DAWN PUTNAM SESSION

Putnam 1967 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1967 — Paper A **5**
1967 A1 5
1967 A2 7
1967 A3 9
1967 A4 11
1967 A5 14
1967 A6 17

Putnam 1967 — Paper B **23**
1967 B1 23
1967 B2 25
1967 B3 28
1967 B4 31
1967 B5 33
1967 B6 36

Putnam 1967 — Paper A

1967 A1

1967 A1. Let $f(x) = a_1 \sin x + a_2 \sin 2x + \cdots + a_n \sin nx$, where $a_1, a_2, \cdots, a_n$ are real numbers and where n is a positive integer. Given that $|f(x)| \leq |\sin x|$ for all real x, prove that $|a_1 + 2a_2 + \cdots + na_n| \leq 1$.

Fields: analysis, trigonometric polynomials, differentiation **Position:** A1 (opener) **Answer:** The bound $|a_1 + 2a_2 + \cdots + na_n| \leq 1$ follows because that quantity is exactly $f'(0)$, and the hypothesis $|f(x)| \leq |\sin x|$ passes to the limit $x \to 0$ unchanged.

Layer 1 THE WHAT — surface analysis

The natural attempt. You are handed a sine polynomial and a target quantity, $a_1 + 2a_2 + \cdots + na_n$, that plainly wants you to differentiate: $\sin(kx)$ differentiates to $k\cos(kx)$, so $f'(x) = a_1 \cos x + 2a_2 \cos 2x + \cdots + na_n \cos nx$, and $f'(0) = a_1 + 2a_2 + \cdots + na_n$ falls out in one line. That part is not the problem. The problem is what you do with the hypothesis $|f(x)| \leq |\sin x|$, which is a statement about f's *values*, not about its coefficients. The natural instinct with a sine polynomial and a bound on its coefficients is to reach for the orthogonality relations that produced those coefficients in the first place: $a_k = \frac{2}{\pi}\int_0^\pi f(x)\sin(kx)\,dx$. You write down $a_1 + 2a_2 + \cdots + na_n = \frac{2}{\pi}\int_0^\pi f(x)\left(\sum_{k=1}^n k\sin kx\right)dx$, apply the triangle inequality and the hypothesis $|f(x)| \leq |\sin x|$ to pull the bound inside the integral, and are left needing $\frac{2}{\pi}\int_0^\pi |\sin x|\,|\sum k\sin kx|\,dx \leq 1$. This is true but is now a harder inequality than the one you started with – you have replaced a clean target with an integral of a sum of n oscillating terms against $|\sin x|$, and bounding that sum termwise (via $|\sin kx| \leq 1$, say) throws away far too much to recover the sharp constant 1. You can feel that the machinery is heavier than the conclusion deserves, but the orthogonality route is the one every sine-series problem has trained you to reach for first.

The model chain. Do not decompose the sum into its individual coefficients at all – work with f as a single function and extract the target quantity as a limit. Since $f(0) = 0$ (every term is a sine), the difference quotient

$$\frac{f(x) - f(0)}{x} = \frac{f(x)}{x}$$

converges to $f'(0) = a_1 + 2a_2 + \cdots + na_n$ as $x \to 0$. Now bring in the hypothesis at exactly this limit: for every $x \neq 0$,

$$\left|\frac{f(x)}{x}\right| = \left|\frac{f(x)}{\sin x}\right| \cdot \left|\frac{\sin x}{x}\right| \leq 1 \cdot \left|\frac{\sin x}{x}\right|,$$

using $|f(x)| \leq |\sin x|$, i.e. $|f(x)/\sin x| \leq 1$ wherever $\sin x \neq 0$. As $x \to 0$, the right-hand side $\to 1$ and the left-hand side $\to |f'(0)|$, and a non-strict inequality survives the limit. Hence

$$|a_1 + 2a_2 + \cdots + na_n| = |f'(0)| \leq 1,$$

with no coefficient ever isolated, no orthogonality integral written, and no case-work on n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $f'(0) = a_1 + 2a_2 + \cdots + na_n$ in the first line – that identification is not where they split. They split immediately after, on how to use $|f(x)| \leq |\sin x|$: the natural attempt treats it as a fact about the whole function to be fed into the orthogonality formula that recovers each a_k separately; the model treats it as a fact that can be evaluated at the single point where the target quantity already lives, namely the derivative at $x = 0$, and takes it there directly.

> *I already have the target as $f'(0)$ – a value built entirely out of what f does* near $x = 0$. *I don't need to know anything about f away from 0, so I shouldn't route the hypothesis through an integral over the whole interval $[0, \pi]$; I should ask what the hypothesis says right at the point I care about, which means dividing by x and taking a limit, exactly the way the derivative itself is built.*

Layer 3 THE WHY

The reason the limit route is not a trick but the forced move is that $f'(0)$ is, by definition, $\lim_{x\to 0} \frac{f(x)-f(0)}{x}$ – and definitions are where "of course" arguments live. The problem does not hand you $f'(0)$ as an abstract symbol; it hands you $a_1 + 2a_2 + \cdots + na_n$, and the only reason that sum equals $f'(0)$ is the definition of the derivative applied to a sine series. Once you have committed to that identification, the honest next question is not "how were the a_k built?" (orthogonality) but "how is $f'(0)$ itself built?" (a limit of a ratio), because that is the actual definition of the object you are trying to bound. The hypothesis $|f(x)| \leq |\sin x|$ is a pointwise statement, true at every real x including every x in a punctured neighbourhood of 0 – so it is available to divide through by x at every one of those points, and a limit of true pointwise inequalities (non-strict) is again true. Dividing by $\sin x$ rather than by x directly is the one extra step that makes the comparison exact: $f(x)/x = (f(x)/\sin x) \cdot (\sin x/x)$ splits the ratio into a factor bounded by the hypothesis and a factor with the classical limit 1, so the bound 1 in the problem's conclusion is not an artefact of estimation but is inherited term-for-term from the sharpest possible reading of $|f(x)| \leq |\sin x|$. Orthogonality, by contrast, is built to recover *individual* Fourier coefficients from global information about f; it is the right tool when you need a_k itself, and the wrong tool when you need a specific linear combination that is secretly a local, single-point object. The problem's phrasing – "for all real x" attached to a global-looking inequality, married to a target quantity that is visibly a derivative – is the cue that the generality of the hypothesis is only ever going to be cashed locally, at one point, and the machinery should match that scale.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape of the target: $a_1 + 2a_2 + \cdots + na_n$ is the coefficient pattern you get from differentiating $\sum a_k \sin kx$ term by term and evaluating at $x = 0$. On seeing it, write down $f'(0) = a_1 + 2a_2 + \cdots + na_n$ before doing anything else.

2. That identification suggests working with the derivative as a limit of a difference quotient, rather than with the Fourier coefficients individually – because the target is now a single local quantity, not a weighted sum of separately-extracted pieces.

3. Hesitation occurs right after computing $f'(0)$, when the instinct is to ask "how do I recover each a_k from f?" and reach for orthogonality. Reinforce against it by asking instead: "is the quantity I actually need already a value of f or of f' at a single point?" If yes, the hypothesis should be applied at that point, not integrated over the whole domain.

4. The trap is over-engineering: applying the full machinery used to construct Fourier coefficients (integration, orthogonality) to a problem that only needs the value of a derivative at one point. The heavier tool produces a true but strictly weaker and harder-to-tighten bound.

5. Recognise this faster next time by checking, as soon as a target quantity is identified as $f^{(k)}(0)$ or $f(0)$ for some specific point, whether the hypothesis is already stated as a pointwise inequality valid in a neighbourhood of that point – if so, a limiting/difference-quotient argument is available and is almost always cheaper than reconstructing the object from series coefficients.

6. Decision rule: when a target linear combination of coefficients of a series is recognisable as a derivative (or a value) of the series at a specific point, and the hypothesis is a pointwise bound on the series holding near that point, bound the derivative directly via the difference-quotient limit, comparing against the hypothesis's own limiting behaviour at that point – do not route through coefficient-extraction machinery (orthogonality, Parseval) unless the individual coefficients themselves are the target.

1967 A2

1967 A2. Define S_0 to be 1. For $n \geq 1$, let S_n be the number of $n \times n$ matrices whose elements are nonnegative integers with the property that $a_{ij} = a_{ji}$, $(i, j = 1, 2, \cdots, n)$ and where $\sum_{j=1}^{n} a_{ij} = 1$, $(i = 1, 2, \cdots, n)$. Prove
(a) $S_{n+1} = S_n + nS_{n-1}$,
(b) $\displaystyle\sum_{n=0}^{\infty} S_n \frac{x^n}{n!} = \exp(x + x^2/2)$,
where $\exp(x) = e^x$.

Fields: combinatorics, matrices, exponential generating functions, involutions **Position:** A2 (mid-paper) **Answer:** S_n is the number of involutions on n labelled elements; part (a) is the standard involution recursion and part (b) is the corresponding exponential generating function $\exp(x + x^2/2)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem is dressed as matrix combinatorics – nonnegative integer entries, symmetric, row sums 1 – and the natural instinct is to attack it as stated, at the level of matrix entries. You start building the matrix row by row: row 1 has nonnegative integers summing to 1, so exactly one entry is 1 and the rest are 0. Fine, but which entry? You case-split on where the 1 in row 1 sits, say column k. If $k = 1$, the rest of the matrix (rows/columns $2, \ldots, n$) is a free $(n-1) \times (n-1)$ instance of the same problem, contributing S_{n-1}. If $k \neq 1$, symmetry forces $a_{k1} = 1$ too, which uses up the entire 1 in row k's sum, so row k is fully determined and rows $1, k$ can both be deleted along with columns $1, k$, leaving a free $(n-2) \times (n-2)$ instance. This is exactly the right case-split, and if you push it through carefully you do land on $S_n = S_{n-1} + (n-1)S_{n-2}$ (a shift of the stated recursion). The natural attempt is not wrong – but it is easy to get lost inside it: you are tracking row sums, column sums, symmetry, and nonnegativity simultaneously, re-deriving from scratch, entry by entry, a fact that is really about a much more familiar object once you see it, and part (b)'s exponential generating function then looks like it has come from nowhere, forcing you into an unmotivated differential-equation manipulation of $F(x) = \sum S_n x^n/n!$ with no sense of why $\exp(x + x^2/2)$ specifically should appear.

The model chain. Translate the matrix condition into combinatorial language before doing anything else. A row of nonnegative integers summing to 1 has exactly one entry equal to 1; since this holds for every row, the matrix is 0-1 valued with exactly one 1 in every row. A 0-1 matrix with exactly one 1 per row *and*, by the same argument applied to columns (the column sums are automatically 1 too, since each column's single 1 comes from whichever row put it there, and no two rows can share a column because that would force a second 1 somewhere), exactly one 1 per column, is precisely a **permutation matrix**: $a_{ij} = 1$ iff $\sigma(i) = j$ for some

permutation σ of $\{1,\dots,n\}$. The extra condition $a_{ij} = a_{ji}$ says $\sigma(i) = j \iff \sigma(j) = i$, i.e. $\sigma^{-1} = \sigma$: σ **is an involution** (a permutation that is its own inverse, equivalently a product of disjoint fixed points and 2-cycles). So S_n is exactly the number of involutions on n labelled points.

(a) Look at where the involution sends the element n. Either $\sigma(n) = n$ (a fixed point), and the rest of σ is an arbitrary involution on the remaining $n-1$ elements, contributing S_{n-1}; or $\sigma(n) = k$ for some $k \neq n$, which forces $\sigma(k) = n$ (a transposition), and there are $n-1$ choices of k, each leaving an arbitrary involution on the remaining $n-2$ elements, contributing $(n-1)S_{n-2}$. Hence $S_n = S_{n-1} + (n-1)S_{n-2}$, which is $S_{n+1} = S_n + nS_{n-1}$ after shifting the index by one.

(b) An involution is exactly a partition of $\{1,\dots,n\}$ into blocks of size 1 or 2 (fixed points and transpositions), with no further structure needed on a block since a block of size 1 or 2 has a unique involution-restriction to it. The exponential formula for labelled structures says: if a "connected piece" of size m can be built in c_m ways, the EGF for structures built from disjoint unions of such pieces is $\exp\left(\sum_m c_m x^m/m!\right)$. Here a piece of size 1 can be built in $c_1 = 1$ way (a fixed point) and a piece of size 2 in $c_2 = 1$ way (the unique transposition on 2 labelled points; there is only one way to pair two specific labels), and no pieces of size ≥ 3 occur. So the EGF is $\exp(x^1/1! \cdot 1 + x^2/2! \cdot 1) = \exp(x + x^2/2)$, i.e. $\sum_{n\geq 0} S_n x^n/n! = \exp(x + x^2/2)$, exactly as claimed. (Equivalently, differentiate $F(x) = \sum S_n x^n/n!$ term by term using the recursion from (a) to get $F'(x) = (1+x)F(x)$, hence $F(x) = F(0)e^{x+x^2/2} = e^{x+x^2/2}$ since $F(0) = S_0 = 1$ – the same content, reached via a differential equation instead of the exponential formula.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains ultimately produce the same recursion. They split at the very first decision: whether to reason about entries of the matrix directly, or to first recognise the matrix as encoding a more familiar combinatorial object. The natural attempt stays at the level of "a 1 in row 1, column k" throughout; the model spends one line converting "symmetric 0/1 matrix, one 1 per row" into "permutation matrix," and one more converting "symmetric permutation matrix" into "involution," before ever touching a recursion.

> *A row of nonnegative integers summing to* 1 *is just a single* 1 *in an unknown position and zeros elsewhere – that is not really a matrix condition, it is a function condition: each row index is being sent to a column index. I should stop thinking about the entries a_{ij} as numbers to case-split on and start thinking about the function $i \mapsto$ (the column of row i's* 1*), which turns the whole object into a permutation, and then the symmetry condition becomes a condition on that permutation, not on the array.*

Layer 3 THE WHY

The identification "row sums 1, nonnegative integers $\Rightarrow$ permutation matrix" is not an observation about this particular problem; it is forced by the definitions involved. Nonnegative integers summing to 1 can only be the multiset $\{1,0,0,\dots,0\}$ in some order – there is no other way to write 1 as a sum of nonnegative integers with more than one term unless all but one term vanish. So "row sum 1" is definitionally equivalent to "exactly one 1 per row," and once every row has exactly one 1, the map sending each row to the column of its 1 is automatically a function from $\{1,\dots,n\}$ to itself; symmetry (which is column information folded back into row information) is exactly what makes this function a bijection with $\sigma^{-1} = \sigma$, because $a_{ij} = a_{ji}$ says "i points to j" and "j points to i" are the same fact. The reason this reduction is worth making *before* attempting the recursion, rather than discovering it by accident partway through a case-split, is that it converts an unfamiliar linear-algebra-flavoured counting problem into an extremely well-trodden one (counting involutions), for which both the recursion and the generating function are standard consequences of a single combinatorial fact: an involution is

nothing but a partition into singletons and pairs. The generating function $\exp(x + x^2/2)$ then stops being a mysterious answer to reverse-engineer from a differential equation and becomes the direct statement of the exponential formula: x is the EGF of "a single labelled point, no further choices," and $x^2/2$ is the EGF of "a pair of labelled points forming the unique 2-cycle on them," and a disjoint union of independent labelled pieces always exponentiates its EGF. Where the official solution presents the differential-equation route to $F(x) = \exp(x + x^2/2)$ without saying why that particular exponent should appear, this is where the archive earns its existence: the exponent $x + x^2/2$ is not solved for, it is read off directly from the two ways an involution can use one more label.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "nonnegative integers... row sums equal to 1." On seeing it, write down immediately: a nonnegative integer row summing to 1 has exactly one nonzero entry, itself equal to 1 – this is a 0-1 matrix with one 1 per row before any other condition is even used.

