DAWN PUTNAM SESSION

Putnam 1974 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1974 — Paper A **5**
1974 A1 5
1974 A2 7
1974 A3 10
1974 A4 12
1974 A5 15
1974 A6 17

Putnam 1974 — Paper B **21**
1974 B1 21
1974 B2 24
1974 B3 27
1974 B4 30
1974 B5 33
1974 B6 36

Putnam 1974 — Paper A

1974 A1

1974 A1. Call a set of positive integers "conspiratorial" if no three of them are pairwise relatively prime. (A set of integers is "pairwise relatively prime" if no pair of them has a common divisor greater than 1.) What is the largest number of elements in any "conspiratorial" subset of the integers 1 through 16?

Caveat. The MAA 1965–1984 OCR renders the range as "the integers I through 16," with the digit 1 misread as the letter I by the column scanner. The statement above restores the digit; no other wording or mathematical content was altered. This is confirmed against the book's own solution, which begins "A conspiratorial subset (CS) of $\{1, 2, \ldots, 16\}$" and proceeds to the answer 11.

Fields: Combinatorics, number theory (coprimality), pigeonhole **Position:** A1 (opener) **Answer:** 11; e.g. $\{2, 3, 4, 6, 8, 9, 10, 12, 14, 15, 16\}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The condition is about triples, so the obvious first move is to start building a big set by hand and checking it. You might notice quickly that even numbers are "safe" to pile in together, since no two of them are coprime, and start with $\{2, 4, 6, 8, 10, 12, 14, 16\}$ — eight numbers, no two coprime, so certainly no three pairwise coprime. Then you try to add more: 9 is fine, since $\gcd(9, 6) = 3 \neq 1$ and $\gcd(9, 12) = 3 \neq 1$, though $\gcd(9, 2) = 1$, $\gcd(9, 4) = 1$ — you have to check that adding 9 doesn't create a coprime pair that then pairs with a THIRD element to complete a coprime triple. Add 3, add 15: now you have eleven numbers and, after a fair amount of checking pairs and triples by hand, you believe none of them are pairwise coprime as a triple. You conjecture 11 is the answer. But you have no way to know, from this construction alone, whether 12 is achievable — maybe you just weren't clever enough in your search. Proving 12 is *impossible* by exhaustively ruling out every 12-element subset of $\{1, \ldots, 16\}$ is completely infeasible by hand: there are $\binom{16}{12} = 1820$ of them, and each requires checking up to $\binom{12}{3} = 220$ triples for pairwise coprimality. The construction gives a lower bound; it gives nothing that resembles a proof of optimality.

The model chain. Read the forbidden condition as a statement about triples and ask: what is the largest set of integers in $\{1, \ldots, 16\}$ that is itself *pairwise coprime* — call it the poison set T? Any three integers drawn from a pairwise coprime family are automatically pairwise coprime, so if a conspiratorial set S contained three elements of T, that would violate S's own defining property. Hence $|S \cap T| \leq 2$ for every conspiratorial S. The primes at most 16 are $2, 3, 5, 7, 11, 13$ — six of them — and together with 1 (which is coprime to everything) they form a pairwise coprime set $T = \{1, 2, 3, 5, 7, 11, 13\}$ with $|T| = 7$; moreover this is the largest such family in $\{1, \ldots, 16\}$, since every other integer in the range is composite with a prime factor already represented in

T, so it cannot be added to T without breaking pairwise coprimality. Since $|S \cap T| \leq 2$, S can contain at most 2 of T's 7 elements, hence must omit at least 5 of them, giving

$$|S| \leq 16 - (7 - 2) = 11.$$

For the matching construction, take $S = \{2, 3, 4, 6, 8, 9, 10, 12, 14, 15, 16\}$ — this is $\{1, \ldots, 16\}$ with $\{1, 5, 7, 11, 13\}$ removed, exactly the 5 "extra" elements of T beyond the allowed 2 (here $2, 3 \in T \cap S$). Every element of S is a multiple of 2 or a multiple of 3 (the multiples of 2 are $\{2, 4, 6, 8, 10, 12, 14, 16\}$, the remaining multiples of 3 are $\{3, 9, 15\}$, and this covers all 11 elements). So S is partitioned into exactly two classes by which of $2, 3$ divides it. Among any three elements of S, two must lie in the same class by pigeonhole (two classes, three elements), and two elements in the same class share that class's prime factor, so they are not coprime. Hence no three elements of S are pairwise coprime: S is conspiratorial, with $|S| = 11$, matching the bound exactly.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains open by noticing the problem forbids a structure inside triples. They split before any specific numbers are chosen, on the question of what to search for first: the natural attempt searches directly for a good S, treating the problem as a construction problem to be solved by trial and adjustment. The model attempt instead searches for the largest possible OFFENDING structure — a pairwise coprime family — inside the universe $\{1, \ldots, 16\}$ itself, and turns that into an upper bound before ever committing to a specific S.

> *If I can find seven numbers in this range that are pairwise coprime, then no conspiratorial set can hold more than two of them — so counting how many numbers avoid that pairwise-coprime family gives me an upper bound for free, before I ever have to check a specific eleven-element set by hand.*

Layer 3 THE WHY

The forbidden condition is stated as "no three of them are pairwise relatively prime" — a condition on the ABSENCE of a 3-element substructure inside S. Whenever a combinatorial condition forbids a k-element substructure of some type, the reflex that avoids brute enumeration is to ask how large the largest instance of that TYPE is in the whole universe, not in S: if T is a pairwise coprime subset of $\{1, \ldots, 16\}$ with $|T| = t$, then any S satisfying the forbidden-triple condition can intersect T in at most 2 elements, because 3 elements of T are pairwise coprime purely by virtue of belonging to T, with no reference to what else is in S. This is what makes the bound free: it costs nothing to derive once T is found, because T's pairwise coprimality is a property of T alone. Why is $T = \{1, 2, 3, 5, 7, 11, 13\}$ the largest such family, and why is that the right family to look for? Every integer n in $\{2, \ldots, 16\}$ has at least one prime factor $q \leq 13$, and a pairwise coprime family can contain at most one multiple of q for each prime q — otherwise those two multiples share q. So a pairwise coprime subset of $\{1, \ldots, 16\}$ is, at most, one representative per prime ≤ 16 plus the element 1: exactly $6 + 1 = 7$, achieved by taking the primes themselves (the "purest" representatives, sharing no factors with anything else in range). This is the same idea, mirrored: the poison set is built by picking ONE number per prime, and the matching construction S is built by picking numbers covered by only TWO primes (2 and 3) and taking ALL of them. The upper bound comes from a family with one member per obstruction; the achieving set comes from a partition into few enough classes that pigeonhole forces a collision. Both are the same coprimality structure read in two directions — forbidding versus covering — and seeing that duality is what turns a search problem into an extremal one that closes in two short arguments instead of one long, unverifiable construction.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "no three of them are pairwise relatively prime" — a forbidden-triple condition. On seeing it, write down: find the largest pairwise coprime

family in the ambient set $\{1, \ldots, 16\}$; that family is the enemy structure whose size controls everything.

2. That cue suggests a pigeonhole bound: a conspiratorial set can meet any pairwise coprime family in at most 2 elements (three would themselves be pairwise coprime), so $|S| \leq 16 - (|T| - 2)$ once the largest such family T is identified.

3. The hesitation comes after the bound 11 is derived: it is not automatic that 11 is achievable, and checking a proposed 11-element set directly by hand means auditing up to $\binom{11}{3} = 165$ triples. It is reinforced away by finding a coarser structure — partition the candidate set into classes by a small shared prime factor (here 2 and 3) so that pigeonhole with only two classes forces any three elements to include a repeated class, certifying the whole set at once.

4. The trap is treating the problem as a pure search: building a plausible large set by hand, believing it by inspection, and having no route to prove that a larger one does not exist, which leaves the answer as a conjecture rather than a solved extremal problem.

5. To recognise this faster: build a private reflex that the phrase "no k of them are pairwise [property]" should immediately trigger "what is the largest pairwise-[property] family in the universe," the same reflex whether the property is coprimality, distance, or any other pairwise relation.

6. Decision rule: for a "no k elements are pairwise-P" extremal problem, first bound $|S|$ by locating the largest pairwise-P family in the ambient universe (pigeonhole caps S's intersection with it at $k - 1$), then search for a partition of the universe into $k - 1$ classes each internally violating P, since the union of any $k - 1$ such classes automatically satisfies the forbidden condition and often matches the bound exactly.

1974 A2

1974 A2. A circle stands in a plane perpendicular to the ground and a point A lies in this plane exterior to the circle and higher than its bottom. A particle starting from rest at A slides without friction down an inclined straight line until it reaches the circle. Which straight line allows descent in the shortest time? [Assume that the force of gravity is constant over the region involved, there are no relativistic effects, etc.] The starting point A and the circle are fixed; the stopping point B is allowed to vary over the circle. Note: the answer may be given in any form which specifies the line of descent in an unambiguous manner; it is not required to find the coordinates of the point B.

Fields: Mechanics (kinematics on an incline), Euclidean geometry (power of a point) **Position:** A2
Answer: The optimal line is the one through A and the lowest point of the circle; the particle slides down this line to the near intersection B of that same line with the circle.

Layer 1 THE WHAT — surface analysis

The natural attempt. Set up coordinates with the circle's centre at the origin and A at a fixed point outside the circle. A candidate line through A is determined by one parameter, its angle θ from the horizontal. Intersecting this line with the circle's equation gives a quadratic in the parameter along the line, and its smaller positive root is the distance $b(\theta)$ from A to the near intersection point B — an expression built from the quadratic formula, involving a square root of a trigonometric expression in θ. Standard kinematics for frictionless descent from rest over a distance b along an incline at angle θ from the horizontal gives $t^2 = 2b/(g \sin\theta)$. You now have $t^2(\theta) = 2b(\theta)/(g \sin\theta)$, a single-variable function built from a square root nested inside a quotient. Differentiating it means applying the quotient rule and the chain rule through the square root, then setting the result to zero and simplifying a trigonometric expression that carries no obvious geometric meaning. It can be done, and the algebra does eventually converge

on a critical angle, but the manipulation is long, a single sign error in the implicit differentiation is easy to make and hard to catch — because the intermediate expressions no longer look like anything recognisable — and even after finding a critical point you still owe yourself a second-derivative or boundary check to be sure it is a minimum and not the direction that sends the line past the circle entirely.

The model chain. Extend the candidate line through A to its full intersection with the circle: it meets the circle in two points in general, the near one B (where the particle actually stops, at distance $b = AB$) and the far one C (at distance $c = AC$). Because A is a fixed point and the circle is fixed, the product $b \cdot c = AB \cdot AC$ is the power of the point A with respect to the circle — a constant, independent of which line through A is chosen, by the standard secant form of the power-of-a-point theorem. From kinematics, $t^2 = 2b/(g \sin\theta)$ where θ is the line's inclination from the horizontal. Writing $k = bc$ for the constant power of A, substitute $b = k/c$:

$$t^2 = \frac{2k}{g\, c \sin\theta}.$$

Since k is fixed, minimising t^2 is exactly maximising $c \sin\theta$. Now observe what $c \sin\theta$ actually is: it is the vertical component of the segment AC, i.e. the drop in height from A down to C. But the drop in height from A to C depends only on the height of the point C itself — it is simply $y_A - y_C$ — and not at all on which line was used to reach C; the length c and the angle θ were only ever a factorisation of that single number, height dropped. So maximising $c \sin\theta$ over all directions θ is the same as asking, over all points C on the circle, which one has the smallest height y_C — and that is visibly the lowest point of the circle, with no calculus required. Hence the optimal line is the one joining A to the lowest point of the circle, and the particle slides down it to the near intersection B.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree, from the first line of reasoning, that the target is $t^2 \propto b/\sin\theta$ where b is the near intersection distance. They split at how b itself is handled. The natural attempt treats b as a function of θ alone, obtained by intersecting a parametrised line with the circle's equation, and differentiates that composite expression directly. The model treats b as one member of a PAIR (b, c) — the two lengths cut by the same line from the same fixed point A — and eliminates b using the constant product bc, rewriting the target as a function of c and θ instead.

> *b alone, as a function of the incline's angle, is an ugly quadratic-formula expression — but b times the OTHER intersection length is a constant, independent of the line entirely, because that is exactly what the power of a point says. I should trade b for c using that constant product, not differentiate b directly.*

> *$c \sin\theta$ is nothing but the height A drops before reaching C — and that drop is already fixed the moment I pick which point C is; it does not care what road got me there. So I do not need calculus at all, I just need the lowest point of the circle.*

Layer 3 THE WHY

Why should minimising a descent time reduce to a fact about a circle's chords? Because the problem couples two fixed objects — an exterior point A and a circle — through a VARYING LINE through A, and the quantity to be minimised is built from one of the two lengths that line cuts on the circle. Whenever a fixed external point and a fixed circle are joined by a family of lines, the two intersection lengths measured from that point are not independent: their product is the power of the point, $AO^2 - r^2$, a number depending only on A and the circle, never on the

line's direction. The reflex this should trigger is immediate and content-free of the physics: any time a variable secant from a fixed external point meets a fixed circle in two points, write down the product of the two distances before doing anything else, because it lets you eliminate one unknown for free. This is exactly the same invariant used in 1976 A-1's minimisation, read in the opposite direction — there, the product itself was the quantity to be minimised; here, the product is a nuisance to be divided out so that the truly awkward variable, b, disappears from the expression entirely, leaving only c and θ.

The second, deeper recognition is why $c \sin\theta$ needed no further algebra at all. In a uniform gravitational field, the vertical drop between two fixed points is simply the difference of their heights; it does not care about the path, the angle, or the length of any particular line joining them. A line's length and its angle from the horizontal are not two independent quantities once its two endpoints are named — their product, length times sine of inclination, IS the vertical separation of the endpoints, nothing more. So the expression $c \sin\theta$, which arrived purely as leftover algebra from substituting $b = k/c$, was secretly never a two-variable trigonometric quantity to be optimised over θ at all: it was always a one-variable quantity depending only on the IDENTITY of the point C. The moment the optimisation variable is re-read as "which point on the circle" rather than "which angle," the extremum stops being a calculus problem and becomes an inspection: among all points of a circle, the lowest one is visibly extremal by the definition of "lowest," the same way STYLE.md's own worked circle example insists that a circle being the set of points equidistant from a centre makes the centre obviously the object of interest. Here, recognising that a formula's variable secretly denotes a POINT rather than an ANGLE is what converts a derivative into a glance.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the coexistence of a fixed exterior point, a fixed circle, and a line through the point free to rotate, with the target quantity built from one of the two lengths that line cuts on the circle. On seeing it, write down: the product of the two intersection lengths from that fixed point is constant (power of the point); use it before parametrising anything.

2. That cue suggests eliminating the awkward length (the near intersection, entangled with a quadratic-formula square root) in favour of the other length via the constant product, turning a messy two-unknown optimisation into a clean one-unknown one.

3. The hesitation occurs at recognising that the simplified target, $c \sin\theta$, is not merely algebra to differentiate but a physical quantity — a height drop — already determined by the endpoint alone. It is reinforced away by explicitly asking what $c \sin\theta$ MEANS physically (a vertical distance) rather than treating it as an abstract trigonometric product awaiting a derivative.

4. The trap is committing to a coordinate parametrisation by angle and differentiating through a nested square root before checking whether the two intersection lengths obey a hidden invariant; once inside that calculus, the algebra conceals the geometry, and a sign slip in the implicit differentiation is both easy to make and impossible to sanity-check against anything visual.

5. To recognise this faster: whenever "time to slide down a straight incline from rest" appears, immediately write $t^2 \propto (\text{distance}) / \sin(\text{angle from horizontal})$, then ask whether that distance is one of a PAIR of lengths tied to a fixed external point and a fixed circle — if so, reach for power of a point before reaching for a derivative.

