DAWN PUTNAM SESSION

Putnam 2004 — Papers A and B

Putnam 120 Archive

Contents

Putnam 2004 — Paper A **5**
2004 A1 5
2004 A2 7
2004 A3 9
2004 A4 12
2004 A5 15
2004 A6 18

Putnam 2004 — Paper B **21**
2004 B1 21
2004 B2 23
2004 B3 25
2004 B4 27
2004 B5 30
2004 B6 33

Putnam 2004 — Paper A

2004 A1

2004 A1. Basketball star Shanille O'Keal's team statistician keeps track of the number, $S(N)$, of successful free throws she has made in her first N attempts of the season. Early in the season, $S(N)$ was less than 80% of N, but by the end of the season, $S(N)$ was more than 80% of N. Was there necessarily a moment in between when $S(N)$ was exactly 80% of N?

Fields: elementary number theory, discrete analogue of the intermediate value theorem **Position:** A1 — the opener **Answer:** Yes. Define $D(N) = 5S(N) - 4N$; then D starts negative, ends positive, and can only step up by $+1$ or down by -4, which forces it through 0.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct is the continuous intermediate value theorem: a function that is negative somewhere and positive somewhere else, moving smoothly, must cross zero. $S(N)/N$ is not smooth — it is a step function that jumps once per free throw — so the theorem does not apply directly, and the honest first move is to ask exactly how big those jumps can be.

Suppose, to find out, that the claim is false: there is an N with $S(N) < 0.8N$ and the very next attempt already has $S(N+1) > 0.8(N+1)$, skipping over 80% entirely. Write $m = S(N)$, so $m/N < 4/5$ and $(m+1)/(N+1) > 4/5$ (the shot at attempt $N+1$ must have gone in, since S can only increase or stay the same, and it has to have increased to climb from below the line to above it). Clearing denominators in the first inequality gives $5m < 4N$. Clearing the second gives $5(m+1) > 4(N+1)$, i.e. $5m > 4N - 1$, i.e. $5m \geq 4N$ (as both sides are integers, "$> 4N - 1$" means "$\geq 4N$"). Now the two bounds read $5m < 4N$ and $5m \geq 4N$ simultaneously — a direct contradiction, no further massaging needed. This works, but it took two separate roundings and a moment's checking that each inequality was strict in the direction claimed before the contradiction became visible; the risk of an off-by-one error going the wrong way is real.

The model chain. Track the single integer quantity $D(N) = 5S(N) - 4N$ instead of the fraction $S(N)/N$. Since $S(N)$ increases by 0 or 1 on each attempt, $D(N+1) - D(N)$ is either

$$5(0) - 4 = -4 \quad \text{(missed shot)} \qquad \text{or} \qquad 5(1) - 4 = 1 \quad \text{(made shot)}.$$

D starts negative (given: $S(N) < 0.8N$ early means $D(N) < 0$) and is eventually positive ($D(N) > 0$ late in the season). Let N^* be the first attempt count at which $D(N^*) > 0$; this exists because D is eventually positive, and $N^* \geq 2$ because D starts negative. Then $D(N^* - 1) \leq 0$ by minimality of N^*, and $D(N^*) - D(N^* - 1)$ is either -4 or 1. It cannot be -4: that would force $D(N^*) \leq D(N^* - 1) - 4 \leq -4$, contradicting $D(N^*) > 0$. So the step was $+1$, meaning $D(N^*) = D(N^* - 1) + 1$. Since $D(N^*) \geq 1$ (it is a positive integer), $D(N^* - 1) = D(N^*) - 1 \geq 0$. Combined with $D(N^* - 1) \leq 0$, this forces $D(N^* - 1) = 0$, i.e. $5S(N^* - 1) = 4(N^* - 1)$, i.e. $S(N^* - 1)/(N^* - 1) = 4/5$ exactly. The answer is yes.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reduce the problem to the same integer quantity in the end — the contradiction in the natural attempt is $5m < 4N \leq 5m$, and the model's step sizes are exactly -4 and $+1$, the same two numbers hiding inside the natural attempt's cleared-denominator inequalities. The chains split on *where* the integer $D(N) = 5S(N) - 4N$ is introduced. The natural attempt introduces it implicitly, at the last possible moment, as a byproduct of clearing denominators inside a proof by contradiction about one specific pair of consecutive N's. The model introduces it first, as the object to be tracked across the *entire* sequence, before any contradiction is attempted.

> *The problem keeps talking about $S(N)/N$ versus $4/5$, but a fraction is an awkward thing to track step by step because every comparison needs a common denominator. If I multiply through by 5 up front and work with $5S(N) - 4N$ instead, that quantity is an integer, and I can ask a much simpler question about it: by how much does it change when one free throw is attempted? That question has exactly two possible answers, -4 or $+1$, and once I know that, I don't need to set up a contradiction about a specific bad pair of consecutive attempts at all — I can just watch $D(N)$ walk from negative to positive and ask what it is forced to do at the moment it first turns positive.*

Layer 3 THE WHY

Why does clearing the denominator *before* setting up any contradiction beat clearing it *during* the contradiction argument? Because the quantity $S(N)/N$ compared against $4/5$ is really asking about the sign of $S(N)/N - 4/5$, and the only reason that comparison feels fraction-bound is that $4/5$ was written as a fraction. Multiply the whole inequality by the positive integer $5N$ and the comparison becomes $5S(N)$ versus $4N$ — two integers, and a comparison between two integers is a statement about an integer difference, which is the object that is actually well-behaved under the "add one free throw" operation. This is not a trick imported from outside the problem: the problem itself only ever changes $S(N)$ by 0 or 1 per step, so the natural unit to reason in is "change per step", and $D(N) = 5S(N) - 4N$ is the smallest integer rescaling of $S(N)/N - 4/5$ for which "change per step" has a clean, small, and exhaustive answer.

Once that reformulation is made, the rest is forced. An integer-valued function that changes by $+1$ or -4 per step and is eventually positive after starting negative cannot leap over 0: a leap over 0 from below could only happen via a $+1$ step landing on a positive value from a negative one only if the value beforehand was -1 or higher already (which is not a leap), or it must pass through 0 on the way up. The only way to skip 0 entirely is a jump of size ≥ 2 while ascending, but the only ascending jump available is $+1$. So the specific numbers in the problem — $80\% = 4/5$, and free throws worth 0 or 1 — are exactly what make the ascending step size equal to 1, which is why the crossing is guaranteed. The remark in Kedlaya's solution, that the same argument works for any $(n-1)/n$ but no other target fraction, is this fact seen from the other side: for a target fraction p/q in lowest terms with $q \neq p + 1$... i.e. $q - p \geq 2$, clearing denominators gives an ascending step of size $q - p \geq 2$ on a made shot, which *can* leap over 0, so the guarantee fails. $4/5$ works only because $5 - 4 = 1$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a percentage target (80%) being compared against a running integer count $S(N)$ that only ever increases by 0 or 1. On seeing it, write down: rescale by the denominator so the comparison is between two integers, not a fraction and a decimal.

2. That cue suggests tracking the integer difference $D(N) = 5S(N) - 4N$ as a function of N and asking what values $D(N+1) - D(N)$ can take — a discrete step-size analysis, not a continuity argument and not (at first) a proof by contradiction.

3. Hesitation occurs right after noticing $S(N)/N$ is not continuous: the pull is to conclude the intermediate value theorem simply fails and abandon the "crossing" intuition altogether.

What reinforces the correct move is asking what discrete replacement for continuity is available — namely, bounding how large each step can be, which is the integer analogue of a bound on the derivative.

4. The trap is treating the contradiction found in the natural attempt (from one specific bad pair $N, N+1$) as the whole story and stopping there, without noticing that the same computation — step size $+1$ or -4 — explains *why* a crossing must exist, for every sequence with this property, not just the one under contradiction.

5. Recognise it faster by treating any "percentage crosses a threshold as a running count updates by fixed increments" problem as an instruction to clear denominators immediately and ask for the resulting integer's possible step sizes, before reaching for continuity language.

6. Reusable rule: when a rational threshold p/q (in lowest terms) is compared against a ratio that updates by adding 0 or 1 to the numerator each step, the rescaled integer quantity $q \cdot (\text{numerator}) - p \cdot (\text{denominator})$ changes by $-p$ (no increment) or $q - p$ (increment) each step. The threshold is guaranteed to be hit exactly whenever $q - p = 1$; when $q - p \geq 2$ it can be skipped, and a counterexample should be sought instead of a proof.

▷ **Cue:** a running percentage $S(N)/N$ is compared against a fixed target fraction p/q while $S(N)$ increases by 0 or 1 per step $\longrightarrow$ **Move:** clear denominators to the integer $D(N) = qS(N) - pN$ and classify its two possible step sizes, $-p$ and $q - p$

▷ **Cue:** an ascending integer sequence changes only by steps drawn from a small fixed set $\longrightarrow$ **Move:** it cannot skip 0 while ascending unless some ascending step has size ≥ 2; check the step-size set before assuming a crossing is guaranteed

2004 A2

2004 A2. For $i = 1, 2$ let T_i be a triangle with side lengths a_i, b_i, c_i, and area A_i. Suppose that $a_1 \leq a_2, b_1 \leq b_2, c_1 \leq c_2$, and that T_2 is an acute triangle. Does it follow that $A_1 \leq A_2$?

Fields: Euclidean geometry, the Law of Sines, monotonicity arguments **Position:** A2 — second problem, first genuine obstacle **Answer:** Yes. Some angle of T_1 is at most the corresponding angle of T_2 (pigeonhole on the angle sum); since T_2 is acute, that lets $A_1 = \frac{1}{2} b_1 c_1 \sin(\text{angle}) \leq \frac{1}{2} b_2 c_2 \sin(\text{angle}) = A_2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Area from three side lengths means Heron's formula, so write

$$16A_i^2 = (a_i + b_i + c_i)(-a_i + b_i + c_i)(a_i - b_i + c_i)(a_i + b_i - c_i),$$

and try to show the right-hand side is at least as large for (a_2, b_2, c_2) as for (a_1, b_1, c_1), given $a_1 \leq a_2$ etc. The first factor $a + b + c$ is clearly monotone: increasing any side increases it. But the other three factors are each a side minus the other two, and increasing a side can only help one of the three "minus" factors while doing nothing (or being irrelevant) to the other two containing that side positively or negatively depending on position. Concretely, take the factor $-a_i + b_i + c_i$: increasing a_i while holding b_i, c_i fixed *decreases* this factor. So termwise monotonicity in (a, b, c) fails for Heron's formula in general — it is entirely possible to increase every side of a triangle and shrink one of these bracket terms.

The rescue attempt is to note that the four factors, multiplied together, are related to a product of sin of half-angles or similar, and try to control the product rather than each factor separately. This can be pushed through with enough algebraic stamina (bounding the product of the three "acute-adjacent" factors together, using that T_2's acuteness caps how negative any partial expression can get), but it requires tracking three coupled inequalities simultaneously and the acuteness hypothesis has to be smuggled in at the level of a factored quartic, which is not where it naturally lives.

The model chain. Every triangle's three interior angles sum to π. If *every* angle of T_1 exceeded the corresponding angle of T_2 (matching up $a_i \leftrightarrow$ its opposite angle, and similarly for b_i, c_i), the three inequalities would sum to $\pi > \pi$ — absurd. So at least one angle of T_1 is at most the corresponding angle of T_2. Say, without loss of generality, this is the angle P_1 opposite a_1, against P_2 opposite a_2: $P_1 \leq P_2$.

Since T_2 is acute, $P_2 < \pi/2$. Combined with $P_1 \leq P_2$, this puts P_1 in $[0, \pi/2)$ too, where sin is increasing. So $\sin P_1 \leq \sin P_2$.

Now use the area formula built from the two sides adjacent to the chosen angle: $A_1 = \frac{1}{2} b_1 c_1 \sin P_1$ and $A_2 = \frac{1}{2} b_2 c_2 \sin P_2$. Every one of the three factors compares favourably: $b_1 \leq b_2$, $c_1 \leq c_2$, $\sin P_1 \leq \sin P_2$, and all six quantities are positive, so the product of the smaller ones is at most the product of the larger ones:

$$A_1 = \tfrac{1}{2} b_1 c_1 \sin P_1 \;\leq\; \tfrac{1}{2} b_2 c_2 \sin P_2 = A_2.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains open by reaching for the standard formula that turns three side lengths into an area. They split at the choice of *which* area formula to reach for. The natural attempt reaches for Heron's formula, which expresses A purely in terms of (a, b, c) with no angle in sight — and pays for that by needing to compare a product of four differently-signed linear combinations of the sides, only one of which is obviously monotone. The model reaches for $A = \frac{1}{2}(\text{side})(\text{side})\sin(\text{included angle})$, which keeps an angle explicitly in the formula, and it is precisely that angle where the acuteness hypothesis has somewhere natural to attach.

> *Heron's formula throws away the angles entirely, but the hypothesis I've been given — T_2 is acute — is a statement about angles, not about linear combinations of side lengths. If the only tool I have with an angle already built into it is $A = \frac{1}{2} bc \sin(\text{angle})$, I should use that formula instead of Heron's, even though it looks less symmetric, because acuteness will then plug in exactly where I need* sin *to be increasing, instead of getting buried inside a product of four bracketed terms where I can't see it act.*

> *I have three side comparisons but I don't know which pair of corresponding* angles *compares favourably. Rather than trying to prove all three angle comparisons go the right way (which is false in general — a triangle can have all larger sides yet a smaller angle at one particular vertex is impossible, but I haven't shown that), I only need* one *of the three to go my way, because the area formula only needs one angle. And the angle sum being fixed at π for both triangles is exactly the fact that forces at least one of the three comparisons to hold: they cannot all fail simultaneously.*

Layer 3 THE WHY

Why does the pigeonhole argument on angle sums succeed where a direct side-by-side comparison of all three angles would not even be true in general? Because the problem supplies three independent inequalities on the sides but the three angles of a triangle are not independent — they are constrained to sum to the fixed constant π. A fixed sum is exactly the structure pigeonhole needs: if three quantities summing to π were each to exceed three other quantities also summing to π, the sums themselves would have to satisfy a strict inequality, and $\pi > \pi$ is false. This is not a geometric fact about triangles specifically; it is the general principle that a constraint of fixed total forbids all parts from moving the same direction at once, and it works here because — and only because — the problem statement handed over a shared, fixed sum (π) for both triangles' angle sets. Comparing three unconstrained quantities termwise never comes with this kind of guarantee, which is exactly the wall the Heron's-formula route runs into: the three bracket factors $-a+b+c$, $a-b+c$, $a+b-c$ have no shared fixed total, so nothing stops one of them shrinking while the sides that build it grow.