2. That cue suggests reformulating the matrix as a function $i \mapsto \sigma(i)$ recording the column of row i's 1; the further symmetry condition then suggests checking what it says about σ, namely $\sigma = \sigma^{-1}$, i.e. an involution – turning the problem into counting involutions, a named and well-studied object.

3. Hesitation occurs right after correctly identifying the case-split on row 1's entry, when it is tempting to just push the matrix case-work through to the recursion without stepping back. Reinforce against it by asking: does this array condition have a name once translated into functions/permutations? If a translation exists, do it before case-splitting, since the case-split becomes cleaner (and its generating function becomes obvious) on the translated object.

4. The trap is deriving the correct recursion by brute matrix case-work in part (a), then treating part (b)'s generating function as an unrelated algebraic exercise to be solved by manipulating a differential equation for $F(x)$, with no combinatorial meaning attached to $\exp(x + x^2/2)$ until after the algebra is done.

5. Recognise this faster next time by treating "symmetric 0-1 matrix with one entry per row and column" as a fixed vocabulary item meaning "involution," the way "sum of nonnegative integers equal to 1" should immediately mean "indicator of a single choice" – both translations should happen in the first thirty seconds, before any case-split.

6. Decision rule: for any combinatorial object built from a matrix or array with strong local constraints (each row/column sums to a small fixed number, symmetry, 0-1 entries), first ask whether the array is really encoding a permutation, a graph, or a set partition in disguise; if so, derive both the recursion and its generating function from the exponential formula applied to that named object's building blocks, rather than manipulating array entries or a differential equation directly.

1967 A3

1967 A3. Consider polynomial forms $ax^2 - bx + c$ with integer coefficients which have two distinct zeros in the open interval $0 < x < 1$. Exhibit with a proof the least positive integer value of a for which such a polynomial exists.

Fields: algebra, quadratics, Diophantine bounding via integer values of a polynomial **Position:** A3 (mid-paper) **Answer:** The least such a is 5, achieved by $5x^2 - 5x + 1$, whose roots are $\frac{5 \pm \sqrt{5}}{10}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You write $f(x) = ax^2 - bx + c$ with roots $r, s \in (0,1)$, $r \neq s$, $a > 0$ (wanting a positive integer). The standard toolkit for "two distinct real roots in an interval"

is the location-of-roots conditions you have drilled for inequalities-with-quadratics problems: $af(0) > 0$, $af(1) > 0$, the discriminant $b^2 - 4ac > 0$, and the vertex $x = b/(2a)$ strictly between 0 and 1. You write all four out in terms of a, b, c and start testing small positive integers $a = 1, 2, 3, 4$ by hand, searching for integers b, c satisfying all four simultaneously. For $a = 1$: $f(0) = c > 0$, $f(1) = 1 - b + c > 0$, $0 < b/2 < 1$ so $b = 1$, and $b^2 - 4ac = 1 - 4c > 0$ forces $c \leq 0$, contradicting $c > 0$ – fails. You repeat for $a = 2, 3, 4$, each time working through the four conditions by hand, and each time everything collapses for the same underlying reason but you have to re-derive the contradiction from scratch in each case, because you have not isolated *why* it keeps failing. Eventually $a = 5$ works (e.g. $b = 5, c = 1$), and you have your answer – but you have proved only that $a = 1, 2, 3, 4$ fail by direct search, not that they *must* fail, and you have no single inequality that says "$a \geq 5$" in one line; the argument is four separate finite verifications bolted together, which is correct but does not generalise and does not explain the number 5 itself.

The model chain. Do not search integers b, c at all. Instead, evaluate f at the two integer points already forced into the problem by the interval's endpoints, $x = 0$ and $x = 1$, and multiply. Writing $f(x) = a(x - r)(x - s)$, we get $f(0) = ars$ and $f(1) = a(1 - r)(1 - s)$, so

$$f(0)\, f(1) = a^2\, r(1 - r)\, s(1 - s).$$

Since $0 < r < 1$ and $0 < s < 1$, both $r(1 - r)$ and $s(1 - s)$ lie in $\left(0, \frac{1}{4}\right]$ by AM–GM (equality only at $r = \frac{1}{2}$, resp. $s = \frac{1}{2}$), and since $r \neq s$ they cannot both equal $\frac{1}{2}$ simultaneously, so the product is strictly less than $\frac{1}{16}$:

$$0 < f(0)f(1) < \frac{a^2}{16}.$$

On the other hand, $f(0) = c$ and $f(1) = a - b + c$ are *integers* (as $a, b, c \in \mathbb{Z}$), and both are positive (each equals a times a positive quantity, with $a > 0$), so their product is a *positive integer*, hence $f(0)f(1) \geq 1$. Combining,

$$1 \leq f(0)f(1) < \frac{a^2}{16} \implies a^2 > 16 \implies a > 4 \implies a \geq 5,$$

since a is a positive integer. Finally exhibit $a = 5$: take $b = 5, c = 1$, so $f(x) = 5x^2 - 5x + 1$ has discriminant $25 - 20 = 5 > 0$ and roots $\frac{5 \pm \sqrt{5}}{10} \approx 0.276, 0.724$, both in $(0, 1)$ and distinct. Hence the least such a is $\boxed{5}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains correctly set up the standard location-of-roots conditions ($af(0) > 0$, $af(1) > 0$, discriminant positive, vertex in $(0, 1)$) as the governing facts. They split on what to do with $f(0)$ and $f(1)$ once known positive: the natural attempt treats them as two of several separate real-number inequalities to be satisfied by searching integer b, c for each candidate a in turn; the model instead multiplies $f(0)$ and $f(1)$ together into a single quantity, bounds that product from above using only the fact that $r, s \in (0, 1)$ (never touching b), and separately uses integrality to bound it from below by 1 – collapsing four conditions and a search over a into one inequality solved once.

> *$f(0)$ and $f(1)$ are not just positive reals here – they are positive* integers, *because a, b, c are integers. A positive integer is at least 1, which is a much stronger statement than "positive." If I multiply $f(0)$ and $f(1)$ together, I get a single positive integer that I can also bound above purely from the roots living in $(0, 1)$ (via AM–GM on $r(1 - r)$), without ever needing to know b – so I get a numerical inequality in a alone, in one line, instead of four separate conditions to search over.*

Layer 3 THE WHY

The whole argument is powered by the gap between "positive real number" and "positive integer": the four location-of-roots conditions only ever say $f(0) > 0$ and $f(1) > 0$ as *real* inequalities, and a real inequality $f(0)f(1) > 0$ gives nothing quantitative – it is consistent with $f(0)f(1)$ being arbitrarily small. It is the problem's own hypothesis, integer coefficients, that upgrades ">0" to "≥ 1" for a product of two integers, and that upgrade is precisely what turns a qualitative root-location fact into a numerical bound on a. This is why multiplying $f(0)$ and $f(1)$ is the natural move once you notice they are integers: multiplication is the operation that turns two separate "≥ 1" facts (if you had them) into one, but here it does something sharper – it is the operation for which the upper bound (via AM–GM, $r(1-r) \leq \frac{1}{4}$) is cheapest to compute, because $f(0)f(1) = a^2\, r(1-r)s(1-s)$ packages both roots into a single symmetric expression that AM–GM handles in one step per factor. The strict inequality $r(1-r)s(1-s) < \frac{1}{16}$ rather than $\leq \frac{1}{16}$ is where the problem's insistence on *two distinct* zeros is spent: if $r = s = \frac{1}{2}$ were allowed, the bound would only be $\leq \frac{1}{16}$ and $a^2 \geq 16$ would permit $a = 4$, so distinctness is not a throwaway hypothesis – it is exactly the fact that excludes $a = 4$ and forces the strict jump to $a \geq 5$. Where a search over $a = 1, 2, 3, 4$ shows failure by exhaustion without ever saying why 4 specifically is the last one to fail, this bound shows it directly: $a = 4$ fails because it is the unique integer just below the threshold 4 that the distinctness of the roots forces the product $f(0)f(1)$ past.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "integer coefficients" combined with "roots strictly inside $(0, 1)$." On seeing it, write down: $f(0)$ and $f(1)$ are forced to be positive integers (not just positive reals) by the location conditions, so any inequality built from them is really an inequality between integers.

2. That cue suggests evaluating f at the interval's endpoints and multiplying, rather than searching integer coefficients directly – since $f(0)f(1)$ is a single integer quantity that can be bounded above using the roots alone (via AM–GM) and bounded below using integrality alone (it is ≥ 1).

3. Hesitation occurs right after writing down the four standard location-of-roots conditions, when the reflex is to start plugging in small integer values of a and searching for b, c. Reinforce against it by asking: is there a single algebraic combination of these conditions (not a search) that turns directly into a numerical bound on a?

4. The trap is proving $a = 1, 2, 3, 4$ fail one at a time by exhaustive search, which is correct but supplies no reason the pattern stops at 4 and gives no route to problems with a larger interval or different coefficient bounds; it also risks a computational slip in one of the four cases going unnoticed.

5. Recognise this faster next time by treating "roots forced into a bounded open interval" plus "integer coefficients" as an immediate cue to evaluate the polynomial at the interval's endpoints and use AM–GM (or a similar symmetric bound) on the resulting product, converting a location-of-roots problem into a single Diophantine inequality.

6. Decision rule: when a polynomial with integer coefficients is required to have all its roots strictly inside a bounded interval (p, q), evaluate the polynomial at p and q, multiply the two integer values together, bound the product above using AM–GM (or the interval's own structure) on the root locations, and bound it below by 1 using integrality – this converts a search over coefficients into one inequality in the leading coefficient alone.

1967 A4

1967 A4. Show that if $\lambda > \frac{1}{2}$ there does not exist a real-valued function u such that for all x in the closed interval $0 \le x \le 1$, $u(x) = 1 + \lambda \int_x^1 u(y)\, u(y-x)\, dy$.

Caveat. The MAA OCR renders the threshold as "$X > i$" and the integral as "λf_x^1"; the statement is restored as $\lambda > \frac{1}{2}$ with $\int_x^1$. Identity check: the book's solution under 1967 A-4 integrates the equation over $0 \le x \le 1$, obtains $\lambda a^2 - 2a + 2 = 0$ for $a = \int_0^1 u$, and observes the roots are non-real exactly when $\lambda > \frac{1}{2}$, which proves this statement. The problem tacitly assumes u is regular enough for the integral to exist; I take u continuous on $[0,1]$ (the argument goes through verbatim for any Lebesgue-integrable u, using Tonelli–Fubini and absolute continuity of $\int_0^y u$). Nothing is claimed here about existence of solutions when $\lambda \le \frac{1}{2}$; the problem does not ask for it, and the argument below gives only a necessary condition.

Fields: real analysis (integral equations, Fubini, change of variables) **Position:** A4 — mid-paper; short once the right scalar is found, impenetrable before **Answer:** No such u exists. Integrating the equation over $x \in [0,1]$ and writing $a = \int_0^1 u$ gives $a = 1 + \frac{\lambda}{2} a^2$, a real quadratic in a with discriminant $4 - 8\lambda < 0$ when $\lambda > \frac{1}{2}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. A nonlinear integral equation for an unknown function. The trained reflex is to turn it into something you know how to solve. Three standard moves come to mind, and a strong student tries them in roughly this order.

First, guess. Try $u \equiv c$: the right side is $1 + \lambda c^2(1-x)$, which depends on x, so no constant works. Try u linear, $u(x) = p + qx$: the right side becomes a cubic in x, and matching coefficients forces $q = 0$ and then fails as before. Guessing dies.

Second, differentiate. The limit x appears both in the lower limit and inside the integrand, so

$$u'(x) = -\lambda\, u(x)u(0) - \lambda \int_x^1 u(y)\, u'(y-x)\, dy,$$

which is not a differential equation at all: it still has an integral in it, now of u'. Differentiating again makes it worse.

Third, iterate. Picard iteration $u_{n+1} = 1 + \lambda \int_x^1 u_n(y) u_n(y-x)\, dy$ from $u_0 = 1$ gives $u_1 = 1 + \lambda(1-x)$, $u_2 = \ldots$, and for small λ a contraction argument would even produce a solution. But the problem asks you to prove *non*-existence, and a failed iteration proves nothing: divergence of one scheme is not impossibility.

All three attempts share one feature. They try to find u. At this point the student is stuck with no wall to push against, which is worse than a wall.

The model chain. Suppose, for contradiction, that u is continuous on $[0,1]$ and satisfies the equation for every $x \in [0,1]$. Integrate both sides over $x \in [0,1]$ and put $a = \int_0^1 u(x)\, dx$:

$$a = 1 + \lambda \int_0^1 \int_x^1 u(y)\, u(y-x)\, dy\, dx.$$

The double integral is over the triangle $T = \{(x,y) : 0 \le x \le y \le 1\}$. Change variables to $s = y$, $t = y - x$; the map $(x,y) \mapsto (s,t)$ has Jacobian determinant of absolute value 1, and it carries T onto $T' = \{(s,t) : 0 \le t \le s \le 1\}$. So

$$\iint_T u(y)u(y-x)\, dx\, dy = \iint_{T'} u(s)u(t)\, ds\, dt.$$

The square $[0,1]^2$ is T' together with its mirror image $\{t \geq s\}$, overlapping only on the diagonal, which has area 0. The integrand $u(s)u(t)$ is symmetric in s, t, so each half carries the same integral, and

$$\iint_{T'} u(s)u(t)\,ds\,dt = \tfrac{1}{2}\Big(\int_0^1 u\Big)^2 = \tfrac{1}{2}a^2.$$

(Equivalently, with $U(y) = \int_0^y u$, the inner integral is $U(y)$ and $\int_0^1 u(y)U(y)\,dy = \big[\tfrac{1}{2}U^2\big]_0^1 = \tfrac{1}{2}a^2$; this is the book's route.)

