Mathematical Association of America, which is itself an icosahedron.]

Fields: combinatorics: counting with constraints; linear algebra over $\mathbb{F}_3$: rank–nullity, fibres of a linear map; geometry: incidence structure of the icosahedron **Position:** A6 — nominally the hardest on the

paper. It collapses once the colours are made into numbers mod 3; without that step it is a genuinely difficult dependency count. **Answer:** $2^{20} \cdot 3^{10} = 61\,917\,364\,224$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Start local. Fix a face and suppose two of its edges are already coloured. If they agree (say both red), the third must be white or blue: 2 choices. If they differ (say red and white), the third must repeat one of them, red or white: 2 choices. Either way, exactly one of the three colours is forbidden for the third edge. That is a striking fact and you should write it down.

It suggests a heuristic. Colour all 30 edges freely: 3^{30} ways. Each face independently "rejects" one colour in three for its last edge, so it keeps a fraction $\frac{2}{3}$. Twenty faces:

$$3^{30}\left(\tfrac{2}{3}\right)^{20} = 2^{20} \cdot 3^{10}.$$

Now try to make it a proof. Order the edges so that faces are completed one at a time, each by a fresh edge with exactly 2 admissible colours. It goes well at first — a band of faces can be built edge by edge — but the icosahedron is a closed surface. Late in the process a single new edge completes *two* faces at once, and it must avoid the forbidden colour of both. That leaves 2 admissible colours if the two forbidden colours coincide and 1 if they do not, and whether they coincide depends on everything coloured earlier. The factor-per-face bookkeeping breaks exactly where the constraints close up.

And this is not a technicality. Run the same heuristic on the octahedron (12 edges, 8 faces): it predicts $3^{12}(\frac{2}{3})^8 = 2^8 \cdot 3^4 = 20\,736$. A direct computer enumeration of all 3^{12} colourings gives 20 898. The heuristic is wrong there. So whether $2^{20} \cdot 3^{10}$ is right for the icosahedron is the whole problem.

The model chain. *Step 1: colours as numbers.* Identify red, white, blue with $0, 1, 2 \in \mathbb{F}_3$ (integers mod 3), in any order. For $x, y, z \in \mathbb{F}_3$:

$$x + y + z = 0 \iff x = y = z \text{ or } \{x, y, z\} = \{0, 1, 2\}.$$

Proof: if all equal, the sum is $3x = 0$; if all different, it is $0 + 1 + 2 = 3 = 0$; if exactly two agree, say $x = x \neq y$, the sum is $2x + y = y - x \neq 0$. So a face is correctly coloured (two the same, one different) if and only if the sum of its three edge values is *nonzero*.

Step 2: a linear map. Let E be the 30 edges and F the 20 faces. Define $T : \mathbb{F}_3^E \to \mathbb{F}_3^F$ by sending a colouring to the vector of face sums. T is linear. The colourings we want are exactly the preimages under T of the 2^{20} vectors in $\mathbb{F}_3^F$ with no zero entry.

Step 3: T is surjective. Take a vertex v. It has degree 5: edges $e_1, \ldots, e_5$ in cyclic order around v, and faces $F_1, \ldots, F_5$ around v with F_i containing e_i and e_{i+1} (indices mod 5). Give $e_1, \ldots, e_5$ the values $2, 1, 2, 1, 2$ and every other edge 0. Then $F_1, \ldots, F_4$ have sums $2 + 1$, $1 + 2$, $2 + 1$, $1 + 2$, all 0; F_5 has sum $e_5 + e_1 = 2 + 2 = 1$; and every other face contains none of the e_i (each e_i lies in exactly two faces, both around v), so its sum is 0. Thus T hits the basis vector of F_5. Every face arises as F_5 for a suitable vertex and cyclic labelling (take any vertex of the face and start the labelling so the face comes last), so every basis vector of $\mathbb{F}_3^F$ is in the image, and T is onto.

Step 4: count. By rank–nullity, $\dim \ker T = 30 - 20 = 10$, so $|\ker T| = 3^{10}$. For a linear map, each nonempty fibre $T^{-1}(y)$ is a translate $x_0 + \ker T$, so it has exactly 3^{10} elements; by surjectivity every fibre is nonempty. Therefore the number of colourings is

$$2^{20} \cdot 3^{10} = 1\,048\,576 \cdot 59\,049 = 61\,917\,364\,224.$$

(A computer check confirms the 20×30 face–edge incidence matrix of the icosahedron has rank 20 over $\mathbb{F}_3$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths reach the same local observation: given two edges of a face, exactly one colour is forbidden for the third. At that line the natural attempt treats the observation as a probability ($\frac{2}{3}$ per face) and multiplies. The model asks *what function* of the two given colours the forbidden colour is — and finds it is $-(x+y)$ in $\mathbb{F}_3$, which makes every constraint linear, and turns the independence question into a rank question.

> *Exactly one colour is forbidden, whatever the other two are. If the two agree, it is that colour; if they differ, it is the third one. That is the rule $z \neq -(x+y)$ mod 3. So I should call the colours $0,1,2$ and the condition is "face sum nonzero".*

> *The heuristic $3^{30}(2/3)^{20}$ is exact precisely when the twenty face-sums are independent — when every pattern of face-sums is equally achievable. For a linear map that is one statement: surjective. So all I have to prove is that I can hit each face on its own.*

> *To hit one face alone I need edge values that cancel on every face but one. Around a vertex of degree 5, alternate $2,1,2,1,\ldots$: adjacent pairs cancel, and because 5 is odd the alternation fails exactly once.*

Layer 3 THE WHY

Why must the colours become elements of $\mathbb{F}_3$? Go back to what the natural attempt found: for every pair (x,y) of colours there is exactly one forbidden z, and the rule is $z = x$ if $x = y$, and z the remaining colour if $x \neq y$. A rule that assigns one "completing" element to every pair is a binary operation on a three-element set, and it is natural to ask whether it is arithmetic in disguise. It is: label the colours $0,1,2$ in any order and the rule is exactly $z = -(x+y)$ mod 3, because $-2x = x$ mod 3, and for $x \neq y$ the residue $-(x+y)$ is the third one (the three residues sum to 0). So the problem statement's "two of the same colour and a third different" is, verbatim, the complement of "the three sum to zero mod 3". Kedlaya's solution begins by identifying the colours with $\mathbb{F}_3$. Here is where that comes from: it is the only way to write the forbidden colour as a formula in the other two, and once the forbidden value is a linear formula the whole system of constraints is a linear map.

Why is surjectivity the crux, and not a detail? A count of the form "free choices times a factor per constraint" is exactly the statement that the constraints are independent. For a linear map T, independence of the constraints means that each target pattern of face-sums is realised equally often, which happens if and only if T is onto: an onto map has equal fibres $|\ker T| = 3^{30-20}$, while a map that misses part of the target concentrates the colourings on the image. The octahedron shows the difference is real. There, each vertex has degree 4, the alternating assignment $1,2,1,2$ around a vertex cancels on *all four* faces, and the face–edge map has rank 7, not 8; only 86 of the $2^8 = 256$ nowhere-zero face patterns are in the image, giving $86 \cdot 3^{12-7} = 20\,898$ — the number the enumeration found. The parity of the vertex degree is what decides it. On the icosahedron the degree is 5, the alternation around a vertex cannot close up, and the single failure is the unit vector that proves surjectivity.

Why is this problem an A6 at all, then? Because every step after the encoding is standard linear algebra, but the encoding is not suggested by anything in the statement except the local "exactly one forbidden colour" fact, and that fact only appears if you compute it. The student who starts with case analysis on the icosahedron's faces never computes it, and is lost in a 30-edge dependency graph. The Blueprint's warning about zombie maths applies squarely: the cheap move is to look at one face first.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the face condition "two edges of the same color and a third edge of a different color" with exactly three colours. Write down: *for two given edges, how many colours are allowed on the third? Always* 2*; exactly one is forbidden.*

2. Encode the colours as $\mathbb{F}_3$ so the forbidden colour is $-(x+y)$; the constraints become "a linear map is nonzero in every coordinate"; count by rank–nullity and fibre sizes.

3. Hesitation occurs after the heuristic $3^{30}(2/3)^{20}$ appears: it looks right and you want to write it. Reinforce it away by testing the heuristic on a smaller closed surface (the octahedron, or the tetrahedron) where brute force is possible, and by naming what the heuristic assumes: independence, that is, surjectivity.

