DAWN PUTNAM SESSION

Putnam 1968 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1968 — Paper A **5**
1968 A1 5
1968 A2 7
1968 A3 10
1968 A4 12
1968 A5 15
1968 A6 18

Putnam 1968 — Paper B **23**
1968 B1 23
1968 B2 25
1968 B3 27
1968 B4 30
1968 B5 32
1968 B6 35

Putnam 1968 — Paper A

1968 A1

1968 A1. Prove
$$\frac{22}{7} - \pi = \int_0^1 \frac{x^4(1-x)^4}{1+x^2}\,dx.$$

Fields: Integral calculus, definite integrals **Position:** A1 (opener) **Answer:** The identity holds exactly; it follows from one polynomial long division of the numerator by $1+x^2$, giving $x^6 - 4x^5 + 5x^4 - 4x^2 + 4 - \dfrac{4}{1+x^2}$, whose termwise integral over $[0,1]$ is $\frac{1}{7} - \frac{2}{3} + 1 - \frac{4}{3} + 4 - \pi = \frac{22}{7} - \pi$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The integrand looks like it wants to be attacked piece by piece. Expand the numerator: $x^4(1-x)^4 = x^4 - 4x^5 + 6x^6 - 4x^7 + x^8$. A natural next move is to integrate each power of x against $1+x^2$ separately, rather than combining them first – after all, a sum of integrals is the sum of the integrals, so why not split immediately?

For the even powers this is harmless: $\int x^4/(1+x^2)\,dx$, $\int x^6/(1+x^2)\,dx$, $\int x^8/(1+x^2)\,dx$ each reduce, by writing $x^{2k} = (x^2+1-1)^k$ expanded binomially or by repeated reduction $x^{2k}/(1+x^2) = x^{2k-2} - x^{2k-2}/(1+x^2)$, to a polynomial minus (eventually) $1/(1+x^2)$, whose antiderivative is $\arctan x$. But the numerator also has odd powers, $-4x^5$ and $-4x^7$, and $\int x^{2k+1}/(1+x^2)\,dx$ is a different animal: the standard reduction is $x^{2k+1}/(1+x^2) = x^{2k-1} - x^{2k-1}/(1+x^2)$, bottoming out at $\int x/(1+x^2)\,dx = \frac{1}{2}\ln(1+x^2)$. So term-by-term integration produces, alongside the polynomial and arctan pieces, several logarithm terms with coefficients coming from the $-4x^5$ and $-4x^7$ contributions. The final claimed answer, $22/7 - \pi$, contains no logarithm at all – so if this route is followed honestly, every one of those log terms must cancel against each other when the five separate antiderivatives are summed and evaluated at $x=1$ and $x=0$ (at $x=0$ they vanish; at $x=1$, $\ln(1+x^2) = \ln 2$ for every one of them, so the cancellation is really a statement that the coefficients sum to zero). That cancellation is true, but discovering it this way means doing five separate reduction computations, keeping track of five separate coefficients of $\ln 2$, and verifying they sum to zero – a real chance for an arithmetic slip to leave a phantom $\ln 2$ sitting in the final answer.

The model chain. The integrand is a single rational function $N(x)/(1+x^2)$ with $\deg N = 8 > 2 = \deg(1+x^2)$. Before doing anything else, divide: perform the polynomial long division of $N(x) = x^8 - 4x^7 + 6x^6 - 4x^5 + x^4$ by $1+x^2$ as one object.

$$x^8 - 4x^7 + 6x^6 - 4x^5 + x^4 = (1+x^2)(x^6 - 4x^5 + 5x^4 - 4x^2 + 4) - 4.$$

(Check: $(1+x^2)(x^6 - 4x^5 + 5x^4 - 4x^2 + 4) = x^6 - 4x^5 + 5x^4 - 4x^2 + 4 + x^8 - 4x^7 + 5x^6 - 4x^4 + 4x^2 = x^8 - 4x^7 + 6x^6 - 4x^5 + x^4$, matching $N(x)$ exactly, so the remainder is the constant -4.) Hence
$$\frac{x^4(1-x)^4}{1+x^2} = x^6 - 4x^5 + 5x^4 - 4x^2 + 4 - \frac{4}{1+x^2}.$$

Every term on the right is elementary to integrate, and only the last one is transcendental:

$$\int_0^1 \left(x^6 - 4x^5 + 5x^4 - 4x^2 + 4\right) dx = \frac{1}{7} - 4\cdot\frac{1}{6} + 5\cdot\frac{1}{5} - 4\cdot\frac{1}{3} + 4 = \frac{1}{7} - \frac{2}{3} + 1 - \frac{4}{3} + 4 = \frac{22}{7},$$

using $\frac{1}{7} - \frac{2}{3} - \frac{4}{3} + 1 + 4 = \frac{1}{7} - 2 + 5 = \frac{1}{7} + 3 = \frac{22}{7}$, and

$$\int_0^1 \frac{4}{1+x^2}\, dx = 4\arctan x\Big|_0^1 = 4\cdot\frac{\pi}{4} = \pi.$$

Subtracting, $\displaystyle\int_0^1 \frac{x^4(1-x)^4}{1+x^2}\, dx = \frac{22}{7} - \pi$, exactly as claimed. No logarithm ever appears, because the division was done on the whole numerator at once.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start by expanding $x^4(1-x)^4$ into the octic $x^8 - 4x^7 + 6x^6 - 4x^5 + x^4$. They split at the very next step: what to do with that polynomial before integrating against $1 + x^2$. The natural attempt integrates it monomial by monomial, treating $\int x^n/(1+x^2)\, dx$ as five independent sub-problems. The model divides the entire octic by $1 + x^2$ as a single polynomial first, obtaining one polynomial quotient and one constant remainder, and only then integrates.

> *This is one rational function with numerator degree* 8 *and denominator degree* 2 *– the textbook condition for polynomial long division – so I should reduce it to a polynomial plus a proper fraction* before *I integrate anything, not split it into five fractions and integrate each on its own; whatever logarithms would appear from doing it piecemeal have to cancel anyway, since the true antiderivative of a proper fraction over* $1 + x^2$ *is pure arctan, and I would rather that cancellation happen for free inside one division than have to verify it by hand across five separate reductions.*

▷ **Cue:** a definite integral $\int P(x)/(1+x^2)\, dx$ where P is a polynomial of degree at least 2, especially one presented as an unexpanded product like $x^4(1-x)^4$ $\longrightarrow$ **Move:** expand P once, then perform a single polynomial long division of the whole numerator by $1 + x^2$ to get quotient (integrate termwise) plus a remainder of degree < 2 over $1 + x^2$ (integrate as arctan, plus at most one log term from any linear part of the remainder) – never integrate the individual monomials of P against $1 + x^2$ separately

Layer 3 THE WHY

The reason the piecemeal route manufactures logarithms it does not need is a fact about partial fractions over $\mathbb{R}$: a proper rational function $R(x)/(1+x^2)$ with $\deg R < 2$ has an antiderivative built only from $\arctan x$ and, if R has a linear term, from $\ln(1+x^2)$ – because $1 + x^2$ has no real roots, so there is no linear factor of the denominator to blame a logarithm on, except through the linear part of the numerator itself. When the whole octic is divided by $1 + x^2$ in one pass, the algebra of the division automatically finds the unique proper remainder, and here that remainder's linear coefficient is exactly 0 (the remainder is the pure constant -4) – which is why the final integral is $22/7 - \pi$ with no log term at all. But a single monomial like x^5 is not itself a proper fraction relative to $1 + x^2$ in a way that respects this cancellation: reducing $x^5/(1+x^2)$ on its own bottoms out at the base case $x/(1+x^2)$, whose antiderivative genuinely is $\frac{1}{2}\ln(1+x^2)$, because in isolation x^5 does carry that behaviour. The log terms that appear from the odd-degree monomials are real features of those monomials individually; they only vanish once recombined with the log terms from the other monomials, because the sum of the numerators (the actual octic) happens to have zero linear coefficient in its remainder. Doing the division on the sum first means this cancellation is not a fact you verify after the computation – it is a fact the division computes on your behalf, since long division respects addition: dividing

$\sum c_i x^i$ by $1 + x^2$ termwise-summing the quotients and remainders gives exactly the same answer as dividing the whole sum at once. The Putnam solution's own remark – "by division, rewrite the integrand as a polynomial plus a rational function with numerator of degree less than 2" – states this as a single instruction precisely because it collapses what would otherwise be several separate log-tracking computations into one.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is $x^4(1-x)^4$ sitting over $1 + x^2$: a numerator of degree 8 over a denominator of degree 2. On seeing it, write down: this is a rational function whose numerator degree exceeds the denominator's, so the standard first move is division, not term-by-term expansion against the denominator.

2. That cue suggests one polynomial long division of the fully expanded numerator by $1 + x^2$, yielding a quotient (elementary, termwise integrable) and a remainder of degree less than 2 over $1 + x^2$ (integrable via arctan, with a possible $\ln(1 + x^2)$ only if the remainder has a nonzero linear term).

3. Hesitation appears right after expanding $x^4(1-x)^4$ into five monomials: it is tempting to integrate $x^4/(1 + x^2)$, $x^5/(1 + x^2)$, and so on immediately, one at a time, since each looks individually tractable. It is reinforced away by noticing that the odd-power monomials, taken alone, genuinely produce $\ln(1 + x^2)$ terms – and the answer $22/7 - \pi$ has none – so integrating termwise commits you to proving several logarithm coefficients cancel, work the single combined division does automatically.

4. The trap is doing five separate reductions and five separate evaluations, each liable to a sign slip, and trusting that the logs will cancel "by symmetry" without checking – or worse, mis-cancelling them and reporting a spurious $\ln 2$ term in the final value.

5. To recognise this faster: treat "numerator degree $\geq$ denominator degree" as an immediate trigger for long division, the same reflex as converting an improper fraction to a mixed number before doing further arithmetic with it – do it once, on the whole numerator, before any term is integrated.

6. Decision rule: for $\int P(x)/Q(x)\,dx$ with Q an irreducible quadratic and $\deg P \geq 2$, always perform one polynomial long division of the entire P by Q before integrating anything. The quotient integrates termwise with no transcendental functions; the remainder (degree < 2) contributes at most one arctan term and one log term, and dividing the sum at once guarantees any impossible extra transcendental terms that piecemeal work would introduce are never generated in the first place.

1968 A2

1968 A2. Given integers a, b, c, d, e, and f with $ad \neq bc$, and given a real number $\varepsilon > 0$, show that there exist rational numbers r and s for which

$$0 < |ra + sb - e| < \varepsilon, \qquad 0 < |rc + sd - f| < \varepsilon.$$

Caveat. The MAA 1965–1984 OCR renders the hypothesis as "a, b, e, c, d, and f with ad -Fbc" and the tolerance as a real number "e" – the same letter as the integer target e in the display. This is a font-substitution collision, not a different problem: the book's own solution disambiguates it completely, solving $ar + bs = e + p$, $cr + ds = f + p$ for a rational p with $0 < p < \varepsilon$ (writing E for the tolerance), which only makes sense if e and f are the two integer targets and ε (rendered "e" by the OCR) is the separate real tolerance. The statement above restores that distinction with ε for the tolerance; the six integers a, b, c, d, e, f and the nonvanishing condition $ad \neq bc$ are otherwise exactly as printed. No mathematical content

was altered.

Fields: Algebra, linear systems, inequalities (miscellaneous) **Position:** A2 **Answer:** Such r, s always exist: solve $ar + bs = e + p$, $cr + ds = f + p$ for any rational p with $0 < p < \varepsilon$; the invertibility of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ (from $ad \neq bc$) makes r, s rational, and both target quantities equal p simultaneously.

Layer 1 THE WHAT — surface analysis

The natural attempt. The statement has the shape of a density argument: as r, s range over the rationals, the point $(ra + sb,\ rc + sd)$ ranges over a set, and the claim is that this set gets arbitrarily close to (e, f) without hitting it. A natural first reaction is to attack the two conditions separately. Fix $s = 0$ for a moment and ask how close ra can get to e for rational r – trivially, as close as you like, since the rationals are dense in $\mathbb{R}$ and $a \neq 0$ (if $a = 0$ this needs $b \neq 0$ instead, already a case split). So treat r as free to approximate e/a arbitrarily well, then separately ask what s has to do to fix up $rc + sd$ to land near f. But now r has already been chosen (to serve the first inequality), so s has only one degree of freedom left to hit the second target – and generically rc will already be off from f by some fixed nonzero amount, forcing $s \approx (f - rc)/d$, which changes s but leaves r untouched, so the first inequality's closeness is not obviously preserved. Worse, chasing $0 < \cdot$ strictly (not just $< \varepsilon$) means you cannot simply let both errors go to 0: you need both quantities to land in a punctured neighbourhood of 0 simultaneously, at the same ε-scale, and juggling two independent approximation processes – one for each coordinate, using up the two degrees of freedom r, s asymmetrically – becomes a genuine two-variable diophantine approximation problem with no clean stopping point. It is not that this approach fails; density arguments of this flavour can be made to work, but only after considerably more machinery (simultaneous rational approximation, or an appeal to the density of $\{(r, s) \in \mathbb{Q}^2\}$ under an invertible linear map with irrational ratios), for a problem that is asking for something far more elementary.

The model chain. Do not aim at the two targets e and f separately. Aim at a single shared quantity instead. Pick any rational p with $0 < p < \varepsilon$ – for instance $p = \varepsilon/2$ if ε is rational, or any rational in $(0, \varepsilon)$ otherwise, which exists because the rationals are dense in $\mathbb{R}$. Now solve the linear system

$$ar + bs = e + p, \qquad cr + ds = f + p$$

for r, s. Since $ad - bc \neq 0$, this system has a unique solution, given by Cramer's rule:

$$r = \frac{(e + p)d - b(f + p)}{ad - bc}, \qquad s = \frac{a(f + p) - c(e + p)}{ad - bc}.$$

Every quantity appearing here – a, b, c, d, e, f, p and the nonzero denominator $ad - bc$ – is rational, so r and s are rational. By construction,

$$ra + sb - e = (e + p) - e = p, \qquad rc + sd - f = (f + p) - f = p.$$

Both quantities equal the same p, and $0 < p < \varepsilon$ by choice. So $0 < |ra + sb - e| < \varepsilon$ and $0 < |rc + sd - f| < \varepsilon$ hold simultaneously, with a single rational r, s, and the proof is complete.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both readings recognise the target as a pair of near-misses controlled by two free rational parameters r, s. They split at what the two inequalities are asked to control *together*. The natural attempt treats $ra + sb - e$ and $rc + sd - f$ as two independent quantities, to be driven close to 0 by (in effect) two separate approximation arguments competing for the same two degrees of freedom. The model chain declares in advance that both quantities will be forced to equal the exact same number p, and only then solves for r, s.