6. Decision rule: when optimising a physical quantity built from a variable line's two intersection points with a fixed circle, measured from a fixed external point, trade the inconvenient intersection length for the other one via the constant power of the point, then check whether the resulting simplified expression is secretly a purely geometric quantity (a height, a distance) depending only on an endpoint's identity, before ever differentiating with respect to the line's angle.

1974 A3

1974 A3. A well-known theorem asserts that a prime $p > 2$ can be written as the sum of two perfect squares ($p = m^2 + n^2$, with m and n integers) if and only if $p \equiv 1 \pmod 4$. Assuming this result, find which primes $p > 2$ can be written in each of the following forms, using (not necessarily positive) integers x and y: (a) $x^2 + 16y^2$; (b) $4x^2 + 4xy + 5y^2$.

Fields: Number theory, quadratic forms, modular arithmetic **Position:** A3 **Answer:** (a) exactly the primes $p \equiv 1 \pmod 8$; (b) exactly the primes $p \equiv 5 \pmod 8$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Test small cases directly: $p = 17 \equiv 1 \pmod 8$ gives $17 = 1 + 16 = 1^2 + 16 \cdot 1^2$, so $(x, y) = (1, 1)$ works for form (a); $p = 41 \equiv 1 \pmod 8$ gives $41 = 25 + 16 = 5^2 + 16 \cdot 1^2$. Trying form (b) on $p = 5 \equiv 5 \pmod 8$: $5 = 4(1)^2 + 4(1)(0) + 5(0)^2$... that gives 4, not 5; try $x = 0, y = 1$: $5(1)^2 = 5$, works. Try $p = 13 \equiv 5 \pmod 8$: hunting for x, y with $4x^2 + 4xy + 5y^2 = 13$ by trial, $x = 1, y = 1$ gives $4 + 4 + 5 = 13$, works. After several such checks you start to believe (a) wants $p \equiv 1 \pmod 8$ and (b) wants $p \equiv 5 \pmod 8$, and you set out to prove it by pure residue-chasing: compute $x^2 + 16y^2 \pmod{16}$ or $\pmod 8$ for all residues of x, y, and separately compute $4x^2 + 4xy + 5y^2 \pmod 8$ for all residues of x, y, tabulating which residues each form can hit. This is finite and will eventually work, but it means re-deriving, from scratch, a classification that ignores the theorem the problem explicitly handed you — and the problem opens with "assuming this result," language that exists precisely because the setter expects you to route through it. Grinding out two independent residue tables for $x^2 + 16y^2$ and $4x^2 + 4xy + 5y^2$ without ever writing $p = m^2 + n^2$ is solving a harder, self-contained problem than the one that was actually asked.

The model chain. Notice both target forms are disguised sums of two squares: $x^2 + 16y^2 = x^2 + (4y)^2$, and $4x^2 + 4xy + 5y^2 = (2x + y)^2 + (2y)^2$ (expand: $(2x + y)^2 + (2y)^2 = 4x^2 + 4xy + y^2 + 4y^2 = 4x^2 + 4xy + 5y^2$, confirmed). So both are instances of $m^2 + n^2$ with the extra constraint that the EVEN one of m, n is a multiple of 4 — form (a) has $n = 4y$ directly divisible by 4, form (b) has $n = 2y$ with $m = 2x + y$, and since p is odd, m, n are of opposite parity, so here m is odd and $n = 2y$ is even but not necessarily divisible by 4. This says the two forms are exactly the two possible "grades" of evenness for n in the given theorem's representation $p = m^2 + n^2$: either $4 \mid n$ (form (a)) or $n \equiv 2 \pmod 4$ (form (b)).

Suppose $p = m^2 + n^2$ with p an odd prime, so m is odd and n is even; write $n = 2v$. Odd squares are always $\equiv 1 \pmod 8$ (check $1^2, 3^2, 5^2, 7^2 \equiv 1, 9, 25, 49 \equiv 1, 1, 1, 1 \pmod 8$), so $m^2 \equiv 1 \pmod 8$ and

$$p = m^2 + 4v^2 \equiv 1 + 4v^2 \pmod 8.$$

If v is even, $4v^2 \equiv 0 \pmod 8$, so $p \equiv 1 \pmod 8$; writing $v = 2y$ gives $n = 4y$, i.e. $p = m^2 + 16y^2$, form (a) with $x = m$. If v is odd, $4v^2 \equiv 4 \pmod 8$ (since $v^2 \equiv 1 \pmod 8$ for odd v), so $p \equiv 5 \pmod 8$; write $m = 2x + v$ for the integer $x = (m - v)/2$ (an integer since m, v are both odd, so $m - v$ is even), giving $p = (2x + v)^2 + 4v^2 = 4x^2 + 4xv + v^2 + 4v^2 = 4x^2 + 4xv + 5v^2$, form (b) with $y = v$.

Conversely, if $p = x^2 + 16y^2$, then x must be odd (else p is even), so $x^2 \equiv 1 \pmod 8$ and $16y^2 \equiv 0 \pmod 8$, giving $p \equiv 1 \pmod 8$. If $p = 4x^2 + 4xy + 5y^2 = (2x + y)^2 + 4y^2$, then $2x + y$ must be odd (else p is even), forcing y odd, so $4y^2 \equiv 4 \pmod 8$ and $(2x + y)^2 \equiv 1 \pmod 8$, giving $p \equiv 5 \pmod 8$. Since $p \equiv 1 \pmod 4$ splits into exactly $p \equiv 1 \pmod 8$ or $p \equiv 5 \pmod 8$, these two cases are exhaustive and mutually exclusive, so: (a) $p = x^2 + 16y^2$ is solvable exactly when $p \equiv 1 \pmod 8$; (b) $p = 4x^2 + 4xy + 5y^2$ is solvable exactly when $p \equiv 5 \pmod 8$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains notice immediately, on reading the problem, that it hands you a theorem about $m^2 + n^2$ right before asking about two forms that look adjacent to it. They split on what to do with that observation. The natural attempt treats $x^2 + 16y^2$ and $4x^2 + 4xy + 5y^2$ as new objects, to be classified from scratch by direct computation on x, y. The model attempt asks whether each target form IS a sum of two squares in algebraic disguise, and if so, folds the entire problem into the machinery of the given theorem, refining it by exactly one extra bit of information: the divisibility of the even variable by 4.

> *$x^2 + 16y^2$ is $x^2 + (4y)^2$, and $4x^2 + 4xy + 5y^2$ is $(2x + y)^2 + (2y)^2$ — both are just $m^2 + n^2$ with an extra constraint that the even one of m, n is a multiple of 4 rather than merely even. I am not being asked a new question; I am being asked to split the theorem I was handed one level deeper, along the parity of $n/2$.*

Layer 3 THE WHY

Why does that recognition have to occur, and why is splitting along $v = n/2$'s parity the right level of refinement rather than, say, mod 16? The problem's own phrasing, "assuming this result, find which primes...," is the strongest possible signal that the intended path runs THROUGH the stated theorem: any solution that proves the classification without ever writing $p = m^2 + n^2$ is answering a strictly harder question than the one being asked, since it re-derives structure the setter already gave away for free. Once that lens is accepted, the specific refinement is forced by counting: an odd prime $p = m^2 + n^2$ has m, n of opposite parity, and the even one, n, has exactly two possible "grades" that matter modulo 8 — divisible by 4, or congruent to 2 $\pmod 4$ — because m^2 itself is already pinned to $1 \pmod 8$ the instant m is odd (every odd square is $1 \pmod 8$, a basic and checkable fact, not a coincidence: odd numbers are ± 1 or $\pm 3 \pmod 8$, and both 1^2 and 3^2 reduce to $1 \pmod 8$). With m^2 fixed at $1 \pmod 8$, the ENTIRE value of $p \pmod 8$ is determined by $4v^2 \pmod 8$ alone, and that in turn depends only on whether v is even or odd — there is no finer distinction available modulo 8, because $v^2 \pmod 8$ only ever takes the values $0, 1, 4$ depending on whether v is even, odd, or $\equiv 2 \pmod 4$, but multiplied by 4 and reduced mod 8 this collapses to just two outcomes (0 or 4) matching exactly the two parities of v.

This is also why the problem has exactly two parts and not three or one: $p \equiv 1 \pmod 4$ splits into precisely two residues mod 8 (namely 1 and 5), so the theorem's single hypothesis "$p \equiv 1 \pmod 4$" was always secretly two cases hiding one modulus down, and the two target forms (a) and (b) are nothing but those two cases written out algebraically. Seeing a problem's two parts as the two branches of one case split — rather than as two unrelated sub-problems — is what turns a full residue-classification exercise into bookkeeping on a single congruence you already knew. The converse direction (a representation of the target shape implies the corresponding residue) has to be checked separately and by the same mod-8 tracking run in reverse, because the forward direction alone shows existence of a representation given the residue, but does not rule out the OTHER form also managing to represent the same prime; only the converse pins the classification down to an "if and only if."

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that both target forms, $x^2 + 16y^2$ and $4x^2 + 4xy + 5y^2$, complete or factor into the shape $m^2 + n^2$ with an extra divisibility condition on the even term. On seeing it, write down: these are refinements of the given sum-of-two-squares theorem, not new problems to classify from scratch.
2. That cue suggests writing $p = m^2 + n^2$ with m odd and $n = 2v$, then asking what property of v corresponds to each target form, rather than computing residues of $x^2 + 16y^2$ and $4x^2 + 4xy + 5y^2$ directly.

3. The hesitation is over which modulus resolves the two cases — mod 4, mod 8, or mod 16. It is reinforced away by computing odd squares mod 8 explicitly (they are all 1), which shows p mod 8 is pinned to exactly two possible values once $p \equiv 1 \pmod 4$, making mod 8 visibly the finest modulus that still separates into exactly two cases.

4. The trap is treating "assuming this result" as throwaway framing rather than as an instruction, and consequently re-deriving a mod-8 (or worse, mod-16) classification of $x^2 + 16y^2$ and $4x^2 + 4xy + 5y^2$ from first principles by residue-chasing on x, y directly — solvable, but several times the work, and with no theorem to check the answer against if an arithmetic slip is made.

5. To recognise this faster: whenever a problem says "assuming [theorem]" and then asks about a form that can be completed or factored into the exact shape of that theorem's conclusion, budget zero time for direct computation and substitute the theorem's own variables immediately.

6. Decision rule: when handed a classification theorem and then asked about representability by a form that algebraically rewrites as an instance of that theorem's expression with an extra congruence constraint on one variable, refine the theorem's own proof one modulus deeper along exactly that constrained variable, rather than starting the classification over from the raw form.

1974 A4

1974 A4. An unbiased coin is tossed n times. What is the expected value of $|H - T|$, where H is the number of heads and T is the number of tails? In other words, evaluate in closed form:

$$\frac{1}{2^{n-1}} \sum_{k<n/2} (n-2k)\binom{n}{k}.$$

(In this problem, "closed form" means a form not involving a series. The given series can be reduced to a single term involving only binomial coefficients, rational functions of n and 2^n, and the greatest integer function $[x]$.)

Caveat. The MAA OCR garbles the displayed sum (it reads "2•1_, 2 : $(n-2k)(k^n)$, $k < n/2$"). It is restored here as $\frac{1}{2^{n-1}} \sum_{k<n/2}(n-2k)\binom{n}{k}$, which is the only reading consistent with the verbal question: $E|H-T| = 2^{-n}\sum_k |n-2k|\binom{n}{k}$, and pairing k with $n-k$ doubles the terms with $k < n/2$. Identity check: the book's solution under 1974 A-4 manipulates exactly $\sum_{k<n/2}(n-2k)\binom{n}{k}$ and ends with $\frac{n}{2^{n-1}}\binom{n-1}{[(n-1)/2]}$, so the label and statement match. The closed form below was checked against the defining sum in exact arithmetic for $n = 1, \ldots, 15$.

Fields: combinatorics, probability (binomial sums) **Position:** A4 — mid-paper; one identity and a telescope, but only if you see the identity **Answer:** $E|H-T| = \dfrac{n}{2^{n-1}}\binom{n-1}{[(n-1)/2]}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You write the expectation from the definition. $H = k$ with probability $\binom{n}{k}/2^n$, and then $|H - T| = |k - (n-k)| = |n-2k|$, so

$$E|H-T| = \frac{1}{2^n}\sum_{k=0}^{n} |n-2k|\binom{n}{k}.$$

The absolute value is the nuisance, so you split at $k = n/2$ and use the symmetry $\binom{n}{k} = \binom{n}{n-k}$: the terms with $k > n/2$ mirror those with $k < n/2$, and the term $k = n/2$ (if n is even) is zero.

That gives the problem's own sum, $\frac{1}{2^{n-1}}\sum_{k<n/2}(n-2k)\binom{n}{k}$. So far the problem has done the work for you.

Now the standard reflex: split $n-2k$ into n and $-2k$ and evaluate each partial sum separately,

$$n\sum_{k<n/2}\binom{n}{k} \;-\; 2\sum_{k<n/2}k\binom{n}{k}.$$

The first is half of 2^n minus a correction for the middle term: for n odd it is 2^{n-1}, for n even it is $2^{n-1}-\frac{1}{2}\binom{n}{n/2}$. The second, using $k\binom{n}{k}=n\binom{n-1}{k-1}$, becomes $2n\sum_{j<n/2-1}\binom{n-1}{j}$, a half-sum of the row $n-1$ with its own parity-dependent correction. You now have four cases (parity of n times two sums), each ending in a middle binomial coefficient, and you have to push the case analysis through without dropping an index. It can be done — the corrections do combine — but this is the zombie route: a page of case-bookkeeping that obscures the fact that the answer is a single term.

The model chain. Write $n-2k=(n-k)-k$ and absorb each piece into the binomial coefficient using the two "committee with a chair" identities

$$(n-k)\binom{n}{k}=n\binom{n-1}{k},\qquad k\binom{n}{k}=n\binom{n-1}{k-1},$$

valid for all integers k with the convention $\binom{n-1}{-1}=0$. (Both are checked by writing out factorials: $(n-k)\frac{n!}{k!(n-k)!}=n\frac{(n-1)!}{k!(n-1-k)!}$, and $k\frac{n!}{k!(n-k)!}=n\frac{(n-1)!}{(k-1)!(n-k)!}$.) Hence

$$(n-2k)\binom{n}{k}=n\left[\binom{n-1}{k}-\binom{n-1}{k-1}\right].$$

Let $K=[(n-1)/2]$, the largest integer k with $k<n/2$. Summing over $0\le k\le K$ telescopes:

$$\sum_{k<n/2}(n-2k)\binom{n}{k}=n\sum_{k=0}^{K}\left[\binom{n-1}{k}-\binom{n-1}{k-1}\right]=n\binom{n-1}{K}.$$

Therefore

$$E|H-T|=\frac{n}{2^{n-1}}\binom{n-1}{[(n-1)/2]}.$$

Checks: $n=1$ gives 1 (one toss, $|H-T|=1$ always); $n=2$ gives 1 (HH, TT give 2, HT, TH give 0); $n=3$ gives $\frac{3\cdot2}{4}=\frac{3}{2}$, matching $\frac{2\cdot3+6\cdot1}{8}$. For even $n=2m$ this can be rewritten as $m\binom{2m}{m}/2^{2m-1}$, and Stirling's formula shows it grows like $\sqrt{2n/\pi}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $\sum_{k<n/2}(n-2k)\binom{n}{k}$. At that line the natural attempt separates n from $-2k$ and evaluates two partial sums of binomial rows. The model separates $n-2k$ as $(n-k)-k$ and turns each piece into a binomial coefficient of row $n-1$, so that the summand becomes a difference of consecutive terms.