4. The trap is the per-face factor. It gives the correct answer here, so an unjustified submission would have the right number and no proof; on the octahedron the same reasoning gives a wrong number. A secondary trap is counting colourings face by face and forgetting that each edge lies in two faces.

5. The slogan: *if every local constraint forbids exactly one value, find the group in which that value is a linear function of the others; then the count is (number of allowed targets) $\times$ (kernel size), provided the map is onto.* This applies to colourings, Latin-square-type conditions, lights-out puzzles, and parity constraints on graphs.

6. The reusable decision rule: **when a count of constrained configurations looks like $q^{\#\text{variables}}$ times a fixed fraction per constraint, the claim is really that a linear (or homomorphic) map is surjective — prove that, by exhibiting a preimage for each basis vector, and do not multiply fractions until you have.** Corollary: to hit a single coordinate, look for an alternating assignment around a cycle of odd length.

▷ **Cue:** each local constraint in a counting problem forbids exactly one value, whatever the other values are $\longrightarrow$ **Move:** find an encoding (usually $\mathbb{Z}/q$) in which the forbidden value is a linear function of the others, and rewrite the constraints as a linear map

▷ **Cue:** three values in $\{0,1,2\}$, condition "not all equal and not all different" $\longrightarrow$ **Move:** this is "sum $\neq 0 \bmod 3$"

▷ **Cue:** count $= q^n \times (\text{fraction})^m$ looks right by independence $\longrightarrow$ **Move:** the independence claim is surjectivity of a linear map $\mathbb{F}_q^n \to \mathbb{F}_q^m$; each fibre then has q^{n-m} elements

▷ **Cue:** need to show a linear map on edge variables of a graph or polyhedron hits each face (or vertex) coordinate $\longrightarrow$ **Move:** alternate $+,-$ (here $2,1$) around a vertex or cycle; odd length makes the cancellation fail at exactly one place, giving a unit vector

▷ **Cue:** a heuristic count on a large symmetric object $\longrightarrow$ **Move:** test the same heuristic on a smaller object of the same kind by brute force before trusting it

Putnam 2017 — Paper B

2017 B1

2017 B1. Let L_1 and L_2 be distinct lines in the plane. Prove that L_1 and L_2 intersect if and only if, for every real number $\lambda \neq 0$ and every point P not on L_1 or L_2, there exist points A_1 on L_1 and A_2 on L_2 such that $\overrightarrow{PA_2} = \lambda \overrightarrow{PA_1}$.

Fields: Geometry (affine transformations of the plane, homothety) **Position:** B1 – the opener of the afternoon session **Answer:** The biconditional is true: L_1, L_2 intersect exactly when the vector condition holds for every $\lambda \neq 0$ and every off-line P.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct is to test the forward direction with the easiest case available. Suppose L_1, L_2 meet at a point X. For $\lambda = 1$, take $A_1 = A_2 = X$: then $\overrightarrow{PA_2} = \overrightarrow{PX} = \overrightarrow{PA_1} = 1 \cdot \overrightarrow{PA_1}$, trivially true for any P not on either line. This settles $\lambda = 1$ in one line and feels like real progress – a concrete, checkable instance of the claim. But the problem demands the vector identity for *every* $\lambda \neq 0$, not just $\lambda = 1$, and the trivial choice $A_1 = A_2 = X$ only works because $\lambda = 1$ makes the two vectors literally the same vector; for $\lambda = 2$, say, A_1 and A_2 cannot both be X (that would force $\overrightarrow{PX} = 2\overrightarrow{PX}$, i.e. $P = X$, contradicting $P \notin L_1$). So the natural attempt has proved a single instance of a universally quantified claim and has no mechanism yet for producing A_1, A_2 for an arbitrary λ – it has found the right building block (the intersection point matters) without the tool that generalises it.

The model chain. Fix P and $\lambda \neq 0$. Consider the homothety (dilation) h centred at P with ratio λ: $h(x) = P + \lambda(x - P)$. Because a homothety scales every displacement from its centre by the same factor λ, it sends any line not through P to a line with the same direction vector – that is, to a line parallel to the original. So $h(L_1)$ is a line parallel to L_1. Since L_1 and L_2 intersect, they are not parallel; since $h(L_1) \parallel L_1$, it follows that $h(L_1)$ is also not parallel to L_2. Two non-parallel lines in the plane meet at a unique point, so $h(L_1)$ and L_2 meet at a unique point A_2. Let $A_1 = h^{-1}(A_2) = P + (A_2 - P)/\lambda$; since $A_2 \in h(L_1)$, its preimage A_1 lies on L_1. By construction $A_2 - P = \lambda(A_1 - P)$, i.e. $\overrightarrow{PA_2} = \lambda \overrightarrow{PA_1}$, exactly as required, for this arbitrary P and λ.

For the converse, suppose L_1 and L_2 do *not* intersect; being distinct lines, they are then strictly parallel and share no point. Take $\lambda = 1$ and let P be any point off both lines. The condition $\overrightarrow{PA_2} = 1 \cdot \overrightarrow{PA_1}$ forces $A_2 - P = A_1 - P$, i.e. $A_2 = A_1$. But $A_1 \in L_1$ and $A_2 \in L_2$, and $L_1 \cap L_2 = \varnothing$, so no point can lie on both – $A_1 = A_2$ is impossible. Hence for this $\lambda = 1$ and this P, no such A_1, A_2 exist, so the "for every λ, P" condition fails. By contraposition, if the condition holds for every λ, P, then L_1, L_2 must intersect.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains open the same way: check $\lambda = 1$, notice the intersection point $X = A_1 = A_2$ works. Both chains agree completely up to that observation. They part at the

sentence "but the problem needs this for every λ." The natural attempt has no answer to that sentence yet; the model chain answers it by replacing the single point X with the whole *line* L_1 as the object being transported, via the map $x \mapsto P + \lambda(x - P)$.

> *I only checked that one point – the intersection – satisfies the vector condition at $\lambda = 1$. But the claim is quantified over every λ, so I cannot keep choosing A_1, A_2 one λ at a time; I need a single rule that produces A_1 from A_2 (or vice versa) for any λ at once, which means I should be asking what happens to the whole line L_1, as a set, under the map $x \mapsto P + \lambda(x - P)$ – not what happens to one of its points.*

> *For the converse, I do not need to show that* no *λ works for parallel lines – some special λ (tied to where P sits between the two parallel lines) might accidentally work, since a well-placed dilation can map L_1 onto L_2 exactly. I only need* one *bad λ, and the cheapest one to test is $\lambda = 1$, because that is the unique value that collapses the scaled vector equation $\overrightarrow{PA_2} = \lambda\overrightarrow{PA_1}$ into the much stronger statement $A_2 = A_1$, a plain point-equality that is independent of where P sits.*

- ▷ **Cue:** a claim quantified as "for every λ" relates two points via a scaled vector from a common centre P $\longrightarrow$ **Move:** recognise $x \mapsto P + \lambda(x - P)$ as a homothety, and reason about the image of an entire line under it, not about individual points
- ▷ **Cue:** disproving a "for every λ" claim in a converse $\longrightarrow$ **Move:** it suffices to exhibit one bad λ; prefer the value that trivialises the algebraic structure entirely – here $\lambda = 1$ turns "scaled vector equality" into "point equality," a condition independent of P

Layer 3 THE WHY

Why does the forward direction need the homothety-of-a-line idea rather than anything weaker? A homothety centred at P is, by definition, the map under which every displacement from P is multiplied by the same scalar λ; this is precisely why it sends a set closed under a fixed direction (a line not through P) to another set with that same direction – a line parallel to the first. That parallelism is not incidental, it is forced by the definition of the transformation itself, so once you know $L_1 \nparallel L_2$, the image $h(L_1) \parallel L_1$ inherits that non-parallelism with L_2 automatically, and non-parallel lines in the plane meeting at a unique point is itself an axiom of plane geometry (two distinct lines meet in at most one point; failing to be parallel means they are not disjoint, so they meet in exactly one). Nothing about P's specific location or λ's specific value is used beyond $\lambda \neq 0$ – the argument is uniform in both, which is exactly what a "for every λ, for every P" claim demands.