> *I have two unknowns, r and s, and I need two quantities to be small – that is exactly enough freedom to solve a 2×2 linear system exactly, not merely approximately, if I am willing to choose* what *they equal rather than only demanding that they be small. So let me not chase two separate limits; let me set both of them equal to one single free parameter $p \in (0, \varepsilon)$ and solve the resulting linear system for r, s outright.*

▷ **Cue:** two rational unknowns need to be chosen so that two linear combinations of them land within ε of two integer targets, and the coefficient matrix has nonzero determinant
⟶ **Move:** do not chase each inequality separately; introduce one shared free parameter $p \in (0, \varepsilon) \cap \mathbb{Q}$, set both linear combinations equal to (target)$+p$, and solve the resulting 2×2 linear system exactly via Cramer's rule – the nonzero determinant guarantees a rational solution

Layer 3 THE WHY

The move works because $ad - bc \neq 0$ is not a vague genericity hypothesis – it is precisely the condition that the matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible over $\mathbb{Q}$, and Cramer's rule expresses that inverse using only the field operations $+, -, \times, \div$ on the entries of M and the target vector. Since $\mathbb{Q}$ is closed under these operations, *any* rational right-hand side pulls back to a rational solution $(r, s) = M^{-1}(\text{target})$ – exactly, not approximately. This converts the problem from an analytic one (approximate two real numbers by two linear combinations of rationals) into an algebraic one (solve a linear system with rational, invertible coefficients), and algebraic solutions are exact for free – there is no limiting process, no density argument, and no risk that improving one coordinate's approximation disturbs the other's, because both coordinates are hit exactly, on the nose, by the same system. The only place density of $\mathbb{Q}$ in $\mathbb{R}$ is used at all is in choosing the single scalar $p \in (0, \varepsilon)$, a one-dimensional approximation problem that is trivial by comparison with approximating a point in $\mathbb{R}^2$. The official solution's brevity – "the easy solution is obtained by selecting a rational number p ... and solving the linear system" – hides exactly this collapse: it is not that the two-inequality problem was solved by two clever approximations, it is that it was never really a two-inequality problem, because forcing both errors to be the *same* number turns "make two things small" into "solve one linear system," and linear systems with a nonzero determinant do not require analysis at all.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the shape $0 < |ra + sb - e| < \varepsilon$ and $0 < |rc + sd - f| < \varepsilon$ together with the hypothesis $ad \neq bc$ stated up front. On seeing two linear forms in two unknowns paired with an explicit nonvanishing determinant, write down: this hypothesis exists so that a 2×2 system in r, s has a unique, exact rational solution – the problem is handing you invertibility before you even ask for it.

2. That cue suggests treating r, s as the unknowns of a linear system rather than as parameters in an approximation process, and choosing the system's right-hand side to force both target quantities to some common controllable value.

3. Hesitation occurs right at the transition from "I need two things to be small" to "I can make two things equal to the same chosen number": it feels like extra structure is being imposed for free, and it is tempting to instead try to make $ra + sb - e$ small first and only then worry about $rc + sd - f$. This is reinforced away by noticing that $ad \neq bc$ gives full rank – two equations, two unknowns, unique solution – so specifying *both* right-hand sides exactly is no harder than specifying one; the system does not care which values you assign, only that a solution exists, so assign values that make both original inequalities trivially true.

4. The trap is treating this as a genuine simultaneous-approximation problem in two real variables and reaching for machinery (density, irrational rotation arguments, or case splits on which of a, b, c, d is nonzero) that the invertible 2×2 system was specifically designed to make unnecessary.

5. To recognise this faster: whenever a problem hands you n unknowns, n target inequalities, and a hypothesis equivalent to "this $n \times n$ coefficient matrix is invertible," check first whether the inequalities can be converted into an exact linear system by choosing a shared slack value, before reaching for any approximation or density argument.

6. Decision rule: given an invertible linear map from $(r, s) \in \mathbb{Q}^2$ (or $\mathbb{Q}^n$) to a target space, and a requirement that the image land within ε of a fixed point without hitting it, do not approximate the target directly – pick a single small rational slack p, aim the linear system at (target)$+p$ exactly, and solve. Invertibility over $\mathbb{Q}$ turns "approximately" into "exactly, for a nearby exact target," which is always solvable.

1968 A3

1968 A3. Prove that a list can be made of all the subsets of a finite set in such a way that (i) the empty set is first in the list, (ii) each subset occurs exactly once, (iii) each subset in the list is obtained either by adding one element to the preceding subset or by deleting one element of the preceding subset.

Fields: Combinatorics, sets, Hamiltonian circuits **Position:** A3 **Answer:** Such a list always exists, for a set of any finite size, and is built by induction: mirror the list for $n-1$ elements, append the new element x to every set in the mirrored copy, and concatenate.

Layer 1 THE WHAT — surface analysis

The natural attempt. The instinct is to try to build the list directly and combinatorially: start from $\emptyset$, and at each step, greedily add or remove a single element to reach a subset not yet listed. For $\{1,2\}$ this is easy by hand: $\emptyset, \{1\}, \{1,2\}, \{2\}$. For $\{1,2,3\}$ it can still be found by trial: $\emptyset, \{1\}, \{1,2\}, \{2\}, \{2,3\}, \{1,2,3\}, \{1,3\}, \{3\}$. This works, and it is genuinely instructive to find one such list by hand – it builds intuition that the answer is really a Hamiltonian path (in fact, since the last and first subsets here differ by one element too, a Hamiltonian *circuit*) on the graph whose vertices are subsets and whose edges join subsets differing by one element – the n-cube graph. But greedy search does not scale: for $n = 4$ there are 16 subsets and a wrong choice partway through can strand you at a subset all of whose single-element-different neighbours are already used, forcing backtracking. Worse, "prove that a list can be made" for *every* finite set demands a construction that works for all n at once, not a list verified by hand for $n = 2, 3, 4$ and asserted to continue – an existence proof by exhaustive search at each size is not a proof for all sizes. What is missing is not cleverness in the search but a way to *build* the size-n solution mechanically out of the size-$(n-1)$ solution, so that the construction is manifestly available at every size by induction rather than rediscovered by trial each time.

The model chain. Proceed by induction on $n = |S|$. For $n = 1$, say $S = \{1\}$, the list $\emptyset, \{1\}$ satisfies (i)–(iii) trivially.

Suppose the claim holds for every set of size $n-1$: there is a list $A_1, A_2, \ldots, A_{2^{n-1}}$ of all 2^{n-1} subsets of a set T of size $n-1$, with $A_1 = \emptyset$ and consecutive terms differing by exactly one element (and, taking the stronger inductive hypothesis that the list is in fact a circuit, $A_{2^{n-1}}$ and A_1 also differ by exactly one element – this stronger circuit form is what makes the induction close, and it is easy to check it holds already at $n = 1$).

Let $S = T \cup \{x\}$ where $x \notin T$, so $|S| = n$. Build the list for S as

$$A_1,\ A_2,\ \ldots,\ A_{2^{n-1}},\ \ A_{2^{n-1}} \cup \{x\},\ A_{2^{n-1}-1} \cup \{x\},\ \ldots,\ A_2 \cup \{x\},\ A_1 \cup \{x\}.$$

This is the original list, followed immediately by the original list *reversed* and with x added to every term. It has $2^{n-1} + 2^{n-1} = 2^n$ entries, and:

- Every subset of S appears exactly once: a subset of S either contains x or does not; those that do not are exactly the A_i (each once, by the inductive hypothesis), and those that do are exactly the $A_i \cup \{x\}$ (each once, for the same reason).
- The list starts at $A_1 = \emptyset$.
- Consecutive terms within the first half differ by one element, by hypothesis; consecutive terms within the second half, $A_j \cup \{x\}$ and $A_{j-1} \cup \{x\}$, differ by one element because A_j and A_{j-1} do (adding the same fixed x to both sides of a one-element difference does not change that difference).
- The join between the two halves, $A_{2^{n-1}}$ to $A_{2^{n-1}} \cup \{x\}$, differs by exactly the element x – one element, as required.

So the concatenated list satisfies (i)–(iii) for S. Moreover it is again a circuit: the last entry is $A_1 \cup \{x\} = \{x\}$, and the first entry is $A_1 = \emptyset$, which differ by exactly the element x – one element, closing the loop and re-establishing the stronger inductive hypothesis for the next step. By induction, the construction succeeds for every finite n.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches agree that the target is a sequence of all 2^n subsets with consecutive (and, in the strong form, cyclically consecutive) terms differing by one element. They split at the method of construction: search directly for the sequence at the given size n, or build the size-n sequence mechanically from a size-$(n-1)$ sequence already assumed to exist.

> *I do not need to find the list for n elements from scratch – I need to find a* rule *that turns any valid list for $n-1$ elements into a valid list for n elements, because then induction does the rest; and the natural rule is: take the last new element x, run the old list once for the subsets without x, then run it* backwards *for the subsets with x, since running it backwards makes the join at the far end – new list to reversed list, and reversed list back to the start – cost exactly one element each time, for free.*

▷ **Cue:** a set of size n is built as (a set of size $n-1$) $\cup\{x\}$, and a Gray-code-style list (each step changes one element) is wanted for the size-n set $\longrightarrow$ **Move:** take the size-$(n-1)$ list as the first half; append the second half as that same list *reversed*, with x added to every entry; the reversal is what makes both junctions (end of first half to start of second, and end of second half back to the start of the whole list) cost exactly one element

Layer 3 THE WHY

The reason reversal is forced, rather than just repeating the list a second time with x added, comes directly from what a "one-element change" means: it is a symmetric, involutive relation (if B differs from A by one element, A differs from B by that same one element), so the only two ways to move between a subset A_i and a $\{x\}$-augmented subset $A_j \cup \{x\}$ by one change are $j = i$ (adding x) or, if already differing by an element within T, $j \neq i$ (which costs at least two changes: fix the T-part, then add x). The construction needs the *end* of the first half, $A_{2^{n-1}}$, to connect to the *start* of the second half by adding only x – so the second half must begin at $A_{2^{n-1}} \cup \{x\}$, not $A_1 \cup \{x\}$. Symmetrically, the construction needs the *end* of the second half to connect back to the *start* of the whole list, $A_1 = \emptyset$, by removing only x – so the second half must end at $A_1 \cup \{x\}$. Given that the second half must start at $A_{2^{n-1}} \cup \{x\}$ and end at $A_1 \cup \{x\}$, while its *internal* steps must still be one-element changes matching the first half's structure (since $A_j \cup \{x\}$ to $A_{j\pm1} \cup \{x\}$ is a one-element change exactly when A_j to $A_{j\pm1}$ is), the only way to visit all of $A_{2^{n-1}}, \ldots, A_1$ in that order, connecting one-element-change to one-element-change throughout, is to traverse the first list *in reverse*. Nothing about reversal is a trick pulled from nowhere: it is the unique traversal order consistent with needing both endpoints of the augmented list to be $A_{2^{n-1}} \cup \{x\}$ and $A_1 \cup \{x\}$ while preserving adjacency. The book's own

solution states the construction ($S_i = T_i$ for $i < t$, then $S_t = T_t \cup \{n\}$, $S_{t+1} = T_{t-1} \cup \{n\}, \ldots$) without motivating the reversal; this is exactly where that motivation comes from, and it is also why the book remarks, correctly, that "this problem is equivalent to finding a Hamiltonian circuit on an n-cube" – the reversed-mirror-and-append rule is precisely the standard recursive construction of the reflected binary Gray code, read as a Hamiltonian circuit on the hypercube graph Q_n.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "each subset ... obtained ... by adding one element ... or deleting one element" together with "prove that a list can be made" for an arbitrary finite set. On seeing a universally quantified existence claim about a structure defined by one-element moves, write down: this smells like induction on $|S|$, since $|S|$ is the only parameter and the one-element-move condition is exactly the kind of local property that survives a controlled extension from $n-1$ to n.
2. That cue suggests looking for a rule that builds the size-n list out of an assumed size-$(n-1)$ list, rather than searching for the size-n list directly – i.e., strengthen "does a list exist" to "does a *circular* list exist" so the inductive step has something to attach the new element's subsets onto at both ends.
3. Hesitation occurs right after splitting subsets of S into those without x (the old list $A_1, \ldots, A_{2^{n-1}}$) and those with x: the default instinct is to append the augmented copy $A_1 \cup \{x\}, \ldots, A_{2^{n-1}} \cup \{x\}$ in the *same* order, which fails at the join (the old list ends at $A_{2^{n-1}}$, but the new half would start at $A_1 \cup \{x\}$, a jump of possibly many elements). It is reinforced away by asking exactly what has to be true at the junction – only one element may change – and noticing that reversing the second half is the unique fix, since it makes the junction land on $A_{2^{n-1}} \cup \{x\}$.
4. The trap is treating this as a search problem to be solved once at each n (or worse, believing a hand-verified example for small n constitutes a general proof), when the actual content of the problem is an inductive construction, and the size of the search space (the n-cube's 2^n vertices) grows too fast for search to serve as a proof technique.
5. To recognise this faster: whenever a "list/arrange all such objects so consecutive ones differ minimally" problem is indexed by a single size parameter n that is built by adjoining one new element to an $(n-1)$-object, reach immediately for "mirror the smaller solution and append the augmented copy in reverse" – this is the reflex that produces a Gray code, and it generalises to any object built by a one-element extension of a smaller one.
6. Decision rule: to construct a Hamiltonian-type list on a structure built recursively as (structure of size $n-1$) + (one new piece), do not search for the size-n list directly. Take the size-$(n-1)$ list as the first half, and as the second half take the same list, reversed, with the new piece attached to every entry – the reversal is forced because it is the unique way to make both the internal join and the wrap-around join cost exactly one change.

1968 A4

1968 A4. Given n points on the sphere $\{(x, y, z) : x^2 + y^2 + z^2 = 1\}$, demonstrate that the sum of the squares of the distances between them does not exceed n^2.