> *Partial sums of a binomial row have no closed form, but the problem promises a single term. So the summand must be a difference of consecutive terms of something, and the sum must telescope. What difference is $(n-2k)\binom{n}{k}$?*

> *$n-2k$ is $(n-k)-k$: the number of tails minus the number of heads. Each of $k\binom{n}{k}$ and $(n-k)\binom{n}{k}$ is n times a binomial coefficient one row up. Write both and subtract.*

Layer 3 THE WHY

The instruction in the problem is the WHY, if you read it as information rather than as a formality: "the given series can be reduced to a single term". A finite sum with variable upper limit K collapses to one term only for a structural reason, and the overwhelmingly common reason is telescoping: $\sum_{k=0}^{K}(a_k - a_{k-1}) = a_K - a_{-1}$. So the question becomes: is $(n-2k)\binom{n}{k}$ a first difference $a_k - a_{k-1}$? Once the question is posed, the answer is forced by the shape of the summand. The factor $\binom{n}{k}$ is a count of k-subsets; a linear factor in front of it means "choose a distinguished element", either inside the subset (weight k) or outside it (weight $n-k$). Choosing the distinguished element first (n ways) and then the rest leaves $\binom{n-1}{k-1}$ or $\binom{n-1}{k}$ choices. That is the entire content of the two absorption identities, and it is why $n-2k$ must be read as $(n-k)-k$ rather than $n-2k$: only the split into "outside minus inside" gives two terms each of which absorbs.

Reading n and $-2k$ separately is the natural attempt's error at the axiom level. The constant n does not absorb into $\binom{n}{k}$ (there is no distinguished element), so it leaves you with a raw partial row sum $\sum_{k\le K}\binom{n}{k}$, and such partial sums are exactly the objects that have no closed form. The problem only works out because, in this particular combination, the two raw partial sums cancel each other's non-closed parts. You can prove that cancellation by cases; the absorption split shows it is not an accident.

There is also a probabilistic reason the answer is a single term, and it is worth knowing as a second path. Let S_j be heads minus tails after j tosses. If $S_j \neq 0$, one more toss moves $|S_j|$ up or down by 1 with equal probability, so the expected change is 0; if $S_j = 0$, the next toss makes $|S_{j+1}| = 1$. Hence $E|S_{j+1}| - E|S_j| = P(S_j = 0)$, and $E|S_n| = \sum_{j<n} P(S_j = 0)$. This is another telescope — the same structure, seen as a walk — and it explains why middle binomial coefficients (return-to-zero probabilities) appear in the answer. The book's solution uses only the algebraic telescope; it states $(n-2k)\binom{n}{k} = n\left[\binom{n-1}{k} - \binom{n-1}{k-1}\right]$ without saying where it comes from. It comes from asking what the summand is a difference of.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the parenthetical "can be reduced to a single term". Write down: "I need $(n-2k)\binom{n}{k} = a_k - a_{k-1}$."

2. Telescoping, with the difference supplied by absorption identities $k\binom{n}{k} = n\binom{n-1}{k-1}$ and $(n-k)\binom{n}{k} = n\binom{n-1}{k}$.

3. Hesitation comes at the split of $n-2k$. The pull is towards n and $2k$ because they are the visible pieces. Reinforce with: "a linear weight on $\binom{n}{k}$ is only useful if it counts a distinguished element; $n-k$ and k do, n does not."

4. The trap is evaluating partial row sums $\sum_{k<n/2}\binom{n}{k}$ and $\sum_{k<n/2} k\binom{n}{k}$ separately, which forces a parity case split with middle-coefficient corrections. It is not wrong, but it hides the one-term answer and invites index errors. A smaller trap: the upper limit is $\lfloor (n-1)/2 \rfloor$, not $\lfloor n/2 \rfloor$; for even n the middle term is excluded (it contributes 0 anyway).

5. Whenever a problem promises that a sum with variable upper limit has a closed form, look for a telescope before anything else. Whenever a polynomial factor multiplies a binomial coefficient, absorb it into a neighbouring row.

6. For $\sum_{k\le K} w(k)\binom{n}{k}$ with w linear, rewrite $w(k)$ as a combination of k and $n-k$, absorb each into row $n-1$, and see whether the result is a difference of consecutive terms. If a sum must collapse to one term, find the a_k with summand $= a_k - a_{k-1}$.

▷ **Cue:** a finite sum with a variable upper limit that the problem says reduces to a single term $\longrightarrow$ **Move:** find a_k with summand $= a_k - a_{k-1}$ and telescope

▷ **Cue:** a linear factor k or $n-k$ multiplying $\binom{n}{k}$ $\longrightarrow$ **Move:** absorb it: $k\binom{n}{k} = n\binom{n-1}{k-1}$, $(n-k)\binom{n}{k} = n\binom{n-1}{k}$

▷ **Cue:** $n-2k$ (or any "outside minus inside" weight) times $\binom{n}{k}$ $\longrightarrow$ **Move:** write it as $(n-k)-k$ so both pieces absorb

▷ **Cue:** expected absolute value $E|X|$ of a symmetric integer variable $\longrightarrow$ **Move:** fold by symmetry to a one-sided sum, doubling it
▷ **Cue:** $E|S_n|$ for a simple random walk $\longrightarrow$ **Move:** increments of $E|S_j|$ equal $P(S_j = 0)$; expect middle binomial coefficients in the answer

1974 A5

1974 A5. Consider the two mutually tangent parabolas $y = x^2$ and $y = -x^2$. [These have foci at $(0, 1/4)$ and $(0, -1/4)$, and directrices $y = -1/4$ and $y = 1/4$, respectively.] The upper parabola rolls without slipping around the fixed lower parabola. Find the locus of the focus of the moving parabola.

Caveat. The statement is clean in the OCR; the book's accompanying figure (marking $(0, 1/4)$ as "the original position of the moving focus" and "the fixed parabola") is lost and is not needed. Identity check: the book's solution under 1974 A-5 uses the fixed focus F, the moving focus M, the tangency point T, and concludes that M lies on $y = 1/4$, so the label and statement match. The reflection formula below was also checked numerically at several contact points.

Fields: geometry (conics, rigid motions) **Position:** A5 — late-paper; almost no computation once the symmetry is seen, a heavy kinematics problem if it is not **Answer:** The moving focus traces the whole line $y = 1/4$, the directrix of the fixed parabola. When the contact point is $(t, -t^2)$, the moving focus is at $(t, 1/4)$, directly above it.

Layer 1 THE WHAT — surface analysis

The natural attempt. "Rolls without slipping" reads as a mechanics instruction, so you set up the motion. Parametrise the contact point on the fixed parabola as $T = (t, -t^2)$. Arc length from the vertex is $s(t) = \int_0^t \sqrt{1 + 4u^2}\, du$, which involves arsinh. The moving parabola must touch at T with the point of it at the same arc length $s(t)$ from its own vertex, and its tangent there must line up with the fixed parabola's tangent, slope $-2t$. So you write the moving parabola as the image of $y = x^2$ under a rotation by some angle $\varphi(t)$ followed by a translation, find φ from the tangent condition, find the translation from the contact condition, and apply both to the focus $(0, 1/4)$.

This is a correct plan, and some of it goes through: the tangent of $y = x^2$ at (t, t^2) has slope $2t$, the fixed tangent has slope $-2t$, so the rotation angle is $-2\arctan 2t$. Then you push $(0, 1/4)$ through a rotation by $-2\arctan 2t$ about the point (t, t^2) and translate (t, t^2) to $(t, -t^2)$. Written out with $\cos(2\arctan 2t) = \frac{1-4t^2}{1+4t^2}$ and $\sin(2\arctan 2t) = \frac{4t}{1+4t^2}$, it works — but it is half a page of trigonometric algebra, you had to justify that the matching points are (t, t^2) and $(t, -t^2)$ (same abscissa), and you have not yet understood why the answer is a straight line.

The model chain. *Step 1: the moving parabola is the mirror image of the fixed one in the common tangent.* Let P be the fixed parabola $y = -x^2$ and P' the moving one, initially $y = x^2$; initially P' is the reflection of P in the common tangent $y = 0$. For a contact point T on P, let ℓ_T be the tangent to P at T and R_T the reflection in ℓ_T. The curve $R_T(P)$ is congruent to P', touches P at T with the same tangent, lies on the other side of ℓ_T, and the arc of $R_T(P)$ from T to its vertex has the same length as the arc of P from T to its vertex — which is exactly the length that has rolled. Moreover $R_T(P) = R_T \circ R_0(P')$, where R_0 is the reflection in $y = 0$, and a composition of two reflections is a rotation or translation, a proper rigid motion. These are precisely the conditions that define rolling without slipping: the moving curve is moved by a proper rigid motion, touches at T, with matching tangent, and the arc lengths rolled out on the two curves

are equal. So at every moment the moving parabola is $R_T(P)$, and its focus is $M = R_T(F)$, where $F = (0, -1/4)$ is the fixed focus.

Step 2: the reflection of the focus in a tangent lies on the directrix. Let D be the foot of the perpendicular from T to the directrix d of P (the line $y = 1/4$). By the focus–directrix definition, $TF = TD$. By the reflection property of the parabola, ℓ_T bisects the angle $\angle FTD$. In the isosceles triangle FTD, the bisector of the apex angle is the perpendicular bisector of the base FD. Hence R_T swaps F and D, so $M = D$.

(If T is the vertex, F, T, D are collinear and ℓ_T is the perpendicular bisector of FD directly.)

Step 3: the locus. With $T = (t, -t^2)$, $D = (t, 1/4)$. As the parabola rolls in both directions, t runs over all of $\mathbb{R}$ (the arc length of P is unbounded both ways), so the locus of M is the whole line $y = 1/4$.

Coordinate check. The tangent at $(t, -t^2)$ is $2tx + y - t^2 = 0$. At F the left side is $-(t^2 + \frac{1}{4})$, so

$$R_T(F) = \left(0, -\tfrac{1}{4}\right) + \frac{2(t^2 + \frac{1}{4})}{4t^2 + 1}(2t, 1) = \left(0, -\tfrac{1}{4}\right) + \tfrac{1}{2}(2t, 1) = \left(t, \tfrac{1}{4}\right).$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from "the moving parabola touches at T with the same tangent". The natural attempt then solves for the rigid motion (rotation angle and translation). The model notices that the configuration is symmetric about the common tangent at every moment, so the rigid motion is a reflection in ℓ_T composed with a fixed reflection, and the focus goes to $R_T(F)$.

> *The two parabolas are congruent and start as mirror images across their tangent. Rolling identical curves keeps them mirror images across the current tangent — the arcs rolled out are equal on both sides. So I never need the rotation angle: the moving focus is the fixed focus reflected in the tangent line.*

> *Reflect the focus in a tangent of a parabola. I know two things about T: TF equals the distance to the directrix, and the tangent bisects the angle between TF and the perpendicular to the directrix. Equal sides and a bisector: the reflection lands on the directrix.*

Layer 3 THE WHY

Why is the mirror picture forced rather than lucky? Because of what "rolling without slipping" means. It is not a force law; it is two geometric conditions: the curves are tangent at the contact point, and the arc length rolled out on each curve is the same. When the two curves are congruent and start in mirror position, there is an obvious candidate that satisfies both conditions at every instant — the mirror image in the current tangent — because reflection preserves arc length and preserves tangency. The only thing to check is that the candidate is reached by a proper motion (you cannot flip the moving parabola over), and that is the fact that two reflections compose to a rotation. So the symmetric picture is not an insight about this problem; it is the definition of rolling applied to congruent curves. The same argument shows that two equal gears, two equal circles, or two equal ellipses rolling on each other stay mirror images across the tangent.

Once the motion is "reflect in the tangent", the question becomes: where does the focus of a parabola go when you reflect it in a tangent? This is where the problem's bracketed hint earns its place. It hands you the foci and directrices, and that is exactly the data the definition of a parabola uses: the set of points equidistant from the focus and the directrix. Then $TF = TD$ for every point T, where D is the foot on the directrix. Two equal segments from T make an isosceles triangle, and reflection in any line through the apex that bisects the apex angle swaps

the base vertices. The reflection property — light from the focus reflects off the parabola parallel to the axis — says the tangent is exactly that bisector. (It is not an extra fact: the reflection property is itself proved from $TF = TD$, by showing that every other point of ℓ_T is closer to the directrix than to F, so ℓ_T meets the parabola only at T.) Thus $M = D$: the moving focus sits on the fixed directrix, directly above the contact point.

The book's solution is three sentences: reflection property, congruence, so MT is vertical and $MT = FT$, so M is on the directrix. It uses the mirror symmetry without saying that it is what rolling congruent curves means. That unspoken step is the whole problem; the rest is the definition of a parabola.

Notice also what the natural attempt never touches: arc length. The problem's phrase "without slipping" seems to demand the arsinh integral, but the mirror argument uses only that arc length is *the same* on both curves, never what it is. Equal-by-symmetry is always cheaper than equal-by-computation.

Layer $+\alpha$ BACKTESTING

1. The first cue is "two mutually tangent parabolas" that are reflections of each other in the x-axis, together with the given foci and directrices. Write down: "moving parabola = reflection of fixed parabola in the tangent at T; moving focus = reflection of F in that tangent."

2. A symmetry argument (reflection in the common tangent), then the focus–directrix definition and the reflection property of the parabola.

3. Hesitation comes at whether the mirror image really is the rolled position. Reinforce with the definition: tangency plus equal arc length plus proper motion. Reflection preserves arc length; two reflections make a rotation.

4. The trap is setting up the kinematics: computing arc length, rotation angles, and pushing the focus through them. It works, but it costs a page and hides why the answer is a line. A second trap is giving only a segment or omitting that t ranges over all reals: the parabola is unbounded, so the whole line is traced.

5. When congruent curves roll on each other from a mirror position, the configuration stays mirror-symmetric in the current tangent. When a conic's focus meets a tangent line, think of the reflection property and the directrix.

6. Before solving for a motion, ask whether the motion is a symmetry you can name. Convert "equal by computation" conditions (equal arc lengths) into "equal by symmetry" whenever the two objects are congruent. For a parabola: the reflection of the focus in any tangent is the foot of the perpendicular from the point of tangency to the directrix.

▷ **Cue:** two congruent curves rolling on each other, starting as mirror images across their common tangent $\longrightarrow$ **Move:** the moving curve is always the reflection of the fixed one in the current tangent

▷ **Cue:** "rolls without slipping" $\longrightarrow$ **Move:** translate to: tangent at the contact point, equal arc lengths rolled, proper rigid motion

▷ **Cue:** reflect the focus of a parabola in a tangent line $\longrightarrow$ **Move:** the image is the foot of the perpendicular from the point of tangency to the directrix

▷ **Cue:** a problem that supplies the focus and directrix of a parabola $\longrightarrow$ **Move:** use $TF =$ distance from T to the directrix, not the equation

▷ **Cue:** a locus problem where a rigid motion depends on a parameter $\longrightarrow$ **Move:** identify the motion as a composition of reflections before computing angles

1974 A6

1974 A6. It is well known that the value of the polynomial $(x+1)(x+2)\cdots(x+n)$ is

exactly divisible by n for every integer x. Given n, let $k = k(n)$ be the minimal degree of any monic integral polynomial

$$f(x) = x^k + a_1 x^{k-1} + \cdots + a_k$$

(with integer coefficients and leading coefficient 1) such that the value of $f(x)$ is exactly divisible by n for every integer x.

Find the relationship between n and $k = k(n)$. In particular, find the value of k corresponding to $n = 1000000$.

Caveat. The OCR renders the polynomial as "$/(x) = x^{\bullet} + a_1 x^{\bullet-1} + \cdots + a_{\bullet}$"; it is restored as $f(x) = x^k + a_1 x^{k-1} + \cdots + a_k$, which the parenthetical "(with integer coefficients and leading coefficient 1)" confirms. Identity check: the book's solution under 1974 A-6 proves $n \mid k!$ is necessary and sufficient via $(x+1)\cdots(x+k)$ and the k-th difference, and ends with $k(10^6) = k(5^6) = 25$, so the label and statement match. The value 25 was confirmed by direct computation of the least k with $10^6 \mid k!$.