The converse's choice of $\lambda = 1$ is not arbitrary either. The general vector condition $\overrightarrow{PA_2} = \lambda\overrightarrow{PA_1}$ has two independent parameters, A_1's position along L_1 and the scaling λ; for general λ this is a genuinely two-dimensional search (over A_1) tied to a moving target (the homothety of L_1), and whether it succeeds for parallel lines can depend delicately on how far P is from each line. Only at $\lambda = 1$ does the equation stop being about vectors at all and become a statement about a single point lying on both lines – turning a geometric search into the purely set-theoretic fact that $L_1 \cap L_2 = \varnothing$ for distinct parallel lines. Choosing $\lambda = 1$ replaces an argument that would have to track distances with an argument that needs nothing but the definition of "parallel and distinct."

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the quantifier structure "for every $\lambda \neq 0$ and every point P." On seeing it, write down: forward direction needs one construction that works uniformly in λ, P; converse only needs one counterexample (λ, P).
2. That cue suggests representing the vector condition as a transformation of the whole plane – specifically the homothety $x \mapsto P + \lambda(x - P)$ – and asking what it does to the line L_1 as a set, rather than searching point by point.

3. Hesitation occurs right after verifying the easy case $\lambda = 1$ at the intersection point: it is tempting to treat this as "essentially done" since it feels like the crux of the problem. It is reinforced away by explicitly re-reading the quantifier and asking, "does my construction depend on $\lambda = 1$, or does it generalise?"

4. The trap is proving a universally quantified statement by checking one convenient instance and treating it as representative. A second trap sits in the converse: assuming that *no* λ can work for parallel lines, when in fact a specially tuned λ (matched to P's position between the lines) can make the homothety image of L_1 coincide with L_2 exactly – only $\lambda = 1$ is guaranteed to fail regardless of P.

5. Recognition would be faster next time by treating any "for every parameter t, [vector/algebraic condition]" claim as automatically splitting into two different economies: proving it needs a construction valid uniformly in t (often a named transformation – homothety, rotation, reflection – applied to a whole line or curve); disproving it needs only one adversarial value of t, ideally the value that collapses the condition to something structurally simpler.

6. The reusable decision rule: when a geometric claim is quantified over a scaling parameter λ multiplying a vector from a fixed point P, read $x \mapsto P + \lambda(x - P)$ as a homothety and ask what it does to lines (parallelism is preserved, distances scale by $|\lambda|$); to disprove such a claim, test $\lambda = 1$ first, since it is the value that turns "scaled vector equality" into plain point equality, the cheapest and most P-independent condition available.

2017 B2

2017 B2. Suppose that a positive integer N can be expressed as the sum of k consecutive positive integers

$$N = a + (a+1) + (a+2) + \cdots + (a+k-1)$$

for $k = 2017$ but for no other values of $k > 1$. Considering all positive integers N with this property, what is the smallest positive integer a that occurs in any of these expressions?

Fields: Number theory (arithmetic of consecutive sums, factorisation and parity, divisor structure) **Position:** B2 **Answer:** The smallest such a is 16.

Layer 1 THE WHAT — surface analysis

The natural attempt. The algebra is forced immediately: summing an arithmetic progression of k terms starting at a gives $N = ka + \binom{k}{2} = \frac{k(2a+k-1)}{2}$, so $2N = k(2a+k-1)$. Plugging in $k = 2017$: $2N = 2017(2a + 2016)$, so $N = 2017(a + 1008)$ – this much is one line of algebra and gives a formula for N in terms of a. The natural next step, faced with "for no other $k > 1$", is to try to rule out competing values of k directly: fix a candidate $N = 2017(a + 1008)$, and for each $k = 2, 3, 4, \ldots$ ask whether $2N/k - (k-1)$ is even and positive (giving an integer $a' \geq 1$). This can be made to work for small k by brute inequality-checking, and a determined solver might even notice that k cannot exceed roughly $\sqrt{2N}$ for a' to stay positive, bounding the search. But the search is still over an unbounded (or at best large, N-dependent) list of individual k values, checked one at a time with no structural reason to expect the pattern to close off cleanly – there is no guarantee that eliminating $k = 2, \ldots, 100$ says anything about $k = 101$, and the argument would have to be redone for every candidate N. This is workable in principle but is exactly the kind of undirected, per-case elimination that the problem's phrasing ("for no other values of k", over an infinite range) is designed to punish.

The model chain. Return to $2N = k(2a+k-1)$ and look at the two factors together rather than one at a time: their difference is $(2a+k-1) - k = 2a - 1$, which is odd for every integer a. So the two factors k and $2a+k-1$ *always* have opposite parity, for any valid representation. This means representations of N as a sum of $k \geq 2$ consecutive positive integers correspond exactly

to factorisations $2N = mn$ with $m < n$ of opposite parity (taking $k = m$, and $a = (n - m + 1)/2$, which is then automatically a positive integer). In particular, the number of valid $k > 1$ for a given N is governed entirely by how many divisor pairs of $2N$ have opposite parity – a purely multiplicative, finite question about N's odd divisors, not a search over k.

Now specialise: $N = 2017(a + 1008)$ with 2017 prime. Write $M = a + 1008$, so $2N = 2 \cdot 2017 \cdot M$. The representation with $k = 2017$ always exists (it corresponds to the factor pair $(2017, 2M)$, of opposite parity since 2017 is odd and $2M$ is even). For no *other* $k > 1$ to work, no other opposite-parity factor pair of $2N$ with smaller factor > 1 may exist. If M had any odd prime factor p, then p would be an odd divisor of $2N$ distinct from 2017 (unless $p = 2017$ itself, which is allowed and does not create a new pair), pairing with the even cofactor $2N/p$ to give another valid, distinct representation. So M must have no odd prime factor at all: M must be a power of 2. Since $a \geq 1$, $M = a + 1008 \geq 1009$; the smallest power of 2 that is at least 1009 is $1024 = 2^{10}$. Hence $M = 1024$, giving $a = 1024 - 1008 = 16$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the identical formula $2N = k(2a + k - 1)$ in the first line – that part is common ground and takes no real thought. They split at the sentence "but for no other values of $k > 1$." The natural attempt reads this as a statement to be verified k by k; the model chain reads the same sentence as a statement about how many opposite-parity factor pairs $2N$ has, and switches from checking values of k to counting divisors of $2N$.

> *The two factors k and $2a + k - 1$ in $2N = k(2a + k - 1)$ are not independent numbers I get to choose separately – their difference is $2a - 1$, always odd, so they are forced into opposite parity no matter what k or a is. That means every value of k that could possibly work corresponds to an actual factor pair of the fixed number $2N$, so I should stop asking "does $k = 2, 3, 4, \ldots$ work?" and start asking "how many opposite-parity factor pairs does $2N$ have?" – because those are the same question, and the second one is finite and structural.*

> *Since 2017 is prime and already accounts for the factor pair I am required to keep, the only way to create an unwanted* second *pair is through an odd prime factor of M other than 2017 itself; so uniqueness of the representation is not a statement about a directly, it is the single structural requirement that $M = a + 1008$ have no odd prime factor, i.e. that M be a power of 2.*

- ▷ **Cue:** a problem states a sum-of-consecutive-integers count is prescribed exactly ("works for $k = K$ but for no other $k > 1$") $\longrightarrow$ **Move:** rewrite as $2N = k(2a + k - 1)$, note the two factors always have opposite parity, and translate the uniqueness condition into a statement about the odd divisors of $2N$ rather than searching over k
- ▷ **Cue:** a number is required to have no other divisor of a certain kind beyond one already fixed $\longrightarrow$ **Move:** express the number as (fixed prime)$\times M$ and require M to avoid the forbidden divisor type entirely – here, M must be a pure power of 2, i.e. have no odd prime factor

Layer 3 THE WHY

Why does the parity of the two factors matter at all, rather than being a side observation? Because it is the mechanism that converts an additive question (which k let N be written as a sum of k consecutive positive integers) into a multiplicative one (which factorisations of $2N$ exist with a certain parity pattern) – and multiplicative questions about a fixed integer are governed by its prime factorisation, a finite and fully computable object, whereas the additive question as posed has no visible bound on k at all. The identity $(2a + k - 1) - k = 2a - 1$ is forced purely by the shape of an arithmetic series (the difference between "twice the average term" and "the number of terms" is always $2a - 1$, an odd number, regardless of which series you started with); this is not a coincidence of the number 2017, it is a structural fact about every such representation, of every N, for every k. Once that fact is in hand, uniqueness of the

representation stops being a claim about the infinite list of integers $k > 1$ and becomes a claim about the finite list of odd prime factors of $2N$ – specifically, that 2017 must be the *only* one. That is why the problem reduces to $M = a + 1008$ being a power of 2: a power of 2 is, by definition, the unique kind of positive integer with no odd prime factor at all, so it is exactly the condition needed for 2017 to remain the sole source of an odd, non-trivial divisor of N. Minimising a then becomes the entirely mechanical last step of finding the smallest power of 2 at least as large as the forced lower bound 1009.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the algebraic identity $2N = k(2a + k - 1)$ together with the observation that its two factors differ by $2a - 1$, an odd number – this should be written down immediately as "every valid k corresponds to an opposite-parity factor pair of $2N$."