Once one favourable angle comparison $P_1 \leq P_2$ is secured, acuteness of T_2 is what converts it into a sin comparison: sin is increasing exactly on $[0, \pi/2)$, so the comparison $P_1 \leq P_2$ only survives being fed into sin if P_2 (the larger of the pair) stays inside that increasing interval. Without the acuteness hypothesis, P_2 could sit in $(\pi/2, \pi)$, where sin is decreasing, and $P_1 \leq P_2$ would give no information about $\sin P_1$ versus $\sin P_2$ at all. This is exactly why the problem specifies T_2 acute and not T_1: acuteness needs to hold for whichever triangle contributes the *larger* angle in the pigeonholed pair, and the pigeonhole argument, run this way round, always produces T_2's angle as the larger one.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a hypothesis about *angles* (acuteness) attached to a claim about *sides and area*. On seeing it, write down: use an area formula that keeps an angle visible, not one (like Heron's) that eliminates angles entirely, since a hypothesis about angles needs an angle to act on.

2. That cue suggests $A = \frac{1}{2}(\text{side})(\text{side})\sin(\text{included angle})$, and immediately raises the question of which angle to use — which in turn suggests exploiting the fixed angle sum π via pigeonhole to guarantee at least one favourable angle comparison exists.

3. Hesitation occurs when Heron's formula is the first (and very natural) reflex for "area from three sides", and the algebra almost works after enough manipulation, so there is a pull to keep pushing it through rather than abandon a partially-successful computation. What reinforces the switch is noticing explicitly that the acuteness hypothesis has not yet been used anywhere in the Heron's-formula attempt — an unused hypothesis after real effort is a sign the wrong formula was chosen.

4. The trap is assuming all three angle comparisons $P_1 \leq P_2$, $Q_1 \leq Q_2$, $R_1 \leq R_2$ must hold simultaneously because all three side comparisons do — they need not, and the proof only asserts (and only needs) at least one.

5. Recognise it faster by treating "prove an inequality about triangles given an angle-type hypothesis (acute, right, obtuse)" as an instruction to route through an angle-carrying formula (Law of Sines, Law of Cosines) rather than a side-only formula (Heron's), before any algebra starts.

6. Reusable rule: when several quantities summing to a fixed total (angles of a triangle, probabilities, proportions) are each claimed to compare unfavourably against a matched set with the same fixed total, pigeonhole guarantees at least one comparison must actually go the other way — use that guaranteed favourable instance rather than attempting to prove all comparisons hold.

▷ **Cue:** an inequality hypothesis is stated in terms of angles (acute/right/obtuse) but the target inequality is stated in terms of sides and area $\longrightarrow$ **Move:** choose a formula that keeps an angle explicit (Law of Sines area $\frac{1}{2}ab\sin C$) instead of a side-only formula (Heron's), so the angle hypothesis has something to act on

▷ **Cue:** several quantities with a shared fixed sum are each claimed to compare unfavourably against a matched set with the same fixed sum $\longrightarrow$ **Move:** apply pigeonhole to the sum: not all comparisons can go the same way, so at least one favourable instance is guaranteed

2004 A3

2004 A3. Define a sequence $\{u_n\}_{n=0}^{\infty}$ by $u_0 = u_1 = u_2 = 1$, and thereafter by the condition that

$$\det \begin{pmatrix} u_n & u_{n+1} \\ u_{n+2} & u_{n+3} \end{pmatrix} = n!$$

for all $n \geq 0$. Show that u_n is an integer for all n. (By convention, $0! = 1$.)

Fields: combinatorics, recurrences, double factorials **Position:** A3 — third problem, first that demands a closed form **Answer:** Yes: $u_n = (n-1)!!$, the double factorial (product of all positive integers of the same parity as $n-1$, down to 1 or 2, with the convention $(-1)!! = 1$), which is manifestly an integer and can be checked to satisfy both the initial conditions and the determinant recurrence.

Layer 1 THE WHAT — surface analysis

The natural attempt. The determinant condition expands to $u_n u_{n+3} - u_{n+1}u_{n+2} = n!$, which rearranges to

$$u_{n+3} = \frac{n! + u_{n+1}u_{n+2}}{u_n}.$$

This defines u_{n+3} from the three previous terms, provided $u_n \neq 0$, and the obvious plan is induction: assume $u_0, \ldots, u_{n+2}$ are positive integers, and try to show u_n divides $n! + u_{n+1}u_{n+2}$ exactly. That divisibility is not visible from the recurrence alone — nothing about the formula says why u_n should divide that particular combination, and trying to prove it by strong induction on the divisibility statement itself, without more structure, goes in a circle: the inductive hypothesis "u_n divides $(\cdot)$" doesn't obviously propagate to "u_{n+3} divides $(\cdot\cdot)$" without already knowing what u_{n+3} is.

The fallback is to compute: $u_0 = u_1 = u_2 = 1$. From $n = 0$: $u_0u_3 - u_1u_2 = 0! = 1$, so $u_3 = 2$. From $n = 1$: $u_1u_4 - u_2u_3 = 1! = 1$, so $u_4 = 3$. From $n = 2$: $u_2u_5 - u_3u_4 = 2! = 2$, so $u_5 = 8$. From $n = 3$: $u_3u_6 - u_4u_5 = 3! = 6$, so $2u_6 - 24 = 6$, $u_6 = 15$. The sequence so far is $1, 1, 1, 2, 3, 8, 15, \ldots$ — every term so far is a positive integer, but this is evidence, not proof, and the pattern $1, 2, 3, 8, 15$ needs to be recognised before it can be exploited.

The model chain. The values $u_2, u_3, u_4, u_5, u_6 = 1, 2, 3, 8, 15$ are $1!!, 2!!, 3!!, 4!!, 5!!$ where $k!!$ is the double factorial: $k!! = k(k-2)(k-4)\cdots$ down to 2 (if k even) or 1 (if k odd), with $0!! = (-1)!! = 1$ by convention. So conjecture $u_n = (n-1)!!$ for all $n \geq 0$, and verify this satisfies the given recurrence for every parity of n.

Write $n = 2m$ (even case). Using $(2k)!! = 2^k k!$ and $(2k-1)!! = (2k)!/(2^k k!)$:

$$u_n = (2m-1)!! = \frac{(2m)!}{2^m m!}, \quad u_{n+1} = (2m)!! = 2^m m!, \quad u_{n+2} = (2m+1)!! = \frac{(2m+1)!}{2^m m!}, \quad u_{n+3} = (2m+2)!! = 2$$

Then

$$u_n u_{n+3} = \frac{(2m)!}{2^m m!} \cdot 2^{m+1}(m+1)! = 2(m+1)(2m)!, \qquad u_{n+1}u_{n+2} = 2^m m! \cdot \frac{(2m+1)!}{2^m m!} = (2m+1)!,$$

so

$$u_n u_{n+3} - u_{n+1}u_{n+2} = (2m)!\big[2(m+1) - (2m+1)\big] = (2m)! \cdot 1 = n!.$$

The odd case $n = 2m+1$ is the same computation with the roles of the even/odd double factorials swapped, and gives $u_n u_{n+3} - u_{n+1}u_{n+2} = (2m+1)! = n!$ by an identical cancellation. Since $u_n = (n-1)!!$ agrees with $u_0 = u_1 = u_2 = 1$ (as $(-1)!! = 0!! = 1!! = 1$) and satisfies the same three-term-forward recurrence as the sequence in the problem, the two sequences coincide for all n by induction. As $(n-1)!!$ is a finite product of positive integers, u_n is a positive integer for every n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree the determinant condition rearranges to $u_{n+3} = (n! + u_{n+1}u_{n+2})/u_n$, and both compute the same first several terms $1, 1, 1, 2, 3, 8, 15, \ldots$ to get traction. They split immediately after that computation. The natural attempt tries to prove integrality *from the recurrence itself*, treating "u_n divides $n! + u_{n+1}u_{n+2}$" as a divisibility statement to be established by induction on the recurrence's own terms. The model instead stops computing new terms and asks what closed-form sequence the computed values already match, then switches the entire proof strategy to verifying that closed form satisfies the recurrence — proving a different, stronger statement (an explicit formula) rather than the original weaker one (mere integrality).

> *Trying to prove $u_n \mid (n! + u_{n+1}u_{n+2})$ directly from the recurrence is circular: I don't have an independent description of u_n to reason about its divisors, only the recurrence that produced it, so I have no leverage on the divisibility beyond what the recurrence already says. The five numbers I've computed — $1, 2, 3, 8, 15$ — are small and structured, small enough that I should stop pushing the recurrence forward and instead ask whether they already are a named sequence. $1, 3, 15$ are the odd double factorials and $1, 2, 8$ are... wait, $2 = 2!!$ and $8 = 4!!$ — the whole list is $(n-1)!!$. If I can guess the closed form from five terms, I can prove integrality by verifying the closed form satisfies the recurrence, which is an algebraic identity I can check directly, with no circularity, because I am no longer trying to reason about the divisors of an object defined only recursively.*

Layer 3 THE WHY

Why does "guess the closed form from small cases, then verify the identity algebraically" bypass the circularity that a direct inductive attack on the divisibility falls into? Because integrality is a *negative* kind of information — it says a certain division has no remainder — and negative information about an object defined purely by a recurrence is exactly the hardest thing to propagate inductively, since the inductive hypothesis at step n ("no remainder so far") does not by itself constrain the algebraic form of u_{n+3} enough to guarantee the *next* division is also clean. A closed form, by contrast, supplies *positive* information: $(n-1)!!$ is manifestly an integer because it is a literal product of positive integers, with no division anywhere in its own definition. The moment the induction is restated as "does this explicit product satisfy the recurrence" rather than "does this recursively-defined quantity stay an integer", the question becomes a finite algebraic identity between factorial-type expressions, which is checkable by direct computation (as carried out above) rather than a divisibility claim requiring its own separate induction.

This is also why the determinant was stated with $n!$ on the right rather than left as an abstract recurrence: $n!$ and $(n-1)!!$ are both classically expressed via the even/odd split $2^k k!$ versus $(2k)!/(2^k k!)$, so the problem is quietly pointing toward the double-factorial identity $u_n u_{n+3} - u_{n+1}u_{n+2} = n!$ being a statement that becomes transparent exactly when both sides are converted to ordinary factorials with a fixed parity split. Kedlaya's solution states the closed form and the key identity $v_{n+3}v_n = (n+2)(n)(n-1)!$ without deriving where it comes from; the derivation above is that this identity is equivalent, after the substitution $(2k)!! = 2^k k!$, to the statement that $(2m)! \cdot [2(m+1) - (2m+1)] = (2m)!$, i.e. that a specific linear combination of consecutive integers collapses to 1 — which is arithmetic, not insight, once the parity split has been made.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a three-term-forward recurrence (here via a 2×2 determinant) whose right-hand side is $n!$. On seeing it, write down: compute the first five or six terms explicitly before attempting any general argument, since a recurrence this specific is very likely to have a closed form worth guessing.
2. That cue suggests, once the terms $1, 1, 1, 2, 3, 8, 15$ are in hand, searching them against known combinatorial sequences (factorials, double factorials, binomial coefficients) rather than continuing to push the recurrence forward term by term.
3. Hesitation occurs at the divisibility step $u_n \mid (n! + u_{n+1}u_{n+2})$: it looks like it should be provable by induction because everything else about the problem is inductive in flavour, and abandoning that instinct feels like giving up on the "natural" method. What reinforces the switch is noticing explicitly that the inductive hypothesis available (previous terms are integers) says nothing about the specific divisor u_n dividing the specific numerator $n! + u_{n+1}u_{n+2}$ — the induction has no purchase on divisibility as stated.
4. The trap is treating "the first five terms are integers" as strong enough evidence to stop looking for a closed form and instead try to push the inductive divisibility argument through by

sheer persistence — the pattern $1, 1, 1, 2, 3, 8, 15$ is exactly recognisable as $(n-1)!!$, and the moment of recognising it is worth more than several more terms of blind computation.

5. Recognise it faster by treating any recurrence problem whose closed-form status is in doubt as a standing instruction: compute five to six terms immediately, and check them against double factorials, Catalan numbers, and simple factorial ratios before attempting an inductive proof on the recurrence's own terms.

6. Reusable rule: when a proof target is a negative property (integrality, divisibility, non-vanishing) of a recursively defined sequence, and a direct induction on that property is circular (the inductive hypothesis doesn't constrain the next step enough), search for a closed form from small cases and re-pose the target as a positive algebraic identity between the closed form and the recurrence — positive identities are checkable by direct computation, negative properties usually are not.

▷ **Cue:** a recursively defined sequence's integrality (or another negative property) resists direct induction on the recurrence itself $\longrightarrow$ **Move:** compute several terms, guess a closed form, and re-pose the proof as verifying that closed form satisfies the same recurrence and initial conditions

▷ **Cue:** a determinant or product recurrence produces the value $n!$ (or another factorial-type quantity) on the right-hand side $\longrightarrow$ **Move:** expect the closed form to itself be built from factorials or double factorials, and test the even/odd parity split $(2k)!! = 2^k k!$, $(2k-1)!! = (2k)!/(2^k k!)$ first

2004 A4

2004 A4. Show that for any positive integer n there is an integer N such that the product $x_1 x_2 \cdots x_n$ can be expressed identically in the form

$$x_1 x_2 \cdots x_n = \sum_{i=1}^{N} c_i (a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n)^n$$

where the c_i are rational numbers and each a_{ij} is one of the numbers $-1, 0, 1$.

Fields: algebra (polynomial identities); symmetrisation / character sums **Position:** A4 — mid-paper, but the shortest problem on it once the right sum is written down. The difficulty is entirely in choosing what to sum. **Answer:** Yes, with $N = 2^n$: $x_1 \cdots x_n = \dfrac{1}{2^n\, n!} \displaystyle\sum_{e \in \{\pm 1\}^n} (e_1 \cdots e_n)\,(e_1 x_1 + \cdots + e_n x_n)^n$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You start with $n = 2$, because that one you know:

$$xy = \tfrac{1}{4}\left[(x+y)^2 - (x-y)^2\right].$$

Good. Now $n = 3$. You try the same idea: combine cubes of $x+y+z$, $x+y-z$, $x-y+z$, $-x+y+z$ and hope the unwanted monomials cancel. Expanding $(x+y+z)^3$ produces ten monomials (x^3, x^2y, …, $6xyz$); each of the other three cubes produces the same ten with various signs. You set up unknown coefficients $c_1, \ldots, c_4$ and write ten linear equations (one per monomial) in four unknowns. It happens to work: $(x+y+z)^3 - (x+y-z)^3 - (x-y+z)^3 - (-x+y+z)^3 = 24xyz$ — though you only find that after an honest page of expansion, and the sign pattern looks arbitrary. For $n = 4$ there are 35 monomials of degree 4 in four variables and you have no idea which linear forms to take. The linear-system approach does not generalise, because you have no reason to believe the system is solvable for general n, and no way of writing its solution. This is the wall: you are solving for coefficients monomial by monomial, and the number of monomials grows like $\binom{2n-1}{n}$.