Caveat. The MAA OCR renders the bound as "nt" and the sphere's braces as "I"; the statement is restored as n^2 on the unit sphere $x^2 + y^2 + z^2 = 1$. Identity check: the book's solution under 1968 A-4 represents the points as unit vectors v_i and proves $\sum |v_i - v_j|^2 = n^2 - \left|\sum v_i\right|^2$, which is exactly this statement. "The distances between them" is read as one

term per unordered pair $\{i, j\}$, $i < j$; this is the reading under which n^2 is the sharp constant (over ordered pairs the sum doubles to at most $2n^2$).

Fields: geometry; linear algebra (vectors, dot products); inequalities **Position:** A4 — mid-paper, but short once the right object is written down; the difficulty is refusing to use coordinates **Answer:** The bound holds, with the exact identity $\sum_{i<j} |v_i - v_j|^2 = n^2 - \left|\sum_i v_i\right|^2$; equality holds precisely when the centroid of the points is the centre of the sphere.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem mentions $x^2 + y^2 + z^2 = 1$, so the first instinct is coordinates. Write the points as (x_i, y_i, z_i) and the squared distance as

$$(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2.$$

Now try small cases. For $n = 2$ the squared distance is at most $(\text{diameter})^2 = 4 = n^2$, with equality for antipodal points. For $n = 3$ you want to show three squared side lengths of a triangle inscribed in the sphere sum to at most 9; the equilateral triangle on a great circle has side $\sqrt{3}$ and gives exactly 9. Promising, but how do you prove it for three points, let alone n?

The coordinate route: expand every bracket, sum over pairs, and collect. The x-part is $\sum_{i<j}(x_i - x_j)^2$, which by a standard identity equals $n\sum x_i^2 - \left(\sum x_i\right)^2$. Most students do not have that identity to hand and grind it out, and many try instead to maximise pair by pair ("each distance is at most 2, so the sum is at most $4\binom{n}{2} = 2n(n-1)$"). That bound is true and useless: for $n \geq 3$ it exceeds n^2. The pairs cannot all be diameters at once, and a pair-by-pair argument has no way to see that. The student is left either with a wall of coordinate algebra or with a bound that is too weak because it ignores how the pairs interact.

The model chain. Represent the points as vectors $v_1, \ldots, v_n \in \mathbb{R}^3$ with $|v_i| = 1$, that is $v_i \cdot v_i = 1$. Put $s = v_1 + \cdots + v_n$.

For each pair,

$$|v_i - v_j|^2 = (v_i - v_j) \cdot (v_i - v_j) = v_i \cdot v_i + v_j \cdot v_j - 2\, v_i \cdot v_j = 2 - 2\, v_i \cdot v_j.$$

Summing over the $\binom{n}{2}$ unordered pairs,

$$D := \sum_{i<j} |v_i - v_j|^2 = n(n-1) - 2\sum_{i<j} v_i \cdot v_j.$$

Now expand the one quantity that contains every cross term exactly once:

$$|s|^2 = s \cdot s = \sum_i v_i \cdot v_i + 2\sum_{i<j} v_i \cdot v_j = n + 2\sum_{i<j} v_i \cdot v_j.$$

So $2\sum_{i<j} v_i \cdot v_j = |s|^2 - n$, and substituting,

$$D = n(n-1) - \left(|s|^2 - n\right) = n^2 - |s|^2 \leq n^2.$$

Equality holds if and only if $s = 0$, that is, the centroid s/n is the centre of the sphere. For every $n \geq 2$ equality is attainable (antipodal pair; equilateral triangle on a great circle; regular tetrahedron; for general n, n equally spaced points on a great circle), so n^2 is the best possible constant. The dimension three plays no role: the identity holds on the unit sphere of any $\mathbb{R}^d$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down the squared distance of one pair. The natural attempt writes it in coordinates, $(x_i - x_j)^2 + \cdots$, and then either bounds each pair separately or expands everything. The model writes it as a dot product, $|v_i - v_j|^2 = 2 - 2\, v_i \cdot v_j$, and notices that the only unknowns left are the cross terms $v_i \cdot v_j$, all of which live inside one square, $|v_1 + \cdots + v_n|^2$.

> *Squared length is a dot product with itself. On the unit sphere $v \cdot v = 1$, so $|v_i - v_j|^2$ is 2 minus twice a cross term, and the sum over all pairs is a constant minus twice the sum of all cross terms.*

> *Where have I seen "the sum of all cross terms $v_i \cdot v_j$" before? In the expansion of $(v_1 + \cdots + v_n)^2$. So the whole sum is controlled by the length of $v_1 + \cdots + v_n$, and a square is never negative.*

> *Each distance being at most 2 gives $2n(n-1)$, which is too big. So the bound must come from the pairs interacting, and the object that records how they interact is the sum of the vectors.*

Layer 3 THE WHY

Why is the dot product the natural language here and not a trick? Go back to what the problem is made of. A sphere centred at the origin is the set of points at distance 1 from the origin; in vector language, $v \cdot v = 1$. A squared distance is $|u - w|^2 = (u - w) \cdot (u - w)$. Both the constraint and the objective are therefore statements about dot products. Coordinates split each dot product into three identical copies, one per axis, and the work triples without adding information. When the constraint and the objective are both quadratic and rotation-invariant, the coordinate-free form is the honest description; coordinates are a costume.

Next, why must the answer involve the sum s? The objective D is symmetric in the n points and quadratic. Once you use $v_i \cdot v_i = 1$, the only degrees of freedom left are the $\binom{n}{2}$ cross terms. A symmetric linear combination of all cross terms with equal weights is, up to constants, exactly what $|s|^2$ expands to. There is no other symmetric quadratic invariant available. So the identity $D = n^2 - |s|^2$ is not discovered by luck; it is the only place the computation can go.

This also explains the equality case geometrically. The quantity D is n times the "moment of inertia" of the points about their centroid: the classical identity $\sum_{i<j} |v_i - v_j|^2 = n \sum_i |v_i - g|^2$, with $g = s/n$, is the parallel-axis theorem in disguise. The moment of inertia about the origin is $\sum |v_i|^2 = n$, and moving the reference point from the origin to the centroid subtracts $n|g|^2 = |s|^2/n$. Multiply by n and you recover $D = n^2 - |s|^2$. The spread of a cloud of points is largest, for a fixed distance from the origin, when the cloud is balanced about the origin. That sentence is the whole problem.

The official solution in the MAA volume says that expanding "the case $n = 3$ suggests the following general identities" and writes them down. Here is where the suggestion comes from: in the $n = 3$ expansion you see $6 - 2(v_1 \cdot v_2 + v_1 \cdot v_3 + v_2 \cdot v_3)$, and the bracket is half of $|v_1 + v_2 + v_3|^2 - 3$. The small case is being used to spot the cross-term sum, and the cross-term sum is recognised as a piece of a square. Small cases are useful here precisely because they show the pattern of which terms appear, not because they suggest a numerical answer.

Finally, why is the pair-by-pair bound doomed? It treats the $\binom{n}{2}$ distances as independent, but they are $\binom{n}{2}$ functions of only n points. Whenever an extremal problem sums a quantity over all pairs, the pairs are coupled, and the best bound comes from an identity that sees all of them at once. The inequality $|s|^2 \geq 0$ is that global view: it is the statement that the points cannot all be far from each other, because they cannot all be on opposite sides of each other.

Layer $+\alpha$ BACKTESTING

1. The cue is "sum of the squares of the distances" together with points on a sphere centred at the origin. Write $|v_i - v_j|^2 = 2 - 2\, v_i \cdot v_j$ immediately.

2. Vector algebra with dot products: express the objective through the cross terms and recognise the cross-term sum inside $|v_1 + \cdots + v_n|^2$.

3. Hesitation comes after writing $D = n(n-1) - 2\sum_{i<j} v_i \cdot v_j$, when the cross terms look uncontrollable individually (each can be as small as -1). Reinforce with: "I cannot bound the cross terms one at a time, but their sum is a square minus a constant."

4. The trap is bounding each distance by the diameter, giving $2n(n-1)$, which exceeds n^2 for $n \geq 3$. The secondary trap is expanding in x, y, z coordinates and losing the structure in triplicate algebra.

5. Any sum over all pairs of a quadratic quantity should trigger the expansion of $\left(\sum_i v_i\right)^2$ (or $\left(\sum x_i\right)^2$ for numbers). The identity $\sum_{i<j}(x_i - x_j)^2 = n\sum x_i^2 - \left(\sum x_i\right)^2$ is the scalar version and is worth knowing cold.

6. When both the constraint and the objective are squared lengths, work with dot products, not coordinates. When the objective sums over all pairs, look for a single square that contains every cross term, and let its nonnegativity supply the bound; the equality case is "that square is zero".

▷ **Cue:** sum of squared distances between points on a sphere centred at the origin $\longrightarrow$ **Move:** write $|v_i - v_j|^2 = 2 - 2\,v_i \cdot v_j$ using $v \cdot v = 1$

▷ **Cue:** a sum of all cross terms $\sum_{i<j} v_i \cdot v_j$ (or $\sum_{i<j} x_i x_j$) $\longrightarrow$ **Move:** rewrite it as $\frac{1}{2}\left(|\sum v_i|^2 - \sum |v_i|^2\right)$

▷ **Cue:** sum over all pairs of $(x_i - x_j)^2$ $\longrightarrow$ **Move:** use $\sum_{i<j}(x_i - x_j)^2 = n\sum x_i^2 - (\sum x_i)^2$

▷ **Cue:** pair-by-pair bound is too weak $\longrightarrow$ **Move:** the pairs are coupled; find one global square whose nonnegativity couples them

▷ **Cue:** constraint and objective both rotation-invariant squared lengths $\longrightarrow$ **Move:** drop coordinates; work in dot products

▷ **Cue:** equality case of a bound proved by "a square ≥ 0" $\longrightarrow$ **Move:** set the square to zero; here the centroid is the centre

1968 A5

1968 A5. Let V be the collection of all quadratic polynomials P with real coefficients such that

$$|P(x)| \leq 1 \quad \text{for all } x \text{ on the closed interval } [0,1].$$

Determine

$$\sup\{\, |P'(0)| : P \in V \,\}.$$

Caveat. Identity check: the book's solution under 1968 A-5 writes $P(x) = ax^2 + bx + c$, derives $P'(0) = b = 4P(\frac{1}{2}) - 3P(0) - P(1)$, bounds by $4 + 3 + 1 = 8$ and exhibits $8x^2 - 8x + 1$; this solves exactly the statement above. The OCR text is clean apart from glyph substitutions. Whether "quadratic" means degree exactly 2 or at most 2 does not affect the answer, since the extremal polynomial has degree 2.

Fields: analysis; polynomial approximation (Chebyshev extremal problems); inequalities **Position:** A5 — late in the paper, but the obstacle is the choice of three sample points, not length **Answer:** The supremum is 8, and it is attained: $P(x) = 8x^2 - 8x + 1$ (the Chebyshev polynomial $T_2(2x-1)$) has $|P| \leq 1$ on $[0,1]$ and $|P'(0)| = 8$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Write $P(x) = ax^2 + bx + c$, so $P'(0) = b$. The goal is the largest $|b|$ subject to $|ax^2 + bx + c| \leq 1$ on $[0,1]$. By replacing P with $-P$ we may aim to maximise b.

The natural move is to translate the constraint. $|P(0)| \leq 1$ gives $|c| \leq 1$. $|P(1)| \leq 1$ gives $|a + b + c| \leq 1$. The third condition concerns the vertex at $x = -b/(2a)$, and only if it lies in $[0,1]$: then $|c - b^2/(4a)| \leq 1$. Now the student splits into cases: $a > 0$ or $a < 0$, vertex inside or outside the interval, and in each case tries to optimise b.

Take the promising case $b > 0$, $a < 0$ (the graph rises from 0 and must come back down). The vertex is at $-b/(2a) > 0$. If the vertex lies beyond 1, P is increasing on $[0,1]$ and $b = P(1) - P(0) - a$, which is bounded only once a is; if the vertex lies inside, the vertex condition enters. The student is already juggling three parameters under inequality constraints. Someone persistent guesses that the extremal P should touch ± 1 as often as possible, tries $P(0) = -1$, $P(\text{vertex}) = 1$, $P(1) = -1$, and solves to find the vertex at $\frac{1}{2}$ and $P(x) = -8x^2 + 8x - 1$, with $b = 8$. This finds the right answer. What it does not do is prove that 8 is an upper bound: "the best polynomial touches the bounds three times" is a heuristic, and the case analysis needed to make it rigorous is long and error-prone.

The model chain. Let $P(x) = ax^2 + bx + c$. Evaluate at $0, \frac{1}{2}, 1$:

$$P(0) = c, \qquad P(\tfrac{1}{2}) = \tfrac{a}{4} + \tfrac{b}{2} + c, \qquad P(1) = a + b + c.$$

Eliminate a and c: $4P(\frac{1}{2}) = a + 2b + 4c$, so

$$4P(\tfrac{1}{2}) - 3P(0) - P(1) = (a + 2b + 4c) - 3c - (a + b + c) = b = P'(0).$$

If $P \in V$, each of $|P(0)|, |P(\frac{1}{2})|, |P(1)|$ is at most 1, so by the triangle inequality

$$|P'(0)| \le 4|P(\tfrac{1}{2})| + 3|P(0)| + |P(1)| \le 4 + 3 + 1 = 8.$$

For sharpness take $P(x) = 8x^2 - 8x + 1 = 2(2x-1)^2 - 1$. Since $(2x-1)^2 \in [0,1]$ for $x \in [0,1]$, $P(x) \in [-1,1]$ there, so $P \in V$; and $P'(0) = -8$. Hence

$$\sup\{|P'(0)| : P \in V\} = 8,$$

and the supremum is a maximum. Equality in the triangle inequality forces $P(\frac{1}{2})$, $-P(0)$, $-P(1)$ to share a sign and all to have modulus 1, so the extremal polynomials are exactly $\pm(8x^2 - 8x + 1)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write $P = ax^2 + bx + c$ and $P'(0) = b$. At that point the natural attempt converts "$|P| \le 1$ on $[0,1]$" into conditions on the coefficients (endpoints and vertex) and optimises. The model does the reverse: it keeps the information as values of P and writes the target b as a fixed linear combination of three values $P(0), P(\frac{1}{2}), P(1)$, each of which is directly bounded by the hypothesis.