Fields: number theory, polynomials (finite differences) **Position:** A6 — last problem of the paper, but among the gentler A6s once finite differences are in hand **Answer:** $k(n)$ is the least positive integer k with $n \mid k!$ (with $k(1) = 0$). For $n = 10^6$, $k = 25$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem opens by handing you $(x+1)(x+2)\cdots(x+n)$: degree n, divisible by n. So $k(n) \le n$. The obvious next question is how much lower you can go, and the natural attempt goes prime by prime. For $n = p$ prime, Fermat gives $x^p - x \equiv 0 \pmod p$, degree p; can you beat p? You try $x^2 + x = x(x+1)$ for $n = 2$ (works, degree 2) and try to find something of degree 2 for $n = 3$: $x^2 + ax + b$ must vanish mod 3 at $x = 0, 1, 2$, but a nonzero quadratic mod 3 has at most 2 roots, so no. For $n = 4$ you experiment: $x(x+1)$ is only divisible by 2; $x(x+1)(x+2)$ is divisible by 6, so by 2 but not always by 4 ($x = 1$ gives 6). Degree 4: $x(x+1)(x+2)(x+3)$ is divisible by 24, so by 4. Is there a monic cubic divisible by 4 everywhere? You test candidates and they fail, but testing is not proof.

At this point the data are $k(2) = 2$, $k(3) = 3$, $k(4) = 4$, and you might guess $k(n) = n$ — wrong, since $k(6) = 3$ ($x(x+1)(x+2)$ is divisible by 6). Or you try to build $k(n)$ from the prime factorisation with Chinese Remainder arguments and root-counting mod p^a, which gets messy fast because root-counting fails modulo prime powers. The attempt stalls on the lower bound: you can build polynomials, but you cannot prove a smaller degree is impossible.

The model chain. *Claim.* n divides $f(x)$ for every integer x, for some monic integer f of degree k, if and only if $n \mid k!$. Hence $k(n) = \min\{k \ge 0 : n \mid k!\}$.

Sufficiency. Let $p_k(x) = (x+1)(x+2)\cdots(x+k)$, monic of degree k with integer coefficients. For an integer $x \ge 0$, $p_k(x) = k!\binom{x+k}{k}$. For $-k \le x \le -1$, one factor is 0. For $x \le -k-1$, write $x = -y$ with $y \ge k+1$; then $p_k(x) = (-1)^k(y-1)(y-2)\cdots(y-k) = (-1)^k k!\binom{y-1}{k}$. In every case $k! \mid p_k(x)$, so if $n \mid k!$ then $n \mid p_k(x)$ for all integers x.

Necessity. Define the forward difference $(\Delta g)(x) = g(x+1) - g(x)$. If g has degree $d \ge 1$ and leading coefficient c, then Δg has degree $d-1$ and leading coefficient cd, because $(x+1)^d - x^d = dx^{d-1} + (\text{lower terms})$ and lower-degree terms of g contribute only lower-degree terms. Applying this k times to monic f of degree k gives $\Delta^k f = k!$, a constant. On the other hand, by induction on j,

$$(\Delta^k f)(0) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(j),$$

an integer combination of values of f. If n divides every $f(j)$, it divides this combination, so $n \mid k!$.

The value at $n = 10^6 = 2^6 \cdot 5^6$. We need $v_5(k!) \geq 6$ and $v_2(k!) \geq 6$, where $v_p(k!) = \sum_{i \geq 1} [k/p^i]$ (Legendre). $v_5(24!) = 4$ and $v_5(25!) = 5 + 1 = 6$, while $v_2(25!) = 12 + 6 + 3 + 1 = 22 \geq 6$. So the least k is 25: $k(10^6) = 25$.

(Examples: $k(p) = p$ for p prime; $k(4) = 4$; $k(6) = 3$; $k(1) = 0$ with $f = 1$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains know that $(x+1) \cdots (x+k)$ gives an upper bound. The natural attempt then tries to prove lower bounds by root-counting modulo primes and prime powers. The model applies the difference operator k times and reads off that $k!$ is an integer combination of the values $f(0), \ldots, f(k)$.

> *Divisible by n for every integer x is a statement about values, not coefficients. What can I compute from values alone that sees the leading coefficient? Differences: k differences of a monic degree-k polynomial give k!, and each difference is an integer combination of values.*

> *The hint's product $(x+1) \cdots (x+n)$ is divisible by far more than n — by n!. So the real invariant is k!, and the question is which k! is the first to contain n.*

Layer 3 THE WHY

Why are finite differences the natural tool rather than a trick? Because of what the hypothesis is. "$n \mid f(x)$ for every integer x" constrains f only through its values on $\mathbb{Z}$. Any lower bound on the degree must therefore come from an operation that takes values as input and returns information about the degree or the leading coefficient. For polynomials over the reals that operation is differentiation, and the k-th derivative of a monic degree-k polynomial is $k!$. But derivatives are not integer combinations of values. The discrete analogue is: replace $f'(x)$ by $f(x+1) - f(x)$. It lowers degree by one and multiplies the leading coefficient by the degree, exactly as differentiation does, and — the point — it is built from values with integer coefficients, so divisibility by n passes through it. The $k!$ appears for the same reason in both settings: each step pulls down one factor of the current degree.

Seen this way the sufficiency half is not separate luck. The values of $p_k(x) = (x+1) \cdots (x+k)$ are $k!$ times binomial coefficients, because p_k is $k!$ times the polynomial $\binom{x+k}{k}$, and binomial-coefficient polynomials are exactly the ones whose differences are again binomial-coefficient polynomials: $\Delta\binom{x}{k} = \binom{x}{k-1}$ (Pascal's rule). The integer-valued polynomials are the integer combinations of $\binom{x}{j}$, and a monic polynomial of degree k is $k!\binom{x}{k}$ plus lower terms. So $k!$ is the natural "denominator" at degree k: the necessity argument shows n can divide at most $k!$, the construction shows it reaches $k!$. The two halves are the two sides of one fact about the basis $\binom{x}{j}$.

The problem's opening sentence is a deliberately weak hint. $(x+1) \cdots (x+n)$ is divisible by n because it is divisible by $n!$. If you ask "by what exactly is it divisible?" you are led to $k!$ and the answer shape $n \mid k!$ at once. The book's solution states "the k-th difference $k!$ of a monic integral polynomial of degree k is divisible by any common divisor of all the values" in one line; the WHY is that differencing is the value-only version of differentiation.

Finally, $k(n)$ as a function is governed by the largest prime-power demand: for 10^6 the factor 5^6 dominates because 5s are scarcer in factorials than 2s. Legendre's formula settles it without computing 25!. Note 25, not 30: $25 = 5^2$ contributes two 5s.

Layer $+\alpha$ BACKTESTING

1. The first cue is "divisible by n for every integer x" together with "monic" and "degree". Write down: "values only; take k differences; $\Delta^k f = k!$."

2. Finite differences for the lower bound, the product of k consecutive integers for the construction, Legendre's formula for the numerical case.

3. Hesitation comes at proving that no smaller degree works, because examples cannot prove impossibility. Reinforce with: "an impossibility for all x needs an identity among values; find the combination of values that equals something fixed."

4. Two traps. First, generalising $k(2) = 2$, $k(3) = 3$, $k(4) = 4$ to $k(n) = n$; $n = 6$ kills it. Second, attempting lower bounds by counting roots modulo prime powers, where Lagrange's theorem fails (for example $x^2 - 1$ has four roots mod 8). Also, in the numerical part, counting $v_5(25!)$ as 5 instead of 6.

5. When a polynomial condition holds at every integer, think Δ and the basis $\binom{x}{j}$. When a hint gives a divisibility fact, ask what the exact divisor is; it is usually larger and more informative.

6. To bound the degree of a polynomial from a condition on its integer values, apply Δ as many times as the degree: Δ^k of a degree-k polynomial with leading coefficient c is $ck!$, and it is an integer combination of $k+1$ consecutive values. To find the least k with $n \mid k!$, use Legendre's formula on the prime power of n that is hardest to satisfy.

▷ **Cue:** a polynomial with integer coefficients whose values at all integers are divisible by n $\longrightarrow$ **Move:** take the k-th forward difference: $\Delta^k f = (\text{leading coeff}) \cdot k!$, an integer combination of values

▷ **Cue:** need a monic polynomial divisible by n at every integer $\longrightarrow$ **Move:** use $(x+1)(x+2)\cdots(x+k) = k!\binom{x+k}{k}$

▷ **Cue:** a hint states a weak divisibility ("divisible by n") $\longrightarrow$ **Move:** ask for the exact divisor (here $n!$); it points to the invariant

▷ **Cue:** least k with $n \mid k!$ $\longrightarrow$ **Move:** Legendre's formula $v_p(k!) = \sum_i \lfloor k/p^i \rfloor$ on each prime power of n; the scarcest prime decides

▷ **Cue:** proving no polynomial of smaller degree works $\longrightarrow$ **Move:** find a fixed quantity that is an integer combination of values; do not count roots modulo prime powers

Putnam 1974 — Paper B

1974 B1

1974 B1. Which configurations of five (not necessarily distinct) points p_1, p_2, p_3, p_4, p_5 on the circle $x^2 + y^2 = 1$ maximize the sum of the ten distances $\sum_{i<j} d(p_i, p_j)$? [Here $d(p, q)$ denotes the straight-line distance between p and q.]

Caveat. The 1965–1984 MAA scan renders this problem with dropped subscripts and a mangled summation ("the sum of the ten distances $2d(p, p,)$"). The statement above is reconstructed from the surrounding legible text and confirmed against the printed solution, which explicitly maximizes $\sum d(p_i, p_j)$ over five points on the unit circle and concludes with the regular pentagon; the reconstruction is not in doubt, only the OCR's rendering of the subscripts.

Fields: combinatorial geometry, optimization on the circle **Position:** B1 – opening problem of the second session **Answer:** The sum is maximized exactly when the five points are the vertices of a regular pentagon.

Layer 1 THE WHAT — surface analysis

The natural attempt. Five points, ten distances, and a request for the maximizing configuration: the reflex is to say "by symmetry, it must be the regular pentagon" and stop, or – if you distrust hand-waving – to set it up as calculus. Write $p_k = (\cos\theta_k, \sin\theta_k)$ and use the chord-length identity $d(p_i, p_j) = 2\left|\sin\frac{\theta_i - \theta_j}{2}\right|$, so the objective is $F(\theta_1, \ldots, \theta_5) = \sum_{i<j} 2\left|\sin\frac{\theta_i - \theta_j}{2}\right|$, a function on the 5-torus. You reach for Lagrange multipliers or a direct partial-derivative attack: $\partial F/\partial\theta_k = 0$ for each k. But F involves ten absolute-value terms, and each $\partial/\partial\theta_k$ pulls in a sum of four signed cosine terms (one for each of the four other points), giving five coupled trigonometric equations in five unknowns with sign ambiguities from the absolute values baked in. You can verify, after real effort, that the regular pentagon ($\theta_k = 2\pi k/5$) is a critical point – the equations become symmetric and cancel – but you have proved only that it is *a* critical point of a function on a compact five-dimensional domain, not that it is the maximum rather than a saddle or a different, unequal-spacing critical configuration that happens to also satisfy the symmetric equations. The direct approach answers "is this configuration stationary?" when the problem asks "which configuration is optimal?" – and closing that gap by comparing all critical points of a five-variable trigonometric system is not something you want to attempt by hand in contest time.

The model chain. Existence of a maximum is immediate: the domain (five points ranging independently over the compact circle C) is $C \times C \times C \times C \times C$, compact, and F is continuous, so a maximum is attained. The content is identifying it, and the route that works never differentiates in five variables at once – it splits the sum and moves one point at a time.

Since $\binom{5}{2} = 10$ and $5 \times 2 = 10$, the ten pairs of $\{p_1, \ldots, p_5\}$ split exactly into two Hamiltonian 5-cycles of the complete graph K_5: the "index-adjacent" cycle $S = d(p_1, p_2) + d(p_2, p_3) +$

$d(p_3,p_4)+d(p_4,p_5)+d(p_5,p_1)$ and the "index-skip" cycle $T = d(p_1,p_3)+d(p_3,p_5)+d(p_5,p_2)+d(p_2,p_4)+d(p_4,p_1)$, and $F = S + T$ exactly, with no pair counted twice or missed, because every edge of K_5 lies in exactly one of these two 5-cycles.

Now fix any four of the five points and let p be the fifth, appearing in S paired with its two cycle-neighbours a, b (fixed). The only part of S that moves is $D = d(p,a) + d(p,b)$. Since p, a, b all lie on the unit circle, the unit circle is the circumcircle of triangle pab, so the Law of Sines gives side lengths directly in terms of opposite angles with circumradius 1: $d(p,a) = 2\sin\beta$ and $d(p,b) = 2\sin\alpha$, where $\alpha = \angle pab$ and $\beta = \angle pba$. Hence $D = 2(\sin\alpha + \sin\beta)$. Because a, b are fixed, the inscribed angle $\angle apb$ subtending the fixed arc ab is fixed (inscribed angle theorem), so $\alpha + \beta = \pi - \angle apb =: \sigma$ is a fixed constant as p varies. Write $D = 2\left(\sin\alpha + \sin(\sigma - \alpha)\right) = 4\sin\frac{\sigma}{2}\cos\left(\alpha - \frac{\sigma}{2}\right)$, using the sum-to-product identity. This is maximized exactly when $\cos(\alpha - \sigma/2) = 1$, i.e. $\alpha = \sigma/2 = \beta$ – that is, exactly when p is placed so that triangle pab is isosceles with $pa = pb$, meaning p sits at the angular midpoint between a and b on the arc it occupies.

So: at a maximum of S (with the other four points held fixed), every point must be the angular midpoint of its two S-neighbours. Repeating verbatim with T in place of S: at a maximum of $F = S + T$, every point must *simultaneously* be the angular midpoint of its two S-neighbours and the angular midpoint of its two T-neighbours. Checking the regular pentagon directly: placing p_k at angle $2\pi k/5$, the S-cycle is the pentagon's edges (arc 72°) and the T-cycle is the pentagon's diagonals (arc 144°, the pentagram), and by the pentagon's five-fold symmetry every vertex is trivially the angular midpoint of its neighbours in both cycles. This configuration exists, is admissible, and satisfies the necessary condition derived above at every point; combined with the compactness argument guaranteeing a maximum exists, the regular pentagon is the maximizer.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both routes agree that a maximizer exists and that the regular pentagon is a natural guess. They split at the question "how do I turn the guess into a proof?" The natural attempt tries to verify the guess by differentiating the ten-term sum in all five variables simultaneously – checking that the regular pentagon is *a* critical point of the whole mess. The model chain instead asks whether the ten-term sum can be repackaged as *two* five-term sums, each of which reduces to a one-point, two-neighbour optimization that a single application of the Law of Sines settles completely.

> *Ten pairs from five points and five index-adjacent pairs is exactly half – I should check whether the other five pairs also form a clean structure, because if the total sum splits into two sums that each individually reduce to "optimize one point against two fixed neighbours," I have turned a five-variable problem into five copies of a one-variable problem.*

> *p, a, b are three points on the very circle I'm optimizing over – that circle is not just the domain, it is also the circumcircle of every triangle I can draw from three of my points, which means the Law of Sines gives me side lengths directly from angles with no unknown circumradius to solve for.*

Layer 3 THE WHY

Why does splitting into two 5-cycles work, and why is it not a lucky coincidence for $n = 5$? Because K_5 has 10 edges and every vertex has degree 4; a graph decomposes into 2-regular subgraphs (unions of cycles) whenever its degree is even, and a Hamiltonian decomposition into two disjoint 5-cycles is the coarsest such decomposition available for K_5 specifically. The reason this matters for the calculus is that a five-term cyclic sum, unlike the original ten-term sum, has the property that any single point appears in exactly two terms – so freezing the other four points isolates a genuine one-variable optimization at that point. The original ten-term sum does

not have this property for any individual point (each point appears in four terms, paired with all four others at once), which is exactly why differentiating it directly produces four-term coupled equations instead of a clean two-term one. Recognising the 2-regular decomposition is what converts "optimize five things jointly" into "optimize one thing against two fixed neighbours, five times over" – the reduction in dimension is the whole content of the model chain.