The model chain. *Claim.* For every $n \geq 1$,

$$x_1 \cdots x_n = \frac{1}{2^n\, n!} \sum_{e \in \{-1,1\}^n} (e_1 e_2 \cdots e_n)\,(e_1 x_1 + \cdots + e_n x_n)^n. \tag{$*$}$$

This is of the required form with $N = 2^n$, $a_{ij} = e_j \in \{\pm 1\}$ and $c_i = \pm\frac{1}{2^n n!} \in \mathbb{Q}$.

Proof. By the multinomial theorem,

$$(e_1 x_1 + \cdots + e_n x_n)^n = \sum_{|\alpha|=n} \binom{n}{\alpha} e_1^{\alpha_1} \cdots e_n^{\alpha_n}\, x_1^{\alpha_1} \cdots x_n^{\alpha_n},$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ runs over non-negative integer vectors with $\alpha_1 + \cdots + \alpha_n = n$ and $\binom{n}{\alpha} = \frac{n!}{\alpha_1! \cdots \alpha_n!}$. Multiply by $e_1 \cdots e_n$ and sum over all e. The coefficient of x^α becomes

$$\binom{n}{\alpha} \sum_e \prod_{j=1}^n e_j^{\alpha_j + 1} = \binom{n}{\alpha} \prod_{j=1}^n \Big(\sum_{e_j = \pm 1} e_j^{\alpha_j + 1} \Big),$$

because the sum over the product set $\{\pm 1\}^n$ of a product of one-variable functions factorises. Each factor is $1^{\alpha_j+1} + (-1)^{\alpha_j+1}$, which is 2 if α_j is odd and 0 if α_j is even. So only monomials with *every* α_j odd survive. Every α_j odd means every $\alpha_j \geq 1$; with n of them summing to n, that forces $\alpha = (1, \ldots, 1)$. Its coefficient is $\binom{n}{1,\ldots,1} \cdot 2^n = n!\, 2^n$. Dividing by $2^n n!$ gives $(*)$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from $xy = \frac{1}{4}[(x+y)^2 - (x-y)^2]$. They split at the reading of that formula (†). The natural attempt reads it as "some combination that happened to cancel" and, for $n = 3$, looks for another combination by solving equations. The model reads it as "sum over the sign of y, weighted by that sign" and asks what that weighting does to each monomial.

> *In $\frac{1}{4}[(x+y)^2 - (x-y)^2]$ I am summing over the sign $e = \pm 1$ of y, with weight e. That weighting keeps exactly the terms that are odd in y. I want to keep only the terms odd in every variable at once — so flip every variable's sign and weight by the product of the signs.*

> *Odd in each of n variables, total degree n: every exponent is at least 1 and they add to n, so every exponent is exactly 1. The only survivor is $x_1 \cdots x_n$. I don't need to expand anything else.*

Layer 3 THE WHY

Why is the sign-weighted sum not a trick? Start from what the target is. Among all degree-n monomials in $x_1, \ldots, x_n$, the product $x_1 \cdots x_n$ is characterised by a symmetry: it is the only one that changes sign whenever any single variable changes sign. That is the property to exploit, because the right-hand side of the problem is built from powers of linear forms, and every power $(\sum a_j x_j)^n$ contains *all* monomials at once. You cannot pick out one monomial from one power; you can only kill the others by averaging. And averaging over a symmetry is exactly the tool that kills everything not transforming the way you want.

Here is the mechanism at the level of the definition. For a polynomial $P(x_1, \ldots, x_n)$, the operation

$$P \longmapsto \tfrac{1}{2}\big[P(x_1, \ldots, x_j, \ldots) - P(x_1, \ldots, -x_j, \ldots)\big]$$

keeps each monomial whose exponent of x_j is odd and deletes each monomial whose exponent is even. This is nothing more than the decomposition of a function into odd and even parts, which you know from A-level: $f(t) = \frac{f(t)+f(-t)}{2} + \frac{f(t)-f(-t)}{2}$. Applying the odd-part operator in each of the n variables in turn is the sum in $(*)$ — the factor $e_1 \cdots e_n$ is precisely the "minus

sign when the variable is flipped" from each operator, multiplied out. So $(*)$ is the odd-part projection in every variable, applied to $(x_1 + \cdots + x_n)^n$. Once you see that, the only calculation left is "which monomials of degree n are odd in every variable?", and that is a one-line counting argument.

Notice what the linear forms are doing. Applying a sign change $x_j \mapsto -x_j$ to $(x_1 + \cdots + x_n)^n$ gives $(x_1 + \cdots - x_j + \cdots + x_n)^n$ — still an n-th power of a linear form with coefficients ± 1. The symmetry group (sign flips) maps the allowed building blocks to allowed building blocks. That is why the problem permits $a_{ij} \in \{-1, 0, 1\}$: those coefficients are closed under the operation you need.

Kedlaya's solution proves $(*)$ differently: the right side vanishes at $x_1 = 0$ (terms pair off under flipping e_1), so it is divisible by x_1, similarly by every x_j, hence it is a scalar times $x_1 \cdots x_n$, and the scalar is read off. That argument is correct and short, but it hides the reason the formula was written. The odd-part projection explains where the factor $e_1 \cdots e_n$ comes from; the divisibility argument only checks it afterwards. His second parenthetical ("odd as a polynomial in each individual variable") is the same idea as the one above, stated as a verification.

The rationality of c_i also has a WHY. The only denominator is $2^n n!$: the 2^n from averaging and the $n!$ from the multinomial coefficient of $x_1 \cdots x_n$. The Remark in the official solutions notes that the construction fails in characteristic $p \leq n$ for exactly this reason — the coefficient $n!$ vanishes there, so no combination of n-th powers can ever produce $x_1 \cdots x_n$. Over $\mathbb{Q}$ there is no obstacle.

Caveat. Verified independently: identity $(*)$ and the $\{0, 1\}$ variant

$$x_1 \cdots x_n = \frac{1}{n!} \sum_{e \in \{0,1\}^n} (-1)^{n - \sum e_i} \Big(\sum e_i x_i \Big)^n$$

were checked by symbolic expansion for $n = 1, \ldots, 5$. The general proof above is complete; the computation is only a guard against sign slips.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the $n = 2$ polarisation identity you already know, $xy = \frac{1}{4}[(x+y)^2 - (x-y)^2]$. On seeing the problem, write it down and rewrite it as $\frac{1}{4} \sum_{e=\pm 1} e\,(x + ey)^2$. The rewrite is the whole problem.

2. Averaging over sign changes with a sign weight (odd-part projection), applied in every variable. The target monomial is the unique one of degree n odd in each variable.

3. Hesitation comes at $n = 3$, when you are tempted to expand cubes and solve for coefficients. It is reinforced away by refusing to expand: ask what the $n = 2$ weighting does to a general monomial, not to these particular ones.

4. The trap is linear algebra on monomials: set up unknowns, write one equation per monomial, solve. It works for $n = 3$ and gives no route to general n. This is zombie maths — the most primitive expansion, begun before asking what structure selects the target.

5. Slogan: *to isolate one term from an expression that contains every term, average over a symmetry under which your term behaves differently from all others.*

6. Decision rule: **when a target is characterised by how it transforms (odd in each variable, invariant under a rotation, periodic with a given period), build it by summing the transformed versions of a generic expression, weighted by the target's own transformation factor; then only the terms that transform like the target survive.** The same rule gives roots-of-unity filters for $\sum \binom{n}{3k}$ and Fourier coefficients.

▷ **Cue:** you must express a monomial (or any specific term) as a combination of powers of linear forms, or of any expressions that each contain every term $\longrightarrow$ **Move:** average the generic expression over sign flips (or roots of unity), weighted by the character of the target term

- ▷ **Cue:** the target is the unique term that is odd in each variable separately $\longrightarrow$ **Move:** apply $P \mapsto \frac{1}{2}[P(\dots, x_j, \dots) - P(\dots, -x_j, \dots)]$ for every j; the product of signs $e_1 \cdots e_n$ is the combined weight
- ▷ **Cue:** a known small case like $xy = \frac{1}{4}[(x+y)^2 - (x-y)^2]$ $\longrightarrow$ **Move:** rewrite it as a sum over a symmetry with a weight before trying to generalise; the rewritten form generalises, the raw form does not
- ▷ **Cue:** a sum over the product set $\{\pm 1\}^n$ of a product $\prod_j \phi_j(e_j)$ $\longrightarrow$ **Move:** factorise it as $\prod_j \sum_{e_j = \pm 1} \phi_j(e_j)$

2004 A5

2004 A5. An $m \times n$ checkerboard is colored randomly: each square is independently assigned red or black with probability $1/2$. We say that two squares, p and q, are in the same connected monochromatic region if there is a sequence of squares, all of the same color, starting at p and ending at q, in which successive squares in the sequence share a common side. Show that the expected number of connected monochromatic regions is greater than $mn/8$.

Fields: probability (linearity of expectation); graph theory (components, edges, cycles) **Position:** A5 — hard end of the paper. The expectation of a component count is not a sum of local quantities, and the problem is to find a local quantity that bounds it. **Answer:** The expected number of regions is at least $\dfrac{mn+3m+3n+1}{8} > \dfrac{mn}{8}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You know linearity of expectation is the only practical tool: the number of regions is a horribly non-local function of the colouring, but expectations of sums are sums of expectations. So you look for a way to write the number of regions as a sum. The first idea is (regions) $= \sum_{\text{squares } s} 1/|\text{region containing } s|$. True, but $\mathbb{E}[1/|\text{region}|]$ is as hard as the original problem.

The second idea is from graph theory. Build the graph G whose vertices are the mn squares, with an edge between two side-adjacent squares of the same colour. Regions are exactly connected components of G. For any graph, adding one edge lowers the number of components by at most 1, so

$$\text{components} \;\geq\; V - E.$$

Now $V = mn$ is fixed and E is a sum of indicators: there are $m(n-1) + (m-1)n$ adjacent pairs, each monochromatic with probability $\frac{1}{2}$. So

$$\mathbb{E}[\text{components}] \;\geq\; mn - \tfrac{1}{2}\big(2mn - m - n\big) = \tfrac{m+n}{2}.$$

That is linear in the side lengths, not in the area. For a 100×100 board it gives 100; you need 1250. This is the wall: the bound $V - E$ is exact for forests and throws away every cycle, and a random two-colouring of a grid is full of cycles.

The model chain. *Setup.* Let G be as above. Let A be the number of edges of G and B the number of monochromatic 2×2 blocks (four squares in a 2×2 arrangement, all one colour). Index the block whose top-left square is in row i, column j by (i, j), $1 \leq i \leq m-1$, $1 \leq j \leq n-1$.

Step 1: components $\geq mn - A + B$. For each monochromatic block (i, j), delete from G the edge joining its two bottom squares, $(i+1, j)$–$(i+1, j+1)$. Distinct blocks give distinct edges (the edge determines (i, j)), so this deletes exactly B edges; call the result G'. I claim G' has the same components as G. It suffices that the two ends of every deleted edge are joined by a path in G'. Vertical edges are never deleted. Induct on i. For block (i, j) the path

$(i+1,j) \to (i,j) \to (i,j+1) \to (i+1,j+1)$ uses two vertical edges (present in G') and the top edge (i,j)–$(i,j+1)$. If $i = 1$ that top edge lies in row 1, which is never the bottom row of a block, so it is present. If $i > 1$ it is either present or it is the deleted bottom edge of block $(i-1,j)$, whose ends are joined in G' by the induction hypothesis. Either way $(i+1,j)$ and $(i+1,j+1)$ are joined in G'. Hence

$$\text{components}(G) = \text{components}(G') \;\geq\; mn - (A - B).$$

Step 2: expectations. By linearity,

$$\mathbb{E}[A] = \tfrac{1}{2}\big(m(n-1) + (m-1)n\big), \qquad \mathbb{E}[B] = \tfrac{2}{16}(m-1)(n-1) = \tfrac{1}{8}(m-1)(n-1),$$

since a block is monochromatic with probability $2/2^4$. Therefore

$$\mathbb{E}[\text{components}] \;\geq\; mn - \tfrac{2mn-m-n}{2} + \tfrac{(m-1)(n-1)}{8} = \frac{mn+3m+3n+1}{8} > \frac{mn}{8}. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach components $\geq V - E$ and compute $\mathbb{E}[\text{components}] \geq \frac{m+n}{2}$. At that line (†) the natural attempt concludes the graph approach is too weak and goes looking for a different idea. The model asks what exactly the inequality threw away, and puts some of it back.

> *$V - E$ is exact for a forest. Every time an edge closes a cycle, I've subtracted 1 for nothing. So the loss is the number of independent cycles. I can't count those, but I can count the cheapest cycles there are — monochromatic 2×2 squares — and each has probability $\frac{1}{8}$, which is exactly the constant the problem is asking for.*

> *The target is $mn/8$ and each 2×2 block is monochromatic with probability $2/16 = 1/8$. The area term in the answer has to come from something with one per unit area. That's the blocks.*

Layer 3 THE WHY

Why is "correct $V - E$ by cycles" the natural move rather than an inspiration? Go to the definition. For a graph, components $= V - E + \beta$ where β is the cycle rank: the number of edges you can delete without disconnecting anything (build a spanning forest; the edges left over are β in number, and each closed a cycle when added). This is not a deep theorem; it is the same edge-by-edge argument that gives $V - E$, with the bookkeeping kept. So the inequality components $\geq V - E$ is the statement $\beta \geq 0$, and it is weak exactly when β is large. The question "how do I improve the bound?" therefore becomes "how do I find many edges I can delete without disconnecting anything?" — and to use linearity of expectation, the count of such edges must be a sum of local indicators.

What local configuration guarantees a deletable edge? A cycle. The shortest cycles in a grid graph are the 2×2 blocks, and a block is a cycle in G exactly when it is monochromatic. That is the most local cycle the geometry allows, and its probability is a constant, $\frac{1}{8}$. The problem's constant $\frac{1}{8}$ is a strong hint that it is the right one: the area term $\frac{mn}{8}$ is $\mathbb{E}[B]$ to leading order, and $V - \mathbb{E}[A]$ cancels the rest of the mn exactly. Reading the answer's constant as a probability is a legitimate step, not a guess: on a competition paper, a clean constant in an expectation bound is almost always $\Pr(\text{some small event})$.