2. That cue suggests translating the whole problem into a divisor-counting statement about $2N$: the number of valid $k > 1$ equals the number of opposite-parity factor pairs of $2N$ with smaller factor greater than 1, which in turn is controlled by the odd divisors of N.

3. Hesitation occurs right after deriving $N = 2017(a + 1008)$, when it is tempting to start testing small values of k by hand to "see the pattern." It is reinforced away by asking, before testing anything, "is there a reason the set of working k is finite and structured, rather than something I need to search?" – which the opposite-parity observation answers immediately.

4. The trap is treating "for no other values of $k > 1$" as an infinite case-check rather than recognising it as a divisor-structure condition on a single fixed number, $2N$. A second, subtler trap is forgetting that the factor pair for $k = 2017$ itself must be preserved – the requirement is not "M has no odd factors at all in an absolute sense" but "$2N$ has no odd divisor *other than* 2017," which is precisely "M has no odd prime factor."

5. Recognition would be faster next time by treating any "exactly one representation with parameter $k = K$" statement in an additive/combinatorial identity as a signal to look for the underlying bijection with a divisor or factorisation set, then count divisors instead of testing parameter values.

6. The reusable decision rule: when a sum-of-consecutive-terms formula produces an identity of the form (total)$\times 2$ =(count)$\times$(twice the average), check the parity relation between the two factors first; it usually forces a bijection between valid parameter values and factor pairs of a fixed number, turning an unbounded case search into a finite question about that number's prime factorisation.

2017 B3

2017 B3. Suppose that $f(x) = \sum_{i=0}^{\infty} c_i x^i$ is a power series for which each coefficient c_i is 0 or 1. Show that if $f(2/3) = 3/2$, then $f(1/2)$ must be irrational.

Fields: Number theory / analysis (rationality and eventually periodic base expansions, power series manipulation) **Position:** B3 **Answer:** $f(1/2)$ is irrational – proved by contradiction: rationality of $f(1/2)$ would force the coefficient sequence (c_i) eventually periodic, and evaluating the resulting closed form at $x = 2/3$ produces a fraction with odd denominator, which cannot equal $3/2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The coefficients are constrained to $\{0, 1\}$ and $f(2/3) = 3/2$ is a single numeric fact, so the obvious first move is to try to squeeze information out of that one equation directly: bound $f(2/3) = \sum c_i (2/3)^i \leq \sum (2/3)^i = 3$ and note it is exactly $3/2$, roughly half the maximum – suggestive but not obviously restrictive, since many different $0/1$ sequences achieve a given partial sum. A more concrete instinct is to try to *construct* explicit sequences (c_i) with $f(2/3) = 3/2$ – for instance a purely periodic pattern such as $c_i = 1$ for i even, $c_i = 0$ for i

odd, which gives $f(2/3) = \sum(2/3)^{2k} = \frac{1}{1-4/9} = \frac{9}{5}$, not $3/2$; adjusting the period and pattern by trial can eventually locate a periodic sequence that does hit $3/2$ exactly, and for that specific sequence one can compute $f(1/2)$ directly and check it is irrational. This is a genuine partial success – it produces *examples* consistent with the claim – but it proves nothing about *every* $0/1$ sequence with $f(2/3) = 3/2$, only the particular ones guessed. The problem asks for a universal statement ("must be irrational"), and guessing periodic examples never touches the general, unconstrained sequences that the claim also has to cover; worse, assuming periodicity as a starting hypothesis rather than deriving it from "$f(1/2)$ is rational" would be circular if used as the whole proof.

The model chain. Argue by contradiction: suppose $f(1/2) = \sum_{i=0}^{\infty} c_i 2^{-i}$ is rational. This sum is exactly a base-2 digit expansion of a real number (shifted by one place relative to the usual convention, since it starts at 2^{-0} rather than 2^{-1}, but the same theorem applies): a real number has an eventually periodic base-2 expansion with digits in $\{0, 1\}$ if and only if it is rational. So there exist integers $n \geq 0$ and $m \geq 1$ such that $c_i = c_{i+m}$ for all $i \geq n$. This turns the infinite, otherwise unconstrained sequence (c_i) into finite data: the first n digits, plus one repeating block of length m.

With that structure, f becomes a genuine rational function of x:

$$f(x) = \sum_{i=0}^{n-1} c_i x^i \; + \; \frac{x^n}{1 - x^m} \sum_{i=0}^{m-1} c_{n+i} x^i.$$

Write $P(x) = \sum_{i=0}^{n-1} c_i x^i$ (a polynomial with $0/1$ coefficients, degree $< n$) and $Q(x) = \sum_{i=0}^{m-1} c_{n+i} x^i$ (also $0/1$ coefficients, degree $< m$). Evaluate at $x = 2/3$:

$$f(2/3) = P(2/3) \; + \; \frac{(2/3)^n}{1 - (2/3)^m} Q(2/3) = \frac{3}{2}.$$

Track denominators only, not exact values. $P(2/3)$, written over a common denominator, has denominator dividing 3^{n-1}: odd. Likewise $Q(2/3)$ has denominator dividing 3^{m-1}: odd. The remaining factor is

$$\frac{(2/3)^n}{1 - (2/3)^m} = \frac{2^n/3^n}{(3^m - 2^m)/3^m} = \frac{2^n\, 3^{m-n}}{3^m - 2^m},$$

and $3^m - 2^m$ is odd, since 3^m is odd and 2^m is even. So every term contributing to $f(2/3)$, once written in lowest terms, has an *odd* denominator: $P(2/3)$'s denominator is a power of 3, and the second term's denominator is (a power of 3) times $(3^m - 2^m)$, both odd. A sum of fractions with odd denominators has an odd denominator (any common denominator built from their product is odd). So $f(2/3) = a/b$ for some integers a, b with b odd. But $f(2/3) = 3/2$, so $2a = 3b$; the left side is even, and the right side is odd $\times$ odd $=$ odd (since b is odd), which is impossible unless both sides are 0 – and they are not. This contradiction shows the assumption that $f(1/2)$ is rational cannot hold, so $f(1/2)$ is irrational.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same raw material: the coefficients are $0/1$, and one value of f is known exactly. They part at the very first move after that: the natural attempt tries to extract information by *choosing* or *guessing* sequences (c_i) (an existence mindset, constructive and example-driven); the model chain instead *assumes* the conclusion is false ($f(1/2)$ rational) and asks what that assumption, alone, forces about the otherwise-free sequence (c_i) – a necessity mindset. The bifurcating line is the move from "what sequences make $f(2/3) = 3/2$ true?" to "what must (c_i) look like if $f(1/2)$ happens to be rational?"