Why does fixing $\alpha + \beta$ pay off, rather than treating α, β as two independent unknowns? Because a two-variable optimization with a rigid linear constraint is a bounded one-variable optimization in disguise – and the inscribed angle theorem is precisely what supplies that constraint for free from the geometry: a, b fixed means the arc between them is fixed, and any point p on the remaining arc sees that fixed arc at a fixed angle. This is the same fact underlying every "angle subtended by a fixed chord is constant" argument in circle geometry, applied here not to prove a static angle-chasing fact but to reduce a genuine two-degree-of-freedom optimization (α and β free) to one degree of freedom (α free, $\beta = \sigma - \alpha$ determined).

Why does the maximizer of $\sin\alpha + \sin(\sigma - \alpha)$ land at $\alpha = \beta$, and why is this "of course" rather than a trick? Because $\sin\alpha + \sin(\sigma - \alpha) = 2\sin\frac{\sigma}{2}\cos(\alpha - \sigma/2)$ makes the dependence on α visible as a single cosine factor with a fixed positive coefficient $2\sin(\sigma/2)$ in front (since $0 < \sigma < \pi$ throughout the admissible range); a cosine is maximized when its argument is zero, which forces $\alpha = \sigma/2$ directly. The isosceles condition $pa = pb$ is not imposed as an aesthetic preference for symmetry – it falls out as the unique zero of a derivative once the constant σ is fixed.

Caveat. The printed solution asserts this local-symmetry condition, checks that the regular pentagon satisfies it, and stops there; it does not exclude the possibility of some other, non-regular configuration that also satisfies "every point is the angular midpoint of its neighbours in both S and T simultaneously" for a different choice of labelling or a non-convex arrangement of the five points on the circle. The existence-of-maximum argument guarantees *some* maximizer satisfies the necessary condition; ruling out other configurations satisfying the same necessary condition (and confirming the regular pentagon is the strict global maximum rather than merely a local one) is not supplied by the official solution and is not reconstructed independently here.

Layer $+\alpha$ BACKTESTING

1. The cue is "five points on a circle, sum of all pairwise distances, maximize" – ten pairwise terms from five points. On seeing it, write down: $\binom{5}{2} = 10 = 5 \cdot 2$, so check whether the pair-set splits into two disjoint Hamiltonian cycles, because a cyclic sum isolates one point against two neighbours at a time.

2. That cue suggests abandoning a joint five-variable optimization in favour of a two-cycle decomposition, then applying the Law of Sines to each one-point sub-problem, using the fact that points constrained to a fixed circle automatically have that circle as the circumcircle of any triangle drawn from three of them.

3. Hesitation occurs right after writing down the ten-term sum and reaching for $\partial F/\partial\theta_k = 0$; it is reinforced away by asking, before differentiating anything, "does this sum decompose into smaller sums where each point appears only twice, rather than four times?"

4. The trap is treating "the regular pentagon is a critical point" as equivalent to "the regular pentagon is the maximum" – a five-variable critical-point check, even when it succeeds, has not ruled out saddle points or other equal-status critical configurations, and existence-of-maximum (from compactness) has to be invoked separately and married to the local necessary condition.

5. Any time a sum runs over all $\binom{n}{2}$ pairs of n points constrained to a common circle (or more generally a common circumcircle-bearing configuration), check whether $\binom{n}{2}$ factors as (number of disjoint Hamiltonian cycles)$\times n$ before attempting direct multivariable calculus.

6. Reusable rule: when optimizing a sum over all pairs of n points on a circle, decompose the complete graph K_n into 2-regular subgraphs (cycles) so that each point appears in only two terms per subgraph; then optimize one point against two fixed neighbours using the Law of Sines in the triangle they form with the common circle as circumcircle, where the angle sum at the moving point is fixed by the inscribed angle theorem – this converts an n-variable problem into n copies of a one-variable problem.

▷ **Cue:** a sum runs over all pairwise distances of n points on a common circle $\longrightarrow$ **Move:** check whether $\binom{n}{2}$ decomposes into disjoint Hamiltonian cycles of K_n (each point in exactly two terms per cycle), and optimize one point at a time against its two cycle-neighbours

▷ **Cue:** three points lie on a circle of known radius R and an angle-versus-side relationship is needed $\longrightarrow$ **Move:** apply the Law of Sines directly with that circle as circumcircle: side $= 2R\sin(\text{opposite angle})$, no separate circumradius computation required

▷ **Cue:** two angles α, β satisfy a fixed sum $\alpha + \beta = \sigma$ and $\sin\alpha + \sin\beta$ is to be optimized $\longrightarrow$ **Move:** rewrite as $2\sin(\sigma/2)\cos(\alpha - \sigma/2)$ via sum-to-product; the maximum is forced at $\alpha = \beta = \sigma/2$ with no calculus needed beyond recognising the cosine factor

1974 B2

1974 B2. Let $y(x)$ be a continuously differentiable real-valued function of a real variable x. Show that if $(y')^2 + y^3 \to 0$ as $x \to +\infty$, then $y(x)$ and $y'(x) \to 0$ as $x \to +\infty$.

Caveat. The scan renders the hypothesis as "$(y')^2 + y' \to 0$", with the exponent on the second term lost to OCR. The printed solution repeatedly uses y^3 (writing "y^3 and $(y')^2 + y^{3}$", "the hypothesis $(y')^2 + y^3 \to 0$") and constructs a comparison differential equation whose exponent only produces the required finite-time blow-up when the power on y is odd and greater than 1 – confirmed independently below. The exponent is therefore reconstructed as 3, not the OCR's dropped digit.

Fields: real analysis (ODE comparison, boundedness arguments) **Position:** B2 **Answer:** Both conclusions hold: $y(x) \to 0$ and $y'(x) \to 0$ as $x \to +\infty$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The hypothesis $(y')^2 + y^3 \to 0$ looks like it should immediately bound y: since $(y')^2 \geq 0$, you have $y^3 \leq (y')^2 + y^3 \to 0$, so $\limsup_{x\to\infty} y(x) \leq 0$. That feels like most of the work. You then try to get a lower bound the same way – but there is no analogous trick, because $(y')^2 + y^3 \to 0$ places no direct lower bound on y^3 alone; the two terms can trade off against each other, with y^3 arbitrarily negative provided $(y')^2$ compensates. You might then try assuming, without justification, that y is eventually monotonic (since it "should" settle down), and differentiate the hypothesis: $2y'y'' + 3y^2y' = 0$ wherever $y' \neq 0$, i.e. $y'(2y'' + 3y^2) = 0$, giving $y'' = -\frac{3}{2}y^2$ whenever $y' \neq 0$ – a clean autonomous ODE that tempts you to solve it and read off the limit directly. But this substitution has silently assumed y is twice differentiable (never given) and, worse, has thrown away exactly the case that turns out to be the hard one: what if y' vanishes infinitely often, or what if y is unbounded below while y' compensates without ever being single-signed? The upper bound you got for free was the easy half of the problem; the natural attempt has no answer for why y cannot dive to $-\infty$ while y' races to $+\infty$ fast enough to keep the sum near 0 forever.

The model chain. Split on whether y' vanishes infinitely often as $x \to \infty$.

Case A: $y'(x_n) = 0$ for a sequence $x_n \to \infty$ enumerating all zeros of y' beyond some point. At each x_n, the hypothesis gives $y(x_n)^3 = (y'(x_n))^2 + y(x_n)^3 \to 0$, so $y(x_n) \to 0$. Since the x_n

are exactly the critical points of y (all local maxima and minima), y is monotonic on each interval $[x_n, x_{n+1}]$, hence bounded between $y(x_n)$ and $y(x_{n+1})$ there. Both endpoint values $\to 0$, so for any $\varepsilon > 0$ there is N with $|y(x_n)| < \varepsilon$ for all $n \geq N$, forcing $|y(x)| < \varepsilon$ for all $x \geq x_N$. Hence $y(x) \to 0$, and then $(y'(x))^2 = \left[(y')^2 + y^3\right] - y^3 \to 0 - 0 = 0$, so $y'(x) \to 0$ as well.

Case B: y' has only finitely many zeros beyond some point x_0, so y' is single-signed on (x_0, ∞) by continuity (an intermediate-value-theorem sign change would force a new zero).

Subcase B(a): $y' > 0$ on (x_0, ∞). If y were unbounded above, then $y^3 \to +\infty$ (since y increasing and unbounded), forcing $(y')^2 + y^3 \to +\infty \neq 0$, contradiction. So y is increasing and bounded above, hence converges to a finite limit L. Then $y^3 \to L^3$, and since $(y')^2 + y^3 \to 0$, $(y')^2 \to -L^3$. But $(y')^2 \geq 0$ forces $L \leq 0$. If $L < 0$ strictly, then $(y')^2 \to -L^3 > 0$, so (since $y' > 0$) $y' \to \sqrt{-L^3} > 0$; but a positive function bounded away from 0 makes y unbounded, contradicting $y \to L$ finite. Hence $L = 0$ and $(y')^2 \to 0$: both conclusions hold.

Subcase B(b): $y' < 0$ on (x_0, ∞). If y is bounded below, the mirror argument to B(a) gives $y \to 0$, $y' \to 0$. Suppose instead $y \to -\infty$. For any $\varepsilon > 0$ there is X with $(y')^2 + y^3 < \varepsilon$ for $x > X$, i.e. $(y')^2 > -y^3 - \varepsilon = |y|^3 - \varepsilon$ (since $y < 0$). Once $|y|^3 \geq 2\varepsilon$ this gives $(y')^2 > \frac{1}{2}|y|^3$, and since $y' < 0$, $-y' > \frac{1}{\sqrt{2}}|y|^{3/2}$. Writing $z = -y \to +\infty$, this is $z' > \frac{1}{\sqrt{2}} z^{3/2}$ for x beyond some point where z is already large. But the comparison equation $w' = \frac{1}{\sqrt{2}} w^{3/2}$ is separable: $\int w^{-3/2} dw = \frac{1}{\sqrt{2}} \int dx$ gives $-2w^{-1/2} = \frac{x}{\sqrt{2}} + C$, so $w^{-1/2} \to 0$ (i.e. $w \to \infty$) at a *finite* value of x. Since z grows at least as fast as such a w once matched at a starting point, $z = -y$ must reach $+\infty$ at some finite $x^* < \infty$ – but y is differentiable, hence finite, for every real x. Contradiction. So y cannot be unbounded below, and Subcase B(b) reduces to the bounded case, giving $y \to 0$, $y' \to 0$.

All cases give $y(x) \to 0$ and $y'(x) \to 0$ as $x \to +\infty$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same free observation: $y^3 \leq (y')^2 + y^3 \to 0$, so $\limsup y \leq 0$. They split on how to handle the lower bound. The natural attempt tries to differentiate the hypothesis and solve an autonomous ODE, silently assuming smoothness and monotonicity it hasn't earned. The model chain instead asks what *kind* of behaviour could make $y \to -\infty$ consistent with the hypothesis, and answers it with a differential inequality that forces a finite-time blow-up – contradicting that y is defined for all x.

> *The easy half of this problem (bounding y above) came from nothing more than $(y')^2 \geq 0$; the fact that it was easy is itself a signal that the real content is in ruling out $y \to -\infty$, not in re-deriving what I already have.*

> *If $y \to -\infty$, the hypothesis forces $(y')^2$ to grow like $|y|^3$ to keep the sum near zero – that is not just "y' is large," it is a specific growth rate, and a specific growth rate on y' in terms of y is exactly the data needed to compare against a solvable separable ODE.*

> *A first-order autonomous inequality $z' \geq cz^p$ with $p > 1$ does not merely make z grow fast – it makes z reach infinity in finite time, because $\int^\infty z^{-p}\, dz$ converges for $p > 1$; that convergence, not the size of the growth rate, is what contradicts y being defined for all real x.*

Layer 3 THE WHY

Why does the case split on zeros of y' matter, rather than treating y uniformly? Because the hypothesis $(y')^2 + y^3 \to 0$ constrains a *combination* of y and y', not y alone, and the only way to extract information about y by itself is to catch it at the moments where $y' = 0$ – points where the combination collapses to y^3 alone. If those moments occur infinitely often and cofinally, the

problem is essentially solved by squeezing y between consecutive critical values. If they do not, y' is eventually single-signed, and single-signedness is precisely what makes y monotonic – turning "does y converge?" into "is y bounded?", a question calculus already knows how to answer for monotonic functions.

Why does boundedness above follow trivially but boundedness below require real work? Because $(y')^2$ is a square: it is automatically nonnegative, so it can only ever *help* push y^3 down towards the hypothesis's target of 0, never make y^3 need to be more negative than the hypothesis allows. There is no symmetric fact bounding y^3 from below by $-(y')^2$-something; the hypothesis genuinely permits $y \to -\infty$ as far as sign considerations alone are concerned, and only a rate argument – comparing the forced growth of $|y'|$ against a differential equation with finite-time blow-up – can close off that possibility.

Why is the exponent $3/2$ (equivalently, the fact that $(y')^2$ scales like $|y|^3$) exactly what produces blow-up rather than mere unboundedness? Because for $z' = cz^p$ with $c > 0$, $\int_a^\infty z^{-p}\,dz$ converges precisely when $p > 1$; the antiderivative of z^{-p} is $\frac{z^{1-p}}{1-p}$, which tends to a *finite* limit as $z \to \infty$ exactly when $1 - p < 0$. Here $p = 3/2 > 1$, so the "time" needed for z to traverse from any finite starting value to ∞ is itself finite – this is the textbook fact that $y' = y^2$ (and any power exceeding 1) blows up in finite time, in contrast to $y' = y$ (exponential growth, which takes infinite time to reach ∞). The cube in the original hypothesis is what supplies an exponent above 1 once the square root is taken; had the hypothesis instead been $(y')^2 + y \to 0$ (linear, not cubic), the comparison equation would have exponent $1/2 < 1$ and would *not* blow up in finite time, and indeed $y \to -\infty$ while $y' \to 0$ slowly is genuinely possible under that weaker hypothesis. The problem's specific choice of y^3 is what makes the conclusion true.

Caveat. The exact numerical constant in the official comparison ODE is unrecoverable from the OCR (the fragment reads "$y' = -(1/2)|y|^{\cdots}$"); the constant used above ($1/\sqrt{2}$) was derived independently from the hypothesis and is sufficient for the argument, but the printed original may use a different (equally valid) constant obtained by a slightly different choice of threshold.

Layer $+\alpha$ BACKTESTING

1. The cue is a hypothesis of the form $(y')^2 + (\text{odd power of } y) \to 0$. On seeing it, write down immediately: $(y')^2 \geq 0$ gives a one-sided bound on y for free, and the real question is always the other side – can y run away to $-\infty$ while y' compensates?

2. That cue suggests splitting on whether y' vanishes infinitely often (squeeze y between consecutive critical values) versus eventually single-signed (reduce to monotonic boundedness, then rule out unboundedness via a comparison ODE).

3. Hesitation occurs right after getting the free upper bound $\limsup y \leq 0$ and being tempted to declare victory or to differentiate the hypothesis assuming unjustified smoothness; it is reinforced away by asking explicitly, "have I shown a lower bound on y, or have I only used that squares are nonnegative once?"