The one point where rigour is required is the one the official solution states in a single clause: "if we remove the bottom edge of each 4-cycle, the resulting graph has the same number of connected components". Removing edges one at a time is not obviously safe, because the top

edge of one block is the bottom edge of the block above it — so a cycle you rely on may already have been cut. The induction on the row index in Step 1 is what makes it correct: deletions only ever remove horizontal edges, and the vertical edges carry every repair path up to a row where the top edge is known to be reconnected. Write this; it is where a grader looks.

Why is the natural attempt's first idea ($\sum 1/|\text{region}|$) the wrong kind of sum? Because it is exact. Exact expressions for global quantities are usually as hard as the quantity; the tool that makes linearity useful is an *inequality* whose right side is local. "Components $\geq$ $V - E +$ (local cycles)" is such an inequality.

Kedlaya's second solution (Elkies) is the same idea organised sequentially: add squares in reading order; a new square adds a component with probability $\frac{1}{4}$ (differs from both the square above and the square to the left) and can remove one only when it merges two regions, which requires the square above and to the left to agree with it while the diagonal square differs — probability $\frac{1}{8}$. The net $\frac{1}{4} - \frac{1}{8} = \frac{1}{8}$ per square is again the blocks' $\frac{1}{8}$, seen one square at a time.

Caveat. Verified independently: for every colouring of every board with $mn \leq 16$ (all $m, n \leq 4$), the inequality components $\geq mn - A + B$ holds, and the exact expectation was computed by full enumeration. For $m \leq 2$ or $n \leq 2$ the bound $\frac{mn+3m+3n+1}{8}$ is attained exactly (e.g. $\frac{11}{4}$ for 2×3); for 3×3 the exact value is $\frac{897}{256}$ against the bound $\frac{7}{2}$, and for 4×4 it is $\frac{168529}{32768} \approx 5.143$ against $\frac{41}{8} = 5.125$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "expected number of connected regions": a global count under independent local randomness. Write down at once: *linearity of expectation needs a sum of local indicators; find an inequality components $\geq$ (local sum).* Then write components $\geq V - E$.

2. Graph bookkeeping: components $= V - E + \beta$, then bound the cycle rank β below by a count of local cycles (monochromatic 2×2 blocks).

3. Hesitation occurs when $V - E$ gives $\frac{m+n}{2}$ and the approach looks dead. Reinforce it away by asking what the inequality discarded; the answer (cycles) names the correction term.

4. Two traps. First, abandoning a correct framework because its first bound is weak. Second, writing "remove one edge from each 4-cycle" without checking that the removals do not interfere: adjacent blocks share edges, and the repair paths must be shown to survive.

5. Slogan: *when an inequality is too weak, name exactly what it threw away, then put back the cheapest local piece of it.* And: *a clean constant in an expected-value bound is the probability of a small local event — find the event.*

6. Decision rule: **to bound the expectation of a global count (components, regions, cycles) under independent local randomness, express the count as a local sum plus a non-negative error, then bound the error below by another local sum; take expectations of both sums by linearity.**

▷ **Cue:** expected number of connected components / regions / clusters of a random configuration $\longrightarrow$ **Move:** build the graph; use components $= V - E + \beta$, bound the cycle rank β below by a count of local cycles whose deleted edges are distinct, and take expectations term by term

▷ **Cue:** a bound from $V - E$ comes out linear in the perimeter when the target is linear in the area $\longrightarrow$ **Move:** the loss is the cycle rank; add back the shortest cycles, which occur with constant density

▷ **Cue:** the problem's constant (here $\frac{1}{8}$) is a small power of $\frac{1}{2}$ $\longrightarrow$ **Move:** find the local event whose probability it is (here, a monochromatic 2×2 block, $2/2^4$)

▷ **Cue:** "delete one edge from each of these cycles" $\longrightarrow$ **Move:** check the cycles do not share the deleted edges or the repair paths; order the deletions (here by row) and induct

2004 A6

2004 A6. Suppose that $f(x,y)$ is a continuous real-valued function on the unit square $0 \le x \le 1, 0 \le y \le 1$. Show that

$$\int_0^1 \left(\int_0^1 f(x,y)dx \right)^2 dy + \int_0^1 \left(\int_0^1 f(x,y)dy \right)^2 dx \le \left(\int_0^1 \int_0^1 f(x,y)dx\,dy \right)^2 + \int_0^1 \int_0^1 [f(x,y)]^2\,dx\,dy.$$

Fields: analysis (integral inequalities); linear algebra (orthogonal decomposition) **Position:** A6 — the last problem of the morning, but short once the right decomposition is written; the difficulty is that every standard inequality points the wrong way. **Answer:** True, and in fact RHS − LHS $= \int_0^1\int_0^1 \big(f(x,y) - g(x) - h(y) + c\big)^2\,dx\,dy \ \ge\ 0$, where $g(x) = \int_0^1 f(x,y)\,dy$, $h(y) = \int_0^1 f(x,y)\,dx$, $c = \iint f$.

Layer 1 THE WHAT — surface analysis

Notation for both chains: $g(x) = \int_0^1 f(x,y)\,dy$ (the row average), $h(y) = \int_0^1 f(x,y)\,dx$ (the column average), $c = \int_0^1\int_0^1 f$, and $\|\phi\|^2$ for the integral of ϕ^2 over its domain. The inequality reads

$$\|h\|^2 + \|g\|^2 \ \le\ c^2 + \|f\|^2. \tag{$*$}$$

The natural attempt. You see squares of integrals on the left, integrals of squares on the right: Cauchy–Schwarz. By Cauchy–Schwarz (or Jensen), $g(x)^2 = \big(\int_0^1 f(x,y)\,dy\big)^2 \le \int_0^1 f(x,y)^2\,dy$, so $\|g\|^2 \le \|f\|^2$; likewise $\|h\|^2 \le \|f\|^2$. Adding: LHS $\le 2\|f\|^2$. To finish you would need $2\|f\|^2 \le c^2 + \|f\|^2$, i.e. $\|f\|^2 \le c^2$. But Cauchy–Schwarz says $c^2 \le \|f\|^2$ — the opposite. So each of your two estimates is true and their sum is too generous: you have spent the one allowance of $\|f\|^2$ twice. The wall is real: no inequality applied to g and h separately can work, because it cannot see that the two terms compete for the same budget.

Next you try a test function. $f = 1$: LHS $= 2$, RHS $= 2$. Equality. $f(x,y) = x$: LHS $= \frac14 + \frac13$, RHS $= \frac14 + \frac13$. Equality again. $f = xy$: LHS $= \frac{1}{12} + \frac{1}{12} = \frac16$, RHS $= \frac{1}{16} + \frac19 = \frac{25}{144}$, strict. You now suspect equality holds for a whole family, but you don't yet know what to do with that.

The model chain. Define the *interaction part*

$$r(x,y) = f(x,y) - g(x) - h(y) + c.$$

Claim. $\|r\|^2 = \|f\|^2 + c^2 - \|g\|^2 - \|h\|^2$, i.e. RHS − LHS $= \|r\|^2$. This proves $(*)$.

Proof. All functions are continuous on $[0,1]^2$, so every integral below exists and Fubini applies. Expand r^2 and integrate term by term over the square:

$$\|r\|^2 = \|f\|^2 + \|g\|^2 + \|h\|^2 + c^2 - 2\iint fg - 2\iint fh + 2c\iint f + 2\iint gh - 2c\iint g - 2c\iint h.$$

Evaluate each cross term by integrating in the variable that g or h does not depend on:

- $\iint f(x,y)g(x)\,dy\,dx = \int_0^1 g(x)\big(\int_0^1 f(x,y)\,dy\big)dx = \|g\|^2$; likewise $\iint fh = \|h\|^2$.
- $\iint g(x)h(y)\,dx\,dy = \big(\int_0^1 g\big)\big(\int_0^1 h\big) = c\cdot c = c^2$, since $\int_0^1 g = \int_0^1 h = c$ by Fubini.
- $\iint f = c$, $\iint g = c$, $\iint h = c$.

Substituting: $\|r\|^2 = \|f\|^2 + \|g\|^2 + \|h\|^2 + c^2 - 2\|g\|^2 - 2\|h\|^2 + 2c^2 + 2c^2 - 2c^2 - 2c^2 = \|f\|^2 - \|g\|^2 - \|h\|^2 + c^2$. Since $\|r\|^2 \geq 0$, $(*)$ follows, with equality exactly when $r \equiv 0$, i.e. $f(x,y) = a(x) + b(y)$ for some continuous a, b. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains notice squares-of-integrals against integrals-of-squares. The split is at the test functions (†): equality for $f = 1$ and for $f = x$. The natural attempt treats these as sanity checks and goes back to hunting for an inequality. The model treats them as the specification of the answer: a non-negative quadratic quantity that vanishes on a subspace should be the squared size of the part of f outside that subspace.

> *Cauchy–Schwarz on g and on h separately spends $\|f\|^2$ twice. Whatever works has to split $\|f\|^2$ into pieces and give one piece to g, one to h, and one to c — without double counting. That's Pythagoras: I need orthogonal pieces.*

> *Equality for $f = 1$, for $f = x$, and (checking) for any $f = a(x) + b(y)$. RHS − LHS is a quadratic form in f that vanishes on all of those. The obvious guess is that it's $\|f - (\text{best } a(x) + b(y))\|^2$. The best $a(x) + b(y)$ should be built from the row and column averages: $g(x) + h(y) - c$.*

Layer 3 THE WHY

Why is the orthogonal decomposition not a rabbit from a hat? Start with what each side measures. Every term in $(*)$ is a squared length in the space of square-integrable functions on $[0,1]^2$, with inner product $\langle \phi, \psi \rangle = \iint \phi\psi$. And the operations in the problem are projections: averaging over y replaces f by $g(x)$, the closest function of x alone (closest in $\|\cdot\|$); averaging over x gives $h(y)$, the closest function of y alone; averaging over both gives the constant c. The "closest" claims are the continuous version of "the mean minimises the sum of squared deviations", which you already know from statistics. Squared lengths of projections are exactly what Pythagoras controls.

So write f as a sum of four mutually orthogonal pieces, each one a new layer of the averaging:

$$f = \underbrace{c}_{\text{grand mean}} + \underbrace{(g-c)}_{\text{row effect}} + \underbrace{(h-c)}_{\text{column effect}} + \underbrace{r}_{\text{interaction}}.$$

Each piece averages to zero in every direction in which the pieces before it are constant: $g - c$ integrates to 0 in x; $h - c$ integrates to 0 in y; r integrates to 0 in x for each fixed y and in y for each fixed x (check: $\int_0^1 r\,dy = g(x) - g(x) - c + c = 0$). A function of x alone with zero mean is orthogonal to anything that does not depend on x and to anything with zero x-mean for each y; that is all the cross-term calculations say. Pythagoras then gives

$$\|f\|^2 = c^2 + \|g-c\|^2 + \|h-c\|^2 + \|r\|^2, \qquad \|g\|^2 = c^2 + \|g-c\|^2, \qquad \|h\|^2 = c^2 + \|h-c\|^2.$$

Substitute into $(*)$: the left side is $2c^2 + \|g-c\|^2 + \|h-c\|^2$, the right side is $2c^2 + \|g-c\|^2 + \|h-c\|^2 + \|r\|^2$. The inequality is "$0 \leq \|r\|^2$" with everything else cancelling. This is the same bookkeeping as two-way analysis of variance: total sum of squares = grand mean + row + column + interaction.

This also explains why the natural attempt failed. Cauchy–Schwarz gives $\|g\|^2 \leq \|f\|^2$ by discarding $\|h-c\|^2 + \|r\|^2$, and $\|h\|^2 \leq \|f\|^2$ by discarding $\|g-c\|^2 + \|r\|^2$. Adding them counts the row and column effects twice on the right, and the c^2 once too few; the correct accounting needs the decomposition, not two inequalities.

Kedlaya's solution states: the difference is $\frac{1}{4}\iiint\int F^2$ with $F(x,y,z,w) = f(x,y) + f(z,w) - f(x,w) - f(z,y)$. Here is where that comes from. F is a second difference — it subtracts the

w-column from the y-column and the z-row from the x-row — so it annihilates every $a(x) + b(y)$, exactly the equality cases. Substituting $f = c + (g - c) + (h - c) + r$, everything but r cancels in F, and since r has zero row and column integrals, $\iiint F^2 = 4\|r\|^2$. The official identity is the decomposition above, written without naming the pieces. His discretisation to an $n \times n$ matrix is unnecessary: the identity holds directly for integrals.

Caveat. Verified independently: the identity RHS $-$ LHS $= \|r\|^2$ was checked by hand expansion (above) and numerically in its discrete form (means over random $k \times k$ matrices, $k \leq 7$) to agreement at 10^{-15}. The test values for $f = xy$ ($\frac{1}{6}$ against $\frac{25}{144}$) were checked by hand.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is structural: squares of averages on one side, the square of the total average and the average of the square on the other. Write down the names g, h, c at once, and the inequality as $\|g\|^2 + \|h\|^2 \leq c^2 + \|f\|^2$. Then test f depending on one variable only, and record the equality.

2. Orthogonal decomposition (Pythagoras in function space): split f into grand mean, row effect, column effect, interaction, and compare squared lengths.

3. Hesitation occurs after Cauchy–Schwarz fails by a factor of two and there is no other standard inequality to hand. Reinforce it away with the equality cases: when a difference of quadratic expressions vanishes on a whole subspace, it is asking to be a squared distance from that subspace.

4. The trap is applying an inequality to each term separately. When the target has several terms competing for one budget ($\|f\|^2$), term-by-term estimates double-spend it. A second trap is Kedlaya's discretisation, which is correct but adds a limiting argument you do not need.

5. Slogan: *if an inequality between quadratic expressions is tight on a whole family, the difference is a sum of squares that vanishes on that family — find the projection that kills it.*

6. Decision rule: **when an inequality involves squares of averages and averages of squares, decompose the function into orthogonal pieces (mean, then successive deviations) and compare by Pythagoras; do not estimate term by term.** The same rule proves the variance formula $\mathbb{E}[X^2] - (\mathbb{E}X)^2 \geq 0$, Bessel's inequality, and the ANOVA decomposition.