> *The hypothesis is about values of P; the question is about a coefficient. A quadratic is determined by three values, so b is some fixed combination $\alpha P(x_0) + \beta P(x_1) + \gamma P(x_2)$, and then $|b| \le |\alpha| + |\beta| + |\gamma|$. I only need to choose the three points well.*

> *Which three points? The ones where the extremal polynomial will touch ± 1. A quadratic that is as steep as possible at 0 while trapped in a band should bounce off the walls at both ends and in the middle, so use $0, \frac{1}{2}, 1$.*

> *If my guess of the extremal polynomial also achieves the bound, I am done in both directions at once: the combination gives $\le$, the example gives $\ge$.*

Layer 3 THE WHY

Why sample the polynomial rather than constrain its coefficients? Look at what each side of the problem is made of. The hypothesis is infinitely many conditions $|P(x)| \le 1$, one for

each $x \in [0,1]$. The unknowns are three coefficients. The object of interest, $P'(0) = b$, is a *linear* function of P. A linear quantity is bounded under box constraints on values in exactly one way: write it as a combination of values and apply the triangle inequality. Coefficients live in a different coordinate system from the constraints; the vertex condition $|c - b^2/(4a)| \le 1$ is not even linear. Every step spent in coefficient space is a step spent translating back and forth.

The fact that makes the translation possible is Lagrange interpolation: a polynomial of degree at most 2 is determined by its values at any three distinct points, and every linear functional (a coefficient, a derivative at a point, an integral) is a fixed linear combination of those three values. Here the combination is found by differentiating the Lagrange basis at 0. With nodes $0, \frac{1}{2}, 1$,

$$L_0(x) = 2(x - \tfrac{1}{2})(x - 1), \quad L_{1/2}(x) = -4x(x - 1), \quad L_1(x) = 2x(x - \tfrac{1}{2}),$$

and $L_0'(0) = -3$, $L_{1/2}'(0) = 4$, $L_1'(0) = -1$. That is where the coefficients $-3, 4, -1$ in the official solution come from. The MAA solution states $f'(0) = 4f(\frac{1}{2}) - 3f(0) - f(1)$ without comment; this is its source.

Any three nodes give an upper bound, but most give a weak one. Try $0, \frac{1}{3}, 1$: the same computation gives weights $-4, \frac{9}{2}, -\frac{1}{2}$ and the bound 9. So why $0, \frac{1}{2}, 1$? Because the bound $\sum |L_k'(0)|$ is attained only if some $P \in V$ takes the value sign $L_k'(0)$ at each node, which means P must reach ± 1 at all three nodes with alternating signs. A quadratic with $|P| \le 1$ on $[0,1]$ that alternates $-1, +1, -1$ (or the reverse) at three points must have its interior extremum at the middle one and touch the walls at the ends; the only such quadratic is $\pm\big(2(2x-1)^2 - 1\big)$, which alternates at $0, \frac{1}{2}, 1$. So the right nodes are the alternation points of the Chebyshev polynomial $T_2(t) = 2t^2 - 1$, carried from $[-1,1]$ to $[0,1]$ by $t = 2x - 1$. The equioscillating polynomial and the sample points are two faces of one object; finding either one gives you the other.

This is also why the natural attempt's heuristic — "touch the bounds as often as possible" — was the right instinct. It identified P. What it lacked was the second half: once you have a candidate extremal P, its touching points are the nodes at which to sample a general P, and the triangle inequality then turns the guess into a proof. The guess and the proof are the same three numbers.

Finally, the size of the answer. The derivative at an endpoint of a polynomial bounded by 1 can be large because nothing on the left of 0 constrains it. The classical Markov inequality says $|P'(x)| \le \frac{2n^2}{b-a}$ for degree n on $[a,b]$, with equality at the endpoints for Chebyshev polynomials; here $n = 2$ and $b - a = 1$ give 8. You do not need Markov's theorem to solve the problem, but knowing it exists tells you that "Chebyshev polynomial" is the expected answer to any question of the form "largest derivative, or largest leading coefficient, of a polynomial trapped in a band".

Layer $+\alpha$ BACKTESTING

1. The cue is "$|P(x)| \le 1$ for all x on $[0,1]$" combined with a question about a linear quantity ($P'(0)$). Write down $P(0), P(\frac{1}{2}), P(1)$ in terms of a, b, c at once.

2. Interpolation: express $P'(0)$ as a fixed combination of sampled values, then bound each value by 1 via the triangle inequality. Sharpness by exhibiting an equioscillating polynomial.

3. Hesitation comes at the choice of sample points. Reinforce with: "sample at the points where my guessed extremal polynomial touches ± 1". If no guess is available, try the endpoints and midpoint first — by symmetry of the interval they are the default.

4. The trap is working in coefficient space, with a vertex condition that is nonlinear and a case split on the sign of a and the vertex location. The secondary trap is finding $8x^2 - 8x + 1$ by the touching heuristic and presenting it as a proof of maximality.

5. "Polynomial bounded on an interval; bound a coefficient, derivative or other linear functional" is one family. Its answer is almost always a Chebyshev polynomial rescaled to the interval, and its proof is almost always interpolation at that polynomial's alternation points.

6. To bound a linear functional of a degree-n polynomial under $|P| \le 1$ on an interval: guess the extremal (equioscillating) polynomial, take its $n+1$ touching points as nodes, write the functional in the Lagrange basis on those nodes, and apply the triangle inequality. The guess proves the lower bound; the interpolation proves the upper bound.

▷ **Cue:** $|P(x)| \le 1$ on an interval and a question about a coefficient or derivative of P ⟶ **Move:** write the target as a linear combination of values of P at $\deg P + 1$ points and use the triangle inequality

▷ **Cue:** choosing interpolation nodes for an extremal bound ⟶ **Move:** use the points where the conjectured extremal polynomial touches ± 1

▷ **Cue:** polynomial trapped in a band, extremise a leading coefficient or derivative ⟶ **Move:** suspect the Chebyshev polynomial T_n rescaled to the interval ($T_2(2x-1) = 8x^2 - 8x + 1$ on $[0,1]$)

▷ **Cue:** a heuristic "touch the bounds as often as possible" produces a candidate ⟶ **Move:** use its touching points as the nodes of the upper-bound argument, turning the guess into a proof

▷ **Cue:** $f'(0)$ for a quadratic from values at $0, \frac{1}{2}, 1$ ⟶ **Move:** $f'(0) = 4f(\frac{1}{2}) - 3f(0) - f(1)$

▷ **Cue:** constraints on values, question in coefficients ⟶ **Move:** stay in the space of values; translate only the target

1968 A6

1968 A6. Determine all polynomials of the form $\sum_0^n a_i x^{n-i}$ with $a_i = \pm 1$ $(0 \le i \le n,\ 1 \le n < \infty)$ such that each has only real zeros.

Caveat. The MAA OCR renders the sum as "$: E \sim a; x^{n-i}$" and the conditions as "$a.; = \pm 1 (0 \sim i \sim n.\ 1 \sim n < \infty)$"; the statement is restored as above, which is the standard wording. Identity check: the book's solution under 1968 A-6 normalises $a_0 = 1$, uses that the sum of squares of the zeros is $a_1^2 - 2a_2$ and their product of squares is a_n^2, applies AM–GM to get $n \le 3$, and lists twelve polynomials; this solves exactly the statement above. The answer list below was also confirmed by a numerical root check over all ± 1 coefficient vectors up to degree 10.

Fields: algebra (polynomials, Vieta's formulas); inequalities (AM–GM) **Position:** A6 — last problem of the morning paper, though short once the right symmetric function is chosen **Answer:** Exactly twelve polynomials: $\pm(x-1)$, $\pm(x+1)$, $\pm(x^2+x-1)$, $\pm(x^2-x-1)$, $\pm(x^3+x^2-x-1)$, $\pm(x^3-x^2-x+1)$. No such polynomial of degree ≥ 4 has only real zeros.

Layer 1 THE WHAT — surface analysis

The natural attempt. Start small. Replacing P by $-P$ does not change the zeros, so take $a_0 = 1$.

$n = 1$: $x + 1$ and $x - 1$ both work.

$n = 2$: $x^2 + a_1 x + a_2$ has discriminant $1 - 4a_2$. That is -3 when $a_2 = 1$ and 5 when $a_2 = -1$. So $x^2 \pm x - 1$ work and $x^2 \pm x + 1$ do not.

$n = 3$: eight cubics. Checking discriminants is tedious; testing $x = \pm 1$ is faster. $x^3 + x^2 - x - 1 = (x+1)^2(x-1)$ and $x^3 - x^2 - x + 1 = (x-1)^2(x+1)$ work. The others either have one real root and two complex ones or are not visibly factorable, and each needs its own check.

$n = 4$: sixteen quartics, and the checking is no longer mechanical. The student now reaches for a general tool. The obvious one is Descartes' rule of signs: a polynomial with nonzero coefficients has at most $V(P)$ positive zeros and at most $V(P(-x))$ negative zeros, where V counts sign changes in the coefficient list; all n zeros real forces $V(P) + V(P(-x)) = n$.

Here is the wall. For ± 1 coefficients, between positions i and $i+1$ the list for $P(x)$ changes sign exactly when $a_i a_{i+1} = -1$, and the list for $P(-x)$ changes sign exactly when $a_i a_{i+1} = +1$. So every consecutive pair contributes one change to exactly one of the two lists, and $V(P) + V(P(-x)) = n$ *always*. Descartes cannot exclude a single polynomial. The student who invested in it has spent their time and learned nothing, and is back to case-checking with no sense of where it ends.

The model chain. Take $a_0 = 1$ (the case $a_0 = -1$ is the negatives). Let $P(x) = x^n + a_1 x^{n-1} + \cdots + a_n$ have zeros $r_1, \ldots, r_n$, all real by hypothesis. By Vieta,

$$\sum_i r_i = -a_1, \qquad \sum_{i<j} r_i r_j = a_2 \ (n \geq 2), \qquad \prod_i r_i = (-1)^n a_n.$$

Hence

$$\sum_i r_i^2 = \Big(\sum_i r_i\Big)^2 - 2\sum_{i<j} r_i r_j = a_1^2 - 2a_2 = 1 - 2a_2, \qquad \prod_i r_i^2 = a_n^2 = 1.$$

The numbers r_i^2 are nonnegative because the r_i are real. By AM–GM,

$$\frac{1 - 2a_2}{n} = \frac{1}{n}\sum_i r_i^2 \ \geq \ \Big(\prod_i r_i^2\Big)^{1/n} = 1,$$

so $1 - 2a_2 \geq n$. For $n \geq 2$: if $a_2 = 1$ this reads $-1 \geq n$, impossible; so $a_2 = -1$ and $n \leq 3$.

$n = 1$: $x \pm 1$, both with a real zero.

$n = 2$: $a_2 = -1$, and $x^2 \pm x - 1$ has discriminant $5 > 0$; both work.

$n = 3$: $a_2 = -1$ and $1 - 2a_2 = 3 = n$, so AM–GM holds with equality, which forces all r_i^2 equal; their product is 1, so every $r_i \in \{1, -1\}$. The multiset of zeros is then $\{1,1,1\}$, $\{1,1,-1\}$, $\{1,-1,-1\}$ or $\{-1,-1,-1\}$, giving $\sum r_i^2 = 3$ in every case but $\sum_{i<j} r_i r_j = 3, -1, -1, 3$ respectively; only the middle two have $a_2 = -1$. They give

$$(x-1)^2(x+1) = x^3 - x^2 - x + 1, \qquad (x+1)^2(x-1) = x^3 + x^2 - x - 1,$$

both with coefficients ± 1.

$n \geq 4$: excluded by $n \leq 3$.

With the negatives, the answer is the twelve polynomials in the vitals line.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains normalise $a_0 = 1$ and can do $n = 1, 2$. The split comes when a general argument is needed. The natural attempt asks a sign-pattern question (Descartes), which sees only the signs of the coefficients — and every sign pattern here is equally good. The model asks a size question: what do the coefficients say about the zeros that realness constrains? It writes down $\sum r_i^2 = a_1^2 - 2a_2$.

> *"Only real zeros" is a hypothesis I can only use through something that is true for reals and false for complex numbers. The cheapest such fact is $r^2 \geq 0$. So I want a symmetric function of the r_i^2 that I can read off the coefficients: $\sum r_i^2 = a_1^2 - 2a_2$.*

> *Every coefficient is ± 1, so that sum of squares is at most 3, whatever n is. But the product of the squares is 1. Many nonnegative numbers with product 1 cannot have a small sum — that is AM–GM, and it caps n.*

> *Descartes' rule counts sign changes, and here each adjacent pair gives exactly one change to $P(x)$ or to $P(-x)$. A tool that only sees signs cannot distinguish these polynomials; I need a tool that sees sizes.*

Layer 3 THE WHY

Why is $\sum r_i^2$ the right object and not a lucky guess? Start from the two things the problem gives you. The coefficients are known only up to sign, and they are all the same size. The zeros are known to be real. The bridge between coefficients and zeros is Vieta: coefficients are the elementary symmetric functions of the zeros. So any usable statement will be a symmetric function of the zeros, computable from the a_i, that realness constrains.

Realness is a strange hypothesis to use because it is qualitative. What is the first quantitative consequence of "r is real"? That $r^2 \geq 0$, and more strongly that $\sum r_i^2$ cannot be small when the r_i are not all small. For complex zeros this fails: z^2 can be negative or non-real, and a polynomial like $x^n + 1$ has $\sum r_i^2 = 0$ for $n \geq 3$ despite all $|r_i| = 1$. So $\sum r_i^2$ is precisely the quantity that distinguishes real-rooted polynomials from the rest, and it is the lowest-degree power sum available: $\sum r_i = -a_1$ is signed and says nothing about size.

Now the collision. From the coefficients, $\sum r_i^2 = a_1^2 - 2a_2 \leq 3$: bounded independently of n. From the constant term, $\prod r_i^2 = 1$: the zeros cannot all be small. n nonnegative numbers with product 1 have sum at least n. So $n \leq 3$. The problem is a tug-of-war between the top of the polynomial (which fixes the sum of squares) and the bottom (which fixes the product), and AM–GM is the rope. The official solution uses exactly this; what it does not say is why one would compute $a_1^2 - 2a_2$ in the first place, and the answer is the paragraph above: it is the first symmetric function in which realness shows up.

It is instructive to see that a sharper-looking tool fails. Newton's inequalities, which hold for every real-rooted polynomial, force $a_{k-1}a_{k+1} = -1$ for $1 \leq k \leq n-1$ when all coefficients are ± 1. For $n = 4$ that leaves $x^4 + ax^3 - x^2 - ax + 1$ with $a = \pm 1$, which passes every Newton condition. Dividing by x^2 and setting $t = x - 1/x$ gives $t^2 + at + 1$, with discriminant -3, so there are no real zeros at all. Newton's inequalities are local (they compare neighbouring coefficients); the obstruction here is global (it compares the second coefficient with the last). AM–GM between $\sum r_i^2$ and $\prod r_i^2$ is the global comparison, and that is why it works where local conditions do not.