> *The equation $f(1/2) = \sum c_i 2^{-i}$ is not just a numeric value – it* is *a base-2 digit expansion in disguise, and the coefficients c_i are otherwise completely unconstrained $0/1$ choices. The only theorem I know that converts "this number is rational" into a* finite *amount of information about an infinite $0/1$ sequence is the periodicity theorem for base expansions; that is the one lever available here, so if I want to say anything general about (c_i), I have to assume rationality and invoke it, not try to guess sequences.*

> *Once the sequence is periodic, f collapses to a polynomial plus a geometric tail — a rational function of x with denominators built entirely from powers of x and $1 - x^m$. Evaluating a rational function of x at $x = 2/3$ produces a fraction whose denominator is built entirely from powers of 3 and from $3^m - 2^m$; since $2/3$ has 3 in its denominator, every denominator I generate is a product of odd numbers, no matter what n, m or the digits are. That parity fact, not the exact value of $f(2/3)$, is what clashes with $3/2$.*

- ▷ **Cue:** an infinite sum with $0/1$ (or fixed-digit) coefficients is claimed rational at one evaluation point and irrational at another $\longrightarrow$ **Move:** treat the sum as a base-expansion; assume the target value is rational, invoke the eventually-periodic-expansion theorem to reduce the infinite digit sequence to finite periodic data, then rewrite the series as (polynomial) + (geometric tail) and evaluate
- ▷ **Cue:** two evaluation points of a series share the same unknown coefficients, and one point's value is known $\longrightarrow$ **Move:** track only the parity or prime structure of denominators across both evaluations, not exact values – the contradiction is usually an arithmetic-invariant clash (here: odd denominator forced, versus $3/2$'s even denominator), not a size mismatch

Layer 3 THE WHY

Why must "$f(1/2)$ rational" be converted into periodicity, rather than manipulated as a plain equation? Because the coefficients c_i carry infinitely many independent binary degrees of freedom, and no algebraic identity alone can pin down an infinite sequence from a single numerical constraint – there is no elementary way to solve $\sum c_i 2^{-i} = p/q$ for the c_i directly. The periodicity theorem is the unique bridge available because it is not really a fact about power series at all: it is a fact about long division. Dividing p by q produces a sequence of remainders in $\{0, 1, \ldots, q-1\}$, and by the pigeonhole principle some remainder must repeat within q steps, after which the digit pattern must repeat exactly (long division is deterministic given the current remainder). That is the entire content of "rational $\iff$ eventually periodic expansion," and it is what converts an infinite, free sequence into finite data the moment rationality is assumed.

Once periodicity is secured, the specific numbers $2/3$ and $3/2$ finally earn their place in the argument, and this is where the problem's design becomes visible: $x = 2/3$ is chosen so that every rational function built from powers of x and $1 - x^m$ has a denominator built from powers of 3 and from $3^m - 2^m$ – and $3^m - 2^m$ is *always* odd, because 3^m is odd and 2^m is even, an odd-minus-even fact that has nothing to do with the specific digits chosen. That single parity fact is completely general (true for every $m \geq 1$), while $f(2/3) = 3/2$ has an *even* denominator in lowest terms. The clash is therefore forced and unavoidable, independent of which periodic sequence one might have picked – which is exactly why the natural attempt's strategy of hunting for one explicit periodic example was never going to reach a general proof: the contradiction lives in the parity of $3^m - 2^m$ versus the parity of 2, a fact true for *every* eventually periodic sequence at once, not a property of any one sequence found by search.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that $f(1/2) = \sum c_i 2^{-i}$ with $c_i \in \{0, 1\}$ is literally a base-2 digit expansion. On seeing it, write down: "rational value $\iff$ eventually periodic digit sequence" as the one theorem available to constrain the otherwise-free coefficients.

2. That cue suggests a proof by contradiction: assume $f(1/2)$ rational, invoke periodicity to get finite data (n, m), then rewrite $f(x)$ as a polynomial plus a geometric tail and evaluate the resulting closed form at the *other* point, $x = 2/3$.

3. Hesitation occurs right after writing the closed form $P(x) + \frac{x^n}{1-x^m}Q(x)$ and being asked to evaluate it at $2/3$: the nested fractions look intimidating to compute exactly. It is reinforced away by realising the exact value is never needed – only the parity of the denominator, tracked symbolically through powers of 3 and $3^m - 2^m$.

4. The trap is trying to find or assume an explicit periodic sequence and verify the claim for it, which only ever produces examples, not the required universal statement; a second trap is computing $f(2/3)$'s exact numeric value instead of tracking the parity of its denominator, which buries the actually decisive fact (the denominator is always odd) under unnecessary arithmetic.

5. Recognition would be faster next time by treating any power series with digit-valued (0/1, or more generally base-b) coefficients, evaluated at two different rational points where one value's rationality must be linked to the other's, as an automatic trigger for the periodicity theorem, followed by a denominator/parity comparison rather than a value comparison.

6. The reusable decision rule: to prove a sum with digit-valued coefficients is irrational at one point given a known rational value at another, assume rationality at the target point, invoke eventual periodicity to reduce the infinite coefficient sequence to finite data, rewrite the sum in closed form (polynomial plus geometric tail), and compare an arithmetic invariant of the denominators (parity, or which primes divide them) across the two evaluation points rather than comparing numeric values.

2017 B4

2017 B4. Evaluate the sum

$$\sum_{k=0}^{\infty}\left(3\cdot\frac{\ln(4k+2)}{4k+2}-\frac{\ln(4k+3)}{4k+3}-\frac{\ln(4k+4)}{4k+4}-\frac{\ln(4k+5)}{4k+5}\right)$$
$$=3\cdot\frac{\ln 2}{2}-\frac{\ln 3}{3}-\frac{\ln 4}{4}-\frac{\ln 5}{5}+3\cdot\frac{\ln 6}{6}-\frac{\ln 7}{7}$$
$$-\frac{\ln 8}{8}-\frac{\ln 9}{9}+3\cdot\frac{\ln 10}{10}-\cdots.$$

(As usual, $\ln x$ denotes the natural logarithm of x.)

Fields: analysis (series, asymptotics of partial sums) **Position:** B4 – mid-paper; the obstacle is bookkeeping with divergent pieces, not a hard idea **Answer:** $(\ln 2)^2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $\varphi(n) = \frac{\ln n}{n}$. The block has four terms with weights $3, -1, -1, -1$, so the first instinct is to combine each block into a single term and bound it: a Taylor expansion of φ about $4k$ shows each block is $O(\ln k/k^2)$, so the series converges. That settles convergence and nothing else. The next instinct is to find an antiderivative: $\int \frac{\ln x}{x}\,dx = \frac{1}{2}(\ln x)^2$, so perhaps compare each block with an integral. Try it: the block is roughly $3\varphi(4k+2) - \varphi(4k+3) - \varphi(4k+4) - \varphi(4k+5) \approx -\varphi'(4k)\cdot(\text{something})$, and summing φ' terms gives back differences of φ that tend to 0 – a first-order approximation throws away exactly the constant you are asked for. You can push the Taylor expansion further, but every order produces another series of the same shape, and the answer is a constant that no finite expansion delivers. Alternatively one tries to split the series into four separate sums $3\sum\varphi(4k+2) - \sum\varphi(4k+3) - \ldots$, and each one diverges like $\frac{1}{8}(\ln k)^2$. Here most students stop: four infinities, apparently no way to subtract them legitimately.

The model chain. Define the weight $w(n) = 3$ if $n \equiv 2 \pmod 4$ and $w(n) = -1$ otherwise. The K-th partial sum of the series (blocks $k = 0, \dots, K-1$) covers $n = 2, \dots, 4K+1$, and since $\varphi(1) = 0$,

$$S_K = \sum_{n=1}^{4K} w(n)\varphi(n) \; - \; \varphi(4K+1).$$

The last term tends to 0. Now rewrite the weight with indicator functions: $w(n) = -1 + 4\,[2 \mid n] - 4\,[4 \mid n]$ (check on $n \equiv 0,1,2,3 \pmod 4$: $-1, -1, 3, -1$). Put $A(x) = \sum_{n \le x} \varphi(n)$ and $H_x = \sum_{n \le x} \frac{1}{n}$. Because $\ln(2m) = \ln 2 + \ln m$ and $\ln(4m) = 2\ln 2 + \ln m$,

$$\sum_{m \le 2K} \frac{\ln(2m)}{2m} = \tfrac{1}{2}\big(\ln 2\, H_{2K} + A(2K)\big), \qquad \sum_{m \le K} \frac{\ln(4m)}{4m} = \tfrac{1}{4}\big(2\ln 2\, H_K + A(K)\big).$$

Hence, exactly,

$$\sum_{n=1}^{4K} w(n)\varphi(n) = -A(4K) + 2A(2K) - A(K) + 2\ln 2\,\big(H_{2K} - H_K\big). \qquad (\dagger)$$

Two asymptotic facts finish it.