▷ **Cue:** squares of partial integrals (or averages) compared with integrals of squares $\longrightarrow$ **Move:** introduce the averaged functions by name and decompose f as mean + row effect + column effect + interaction; apply Pythagoras

▷ **Cue:** Cauchy–Schwarz applied to each of two terms gives a bound twice too large $\longrightarrow$ **Move:** the terms share one budget; switch from separate inequalities to an exact orthogonal split

▷ **Cue:** equality holds in a quadratic inequality for every f in a subspace (here all $a(x) + b(y)$) $\longrightarrow$ **Move:** guess that the difference is $\|f - Pf\|^2$ where P projects onto that subspace, then verify by expansion

▷ **Cue:** an official solution introduces a mixed second difference like $f(x,y) + f(z,w) - f(x,w) - f(z,y)$ $\longrightarrow$ **Move:** recognise it as the operator that kills additive functions $a(x) + b(y)$; its square integrates to a multiple of the interaction's squared norm

Putnam 2004 — Paper B

2004 B1

2004 B1. Let $P(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_0$ be a polynomial with integer coefficients. Suppose that r is a rational number such that $P(r) = 0$. Show that the n numbers

$$c_n r,\ c_n r^2 + c_{n-1} r,\ c_n r^3 + c_{n-1} r^2 + c_{n-2} r,\ \ldots,\ c_n r^n + c_{n-1} r^{n-1} + \cdots + c_1 r$$

are integers.

Fields: Algebra, number theory (rational root theorem, Gauss's lemma) **Position:** B1 – opening problem of the second session **Answer:** Each of the n partial sums is an integer, because each one can be written simultaneously as a fraction with denominator q^{n-k} and as a fraction with denominator p^k, where $r = p/q$ in lowest terms; coprimality of p and q forces both fractions to be whole numbers.

Layer 1 THE WHAT — surface analysis

The natural attempt. You write $r = p/q$ in lowest terms because the problem is visibly a rational-root-theorem cousin. You substitute into $P(r) = 0$ and clear denominators by multiplying through by q^n:

$$c_n p^n + c_{n-1} p^{n-1} q + \cdots + c_0 q^n = 0.$$

Every term here is manifestly an integer, and the whole sum is zero – an integer. That feels like progress: you have shown *something* about $P(r) = 0$ is an integer statement. But look again at what the problem actually asks for. It does not ask you to show $P(r) = 0$ scaled by q^n is an integer (that's trivial once cleared, and equals 0 anyway). It asks you to show each of n different *partial* sums – $c_n r$, then $c_n r^2 + c_{n-1} r$, and so on up to the second-to-last one – is individually an integer. Multiplying the whole equation by q^n tells you nothing about any single partial sum; it only re-derives the one global fact you already had. You can sit with the cleared equation for a while, try to peel a partial sum back out of it, and find there is no clean way to isolate, say, just $c_n r^3 + c_{n-1} r^2 + c_{n-2} r$ from an equation that only ever gives you the full alternating combination equal to zero.

The model chain. Fix an index k with $0 \le k \le n-1$. Since $P(r) = 0$, dividing by r^k gives

$$c_n r^{n-k} + c_{n-1} r^{n-k-1} + \cdots + c_{k+1} r \;=\; -\Big(c_k + c_{k-1} r^{-1} + \cdots + c_0 r^{-k}\Big).$$

This is one equation, but its two sides invite two different clearings. Write $r = p/q$ with $\gcd(p, q) = 1$. Multiply the left side by q^{n-k}: every term $c_i r^{i-k}$ ($i > k$) becomes $c_i p^{i-k} q^{n-i}$, an integer, since $i - k \ge 1$ and $n - i \ge 0$. So the left side equals A/q^{n-k} for an integer A. Multiply the right side by p^k: every term $c_i r^{i-k}$ ($i \le k$) becomes $c_i p^{k-i} q^i \cdot (\text{power of } p)$, concretely $-c_i p^k r^{i-k} = -c_i p^i q^{k-i}$, an integer since $k - i \ge 0$. So the right side equals B/p^k for an integer B. But the left and right sides are the *same number* V. So $V = A/q^{n-k} = B/p^k$, which gives $Ap^k = Bq^{n-k}$. Since $\gcd(p, q) = 1$, also $\gcd(p^k, q^{n-k}) = 1$, so q^{n-k} divides A, meaning $V = A/q^{n-k}$ is an integer. As k ranges over $0, \ldots, n-1$, the quantity V (which is

$c_n r^{n-k} + \cdots + c_{k+1} r$) runs exactly over the problem's n listed numbers, in reverse order. All are integers.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on the opening move: write $r = p/q$ in lowest terms, then use $P(r) = 0$. They split at what to do with that equation. The natural attempt clears denominators across the *whole* polynomial at once, at the single scale q^n. The model chain instead divides by r^k *before* clearing anything, for every k from 0 to $n-1$ in turn – and then clears the two resulting halves of that equation at two *different* scales, q^{n-k} on one side and p^k on the other.

> *The numbers I need to prove are integers are partial sums, not the whole polynomial evaluated at r – so whatever equation I write down has to isolate a partial sum on one side, and dividing $P(r) = 0$ by r^k before splitting is exactly how I get a partial sum to stand alone.*

> *I have the same quantity written two ways, with denominator q^{n-k} on one side and denominator p^k on the other; since p and q share no common factor, neither does p^k and q^{n-k}, and a number that is simultaneously (integer)/q^{n-k} and (integer)/p^k with those denominators coprime has to actually be an integer.*

Layer 3 THE WHY

Caveat. The chain above is a full derivation I checked line by line against small cases (e.g. $P(x) = 2x^2 - x - 1$ with $r = -\frac{1}{2}$, and $P(x) = 6x^2 - 5x + 1$ with $r = \frac{1}{3}$), and it matches Kedlaya-Ng's published solution. No step is unverified.

Why does splitting at r^k – rather than clearing the whole polynomial – produce two coprime denominators instead of one? Because the two halves of $P(r) = 0$ after dividing by r^k are structurally different: the high-degree half ($i > k$) has every surviving power of r non-negative, so it clears with q^{n-k}; the low-degree half ($i \leq k$) has every surviving power of r *negative or zero* (because you divided by r^k and those terms have degree $\leq k$), so it clears with p^k instead. The division by r^k is precisely what manufactures negative exponents on one side and only non-negative ones on the other – and it is that asymmetry, not anything about the specific coefficients, that hands you two different natural denominators to compare. Clearing the whole polynomial at q^n never creates this asymmetry, because every term in $P(r) = 0$ already has a non-negative power of r; there is only ever one denominator scale available, and one scale can never force integrality of anything smaller than the sum itself. The deeper fact underneath is basic: a rational number equal to (integer)/m and also equal to (integer)/m' with $\gcd(m, m') = 1$ must be an integer, because its reduced denominator divides both m and m' and so must be 1. This is the same mechanism as the rational root theorem itself (where $P(r) = 0$ forces $q \mid c_n$ and $p \mid c_0$ by an identical coprime-denominator argument) — the problem is really asking you to run that mechanism n times, once per truncation point of the polynomial, rather than once for the whole thing.

Layer $+\alpha$ BACKTESTING

1. The cue is "r is rational" sitting next to "integer coefficients" and "$P(r) = 0$." On seeing it, write $r = p/q$ in lowest terms immediately and flag $\gcd(p, q) = 1$ as the object you will exploit – that coprimality, not the polynomial itself, is the engine of the whole proof.

2. The cue suggests the rational-root-theorem family of arguments: clear denominators and use coprimality to force divisibility. The refinement needed here is that you must clear denominators *locally*, at each truncation index k, not globally over the whole equation.

3. Hesitation occurs right after clearing the whole equation at q^n and finding the result trivial (it just reproduces $0 = 0$). That dead end is what should push you to ask "what equation, if I clear it, gives me one specific partial sum rather than the whole thing" – and dividing by r^k first is the answer.

4. The trap is mistaking "I cleared denominators and got an integer equation" for having proven what was asked. The global clearing at q^n is real progress-shaped activity that produces nothing usable, because the object of interest (a partial sum) never appears in isolation in that equation.

5. To recognise this faster: whenever a problem hands you n related quantities to prove integral, and they look like the partial sums of a single defining equation, expect that each one needs its *own* instance of the coprime-denominator argument, indexed by where the truncation happens – not one global clearing.

6. The reusable rule: if a rational number can be written both as (integer)$/m$ and as (integer)$/m'$ with $\gcd(m, m') = 1$, it is an integer. Whenever you have one equation with a rational root, look for a way to split it into two halves that naturally clear at coprime scales – dividing by a power of the root before splitting is the standard way to manufacture that split.

2004 B2

2004 B2. Let m and n be positive integers. Show that

$$\frac{(m+n)!}{(m+n)^{m+n}} < \frac{m!}{m^m}\frac{n!}{n^n}.$$

Fields: Algebra, combinatorics (binomial theorem) **Position:** B2 – second problem of the second session
Answer: The inequality is equivalent to $(m+n)^{m+n} > \binom{m+n}{m} m^m n^n$, and the right side is exactly one term of the binomial expansion of $(m+n)^{m+n} = \sum_i \binom{m+n}{i} m^i n^{m+n-i}$, which has strictly more than one positive term.

Layer 1 THE WHAT — surface analysis

The natural attempt. The expression $m!/m^m$ screams Stirling: $m! \sim \sqrt{2\pi m}\,(m/e)^m$, so $m!/m^m \sim \sqrt{2\pi m}\,e^{-m}$. Multiply the asymptotics for m and n together and compare against the asymptotic for $m+n$, and you can chase down an inequality between the leading exponential terms that plausibly gives you what you want in the limit of large m, n. This is a real path, and it is not stupid – it's the reflex of someone who has just seen $n!/n^n$ and reaches for the tool built exactly for factorial-over-exponential ratios. But it costs you dearly. Stirling's formula is an asymptotic with error terms that need controlling (the standard bounds are $\sqrt{2\pi m}(m/e)^m e^{1/(12m+1)} < m! < \sqrt{2\pi m}(m/e)^m e^{1/(12m)}$), and turning "asymptotically true" into "true for every positive integer m, n, with strict inequality, including $m = n = 1$" means carrying those error bounds through the whole computation and checking they never flip the inequality at small values. You can do it – but you spend the exam reconciling error terms for a problem that, as the model chain shows, needs no calculus and no approximation at all.

The model chain. Expand $(m+n)^{m+n}$ by the binomial theorem, treating it as $(a+b)^N$ with $a = m, b = n, N = m+n$:

$$(m+n)^{m+n} = \sum_{i=0}^{m+n} \binom{m+n}{i} m^i n^{m+n-i}.$$

Every term in this sum is a positive integer (since $m, n \geq 1$). The term at $i = m$ is exactly

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

Because $m, n \geq 1$, we have $0 < m < m+n$, so $i = m$ is a genuinely interior index: the sum has at least the three distinct positive terms $i = 0$, $i = m$, and $i = m+n$ (or at least two, if $m = n+...$ – concretely, whenever $m+n \geq 2$, which holds since $m, n \geq 1$, the sum has more than one term). A sum of two or more positive numbers strictly exceeds any single one of them, so

$$(m+n)^{m+n} > \binom{m+n}{m} m^m n^n = \frac{(m+n)!}{m!\,n!}\, m^m n^n.$$

Multiply both sides by $m!\,n!$ and divide both sides by $(m+n)^{m+n}\, m^m\, n^n$:

$$\frac{m!\,n!}{m^m n^n} > \frac{(m+n)!}{(m+n)^{m+n}},$$

which is exactly the claimed inequality, read right to left.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches start from the same visual cue: a factorial divided by a self-power, for three related quantities m, n, $m+n$. They split on what kind of tool that cue is treated as calling for. The natural attempt reads "factorial over exponential" and reaches for the asymptotic tool built for exactly that shape (Stirling), which is a statement about *growth rates*. The model chain instead reads $(m+n)^{m+n}$ – the term *without* a factorial in it – as an object with its own exact algebraic expansion, and asks what that expansion contains.

> *I don't need an inequality-proving tool at all here, because $(m+n)^{m+n}$ is not just some number to be estimated – it is literally $(m+n)$ raised to a power, with $m+n$ = the sum of the two bases I care about, so the binomial theorem applies to it directly and exactly, with no approximation.*

> *Once I've expanded $(m+n)^{m+n}$ as a sum over i, I should look for the specific term that matches what I'm trying to beat – and $\binom{m+n}{m} m^m n^n$, which is exactly $\frac{(m+n)!}{m!\,n!} m^m n^n$, is sitting right there at $i = m$; the whole inequality is just "a sum of positive terms is bigger than one of them," which needs no estimation, only the observation that the sum has more than one term.*

Layer 3 THE WHY

Caveat. None. This derivation is exact algebra (no approximation, no limiting argument), verified directly by rearrangement, and matches Kedlaya-Ng's published first solution term for term.

Why is $(m+n)^{m+n}$ the right object to expand, rather than $m!/m^m$ or $n!/n^n$ individually? Because the problem compares a quantity built from $m+n$ against a *product* of quantities built from m and n separately, and the binomial theorem is precisely the identity that expresses a power of a *sum* in terms of powers of the two summands – it is the one algebraic tool in this problem whose inputs and outputs line up exactly with the shape of the inequality, term for term, with no residue. Compare this to why Stirling looks tempting but is the wrong register: Stirling estimates $m!$ (a genuinely irregular, non-closed-form quantity) by a smooth function; but nothing here actually requires estimating a factorial in isolation – the factorial that matters, $(m+n)!/(m!\,n!) = \binom{m+n}{m}$, already has an exact combinatorial meaning (a binomial coefficient), and that coefficient's exact home is inside the binomial expansion of $(m+n)^{m+n}$. Once you see the binomial coefficient sitting in the original expression (divide the target inequality through and $\binom{m+n}{m}$ appears immediately), the identity $(a+b)^N = \sum \binom{N}{i} a^i b^{N-i}$ is not a trick pulled from nowhere – it is the unique standard identity that has $\binom{m+n}{m}$, m^m, and n^n all appearing together as a single term, which is exactly the combination the problem needs bounded from above.