Finally, the equality case at $n = 3$ is not a nuisance to be checked by hand; it is information. When AM–GM is tight, all the terms are equal, so the zeros are ± 1. That turns eight cubics into four multisets, of which two survive the $a_2 = -1$ requirement.

Layer $+\alpha$ BACKTESTING

1. The cue is "only real zeros" together with coefficients of fixed size. Write $\sum r_i^2 = a_1^2 - 2a_2$ and $\prod r_i^2 = a_n^2$ immediately.

2. Vieta plus a mean inequality: bound a symmetric function of r_i^2 from the top coefficients, bound another from the constant term, and compare with AM–GM.

3. Hesitation comes after the small cases, when there is no obvious way to deal with general n. Reinforce with: "realness is only usable through $r^2 \geq 0$; which symmetric function of the r_i^2 can I read from the coefficients?"

4. The trap is Descartes' rule of signs, which is the natural real-roots tool but gives $V(P) + V(P(-x)) = n$ for every ± 1 polynomial and so excludes nothing. A second trap is exhaustive case-checking for $n = 4, 5, \ldots$ with no stopping rule. A third is forgetting the negatives $a_0 = -1$ in the final list.

5. Any problem that says "all roots real" and constrains the coefficients should trigger $\sum r_i^2 = e_1^2 - 2e_2 \geq 0$, and, if the product of the roots is also pinned down, AM–GM between $\sum r_i^2$ and $\prod r_i^2$.

6. To use a qualitative hypothesis (real, positive, integer), find its first quantitative consequence and express it in quantities you control. For real roots, that consequence is $r^2 \geq 0$, and the

controllable quantities are the power sums via Vieta. When a sign-counting tool is blind to the problem's data, switch to a size-measuring one.

▷ **Cue:** "all zeros real" with information on the coefficients $\longrightarrow$ **Move:** compute $\sum r_i^2 = a_1^2 - 2a_2$ (monic) and use that it is a sum of nonnegative terms

▷ **Cue:** sum of squares of the roots bounded, product of the roots fixed $\longrightarrow$ **Move:** AM–GM on the r_i^2 bounds the degree

▷ **Cue:** AM–GM holds with equality $\longrightarrow$ **Move:** all terms are equal; read off the roots

▷ **Cue:** coefficients all ± 1 and the question is about real zeros $\longrightarrow$ **Move:** do not use Descartes' rule: $V(P) + V(P(-x)) = n$ for every such polynomial

▷ **Cue:** palindromic or antipalindromic quartic $x^4 + ax^3 + bx^2 \mp ax + 1$ $\longrightarrow$ **Move:** divide by x^2 and substitute $t = x \pm 1/x$ to reduce to a quadratic

▷ **Cue:** "determine all" with a sign symmetry $P \mapsto -P$ $\longrightarrow$ **Move:** normalise the leading coefficient and restore the negatives at the end

Putnam 1968 — Paper B

1968 B1

1968 B1. The temperatures in Chicago and Detroit are x° and y°, respectively. These temperatures are not assumed to be independent; namely, we are given: (i) $P(x^\circ = 70^\circ)$, the probability that the temperature in Chicago is 70°, (ii) $P(y^\circ = 70^\circ)$, and (iii) $P(\max(x^\circ, y^\circ) = 70^\circ)$. Determine $P(\min(x^\circ, y^\circ) = 70^\circ)$.

Caveat. The OCR renders the temperature variables as $x0$ and $y0$ throughout, with the degree sign printed as a zero except where it survives correctly on "70°". Read against the mathematics – a probability problem about the joint distribution of two named random quantities, phrased with a degree symbol reappearing correctly two words later in the same line – this is unambiguously x° and y° (temperature in degrees), not a variable named "$x0$". Reconstructed accordingly; no other statement in this problem is malformed.

Fields: probability, elementary set/event algebra **Position:** B1 – opener of the second half, the easiest problem on the paper **Answer:** $P(\min = 70^\circ) = P(x^\circ = 70^\circ) + P(y^\circ = 70^\circ) - P(\max(x^\circ, y^\circ) = 70^\circ)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The problem hands you three probabilities about two dependent random variables and asks for a fourth, so the reflex is to try to reconstruct the joint distribution: set $p = P(x^\circ = 70^\circ)$, $q = P(y^\circ = 70^\circ)$, $r = P(\max = 70^\circ)$, and attempt to split the sample space into the four exhaustive cases $\{x^\circ < 70^\circ \text{ or } = 70^\circ \text{ or } > 70^\circ\} \times \{\text{same for } y^\circ\}$, hunting for a system of equations linking p, q, r to the unknown $P(x^\circ = 70^\circ, y^\circ = 70^\circ)$ and the tail probabilities $P(x^\circ < 70^\circ)$, etc. This is workable in principle – there genuinely is a consistent joint distribution satisfying any given p, q, r – but it means inventing five or six unknowns for probabilities the problem never asked about and never intends you to compute, then eliminating all of them to recover one combination of p, q, r. It is real casework, entirely undirected: nothing tells the solver in advance that everything except p, q, r will cancel, so the temptation is to keep the full joint table around out of caution, which turns a two-line identity into a page of bookkeeping.

The model chain. Name four *events*, not probabilities: $A = \{x^\circ = 70^\circ\}$, $B = \{y^\circ = 70^\circ\}$, $C = \{\max(x^\circ, y^\circ) = 70^\circ\}$, $D = \{\min(x^\circ, y^\circ) = 70^\circ\}$. Fix any outcome ω, with realised values x, y. Check the two claims $A \cup B = C \cup D$ and $A \cap B = C \cap D$ directly from the definitions of max and min, with no probability involved yet.

If $\omega \in A \cup B$, say $x = 70$ (the case $y = 70$ is symmetric). Either $y \leq 70$, in which case $\max(x, y) = 70$ so $\omega \in C$; or $y \geq 70$, in which case $\min(x, y) = 70$ so $\omega \in D$. Either way $\omega \in C \cup D$. Conversely, $\max(x, y) = 70$ forces both $x \leq 70, y \leq 70$ with at least one equal to 70, so $\omega \in A \cup B$; and $\min(x, y) = 70$ forces both ≥ 70 with at least one equal to 70, likewise giving $\omega \in A \cup B$. So $A \cup B = C \cup D$.

For the intersection: $\omega \in A \cap B$ means $x = y = 70$, which immediately gives max = min = 70, so $\omega \in C \cap D$. Conversely $\omega \in C \cap D$ means $\max(x, y) = \min(x, y) = 70$, which forces

$x = y = 70$ (a set with equal max and min is a single point), so $\omega \in A \cap B$. So $A \cap B = C \cap D$, as sets of outcomes, with no distributional assumption anywhere.

Now apply inclusion–exclusion, which is a probability identity valid for any two events: $P(A) + P(B) = P(A \cup B) + P(A \cap B)$. Substituting the two set identities term by term, $P(A) + P(B) = P(C \cup D) + P(C \cap D) = P(C) + P(D)$ (inclusion–exclusion again, now for C, D). So $P(D) = P(A) + P(B) - P(C)$: $P(\min(x^\circ, y^\circ) = 70^\circ) = P(x^\circ = 70^\circ) + P(y^\circ = 70^\circ) - P(\max(x^\circ, y^\circ) = 70^\circ)$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches start by naming the same four quantities of interest. They split at the very first move after that: the natural attempt asks "what is the joint distribution consistent with these numbers?" and starts assigning probability mass to cells of a table. The model asks a narrower question: "is there a purely set-theoretic (not probabilistic) relationship between the *events* $\{x = 70\}, \{y = 70\}, \{\max = 70\}, \{\min = 70\}$ that holds for every single outcome, regardless of distribution?" That question is answered by writing down what max and min mean and checking two set equalities, with P not yet in the picture at all.

> *I don't need to know the joint distribution to answer this – I need a relationship between these four* events *that is true outcome-by-outcome, because if such a relationship exists, inclusion–exclusion turns it into a relationship between the probabilities for free, and every unknown I was about to invent (the tail probabilities, the individual cell masses) will cancel automatically without my ever naming them.*

Layer 3 THE WHY

The reason a joint-distribution attack is overkill is that max and min are not independent pieces of information about the pair (x, y) – they are a re-encoding of exactly the same two numbers, in the same way that $\{a + b, |a - b|\}$ re-encodes $\{a, b\}$. Whenever a problem swaps between two equivalent encodings of the same underlying data ($\{x, y\}$ versus $\{\max(x, y), \min(x, y)\}$), the productive move is to ask what set identities the two encodings induce on *events*, not to treat the second encoding as new information requiring a fresh distributional model. Here the defining property of max and min – $\max(x, y) = 70$ iff both are ≤ 70 with equality somewhere, $\min(x, y) = 70$ iff both are ≥ 70 with equality somewhere – is precisely what makes $A \cup B = C \cup D$ and $A \cap B = C \cap D$ true by definition rather than by computation, and inclusion–exclusion is the standard tool for converting a set identity between two pairs of events into a probability identity, since $P(A) + P(B) = P(A \cup B) + P(A \cap B)$ holds unconditionally for any two events on any probability space. Once the definitions of max and min are read as event identities rather than numerical facts, the algebra writes itself.

Caveat. The OCR of this problem also drops the source of the original variable names (likely x_C and y_D for the two named cities, subscripted rather than superscripted with degree signs, given the OCR's general unreliability with subscripts and diacritics in this volume); this does not affect the mathematics, since the argument above depends only on there being two real-valued quantities and the number 70, not on their labels.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the appearance of both max and min of the same two variables in one problem; the moment that pairing appears, write down the definitions of max and min as *case splits* on which variable is larger, rather than treating them as opaque functions.
2. That cue suggests looking for a relationship between the four events $\{x = 70\}, \{y = 70\}, \{\max = 70\}, \{\min = 70\}$ that holds outcome-by-outcome, then reaching for inclusion–exclusion to convert it into a probability statement.

3. Hesitation occurs right after naming the four events, when it is tempting to start assigning probability mass to individual outcomes "just to check." The reinforcement is to test the set identity on a couple of concrete numeric examples (e.g. $x = 70, y = 65$; $x = 70, y = 75$) rather than building a table – two or three checks confirm $A \cup B = C \cup D$ without ever invoking a distribution.

4. The trap is assuming that because the problem gives probabilities, the solution must be a computation with numbers; in fact the entire content of the problem is an event identity that happens to be dressed in probabilistic language, and the temptation to distribute mass immediately is what balloons the casework.

5. Recognisable faster next time by naming the pattern: whenever a problem re-expresses the same underlying data in two different ways (here, $\{x, y\}$ versus $\{\max, \min\}$), look first for a set identity linking the two families of events before reaching for a distribution or a computation.

6. Reusable rule: for a probability question stated in terms of max and min (or any other symmetric re-encoding of two variables), write down the events as outcome-level definitions first, check for a set identity ($\cup$, $\cap$, or otherwise) between the two encodings, and let inclusion–exclusion carry that identity to the level of probabilities before inventing a single unknown.

▷ **Cue:** a probability problem stating $\max(X, Y)$ and asking about $\min(X, Y)$ (or vice versa) $\longrightarrow$ **Move:** write out the outcome-level definitions of max/min as case splits and look for a set identity between the two encodings' events, rather than reconstructing the joint distribution

▷ **Cue:** a set identity $A \cup B = C \cup D$ and $A \cap B = C \cap D$ established between two pairs of events $\longrightarrow$ **Move:** apply inclusion–exclusion to both pairs to convert directly into $P(A) + P(B) = P(C) + P(D)$

1968 B2

1968 B2. A is a subset of a finite group G (with group operation called multiplication), and A contains more than one half of the elements of G. Prove that each element of G is the product of two elements of A.

Fields: group theory (finite groups, counting arguments) **Position:** B2 – middle of the second half
Answer: Every $g \in G$ can be written $g = a_1 a_2$ with $a_1, a_2 \in A$; proved via the size of $A \cap gA^{-1}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Fix $g \in G$ and try to count directly. Write out (mentally or on paper) the multiplication table of G: an $n \times n$ grid where row x, column y holds xy, with $n = |G|$. Since G is a group, g appears exactly once in every row and exactly once in every column of the full table (each row and column is a permutation of G). The student restricts attention to the $|A| \times |A|$ sub-block indexed by rows and columns in A, and asks how many times g can appear *outside* that sub-block, then tries to bound it. Outside the A-block, g can appear at most twice per row that has one coordinate in A and one outside – once in the A-row/non-A-column part, once in the non-A-row/A-column part – giving an upper bound of $2(n - |A|)$ appearances of g outside the block. Since g appears exactly n times in total (once per row), it appears at least $n - 2(n - |A|) = 2|A| - n$ times inside the A-block. This works – it is in fact the source's own alternate solution – but it requires carefully justifying the "at most twice per non-A row" bound, keeping track of which rows and columns are being double-counted, and it produces a numeric inequality ($2|A| - n > 0$) as the mechanism, rather than exhibiting the pair (a_1, a_2) directly.

The model chain. Fix $g \in G$. Consider the set $gA^{-1} = \{ga^{-1} : a \in A\}$. The map $a \mapsto ga^{-1}$ is a bijection $G \to G$ (it is the composition of $a \mapsto a^{-1}$, a bijection since every element has a unique inverse, with left multiplication by g, a bijection since G is a group), so $|gA^{-1}| = |A|$.

Now compare A and gA^{-1}, two subsets of G each of size $|A| > |G|/2$. If they were disjoint, $|A| + |gA^{-1}| \leq |G|$, i.e. $2|A| \leq |G|$, contradicting $|A| > |G|/2$. So $A \cap gA^{-1} \neq \varnothing$: there is some element a_1 lying in both sets. Because $a_1 \in gA^{-1}$, there is some $a_2 \in A$ with $a_1 = ga_2^{-1}$. Rearranging, $g = a_1a_2$, with $a_1 \in A$ (directly, from the intersection) and $a_2 \in A$ (by construction). This exhibits the required factorisation of g with no reference to a multiplication table at all.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same numerical fact: $|A| > |G|/2$ means two copies of A's size exceed $|G|$'s size, so two subsets of size $|A|$ cannot be disjoint. They split on *which two subsets of size* $|A|$ to compare. The natural attempt keeps A fixed and tries to isolate g's appearances directly inside the fixed $A \times A$ grid, which forces a counting argument over the whole table. The model instead asks: is there a second subset of G, of the same size as A, built specifically so that any element common to it and A hands me a factorisation of g for free? That question is answered by gA^{-1}.