Hence $a = 1 + \frac{\lambda}{2}a^2$, that is,

$$\lambda a^2 - 2a + 2 = 0.$$

For $\lambda > \frac{1}{2}$ this is a genuine quadratic ($\lambda \neq 0$) with discriminant $4 - 8\lambda < 0$, so it has no real root. But a is a real number, the integral of a real continuous function. Contradiction; no such u exists.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains begin at the same equation. The natural attempt treats it as an equation for the function u and tries to solve it (guess, differentiate, iterate). The model treats it as a family of real equations, one for each x, and adds them all up (integrates over x) to produce one equation for one real number. The split is the line "integrate both sides over $x \in [0,1]$".

> *I'm asked to show something does* not *exist, so I don't need u; I need one consequence of the equation that is impossible. The cheapest consequence is a single number. What number can I extract that the equation controls? Integrate over x and see what the right side becomes.*

> *The integrand $u(y)u(y-x)$ is u at two points; if I rename the two points s and t the region becomes half of a square, and half a square of $u(s)u(t)$ is half of $(\int u)^2$. The equation collapses to a quadratic in $\int u$.*

> *The threshold is $\frac{1}{2}$, which smells of a discriminant. If I get a quadratic in a real quantity, $\lambda > \frac{1}{2}$ should be where it stops having real roots.*

Layer 3 THE WHY

Why is integrating the equation the natural move and not a rabbit from a hat? Start from the logical form of the task. "Show there does not exist" is a universal negative: for every candidate u, something breaks. You cannot check every candidate, so you need a property that every solution would have to satisfy and that no real object can satisfy. The smallest such property is a statement about a single real number. So the first question is not "what is u?" but "what scalar does this equation determine?". That reframing is the whole problem.

Which scalar? There are only a few natural ones for a function on $[0,1]$: a value such as $u(0)$ or $u(1)$, or the integral $\int_0^1 u$. Values are tried first and fail informatively: $u(1) = 1$ (the integral is over a point), and $u(0) = 1 + \lambda \int_0^1 u^2$, which introduces $\int u^2$, another unknown. The integral is the one scalar that an integral equation is built to produce, because integrating the right side over x turns the inner integral into a double integral, and double integrals of products $u(\cdot)u(\cdot)$ are exactly what factorise.

Now the factorisation. The MAA solution says "interchange the order of integration", substitutes $z = y - x$, and moves on. Here is where it comes from. The expression $\int_x^1 u(y)u(y-x)\,dy$ is a correlation: it pairs the value of u at a point with its value a distance x earlier, over all

admissible positions. Integrating over all lags $x \geq 0$ therefore sums $u(s)u(t)$ over *every* ordered pair of points with $t \leq s$, each exactly once. That is the continuous version of the finite identity

$$\sum_{0 \leq j \leq i \leq n} u_i u_j = \tfrac{1}{2}\Big(\big(\textstyle\sum u_i\big)^2 + \sum u_i^2 \Big),$$

and in the continuum the diagonal term vanishes because a line has zero area. So "sum over all lags of the correlation equals half the square of the total" is not a trick; it is counting ordered pairs. Once you see $u(y)u(y-x)$ as "u at two points", the change of variables $s = y$, $t = y - x$ is forced, and the half-square is what the region is.

Finally, why a discriminant? A self-convolution equation of the shape $u = 1 + \lambda(u * u)$ becomes, under any additive functional that turns $*$ into multiplication, a scalar equation $a = 1 + \lambda\kappa a^2$. Quadratics in a real unknown can fail to have real roots, and the problem's threshold $\frac{1}{2}$ is the value of λ where $1 - 4 \cdot \frac{\lambda}{2} \cdot 1$ changes sign. The answer's shape tells you the proof's shape: when a problem's threshold is a clean rational in a parameter multiplying a quadratic term, look for the discriminant.

A last honesty point. This argument proves non-existence for $\lambda > \frac{1}{2}$ and says nothing for $\lambda \leq \frac{1}{2}$. It would be a mistake to write "and for $\lambda \leq \frac{1}{2}$ a solution exists"; that needs a separate argument that neither the book nor this analysis supplies.

Layer $+\alpha$ BACKTESTING

1. The cue is the word "not" in "there does not exist". Write down at once: "assume u exists; find a real number the equation pins down".

2. Integrate the equation over the whole interval to produce an equation for $a = \int_0^1 u$, then evaluate the double integral by symmetry.

3. Hesitation comes at the double integral $\iint u(y)u(y-x)$, which looks unevaluable. Reinforce with "two copies of u at two points: rename the points and look at the region".

4. The trap is trying to solve the equation — guess, differentiate, iterate — when the question only asks you to refute it. Differentiation is the worst of these: it produces an integro-differential equation that is harder than the original.

5. Any non-existence claim about a function satisfying an equation: look for a scalar (integral, value, norm) that the equation forces to satisfy an impossible real equation. A parameter threshold like $\frac{1}{2}$ or $\frac{1}{4}$ is a discriminant in waiting.

6. To prove no function satisfies a functional or integral equation, apply the linear functional that turns the equation's nonlinearity into a product of scalars (for convolution-type terms, integration over everything), then show the resulting scalar equation has no real solution.

▷ **Cue:** "show that there does not exist a function such that" an equation holds $\longrightarrow$ **Move:** do not solve; extract one scalar consequence and show it is impossible

▷ **Cue:** an integral equation containing $\int_x^1 u(y)u(y-x)\,dy$ $\longrightarrow$ **Move:** integrate over $x \in [0,1]$; the double integral is $\frac{1}{2}(\int_0^1 u)^2$

▷ **Cue:** a double integral of $u(\text{point}_1)\,u(\text{point}_2)$ over a triangle $\longrightarrow$ **Move:** rename the two points s, t; the triangle is half the square, and the symmetric integrand gives half of $(\int u)^2$

▷ **Cue:** a parameter threshold such as $\lambda > \frac{1}{2}$ in a problem with a quadratic nonlinearity $\longrightarrow$ **Move:** expect a quadratic in some real quantity whose discriminant changes sign at the threshold

▷ **Cue:** temptation to differentiate an integral equation with the variable in the limit and in the integrand $\longrightarrow$ **Move:** check first whether integrating instead removes the variable entirely

1967 A5

1967 A5. Show that in a convex region in the plane whose boundary contains at most a finite number of straight line segments and whose area is greater than $\frac{\pi}{4}$ there is at least one pair of points a unit distance apart.

Caveat. The MAA OCR renders the area bound as "$r/4$"; it is restored as $\pi/4$. Identity check: the book's solution under 1967 A-5 takes a maximal diameter $2d < 1$, bounds the region between upper and lower boundary functions, and compares the area with $\int_{-1/2}^{1/2} \sqrt{1-4x^2}\,dx = \pi/4$, which proves exactly this statement. The OCR prints the key inequality as "$f(x) + g(-x) < \sqrt{1+4x^2}$"; the sign is a scanning error, and it must be $\sqrt{1-4x^2}$ (with $+$ the final integral would not equal $\pi/4$ and the bound would go the wrong way). As the book's own comment says, the hypothesis about finitely many line segments is not needed; the proof below does not use it.

Fields: geometry (convexity, isodiametric inequality), integral calculus **Position:** A5 — the paper's geometry problem; the constant $\pi/4$ is sharp, so no slack is available **Answer:** True. If no two points are at distance 1, convexity forces every distance to be below 1, so the diameter is $D \le 1$; then the area is at most $\pi D^2/4 \le \pi/4$ (the isodiametric inequality, proved by pairing each upper boundary point with a lower boundary point on the reflected vertical line).

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student sees the reduction at once, and it is correct. If the region contains two points P, Q with $|PQ| \ge 1$, then by convexity the whole segment PQ lies in the region, and the point R on it with $|PR| = 1$ gives the required pair. So the claim reduces to: a convex region all of whose distances are less than 1 has area at most $\pi/4$. And $\pi/4$ is the area of a disc of diameter 1.

So the student writes: "every set of diameter at most 1 fits inside a disc of diameter 1, so its area is at most $\pi/4$." This is the wall, and it is a false sentence. The equilateral triangle of side 1 has diameter 1, but the smallest disc containing it is its circumcircle, of radius $1/\sqrt{3} \approx 0.577 > \frac{1}{2}$. The correct covering statement (Jung's theorem) is that a planar set of diameter D fits in a disc of radius $D/\sqrt{3}$; that gives area at most $\pi/3$, and $\pi/3 > \pi/4$. The covering approach is not merely unproved; it is too weak by a constant.

A second try: compare with the disc by moving area around, "round the region off" by symmetrising. Steiner symmetrisation works (it preserves area and does not increase diameter), but proving that second property from scratch is longer than the whole problem, and most students who try it assert it.

The model chain. *Reduction.* Let K_0 be the region. If K_0 is unbounded, then, being convex with positive area, it contains points at every distance, in particular two at distance more than 1, and the segment argument finishes. So suppose K_0 is bounded, and suppose for contradiction that no two of its points are at distance exactly 1. If $P, Q \in K_0$ had $|PQ| > 1$, the point $R = P + \frac{Q-P}{|Q-P|}$ lies on segment $PQ \subset K_0$ with $|PR| = 1$; so every distance in K_0 is less than 1. Let K be the closure of K_0: it is compact and convex, has the same area (the boundary of a bounded convex set has zero area), and its diameter D satisfies $D \le 1$. We show $\text{area}(K) \le \pi D^2/4 \le \pi/4$, contradicting $\text{area}(K_0) > \pi/4$.

Coordinates. Because K is compact, the diameter is attained: pick $A, B \in K$ with $|AB| = D$. Place $A = (-D/2, 0)$, $B = (D/2, 0)$. Every $P = (x, y) \in K$ has $|x| \le D/2$: if $x > D/2$ then $|PA| \ge x + D/2 > D$, and symmetrically. For each $x \in [-D/2, D/2]$ the vertical line through x meets K in a closed segment (convexity, compactness), which contains $(x, 0)$ because $AB \subset K$. Write it as $\{x\} \times [-g(x), f(x)]$ with $f, g \ge 0$. The function f is concave and g is concave (upper

and lower boundaries of a convex set), so both are integrable, and

$$\text{area}(K) = \int_{-D/2}^{D/2} \big(f(x) + g(x)\big)\, dx = \int_{-D/2}^{D/2} \big(f(x) + g(-x)\big)\, dx,$$

the second form by substituting $x \mapsto -x$ in the g-part, which is legitimate because the interval is symmetric about 0.

The pairing. For each x, the points $P = (x, f(x))$ and $Q = (-x, -g(-x))$ both lie in K. Hence

$$D^2 \geq |PQ|^2 = (2x)^2 + \big(f(x) + g(-x)\big)^2, \qquad \text{so} \qquad f(x) + g(-x) \leq \sqrt{D^2 - 4x^2}.$$

Integrate. With $x = \frac{D}{2}\sin\theta$,

$$\text{area}(K) \leq \int_{-D/2}^{D/2} \sqrt{D^2 - 4x^2}\, dx = \frac{D^2}{2} \int_{-\pi/2}^{\pi/2} \cos^2\theta\, d\theta = \frac{\pi D^2}{4} \leq \frac{\pi}{4}.$$

This contradicts the hypothesis, so some pair of points is at unit distance.

Sharpness. The open disc of diameter 1 is convex, has area exactly $\pi/4$, and has no two points at distance 1. So "greater than" cannot be weakened to "at least".

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach the same reduced claim: diameter at most 1 implies area at most $\pi/4$. The natural attempt tries to prove it by *containment* (fit the set inside a disc). The model proves it by *integration*: slice the region into vertical chords and bound each chord, paired with a partner, by a distance. The split is the choice between "put the set inside the disc" and "compare the set with the disc slice by slice".

> *Containment can't work: the triangle of side* 1 *doesn't fit in a disc of diameter* 1*, yet its area is small. The inequality has to be about area directly. So I slice: area is the integral of chord lengths, and the only thing I know is that every distance is at most* D*.*

> *One vertical chord is not bounded by* D *in a useful way — a chord of length* D *at every* x *gives a square. I need to pair the top at* x *with a bottom somewhere else so that the distance constraint bites. In the disc, the point antipodal to* $(x, f(x))$ *is* $(-x, -f(x))$*. So pair the top at* x *with the bottom at* $-x$*.*

Layer 3 THE WHY

Where does the pairing of $(x, f(x))$ with $(-x, -g(-x))$ come from? The book writes it down without comment. It comes from the extremal case, and this is the "of course" the problem is built on.

The answer $\pi/4$ is the area of a disc of diameter 1. So the disc is the extremal shape, and a proof of an extremal inequality must be tight on the extremal shape: every inequality used must become an equality for the disc. Ask what inequality holds with equality everywhere on a disc of diameter D centred at the origin. The only constraint the problem gives is "all distances at most D", and in a disc that constraint is tight precisely for antipodal pairs — points P and $-P$. On the disc, the top of the chord at x is $(x, f(x))$ and its antipode is $(-x, -f(x))$, the bottom of the chord at $-x$. So the pairing is not chosen; it is read off from the extremal shape. Pair each top point with the bottom point that would be its antipode if the region were a disc, apply the only constraint you have, and you have an inequality that the disc makes an equality. Any other pairing loses tightness and so cannot reach the sharp constant.

Why vertical chords and the diameter along the x-axis? Area has to be computed somehow, and slicing is the only general way to compute the area of an unknown convex set. Putting the

diameter on the axis does two jobs: it confines the region to the strip $|x| \leq D/2$, so the integral has known limits, and it guarantees each slice contains a point of the axis, so $f, g \geq 0$ and the slice has the form $[-g, f]$.

Why the reflection $x \mapsto -x$ in the g-part? Because it is free: over a symmetric interval, $\int g(x)\,dx = \int g(-x)\,dx$. Geometrically, you have reflected the lower half of the region in the y-axis; area does not change, and now each vertical chord of the new figure has length $f(x) + g(-x)$, which is a vertical separation of two points of the original region. That is how a statement about distances in K becomes a statement about chord lengths.

Finally, the reason containment fails is worth keeping. Diameter controls area, not position. A set of diameter 1 need not lie in any disc of diameter 1, but it cannot have more area than one. Inequalities between two measures of size (diameter and area, perimeter and area) are proved by comparing the measures, not by embedding the set in the extremal shape; embedding is a stronger statement and is often false.

The book's comment that the line-segment hypothesis is extraneous is correct: nothing above used it. It was presumably included to reassure contestants that the boundary is well behaved; convexity already guarantees that.