And strictness costs nothing further: a finite sum of positive numbers with two or more terms strictly exceeds any one term, by the ordering axioms on the positive reals – there is no case analysis on m, n needed beyond noting $m, n \geq 1$ guarantees at least one other term exists.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is seeing m^m, n^n, and $(m+n)^{m+n}$ all in the same problem – three separate powers, each base equal to its own exponent. Write down immediately: "this is the shape of a binomial expansion $(a+b)^N$ with $a = m, b = n, N = m+n$," before reaching for anything else.
2. That cue suggests expanding $(m+n)^{m+n}$ by the binomial theorem and hunting for the specific term $\binom{m+n}{m} m^m n^n$ inside it, rather than estimating any factorial.
3. Hesitation occurs at the sight of $m!/m^m$, which looks like it wants Stirling because it is the textbook shape for that estimate. That impulse is reinforced away by first rearranging the target inequality algebraically – multiply both sides by $m!\,n!\,(m+n)^{m+n}$ – and noticing a binomial coefficient $\binom{m+n}{m}$ falls straight out; once a binomial coefficient is visible, the binomial theorem is the natural next tool, not an asymptotic one.
4. The trap is treating "factorial divided by something" as automatically a Stirling problem. Stirling is the right tool when you need the *size* of a factorial in isolation; here the factorials only ever appear in the ratio $(m+n)!/(m!\,n!)$, i.e. as a binomial coefficient, and binomial coefficients have exact combinatorial identities that make approximation unnecessary and strictly weaker.
5. To recognise this faster: whenever an inequality compares N^N to a product of a^a-type terms with $a + b = N$, rearrange first to see whether a binomial coefficient $\binom{N}{a}$ appears; if it does, the binomial theorem answer is usually one line away.
6. The reusable rule: to prove one specific weighted count is strictly less than a total count, express the total as an explicit sum that contains the specific count as one positive term among several – strict inequality then costs nothing beyond confirming the sum has more than one term. This beats asymptotic or calculus estimation whenever an exact combinatorial identity is available.

2004 B3

2004 B3. Determine all real numbers $a > 0$ for which there exists a nonnegative continuous function $f(x)$ defined on $[0, a]$ with the property that the region

$$R = \{(x, y); 0 \leq x \leq a, 0 \leq y \leq f(x)\}$$

has perimeter k units and area k square units for some real number k.

Fields: Real analysis, plane geometry **Position:** B3 – middle problem of the second session **Answer:** $a > 2$. A constant function $f(x) = 2a/(a-2)$ realises it for $a > 2$; for $a \leq 2$ perimeter strictly exceeds area for every admissible f, so no such $a \leq 2$ works.

Layer 1 THE WHAT — surface analysis

The natural attempt. You suspect, correctly, that a constant function should be the extremal case to test first, and quickly find that a rectangle of width a and height c works precisely when $a > 2$, by solving $2a + 2c = ac$. That half is not the hard part. The hard part is the converse: proving that for $a \leq 2$, *no* nonnegative continuous f – however wild – can make perimeter equal area. Here the natural instinct is to reach for calculus: perimeter of the curve $y = f(x)$ from $x = 0$ to $x = a$ "is" $\int_0^a \sqrt{1 + f'(x)^2}\,dx$, plus the two vertical end-segments, plus the base. You start trying to bound this arc-length integral from below by something involving the area $\int_0^a f(x)\,dx$, hoping for a clean isoperimetric-style comparison. Two problems surface almost

immediately. First, f is only assumed *continuous*, not differentiable – the arc-length integral may not even be defined, and the "perimeter" in the problem must mean something more primitive (an infimum over polygonal approximations, or simply "assume it's finite and behaves like length should"). Second, even if you restrict attention to nice f, bounding $\int \sqrt{1+f'^2}$ below in terms of $\int f$ in a way that is tight enough to beat $a \leq 2$ exactly, with no slack, is a real analysis fight you did not sign up for on a "determine all a" problem.

The model chain. Forget the whole curve. Let $P = (x_0, y_0)$ be any point on the graph of f where f attains its maximum on $[0, a]$ (it exists since f is continuous on a closed interval). Two crude bounds, neither requiring f to be anything more than continuous:

Area bound. The region R lies entirely below the horizontal line $y = y_0$ and has width a, so

$$\text{area}(R) \;\leq\; a\, y_0.$$

Perimeter bound. The boundary of R contains a path from $(0,0)$ to P, then from P to $(a,0)$, then the straight base back from $(a,0)$ to $(0,0)$. Path length is always at least the straight-line distance between its endpoints, and the straight-line distance from $(0,0)$ to $P = (x_0, y_0)$ is at least y_0 (the vertical leg alone), and likewise from P to $(a,0)$. The base contributes exactly a. So

$$\text{perimeter}(R) \;\geq\; 2y_0 + a.$$

Now compare the two bounds when $a \leq 2$. Claim: $2y_0 + a > a\, y_0$ for every $y_0 \geq 0$. If $y_0 \leq 1$, both $2y_0 \geq 0$ and $a(1-y_0) \geq 0$, and $a > 0$ forces the sum $2y_0 + a(1-y_0)$ strictly positive, i.e. $2y_0 + a > ay_0$. If $y_0 > 1$, the claim $2y_0 + a > ay_0$ rearranges to $a < 2y_0/(y_0-1)$; the function $y_0/(y_0-1)$ is strictly decreasing for $y_0 > 1$ (its derivative is $-1/(y_0-1)^2 < 0$) with limit 1 as $y_0 \to \infty$, so $2y_0/(y_0-1) > 2 \geq a$ for every $y_0 > 1$. Either way, $2y_0 + a > ay_0$. Chaining the bounds:

$$\text{perimeter}(R) \;\geq\; 2y_0 + a \;>\; a\, y_0 \;\geq\; \text{area}(R),$$

so perimeter strictly exceeds area for every admissible f when $a \leq 2$: no k can equal both.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches agree the construction for $a > 2$ is easy and the real content is the impossibility for $a \leq 2$. They split on how to bound the perimeter of an arbitrary continuous boundary curve. The natural attempt reaches for the calculus machinery built for smooth curves – the arc-length integral – and tries to make it do exact, tight work. The model chain instead asks only for a *lower* bound on length, and gets one for free from the single fact that any path is at least as long as the straight line between its endpoints.

> *I only need an inequality, not an exact perimeter – so I should stop trying to compute or closely estimate the length of the curve, and instead ask what is the crudest true lower bound on that length that costs me nothing in generality.*

> *Any path between two points has length at least the straight-line distance between them, and this needs zero assumptions about smoothness – so if I route the boundary through the single highest point P on the curve, I get a perimeter lower bound in terms of P's height alone, for absolutely any continuous f.*

> *The matching move on the other side is equally crude: the whole region sits under the horizontal line at the maximum height, so area is at most base times that one height – I don't need to integrate f precisely, I need an upper bound, and width times max height is the cheapest one available.*

Layer 3 THE WHY

Caveat. The algebra comparing $2y_0 + a$ against ay_0 for $a \leq 2$, split into the cases $y_0 \leq 1$ and $y_0 > 1$, I verified independently by direct calculus (monotonicity of $y_0/(y_0 - 1)$) rather than by quoting Kedlaya-Ng's terser phrasing; it matches their stated conclusion.

Why does the model chain survive contact with a merely continuous, possibly nowhere-differentiable f, while the natural attempt does not? Because the two bounds it uses are not approximations of the true perimeter and area – they are structural inequalities that hold by definition, independent of any regularity of f. "The region lies under its own maximum height" is not a calculus fact, it is what "maximum" *means*. "A path's length is at least the distance between its endpoints" is not a calculus fact either, it is what "length" has to mean for the word to be coherent at all, whether or not the curve is rectifiable in a technical sense. The problem's phrasing – "continuous function," not "differentiable" – is itself the cue that any intended solution must be blind to smoothness, and the only tools that are automatically blind to smoothness are ones phrased purely in terms of distances and suprema, never derivatives. Once you notice that the maximum point P is the one feature of f guaranteed to exist by continuity alone (extreme value theorem) and nothing more, routing *both* bounds through P stops looking like a trick and starts looking like the only kind of argument the hypotheses actually license. The strict inequality at the end is not an accident of the specific numbers either: it is exactly the boundary case $a = 2$ that separates "achievable" from "not," which is why the constant-function construction needed $a > 2$ strictly – the two halves of the problem are proving the same threshold from opposite sides.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "f is continuous" with no differentiability assumed, sitting next to a question about perimeter and area. Write down immediately: whatever bound I use on the boundary's length must not depend on f being smooth.

2. That cue suggests bounding length by the triangle inequality (straight-line distance $\leq$ path length) rather than by any calculus formula for arc length.

3. Hesitation occurs when you first see "perimeter" and reflexively reach for $\int \sqrt{1 + f'^2}$; that reflex needs to be talked down by re-reading the hypothesis and noticing f was never assumed differentiable, so that formula may not even apply.

4. The trap is assuming a competition problem about area and perimeter must intend smooth, calculus-friendly curves. It doesn't: the intended solution is regularity-free by design, and reaching for calculus here is solving a harder, better-specified problem than the one actually asked.

5. To recognise this faster: whenever a problem states "continuous" pointedly (rather than "differentiable" or "smooth"), and asks about length or perimeter, expect the intended bound to come from the triangle inequality applied to a small number of distinguished points, not from an integral formula.

6. The reusable rule: to disprove that two awkward-to-compute quantities (like length and area of an arbitrary continuous region) can be equal, find one distinguished point (an extremum, guaranteed to exist by a basic theorem like extreme value) and bound both quantities crudely in terms of that single point – then compare the crude bounds. Exact computation is unnecessary; strict inequality of the bounds is enough.

2004 B4

2004 B4. Let n be a positive integer, $n \geq 2$, and put $\theta = 2\pi/n$. Define points $P_k = (k, 0)$ in the xy-plane, for $k = 1, 2, \ldots, n$. Let R_k be the map that rotates the plane counterclockwise by the angle θ about the point P_k. Let R denote the map obtained by applying, in order, R_1,

then $R_2, \ldots,$ then R_n. For an arbitrary point (x, y), find, and simplify, the coordinates of $R(x, y)$.

Fields: geometry of plane isometries; complex numbers; finite sums of roots of unity **Position:** B4 — mid-paper, but one of the gentlest problems on the 2004 paper once the right coordinates are chosen; the cost is entirely in the choice **Answer:** $R(x, y) = (x + n,\ y)$: R is translation by n units to the right.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student writes a rotation about P_k in real coordinates: translate P_k to the origin, rotate, translate back,

$$R_k \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x-k \\ y \end{pmatrix} + \begin{pmatrix} k \\ 0 \end{pmatrix}.$$

Writing M for the rotation matrix, two steps give $M^2 v - M^2\binom{1}{0} + M\binom{1}{0} - M\binom{2}{0} + \binom{2}{0}$, and after n steps

$$R(v) = M^n v + \sum_{k=1}^{n} k\,(I - M)M^{n-k}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

M^n is rotation by 2π, so $M^n = I$: that part is fine. The remaining vector has components that are sums like $\sum_{k=1}^{n} k\cos\big((n-k)\theta\big)$ and $\sum_{k=1}^{n} k\sin\big((n-k)\theta\big)$, then hit by the matrix $I - M$. The student now tries product-to-sum formulae on $\sum_j j\cos j\theta$, or differentiates $\sum_j \sin j\theta$ with respect to θ (illegal here: θ is fixed at $2\pi/n$, and the closed form of $\sum \sin j\theta$ has a $0/0$ at exactly that value). An hour later there is a page of trigonometry and a suspicion that the answer is $(x + n, y)$ from trying $n = 2$ by hand. That is the wall: the structure is right, but the vector sum is being computed in coordinates that hide the fact that it is a geometric-type series.

The model chain. **Step 1: complex coordinates.** Identify (x, y) with $z = x + iy$, so P_k is the real number k. Put $\zeta = e^{i\theta} = e^{2\pi i/n}$. Counterclockwise rotation by θ about k is $z \mapsto k + \zeta(z - k)$, i.e.

$$R_k(z) = \zeta z + k(1 - \zeta).$$

Step 2: the composite is a translation. Each R_k has the form $z \mapsto \zeta z + (\text{const})$, and composing two such maps multiplies the linear coefficients. So $R(z) = \zeta^n z + T = z + T$ for a constant $T \in \mathbb{C}$, since $\zeta^n = 1$. By induction on the number of maps applied,

$$T = (1 - \zeta)\sum_{k=1}^{n} k\,\zeta^{n-k} = (1 - \zeta)\sum_{j=0}^{n-1}(n - j)\,\zeta^j \qquad (j = n - k).$$

(Induction step: if $R_m \circ \cdots \circ R_1(z) = \zeta^m z + (1 - \zeta)\sum_{k=1}^{m} k\zeta^{m-k}$, applying R_{m+1} multiplies by ζ and adds $(m + 1)(1 - \zeta) = (1 - \zeta)(m + 1)\zeta^0$, which is the formula for $m + 1$.)

Step 3: telescope. Write $S = \sum_{j=0}^{n-1}(n - j)\zeta^j$. Then

$$S - \zeta S = \sum_{j=0}^{n-1}(n - j)\zeta^j - \sum_{j=1}^{n}(n - j + 1)\zeta^j = n - \sum_{j=1}^{n}\zeta^j,$$

because for $1 \le j \le n - 1$ the coefficients $(n - j) - (n - j + 1) = -1$, the $j = 0$ term is n, and the $j = n$ term of the second sum is ζ^n with coefficient 1. Since $n \ge 2$, $\zeta \ne 1$, and $\sum_{j=1}^{n}\zeta^j = \zeta\,\frac{\zeta^n - 1}{\zeta - 1} = 0$. Hence $T = (1 - \zeta)S = n$.

Step 4: answer. $R(z) = z + n$, that is, $R(x, y) = (x + n, y)$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on the shape: rotate about P_k means translate, rotate, translate back, and the composite has linear part rotation by $n\theta = 2\pi$. They split at the first line: (N) encodes a rotation as a 2×2 matrix and the offsets as vectors; (M) encodes the rotation as multiplication by the single number ζ. In (N) the offset sum is a pair of trigonometric sums weighted by k; in (M) it is one arithmetico-geometric sum $\sum(n-j)\zeta^j$, and the standard move for those — multiply by the ratio and subtract — finishes in two lines.

> *Every map here is a rotation about a point on the real axis, and the angle is $2\pi/n$ — that is an n-th root of unity staring at me. Rotation in the plane is multiplication by $e^{i\theta}$, so I work in $\mathbb{C}$, where the whole composite is $z \mapsto \zeta^n z + \text{constant}$.*

> *The linear part is $\zeta^n = 1$, so R is a translation and I only need one vector. The vector is $\sum k\zeta^{n-k}$ times $(1-\zeta)$: an arithmetic coefficient times a geometric term. Multiply by ζ and subtract — the $(1-\zeta)$ is already there, doing it for me.*

> *Sanity check with a picture: a regular n-gon with unit sides rolling along the line. After n rolls it is back in its original orientation, one full perimeter n further along.*

- ▷ **Cue:** a composition of rotations (or similarities) in the plane, with angles that are multiples of $2\pi/n$ $\longrightarrow$ **Move:** identify the plane with $\mathbb{C}$; each map is $z \mapsto \zeta z + c$; the composite is $z \mapsto \zeta^{\text{total}} z + C$
- ▷ **Cue:** the rotation angles add up to a multiple of 2π $\longrightarrow$ **Move:** the composite is a translation; compute its vector from one point, or from the constant term only
- ▷ **Cue:** a sum $\sum_j (\text{linear in } j)\, r^j$, possibly already multiplied by $(1-r)$ $\longrightarrow$ **Move:** multiply by r and subtract; the linear coefficients telescope to a plain geometric sum
- ▷ **Cue:** a sum of all n-th roots of unity $\longrightarrow$ **Move:** it is 0 for $n \geq 2$

Layer 3 THE WHY

Why is $\mathbb{C}$ the natural coordinate system and not a trick? Go to the definition. A rotation by θ about the origin is the map that preserves distances to the origin and adds θ to every argument. In polar form $z = re^{i\phi}$, that is literally $re^{i(\phi+\theta)} = e^{i\theta}z$: multiplication by one complex number. The 2×2 matrix is the same map with its real and imaginary parts written out separately, and writing them out separately is what turned one geometric series into two trigonometric sums. The problem's own data point at $\mathbb{C}$ twice: all centres lie on one line (the real axis), and the angle is $2\pi/n$, the argument of a primitive n-th root of unity. When the angle is "2π divided by the number of steps", the question is asking you to use $\zeta^n = 1$.