> *I don't want to count how often g appears in the $A \times A$ block of the table – I want to manufacture a second set of exactly the same size as A, such that any element it shares with A* is itself *half of the equation $g = a_1a_2$, so that finding an intersection point finishes the proof instead of merely bounding a count.*

Layer 3 THE WHY

The problem's whole content is a pigeonhole statement in disguise: two subsets of G, each strictly more than half of G, cannot be disjoint. The natural attempt uses this fact but buries it inside a two-dimensional counting argument over the full $|G| \times |G|$ table, because the multiplication table is the object the equation $g = xy$ naturally lives in, and staring at g's appearances in that table is the default first move. The model chain instead asks what single set, other than A, is guaranteed to have the same cardinality as A and to be directly tied to the target equation $g = a_1a_2$ – and the equation $a_1 = ga_2^{-1}$ says exactly that a_1 ranges over gA^{-1} as a_2 ranges over A, so gA^{-1} *is* the set of "first factors that would work," built by definition rather than discovered by counting. Once the target equation is read backwards as defining a set ($a_1 \in gA^{-1}$ iff $g = a_1a_2$ for some $a_2 \in A$), pigeonhole on $|A| + |gA^{-1}| > |G|$ finishes the proof in one line, because a bijective image of A under $a \mapsto ga^{-1}$ preserves cardinality for free – this is simply the definition of a group (every element invertible, every left translation a bijection), asked to do exactly the work needed and no more.

Caveat. None. Both chains here are independently verified against the mathematics (the bijection $a \mapsto ga^{-1}$ and the counting bound $2|A| - n$ were each checked directly), and both appear in the source – the model chain is the source's primary solution, the natural attempt is its own listed alternate, so no reconstruction across an OCR gap was required for this problem.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "more than one half of the elements": the moment a subset's size is stated as strictly more than half the ambient set's size, write down that two subsets of that size cannot be disjoint (pigeonhole), before doing anything else.
2. That cue suggests looking for two subsets, each of size $> |G|/2$, whose non-disjointness directly encodes the statement to be proved – one of them is given (A); the task becomes constructing the other from the target equation.

3. Hesitation occurs at the step of choosing the second set: it is tempting to reach for A itself twice, or for gA, before checking which one actually encodes $g = a_1a_2$. The reinforcement is to write the target equation as $a_1 = ga_2^{-1}$ first and read off, mechanically, that the set of valid a_1's is $\{ga^{-1} : a \in A\} = gA^{-1}$.

4. The trap is trying to prove the result by directly counting occurrences of g in the $A \times A$ sub-block of the multiplication table; it is correct but requires an auxiliary bound (the "at most twice per row" argument) that the gA^{-1} construction bypasses entirely.

5. Recognisable faster next time by naming the transformation: given a target equation $g = xy$ with x, y constrained to a set A, isolate one variable ($x = gy^{-1}$) and read off the set of values x would need to take as y ranges over A – that set, not the original A twice over, is the right second set for a pigeonhole argument.

6. Reusable rule: whenever a problem asks to show two elements of a constrained set multiply (or add, or combine) to a target, and a size hypothesis (more than half, more than a given fraction) is available, isolate one variable in the target equation to define a second set of the same size built from the target itself, then apply pigeonhole to the two sets rather than counting occurrences in a table.

▷ **Cue:** a size hypothesis of the form "A contains more than half of a finite set G" $\longrightarrow$ **Move:** recall that two subsets of G each of size $> |G|/2$ cannot be disjoint, and look for a second such subset to pair with A

▷ **Cue:** a target equation $g = xy$ with x, y required to lie in a set A $\longrightarrow$ **Move:** isolate one variable ($x = gy^{-1}$) and read off the induced set (gA^{-1}) of values the isolated variable would need, then intersect that set with A

1968 B3

1968 B3. Assume that a 60° angle cannot be trisected with ruler and compass alone. Prove that if n is a positive multiple of 3, then no angle of $360/n$ degrees can be trisected with ruler and compass alone.

Fields: field theory, constructibility, elementary number theory (Euler's ϕ-function) **Position:** B3 – middle of the second half, the paper's field-theory problem **Answer:** An angle of $360/n^\circ$ is trisectible if and only if $3 \nmid n$; in particular it is never trisectible when $3 \mid n$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The hint is doing real work – it hands you the one fact you are allowed to assume, so the instinct is to look for a short geometric argument that reduces the general case straight back to that one fact, using only moves that are always legal with ruler and compass (bisecting an angle, adding two given angles by copying one onto the other). Write $n = 3m$, so the target angle is $360/n = 120/m$ degrees. Suppose, for contradiction, this angle could be trisected: that produces a new angle of $40/m$ degrees. Copying that new angle onto itself m times (always legal – angle addition never needs trisection) gives $m \cdot (40/m) = 40^\circ$. Bisecting 40° (always legal) gives $20^\circ = 60^\circ/3$ – apparently a trisection of 60°, contradicting the hint, and the proof looks finished in four lines.

It is not finished, and the gap is worth stating precisely because it is exactly the gap a strong student will not see until pressed: every one of those four lines used the angle $360/n$ itself as a piece of *given* data, alongside the unit length. But "60° cannot be trisected" is a statement about constructions starting from *unit length alone* – no other angle handed to you for free. Unless $360/n$ happens to be constructible from unit length by itself (which is a much stronger, separate condition on n, governed by which regular polygons are constructible at all), smuggling it in as free data proves nothing about the classical, from-scratch 60° fact. The chain does not fail because a step is wrong; it fails because it silently changes what is being assumed.

The model chain. Work with fields instead of physical constructions. For a positive integer k, let $Q_k = \mathbb{Q}(\cos(360^\circ/k))$, the field generated over the rationals by the cosine of the corresponding central angle; this is the algebraic bookkeeping for "the data available once the angle $360^\circ/k$ is given." Three facts do the work:

(1) $[\,\mathbb{Q}(\cos(360^\circ/k)) : \mathbb{Q}\,] = \phi(k)$, where ϕ is Euler's totient.

(2) Tower law: if $\mathbb{Q} \subseteq Q_k \subseteq Q_{3k}$ (as is the case here, since $\cos(360^\circ/3k)$ generates a field containing $\cos(360^\circ/k)$), then $[Q_{3k} : \mathbb{Q}] = [Q_{3k} : Q_k] \cdot [Q_k : \mathbb{Q}]$.

(3) An angle of $360^\circ/k$ is trisectible – i.e., $\cos(360^\circ/3k)$ is constructible from the data $\{\text{unit length}, \cos(360^\circ/k)\}$ – exactly when the relative degree $[Q_{3k} : Q_k]$ is a power of 2 (the Wantzel-type criterion for ruler-and-compass construction, applied relative to the base field Q_k rather than to $\mathbb{Q}$).

Combining (1) and (2): $[Q_{3k} : Q_k] \cdot \phi(k) = \phi(3k)$, so the relative degree we need is $\phi(3k)/\phi(k)$. Euler's totient is multiplicative on prime powers, and 3 is prime, so $\phi(3k) = 3\phi(k)$ when $3 \mid k$, and $\phi(3k) = 2\phi(k)$ when $3 \nmid k$. Hence the relative degree is exactly 3 when $3 \mid k$, and exactly 2 when $3 \nmid k$.

By fact (3): the angle $360^\circ/k$ is trisectible iff this relative degree is a power of 2. It is 2 (a power of 2) when $3 \nmid k$ – trisectible – and it is 3 (not a power of 2) when $3 \mid k$ – not trisectible. Taking $k = n$ gives exactly the claim: if $3 \mid n$, the angle $360^\circ/n$ is not trisectible. (The case $k = 6$, giving $360^\circ/6 = 60^\circ$ and relative degree 3, recovers the given hypothesis as the smallest instance of the same pattern – it is not an independent input to this argument, but it is a useful sanity check that the machine reproduces it.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from the same reduction $n = 3m$ and the same instinct to relate the general case to the 60° instance. They split at the very first serious question: what exactly does "given the angle $360^\circ/n$" mean as a resource? The natural attempt answers, implicitly, "it means I may draw it and use it in constructions, full stop" – treating it as free extra data outside the unit-length-only world where the 60° fact lives. The model answers instead: "it means I may adjoin $\cos(360^\circ/n)$ to $\mathbb{Q}$ as a base field, and I must track everything – including the 60° fact itself – *relative to* whatever base field is in play." That single reframing is what turns a leaky bisection argument into a provable one.

> *I keep wanting to bisect and add my way from $40/m$ degrees back to 20°, but bisection and addition are legal moves regardless of which base field I'm working over – they don't tell me anything about whether the* one *genuinely hard move, the trisection itself, is legal, and that legality is governed by a field-degree condition, not by which auxiliary angles I happen to have lying around.*

> *The hint isn't there so I can chain elementary constructions back to 60° by hand; it's there to name the smallest instance ($k = 6$) of a pattern I need to prove in general – so my job is to find the invariant (the relative field degree) that makes 60°'s failure and $360^\circ/n$'s failure the* same *failure, not to physically route one angle into the other.*

Layer 3 THE WHY

The classical fact that 60° cannot be trisected is a statement about the field $\mathbb{Q}$: the equation $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$, evaluated at $\theta = 20^\circ$ with $\cos 60^\circ = 1/2$, gives an irreducible cubic $8x^3 - 6x - 1 = 0$ over $\mathbb{Q}$, and a length is constructible from unit length alone exactly when its minimal polynomial over $\mathbb{Q}$ has degree a power of 2 – degree 3 fails that test. The moment the problem instead *gives* you the angle $360^\circ/n$ as raw data, the correct base field is no longer $\mathbb{Q}$ but $Q_n = \mathbb{Q}(\cos(360^\circ/n))$, because that is the precise algebraic content of "you may use this angle freely." Everything downstream – what "trisectible" means, what "constructible" means – has to be re-measured relative to Q_n, by the tower law, which is simply the statement that degrees

of field extensions compose multiplicatively along a chain $\mathbb{Q} \subseteq Q_n \subseteq Q_{3n}$. Euler's totient enters only because it is the known formula for $[\mathbb{Q}(\cos(360^\circ/k)) : \mathbb{Q}]$, i.e. for how large the field generated by a k-th root of unity's cosine actually is; its multiplicativity property ($\phi(3k) = 3\phi(k)$ or $2\phi(k)$ according to whether $3 \mid k$) is exactly what converts "is $3 \mid n$?" into "is the relative degree 2 or 3?", and 2 versus 3 is precisely the power-of-2 test that decided the 60° case in the first place. Seen this way, the 60° hypothesis is not a lemma to be invoked by a geometric trick; it is the $k = 1$-relative (equivalently $k = 6$ absolute) instance of one uniform degree computation that the problem is really asking you to carry out for every k divisible by 3.

Caveat. Fact (3) above – that trisectibility of $360^\circ/k$ from the base data $\{\text{unit}, \cos(360^\circ/k)\}$ is equivalent to the relative degree $[Q_{3k} : Q_k]$ being a power of 2 – is the relative (base-field) generalisation of the classical Wantzel constructibility criterion. I have verified the absolute case ($k = 1$, recovering the 60° irreducible-cubic fact) directly, and the tower-law and Euler-totient computations independently, but I have not re-derived the relative version of Wantzel's theorem itself from the axioms of ruler-and-compass construction here; it is taken as classical field theory, exactly as the problem's own given hypothesis takes the 60° case as classical. Nothing downstream of that citation is unverified.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is being handed a specific numerical instance of a general claim as an assumed hypothesis ("assume 60° cannot be trisected"); write down immediately what integer that instance corresponds to in the general family ($60^\circ = 360^\circ/6$) before trying to use the fact.

2. That cue suggests the intended tool is whatever single computation makes the given instance and the general claim special cases of the same formula – here, a field-degree computation via Euler's ϕ-function – rather than a geometric reduction that physically routes the general angle into the specific one.

3. Hesitation occurs exactly where the elementary bisection-and-addition argument seems to finish in four lines; the reinforcement is to ask, of every step, "relative to what base field (or: using what extra given data) is this move legal?" – if the answer changes partway through the argument, the argument has smuggled in an assumption.

4. The trap is treating "given angle $360^\circ/n$" as equivalent to "constructible from unit length alone": the two are different starting points, and a proof that silently switches between them proves something, but not the stated claim.

5. Recognisable faster next time by naming the object: any constructibility question phrased relative to a *given* (not necessarily from-scratch) angle or length is a relative-extension question, and the fix is to name the base field the given data generates before doing anything else.

6. Reusable rule: when a problem gives you one instance of an impossibility result and asks you to extend it to a family, look for the single algebraic invariant (a field degree, a totient value, a parity) that makes the given instance and the general family literal special cases of the same formula, rather than trying to construct a chain of elementary geometric moves from the general case back to the specific one.

▷ **Cue:** a problem gives one numerical instance of a classical impossibility (e.g. "60° cannot be trisected") and asks for a family $\longrightarrow$ **Move:** identify which integer parameter that instance corresponds to, and look for a single formula (degree of a field extension, value of an arithmetic function) that specialises to the given instance and generalises to the whole family

▷ **Cue:** a construction question about an angle or length that is *given* rather than built from scratch $\longrightarrow$ **Move:** name the base field generated by the given data and re-ask the constructibility question as a *relative* field-extension degree, not an absolute one

1968 B4

1968 B4. Show that if f is real-valued and continuous on $(-\infty, \infty)$ and $\int_{-\infty}^{\infty} f(x)\,dx$ exists, then

$$\int_{-\infty}^{\infty} f\left(x - \frac{1}{x}\right) dx = \int_{-\infty}^{\infty} f(x)\,dx.$$

Caveat. The MAA OCR renders the displayed identity as "J_:f (x - ~) d.t = f_:f(x)dx". The fraction was reconstructed as $1/x$ from the book's own solution, which works throughout with "the graph of $y = x - 1/x$" and the substitutions $x = \frac{1}{2}[y \mp \sqrt{y^2+4}]$; that solution solves exactly the statement above, and this is the standard wording of 1968 B4. The result was also checked numerically for $f(x) = e^{-x^2}(1 + \sin 3x)$: both sides agree to 10^{-15}.

Fields: Real analysis (improper integrals, change of variables); algebra (Vieta) **Position:** B4 (upper middle) **Answer:** The identity holds: each y has exactly two preimages under $x \mapsto x - 1/x$, they sum to y, so their Jacobian weights sum to 1.