4. The trap is mistaking the free half of the inequality for the whole problem: $y^3 \leq (y')^2 + y^3$ is true but only ever constrains y from above, and stopping there leaves the harder, unbounded-below case completely unaddressed.

5. Any time a hypothesis combines $(y')^2$ (or any even power) with an odd power of y tending to 0, immediately test whether the implied growth rate of $|y'|$ versus $|y|$, near a hypothetical blow-up, gives a separable ODE with exponent exceeding 1 – that exponent is the signature of finite-time blow-up.

6. Reusable rule: to rule out $y \to -\infty$ (or $+\infty$) for a function defined on all of $[x_0, \infty)$, derive a differential inequality $|y'| \geq c|y|^p$ from the hypothesis in the region where y is presumed to escape, and check the exponent: $p > 1$ forces finite-time blow-up, contradicting that y is defined for all x; $p \leq 1$ does not, and the unboundedness may be genuinely possible.

- ▷ **Cue:** a hypothesis bounds $(y')^2$ plus an odd power of y as $x \to \infty$ $\longrightarrow$ **Move:** extract the free one-sided bound on y from $(y')^2 \geq 0$ immediately, then treat the opposite-sign unboundedness as the separate, harder case requiring a rate argument
- ▷ **Cue:** ruling out that a differentiable function escapes to $\pm\infty$ on an infinite domain $\longrightarrow$ **Move:** derive a differential inequality $|y'| \geq c|y|^p$ near the presumed blow-up and check whether $p > 1$; if so, compare to the separable ODE $w' = cw^p$, whose solution reaches infinity in finite time, contradicting that y is defined for all x
- ▷ **Cue:** a function's derivative vanishes at a sequence of points tending to infinity, with a known limit for the function's value there $\longrightarrow$ **Move:** use monotonicity between consecutive zeros to squeeze the function's values on the intervening intervals between the (converging) endpoint values

1974 B3

1974 B3. Prove that if α is a real number such that $\cos \pi\alpha = 1/3$, then α is irrational. (The angle $\pi\alpha$ is in radians.)

Fields: number theory (3-adic valuations, orbit counting via the doubling map) **Position:** B3 **Answer:** α is irrational – proved by contradiction, using the doubling formula and a 3-adic valuation argument to exhibit infinitely many distinct values of $\cos(2^m \pi\alpha)$, contradicting the finitely many values available if α were rational.

Layer 1 THE WHAT — surface analysis

The natural attempt. The number $1/3$ looks arithmetically ordinary, so the instinct is to search for special angles: is $\pi\alpha$ a "nice" multiple of π where the cosine is known explicitly – $\pi/3, \pi/4, \pi/6$, sums of these? You check the standard table of cosines of rational multiples of π (Niven's theorem territory, even if you don't know the name) and $1/3$ is not among $0, \pm\frac{1}{2}, \pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{3}}{2}, \pm 1$ – the only rational values $\cos(\theta)$ takes for θ a rational multiple of π. This is suggestive but not a proof: you have shown $1/3$ is absent from a list you've checked by memory or by testing small denominators, not that it is absent from *every* rational multiple of π, and "I couldn't find it" is not "it doesn't exist." You might try assuming $\alpha = p/q$ in lowest terms and manipulating $\cos(\pi p/q) = 1/3$ algebraically – perhaps invoking that $2\cos(\pi p/q)$ is an algebraic integer (a root of a minimal polynomial related to cyclotomic polynomials) and that $2/3$ is not an algebraic integer, since algebraic integers satisfying monic integer polynomials cannot be non-integer rationals. This is in fact a legitimate route (it is exactly how Niven's theorem is usually proved), but reaching it requires machinery – the theory of algebraic integers and the rational root theorem for monic polynomials – that is a large hammer for a Putnam B3, and assembling the precise minimal polynomial of $2\cos(\pi p/q)$ from scratch, under time pressure, is its own substantial detour.

The model chain. Suppose, for contradiction, that $\alpha = r/s$ is rational, with r, s integers and $s > 0$. Let $\theta = \pi\alpha = \pi r/s$. Then for any integer n, $n\theta = n\pi r/s$, and modulo 2π this depends only on nr mod $2s$ (since $2\pi \cdot \frac{k}{2s} = \pi \cdot \frac{k}{s}$ recovers the same form). As n ranges over the integers, nr mod $2s$ takes at most $2s$ distinct residues, so $\cos(n\theta)$ takes **at most** $2s$ **distinct values** as n ranges over all integers – a finite bound depending only on the (fixed) denominator s.

Now produce infinitely many distinct values of $\cos(n\theta)$ directly, using only $n = 2^m$ and the double-angle formula $\cos(2\phi) = 2\cos^2\phi - 1$. Set $c_0 = \cos\theta = 1/3$ and $c_m = \cos(2^m\theta)$, so $c_m = 2c_{m-1}^2 - 1$ for $m \geq 1$. **Claim:** for every $m \geq 0$, $c_m = a_m/3^{2^m}$ for some integer a_m with $3 \nmid a_m$.

Base case $m = 0$: $c_0 = 1/3 = 1/3^{2^0}$, so $a_0 = 1$, and $3 \nmid 1$.

Inductive step: suppose $c_{m-1} = a_{m-1}/3^{2^{m-1}}$ with $3 \nmid a_{m-1}$. Then

$$c_m = 2c_{m-1}^2 - 1 = \frac{2a_{m-1}^2}{3^{2^m}} - 1 = \frac{2a_{m-1}^2 - 3^{2^m}}{3^{2^m}},$$

using $\left(3^{2^{m-1}}\right)^2 = 3^{2^m}$. Set $a_m = 2a_{m-1}^2 - 3^{2^m}$. Since $3 \nmid a_{m-1}$, $a_{m-1}^2 \equiv 1 \pmod 3$ (every nonzero residue mod 3 squares to 1), so $2a_{m-1}^2 \equiv 2 \pmod 3$; and $3^{2^m} \equiv 0 \pmod 3$ since $2^m \geq 1$. Hence $a_m \equiv 2 \pmod 3$, so $3 \nmid a_m$, closing the induction.

Now show the c_m are pairwise distinct. Suppose $c_m = c_{m'}$ for some $m > m' \geq 0$. Then

$$\frac{a_m}{3^{2^m}} = \frac{a_{m'}}{3^{2^{m'}}} \implies a_m = a_{m'} \cdot 3^{2^m - 2^{m'}}.$$

Since $m > m'$, $2^m - 2^{m'} \geq 2^m - 2^{m-1} = 2^{m-1} \geq 1 > 0$, so the right side is divisible by 3. But a_m is not divisible by 3 (shown above) – contradiction. So $c_0, c_1, c_2, \ldots$ are pairwise distinct: an infinite set of distinct values of $\cos(2^m\theta)$.

This contradicts the bound of at most $2s$ distinct values of $\cos(n\theta)$ over all integers n (taking $n = 2^m$ in particular). The contradiction came from assuming α rational, so α is irrational.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both routes start by recognising $1/3$ is not a familiar cosine value. They split on what to do with that recognition. The natural attempt treats "not on my list of known values" as evidence and looks for an algebraic-integer obstruction, which is correct in spirit but requires importing machinery. The model chain instead asks a purely combinatorial question: if α were rational with denominator s, how many distinct values *could* $\cos(n\theta)$ possibly take across all integers n – and then manufactures an explicit infinite family of provably distinct values using nothing but the double-angle formula and a mod-3 count, without ever invoking algebraic number theory by name.

> *If $\alpha = r/s$ is rational, then $\{n\theta \bmod 2\pi : n \in \mathbb{Z}\}$ is a finite set of size at most $2s$ – so $\cos(n\theta)$ can only ever take finitely many values as n ranges over the integers; exhibiting infinitely many distinct values of $\cos(n\theta)$ for any single sequence of n's is therefore already a complete contradiction, and I don't need to know anything else about the structure of that sequence.*

> *The double-angle formula $\cos(2\phi) = 2\cos^2\phi - 1$ turns "start at $1/3$" into a recursion on a rational number, and a recursion on a rational number is exactly the kind of object where I can track a single prime's divisibility (here, 3) through every step instead of tracking the whole fraction.*

> *I don't need to solve the recursion $c_m = 2c_{m-1}^2 - 1$ in closed form or understand its long-run dynamics at all – I only need the much weaker fact that the numerators, reduced over the forced denominators 3^{2^m}, are never divisible by 3, because that single invariant is already enough to force every term to be distinct from every other term.*

Layer 3 THE WHY

Why does rationality of α bound the number of values of $\cos(n\theta)$ at all? Because $\theta = \pi r/s$ means $n\theta \pmod{2\pi}$ is completely determined by $nr \bmod 2s$ (since 2π corresponds to $2s$ in the "k/s" parametrization, $\pi k/s \equiv \pi k'/s \pmod{2\pi}$ exactly when $k \equiv k' \pmod{2s}$) – and there are only $2s$ residues mod $2s$. This is the arithmetic content of "θ is a rational multiple of π": the orbit $\{n\theta \bmod 2\pi\}$ under repeated addition of θ is not merely bounded, it is *finite*, of size dividing $2s$. Irrationality proofs by contradiction that start from "$\alpha = r/s$" and end in a counting

contradiction are exploiting exactly this: rational rotation has a finite, closed orbit; irrational rotation does not. The whole strategy is to show the orbit of θ under doubling ($n = 1, 2, 4, 8, \dots$) is *not* finite, which is incompatible with θ being a rational multiple of π.

Why is the mod-3 tracking exactly the right invariant to prove the orbit under doubling is infinite, rather than some other modulus or some other property? Because the starting value $\cos\theta = 1/3$ has 3 sitting in the denominator, and the doubling formula $c_m = 2c_{m-1}^2 - 1$ squares the numerator and squares the denominator at each step – so the denominator is forced to be exactly 3^{2^m} at every stage (never reducing, since $3 \nmid$ numerator is preserved), and two fractions with denominators that are different powers of the *same prime* can only be equal if the prime divides the numerator of the one with larger denominator after cross-multiplying. This is not a coincidence tailored to $1/3$: the same argument would work verbatim for $\cos\theta = 1/p$ for any prime $p \geq 3$ that keeps $a_{m-1}^2 \not\equiv 0 \pmod{p}$ at every step (which holds automatically once $p \nmid a_0$, since $p \nmid a_{m-1} \Rightarrow p \nmid a_{m-1}^2$), giving a whole family of irrationality results by the identical mechanism.

Why does "$a_{m-1}^2 \equiv 1 \pmod{3}$ whenever $3 \nmid a_{m-1}$" do all the work of the induction? Because modulo a prime, the nonzero residues form a group under multiplication, and mod 3 that group has only two elements, $\{1, 2\}$, both of which square to 1 – so squaring any unit mod 3 is forced to land on 1, with no case analysis needed. This is why 3 (rather than, say, 5, where units square to either 1 or 4) makes the induction a one-line computation instead of a case split: the specific starting fraction $1/3$ was not an arbitrary choice by the problem-setter, it is the smallest prime for which this single-valued-square phenomenon holds, making the resulting proof as short as it is.

Layer $+\alpha$ BACKTESTING

1. The cue is "$\cos(\pi\alpha)$ equals a specific rational number that is not on the standard list of rational-multiple-of-π cosine values; prove α irrational." On seeing it, write down: assume $\alpha = r/s$ rational, and note that this forces $\{n\theta \bmod 2\pi\}$ to be a finite set of size at most $2s$ – so $\cos(n\theta)$ takes only finitely many values over all integers n.

2. That cue suggests hunting for a specific integer sequence n (here $n = 2^m$, via repeated doubling) along which $\cos(n\theta)$ can be shown, by an explicit recursive formula, to take infinitely many distinct values – turning the finite-orbit consequence of rationality directly into a contradiction.

3. Hesitation occurs right after checking that $1/3$ is not a "known" cosine value and being tempted to reach for algebraic-integer machinery; it is reinforced away by asking, "can I get a contradiction from counting alone, using the double-angle formula and one prime's divisibility, without naming any theorem about algebraic integers?"

4. The trap is treating "I don't recognise this value" as a proof technique in itself; recognising that $1/3$ is absent from a memorised table only motivates the conjecture, it does not supply the argument, which has to come from an independent counting or algebraic mechanism.

5. Any time a cosine (or sine, or tangent) of a specific angle is given as a specific rational number and irrationality of the angle (as a multiple of π) is wanted, immediately consider the orbit of that angle under the doubling map via $\cos(2\phi) = 2\cos^2\phi - 1$, and track the divisibility, by the prime appearing in the given value's denominator, of the numerators produced by the recursion.

6. Reusable rule: to prove θ/π is irrational given $\cos\theta = 1/p$ for a prime p, show by induction that $\cos(2^m\theta) = a_m/p^{2^m}$ with $p \nmid a_m$ for all m (using that the nonzero residues mod a prime are closed under squaring away from 0), conclude the $\cos(2^m\theta)$ are pairwise distinct, and contradict the fact that a rational multiple of π has only finitely many values of $\cos(n\theta)$ across all integers n.

▷ **Cue:** $\cos\theta$ equals a specific rational number $1/p$ for a prime p, and irrationality of θ/π is wanted $\longrightarrow$ **Move:** apply the doubling formula $\cos(2\phi) = 2\cos^2\phi - 1$ recursively and

track that the numerator stays coprime to p at every step, forcing infinitely many distinct values of $\cos(2^m\theta)$

- ▷ **Cue:** a hypothetical rational multiple of π, say $\theta = \pi r/s$, needs to be shown impossible $\longrightarrow$ **Move:** note that $\cos(n\theta)$ can then take at most $2s$ distinct values over all integers n, since $n\theta \bmod 2\pi$ depends only on $nr \bmod 2s$; any exhibited infinite family of distinct values of $\cos(n\theta)$ is then an immediate contradiction
- ▷ **Cue:** a recursively defined rational number's numerator and denominator need to be shown coprime to a fixed prime at every step $\longrightarrow$ **Move:** use that the nonzero residues modulo a prime form a group under multiplication, closed under squaring; verify the recursion's squaring step cannot produce a multiple of that prime once the previous term avoided it

1974 B4

1974 B4. In the standard definition, a real-valued function of two real variables $g : \mathbb{R}^2 \to \mathbb{R}^1$ is continuous if, for every point $(x_0, y_0) \in \mathbb{R}^2$ and every $\epsilon > 0$, there is a corresponding $\delta > 0$ such that

$$[(x - x_0)^2 + (y - y_0)^2]^{1/2} < \delta \quad \text{implies} \quad |g(x, y) - g(x_0, y_0)| < \epsilon.$$

By contrast, $f : \mathbb{R}^2 \to \mathbb{R}^1$ is said to be continuous in each variable separately if, for each fixed value y_0 of y, the function $f(x, y_0)$ is continuous in the usual sense as a function of x, and similarly $f(x_0, y)$ is continuous as a function of y for each fixed x_0.

Let $f : \mathbb{R}^2 \to \mathbb{R}^1$ be continuous in each variable separately. Show that there exists a sequence of continuous functions $g_n : \mathbb{R}^2 \to \mathbb{R}^1$ such that

$$f(x, y) = \lim_{n\to\infty} g_n(x, y) \qquad \text{for all } (x, y) \in \mathbb{R}^2.$$

Caveat. The statement is reconstructed from the MAA volume's OCR, which garbles the formulae ($\mathbb{R}^2$, ϵ, δ, the limit). The reconstruction was checked against the book's solution printed under 1974 B-4 (linear interpolation in x between the vertical lines $x = m/n$), which solves exactly this statement.

Fields: real analysis; continuity in several variables; approximation by interpolation **Position:** B4 — middle of the paper; short once you see which variable to discretise **Answer:** Take g_n equal to f on every vertical line $x = m/n$ ($m \in \mathbb{Z}$) and linear in x between consecutive lines. Continuity in y makes each g_n jointly continuous; continuity in x makes $g_n \to f$ pointwise.