Layer $+\alpha$ BACKTESTING

1. The cue is the number $\pi/4$ beside "unit distance": it is the area of a disc of diameter 1. Write down: "the claim is that diameter < 1 forces area $\leq \pi/4$; the disc is the extremal case".

2. Reduce by convexity to a diameter bound, then prove the isodiametric inequality by slicing: diameter on the axis, vertical chords, pair each top point with its would-be antipode.

3. Hesitation comes after writing the area as $\int(f+g)$: a single chord cannot be bounded usefully. Reinforce with "what would be antipodal in the disc?".

4. The trap is "a set of diameter 1 lies in a disc of diameter 1", which is false (equilateral triangle). A second trap is forgetting to pass from "no pair at distance exactly 1" to "all distances below 1", which is where convexity is used.

5. When an extremal inequality's constant is attained by a symmetric shape, find where that shape makes each available constraint tight, and build the proof on those pairs.

6. To prove an inequality between two size measures of a set, never embed the set in the extremal shape; integrate over slices and apply the constraint to pairs of points that are extremal in the equality case.

▷ **Cue:** the area bound is $\pi/4$ and the distance is 1 $\longrightarrow$ **Move:** the disc of diameter 1 is extremal; the problem is the isodiametric inequality

▷ **Cue:** "some pair of points at distance exactly d" in a convex set $\longrightarrow$ **Move:** by convexity it suffices to find a pair at distance $\geq d$; slide along the segment

▷ **Cue:** bound the area of a convex set with known diameter $\longrightarrow$ **Move:** put the diameter on the x-axis, slice vertically, area $= \int(f+g)$

▷ **Cue:** a sharp inequality whose equality case is a disc $\longrightarrow$ **Move:** pair points that would be antipodal in the disc: top at x with bottom at $-x$

▷ **Cue:** temptation to argue "the set fits inside the extremal shape" $\longrightarrow$ **Move:** test it on the equilateral triangle; compare measures instead of embedding

1967 A6

1967 A6. Given real numbers $\{a_i\}$ and $\{b_i\}$, $(i = 1,2,3,4)$, such that $a_1b_2 - a_2b_1 \neq 0$. Consider the set of all solutions (x_1, x_2, x_3, x_4) of the simultaneous equations

$$a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 = 0 \quad \text{and} \quad b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 = 0,$$

for which no x_i $(i = 1,2,3,4)$ is zero. Each such solution generates a 4-tuple of plus and

minus signs $(\operatorname{signum} x_1, \operatorname{signum} x_2, \operatorname{signum} x_3, \operatorname{signum} x_4)$. (a) Determine, with a proof, the maximum number of distinct 4-tuples possible. (b) Investigate necessary and sufficient conditions on the real numbers $\{a_i\}$ and $\{b_i\}$ such that the above maximum number of 4-tuples is attained.

Caveat. The MAA OCR garbles the hypothesis ("$a_1b_2 - a_2b_1 \neq 0$" comes through as "a,b:-atbn•˜O") and the equations ("$+aa$, •0"); both are restored to the standard wording above. Identity check: the book's solution under 1967 A-6 solves for x_1, x_2 in terms of x_3, x_4, draws four lines through the origin in the (x_3, x_4)-plane, and concludes "eight", with the condition $a_ib_j - a_jb_i \neq 0$ for all $i < j$. That solves exactly this statement. The answers below were also checked numerically by sampling the solution plane for several (a, b).

Fields: linear algebra (solution spaces, linear functionals), plane geometry of lines through the origin, combinatorics of sign patterns **Position:** A6 — the last problem of the morning paper, but its difficulty is conceptual, not computational: the entire problem collapses once the solution set is seen as a plane **Answer:** (a) 8. (b) The maximum is attained if and only if all six 2×2 minors $a_ib_j - a_jb_i$ $(1 \leq i < j \leq 4)$ are non-zero.

Layer 1 THE WHAT — surface analysis

The natural attempt. The strong student sees sixteen possible sign patterns and starts testing them. Pick a pattern, say $(+,+,+,+)$: is there a solution with every $x_i > 0$? That is a question about whether the positive orthant meets the solution set, which smells of linear programming. The student tries small examples. With $a = (1,1,1,1)$, $b = (1,-1,2,3)$, the pattern $(+,+,+,+)$ is impossible (the first equation says the x_i sum to zero), and so is $(-,-,-,-)$. So not all sixteen occur. Then the student tries to argue pattern by pattern which ones must fail, and the case analysis explodes: sixteen patterns, arbitrary coefficients, no visible structure. Some students get a feel that "about half" fail and guess 8; few can prove it, and part (b) is out of reach this way, because the conditions come out as a pile of sign inequalities on the coefficients rather than a clean statement.

The wall is real: the student is treating the problem as sixteen separate feasibility questions, and nothing in that framing counts them at once.

The model chain. *The solution set is a plane.* Write $a = (a_1, \ldots, a_4)$, $b = (b_1, \ldots, b_4)$ and $D_{ij} = a_ib_j - a_jb_i$. Since $D_{12} \neq 0$, the vectors a, b are linearly independent, so the solution set $V = \{x \in \mathbb{R}^4 : a \cdot x = b \cdot x = 0\}$ is a 2-dimensional subspace. Concretely, $D_{12} \neq 0$ lets us solve the two equations for x_1, x_2 by Cramer's rule in terms of $s = x_3$, $t = x_4$:

$$x_1 = \frac{D_{23}s + D_{24}t}{D_{12}}, \qquad x_2 = -\frac{D_{13}s + D_{14}t}{D_{12}}, \qquad x_3 = s, \qquad x_4 = t,$$

and every $(s,t) \in \mathbb{R}^2$ gives exactly one solution. So each coordinate x_i is a linear function $\ell_i(s,t)$ on the plane.

Each coordinate contributes a line. If some ℓ_i is identically zero, no solution has all $x_i \neq 0$ and there are no 4-tuples at all. Otherwise each ℓ_i vanishes exactly on a line L_i through the origin, is positive on one open half-plane of L_i and negative on the other. Let k be the number of *distinct* lines among $L_1, \ldots, L_4$; $k \leq 4$.

Upper bound. k distinct lines through the origin cut the plane minus the lines into exactly $2k$ open sectors (the $2k$ rays from the origin, taken in angular order, bound consecutive sectors). Inside one sector no ℓ_i vanishes, and a sector is convex, hence connected, so by the intermediate value theorem each ℓ_i has constant sign on it. Every admissible solution lies in some sector. Hence the number of distinct 4-tuples is at most $2k \leq 8$.

Eight is attained when the four lines are distinct. Suppose $k = 4$. Label the eight sectors $S_1, \ldots, S_8$ in angular order. Crossing from S_m to S_{m+1} passes over one ray, which lies on exactly one line L_i (the lines are distinct), so exactly one sign flips: the sign of ℓ_i. Going once round, the rays met are on $L_{\sigma(1)}, L_{\sigma(2)}, L_{\sigma(3)}, L_{\sigma(4)}, L_{\sigma(1)}, \ldots, L_{\sigma(4)}$ for some ordering σ, since the two rays of a line are opposite and so are four steps apart. Take two sectors $S_m \neq S_{m'}$. One of the two arcs between them crosses $r \leq 4$ rays, and those r rays lie on r *different* lines (any four consecutive rays cover the four lines once each). So exactly $r \geq 1$ coordinates change sign between S_m and $S_{m'}$: the two sign vectors differ. Eight sectors, eight distinct 4-tuples. This proves (a): the maximum is 8.

When is $k = 4$? We need: no $\ell_i \equiv 0$, and no two of ℓ_i, ℓ_j define the same line, i.e. no two are proportional. Both failures say the same thing: ℓ_i and ℓ_j are linearly dependent as functions on V for some $i \neq j$ (a zero function is dependent with anything). Now $\lambda\ell_i + \mu\ell_j \equiv 0$ on V with $(\lambda, \mu) \neq (0, 0)$ means the vector $\lambda e_i + \mu e_j$ is orthogonal to V, i.e. lies in $V^\perp = \operatorname{span}(a, b)$. So it equals $\alpha a + \beta b$ with $(\alpha, \beta) \neq (0, 0)$, and its other two coordinates vanish: with $\{k, l\}$ the complement of $\{i, j\}$,

$$\alpha a_k + \beta b_k = 0, \qquad \alpha a_l + \beta b_l = 0.$$

This 2×2 system has a non-trivial solution iff $D_{kl} = a_k b_l - a_l b_k = 0$. Conversely, if $D_{kl} = 0$, take a non-trivial (α, β); then $\alpha a + \beta b \neq 0$ (independence) and is supported on $\{i, j\}$, giving the dependence. Therefore:

$$\ell_i, \ell_j \text{ dependent on } V \iff D_{kl} = 0, \qquad \{i, j, k, l\} = \{1, 2, 3, 4\}.$$

Running over all six pairs, $k = 4$ iff all six minors D_{ij} are non-zero. This proves (b). (It also recovers the book's condition: from the formulas above, $A_1 = D_{23}/D_{12}$, $B_1 = D_{24}/D_{12}$, $A_2 = -D_{13}/D_{12}$, $B_2 = -D_{14}/D_{12}$, and $A_1 B_2 - A_2 B_1 = D_{34}/D_{12}$ by the identity $D_{12}D_{34} - D_{13}D_{24} + D_{14}D_{23} = 0$.)

The full picture. The possible counts are 0 (some $\ell_i \equiv 0$) and $2k$ for $k = 2, 3, 4$, i.e. $4, 6, 8$. For example $a = (1, 0, 1, 0)$, $b = (0, 1, 0, 1)$ forces $x_3 = -x_1$, $x_4 = -x_2$: two lines, four tuples.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same fact: two independent equations in four unknowns. The natural attempt then asks, sixteen times, "is this sign pattern feasible?". The model asks once, "what does the solution set look like?", answers "a plane", and parametrises it by (x_3, x_4). The split is the sentence "each such solution" — read as "each point of a plane".

> *Two independent equations, four unknowns: the solutions are a plane through the origin. Everything I want is a sign pattern on that plane, so I should draw the plane and ask where each coordinate changes sign.*

> *A coordinate on a plane is a linear function; it changes sign only across one line. Four coordinates, at most four lines, at most eight sectors. So the answer is eight and the only question is when the four lines are genuinely different.*

> *Two coordinates give the same line when some combination of them is killed by the plane — that combination has to be a combination of the two equations. So I ask: when can a combination of a and b vanish in the other two slots? That's a 2×2 determinant.*

Layer 3 THE WHY

Why is the plane the right object? Because of what a sign pattern is. signum x_i is constant exactly on the connected pieces where x_i does not vanish; the definition of the sign function

forces you to ask *where the zero set is*. A linear function on a plane has a line as its zero set — that is the definition of a line through the origin. So the moment you know the solutions form a plane, the problem is literally "how many regions do four lines through the origin make, and are the labels distinct?". That is not a trick; it is the problem restated with its geometry visible.

And why is the plane visible from the statement? The hypothesis $a_1b_2 - a_2b_1 \neq 0$ is placed there for precisely one purpose: it is the determinant that lets you solve for x_1, x_2 given x_3, x_4. The setters hand you the parametrisation. A hypothesis in a problem is a cue: ask what it is *for*. A non-zero 2×2 determinant is for inverting a 2×2 system. Once x_3, x_4 are free, they are coordinates on the solution set, and you are drawing in the (x_3, x_4)-plane whether you meant to or not.

Why do the other minors appear in (b)? Here is the step the book states without motivation ("this is equivalent to $A_1 \neq 0, \ldots, A_1B_2 - A_2B_1 \neq 0$ or, simply, $a_ib_j - a_jb_i \neq 0$"). The "simply" hides the Plücker identity. The structural reason is duality: a linear function on V vanishes identically iff it is a combination of the equations, because $V^\perp = \mathrm{span}(a, b)$ — that is what it means for V to be cut out by a and b. Asking whether x_i and x_j are tied together on V is therefore asking whether a combination of a, b lives only in slots i, j, i.e. is zero in the complementary slots k, l. That is a statement about the $\{k, l\}$ columns of the 2×4 matrix with rows a and b, and it is exactly $D_{kl} = 0$. The complement swap is the whole content: "the pair $\{i, j\}$ is degenerate" is "the minor on the other pair vanishes". The given hypothesis $D_{12} \neq 0$ is itself the instance $\{i, j\} = \{3, 4\}$: it guarantees x_3, x_4 are independent on V, which is why they could serve as coordinates in the first place. The problem's own hypothesis is one sixth of the answer to (b).

Why are the eight labels distinct? Because a line through the origin crosses the circle of directions twice, in opposite directions, so going half-way round crosses each line once. Any two different sectors are joined by an arc of at most half a turn, which crosses each line at most once, so each crossing flips a different sign and nothing flips back.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "the set of all solutions" of two linear equations in four unknowns, together with the hypothesis $a_1b_2 - a_2b_1 \neq 0$. On seeing it, write: "solution set is a 2-dimensional subspace; parametrise by $x_3 = s$, $x_4 = t$; x_1, x_2 linear in s, t".

2. Geometry of linear functionals: each coordinate is a linear function on the plane, sign changes happen only across its zero line, and counting sign patterns becomes counting regions cut by lines through the origin.

3. Hesitation comes at the transition from "sixteen patterns" to "regions". It feels like counting regions ignores which patterns are feasible. It doesn't: every admissible solution sits in a region, and each region has one pattern. Reinforce with the line-picture: draw four lines through a point and label each sector by which side of each line it is on.

4. The trap is part (b). It is tempting to demand only that $A_1, B_1, A_2, B_2 \neq 0$ (no coordinate line is an axis), forgetting that L_1 and L_2 can coincide with each other, or to give the condition in terms of the derived quantities A_i, B_i rather than the given a_i, b_i. A second trap: forgetting the case where some x_i vanishes on the whole plane, which gives zero tuples.

5. Recognise sign-pattern and "which orthants does it meet" questions as questions about zero sets. Whenever a problem asks for the signs of several linear quantities over a family, the family's dimension decides the picture: in a plane, zero sets are lines through the origin; in space, planes through the origin.

6. When a problem constrains a family of objects by linear equations, first compute the dimension of the family and pick free coordinates; then read every quantity in the problem as a function on that small space. And when two linear quantities are asked to be "independent on the solution set", test it by asking whether some combination of the defining equations is supported only on those two slots: that is a minor on the complementary slots.