Layer 1 THE WHAT — surface analysis

The natural attempt. You see $f(x - 1/x)$ and reach for the substitution $u = x - \frac{1}{x}$, $du = \left(1 + \frac{1}{x^2}\right)dx$. You write $dx = \dfrac{du}{1 + 1/x^2}$ and immediately hit the wall: the new integrand still contains x, and you have to express x in terms of u. You solve $x^2 - ux - 1 = 0$ and get $x = \frac{1}{2}\left(u \pm \sqrt{u^2+4}\right)$, and now there is a sign to choose, and whichever sign you pick covers only half the x-line. You also notice the integrand has a problem at $x = 0$ (where $x - 1/x$ blows up), so the left-hand side is improper at three places, not two.

At this point most people try a different route: test with a specific f, say $f(x) = e^{-x^2}$, and try to evaluate $\int e^{-(x-1/x)^2} dx$ directly. Expanding, $(x - 1/x)^2 = x^2 - 2 + 1/x^2$, so you are looking at $e^2 \int e^{-x^2 - 1/x^2} dx$, which is a known but awkward integral. It does come out as $\sqrt{\pi}$, but a special case proves nothing about general f, and after twenty minutes you have a confirmation and no proof.

The model chain. **Set-up.** Let $\phi(x) = x - \frac{1}{x}$ on $x \neq 0$. Since $\phi'(x) = 1 + \frac{1}{x^2} > 0$, ϕ is strictly increasing on each of $(-\infty, 0)$ and $(0, \infty)$. On $(0, \infty)$ it runs from $-\infty$ (as $x \to 0^+$) to $+\infty$; on $(-\infty, 0)$ it runs from $-\infty$ (as $x \to -\infty$) to $+\infty$ (as $x \to 0^-$). So every real y has exactly two preimages, the two roots of $x^2 - yx - 1 = 0$:

$$x_+(y) = \tfrac{1}{2}\left(y + \sqrt{y^2+4}\right) > 0, \qquad x_-(y) = \tfrac{1}{2}\left(y - \sqrt{y^2+4}\right) < 0.$$

By Vieta, $x_+(y) + x_-(y) = y$ for every y. Differentiating,

$$w_+(y) + w_-(y) = 1, \qquad w_\pm(y) := x'_\pm(y) = \tfrac{1}{2}\left(1 \pm \frac{y}{\sqrt{y^2+4}}\right) \in (0, 1).$$

Substitution on compact pieces. For $0 < \varepsilon < 1 < M$, the substitution $x = x_+(y)$ is a C^1 increasing bijection $[\phi(\varepsilon), \phi(M)] \to [\varepsilon, M]$, so

$$\int_{\varepsilon}^{M} f\big(\phi(x)\big)\,dx = \int_{\phi(\varepsilon)}^{\phi(M)} f(y)\,w_+(y)\,dy,$$

and likewise $\int_{-M}^{-\varepsilon} f(\phi(x))\,dx = \int_{\phi(-M)}^{\phi(-\varepsilon)} f(y)\,w_-(y)\,dy$. As $\varepsilon \to 0^+$ and $M \to \infty$ the limits of integration go to $\mp\infty$ and $\pm\infty$ independently, so the four improper pieces of the left side

over $(-\infty,-1],[-1,0),(0,1],[1,\infty)$ converge if and only if the four integrals $\int_{-\infty}^{0} fw_{\pm}$, $\int_{0}^{\infty} fw_{\pm}$ converge.

Convergence (Abel's test by hand). Let $F(y)=\int_0^y f(t)\,dt$. Because $\int_{-\infty}^{\infty} f$ exists, F has finite limits at $\pm\infty$ and is therefore bounded, say $|F|\le C$. Each $w=w_{\pm}$ is C^1, monotone ($w'_+(y)=2(y^2+4)^{-3/2}>0$, $w'_-=-w'_+$) and bounded. Integrating by parts on $[0,B]$,

$$\int_0^B f(y)w(y)\,dy = F(B)w(B)-\int_0^B F(y)w'(y)\,dy.$$

As $B\to\infty$, $F(B)w(B)$ converges (both factors do), and $\int_0^{\infty}|Fw'|\le C\int_0^{\infty}|w'|=C\,|w(\infty)-w(0)|<\infty$, so the last integral converges absolutely. The same argument works on $(-\infty,0]$. All four pieces converge.

Adding. Now every integral in sight converges, so they may be added:

$$\int_{-\infty}^{\infty} f\big(x-\tfrac{1}{x}\big)\,dx=\int_{-\infty}^{\infty} f(y)\,w_+(y)\,dy+\int_{-\infty}^{\infty} f(y)\,w_-(y)\,dy=\int_{-\infty}^{\infty} f(y)\,dy. \qquad \blacksquare$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write $u=x-1/x$ and solve the quadratic $x^2-ux-1=0$. At that line the natural attempt treats the $\pm$ as a nuisance — a branch to be chosen — and stalls. The model chain keeps *both* roots, because the map is two-to-one, and then asks what the two roots have in common: their sum is y.

> *This map is two-to-one, so after substituting I'll get two integrals of $f(y)$ against two different weights. The weights are the derivatives of the two roots of $x^2-yx-1=0$ — and I know the sum of those roots without solving anything: it's y. So the weights add up to 1, and the two integrals add up to $\int f$.*

> *The right-hand side only knows that $\int f$ converges, not $\int|f|$. So I can't split and recombine freely; I have to prove each weighted piece converges, and the tool for "convergent integral times bounded monotone weight" is integration by parts against $F(y)=\int_0^y f$.*

Layer 3 THE WHY

Why should the sum of the preimages be the first thing to look at? Go back to what a change of variables is. When $x\mapsto y=\phi(x)$ is one-to-one, $\int f(\phi(x))\,dx=\int f(y)\,\big|\frac{dx}{dy}\big|\,dy$: every y is counted once, with the weight $|dx/dy|$ measuring how much x-length is packed into a unit of y-length. When ϕ is k-to-one, the same bookkeeping gives $\int f(y)\sum_{\text{preimages}}\big|\frac{dx_i}{dy}\big|\,dy$. That formula is the entire content of the problem. The claim "the integral of $f\circ\phi$ equals the integral of f for *every* f" is then equivalent, by the definition of what it means for two weights to integrate every f the same way, to

$$\sum_i\Big|\frac{dx_i}{dy}\Big|=1 \quad \text{for every } y.$$

So the problem is not asking you to evaluate anything. It is asking you to check a single identity about the preimages. Once that is seen, the preimages are the natural object of interest (in the Blueprint's sense: the circle's centre), and a sum of derivatives of preimages is the derivative of the sum of preimages. The preimages are roots of a quadratic, and the sum of the roots of a quadratic is the one fact about roots you know without the quadratic formula. That is why Vieta appears: not as a trick, but because the question reduces to "what is $x_+ + x_-$?" and Vieta is the cheapest answer to that question.

This also tells you which maps work. Any $\phi(x) = x - \sum_i \frac{a_i}{x-b_i}$ with $a_i > 0$ has the same property (the Cauchy–Schlömilch / Glasser master theorem): clearing denominators gives a polynomial equation in x whose leading two coefficients make the sum of roots equal to y plus a constant, and each branch is increasing. Here $a_1 = 1$, $b_1 = 0$.

The book's solution says only that one "makes the change of variables $x = \frac{1}{2}[y \pm \sqrt{y^2+4}]$" and that "a term involving $y/\sqrt{y^2+4}$ cancels". That cancellation is $w_+ + w_- = 1$. Here is where it comes from: $x_+ + x_- = y$, differentiated.

The second thought — about convergence — is forced by the hypothesis. "$\int_{-\infty}^{\infty} f$ exists" is weaker than absolute integrability; f can be unbounded and oscillate wildly at infinity, and then $f(x - 1/x)$ oscillates wildly near $x = 0$. Near 0 the weight $w_\pm$ is not constant, so convergence of the weighted integral is not free. What rescues it is that the weights are monotone and bounded, and integration by parts converts "$\int f$ converges" into "F is bounded", which is exactly what is needed to control $\int Fw'$ when w' has finite total mass. That is the proof of Abel's test; writing it out costs four lines and removes any dependence on a named theorem.

Layer $+\alpha$ BACKTESTING

1. The cue is $f(x - 1/x)$ inside an integral whose answer is claimed to be $\int f$ for *arbitrary* f. Write down: "$y = x - 1/x$ is two-to-one; the preimages are the roots of $x^2 - yx - 1 = 0$; their sum is y."

2. Change of variables for a many-to-one map: split the domain into monotone branches, substitute on each, and add the Jacobian weights.

3. Hesitation comes at the $\pm$ in the quadratic formula ("which root?"). It is reinforced away by the rule: a two-to-one map needs both roots, and you should reach for their symmetric functions before their explicit forms.

4. Two traps. First, testing a specific f (the Gaussian), which confirms without proving. Second, adding four improper integrals without first showing each converges; with only conditional convergence of $\int f$ that step needs the integration-by-parts argument.

5. Any time a problem says "for every f, $\int f \circ \phi = \int f$", translate it into "the Jacobian weights of ϕ's preimages sum to 1" and look at the preimages as roots of an equation.

6. Decision rule: when a substitution produces an equation with several roots, compute the symmetric function of the roots that the problem needs (sum, product) by Vieta before ever writing the roots down explicitly.

▷ **Cue:** The integrand is $f(\phi(x))$ with ϕ not injective, and the claimed answer is independent of f $\longrightarrow$ **Move:** Split into monotone branches, substitute on each, and prove the sum of the branch Jacobians $|dx_i/dy|$ is identically 1

▷ **Cue:** The preimages of y are the roots of a polynomial in x whose coefficients depend on y $\longrightarrow$ **Move:** Read off the sum of the roots by Vieta and differentiate it, instead of using the explicit root formula

▷ **Cue:** An improper integral $\int f$ is only known to converge, and you need $\int fw$ with w bounded and monotone $\longrightarrow$ **Move:** Integrate by parts against $F(y) = \int_0^y f$: F is bounded and w' has finite total mass

1968 B5

1968 B5. Let p be a prime number. Let J be the set of all 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ whose entries are chosen from $\{0, 1, 2, \ldots, p-1\}$ and satisfy the conditions $a + d \equiv 1 \pmod{p}$, $ad - bc \equiv 0 \pmod{p}$. Determine how many members J has.

Caveat. The MAA OCR renders the matrix as "~~)" and the modulus as "P". The matrix was reconstructed as $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ from the conditions, which name exactly the four entries a, b, c, d, and from the book's solution, which counts solutions of $bc \equiv a(1-a)$ with d determined by a — that solution solves exactly the statement above. The answer $p^2 + p$ was confirmed by exhaustive enumeration for $p = 2, 3, 5, 7, 11, 13$ ($6, 12, 30, 56, 132, 182$ matrices).

Fields: Number theory (congruences, $\mathbb{Z}/p$ as a field); linear algebra (idempotent matrices); counting
Position: B5 (but one of the gentler B5s) **Answer:** $|J| = p^2 + p$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You eliminate d first, because the trace condition gives it for free: $d \equiv 1 - a$. The determinant condition becomes

$$bc \equiv a(1-a) \pmod{p}.$$

So far so good. Now you fix a and count pairs (b, c) with $bc \equiv k$, where $k = a(1-a)$. This needs two counts. If $k \not\equiv 0$, each of the $p-1$ nonzero values of b gives exactly one $c \equiv kb^{-1}$, and $b = 0$ gives nothing: $p-1$ pairs. If $k \equiv 0$, then $b = 0$ or $c = 0$; that is $p + p$ pairs minus the pair $(0,0)$ counted twice: $2p - 1$ pairs. Then you have to know for which a the value k vanishes: $a(1-a) \equiv 0$ forces $a \equiv 0$ or $a \equiv 1$ (because p is prime). Assembling,

$$|J| = 2(2p-1) + (p-2)(p-1) = 4p - 2 + p^2 - 3p + 2 = p^2 + p.$$

This is the book's solution, and it works. It costs two sub-counts, one inclusion–exclusion (the -1 for $(0,0)$), and a final polynomial simplification; it is also where most lost marks come from — answers of $p^2 + p + 2$ (forgetting the -1) and $p^2 - p$ (forgetting that $k = 0$ is special) are the typical casualties.

The model chain. **Reduce.** From $a + d \equiv 1$, $d \equiv 1 - a$ is determined by a, so $|J|$ is the number of triples $(a, b, c) \in (\mathbb{Z}/p)^3$ with

$$bc \equiv a - a^2 \pmod{p}. \tag{$*$}$$

Split on the variable that appears linearly. In $(*)$ the variable c appears once, multiplied by b.

- If $b \not\equiv 0$: b is invertible mod p (here primality is used), so for each of the $p-1$ choices of b and each of the p choices of a there is exactly one $c \equiv b^{-1}(a - a^2)$. That is $p(p-1)$ triples.
- If $b \equiv 0$: $(*)$ reads $a(1-a) \equiv 0$. Since $\mathbb{Z}/p$ has no zero-divisors, $a \equiv 0$ or $a \equiv 1$ (2 choices, distinct because $p \geq 2$), and c is unconstrained (p choices). That is $2p$ triples.

The two cases are disjoint and exhaustive, so

$$|J| = p(p-1) + 2p = p^2 + p. \qquad \blacksquare$$

Structural check. The characteristic polynomial of $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $x^2 - (\operatorname{tr} A)x + \det A = x^2 - x$, so by Cayley–Hamilton $A^2 = A$: the members of J are exactly the idempotent 2×2 matrices over $\mathbb{F}_p$ with trace 1, i.e. the rank-one projections. Such a projection is fixed by its image (a line in $\mathbb{F}_p^2$) and its kernel (a different line), and any ordered pair of distinct lines gives one. The plane $\mathbb{F}_p^2$ has $(p^2-1)/(p-1) = p+1$ lines, so there are $(p+1) \cdot p$ projections. Same answer, no arithmetic.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains reach $bc \equiv a(1-a)$. At that line the natural attempt fixes a — the variable that looks like the "parameter" — and is left counting solutions of a product equation $bc \equiv k$, which splits by whether k is zero. The model fixes b, the coefficient of the linear unknown c, and the equation becomes solvable for c in one step except on the single value $b = 0$.