Layer 1 THE WHAT — surface analysis

The natural attempt. The standard way to approximate a function on the plane by continuous ones is to sample it on a fine grid and interpolate. So the strong student samples f at the lattice points $(i/n, j/n)$ and fills in each little square by bilinear interpolation. The result g_n is continuous (it is a polynomial on each square and the pieces agree on the edges), so the only question is convergence. At a point (x, y) inside a square, $g_n(x, y)$ is a weighted average of the four corner values $f(i/n, j/n)$, and the corners tend to (x, y). If f were jointly continuous, the corner values would tend to $f(x, y)$ and we would be done.

But f is not assumed jointly continuous, and that is the whole point of the problem. Carry the attempt into the standard example. Let $\varphi(u, v) = uv/(u^2 + v^2)$ with $\varphi(0, 0) = 0$; it is continuous in each variable separately, and it equals $\frac{1}{2}\sin 2\theta$ in polar angle θ, so it is discontinuous at the origin. Move the bad point off the lattice: $f(x, y) = \varphi(x - a, y - a)$ with $a = \sqrt{2}$. At (a, a) write

$t = \{na\}$ (the fractional part), so the corners sit at offsets $u, v \in \{-t/n,\ (1-t)/n\}$ from (a,a). Since φ depends only on the direction of (u,v), the bilinear average is

$$g_n(a,a) = \tfrac{1}{2}\big(t^2 + (1-t)^2\big) - \frac{2t^2(1-t)^2}{t^2+(1-t)^2},$$

which is $\frac{1}{5}$ at $t = \frac{1}{4}$ and tends to $\frac{1}{2}$ as $t \to 0$ or $t \to 1$. By Dirichlet's approximation theorem $\{n\sqrt{2}\}$ lies within $1/n$ of 0 or 1 for infinitely many n, so $g_n(a,a)$ is near $\frac{1}{2}$ infinitely often, while $f(a,a) = 0$. (Numerically: $n = 17$ gives 0.457, $n = 1001$ gives 0.061.) The grid scheme fails at exactly the points where the hypothesis is weaker than joint continuity. The wall is structural: any scheme that samples f at points approaching (x,y) *diagonally* asks for joint continuity.

The model chain. Fix $n \geq 1$. For real x, put $m = \lfloor nx \rfloor$ and $\lambda = nx - m \in [0,1)$, so $x = (1-\lambda)\frac{m}{n} + \lambda\frac{m+1}{n}$. Define

$$g_n(x,y) = (1-\lambda)\, f\big(\tfrac{m}{n}, y\big) + \lambda\, f\big(\tfrac{m+1}{n}, y\big)\,.$$

On the closed strip $S_m = [\frac{m}{n}, \frac{m+1}{n}] \times \mathbb{R}$ the same formula, with $\lambda = nx - m \in [0,1]$, gives a function G_m that agrees with g_n on S_m (at $x = \frac{m+1}{n}$ both give $f(\frac{m+1}{n}, y)$).

Step 1: each G_m is continuous on S_m. The map $(x,y) \mapsto f(\frac{m}{n}, y)$ is continuous on $\mathbb{R}^2$: given ϵ, continuity of $f(\frac{m}{n}, \cdot)$ at y_0 gives δ, and $[(x-x_0)^2 + (y-y_0)^2]^{1/2} < \delta$ forces $|y - y_0| < \delta$. The map $(x,y) \mapsto \lambda = nx - m$ is continuous. Sums and products of continuous functions are continuous, so G_m is.

Step 2: g_n is continuous on $\mathbb{R}^2$. If $x_0 \notin \frac{1}{n}\mathbb{Z}$, then near (x_0, y_0) the function g_n coincides with one G_m, so it is continuous there. If $x_0 = \frac{m}{n}$, then near (x_0, y_0) the plane is covered by S_{m-1} and S_m, on which g_n equals G_{m-1} and G_m, and both equal $f(x_0, y_0)$ at (x_0, y_0). Given ϵ, choose δ_1, δ_2 for G_{m-1}, G_m and let $\delta = \min(\delta_1, \delta_2, \frac{1}{n})$: every point within δ lies in one of the two strips, so $|g_n - g_n(x_0, y_0)| < \epsilon$ there.

Step 3: $g_n \to f$ pointwise. Fix (x,y) and $\epsilon > 0$. By continuity of $f(\cdot, y)$ at x there is $\delta > 0$ with $|f(s,y) - f(x,y)| < \epsilon$ whenever $|s - x| < \delta$. For $n > 1/\delta$, both $\frac{m}{n}$ and $\frac{m+1}{n}$ lie within $\frac{1}{n} < \delta$ of x, and $g_n(x,y)$ is a convex combination of $f(\frac{m}{n}, y)$ and $f(\frac{m+1}{n}, y)$, so

$$|g_n(x,y) - f(x,y)| \leq (1-\lambda)\big|f(\tfrac{m}{n}, y) - f(x,y)\big| + \lambda\big|f(\tfrac{m+1}{n}, y) - f(x,y)\big| < \epsilon. \qquad \square$$

The official solution is these three steps in four lines.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree that you should sample f on a grid and interpolate. They split at the choice of grid. The natural attempt discretises *both* variables, sampling at lattice points. The model discretises *one* variable and keeps the other exact: it samples along whole vertical lines $x = m/n$. With lattice points, convergence needs $f(i/n, j/n) \to f(x,y)$ along diagonal approaches, which the hypothesis does not give. With lines, convergence only ever moves in x at a fixed y, which is exactly what the hypothesis gives.

> *My hypothesis only lets me move along horizontal and vertical lines. So I may never approach a point diagonally. If I sample on a grid of points, my samples come in diagonally. I have to keep one coordinate exact.*

> *I have two hypotheses and two jobs. Continuity in x should make the approximation converge; continuity in y should make each approximant continuous. So discretise in x only, and keep y as it is: along each line $x = m/n$ the function is already continuous in y.*

- ▷ **Cue:** a hypothesis weaker than the one the standard method needs (separately continuous instead of jointly continuous; pointwise instead of uniform) ⟶ **Move:** find the exact place the standard method uses the stronger hypothesis and redesign that step so it uses only the weaker one
- ▷ **Cue:** two hypotheses and two things to prove (here: continuity of g_n, convergence $g_n \to f$) ⟶ **Move:** assign each hypothesis to one job before constructing anything; the construction then writes itself
- ▷ **Cue:** approximate a function by continuous ones ⟶ **Move:** piecewise-linear interpolation between sample values; convergence follows because the interpolant is a convex combination of nearby values

Layer 3 THE WHY

Why is the vertical-line construction natural and not a trick? Start from what the hypothesis actually says. Separate continuity is a statement about f restricted to lines parallel to the axes, and nothing else. Any argument that uses f at points approaching (x, y) from a slanted direction is using information you were not given. So the admissible approximation schemes are the ones in which every limiting process runs along a horizontal or a vertical line. That is a strong constraint, and it nearly dictates the answer.

Now look at the two things to prove. Continuity of g_n is a statement about all directions at once, so it cannot be bought from f directly. It can be bought from a structural fact that is true at A-level: a function of y alone that is continuous in y is continuous as a function of (x, y), because moving the point by less than δ moves y by less than δ. So if g_n is built out of *one-variable* continuous pieces — $f(\frac{m}{n}, y)$ is continuous in y, and the weights $\lambda(x)$ are continuous in x — then sums and products make it jointly continuous. Separately continuous functions misbehave only through genuine interaction between x and y; a construction in which each piece depends on one variable has no interaction to misbehave with.

Convergence is a statement about one point (x, y) and a limit in n, so it can run along the horizontal line through (x, y), which the hypothesis allows. Interpolating linearly in x puts $g_n(x, y)$ between two values of $f(\cdot, y)$ taken within $1/n$ of x. That convex-combination bound is the whole reason linear interpolation is chosen: an average of numbers each within ϵ of $f(x, y)$ is within ϵ of $f(x, y)$.

So the design follows from two definitions: continuity of a function of one variable, and the triangle-inequality behaviour of averages. The asymmetry between x and y is not a choice of taste. One variable is spent on regularity and the other on convergence, because each hypothesis can pay for exactly one of them. Swapping the roles (horizontal lines $y = m/n$, interpolating in y) works equally well, and that symmetry is the check that nothing arbitrary happened.

Why must the answer be only a pointwise limit and not a uniform one? Because a uniform limit of continuous functions is continuous, and φ above is separately continuous but not continuous. The problem is asking for the most you can get, and you should expect the convergence to be genuinely pointwise. The book's remark records what this buys: separately continuous functions are pointwise limits of continuous functions (Baire class 1), hence Borel measurable and continuous except on a meagre set.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "continuous in each variable separately" set against the full ϵ–δ definition. Write down: "I may only take limits along lines parallel to the axes."

2. It suggests a sampling-and-interpolation construction in which every limit runs along an axis direction: sample on whole lines, not at points.

3. Hesitation comes at the grid: the lattice is the default, and it works for jointly continuous f, so it feels safe. Reinforce it away by asking, for the candidate g_n, "which values of f does $g_n(x, y)$ use, and how do those sample points approach (x, y)?"

4. The trap is bilinear interpolation on a lattice. It is continuous and converges for every jointly continuous f, so it passes every sanity check except the one that matters. With $f(x,y) = \varphi(x-\sqrt{2}, y-\sqrt{2})$ it fails at $(\sqrt{2},\sqrt{2})$.

5. Whenever a problem hands you a weakened hypothesis next to the standard one, locate the step of the standard proof that uses the difference. That step is the one to redesign; the rest can stay.

6. Reusable decision rule: *when you have two hypotheses and two conclusions, match them one to one before constructing anything, and build the object so that each conclusion draws only on its assigned hypothesis.*

1974 B5

1974 B5. Show that $1 + (n/1!) + (n^2/2!) + \cdots + (n^n/n!) > e^n/2$ for every integer $n \geq 0$.
REMARKS. You may assume as known Taylor's remainder formula:

$$e^x - \sum_{k=0}^{n} \frac{x^k}{k!} = \frac{1}{n!}\int_0^x (x-t)^n e^t\,dt,$$

as well as the fact that

$$n! = \int_0^\infty t^n e^{-t}\,dt.$$

Caveat. The two displayed formulae in the Remarks are reconstructed from badly garbled OCR in the MAA volume. The reconstruction is the only reading consistent with the book's solution under 1974 B-5, which uses exactly these two identities and the substitution $u = n - t$; both are standard and were re-derived independently here.

Fields: real analysis; inequalities; integral representation of sums; (context) the Poisson and Gamma distributions **Position:** B5 — late in the paper; the Remarks hand over the tools, and the difficulty is in using them **Answer:** Convert the sum to an integral: the claim is equivalent to $\int_0^n u^n e^{-u}\,du < \int_n^\infty u^n e^{-u}\,du$, which holds because $u^n e^{-u}$ peaks at $u = n$ and falls off more slowly to the right than to the left: $(n+h)^n e^{-h} > (n-h)^n e^h$ for $0 < h < n$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $T_k = n^k/k!$, so the claim is $\sum_{k\leq n} T_k > \frac{1}{2}\sum_{k\geq 0} T_k$, that is, head $\geq$ tail where the head is $k \leq n$ and the tail is $k > n$. The terms rise until $k = n$ and then fall, so the student pairs the tail term T_{n+j} with the head term T_{n-j} and hopes for $T_{n+j} \leq T_{n-j}$. The ratio is

$$\frac{T_{n+j}}{T_{n-j}} = \frac{n^{2j}}{\prod_{i=1}^{j}(n+i)(n-i+1)} = \prod_{i=1}^{j}\frac{n^2}{n^2+n-i(i-1)}.$$

For small j every factor is below 1 (for $j = 1$ the ratio is $n/(n+1)$), which looks promising. But once $i(i-1) > n$ the factors exceed 1, and the ratio eventually climbs past 1. For $n = 10$ the ratios for $j = 1,\ldots,10$ are $0.91, 0.84, 0.81, 0.83, 0.92, 1.15, 1.69, 3.12, 8.22, 41.1$. On top of that, the tail has infinitely many terms beyond $k = 2n$ with no partner at all. So term-by-term pairing does not close; you are left needing a cumulative estimate of partial products, followed by a bound on the far tail. The honest numbers for $n = 10$ are head ≈ 12842, tail ≈ 9184: true, with room, but the discrete route has no clean handle on it.

The other default is asymptotics: $e^{-n}\sum_{k\leq n} T_k = P(X \leq n)$ for X Poisson with mean n, and the central limit theorem gives $\to \frac{1}{2}$. That explains why $\frac{1}{2}$ is the right constant, and proves nothing for any particular n.

The model chain. For $n = 0$ the left side is $1 > \frac{1}{2}$. Let $n \geq 1$. By the Taylor formula with $x = n$,

$$e^n - \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{n!} \int_0^n (n-t)^n e^t \, dt = \frac{e^n}{n!} \int_0^n u^n e^{-u} \, du$$

(substitute $u = n - t$). So the claim $\sum_{k \leq n} n^k/k! > e^n/2$ is equivalent to $\frac{1}{n!} \int_0^n u^n e^{-u} \, du < \frac{1}{2}$. Write $\phi(u) = u^n e^{-u}$. Since $n! = \int_0^\infty \phi$, the claim is

$$\int_0^n \phi(u) \, du < \int_n^\infty \phi(u) \, du. \qquad (*)$$

Key comparison. For $0 < h < n$ let

$$g(h) = \ln \phi(n+h) - \ln \phi(n-h) = n \ln(n+h) - n \ln(n-h) - 2h.$$

Then $g(0) = 0$ and

$$g'(h) = \frac{n}{n+h} + \frac{n}{n-h} - 2 = \frac{2n^2}{n^2 - h^2} - 2 > 0 \qquad (0 < h < n),$$

so $g(h) > 0$, that is, $\phi(n+h) > \phi(n-h)$ for $0 < h < n$.

Integrate. Substituting $u = n - h$ and $u = n + h$,

$$\int_0^n \phi(u) \, du = \int_0^n \phi(n-h) \, dh < \int_0^n \phi(n+h) \, dh = \int_n^{2n} \phi(u) \, du \leq \int_n^\infty \phi(u) \, du.$$

The strict inequality holds because the integrands are continuous and the inequality between them is strict on $(0, n)$. This is $(*)$. □

The official solution is this argument, stated with no motivation for either the substitution or the comparison $\phi(n+h)$ versus $\phi(n-h)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains want to prove "head beats tail" and both reach for a reflection about $k = n$. They split at the object being reflected. The natural attempt reflects the discrete terms $T_{n \pm j}$, and the reflection fails because the discrete sequence is not shaped the right way far from the centre. The model first spends the Remarks to turn the sum into an integral of $\phi(u) = u^n e^{-u}$, and reflects the continuous density ϕ about its peak, where the comparison holds for every h at once.

> *They have given me Taylor's remainder and the Gamma integral for $n!$. They are not decoration. The remainder turns the missing part of e^n into $\int_0^n$ of something, and $n!$ is $\int_0^\infty$ of the same thing. So the whole question is whether $\int_0^n$ is less than half of $\int_0^\infty$.*

> *$u^n e^{-u}$ has its maximum at $u = n$ exactly. So I am comparing the area left of the peak with the area right of the peak. Reflect: compare $\phi(n-h)$ with $\phi(n+h)$. Take logs, because ϕ is a product of a power and an exponential.*

▷ **Cue:** an inequality between a partial sum of $x^k/k!$ and e^x, especially with the Taylor remainder offered ⟶ **Move:** write the remainder as an integral, substitute to get $\int u^n e^{-u}$, and compare with $n! = \int_0^\infty u^n e^{-u}$

▷ **Cue:** "is the area to the left of the peak less than the area to the right?" ⟶ **Move:** reflect about the peak: compare $\phi(p-h)$ with $\phi(p+h)$ pointwise, via $\ln \phi(p+h) - \ln \phi(p-h)$ and its derivative in h

▷ **Cue:** the problem statement supplies formulae "you may assume" ⟶ **Move:** treat them as the intended first line; ask what each one converts into what

Layer 3 THE WHY

Why is the integral the right object, and why does the reflection work there when it fails for the sum? Go to the definitions.