▷ **Cue:** two independent linear equations in four unknowns; the question is about the whole solution set $\longrightarrow$ **Move:** write the solution set as a plane, parametrise by the two free

variables
▷ **Cue:** count the sign patterns of several linear quantities $\longrightarrow$ **Move:** draw each quantity's zero line; patterns correspond to regions
▷ **Cue:** k distinct lines through the origin in the plane $\longrightarrow$ **Move:** $2k$ sectors; opposite rays four steps apart when $k = 4$, so all labels distinct
▷ **Cue:** when do two linear functions coincide (up to scale) on a solution space V $\longrightarrow$ **Move:** a combination of them lies in $V^{\perp}$, the span of the equations
▷ **Cue:** a combination of the rows of a $2 \times n$ matrix must vanish in two chosen columns $\longrightarrow$ **Move:** the 2×2 minor on those columns is zero
▷ **Cue:** a hypothesis given as a non-zero determinant $\longrightarrow$ **Move:** ask what it inverts: it names the free variables

Putnam 1967 — Paper B

1967 B1

1967 B1. Let $(ABCDEF)$ be a hexagon inscribed in a circle of radius r. Show that if $\overline{AB} = \overline{CD} = \overline{EF} = r$, then the midpoints of $\overline{BC}$, $\overline{DE}$, $\overline{FA}$ are the vertices of an equilateral triangle.

Caveat. The MAA OCR gives the rotation constant as "$w = \cos(\pi/3) + i\sin(\pi/3)$... Since $w^3 = 1$ and $w \neq 1$, $w^2 - w + 1 = 0$." This is internally inconsistent: $w = e^{i\pi/3}$ satisfies $w^3 = e^{i\pi} = -1$, not 1 (the correct order-6 root has $w^6 = 1$). The intended identity is recovered correctly, however, from $w^3 = -1$: factor $w^3 + 1 = (w+1)(w^2 - w + 1) = 0$, and since $w \neq -1$, $w^2 - w + 1 = 0$ follows exactly as the book states. The sign has been silently corrected below; no other content is affected, and the corrected identity is what the rest of the book's own solution in fact uses.

Fields: Euclidean geometry, complex numbers **Position:** B1 (opener) **Answer:** The claim holds; the midpoints form an equilateral triangle because $AB = CD = EF = r$ each encode a fixed 60° rotation about the centre, and rotating the whole configuration by that same angle permutes the three midpoints cyclically.

Layer 1 THE WHAT — surface analysis

The natural attempt. Put the centre O at the origin and assign each vertex a polar angle: A at angle α, and so on around the hexagon. Since $OA = OB = r$ and $AB = r$, triangle OAB is equilateral, so the angle $\angle AOB = 60^\circ$; likewise for $\angle COD$ and $\angle EOF$. The three remaining arcs BC, DE, FA are free (subject only to summing to $360^\circ - 3 \cdot 60^\circ = 180^\circ$). To show the midpoints P, Q, R of BC, DE, FA form an equilateral triangle, you write down their Cartesian coordinates using half-angle averages – the midpoint of a chord is not simply the angular midpoint on the circle, so you must average the two *position vectors*, e.g. $P = \frac{1}{2}(\cos\beta + \cos\gamma,\ \sin\beta + \sin\gamma)$ where β, γ are the angles of B, C. You then need $|PQ| = |QR| = |RP|$ and one angle of 60°, which means expanding six such coordinates, applying sum-to-product formulas ($\cos\beta + \cos\gamma = 2\cos\frac{\beta-\gamma}{2}\cos\frac{\beta+\gamma}{2}$) to each pair, and comparing three unwieldy trigonometric expressions whose only relation is the single constraint that three of the six arc-gaps are fixed at 60° and the other three are free but sum to 180°. The computation is not wrong – it will eventually close – but the free arcs refuse to cancel cleanly, and by the second side-length comparison you are carrying four or five independent angle variables through double-angle expansions with no structural reason to expect the mess to simplify. This is exactly the point at which a correct method becomes too expensive to finish in exam time.

The model chain. Place the circle's centre at the origin of the complex plane and represent A, B, C, D, E, F as complex numbers of modulus r. Since $OA = OB = AB = r$, triangle OAB is equilateral, so B is obtained from A by a fixed rotation of 60° about O: $B = A\omega$, where $\omega = \cos(\pi/3) + i\sin(\pi/3)$. By the same argument, $D = C\omega$ and $F = E\omega$ (all three rotations

in the same sense, since the hexagon's vertices are listed in one consistent cyclic order). Since $\omega^3 = -1$ and $\omega \neq -1$, factoring $\omega^3 + 1 = (\omega+1)(\omega^2-\omega+1)$ gives $\omega^2 - \omega + 1 = 0$, i.e. $\omega^2 = \omega - 1$.

Let P, Q, R be the midpoints of BC, DE, FA:

$$P = \tfrac{1}{2}(B+C) = \tfrac{1}{2}(A\omega + C), \qquad Q = \tfrac{1}{2}(D+E) = \tfrac{1}{2}(C\omega + E), \qquad R = \tfrac{1}{2}(F+A) = \tfrac{1}{2}(E\omega + A).$$

Rotating R about Q by 60° sends it to $Q + \omega(R-Q)$. Compute:

$$Q + \omega(R-Q) = \tfrac{1}{2}(C\omega + E) + \omega\left[\tfrac{1}{2}(E\omega + A) - \tfrac{1}{2}(C\omega + E)\right] = \tfrac{1}{2}\left[C\omega + E + E\omega^2 - E\omega + A\omega - C\omega^2\right].$$

Group the C and E terms: this equals $\frac{1}{2}\left[A\omega + C(\omega - \omega^2) + E(1 - \omega + \omega^2)\right]$. Since $\omega^2 = \omega - 1$, we get $\omega - \omega^2 = \omega - (\omega - 1) = 1$ and $1 - \omega + \omega^2 = 1 - \omega + (\omega - 1) = 0$. So the bracket collapses to $A\omega + C \cdot 1 + E \cdot 0 = A\omega + C$, and

$$Q + \omega(R-Q) = \tfrac{1}{2}(A\omega + C) = P.$$

So the segment QR, rotated 60° about Q, lands exactly on QP; hence $|QP| = |QR|$ and $\angle PQR = 60^\circ$, which forces triangle PQR to be equilateral.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both readings start from the same three equilateral triangles OAB, OCD, OEF forced by $OA = OB = AB = r$ (and cyclically). They split at what to do with that fact. The natural attempt treats "$\angle AOB = 60^\circ$" as a number to feed into coordinate trigonometry – write down six angles, average pairs of position vectors, expand with sum-to-product identities. The model treats "$\angle AOB = 60^\circ$" as an *operation*: B is not just at some angle 60° from A, it is the image of A under a fixed rotation ω, and that operation can be composed and factored algebraically instead of measured.

> *$AB = r = OA = OB$ is not telling me a distance to plug into the cosine rule – it is telling me that B is what you get by rotating A by exactly 60° about the centre. I should carry that rotation as an algebraic operator ω, not throw it away into a numerical angle the moment I write it down.*

> *To show three points are the vertices of an equilateral triangle, I do not need to compute all three side lengths and an angle separately – it is enough to exhibit a single 60° rotation about one vertex that carries a second vertex to the third. That is a strictly weaker, and strictly cheaper, thing to prove.*

- ▷ **Cue:** a chord subtending the same length as the circle's radius appears (a chord of length r in a circle of radius r) $\longrightarrow$ **Move:** recognise the two radii and the chord as an equilateral triangle, and represent the two endpoints as complex numbers related by multiplication by $\omega = e^{i\pi/3}$ (or $e^{-i\pi/3}$), rather than recording only the numerical angle 60°
- ▷ **Cue:** a claim that three constructed points form an equilateral triangle $\longrightarrow$ **Move:** show that rotating one of the points about a second by 60° lands exactly on the third, using complex multiplication by $e^{\pm i\pi/3}$, instead of computing and comparing three side lengths and an angle

Layer 3 THE WHY

The reason the rotation view is not a trick but the literal content of the hypothesis is this: a circle is the set of points equidistant from its centre, so any two points X, Y on a circle of radius r with $XY = r$ automatically satisfy $OX = OY = XY = r$ – triangle OXY is equilateral for the cheapest possible reason, three equal sides. An equilateral triangle, in turn, is exactly the fixed-point set relation of a 60° rotation about any one of its vertices: rotating OXY by 60°

about O carries X to Y (in one of the two orientations). So the hypothesis "$AB = CD = EF = r$" is not three isolated length facts; it is three instances of the same operator, ω, applied at three different places around the hexagon. And the target – "P, Q, R equilateral" – is itself, by the same definitional fact run in reverse, nothing but the statement that a 60° rotation about one of P, Q, R carries a second to the third. Once both the hypothesis and the conclusion are recognised as instances of the identical algebraic object (rotation by ω), the proof is forced to be algebraic: represent the points as complex numbers (the natural carrier of a plane rotation), substitute $B = A\omega$, $D = C\omega$, $F = E\omega$ into the definitions of P, Q, R, and let the single identity $\omega^2 - \omega + 1 = 0$ (itself just the minimal polynomial of a primitive sixth root of unity, forced by $\omega \neq -1$ dividing $\omega^3 + 1 = 0$) absorb all of the algebra. Nothing about the three "free" arcs BC, DE, FA ever needs to be pinned down, because the rotation argument never used their lengths – only the fixed rotations at A, C, E mattered. The natural coordinate attempt fails to see this because it insists on measuring every angle before combining them, when the combination itself is the content the problem is testing.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $\overline{AB} = r$ together with the vertices lying on a circle of radius r. On seeing it, write down immediately: OAB is equilateral, so $\angle AOB = 60^\circ$, and record this as "B is the rotation of A by $\omega = e^{i\pi/3}$," not merely as a numerical angle.

2. That cue suggests representing every vertex as a complex number on $|z| = r$ and encoding the three special sides as $B = A\omega$, $D = C\omega$, $F = E\omega$, then expressing the midpoints P, Q, R purely in terms of A, C, E and ω.

3. Hesitation occurs at the moment you must show P, Q, R are equilateral once you have their formulas: it is tempting to fall back to computing $|PQ|, |QR|, |RP|$ directly, which reintroduces the same trigonometric mess as the coordinate approach. It is reinforced away by remembering that "equilateral" can instead be verified by exhibiting one rotation, $Q + \omega(R - Q) = P$, which is a single algebraic identity to check rather than three lengths and an angle.

4. The trap is measuring $\angle AOB$ as a bare number (60°) and discarding the rotational structure, which then forces a return to full coordinate trigonometry with the three free arcs BC, DE, FA left dangling and unusable.

5. To recognise this faster: whenever a chord equals the radius, say "equilateral triangle, rotation by 60°" in the same breath – never just "60° angle." And whenever a target is "these three points are equilateral," immediately ask whether a single rotation carrying one target point to another is easier to exhibit than three lengths.

6. Decision rule: chord length equal to the radius is a rotation-by-60° fact, not merely a length fact – encode it as multiplication by $e^{\pm i\pi/3}$ on complex-number coordinates. And an equilateral-triangle conclusion should be sought as a single rotation identity between the three points, not as three separately verified side lengths.

1967 B2

1967 B2. Let $0 \leq p \leq 1$ and $0 \leq r \leq 1$ and consider the identities

(a) $(px + (1-p)y)^2 = Ax^2 + Bxy + Cy^2$,

(b) $(px + (1-p)y)(rx + (1-r)y) = \alpha x^2 + \beta xy + \gamma y^2$.

Show that (with respect to p and r) (a) $\max\{A, B, C\} \geq 4/9$, (b) $\max\{\alpha, \beta, \gamma\} \geq 4/9$.

Caveat. The book's OCR renders the inequality symbol as "~" in both parts: "max{A,B,C} ~ 4/9." This is a dropped/garbled $\geq$, confirmed by the solution text itself, which proves a *lower* bound (it exhibits $p = 1/3$ making two of the three quantities equal $4/9$ and argues the third cannot then also fall below $4/9$) – an upper bound of $4/9$ would in any case be false, since $B = 2p(1-p) = 1/2$ at $p = 1/2$. Read as $\geq$ throughout.

Fields: Algebra, inequalities (two-variable optimisation via a hyperbola argument) **Position:** B2 **Answer:** Both maxima are always at least $4/9$, with equality exactly at $p = r = 1/3$ (and its reflection $p = r = 2/3$ for part (a) alone); no choice of $p, r \in [0,1]$ can push all three coefficients below $4/9$ simultaneously.

Layer 1 THE WHAT — surface analysis

The natural attempt. Part (a): with $A = p^2$, $B = 2p(1-p)$, $C = (1-p)^2$, the instinct is calculus – minimise $\max\{A, B, C\}$ over $p \in [0,1]$ by finding where the pieces cross and checking derivatives on each region. This works, but it is three separate one-variable optimisations glued together at case boundaries, and it does not obviously generalise.

Part (b) is worse. With $\alpha = pr$, $\beta = p(1-r) + r(1-p)$, $\gamma = (1-p)(1-r)$, you now have *two* free variables. The calculus instinct is to fix r, differentiate $\max\{\alpha, \beta, \gamma\}$ in p, find critical points, then differentiate the resulting expression in r – a nested optimisation over a function that is not even differentiable everywhere (it is a max of three smooth pieces, with corners on the curves where two of them are equal). Tracking which of α, β, γ is largest as (p, r) ranges over the unit square requires partitioning the square into regions by the curves $\alpha = \beta$, $\beta = \gamma$, $\alpha = \gamma$ (each itself a quadratic curve, not a line), and doing calculus separately in each region. It is a correct plan, but by the time you have found and classified all the boundary curves you are doing more bookkeeping than mathematics, with no guarantee the case analysis closes before time runs out.

The model chain. Part (a): Suppose, for contradiction, that A, B, C are all $< 4/9$. From $A = p^2 < 4/9$, $p < 2/3$; from $C = (1-p)^2 < 4/9$, $p > 1/3$. So $p \in (1/3, 2/3)$. But $B = 2p(1-p) = 2p - 2p^2$ is a downward-opening parabola in p with $B(1/3) = B(2/3) = 4/9$ and B strictly increasing on $[1/3, 1/2]$, so $B(p) > 4/9$ for every p strictly between $1/3$ and $2/3$ – contradicting $B < 4/9$. Hence no p makes all three simultaneously below $4/9$, so $\max\{A, B, C\} \geq 4/9$ always, with equality exactly at $p = 1/3$ or $p = 2/3$.

Part (b): substitute $p' = p - \frac{1}{2}$, $r' = r - \frac{1}{2}$ (centring the square on its symmetry point), so $p', r' \in [-\frac{1}{2}, \frac{1}{2}]$. Direct expansion gives $\beta = p + r - 2pr = \frac{1}{2} - 2p'r'$. Suppose $\alpha = pr < 4/9$ and $\gamma = (1-p)(1-r) < 4/9$; we show $p'r' \leq 1/36$, which gives $\beta \geq 4/9$.