Why is the composite a translation? Because the linear part of $z \mapsto \zeta z + c$ is the multiplier ζ, and multipliers multiply under composition — the constants never feed back into the multiplier. So the only thing that can survive n steps is $\zeta^n = 1$, and a map $z \mapsto z + T$ is a translation. Geometrically: angles of rotations add, and a total turn of 2π is no turn. That reduces the problem from "find a map" to "find a vector".

Why does the vector come out to n? The telescoping of Step 3 is the algebra; the rolling polygon of Kedlaya and Ng's second solution is the reason. Place a regular n-gon with unit sides below the x-axis, with vertices $V_0 = 0$ and $V_1 = 1$. Its interior angle is $\pi - \theta$, so the edge V_1V_2 leaves V_1 at direction $2\pi - \theta$; rotating counterclockwise by θ about $V_1 = P_1$ lays it along the axis, $V_2 \mapsto 2 = P_2$. The polygon has rolled one edge. Repeating, after $R_1, \ldots, R_k$ the vertex V_k sits at $P_k = k$ and V_{k+1} at $k+1$. After all n maps, $V_n = V_0$ sits at n and V_1 at $n+1$. An orientation-preserving isometry is fixed by where it sends two distinct points, so R agrees with $z \mapsto z + n$ everywhere. The centres $P_1, \ldots, P_n$ were placed a unit apart precisely so that they are

the successive contact points of a rolling polygon with unit sides — the problem is a disguised wheel, and a wheel that has made one full turn has moved its perimeter.

Kedlaya's first solution ends "Expanding the product $(1-\zeta)(\zeta^{n-1}+2\zeta^{n-2}+\cdots+n)$ yields $-\zeta^n-\cdots-\zeta+n=n$." Here is where that comes from: the factor $(1-\zeta)$ is the "multiply by the ratio and subtract" step already performed, because each rotation about k contributes $k(1-\zeta)$. You do not have to invent the telescoping; the problem gives it to you, provided you do not expand the $(1-\zeta)$ away.

The natural attempt did no wrong mathematics. It chose coordinates in which the structure (one multiplier, a geometric series) is invisible, and then paid for that choice in trigonometry. That is the zombie maths this problem punishes.

Layer $+\alpha$ BACKTESTING

1. The first cue is "rotates ... by the angle θ" with $\theta = 2\pi/n$ and n rotations. On seeing it, write $R_k(z) = \zeta z + k(1-\zeta)$ with $\zeta = e^{2\pi i/n}$, and note $\zeta^n = 1$.

2. Complex affine maps: compose $z \mapsto \zeta z + c_k$, observe the linear part is $\zeta^n = 1$, so R is a translation, and compute the constant as an arithmetico-geometric sum.

3. Hesitation comes at the sum $\sum k\zeta^{n-k}$: it looks like it needs a differentiated geometric-series formula. It is reinforced away by keeping the factor $(1-\zeta)$ attached and telescoping, and by the rolling-polygon picture, which predicts the answer before any algebra.

4. The trap is real-coordinate matrices, which split one complex sum into two trigonometric sums weighted by k. A second trap is differentiating $\sum \sin j\theta$ in θ, which is invalid at the fixed value $\theta = 2\pi/n$ where the closed form is singular.

5. Ask, before choosing coordinates: *what is the symmetry of each map, and which number system turns that symmetry into one multiplication?* Rotations become multiplication in $\mathbb{C}$; total angle a multiple of 2π means translation.

6. Reusable decision rule: *for a composite of affine maps $x \mapsto \lambda x + c$, the multipliers multiply and the constants form a weighted geometric sum; first decide what the product of multipliers is (often 1, giving a translation), then evaluate the constant by multiply-by-ratio-and-subtract, never term by term.*

2004 B5

2004 B5. Evaluate

$$\lim_{x\to 1^-} \prod_{n=0}^{\infty} \left(\frac{1+x^{n+1}}{1+x^n}\right)^{x^n}.$$

Fields: real analysis (infinite products and series, summation by parts, Riemann sums) **Position:** B5 — genuinely hard in the arena: the product has no closed form, and every factor tends to 1, which invites a wrong answer **Answer:** The limit is $2/e$.

Layer 1 THE WHAT — surface analysis

The natural attempt. First reflex: each factor tends to $\left(\frac{2}{2}\right)^1 = 1$ as $x \to 1^-$, so "the limit is 1". The student distrusts that (infinitely many factors, each near 1, can multiply to anything) and takes logarithms instead. Write $a_n = \ln(1+x^n)$ and

$$\ln \Pi(x) = \sum_{n=0}^{\infty} x^n \big(a_{n+1} - a_n\big).$$

Now the natural move is Taylor expansion: $a_{n+1} - a_n = \ln \frac{1+x^{n+1}}{1+x^n} = \ln\left(1 - \frac{x^n(1-x)}{1+x^n}\right) \approx -\frac{x^n(1-x)}{1+x^n}$, so

$$\ln \Pi(x) \approx -(1-x) \sum_{n\geq 0} \frac{x^{2n}}{1+x^n}.$$

This is actually a promising line — it is a Riemann sum in disguise — but the student who has reached it usually tries to evaluate $\sum x^{2n}/(1+x^n)$ in closed form, expands $\frac{1}{1+x^n}$ as a geometric series, gets a double sum $\sum_n \sum_m (-1)^m x^{(m+2)n}$, swaps, and faces $\sum_m \frac{(-1)^m}{1-x^{m+2}}$, which blows up term by term as $x \to 1$. Multiplying back by $(1-x)$ repairs each term but now the interchange of limit and sum needs uniform-convergence arguments for an alternating series. Worse, the "$\approx$" in the Taylor step was never controlled: the error is $O\big((1-x)^2 x^{2n}\big)$ summed over infinitely many n, which is $O(1-x)$ and does vanish — but only if you check it. The attempt can be pushed through (it is essentially Kedlaya and Ng's first solution), at a cost of a page of justification and several opportunities for a sign error.

The model chain. Fix $0 < x < 1$ and write $a_n = \ln(1+x^n)$, so $0 < a_n \le x^n$ for $n \ge 1$ and $a_0 = \ln 2$. Every series below converges absolutely because its terms are bounded by a constant times x^n.

Step 1: logarithm and summation by parts. The logarithm of the product is

$$\ell(x) = \sum_{n=0}^{\infty} x^n a_{n+1} - \sum_{n=0}^{\infty} x^n a_n = \sum_{n=1}^{\infty} x^{n-1} a_n - \sum_{n=0}^{\infty} x^n a_n = -\ln 2 + \sum_{n=1}^{\infty} \big(x^{n-1} - x^n\big) a_n.$$

(Splitting into two sums is legitimate because both converge.) So

$$\ell(x) = -\ln 2 + \frac{1}{x} \sum_{n=1}^{\infty} \big(x^n - x^{n+1}\big) \ln\big(1+x^n\big).$$

Step 2: recognise a Riemann sum. Put $t_n = x^n$ and $g(t) = \ln(1+t)$. The points $1 = t_0 > t_1 > t_2 > \cdots \to 0$ partition $(0,1]$ into intervals $[t_{n+1}, t_n]$ of length $t_n - t_{n+1} = x^n(1-x) \le 1-x$. The sum in Step 1 is

$$U(x) = \sum_{n=1}^{\infty} (t_n - t_{n+1})\, g(t_n),$$

a Riemann sum for $\int_0^{t_1} g$ using right endpoints.

Step 3: squeeze. g is increasing, so on $[t_{n+1}, t_n]$ we have $g(t_{n+1}) \le g \le g(t_n)$. Integrating and summing over $n \ge 1$ (a countable union of intervals covering $(0, t_1]$),

$$L(x) := \sum_{n \ge 1} (t_n - t_{n+1}) g(t_{n+1}) \;\le\; \int_0^x \ln(1+t)\, dt \;\le\; U(x).$$

The gap is

$$U(x) - L(x) = \sum_{n \ge 1} (t_n - t_{n+1}) \big(g(t_n) - g(t_{n+1})\big) \le (1-x) \sum_{n \ge 1} \big(g(t_n) - g(t_{n+1})\big) = (1-x)\ln(1+x),$$

by telescoping (the partial sums are $g(t_1) - g(t_{N+1}) \to g(x) - g(0)$). Hence $0 \le U(x) - \int_0^x \ln(1+t)\, dt \le (1-x)\ln 2 \to 0$, and since $\int_0^x \ln(1+t)\, dt \to \int_0^1 \ln(1+t)\, dt$,

$$\lim_{x \to 1^-} U(x) = \int_0^1 \ln(1+t)\, dt = \big[(1+t)\ln(1+t) - t\big]_0^1 = 2\ln 2 - 1.$$

Step 4: assemble. $\ell(x) = -\ln 2 + U(x)/x \to -\ln 2 + 2\ln 2 - 1 = \ln 2 - 1$. The exponential is continuous, so the product tends to $e^{\ln 2 - 1} = 2/e$. □

Caveat. Numerical check: exponentiating the logarithmic series, summed to machine precision, gives the product values 0.72389, 0.73462, 0.73565, 0.73575 at $x = 0.9, 0.99, 0.999, 0.9999$, converging to $2/e = 0.735759\ldots$ at rate roughly proportional to $1-x$, as Step 3's error bound predicts.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains take logarithms and arrive at $\sum_n x^n(a_{n+1} - a_n)$. They split at the next line. (N) expands the *difference* $a_{n+1} - a_n$ by Taylor series, producing a new sum with approximation errors and no visible structure. (M) moves the difference off the a's and onto the weights x^n — summation by parts — producing $(x^{n-1} - x^n)\, a_n$, which is a width times a height: a Riemann sum.

> *I have a sum of weight times* difference*:* $\sum x^n(a_{n+1} - a_n)$*. That's the discrete version of* $\int u\, dv$*. Integrate by parts — shift the index on the first half — so the difference lands on* x^n*, where it is simple:* $x^{n-1} - x^n = x^{n-1}(1 - x)$*.*

> *Now I have* $\sum(x^n - x^{n+1}) \ln(1 + x^n)$*. The factor* $x^n - x^{n+1}$ *is the gap between consecutive points* $x^{n+1} < x^n$*, and it goes to* 0 *as* $x \to 1$*. Gap times function value at a point: that's a Riemann sum for* $\int_0^1 \ln(1 + t)\, dt$ *with partition points* $t = x^n$*.*

> *Each factor tends to* 1*, but there are more and more factors that matter as* $x \to 1$*. So the limit is a competition, and a competition between "many" and "small" is what an integral measures.*

- ▷ **Cue:** a sum of the form $\sum w_n(a_{n+1} - a_n)$ ⟶ **Move:** summation by parts: rewrite as $-w_0 a_0 + \sum(w_{n-1} - w_n)a_n$ (plus a boundary term at infinity, zero here)
- ▷ **Cue:** $\lim_{x\to 1^-}(1 - x)\sum_n x^n f(x^n)$, or $\sum(x^n - x^{n+1})f(x^n)$ ⟶ **Move:** a Riemann sum with geometric partition $t = x^n$; the limit is $\int_0^1 f(t)\, dt$
- ▷ **Cue:** an infinite product whose exponent or factors depend on the limit variable ⟶ **Move:** take logarithms first and turn it into a series
- ▷ **Cue:** a Riemann sum over a non-uniform partition of a monotone function ⟶ **Move:** squeeze between upper and lower sums; their gap is at most (mesh)×(total variation)

Layer 3 THE WHY

Why does summation by parts come to mind? Go to what the log of the product is: $\sum x^n\, \Delta a_n$, a weight times a forward difference. That is structurally $\int u\, dv$. Integration by parts exists to move the derivative from the factor you cannot handle to the one you can. Here $\Delta a_n = \ln \frac{1+x^{n+1}}{1+x^n}$ is ugly (a log of a ratio), while $\Delta(x^n) = x^n(x - 1)$ is a single term. So you move the difference onto x^n. The Taylor expansion in the natural attempt tries to simplify Δa_n directly, which is attacking the harder factor.

Why is a Riemann sum the right lens once you have $\sum(x^n - x^{n+1}) \ln(1 + x^n)$? Look at what $x \to 1^-$ does to the points x^n: they fill $(0, 1]$ more and more densely, with consecutive gaps $x^n(1 - x) \to 0$. A sum of "gap × function value at a grid point" over a grid whose mesh shrinks to zero is, by definition, what an integral is the limit of. The problem is not asking you to evaluate a series in closed form; it is an integral wearing a series costume. The telltale is the weight x^n in the exponent of the original product: it was put there precisely so that after summation by parts it becomes the width $x^{n-1} - x^n$. Without that exponent the problem would be a telescoping product, $\prod \frac{1+x^{n+1}}{1+x^n} = \frac{1}{2}$ for every x, and nothing would happen as $x \to 1$.

That last observation is the key to why the naive "each factor $\to 1$" is wrong and what replaces it. Without the exponents the product telescopes to $1/2$. The exponents x^n are close to 1 for small n and close to 0 for large n, so they do not telescope, and the defect is measured by how the weights x^n vary — exactly the $(x^{n-1} - x^n)$ of summation by parts. The answer $2/e$ is $\frac{1}{2}$ (the telescoped part, $-\ln 2$ in the log) corrected by $e^{2\ln 2 - 1} = 4/e$ (the integral).