(i) $A(x) = \frac{1}{2}(\ln x)^2 + C + o(1)$ for a constant C. Proof: let $D(n) = A(n) - \int_1^n \varphi(t)\,dt = A(n) - \frac{1}{2}(\ln n)^2$. Since φ is decreasing on $[e, \infty)$, for $n \ge 3$ we have $D(n+1) - D(n) = \varphi(n+1) - \int_n^{n+1} \varphi < 0$, and $\sum_{j=4}^{n} \varphi(j) \ge \int_4^{n+1} \varphi$ gives $D(n) \ge D(3) - \int_3^4 \varphi$. A decreasing sequence bounded below converges.

(ii) $H_{2K} - H_K \to \ln 2$ (the same argument applied to $1/t$, or $H_n = \ln n + \gamma + o(1)$).

Substitute (i) into $(\dagger)$ with $L = \ln K$, $\ell = \ln 2$. The constants cancel ($-C + 2C - C = 0$) and

$$\tfrac{1}{2}\Big[-(L+2\ell)^2 + 2(L+\ell)^2 - L^2\Big] = \tfrac{1}{2}\big[-4\ell^2 + 2\ell^2\big] = -\ell^2.$$

So $S_K \to -\ell^2 + 2\ell \cdot \ell = (\ln 2)^2$. (Numerically, 0.480453…; I checked this against a direct summation to 10^7 blocks and an Euler–Maclaurin evaluation.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree the series converges and that each of the four residue-class sums diverges. They split at the sentence "these four sums each diverge, so I cannot separate them." The natural attempt treats divergence as a stop sign and goes back to estimating blocks. The model separates them anyway, *at a finite cut-off* $4K$, where every manipulation is a finite rearrangement and therefore exact; divergence only matters when $K \to \infty$, and by then the divergent parts are written in a form where they visibly cancel.

> *Four divergent sums are only a problem if I take the limit first. I'll truncate at* $4K$*, rearrange a finite sum freely, and only then ask what grows.*

> *The weights* $3, -1, -1, -1$ *are a function of* n *mod* 4 *that sums to zero over a period – so I should write them as a combination of "every* n*", "even* n*", "multiples of* 4*". Those subsequences are the original sequence rescaled, and* ln *turns rescaling into adding a constant.*

Layer 3 THE WHY

Why is the indicator rewrite the right first line and not a trick? Because the problem gives you a periodic weight, and a periodic weight is a statement about divisibility. The only residue classes that interact well with $\varphi(n) = \ln n / n$ are the ones closed under the operation φ respects. And φ respects multiplication: $\varphi(2m) = \frac{\ln 2}{2m} + \frac{1}{2}\varphi(m)$. That is the defining property of the

logarithm, $\ln(ab) = \ln a + \ln b$, and nothing else is being used. The classes "n even" and "$4 \mid n$" are exactly the images of all of $\mathbb{N}$ under $m \mapsto 2m$ and $m \mapsto 4m$, so each collapses back to A and H at a rescaled cut-off. The class "$n \equiv 3 \pmod 4$" has no such description, which is why the naive four-way split was hopeless: it chose residue classes that the function cannot see.

Second, why do the divergences cancel rather than leave a mess? Because the weight has mean zero over a period, the leading growth $\frac{1}{2}(\ln x)^2$ enters as $-\frac{1}{2}(\ln 4K)^2 + 2 \cdot \frac{1}{2}(\ln 2K)^2 - \frac{1}{2}(\ln K)^2$, a second difference of $(\ln K + j \ln 2)^2$ in j. A second difference of a quadratic is constant: the L^2 and L terms must die and only $-\ell^2$ survives. The unknown constant C dies because the coefficients $-1, 2, -1$ sum to zero. So the answer is forced to be a polynomial in $\ln 2$ before any computation is done; the computation only finds which one.

Kedlaya's first solution defines $a_k = \varphi(k) - \varphi(k+1)$ and regroups. Here is where that comes from: it is the same move in disguise. Writing the block as a combination of consecutive differences makes the series absolutely convergent, so rearranging is legal without a cut-off; the $\ln 2 \sum(-1)^{k+1}/k$ that appears there is my $2 \ln 2(H_{2K} - H_K)$. His second solution, $L(s) = 1 + \zeta(s)(-1 + 2^{2-s} - 2^{2-2s})$, is the same indicator rewrite performed on n^{-s} – the 2^{-s} and 4^{-s} factors are the rescalings – followed by differentiation at $s = 1$. All three solutions are one idea: the weight is a divisibility pattern and ln is a homomorphism.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the periodic weight $3, -1, -1, -1$ summing to 0. On seeing it, write $w(n) = -1 + 4[2 \mid n] - 4[4 \mid n]$ and check it on the four residues.

2. It suggests truncating at a multiple of the period and expressing every piece through one master partial sum $A(x) = \sum_{n \le x} \ln n / n$ evaluated at rescaled cut-offs.

3. Hesitation occurs at "each sum diverges". What reinforces it away: a finite sum can be rearranged at will, so do all algebra at cut-off $4K$ and let $K \to \infty$ only at the end.

4. The trap is the four-way split by residue class mod 4, which is correct but useless, and the block-by-block Taylor expansion, which proves convergence and delivers no constant.

5. Recognise faster by asking of any weighted series: is the weight periodic with mean zero, and does the summand have a clean scaling law under $n \mapsto dn$? If both, the value is a finite combination of rescaled partial sums.

6. Reusable rule: to sum $\sum w(n) f(n)$ with w periodic of mean zero, decompose w into indicators of $d \mid n$, use the scaling law of f to rewrite each piece as the unweighted partial sum at cut-off N/d, and let the known asymptotic expansion of that one partial sum do the cancellation.

▷ **Cue:** a series whose weights repeat with period q and sum to zero over each period $\longrightarrow$ **Move:** rewrite the weight as a combination of indicators $[d \mid n]$, $d \mid q$

▷ **Cue:** several pieces of a convergent series each diverge $\longrightarrow$ **Move:** truncate at a finite N, rearrange exactly, and let the asymptotic expansion of the partial sums cancel the growth

▷ **Cue:** a summand involving $\ln n$ restricted to multiples of d $\longrightarrow$ **Move:** use $\ln(dm) = \ln d + \ln m$ to reduce to the full sum at cut-off N/d plus $\ln d$ times a harmonic sum

2017 B5

2017 B5. A line in the plane of a triangle T is called an *equalizer* if it divides T into two regions having equal area and equal perimeter. Find positive integers $a > b > c$, with a as small as possible, such that there exists a triangle with side lengths a, b, c that has exactly two distinct equalizers.

Fields: geometry, real analysis (convexity, counting roots), elementary number theory **Position:** B5 – hard-slot, but the difficulty is organisation, not technique **Answer:** $(a, b, c) = (9, 8, 7)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct is to put the triangle in coordinates, write a general line $y = mx + k$, compute where it meets the sides, and impose two equations (area, perimeter) on two unknowns (m, k). This is legitimate and immediately ugly: which two sides the line meets depends on (m, k), the perimeter involves square roots of the intersection coordinates, and the system has to be solved separately in each of three regimes. A second instinct is to guess that the equalizers are "special" lines – medians, angle bisectors – and test them. A median bisects area but, for a scalene triangle, not perimeter; an angle bisector generally does neither. The guesses fail, and the student returns to the coordinate system and runs out of time. The deeper problem is that "exactly two" is a *counting* statement, and neither approach produces anything countable.

The model chain. Label $BC = a$, $CA = b$, $AB = c$ with $a > b > c$, and $p = a + b + c$. A line that splits T into two regions of positive area meets the boundary in two points on two different sides, so it cuts off a triangle at one vertex V. Let x, y be the distances from V to the cut points along the two sides at V, of lengths s, t. The cut-off triangle has area $\frac{1}{2}xy \sin V$, the whole has $\frac{1}{2}st \sin V$; equal areas means $xy = \frac{st}{2}$. The piece at V has perimeter $x + y + \ell$ and the other piece $p - x - y + \ell$ (ℓ the chord), so equal perimeters means $x + y = \frac{p}{2}$. Eliminating y, and noting $y \le t$ forces $x \ge s/2$: equalizers at V correspond to roots of

$$F_V(x) = x + \frac{st}{2x} = \frac{p}{2}, \qquad x \in \left[\frac{s}{2}, s\right].$$

Each F_V is strictly convex, with minimum value $\sqrt{2st}$ at $x = \sqrt{st/2}$. Endpoint values are $F_V(s/2) = \frac{s}{2} + t$ and $F_V(s) = s + \frac{t}{2}$.