If $p'r' \leq 0$ this is immediate. Otherwise p', r' have the same sign; take both positive (the case both negative is identical by the symmetry $p \leftrightarrow 1-p$, $r \leftrightarrow 1-r$, which swaps $\alpha \leftrightarrow \gamma$ and fixes β). Expand $\alpha = (p' + \frac{1}{2})(r' + \frac{1}{2}) = p'r' + \frac{1}{2}(p' + r') + \frac{1}{4} < 4/9$, so

$$p'r' + \tfrac{1}{2}(p' + r') < \tfrac{7}{36}.$$

Since $p', r' > 0$, AM–GM gives $\frac{1}{2}(p' + r') \geq \sqrt{p'r'}$, so with $t = \sqrt{p'r'} \geq 0$,

$$t^2 + t \leq p'r' + \tfrac{1}{2}(p' + r') < \tfrac{7}{36}.$$

The positive root of $t^2 + t - \frac{7}{36} = 0$ is $t = \frac{-1+\sqrt{1+7/9}}{2} = \frac{-1+4/3}{2} = \frac{1}{6}$, so $t < 1/6$, giving $p'r' = t^2 < 1/36$. Hence $\beta = \frac{1}{2} - 2p'r' > \frac{1}{2} - \frac{1}{18} = \frac{4}{9}$. So whenever $\alpha, \gamma < 4/9$, in fact $\beta > 4/9$; equivalently, it is impossible for all three of α, β, γ to fall below $4/9$, so $\max\{\alpha, \beta, \gamma\} \geq 4/9$ for every $p, r \in [0,1]$, with equality at $p = r = 1/3$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both readings begin by writing down A, B, C (or α, β, γ) explicitly in p (and r). They split at the choice of proof strategy for a "show the max is at least $4/9$" claim:

the natural attempt treats it as a minimisation problem to be solved directly (find the p, r that minimises the max, via calculus and case boundaries), while the model treats it as a claim to be proved by *contradiction against a threshold*: assume all three quantities are simultaneously below $4/9$, and derive that this forces one of them above $4/9$ after all.

> *I do not need to find the minimising p – I only need to rule out the possibility that all three quantities are simultaneously small. So instead of optimising, I should assume the bad case (everything $< 4/9$) and squeeze it until it contradicts itself.*

> *The three quantities A, B, C (and α, β, γ) are not independent – they are the coefficients of a perfect square, or of a product of two linear forms, so they satisfy an algebraic relation ($A + B + C = 1$ in part (a); a similar identity underlies part (b)). Once two of them are pinned to a narrow range by the $< 4/9$ assumption, the third is not free – its formula is forced, and I should substitute the resulting narrow range for p (or p, r) directly into it rather than re-deriving bounds on it from scratch.*

- ▷ **Cue:** a claim of the form "the maximum of several expressions summing to (or algebraically tied to) a constant is at least X" $\longrightarrow$ **Move:** assume the contrary – that every expression is below X – and substitute the resulting restricted range of the variable(s) into the remaining expression(s), aiming for a direct contradiction, rather than solving a minimisation problem for the true minimising point
- ▷ **Cue:** two expressions built from the same variable(s) $p, r \in [0, 1]$ symmetric about the midpoint $\frac{1}{2}$ $\longrightarrow$ **Move:** recentre with $p' = p - \frac{1}{2}$, $r' = r - \frac{1}{2}$ before expanding, to expose the symmetry algebraically instead of carrying $\frac{1}{2}$'s through every expansion

Layer 3 THE WHY

Part (a) is really a statement about three quantities that are forced to sum to 1: $A + B + C = p^2 + 2p(1-p) + (1-p)^2 = (p + (1-p))^2 = 1$, since A, B, C are exactly the coefficients of the perfect square $(px + (1-p)y)^2$ and the identity is what you get by setting $x = y = 1$. Three nonnegative numbers summing to 1 cannot all be less than $1/3$ – that is pigeonhole, not calculus. The problem does not ask for the sharper threshold $1/3$; it asks for $4/9$, which is exactly the extra information carried by the specific quadratic shape of $B = 2p(1-p)$ near the boundary of the region where $A, C < 4/9$ forces $p \in (1/3, 2/3)$. So the correct instinct is: use the constraint ($A + B + C = 1$, or the analogous relation in part (b)) to convert "show a max is large" into "show three quantities cannot be simultaneously small," which is a contradiction argument, not an optimisation. This is precisely why the official solution reaches for a *region* in the (p, r)-plane rather than a derivative: the claim "$\alpha, \gamma < 4/9$" is a two-dimensional region bounded by two hyperbolas ($pr = 4/9$ and $(1-p)(1-r) = 4/9$), and the archive's contribution is to replace the book's appeal to "examination of the graphs" (a picture, not a proof) with an explicit algebraic squeeze: centre the square at its symmetry point, and bound the resulting product $p'r'$ using AM–GM on the two hypotheses at once. The half-angle-style substitution $p' = p - \frac{1}{2}$ is forced by the fact that the constraint set (both hyperbolas) is symmetric under $p \mapsto 1 - p$, $r \mapsto 1 - r$ simultaneously – recentring on the fixed point of that symmetry is what turns two coupled quadratic inequalities into one clean bound on a single product $p'r'$.

> **Caveat.** The book's own proof of part (b) is explicitly graphical ("by looking at asymptotes, we see graphically that the region...does not overlap..."); the AM–GM argument given above in the model chain is this analyst's own construction, verified independently by direct algebra (checked at the equality point $p = r = 1/3$, where $\alpha = 1/9$, $\beta = \gamma = 4/9$, confirming the bound is tight), and is offered as a genuine proof where the source supplies only a picture.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that A, B, C are literally the coefficients of a perfect square $(px + (1-p)y)^2$, hence sum to 1 (set $x = y = 1$). On seeing it, write down: three nonnegative numbers summing to a fixed constant cannot all be small – this is the shape of a pigeonhole bound before any calculus is attempted.

2. That cue suggests proof by contradiction against the threshold: assume all three (or in part (b), all three of α, β, γ) are below $4/9$, and show the resulting restriction on p (or p, r) is inconsistent with the third quantity's formula.

3. Hesitation occurs in part (b) once two variables are in play: it is tempting to fall back to case-by-case calculus across the unit square. It is reinforced away by recentring at $p' = p - \frac{1}{2}, r' = r - \frac{1}{2}$ first, which exposes $\beta = \frac{1}{2} - 2p'r'$ as a single clean function of the product $p'r'$, converting a two-variable case analysis into a one-variable bound.

4. The trap is attempting a nested nested-derivative optimisation over the max of three non-differentiable pieces in two variables, which is a legitimate but enormously expensive route with corner cases that must each be checked.

5. To recognise this faster: whenever several expressions built from the same bounded variables are known (or can be shown) to sum to, or be tied by, a fixed algebraic relation, and the claim is a lower bound on their maximum, reach first for "assume all are below the bound and derive a contradiction from the relation," before reaching for calculus.

6. Decision rule: for max-of-several-related-quantities $\geq$ threshold claims, exploit the algebraic relation tying the quantities together (a sum, a product identity, a shared substitution) to convert the problem into a contradiction argument; reserve calculus for when no such relation is visible. When two variables are symmetric about a common midpoint, recentre both before expanding.

1967 B3

1967 B3. If f and g are continuous and periodic functions with period 1 on the real line, then

$$\lim_{n\to\infty} \int_0^1 f(x)\, g(nx)\, dx = \left(\int_0^1 f(x)\, dx \right) \left(\int_0^1 g(x)\, dx \right).$$

Fields: Real analysis (uniform continuity, Riemann sums, periodicity) **Position:** B3 **Answer:** The limit equals the product of the two averages, $\left(\int_0^1 f\right)\left(\int_0^1 g\right)$, because as $n \to \infty$ the rapidly oscillating factor $g(nx)$ contributes only its average value $\int_0^1 g$ over every subinterval on which f is essentially constant.

Layer 1 THE WHAT — surface analysis

The natural attempt. $g(nx)$ oscillates n times as fast as $g(x)$ as $n \to \infty$, and the intuitive picture – rapid oscillation averaging out against a slowly-varying factor – is exactly right. The natural way to try to make this rigorous is integration by parts, since g has an antiderivative to exploit its periodicity. Let $G(x) = \int_0^x g(t)\, dt$ and $c = \int_0^1 g(t)\, dt = G(1) - G(0)$; periodicity of g gives $G(x+1) = G(x) + c$ for all x, so G is "linear plus periodic." If f were differentiable, integration by parts would give

$$\int_0^1 f(x)g(nx)\, dx = \Big[f(x)\frac{G(nx)}{n} \Big]_0^1 - \frac{1}{n} \int_0^1 f'(x)G(nx)\, dx,$$

and the boundary term is $O(1/n)$ while the remaining integral, bounded by $\frac{1}{n} \max|f'| \max|G| \cdot 1 \to 0$, also vanishes – except that this only proves the limit is 0, not $c \int_0^1 f$! The linear-growth part of G (the "$+c\lfloor nx \rfloor$" drift) has been absorbed into the $O(1/n)$ estimate incorrectly, because $G(nx)$ is not bounded when $c \neq 0$ – it grows like cnx. So this route needs to be repaired by first

subtracting off the drift, $\tilde{G}(x) = G(x) - cx$ (now genuinely periodic and bounded), redoing the algebra with $g = \tilde{G}' + c$, and separately tracking the $c \int_0^1 f(x)\,dx \cdot 1$ term that this subtraction produces – doable, but only after noticing the original bound was silently false, and it still assumes f is differentiable, which the problem never grants.

The model chain. f is continuous and periodic with period 1, hence bounded and uniformly continuous on all of $\mathbb{R}$. Fix $\varepsilon > 0$ and let $M = \sup|g|$. By uniform continuity, choose n_0 so that for $n \geq n_0$, $|x - y| \leq 1/n$ implies $|f(x) - f(y)| < \varepsilon$. Fix such an n and partition $[0,1]$ into $[m/n, (m+1)/n]$, $m = 0, \ldots, n-1$:

$$\int_0^1 f(x)g(nx)\,dx = \sum_{m=0}^{n-1} \int_{m/n}^{(m+1)/n} f(m/n)\,g(nx)\,dx + \sum_{m=0}^{n-1} \int_{m/n}^{(m+1)/n} \big[f(x) - f(m/n)\big]g(nx)\,dx.$$

For the first sum, substitute $t = nx$: $\int_{m/n}^{(m+1)/n} g(nx)\,dx = \frac{1}{n}\int_m^{m+1} g(t)\,dt = \frac{1}{n}\int_0^1 g(t)\,dt$, using that g has period 1 so its integral over any unit interval is the same. So the first sum equals

$$\Big(\int_0^1 g(t)\,dt\Big) \cdot \frac{1}{n}\sum_{m=0}^{n-1} f(m/n),$$

and $\frac{1}{n}\sum_{m=0}^{n-1} f(m/n)$ is a Riemann sum for $\int_0^1 f(x)\,dx$ with mesh $1/n \to 0$, so it converges to $\int_0^1 f(x)\,dx$ as $n \to \infty$ (this uses only continuity of f, not differentiability). For the second sum, each integrand satisfies $|f(x) - f(m/n)| < \varepsilon$ (since $x, m/n$ are within $1/n$ of each other and $n \geq n_0$) and $|g(nx)| \leq M$, so each term is bounded in absolute value by $\varepsilon M/n$, and the whole sum by εM. Hence

$$\Big|\int_0^1 f(x)g(nx)\,dx - \Big(\int_0^1 g\Big)\cdot\frac{1}{n}\sum_{m=0}^{n-1} f(m/n)\Big| \leq \varepsilon M$$

for all $n \geq n_0$. Letting $n \to \infty$ (with ε fixed), the Riemann sum converges to $\int_0^1 f$, so

$$\limsup_{n\to\infty}\Big|\int_0^1 f(x)g(nx)\,dx - \Big(\int_0^1 f\Big)\Big(\int_0^1 g\Big)\Big| \leq \varepsilon M.$$

Since $\varepsilon > 0$ was arbitrary and M is fixed, the left side is 0, proving the limit.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both readings agree that the mechanism is "rapid oscillation averages to the mean." They split on how to make that rigorous. Integration by parts tries to control the oscillation by exploiting an antiderivative of g, which needs f differentiable to integrate by parts against and, worse, silently mishandles the unbounded drift term in G when $\int_0^1 g \neq 0$. The uniform-continuity route instead partitions the domain of f into pieces so fine that f is nearly constant on each, and evaluates $g(nx)$'s local average *exactly* (by periodicity) on each piece, never differentiating anything.

> *I do not need an antiderivative of g to control its oscillation – g is periodic, so I can compute the integral of $g(nx)$ over any interval of length exactly $1/n$* exactly*, by direct substitution, with no approximation at all. The only approximation I actually need is on f, and continuity alone (not differentiability) already gives me that, via uniform continuity on a compact interval.*

> *The integration-by-parts route silently assumed $G(nx) = \int_0^{nx} g$ stays bounded as $n \to \infty$, but it only does if $\int_0^1 g = 0$. Before trusting a bound like* $\max|G|$*, I should check whether the function I am bounding can actually grow – here it grows linearly whenever the periodic function does not average to zero, which is exactly the generic case the problem is asking about.*

▷ **Cue:** a rapidly oscillating factor $g(nx)$, with g periodic, multiplying a fixed continuous factor $f(x)$, inside an integral as $n \to \infty$ $\longrightarrow$ **Move:** partition the domain into subintervals of length exactly $1/n$ (one period of $g(nx)$), replace f by its value at one point of each subinterval (justified by uniform continuity, with an explicit ε-error bound), and evaluate $\int g(nx)\,dx$ over each subinterval exactly using periodicity and a substitution $t = nx$, rather than integrating by parts against an antiderivative of g

Layer 3 THE WHY

The move to partition-and-approximate rather than integrate-by-parts is forced by matching each factor to the tool that needs the least from it. g's only relevant property is periodicity with period 1; periodicity is a statement about *integrals over intervals of length* 1 being all equal – $\int_a^{a+1} g = \int_0^1 g$ for every a – which is precisely the fact used to evaluate $\int_{m/n}^{(m+1)/n} g(nx)\,dx$ exactly after the substitution $t = nx$ turns the interval $[m/n, (m+1)/n]$ into $[m, m+1]$, a full period. No derivative of g is needed because none of g's hypotheses supply one. Meanwhile f's only hypothesis is continuity, and the theorem that converts "continuous on a compact interval" into a usable quantitative statement is uniform continuity, which is exactly the license to replace $f(x)$ by $f(m/n)$ on a subinterval with an error that can be made uniformly small by shrinking the subinterval – again requiring no derivative. The official solution's structure (split the integral, evaluate the g-part exactly by substitution, bound the f-error by uniform continuity) is the correct decomposition because it is the unique way to use *exactly* the hypotheses given, no more. The integration-by-parts attempt fails for the same reason overreaching methods usually fail on a Putnam problem: it silently imports an extra hypothesis (differentiability of f, and implicitly boundedness of the antiderivative G) that is not given and, in the case of G, not even true when $\int_0^1 g \neq 0$ – so the method's first success (bounding the boundary term) conceals its later failure (assuming G stays bounded). Recognising periodicity as "integral over any full period is the same constant" rather than "a function I could antidifferentiate" is the structural fact that makes the correct route visible.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $g(nx)$ inside an integral with g periodic and $n \to \infty$: an oscillation frequency going to infinity. On seeing it, write down: periodicity means $\int$ over any interval of length equal to one period of $g(n\cdot)$, namely $1/n$, is exactly $\int_0^1 g$ – an exact fact, not an approximation.