Kedlaya and Ng's second solution (due to Greg Price) sets $t_n = \ln(1+x^n)$ and reads the sum as a Riemann sum for $\int_0^{\ln 2}(1-e^t)\,dt$. Here is where that comes from: it is the same summation-by-parts idea with the roles of the two factors swapped — the difference kept on $\ln(1+x^n)$ and the function value $x^n = e^{t_n} - 1$. Either choice of partition gives $\ln 2 - 1$; the one above has the simpler integrand. Their first solution expands $\ln(1+x^n)$ in a power series and swaps sums, which is correct but needs an absolute-convergence argument, a uniform-convergence argument, and care with a conditionally convergent alternating series at the end. That is the price of not seeing the integral.

Why the squeeze and not "Riemann sums converge"? The textbook theorem is for finite partitions of a closed interval; here the partition is infinite and accumulates at 0. Monotonicity gives a self-contained proof in three lines, with an explicit error $(1-x)\ln 2$, so there is nothing to cite and nothing to worry about.

Layer $+\alpha$ BACKTESTING

1. The first cue is an infinite product with exponent x^n and a limit $x \to 1^-$. On seeing it, write the logarithm $\sum x^n\big(\ln(1+x^{n+1}) - \ln(1+x^n)\big)$ and note it is weight times difference.
2. Summation by parts, then read $\sum(x^n - x^{n+1})f(x^n)$ as a Riemann sum with partition points x^n, giving $\int_0^1 \ln(1+t)\,dt$.
3. Hesitation occurs at the Riemann-sum identification: the partition is infinite and not uniform, and the student is unsure the limit is the integral. It is reinforced away by the upper/lower-sum squeeze, whose gap is at most mesh times total variation, $(1-x)\ln 2$.
4. The trap is "each factor tends to 1, so the product tends to 1". The second trap is Taylor-expanding the difference $a_{n+1} - a_n$, which produces uncontrolled errors and a double series that must be interchanged with care.
5. Recognise it faster by asking of any $\lim_{x\to 1^-}$ of a series in x^n: *is* $(1-x)$ *or* $x^n - x^{n+1}$ *hiding in here? If so, the points* x^n *are a partition of* $(0,1)$ *and the limit is an integral.*
6. Reusable decision rule: *when a sum has the form weight* $\times$ *difference, move the difference to whichever factor has the simpler difference (summation by parts). When a limit makes a grid of sample points fill an interval with vanishing mesh, stop evaluating the sum and compute the integral it approximates, justified by an upper/lower squeeze for monotone integrands.*

2004 B6

2004 B6. Let $\mathcal{A}$ be a non-empty set of positive integers, and let $N(x)$ denote the number of elements of $\mathcal{A}$ not exceeding x. Let $\mathcal{B}$ denote the set of positive integers b that can be written in the form $b = a - a'$ with $a \in \mathcal{A}$ and $a' \in \mathcal{A}$. Let $b_1 < b_2 < \cdots$ be the members of $\mathcal{B}$, listed in increasing order. Show that if the sequence $b_{i+1} - b_i$ is unbounded, then

$$\lim_{x\to\infty} N(x)/x = 0.$$

Fields: additive combinatorics (difference sets, density); elementary counting **Position:** B6 — the last problem of the paper and genuinely hard in the arena; the proof, once found, is short, and everything hinges on turning "$b \notin \mathcal{B}$" into "two translates of $\mathcal{A}$ are disjoint" **Answer:** Proof: for every k one constructs $c_1 < \cdots < c_k$ with no pairwise difference in $\mathcal{B}$; the translates $\mathcal{A} + c_i$ are then disjoint, forcing $\limsup N(x)/x \le 1/k$.

Layer 1 THE WHAT — surface analysis

The natural attempt. A strong student starts with the density question directly: suppose $N(x)/x \not\to 0$, so $N(x) > \delta x$ for some $\delta > 0$ and arbitrarily large x, and try to show the differences are then dense enough to have bounded gaps. By pigeonhole, among the elements of $\mathcal{A}$ up to x there are many pairs at distance at most $1/\delta$, so some small difference occurs —

but that says nothing about *large* differences, and the hypothesis is about gaps among large differences. The student then tries to show every long interval $[y, y+L]$ contains a difference, i.e. that $\mathcal{A}$ meets $\mathcal{A}+t$ for some $t \in [y, y+L]$: the sets $\mathcal{A}+t$ for $t = y, \dots, y+L$ are $L+1$ sets each of density $\geq \delta$ along a subsequence, so if $(L+1)\delta > 1$ two of them intersect. That gives $a + t = a' + t'$, a difference $t - t'$ *between two shifts*, which is at most L — again small, not near y. The wall: the counting produces differences of the size of the window, never differences located in a prescribed far-away window. The attempt is aiming the pigeonhole at the wrong family of sets.

The model chain. If $\mathcal{A}$ is finite, $N(x)$ is bounded and $N(x)/x \to 0$. So assume $\mathcal{A}$ is infinite; then $\mathcal{B}$ is infinite (it contains $a - a_{\min}$ for every $a \neq a_{\min}$), and the hypothesis says: for every M there are M consecutive positive integers none of which lies in $\mathcal{B}$. (Take a gap $b_{i+1} - b_i > M$; the integers $b_i + 1, \dots, b_i + M$ avoid $\mathcal{B}$.)

Step 1: one excluded difference means two disjoint translates. For an integer $c \geq 0$ write $\mathcal{A} + c = \{a + c : a \in \mathcal{A}\}$. If $c < c'$ and $c' - c \notin \mathcal{B}$, then $(\mathcal{A} + c) \cap (\mathcal{A} + c') = \varnothing$: a common element $a + c = a' + c'$ would give $a - a' = c' - c \in \mathcal{B}$.

Step 2: disjoint translates cap the density. Suppose $0 \leq c_1 < c_2 < \cdots < c_k$ are integers with $c_j - c_i \notin \mathcal{B}$ for all $i < j$. By Step 1 the sets $\mathcal{A} + c_i$ are pairwise disjoint subsets of the positive integers. The elements of $\mathcal{A} + c_i$ not exceeding x number $N(x - c_i)$, and they all lie in $\{1, \dots, \lfloor x \rfloor\}$, so

$$\sum_{i=1}^{k} N(x - c_i) \leq x.$$

Removing the interval $(x - c_i, x]$ removes at most c_i integers, so $N(x - c_i) \geq N(x) - c_i$. Hence $kN(x) - \sum_i c_i \leq x$, i.e.

$$\frac{N(x)}{x} \leq \frac{1}{k} + \frac{c_1 + \cdots + c_k}{kx}, \qquad \text{so} \qquad \limsup_{x \to \infty} \frac{N(x)}{x} \leq \frac{1}{k}.$$

Step 3: build arbitrarily long such lists. Start with $c_1 = 0$. Given $0 = c_1 < \cdots < c_k$ with all pairwise differences outside $\mathcal{B}$, use the hypothesis to find a run of $c_k + 1$ consecutive positive integers $u, u+1, \dots, u + c_k$ none of which lies in $\mathcal{B}$, and set $c_{k+1} = u + c_k$. Then $c_{k+1} > c_k$ (as $u \geq 1$), and for each $i \leq k$, $0 \leq c_i \leq c_k$ gives

$$c_{k+1} - c_i = u + (c_k - c_i) \in \{u, \dots, u + c_k\},$$

so $c_{k+1} - c_i \notin \mathcal{B}$. The old differences are unchanged. By induction, lists of every length k exist.

Step 4: conclude. Steps 2 and 3 give $\limsup N(x)/x \leq 1/k$ for every k, so $\limsup N(x)/x = 0$. Since $N(x)/x \geq 0$, the limit exists and equals 0. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains accept the reformulation "$\mathcal{B}$ has arbitrarily long runs of non-members". They split on what a non-member *is*. (N) treats $b \notin \mathcal{B}$ as a fact about $\mathcal{A}$ to be contradicted by counting pairs inside $\mathcal{A}$. (M) reads $b \notin \mathcal{B}$ as "$\mathcal{A}$ and $\mathcal{A} + b$ do not overlap", i.e. as a packing statement: disjoint copies of $\mathcal{A}$ must fit inside $\mathbb{Z}_{>0}$, and k disjoint copies of a set of density δ need $k\delta \leq 1$.

> *$b \notin \mathcal{B}$ means no two elements of $\mathcal{A}$ are b apart. Said the other way round: if I slide $\mathcal{A}$ by b, it lands entirely in the holes of $\mathcal{A}$. So one missing difference already gives me two disjoint copies of $\mathcal{A}$, and density at most $1/2$.*

> *To get density $1/k$ I want k disjoint copies, $\mathcal{A}+c_1,\dots,\mathcal{A}+c_k$. That needs every difference c_j-c_i to be missing from $\mathcal{B}$ — not merely $k-1$ missing numbers, but all $\binom{k}{2}$ differences.*

> *Long runs of non-members are exactly what lets me add one more shift. If I already have shifts in $[0,c_k]$, a new shift c has differences $c-c_i$ filling a window of width c_k ending at c; so I need a run of length c_k+1 avoiding $\mathcal{B}$, and the hypothesis hands me one.*

- ▷ **Cue:** "b is not a difference of two elements of $\mathcal{A}$" $\longrightarrow$ **Move:** rewrite as $\mathcal{A}\cap(\mathcal{A}+b)=\varnothing$: translates are disjoint
- ▷ **Cue:** k pairwise disjoint translates of a set inside $\{1,\dots,x\}$ $\longrightarrow$ **Move:** density at most $1/k$: add up the counts, $\sum N(x-c_i)\le x$
- ▷ **Cue:** need a set of shifts whose pairwise differences all avoid a set $\mathcal{B}$, and $\mathcal{B}$ has arbitrarily long gaps $\longrightarrow$ **Move:** build the shifts greedily: each new shift sits at the top of a gap longer than the largest shift so far
- ▷ **Cue:** show a density is 0 $\longrightarrow$ **Move:** show lim sup $\le 1/k$ for every k, one k at a time

Layer 3 THE WHY

Why is "translates are disjoint" the natural reading and not a trick? Go to the definition of $\mathcal{B}$: $b\in\mathcal{B}$ iff $a=a'+b$ for some $a,a'\in\mathcal{A}$, iff $\mathcal{A}$ and $\mathcal{A}+b$ share an element. That is a line of logic, not a rewriting chosen for convenience: membership of b in the difference set *is* the statement that $\mathcal{A}$ meets its own b-translate. So non-membership is disjointness, and disjointness of sets is the most direct route to an upper bound on size, because sizes of disjoint sets add. The conclusion to be proved is an upper bound on the size of $\mathcal{A}$; the hypothesis, read this way, supplies disjoint copies of $\mathcal{A}$. The two are made for each other.

Why do we need all pairwise differences and not only consecutive ones? Because disjointness of k sets is a pairwise condition: $\mathcal{A}+c_1$ and $\mathcal{A}+c_3$ can collide even if both avoid $\mathcal{A}+c_2$. This is where the natural attempt goes wrong in a different way when it gets close: the student collects k separate missing differences $b^{(1)},\dots,b^{(k)}$ and hopes for density $1/(k+1)$, but $\mathcal{A}+b^{(1)}$ and $\mathcal{A}+b^{(2)}$ collide whenever $b^{(2)}-b^{(1)}\in\mathcal{B}$. The hypothesis is about *runs* of missing differences precisely because one needs a whole window of them: the new shift must avoid $\mathcal{B}$ at k different distances simultaneously, and those distances form a translate of the old shift set, which lies in an interval of length c_k.

Why does the natural attempt stall? It uses the hypothesis about $\mathcal{B}$ only through a contradiction at the very end, so all its counting happens inside $\mathcal{A}$, where differences are forced to be small (pigeonhole finds close pairs). The information in the problem is about *large* missing differences. The model uses that information at the start, as a tool that manufactures disjoint sets, so the large size of the shifts costs nothing: in Step 2 the shifts only contribute $\sum c_i/x\to 0$.

Kedlaya and Ng's first solution (after Dan Bernstein) builds a mixed-radix system $b_0=1,b_1,\dots,b_n$ with $b_i/(2b_{i-1})$ an integer and all $|\sum e_ib_i|\notin\mathcal{B}$, then partitions the integers into blocks of 2^n elements that pairwise differ by excluded amounts. Here is where that comes from: it is the same packing idea, run inside $\mathcal{A}$ rather than on translates of it, with the shifts chosen as subset sums so that the clumps tile the line. The translate version in their closing remark (the one given above) does the same job with less bookkeeping: translates of $\mathcal{A}$ need not tile anything, because disjointness alone gives the count. Their second solution (Paul Pollack) is extremal: take $L=\sup\limsup N(x)/x$ over all admissible $\mathcal{A}$, pick $\mathcal{A}$ nearly attaining it, and show $\mathcal{A}\cup(\mathcal{A}+m)$ for some $m\notin\mathcal{B}$ is still admissible (its difference set lies in $\mathcal{B}\cup(\mathcal{B}\pm m)\cup\{1,\dots,m\}$, whose gaps are still unbounded) but has density almost $2L$. That is the doubling step of the model chain, applied once and bootstrapped by extremality instead of by induction.

The difficulty of the problem is entirely in the reading of $b\notin\mathcal{B}$. After that, Steps 2 and 3 are routine counting and a greedy construction — which is why this is a hard B6 with a short proof.

Layer $+\alpha$ BACKTESTING

1. The first cue is the definition "$b = a - a'$ with $a, a' \in \mathcal{A}$" together with a hypothesis about numbers *not* of that form. On seeing it, write $b \notin \mathcal{B} \iff \mathcal{A} \cap (\mathcal{A} + b) = \varnothing$.

2. Packing: find k shifts with pairwise differences outside $\mathcal{B}$, so k disjoint translates of $\mathcal{A}$ fit in $\{1, \ldots, x\}$ up to an error of $\sum c_i$, giving density at most $1/k$.

3. Hesitation occurs when choosing the next shift: it must avoid $\mathcal{B}$ at k distances at once. It is reinforced away by noticing that those distances form a block of width c_k, so a run of $c_k + 1$ non-members of $\mathcal{B}$ — which the hypothesis provides — is exactly what is needed.

4. The trap is counting pairs inside $\mathcal{A}$ (pigeonhole), which only finds small differences, while the hypothesis concerns large ones. A second trap is using k unrelated missing differences as shifts, forgetting that disjointness is pairwise.

5. Recognise it faster by translating every statement about a difference set into a statement about overlaps of translates, before doing anything else: $t \in \mathcal{A} - \mathcal{A}$ *iff* $\mathcal{A}$ *meets* $\mathcal{A} + t$.

6. Reusable decision rule: *to bound the size of a set from above, look for disjoint copies of it; the hypothesis "this difference never occurs" produces disjoint translates, and k disjoint copies of a set of density δ force $k\delta \leq 1$. When the copies must be pairwise disjoint, build their positions greedily using the longest available windows of forbidden differences.*