First, what the Remarks are for. The target compares a truncated exponential series with e^n, so the missing piece $e^n - \sum_{k\le n} n^k/k!$ is the Taylor remainder, and the Remarks give it as an integral. Once it is an integral, $n!$ in the same shape ($\int t^n e^{-t}$) is the natural thing to compare it with, because the factor $\frac{1}{n!}$ is sitting in front. The substitution $u = n - t$ is forced by wanting the two integrands to match: $(n-t)^n e^t$ becomes $e^n\, u^n e^{-u}$. So the chain "remainder → integral → same integrand as $n!$" is not inspiration; it is what the given formulae do when you apply them to the given sum.

Second, why $\phi(u) = u^n e^{-u}$ has its peak at n: $\ln\phi = n\ln u - u$ has derivative $n/u - 1$, which vanishes at $u = n$. So $(*)$ is a question about the shape of ϕ on either side of its peak, and reflection about the peak is the direct way to compare the two sides.

Third, why the right side is heavier. The rate at which $\ln\phi$ falls away from the peak is $|n/u - 1|$. At distance h to the left this is $h/(n-h)$; at distance h to the right it is $h/(n+h)$, which is smaller. So ϕ falls faster to the left than to the right, at every distance, and $g'(h) > 0$ is exactly that statement integrated. Underneath it is the convexity of $1/u$: $\ln(n+h) - \ln(n-h) = \int_{n-h}^{n+h} du/u$, and the average of a convex function over an interval centred at n exceeds its value $1/n$ at the centre, which gives $n\ln\frac{n+h}{n-h} > 2h$ with no calculus beyond that one fact. The right tail also continues past $2n$ with nothing to match it, which only helps.

Fourth, why the discrete version fails. The terms T_k are samples of a curve that is not symmetric, and the discrete ratios T_{n+j}/T_{n-j} compare values at lattice points without accounting for the fact that T_k is ϕ divided by $k!$, whose discrete structure (the factor $(n+i)(n-i+1)$) does not reflect cleanly. The integral representation removes the factorials, and what remains is a smooth one-peaked function where reflection is a one-line derivative. That is why the problem supplies the formulae: it is telling you to leave the discrete world.

Context, as a check that $\frac{1}{2}$ is sharp: $e^{-n}\sum_{k\le n} n^k/k! = P(X \le n)$ for X Poisson with mean n, and this tends to $\frac{1}{2}$. Ramanujan's refinement writes $\frac{1}{2}e^n = \sum_{k<n} n^k/k! + \theta_n\, n^n/n!$ with $\frac{1}{3} < \theta_n < \frac{1}{2}$, which implies the problem at once. It is far harder; the problem asks for the half of it that the integral trick gives.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the Remarks: "You may assume as known Taylor's remainder formula ... as well as $n! = \int_0^\infty t^n e^{-t}\,dt$." Write down $e^n - \sum_{k\le n} n^k/k! = \frac{1}{n!}\int_0^n (n-t)^n e^t\,dt$ at once.

2. It suggests converting the discrete inequality to $\int_0^n u^n e^{-u}\,du < \int_n^\infty u^n e^{-u}\,du$ and comparing the integrand on the two sides of its peak.

3. Hesitation occurs after the substitution, when you have $(*)$ and wonder how to compare two integrals over different ranges. Reinforce it away with the question "where is the peak?": the answer $u = n$ is the split point of $(*)$, which says reflect.

4. The trap is working with the discrete terms: pairing T_{n+j} with T_{n-j} works for small j and fails for large j, which draws you into estimates of partial products. The second trap is the CLT, which gives the limit and not the inequality.

5. When a sum resembles a probability or a truncated series, look for an integral identity that removes the factorials. Once the object is a smooth one-peaked function, left-versus-right questions are reflection questions.

6. Reusable decision rule: *to compare the mass on two sides of the peak of a unimodal function, reflect about the peak and compare* $\ln\phi(p+h)$ *with* $\ln\phi(p-h)$*; the sign of the derivative in* h *settles it for every* h *at once.*

1974 B6

1974 B6. For a set with n elements, how many subsets are there whose cardinality (the number of elements in the subset) is respectively $\equiv 0 \pmod 3$, $\equiv 1 \pmod 3$, $\equiv 2 \pmod 3$? In other words, calculate

$$s_{i,n} = \sum_{k \equiv i \,(\mathrm{mod}\ 3)} \binom{n}{k}$$

for $i = 0, 1, 2$. Your result should be strong enough to permit direct evaluation of the numbers $s_{i,n}$ and to show clearly the relationship of $s_{0,n}$ and $s_{1,n}$ and $s_{2,n}$ to each other for all positive integers n. In particular, show the relationships among these three sums for $n = 1000$. [An illustration of the definition of $s_{i,n}$ is $s_{0,6} = \binom{6}{0} + \binom{6}{3} + \binom{6}{6} = 22$.]

Caveat. The OCR of the MAA volume drops the subscripts and binomial coefficients in the definition and in the bracketed illustration. The illustration is reconstructed as $s_{0,6} = \binom{6}{0} + \binom{6}{3} + \binom{6}{6} = 22$, the only choice of three binomial coefficients consistent with the printed total 22 and the definition; the reconstruction was checked against the book's solution under 1974 B-6 (period-6 table, $s_{0,n} + s_{1,n} + s_{2,n} = 2^n$, and $s_{0,1000} = s_{1,1000} = s_{2,1000} - 1 = (2^{1000} - 1)/3$). The book's period-6 table is itself garbled; the table below was recomputed and checked by machine for $n \le 12$ and at $n = 1000$.

Fields: combinatorics; binomial coefficients; roots of unity; linear recurrences **Position:** B6 — last on the paper, but a standard technique closes it in a few lines; the problem's real test is whether you know the roots-of-unity filter **Answer:** $s_{i,n} = \frac{1}{3}\left(2^n + 2\cos\frac{(n-2i)\pi}{3}\right)$. Two of the three sums are always equal and the third differs from them by 1, in a pattern of period 6 in n. For $n = 1000$: $s_{0,1000} = s_{1,1000} = \frac{1}{3}(2^{1000} - 1)$ and $s_{2,1000} = s_{0,1000} + 1 = \frac{1}{3}(2^{1000} + 2)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student knows $\sum_k \binom{n}{k} = 2^n$ and $\sum_k (-1)^k \binom{n}{k} = 0$, which split the subsets by parity. Mod 3 is new, so the student reaches for Pascal's rule $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$. Summing over $k \equiv i$ gives

$$s_{i,n+1} = s_{i,n} + s_{i-1,n} \qquad \text{(indices mod 3)}.$$

Start from $(s_0, s_1, s_2) = (1, 0, 0)$ at $n = 0$ and iterate:

$$(1,0,0),\ (1,1,0),\ (1,2,1),\ (2,3,3),\ (5,6,5),\ (11,11,10),\ (22,21,21),\ (43,43,42), \ldots$$

Every row has two equal entries and a third differing by 1, and the position and sign of the odd one out cycle with period 6. The student conjectures the table, proves it by induction using the recurrence (six cases), and reads off $n = 1000 \equiv 4 \pmod 6$ using $s_0 + s_1 + s_2 = 2^n$.

This is the official solution, and it is correct. Its cost: the pattern must be spotted from data, the induction runs over six cases, the "direct evaluation" the problem asks for comes out as a case table rather than a formula, and nothing in the argument explains why the period is 6 or why the deviations are $\pm 1, \pm 2$.

The model chain. Let $\omega = e^{2\pi i/3}$, so $\omega^3 = 1$ and $1 + \omega + \omega^2 = 0$. For an integer m,

$$\frac{1}{3}\sum_{j=0}^{2} \omega^{jm} = \begin{cases} 1, & 3 \mid m, \\ 0, & 3 \nmid m, \end{cases}$$

since for $3 \nmid m$ the three terms are $1, \omega^m, \omega^{2m}$, which are $1, \omega, \omega^2$ in some order. Apply this with $m = k - i$ and sum against $\binom{n}{k}$:

$$s_{i,n} = \frac{1}{3}\sum_{j=0}^{2} \omega^{-ij} \sum_k \binom{n}{k} \omega^{jk} = \frac{1}{3}\sum_{j=0}^{2} \omega^{-ij}(1+\omega^j)^n.$$

Now $1+\omega = -\omega^2 = e^{i\pi/3}$ and $1+\omega^2 = -\omega = e^{-i\pi/3}$, while $1+\omega^0 = 2$. So

$$s_{i,n} = \frac{1}{3}\left(2^n + e^{i(n-2i)\pi/3} + e^{-i(n-2i)\pi/3}\right) = \frac{1}{3}\left(2^n + 2\cos\frac{(n-2i)\pi}{3}\right).$$

Write $s_{i,n} = \frac{1}{3}(2^n + \varepsilon_{i,n})$ with $\varepsilon_{i,n} = 2\cos\frac{(n-2i)\pi}{3} \in \{2, 1, -1, -2\}$, depending only on n mod 6:

n mod 6	0	1	2	3	4	5
ε_0	2	1	−1	−2	−1	1
ε_1	−1	1	2	1	−1	−2
ε_2	−1	−2	−1	1	2	1

In every column the entries sum to 0 (as they must, since $s_0 + s_1 + s_2 = 2^n$), two are equal and the third differs from them by 3; so two of the $s_{i,n}$ are equal and the third differs by exactly 1. Check: $n = 6$ gives $s_{0,6} = (64+2)/3 = 22$.

For $n = 1000$: $1000 = 6 \cdot 166 + 4$, so $\varepsilon = (-1, -1, 2)$ and

$$s_{0,1000} = s_{1,1000} = \frac{2^{1000}-1}{3}, \qquad s_{2,1000} = \frac{2^{1000}+2}{3} = s_{0,1000} + 1. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin from "split the 2^n subsets by size mod 3" and both know the parity case. They split at how the parity case is generalised. The natural attempt generalises its *proof method* (Pascal's rule, then induction) and has to discover the answer from data. The model generalises its *mechanism*: parity was filtered by evaluating $(1+x)^n$ at $x = \pm 1$, the square roots of unity, so mod 3 is filtered by evaluating at the cube roots of unity. The formula then arrives before any table.

> *Parity comes from $(1+1)^n$ and $(1-1)^n$: adding them kills the odd k. That works because ± 1 are the square roots of 1. For mod 3 I need the cube roots of 1. Put $x = 1, \omega, \omega^2$ into $(1+x)^n$ and average with the right twists.*

> *$1+\omega$ is a sixth root of unity, of modulus 1. So the two correction terms never grow; they rotate by 60° each step. That is why the pattern has period 6 and why the sums differ from $2^n/3$ by less than 1.*

- ▷ **Cue:** a sum of $\binom{n}{k}$ (or of coefficients of any polynomial) over k in a residue class mod m ⟶ **Move:** roots-of-unity filter: $\sum_{k\equiv i} a_k = \frac{1}{m}\sum_{j=0}^{m-1} \omega^{-ij} P(\omega^j)$ with $\omega = e^{2\pi i/m}$
- ▷ **Cue:** a table computed from a recurrence shows a period ⟶ **Move:** write the recurrence as a matrix and find its eigenvalues; a period comes from eigenvalues that are roots of unity (up to a common scaling)
- ▷ **Cue:** "show clearly the relationship" between several counts that sum to a known total ⟶ **Move:** write each as (total/m) + (small correction) and study the corrections

Layer 3 THE WHY

Why the cube roots of unity, and why is the filter not a trick? Start with what is being asked: pick out the terms of $(1+x)^n = \sum \binom{n}{k} x^k$ whose exponent lies in a residue class mod 3. The

only information you have about an exponent k when you substitute a number x is x^k. So you need numbers x for which x^k depends on k only through k mod 3. That is the definition of $x^3 = 1$: the cube roots of unity are exactly the numbers that cannot see anything about k except its residue. Once you are restricted to $1, \omega, \omega^2$, the three values x^k are the three "characters" of $\mathbb{Z}/3$, and averaging them with twists ω^{-ij} picks out one residue, because $1 + \omega + \omega^2 = 0$ (the three points are symmetric about the origin). Parity is the same construction with $x = \pm 1$; you already knew it, and the model's only step is to name its mechanism.

Why is the answer so close to $2^n/3$? The $j = 0$ term contributes $2^n/3$, the fair share. The other two terms involve $(1 + \omega)^n$ and $(1 + \omega^2)^n$, and $|1 + \omega| = 1$: geometrically, 1 and ω are unit vectors 120° apart, so their sum is a unit vector bisecting them, at angle 60°. So the corrections have modulus 1 for every n, they are complex conjugates (so their sum is real), and they rotate by 60° per step. Six steps return them to the start: period 6. Their sum is $2\cos\theta$ with θ a multiple of 60°, hence one of $\pm 1, \pm 2$. That is the whole table, derived rather than observed.

The same fact explains the official induction. The recurrence $s_{i,n+1} = s_{i,n} + s_{i-1,n}$ says the vector (s_0, s_1, s_2) is multiplied by $I + P$ each step, where P is the cyclic shift. The eigenvalues of P are $1, \omega, \omega^2$ (its eigenvectors are the characters above), so the eigenvalues of $I + P$ are $2,\ 1 + \omega,\ 1 + \omega^2$. The component along $(1, 1, 1)$ doubles each step, which is the $2^n/3$; the components in the plane $s_0 + s_1 + s_2 = 0$ are rotated by $\pm 60^\circ$ without stretching. The student who built the table was watching this rotation without the eigenvalues that name it. Kedlaya-style archives would give the filter in one line; the MAA book gives the table. The filter is the reason the table exists.

Finally, a check that the answer meets the problem's demand for "direct evaluation": the formula gives each $s_{i,n}$ in closed form, and since $|\varepsilon_{i,n}| \le 2 < 3$, each $s_{i,n}$ is the integer nearest to $2^n/3$ whenever $|\varepsilon_{i,n}| = 1$, and the other integer within 1 when $|\varepsilon_{i,n}| = 2$. No subset count is further than $\frac{2}{3}$ from a fair third.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "subsets whose cardinality is $\equiv 0, 1, 2 \pmod 3$", i.e. binomial coefficients summed over a residue class. Write down $(1 + x)^n = \sum \binom{n}{k} x^k$ and $\omega = e^{2\pi i/3}$.

2. It suggests the roots-of-unity filter: evaluate $(1 + x)^n$ at $x = 1, \omega, \omega^2$ and combine with weights ω^{-ij}.

3. Hesitation comes at $(1 + \omega)^n$, which looks messy. Reinforce it away by drawing 1 and ω as vectors: their sum is $e^{i\pi/3}$, a sixth root of unity, and the powers are immediate.

4. The trap is not the Pascal-table route (it works) but stopping at the table: presenting six cases spotted from data without a formula, which is weaker than the "direct evaluation" the problem requests, and invites arithmetic slips in reading off $n = 1000$ (which is $\equiv 4 \pmod 6$, not $\equiv 1 \pmod 3$ read as a period-3 pattern).

5. Whenever coefficients are selected by a congruence condition on the index, the m-th roots of unity are the tool, because they are exactly the numbers whose powers depend only on the index mod m.

6. Reusable decision rule: *if a quantity splits into m parts that sum to a known total, write each part as total/m plus a correction, and compute the corrections with the m-th roots of unity; if a recurrence produces a periodic pattern, look for an eigenvalue that is a root of unity.*