2. That cue suggests partitioning $[0,1]$ into n subintervals of length $1/n$ (each a full period of $g(nx)$) and, on each, replacing the slowly varying factor f by a single sample value, controlled by uniform continuity.

3. Hesitation occurs when trying to make the oscillation-averages-out intuition rigorous: it is tempting to integrate by parts against an antiderivative of g, which looks standard for oscillatory integrals. It is reinforced away by checking, before committing, whether the antiderivative stays bounded – here it does not, whenever $\int_0^1 g \neq 0$, exposing that the method silently needs an extra hypothesis (a differentiable f, and a mean-zero g) that the problem does not give.

4. The trap is trusting an oscillatory-integral technique (integration by parts, Riemann–Lebesgue-style reasoning) imported from problems where the oscillating function has mean zero, without checking that this problem's g is allowed to have a nonzero mean – which is in fact the generic, interesting case here.

5. To recognise this faster: when an oscillation frequency $n \to \infty$ multiplies a periodic function inside an integral, check first whether the oscillating function's integral over one full period can be evaluated exactly by a linear substitution – if so, that is cheaper and more robust than any antiderivative-based technique, and it uses only continuity of the other factor, not differentiability.

6. Decision rule: to average out a rapid periodic oscillation $g(n\cdot)$ against a slowly varying continuous factor f, partition the domain into intervals of exactly one oscillation period, evaluate the oscillating factor's contribution on each interval exactly via a linear change of variable and periodicity, and bound the slowly-varying factor's deviation from a constant by uniform continuity – never assume extra smoothness (e.g. differentiability) that the hypotheses do not supply.

1967 B4

1967 B4. (a) A certain locker room contains n lockers numbered $1, 2, 3, \cdots, n$ and all are originally locked. An attendant performs a sequence of operations $T_1, T_2, \cdots, T_n$ whereby with the operation T_k, $1 \leq k \leq n$, the condition of being locked or unlocked is changed for all those lockers and only those lockers whose numbers are multiples of k. After all the n operations have been performed it is observed that all lockers whose numbers are perfect squares (and only those lockers) are now open or unlocked. Prove this mathematically.

(b) Investigate in a meaningful mathematical way a procedure or set of operations similar to those above which will produce the set of cubes, or the set of numbers of the form $2m^2$, or the set of numbers of the form $m^2 + 1$, or some nontrivial similar set of your own selection.

Fields: number theory (divisor counting); parity arguments **Position:** B4 — middle of the paper; part (a) is a classic, part (b) is open-ended and rewards understanding over computation **Answer:** (a) Locker m is toggled once for each divisor of m, and m has an odd number of divisors exactly when m is a square. (b) Toggle multiples of $2k$ to get $\{2m^2\}$; toggle i when $k \mid i-1$ to get $\{m^2+1\}$; perform T_k only for those k whose prime exponents are all $\equiv 0$ or $1 \pmod 3$ to get the cubes.

Caveat. The OCR renders the second target set as "$2m1$"; the book's own solution ("the set of numbers of the form $2m^2$ are obtained by having T_k change lockers whose numbers are multiples of $2k$") confirms $2m^2$, and that is used here. The book gives no procedure for cubes; the one below is the archive's own, proved in full and checked by computer for $m \leq 3000$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Locker m changes state once for every $k \leq n$ with $k \mid m$; since $m \leq n$, every divisor of m is some $k \leq n$. So locker m is changed $d(m)$ times, where $d(m)$ is the number of positive divisors, and it ends open exactly when $d(m)$ is odd. The strong student now reaches for the divisor formula: write $m = p_1^{a_1} \cdots p_r^{a_r}$, so

$$d(m) = (a_1 + 1)(a_2 + 1) \cdots (a_r + 1).$$

A product of integers is odd iff every factor is odd, so $d(m)$ is odd iff every a_i is even, iff m is a square. Part (a) is done, correctly.

Then part (b) arrives, and the student tries to repeat the move for cubes: "I want a toggle count that is odd iff every a_i is a multiple of 3." The first guesses fail. Toggle multiples of k^3: then m is toggled $\prod(\lfloor a_i/3 \rfloor + 1)$ times, odd iff every $\lfloor a_i/3 \rfloor$ is even, which catches $m = 2$ (since $\lfloor 1/3 \rfloor = 0$). Toggle only when k is a cube: same count, same failure. Toggle multiples of k

only when $k^3 \le m$: no multiplicative structure at all. The wall is real — the student is guessing procedures and checking them, with no principle telling them what the count per prime should be.

The model chain. *Part (a).* Locker m ($1 \le m \le n$) is toggled once for each divisor k of m, so it ends unlocked iff $d(m)$ is odd. Pair each divisor k with m/k. This is an involution on the set of divisors; its fixed points are the k with $k = m/k$, i.e. $k^2 = m$. The non-fixed divisors come in pairs, so $d(m)$ is odd iff there is a fixed point, iff m is a perfect square.

Part (b), the principle. Let K be any set of positive integers and perform T_k (toggle all multiples of k) only for $k \in K$. Locker m is toggled $N_K(m) = \#\{k \in K : k \mid m\}$ times. If membership of K is decided prime by prime — $k \in K$ iff each exponent e_p of k lies in an allowed set $E \subseteq \{0, 1, 2, \dots\}$ with $0 \in E$ — then choosing a divisor $k \mid m$ in K means choosing each exponent independently, so

$$N_K(m) = \prod_{p^a \,\|\, m} c(a), \qquad c(a) = \#\{e \in E : 0 \le e \le a\}.$$

Parity of the product is controlled by the parity of each $c(a)$.

Cubes. Take $E = \{e : e \equiv 0 \text{ or } 1 \pmod 3\}$. Write $a = 3q + s$ with $s \in \{0, 1, 2\}$. The allowed exponents up to $3q - 1$ are two per block of three, $2q$ in all; then $3q$ is allowed, and $3q + 1$ is allowed if $s \ge 1$. So $c(a) = 2q + 1$ if $s = 0$ and $c(a) = 2q + 2$ if $s \in \{1, 2\}$. Thus $c(a)$ is odd iff $3 \mid a$, and $N_K(m)$ is odd iff every exponent of m is divisible by 3, iff m is a cube. Concretely, K is the set of integers $s^3 t$ with t squarefree: $K = \{1, 2, 3, 5, 6, 7, 8, 10, 11, 13, 14, 15, 16, \dots\}$, omitting $4, 9, 12, 18, \dots$. (The same computation with $E = \{e : e \bmod r \in \{0, 1\}\}$ produces the r-th powers; for $r = 2$ every e is allowed and part (a) reappears.)

Numbers $2m^2$. Let T_k toggle the multiples of $2k$. An odd locker is never touched. An even locker i is toggled $\#\{k : 2k \mid i\} = d(i/2)$ times, odd iff $i/2$ is a square by part (a). So the open lockers are exactly $2m^2 \le n$.

Numbers $m^2 + 1$. Let T_k toggle locker i when $k \mid i - 1$, except that locker 1 is toggled only by T_1. For $i \ge 2$ locker i is toggled $d(i - 1)$ times (all divisors of $i - 1 \le n - 1$ are available), odd iff $i - 1$ is a square; locker $1 = 0^2 + 1$ is toggled once. So the open lockers are exactly $m^2 + 1 \le n$, $m \ge 0$. □

Layer 2 THE HOW — the bifurcation

Bifurcating point. Part (a) forks at the words "$d(m)$ is odd". The natural attempt computes $d(m)$ by the product formula; the model pairs divisors. Both finish (a). The fork only shows its cost in (b): the product formula, used as a black box, gives the student nothing to vary, while the pairing and the per-prime count $c(a)$ turn "design a procedure" into "choose a set of allowed exponents".

> *I am being asked to count how many times something happens, and only its parity matters. So I look for a pairing that cancels everything except a few fixed points. Divisors pair as $k \leftrightarrow m/k$; the fixed point is $\sqrt{m}$.*

> *For cubes I will stop guessing procedures. The count is a product over primes of a count $c(a)$, so all I have to design is which exponents a toggling index may use, so that $c(a)$ is odd exactly when $3 \mid a$. Two allowed residues out of three does it.*

- ▷ **Cue:** a count where only the parity matters ("open or closed", "toggled", "sign") ⟶ **Move:** find an involution on the objects being counted; parity equals parity of the fixed points
- ▷ **Cue:** a divisor-type count $\#\{k \mid m : k \text{ has property } P\}$ with P decided prime by prime ⟶ **Move:** it is multiplicative: a product over $p^a \| m$ of the number of allowed exponents up to a
- ▷ **Cue:** "produce the same effect on a shifted or scaled set" ⟶ **Move:** transport the solved case: shift indices ($i \mapsto i - 1$) or scale ($i \mapsto i/2$) and reuse (a)

Layer 3 THE WHY

Why pairing rather than the formula? Because the question never asked for $d(m)$; it asked for $d(m) \bmod 2$. Parity is what survives when you throw away pairs, so whenever parity is the target the natural object is an involution on the things being counted. Divisors of m come with a built-in involution, $k \mapsto m/k$, because "k divides m" is a statement about a factorisation $m = k \cdot (m/k)$, and a factorisation has two sides. The fixed points are the factorisations whose two sides agree. That is where the square comes from: a square is, by definition, a number that factorises with equal sides.

Why does part (b) work for cubes with the rule "exponents $\equiv 0, 1 \pmod 3$"? Go back to what the attendant's procedure is. Choosing T_k to act on multiples of k means locker m records the number of indices k that divide it. A divisor of m is nothing but an independent choice of exponent at each prime, from 0 up to a. So the count factorises over primes, and the question at a single prime is: for which allowed-exponent sets E does $\#(E \cap [0, a])$ switch parity with period 3? Walk e upwards from 0: each allowed exponent flips the parity. We need the parity to be odd at $a = 0$ (so $0 \in E$), even at $a = 1$ (so $1 \in E$), even at $a = 2$ (so $2 \notin E$), odd again at $a = 3$ (so $3 \in E$), and so on: each membership is forced in turn, and the forced pattern is "flip at residues 0 and 1, not at 2". So the rule is not a trick; it is the only solution, and it is visible the moment you think of the count as a walk that flips parity at allowed exponents. Part (a) is the case $r = 2$: two flips per period of length 2 means every exponent is allowed, which is "all k". The general statement explains the special one.

The other two targets reuse (a) by a change of variable. $\{2m^2\}$ is the image of $\{m^2\}$ under $x \mapsto 2x$, and toggling multiples of $2k$ is the same procedure conjugated by that map. $\{m^2 + 1\}$ is the image under $x \mapsto x + 1$, and "$k \mid i - 1$" is the conjugated procedure; locker 1 corresponds to 0, which every k divides, hence the need for a special rule there. The book's stipulation that "locker number 1 is changed only by T_1" is exactly the repair for 0 having infinitely many divisors.

Layer $+\alpha$ BACKTESTING

1. The first cue is "locked or unlocked is changed": a toggle, so only the parity of a count matters. Write down: locker m is changed $d(m)$ times, and I need $d(m) \bmod 2$.

2. Parity of a count suggests an involution with few fixed points: pair k with m/k.

3. Hesitation comes in (b), in the gap between guessing procedures and designing one. It is reinforced away by writing the count as $\prod c(a)$ and designing $c(a)$ at a single prime.

4. The trap is "cubes, so use k^3": toggling multiples of k^3 gives parity $\prod(\lfloor a/3 \rfloor + 1)$, which is odd for $m = 2$. A second trap is forgetting locker 1 in the $m^2 + 1$ version, where 0 is divisible by every k.

5. Recognise faster by asking, of any toggling or sign problem, "what is being counted, and does that set carry a natural involution?". For designed-count problems, ask "is the count multiplicative?" before guessing.

6. Reusable decision rule: *when only a parity is asked for, look for an involution and count its fixed points; when you must design a count with a prescribed parity pattern, reduce to one prime (or one coordinate) and choose which steps flip the parity.*

1967 B5

1967 B5. Show that the sum of the first n terms in the binomial expansion of $(2 - 1)^{-n}$ is $\frac{1}{2}$, where n is a positive integer.

Fields: combinatorics (binomial coefficients); series; probability **Position:** B5 — late in the paper, but short once the sum is read as a probability **Answer:** With $A_n = \sum_{i=0}^{n-1} \binom{n+i-1}{i} 2^{-n-i}$, one has $A_n = \frac{1}{2}$ for every $n \geq 1$: it is the probability that a fair coin shows n heads before n tails.

Caveat. The OCR reads "$(2-1)$-,. is;,"; the exponent $-n$ and the value $\frac{1}{2}$ are restored from the book's solution, which defines A_n as the sum of the first n terms of the expansion of $(2-1)^{-n}$ and concludes $A_n = \frac{1}{2}$. The value was checked in exact arithmetic for $n \le 11$.