Vertex B ($s = c$, $t = a$): $F(c/2) - \frac{p}{2} = \frac{a-b}{2} > 0$ and $F(c) - \frac{p}{2} = \frac{c-b}{2} < 0$. A continuous function changing sign has a root; a convex one crossing from above to below on an interval crosses exactly once. One equalizer, always.

Vertex A ($s = c$, $t = b$): $F(c/2) - \frac{p}{2} = \frac{b-a}{2} < 0$ and $F(c) - \frac{p}{2} = \frac{c-a}{2} < 0$. A convex function lies below its chord, so $F < \frac{p}{2}$ throughout. No equalizer.

Vertex C ($s = a$, $t = b$): $F(a/2) - \frac{p}{2} = \frac{b-c}{2} > 0$ and $F(a) - \frac{p}{2} = \frac{a-c}{2} > 0$. The minimiser $\sqrt{ab/2}$ lies in $[a/2, a]$ because $b < 2a$ and $b > a/2$ (the latter from $a < b + c < 2b$). So there are 0, 1 or 2 roots according as $\sqrt{2ab} >, =, < \frac{p}{2}$.

No root sits at an endpoint (all endpoint values differ from $\frac{p}{2}$), so no line is counted in two families. The total is $1 + \{0, 1, 2\}$, and it equals 2 exactly when

$$8ab = (a + b + c)^2.$$

Now minimise a. We need $2ab$ a perfect square, $2 \le b < a$, and $a < b + c < 2b$. For $a \le 9$, $ab \le 72$ and $ab \in \{8, 18, 32, 50, 72\}$ gives $(a, b) \in \{(4, 2), (8, 4), (6, 3), (9, 2), (9, 8)\}$; only $(9, 8)$ has $a < 2b$. Then $p = \sqrt{576} = 24$, $c = 7$, and $9 < 8 + 7$. So $(9, 8, 7)$. (I confirmed by exhaustive search over $a \le 12$ that $(9, 8, 7)$ is the only integer solution with $a \le 12$, and numerically that the vertex-C equation for $(9, 8, 7)$ has minimum exactly $p/2$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start by describing a line. They split at the choice of parameters. The natural attempt parametrises the line itself (m, k); the model parametrises the *cut-off triangle* by its two legs (x, y) at the vertex it contains. With that choice, both conditions become polynomial ($xy = st/2$, $x + y = p/2$), the three regimes become three instances of one function, and "how many equalizers" becomes "how many roots of $x + K/x = p/2$ on an interval".

> *A line through a triangle chops off a corner. I'll name the line by how far along the two sides of that corner it cuts – then area is a product and perimeter is a sum, and I get $xy = \text{const}$, $x + y = \text{const}$.*

> *They ask for* exactly *two. Each corner gives a convex equation with* 0, 1 *or* 2 *roots. An exact count that isn't the generic one has to come from a tangency, so the answer is an equation* $\min F = p/2$, *not an inequality.*

Layer 3 THE WHY

Why parametrise by the corner? Because both conditions in the definition of equalizer are statements about lengths along the sides. Perimeter is a sum of boundary lengths; the chord is shared by both pieces and cancels – that is the whole content of "equal perimeter". What is left, $x + y = p/2$, is linear in the legs. Area of a triangle with a fixed angle is $\frac{1}{2}xy \sin V$, the product of the legs. So in the coordinates (x, y) the problem is "a rectangle's side lengths with given sum and given product", which is a quadratic, $x^2 - \frac{p}{2}x + \frac{st}{2} = 0$. The line's slope and intercept are the wrong coordinates because the problem never mentions them.

Why convexity? Once each family is a function $x + K/x$ on an interval, the question "how many roots" is answered by endpoint signs plus shape, and $x + K/x$ is the textbook convex function (second derivative $2K/x^3 > 0$). Knowing that the answer comes from convexity and endpoints tells you what to compute: six endpoint values, all of the form $\frac{\text{side}-\text{side}}{2}$, whose signs are read straight off $a > b > c$. The ordering hypothesis $a > b > c$ is in the problem precisely so that these signs are determined.

Why tangency? The generic counts are $1 + 0$ and $1 + 2$; the number 2 sits on the boundary between two open regions of shape space. Whenever a problem asks for a count that separates two generic counts, the configuration is degenerate, and degeneracy of a convex equation means its minimum equals the target. That converts "exactly two" into the Diophantine condition $8ab = p^2$, and the minimisation is then a finite check.

Kedlaya's solution follows this route and states the three endpoint comparisons tersely; the addition here is the reason the parametrisation is forced and the reason the answer must be an equality. A cross-check worth knowing: for any line cutting off the corner at B, splitting into triangles from the incentre I gives area of the corner piece $= \frac{r}{2}(x + y) \pm \frac{1}{2}\ell\, d$, where r is the inradius and d the distance from I to the line. Equal area and equal perimeter together force $d = 0$: every equalizer passes through the incentre. That is why there are so few of them.

Layer $+\alpha$ BACKTESTING

1. The first cue is "divides T into two regions": a line through a triangle cuts off a corner. Write down x, y (legs at that corner) and the two conditions $xy = st/2$, $x + y = p/2$.

2. It suggests reducing each corner to a one-variable equation $x + K/x = p/2$ on the interval $[s/2, s]$, and counting roots by convexity and endpoint signs.

3. Hesitation: "which sides does the line meet?" This dissolves once you treat the three corners as three instances of the same computation and check that endpoint cases are never roots.

4. The trap is coordinates with a general line, or testing medians and angle bisectors. Also: forgetting the range restriction $x \geq s/2$ (from $y \leq t$), which is what makes the endpoint analysis work.

5. Recognise faster by reading "exactly n" as a boundary case between two generic counts and looking for the tangency condition at once.

6. Reusable rule: when an object must satisfy one sum condition and one product condition, parametrise by the two quantities whose sum and product appear; count solutions of the resulting convex equation from endpoint signs, and a non-generic count forces the minimum to equal the target.

▷ **Cue:** a line splitting a triangle's area and perimeter $\longrightarrow$ **Move:** parametrise by the two legs x, y at the cut-off corner: area gives xy, perimeter gives $x + y$

▷ **Cue:** "exactly two" (or exactly n) solutions where generic counts are $n-1$ and $n+1$ $\longrightarrow$ **Move:** impose tangency: minimum of the convex function equals the target
▷ **Cue:** counting roots of $x + K/x = \text{const}$ on an interval $\longrightarrow$ **Move:** evaluate at the endpoints, use convexity; locate the minimiser $\sqrt{K}$

2017 B6

2017 B6. Find the number of ordered 64-tuples $(x_0, x_1, \ldots, x_{63})$ such that $x_0, x_1, \ldots, x_{63}$ are distinct elements of $\{1, 2, \ldots, 2017\}$ and

$$x_0 + x_1 + 2x_2 + 3x_3 + \cdots + 63x_{63}$$

is divisible by 2017.

Fields: combinatorics, number theory (linear equations over $\mathbb{F}_p$, group actions, recursion) **Position:** B6 – genuinely hard; the main term is a guess anyone can make, the correction term is the problem **Answer:** $\frac{2016!}{1953!} - 63! \cdot 2016$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Since 2017 is prime (trial division by primes up to 43), identify $\{1, \ldots, 2017\}$ with $\mathbb{F}_p$, $p = 2017$. There are $p(p-1)\cdots(p-63) = \frac{2017!}{1953!}$ tuples of distinct elements, and "one in p of them should have the sum $\equiv 0$", giving $\frac{2016!}{1953!}$. That heuristic is the right main term and the wrong answer. To make it rigorous the natural move is: choose $x_1, \ldots, x_{63}$ distinct ($\frac{2017!}{1954!}$ ways), and then x_0 is forced, $x_0 \equiv -(x_1 + 2x_2 + \cdots + 63x_{63})$. Subtract the cases where the forced x_0 collides with some x_j. But $x_0 = x_j$ means $(1 + c_j)x_j + \sum_{i \neq 0,j} c_i x_i \equiv 0$ – a new equation of the same type with two coefficients merged, whose own count needs its own collision correction, and so on down every way of merging variables. This is inclusion–exclusion over set partitions of $\{0, \ldots, 63\}$. It can be completed (Kedlaya's first solution does so via Möbius coefficients μ_π), but a student who starts here in the exam is building the lattice of partitions from scratch, and most hit the wall at the second level of merging.

The model chain. Write the form as $L(x) = \sum_{i=0}^{63} c_i x_i$ with $c_0 = 1$, $c_i = i$ for $i \geq 1$. The coefficients sum to $1 + \frac{63 \cdot 64}{2} = 2017 \equiv 0 \pmod{p}$.

Step 1: translation. For $t \in \mathbb{F}_p$, the map $x \mapsto x + t(1, \ldots, 1)$ preserves distinctness and satisfies $L(x + t\mathbf{1}) = L(x) + t\sum c_i = L(x)$. It acts on the set of distinct tuples with orbits of size p, and each orbit contains exactly one tuple with $x_0 = 0$ (take $t = -x_0$). Hence

$$\#\{\text{solutions}\} = p \cdot \#\{\text{solutions with } x_0 = 0\}.$$

With $x_0 = 0$, the remaining $x_1, \ldots, x_{63}$ are distinct and *nonzero*, and satisfy $\sum_{i=1}^{63} c_i x_i \equiv 0$, where now $\sum_{i \geq 1} c_i = 2016 \not\equiv 0$.

Step 2: a recursion. For $k \geq 1$ and positive integers $a_1, \ldots, a_k$ with $A = a_1 + \cdots + a_k < p$, let $f(k)$ be the number of $(x_1, \ldots, x_k)$, pairwise distinct and nonzero in $\mathbb{F}_p$, with $\sum a_i x_i \equiv 0$. Claim: $f(1) = 0$ and

$$f(k) = \frac{(p-1)!}{(p-k)!} - k\, f(k-1) \qquad (k \geq 2),$$

independently of the a_i. Proof by induction. $f(1) = 0$ since $p \nmid a_1$. For $k \geq 2$, translate again: among the $\frac{p!}{(p-k)!}$ distinct k-tuples, orbits have size p and $\sum a_i(x_i + t) = \sum a_i x_i + At$ with A invertible, so exactly one member of each orbit is a solution: $\frac{(p-1)!}{(p-k)!}$ solutions. Now remove those with a zero entry. At most one entry is zero; if $x_j = 0$, the other $k-1$ entries are distinct,

nonzero, and solve an equation whose coefficients sum to $A - a_j \in (0, p)$. By induction there are $f(k-1)$ of these for each j.

Step 3: solve the recursion. Put $g(k) = \frac{p\,f(k)}{k!}$. Multiplying the recursion by $p/k!$ gives $g(k) = \binom{p}{k} - g(k-1)$, $g(1) = 0$. Using $\binom{p}{k} = \binom{p-1}{k} + \binom{p-1}{k-1}$, the quantity $g(k) - \binom{p-1}{k}$ changes sign at each step: $g(k) - \binom{p-1}{k} = -\big(g(k-1) - \binom{p-1}{k-1}\big)$. Starting from $g(1) - \binom{p-1}{1} = -(p-1)$,

$$g(k) = \binom{p-1}{k} + (-1)^k (p-1).$$

Step 4: assemble. The answer is $p\,f(63) = 63!\,g(63) = 63!\left(\binom{2016}{63} - 2016\right) = \frac{2016!}{1953!} - 63! \cdot 2016$.

I verified the general formula $(n-1)!\big(\binom{p-1}{n-1} + (-1)^{n-1}(p-1)\big)$ by exhaustive enumeration in the two smaller cases of the same shape ($1 + \binom{n}{2} = p$ prime): $p = 7$, $n = 4$ gives 84, and $p = 11$, $n = 5$ gives 5280; both match. The sign $(-1)^{n-1}$ matters – it is negative for $n = 64$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths identify the answer as roughly $\frac{1}{p}$ of all distinct tuples. They split at the question "what breaks the $1/p$ heuristic?". The natural attempt fixes 63 variables, solves for the last, and corrects collisions one at a time. The model looks at the coefficients first, finds they sum to 0 mod p, and uses the resulting symmetry to kill one variable exactly, leaving an equation whose coefficients do *not* sum to 0 – where the $1/p$ heuristic becomes an exact orbit count.

> *The coefficients add up to* $1 + 1 + 2 + \cdots + 63 = 2017$. *That is the modulus. So shifting every* x_i *by the same amount doesn't change the sum mod* 2017 *– I can slide* x_0 *to* 0 *for free and multiply by* 2017 *at the end.*

> *Once* $x_0 = 0$ *the other coefficients sum to* 2016, *which is invertible, so now a shift* does *move the sum, through every residue exactly once. That's an exact count – except that the shift can land a variable on* 0, *and those bad tuples are the same problem one size smaller.*

Layer 3 THE WHY

Why look at the sum of the coefficients at all? Because the heuristic "1 in p" has a precise meaning: it holds exactly whenever some group acts on your set freely and moves the value of L through all residues equally. The cheapest group preserving "distinct" is translation, $x \mapsto x + t\mathbf{1}$, and translation changes L by $t\sum c_i$. So the single number $\sum c_i$ mod p decides everything: nonzero, and the heuristic is exact; zero, and translation is a symmetry of L that can be quotiented out instead. The problem-setter chose coefficients $1, 1, 2, \ldots, 63$ – an odd-looking list with 1 repeated – precisely so that the sum is $1 + \binom{64}{2} = 2017$. When a competition list looks irregular, its sum or product is usually the reason; that duplicated 1 is the cue, not decoration.

Why does the correction have the shape $(-1)^k(p-1)$? Because the recursion $g(k) = \binom{p}{k} - g(k-1)$ is an alternating sum, and alternating sums of binomial coefficients telescope via Pascal's rule. The "exact orbit count minus the tuples that hit 0" structure is inclusion–exclusion done one variable at a time: excluding the value 0 from k slots, where at most one slot can take it. Distinctness is what makes the collision set small (at most one zero), and small collision sets are what make a one-step recursion possible where Kedlaya's first solution needs the full partition lattice. His second solution (credited to AoPS user ABCDE) is this route; the translation reduction there is stated in a single sentence, and the reason it exists is the coefficient sum.

Last, why is the main term $\frac{2016!}{1953!}$ exactly the naive guess? Because $63!\binom{2016}{63} = \frac{2016!}{1953!}$ is the "$1/p$ of all distinct tuples" count, and the whole correction $-63! \cdot 2016$ is the price of the one degenerate orbit type the translation symmetry creates. The answer tells you its own proof.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the coefficient list $1, 1, 2, \ldots, 63$ with its repeated 1. Write down its sum: 2017, the modulus.

2. It suggests translation invariance of L and a reduction to $x_0 = 0$, followed by an orbit-counting recursion on the number of variables.

3. Hesitation occurs when the orbit count for the reduced problem must exclude tuples with a zero entry. Reinforce: those tuples are an instance of the same problem with one fewer variable, and distinctness allows at most one zero, so the correction is a single term $k\, f(k-1)$.

4. The trap is the $1/p$ heuristic: $\frac{2016!}{1953!}$ looks like a finished answer. Also the sign: $(-1)^{63}$ is negative.

5. Recognise faster by computing the sum of the coefficients of any linear condition mod p on distinct values before doing anything else; zero means a translation symmetry, nonzero means an exact equidistribution.

6. Reusable rule: to count distinct tuples satisfying a linear congruence mod a prime, find a group action (translation, scaling) that preserves distinctness and track how it moves the linear form; quotient it out when it fixes the form, use it to equidistribute when it moves the form, and handle the fixed-value exclusions by a recursion in the number of variables.

▷ **Cue:** a linear congruence mod p whose coefficients sum to a multiple of p $\longrightarrow$ **Move:** the form is translation-invariant: fix one variable to 0 and multiply by p

▷ **Cue:** a linear congruence mod p whose coefficients sum to a nonzero residue, over distinct tuples $\longrightarrow$ **Move:** translation orbits each contain exactly one solution: count is (number of tuples)$/p$

▷ **Cue:** a count that excludes a single forbidden value from k distinct slots $\longrightarrow$ **Move:** at most one slot can hit it; subtract k times the $(k-1)$-variable count and recurse