Layer 1 THE WHAT — surface analysis

The natural attempt. First decide which expansion is meant. Expanding $(a+b)^{-n}$ with $a=2$, $b=-1$ in powers of $b/a = -\frac{1}{2}$ converges; the other order does not. Using $\binom{-n}{i}(-1)^i = \binom{n+i-1}{i}$,

$$(2-1)^{-n} = 2^{-n}\left(1-\tfrac{1}{2}\right)^{-n} = \sum_{i\ge 0}\binom{n+i-1}{i}2^{-n-i}, \qquad A_n := \sum_{i=0}^{n-1}\binom{n+i-1}{i}2^{-n-i}.$$

Check: $A_1 = \frac{1}{2}$, $A_2 = \frac{1}{4}+\frac{2}{8} = \frac{1}{2}$, $A_3 = \frac{1}{8}+\frac{3}{16}+\frac{6}{32} = \frac{1}{2}$. The pattern invites induction: show $A_n = A_{n-1}$. Split with Pascal's rule, $\binom{n+i-1}{i} = \binom{n+i-2}{i} + \binom{n+i-2}{i-1}$. The first piece is

$$\sum_{i=0}^{n-1}\binom{(n-1)+i-1}{i}2^{-n-i} = \tfrac{1}{2}\Big[A_{n-1} + \tbinom{2n-3}{n-1}2^{-2n+2}\Big],$$

because A_{n-1} stops at $i = n-2$ and the $i = n-1$ term is left over. The second piece, with $j = i-1$, is

$$\sum_{j=0}^{n-2}\binom{n+j-1}{j}2^{-n-j-1} = \tfrac{1}{2}\Big[A_n - \tbinom{2n-2}{n-1}2^{-2n+1}\Big],$$

because it is missing the top term of A_n. Adding,

$$A_n = \tfrac{1}{2}A_{n-1} + \tfrac{1}{2}A_n + 2^{-2n}\Big[2\tbinom{2n-3}{n-1} - \tbinom{2n-2}{n-1}\Big].$$

The bracket vanishes, since $\binom{2n-2}{n-1} = \binom{2n-3}{n-1} + \binom{2n-3}{n-2}$ and the two terms on the right are equal (their lower indices add to $2n-3$). So $A_n = A_{n-1}$ for $n \ge 2$ and $A_n = A_1 = \frac{1}{2}$. This is the book's first solution. It is correct, and every line of it is a place to drop a power of 2 or an index.

The model chain. Toss a fair coin repeatedly. For $0 \le i \le n-1$, the event "the n-th head arrives at toss $n+i$, with exactly i tails before it" has probability

$$\binom{n+i-1}{i}2^{-(n+i)},$$

since the first $n+i-1$ tosses must contain exactly i tails in some order and toss $n+i$ must be a head. These events are disjoint, and their union is the event H: *n heads appear before n tails*. (If n heads come first, at that moment there are $i \le n-1$ tails.) So $A_n = P(H)$.

Let T be the event that n tails appear before n heads. Among the first $2n-1$ tosses there are either at least n heads or at least n tails, not both, so exactly one of H, T happens: $P(H)+P(T) = 1$. Swapping heads and tails is a bijection of outcome sequences preserving probability, so $P(H) = P(T)$. Hence $A_n = P(H) = \frac{1}{2}$. □

(The full series sums to 1, so the tail $\sum_{i\ge n}$ is also $\frac{1}{2}$: the expansion splits exactly in half at the n-th term.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down $A_n = \sum_{i=0}^{n-1}\binom{n+i-1}{i}2^{-n-i}$. At that line the natural attempt reads it as an algebraic object and manipulates it; the model reads each term as a probability, $\binom{n+i-1}{i}$ arrangements times $2^{-(n+i)}$ per sequence, and asks what event the sum describes.

> *Every term is a count of arrangements times 2 to minus the length. That is the probability of some coin-toss event. $\binom{n+i-1}{i}$ with the last place fixed is "the n-th head lands on toss $n+i$". So the sum is the chance of n heads before n tails, and that is $\frac{1}{2}$ by symmetry.*

> *The answer is $\frac{1}{2}$ for every n. A constant $\frac{1}{2}$ is the signature of a symmetry between two outcomes, not of a computation. I should look for the two outcomes.*

- ▷ **Cue:** a sum $\sum (\cdot) p^a q^b$ with $p+q=1$ (especially $p=q=\frac{1}{2}$) $\longrightarrow$ **Move:** read each term as the probability of a sequence event; the sum is the probability of a union
- ▷ **Cue:** $\binom{n+i-1}{i}$ with n fixed and i varying $\longrightarrow$ **Move:** negative binomial: the n-th success on trial $n+i$
- ▷ **Cue:** a claimed constant answer $\frac{1}{2}$ independent of n $\longrightarrow$ **Move:** look for a symmetry swapping two complementary events

Layer 3 THE WHY

Why is the probabilistic reading not a trick? Look at what a binomial coefficient is and what 2^{-N} is. $\binom{N}{i}$ counts the words of length N with i letters T and $N-i$ letters H. 2^{-N} is the probability of any one word of length N under a fair coin. So $\binom{N}{i}2^{-N}$ is, by definition, the probability that N fair tosses give exactly i tails. The problem's terms are that, with one extra constraint visible in the indices: the top is $n+i-1$ but the power of 2 is $n+i$, so one toss is spoken for and not being arranged. A toss that is fixed rather than arranged is a stopping toss, and the only natural stopping rule with n fixed is "stop at the n-th head". The expansion of $(1-x)^{-n}$ is the generating function of exactly this waiting time — that is what negative binomial means — so the problem has literally handed you a probability distribution in disguise and asked for a partial sum.

Why $\frac{1}{2}$ and why cut at n terms? The cut $i \le n-1$ says "fewer than n tails when the n-th head arrives", which is a race to n between heads and tails. A race between two symmetric players is won by each with probability $\frac{1}{2}$, provided it cannot be drawn and always ends. It cannot be drawn and always ends because $2n-1$ tosses decide it (pigeonhole). That is the whole content; the algebraic proof is re-deriving this symmetry one Pascal step at a time, and the vanishing boundary bracket $2\binom{2n-3}{n-1} - \binom{2n-2}{n-1} = 0$ is the symmetry $\binom{2n-3}{n-1} = \binom{2n-3}{n-2}$ surfacing in coordinates.

The book's alternate solution uses a random walk to the boundary of a square, first hitting the side $x=n$ versus the side $y=n$. That is the same race drawn in the plane: x counts heads, y counts tails.

Layer $+\alpha$ BACKTESTING

1. The first cue is the weight 2^{-n-i} sitting next to a binomial coefficient: write "this is $\binom{N}{i}/2^N$ with one toss fixed".

2. A probability model (negative binomial / race to n) with a symmetry argument.

3. Hesitation: "is a probabilistic proof rigorous here?". It is: the sample space of the first $2n-1$ tosses is finite and uniform, and every step is counting. Writing the space explicitly removes the doubt.

4. The trap is the induction with Pascal's rule, which works but has two boundary terms with different powers of 2; one slip and the recursion looks false. A second trap is expanding in the wrong order, in powers of $\frac{2}{-1}$, which diverges.

5. Any sum of $(\cdot)\cdot 2^{-(\cdot)}$, or more generally $(\cdot)p^a q^b$ with $p+q=1$, should be tried as a probability before being manipulated. A constant answer across all n points to symmetry.

6. Reusable decision rule: *when every term of a sum is a count times a probability of a single outcome, name the event the sum describes; the event usually has a structural reason to have the stated probability.*

1967 B6

1967 B6. Let f be a real-valued function having partial derivatives and which is defined for $x^2 + y^2 \leq 1$ and is such that $|f(x, y)| \leq 1$. Show that there exists a point (x_0, y_0) in the interior of the unit circle such that

$$\left(\frac{\partial f}{\partial x}(x_0, y_0)\right)^2 + \left(\frac{\partial f}{\partial y}(x_0, y_0)\right)^2 \leq 16.$$

Fields: analysis (several variables); extremal principle **Position:** B6 — last problem of the paper; hard to start, short once the auxiliary function is found **Answer:** Minimise $g = f + 2(x^2 + y^2)$ over the closed disc; the minimum is attained at an interior point (x_0, y_0), where $\nabla f = -4(x_0, y_0)$, so $|\nabla f|^2 = 16(x_0^2 + y_0^2) < 16$.

Caveat. Two points about the source. (1) The OCR renders the inequality sign as ";;!"; the book's solution (which produces $|\partial f/\partial x| = 4|x_0|$, $|\partial f/\partial y| = 4|y_0|$ at an interior point) confirms "≤ 16". (2) The proof, like the book's, uses the extreme value theorem, so it needs f continuous on the closed disc. "Having partial derivatives" alone does not imply continuity in two variables. The standard reading, followed here, is that f is differentiable (for instance C^1) on the closed disc, hence continuous. The archive has not verified whether the statement survives under the bare hypothesis that partial derivatives exist.

Layer 1 THE WHAT — surface analysis

The natural attempt. The student's first instinct is the one-variable mean value theorem. Along the diameter $y = 0$, $h(t) = f(t, 0)$ satisfies $|h(1) - h(-1)| \leq 2$, so some $t_0 \in (-1, 1)$ has $|f_x(t_0, 0)| = \frac{1}{2}|h(1) - h(-1)| \leq 1$. Excellent — except that $f_y(t_0, 0)$ is completely uncontrolled. Try again along $x = t_0$ to control f_y: that picks a different point, where f_x is now uncontrolled. Any mean-value argument along a line controls one directional derivative at one point, and the problem needs both partial derivatives small *at the same point*.

The next instinct is contradiction: suppose $|\nabla f|^2 > 16$ everywhere inside. Then f changes quickly in some direction at every point, so it should run out of room between -1 and 1. Making this precise requires following a curve along which f keeps increasing (a gradient curve), which needs existence theory for differential equations. It can be done (see the remark in Layer 3), but in an exam it stalls.

The model chain. Assume f is continuous on the closed disc $D = \{x^2 + y^2 \leq 1\}$ and has partial derivatives in the interior. Define

$$g(x, y) = f(x, y) + 2(x^2 + y^2).$$

Compare values. At the centre, $g(0, 0) = f(0, 0) \leq 1$. On the boundary circle, $x^2 + y^2 = 1$ and $f \geq -1$, so $g \geq -1 + 2 = 1$.

g is continuous on the compact set D, so it attains a minimum value m at some point P. Since $m \leq g(0, 0) \leq 1$, if P is on the boundary then $1 \leq g(P) = m \leq g(0, 0) \leq 1$, so $g(0, 0) = m$ and the origin is also a minimiser. Either way there is a minimiser (x_0, y_0) in the interior.

At an interior minimum each partial derivative that exists is zero: $t \mapsto g(x_0 + t, y_0)$ has a minimum at $t = 0$ in an open interval, so its derivative there vanishes, and likewise in y. Since $g_x = f_x + 4x$, $g_y = f_y + 4y$,

$$f_x(x_0, y_0) = -4x_0, \qquad f_y(x_0, y_0) = -4y_0,$$

and therefore

$$f_x(x_0, y_0)^2 + f_y(x_0, y_0)^2 = 16(x_0^2 + y_0^2) < 16. \qquad \square$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains want a point where the gradient is small. The natural attempt tries to *find* such a point by moving along lines (mean value theorem), and meets the one-direction-at-a-time wall. The model *manufactures* a point where the gradient is known exactly: a critical point of a modified function. The split is the line "Define $g = f + 2(x^2 + y^2)$".

> *The only points where I know both partial derivatives at once are critical points. f need not have one. So I add something whose gradient I know, to force a critical point to exist, and read ∇f off it.*

> *To force an interior minimum I need the modified function to be bigger on the boundary than at the centre. f can fall by 2 from centre to edge, so I add a bowl $c(x^2 + y^2)$ that rises by at least 2: $c = 2$. Its gradient is $4(x, y)$, of length less than 4, and $4^2 = 16$.*

- ▷ **Cue:** "show there is a point where the derivative (or gradient) satisfies a bound", with only a bound on the function given $\longrightarrow$ **Move:** add an auxiliary term with known derivative so that the new function has an interior extremum; read the derivative off the critical-point equation
- ▷ **Cue:** need an interior extremum on a closed region $\longrightarrow$ **Move:** compare the value at one interior point with the values on the boundary
- ▷ **Cue:** a mean-value argument controls one direction at a time $\longrightarrow$ **Move:** switch from lines to a two-dimensional extremum

Layer 3 THE WHY

The book's solution begins by setting $g = f + 2(x^2 + y^2)$. Here is where that comes from.

Start from the definition of what you know. You know nothing about f' anywhere; you know $|f| \leq 1$. The only theorem that converts information about *values* into exact information about *derivatives* at a point is: at an interior extremum, the derivative is zero. (Rolle's theorem is this; the mean value theorem is Rolle applied to f minus a linear function.) So a proof must produce an interior extremum of something. f itself may have none — $f(x, y) = x$ is monotone in x — so the something has to be f plus a correction, and to learn about ∇f the correction's gradient must be known. This is the two-variable mean value theorem in disguise: in one variable, subtracting the chord makes endpoint values equal and forces a critical point; here, adding a bowl makes the boundary values beat the centre and forces a critical point.

Why a bowl and why 2? Radial symmetry of the disc says the correction should depend on $x^2 + y^2$; the simplest such function is $c(x^2 + y^2)$. For an interior minimum you need $\min_{\partial D} g \geq g(\text{centre})$. The worst case is $f = -1$ on the edge and $f = 1$ at the centre, a drop of 2, so the bowl must rise by 2 over radius 1: $c = 2$. Then $\nabla(\text{bowl}) = 4(x, y)$, of length < 4, and at the critical point ∇f equals minus that. The constant 16 in the problem is announcing $c = 2$: $16 = (2c)^2$.

Why is 16 generous? The bowl argument wastes a lot: it uses a whole radius to absorb a drop of 2. The truth is much stronger. *Remark (sketch, for $f \in C^1$ on the open disc).* Suppose $|\nabla f| > 1$ everywhere inside. Through the origin run a curve that always moves in the direction $\nabla f / |\nabla f|$ at unit speed (it exists by Peano's theorem, since that field is continuous), extended as far as possible both ways. Along it f increases at rate $|\nabla f| > 1$, so the whole curve has length less than 2, because f varies by at most 2. A maximal curve of finite length has endpoints, and they cannot be interior points (it could be extended), so both ends lie on the boundary circle; from the origin that needs length at least 1 each way, total at least 2. Contradiction. So some interior point has $|\nabla f| \leq 1$, and $f = x$ shows 1 is best possible. This is the gradient-curve idea from the natural attempt, finished; it is not needed for the problem, and the existence step is exactly why the bowl is the exam-room choice.

Layer $+\alpha$ BACKTESTING

1. The first cue is "there exists a point where the derivative is at most…" with only a bound on $|f|$ supplied. Write: I need an interior critical point of $f + (\text{something})$.
2. The extremal principle: build $g = f + \phi$ with an interior minimum; at it, $\nabla f = -\nabla\phi$.
3. Hesitation is at choosing ϕ. Reinforce it away by writing the boundary-versus-centre inequality you need, $\min_{\partial D} g \geq g(0)$, and solving it for the constant.
4. The trap is the one-dimensional mean value theorem along lines, which controls one partial derivative at a time. A rigour trap is to claim the minimum is interior without comparing with the boundary, or to ignore that the extreme value theorem needs continuity.
5. Any "some point has small derivative" statement is a Rolle-type statement. Ask which auxiliary function makes the extremum interior; its derivative gives the bound, and the bound's constant tells you the auxiliary function.
6. Reusable decision rule: *to prove a derivative is small somewhere, add a function with known derivative that makes an extremum interior; then read off the derivative at the extremum. Size the added function by comparing boundary and interior values.*