> *I have one equation in three unknowns. Which unknown can I solve for outright? c appears to the first power, multiplied by b. So I case on whether b is zero: if not, c is forced and a, b are free; if so, the equation collapses to a condition on a alone.*

> *Trace* 1, *determinant* 0 — *that is the characteristic polynomial* $x^2 - x$. *These matrices satisfy* $A^2 = A$. *I am counting projections onto a line along another line.*

Layer 3 THE WHY

Why is "solve for the linear variable" the natural first move and not a trick? Go back to what counting solutions of one equation in several unknowns means. You are counting the points of a set $\{(a, b, c) : F(a, b, c) = 0\}$. If F can be written as $F = b\,c - g(a)$, then on the region where the coefficient b of c is invertible, the set is the *graph* of a function $c = g(a)/b$ over the (a, b)-plane, and a graph over a region has exactly as many points as the region. All the difficulty is squeezed into the locus where the coefficient vanishes. That is the whole content of the method: count the generic part as a graph, then handle the degenerate slice separately, where the equation always becomes simpler (here: $g(a) = 0$ with c free). Fixing a instead does not make anything a graph — $bc \equiv k$ is a hyperbola, and counting points on a hyperbola is exactly the sub-problem you then have to solve, twice.

The word *prime* in the statement is the other half of the WHY. It is used twice, and only twice: to invert b, and to conclude $a(1-a) \equiv 0 \Rightarrow a \in \{0, 1\}$. Both are the field axioms of $\mathbb{Z}/p$ (every nonzero element is a unit; no zero-divisors). Drop primality and both fail. Enumeration confirms that the formula breaks: modulo 4 there are 24 such matrices, not 20; modulo 6 there are 72, not 42 ($72 = 6 \cdot 12$, the product of the counts mod 2 and mod 3, by the Chinese remainder theorem). So the prime hypothesis is not decoration; it is the licence for the division step, and a solution that never visibly uses it is suspect.

Where does the projection picture come from, and why is it worth writing down even though the counting proof is complete? The two conditions are the two coefficients of the characteristic polynomial. Any time a problem prescribes trace and determinant of a 2×2 matrix, it is prescribing $x^2 - (\mathrm{tr})x + \det$, and Cayley–Hamilton turns that into a matrix identity. Here the identity is $A^2 = A$, and an idempotent is a geometric object: it is determined by where it projects (image) and what it kills (kernel). Counting geometric objects in $\mathbb{F}_p^2$ reduces to counting lines, and there are $p + 1$ of them (each line through the origin contains $p - 1$ nonzero vectors, and the $p^2 - 1$ nonzero vectors are partitioned among the lines). Thus $p^2 + p = (p+1)p$ is not a coincidence of algebra: it factors because it is "choose a line, then choose a different line". That factorisation is also a check on your arithmetic — an answer that does not factor as $p(p+1)$ should make you re-count.

The book's solution fixes a. It is correct, and it is the more laborious route; the extra cost is the $2p - 1$ inclusion–exclusion and the separate treatment of $a \in \{0, 1\}$. Casing on b merges both into the single line "$b = 0$ forces $a \in \{0, 1\}$ and frees c".

Layer $+\alpha$ BACKTESTING

1. The first cue is the trace condition $a + d \equiv 1$: write $d = 1 - a$ immediately and reduce to $bc \equiv a(1-a)$. The second cue is the pair (trace, determinant): write down the characteristic polynomial $x^2 - x$.

2. Counting points on one equation in several unknowns over a field: solve for a variable that appears linearly, count the generic part as a graph, then handle the slice where its coefficient vanishes. Alternatively, Cayley–Hamilton to recognise the matrices as projections and count lines.

3. Hesitation comes at "which variable do I fix?" after reaching $bc \equiv a(1-a)$. Reinforce it away with the rule: fix the coefficient of the linear unknown, not the variable that looks like a parameter.

4. The trap is the product equation $bc \equiv 0$: it has $2p-1$ solutions, not $2p$ and not $p-1$, and whichever route you take the degenerate case must be counted separately. A second, quieter trap is never using that p is prime; the count is false for composite moduli.

5. Whenever a 2×2 matrix is constrained by its trace and determinant, rewrite the constraint as its characteristic polynomial and apply Cayley–Hamilton before counting anything. Whenever an equation is linear in one unknown, count by solving for that unknown.

6. Decision rule: to count solutions of $F = 0$ over a finite field, find the unknown in which F is linear, split on whether its coefficient is zero, and count the nonzero-coefficient part as a graph. If the final count does not factor the way the structure suggests, recount.

▷ **Cue:** An equation to be counted over $\mathbb{Z}/p$ has an unknown appearing only to the first power, as in $bc \equiv g(a)$ $\longrightarrow$ **Move:** Case on whether its coefficient is zero; when it is not, that unknown is determined and the remaining unknowns are free

▷ **Cue:** A 2×2 matrix is constrained by its trace and determinant $\longrightarrow$ **Move:** Write the characteristic polynomial and apply Cayley–Hamilton; trace 1 and determinant 0 means $A^2 = A$

▷ **Cue:** The problem says "p prime" in a congruence count $\longrightarrow$ **Move:** Locate the division step or the no-zero-divisors step that needs it; if the solution never uses it, the solution is incomplete

▷ **Cue:** Counting objects in $\mathbb{F}_p^2$ (projections, subspaces, pairs of directions) $\longrightarrow$ **Move:** Count lines through the origin: $(p^2-1)/(p-1) = p+1$

1968 B6

1968 B6. A set of real numbers is called compact if it is closed and bounded. Show that there does not exist a sequence $\{K_n\}_{n=0}^{\infty}$ of compact sets of rational numbers such that each compact set of rationals is contained in at least one K_n.

Caveat. The MAA OCR renders the sequence as "$\{K_n I :' \,..o$"; this was read as $\{K_n\}_{n=0}^{\infty}$. The book's solution under the label B-6 ("For each n, there is a rational $r_n \notin K_n$ with $0 \le r_n < 1/n \ldots$ Let $S = \{0, r_1, r_2, \ldots\}$") solves exactly this statement, and the book's index files it under "Real analysis — compact sets, 1968 B-6". The indexing from 0 is immaterial; the book's own solution uses $1/n$ and so tacitly starts at $n = 1$. Below the witnesses are taken in $[0, \frac{1}{n+1}]$ so that K_0 is handled too.

Fields: Real analysis; point-set topology of $\mathbb{R}$ (closed sets, compactness, density of $\mathbb{Q}$) **Position:** B6 (last problem of the paper, but short once the right frame is found) **Answer:** No such sequence exists: for each n pick a rational $r_n \in [0, \frac{1}{n+1}] \setminus K_n$ (possible because a closed set of rationals contains no interval's worth of rationals); then $S = \{0\} \cup \{r_n\}$ is a compact set of rationals contained in no K_n.

Layer 1 THE WHAT — surface analysis

The natural attempt. A statement of the form "there does not exist a sequence such that…" makes most strong students do one of two things first.

Counting. The K_n are countably many; surely the compact sets of rationals are "more" than that? They are — there are continuum many (every subset A of $\{1/m : m \geq 1\}$ gives a compact set $A \cup \{0\}$). But counting proves nothing here, because *one* K_n can contain continuum many compact sets: $K = \{0\} \cup \{1/m\}$ contains all of those $A \cup \{0\}$ at once. So the cardinality comparison is $\aleph_0 \times 2^{\aleph_0}$ against $2^{\aleph_0}$, and the wall is immediate.

Trying to build the sequence. The second instinct, and the better one, is to try to prove the opposite and see what breaks. Over $\mathbb{R}$ it is easy: $K_n = [-n, n]$ works, since every bounded set lies in some $[-n, n]$. Over $\mathbb{Q}$ you try the analogue $\mathbb{Q} \cap [-n, n]$, and it is not closed: its closure is all of $[-n, n]$, which contains irrationals. So you shrink to something finite that grows: $K_n = \{p/q \in [-n, n] : 1 \leq q \leq n\}$. Each K_n is finite, hence compact, and every rational lies in some K_n. Then you test it on the most basic infinite compact set of rationals, $\{0\} \cup \{1/m : m \geq 1\}$, and it fails: the denominators are unbounded, so no finite K_n holds it. You patch by throwing in convergent sequences, $K_n = \{0\} \cup \{1/m\} \cup \dots$, and then $\{0\} \cup \{1/m^2 + \sqrt{}\text{-free shifts} \dots\}$ — you can always find a sequence your current patch missed. After fifteen minutes you are convinced the answer is "no" and you have no proof, only a feeling that the patches never end.

The model chain. Let $K_0, K_1, K_2, \dots$ be any sequence of compact sets, each consisting only of rational numbers. We construct a compact set $S \subseteq \mathbb{Q}$ with $S \not\subseteq K_n$ for every n.

Step 1: each K_n misses rationals arbitrarily close to 0. Fix n and put $\delta_n = \frac{1}{n+1}$. Suppose, for contradiction, that $\mathbb{Q} \cap [0, \delta_n] \subseteq K_n$. Every real $x \in [0, \delta_n]$ is a limit of rationals in $[0, \delta_n]$ (density of $\mathbb{Q}$), and K_n is closed, so it contains all its limits; hence $[0, \delta_n] \subseteq K_n$. But $[0, \delta_n]$ contains irrationals (for instance $\delta_n/\sqrt{2}$), and K_n contains only rationals. Contradiction. So there is a rational

$$r_n \in [0, \delta_n] \quad \text{with} \quad r_n \notin K_n.$$

Step 2: the diagonal set. Put $S = \{0\} \cup \{r_n : n \geq 0\}$. Then:

- $S \subseteq \mathbb{Q}$, since 0 and every r_n are rational.
- S is bounded: $S \subseteq [0, 1]$.
- S is closed. Let x be a limit of a sequence (s_k) of points of S. If some point of S occurs infinitely often among the s_k, then x is that point, so $x \in S$. Otherwise each point occurs only finitely often, so $s_k = r_{n_k}$ (or 0) with $n_k \to \infty$ along the terms that are not 0; since $0 \leq r_{n_k} \leq \frac{1}{n_k+1} \to 0$, we get $x = 0 \in S$.

So S is a compact set of rationals.

Step 3: S escapes every K_n. For each n, $r_n \in S$ and $r_n \notin K_n$, so $S \not\subseteq K_n$. Hence no sequence of compact sets of rationals can contain every compact set of rationals. ■

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains get as far as "the K_n are given; I must defeat them". At that line the natural attempt treats the sequence as something to *outnumber* (counting) or to *improve* (patching). The model chain treats it as something to *defeat one member at a time*, with a single set that carries one witness point per K_n. The split is the sentence "for each n, pick $r_n \notin K_n$".

> *I'm being handed a countable list and asked to show it misses something. That's Cantor: I build one object that disagrees with the n-th list entry at an n-th place. Here "disagree with K_n" means "contain a point that K_n doesn't", so I need one witness $r_n \notin K_n$ for each n, and then I throw them all into one set.*

> *Throwing arbitrary witnesses together won't give a compact set. The only countable compact sets I*

> *can control are convergent sequences with their limit added. So I'll choose every witness near one rational point — say 0 — with $r_n \to 0$, and put 0 in.*

> *Can I find $r_n \notin K_n$ inside $[0, \frac{1}{n+1}]$? If not, K_n swallows every rational in that interval, and being closed it swallows the whole interval, irrationals and all. That can't happen to a set of rationals. So the witness is always there.*

Layer 3 THE WHY

Why is diagonalisation the natural move, and not a trick? Read the statement's logical shape. "There is a sequence (K_n) such that for every compact C there is an n with $C \subseteq K_n$" — the negation is: for every sequence (K_n) there is a compact C such that for every n, $C \not\subseteq K_n$. The quantifier "for every n" sits *inside* the choice of C. You are building one C that must fail each of countably many tests. The only honest way to fail countably many tests with one object is to allot each test its own piece of the object. That is all diagonalisation is. When Cantor lists the reals and changes the n-th digit of the n-th one, he is doing exactly this; here the "digit" is a point and "changing it" is choosing a point outside K_n.

Why do the witnesses have to converge? Go back to the definition the problem hands you: compact means closed and bounded. Bounded is free. Closed means every limit of points of S is in S. A countable set of rationals is closed only if all its accumulation points are rational and included. The simplest way to guarantee that is to have *exactly one* accumulation point, and to choose it yourself: pick a rational target (0 is the obvious one), force the witnesses to converge to it, and add it. So "$r_n \to 0$" is not decoration; it is what the word "closed" demands once you are building a countable set.

Why must the witness exist near 0? This is the only place the word "rational" in "compact sets of rational numbers" is used, and it has to be used somewhere, because the statement is false over $\mathbb{R}$ ($K_n = [-n, n]$). What distinguishes $\mathbb{Q}$ is that it is dense with holes everywhere: any closed set that contains all rationals of an interval contains the interval, and every interval contains irrationals. So a closed set of rationals contains no interval of rationals — it is "thin" near every point, and in particular near 0. That thinness is exactly the room the witnesses need.

The book's solution is three sentences and gives none of this. It says "for each n there is a rational $r_n \notin K_n$ with $0 \leq r_n < 1/n$" without saying why one should look for such a thing. Here is where it comes from: the negated statement asks for one set failing countably many tests, which is a diagonal construction; the word "compact" forces the diagonal points to converge; and the word "rational" is what makes a point outside K_n available in every neighbourhood of the chosen limit. Each step is read off a word in the statement.

One more check on the model: it does not matter that 0 itself might lie in some K_n, or even in all of them. S needs only one point outside each K_n, and r_n is that point. (It is also allowed that $r_n = 0$ for some n; then $0 \notin K_n$ and nothing changes.)

Layer $+\alpha$ BACKTESTING

1. The cue is "there does not exist a sequence $K_0, K_1, \ldots$ such that every X is contained in some K_n". Write down the negation with the quantifiers in order: "given any (K_n), build one C with, for each n, a point $c_n \in C \setminus K_n$."

2. Diagonalisation: one witness per member of the countable family, all packed into a single object — here a convergent sequence of rationals together with its rational limit.

3. Hesitation comes at two points: whether a witness near 0 must exist, and whether the resulting set is closed. The first is settled by "closed + contains all rationals of an interval $\Rightarrow$ contains the interval". The second by the standard fact that a convergent sequence together with its limit is compact; prove it once, carefully, and it becomes a reflex.

4. The trap is counting. Compact sets of rationals are uncountably many, which tempts you to say "countably many K_n can't cover them"; but a single K_n can contain uncountably many

compact subsets, so the cardinalities do not separate. The second trap is to spend the time patching an attempted covering family rather than asking why every patch fails.

5. Whenever a problem says "no countable family can dominate/contain/cover all X", stop comparing sizes and start diagonalising: allot the n-th member its own private defeat inside one constructed X, and then check that the constraint on X (here, closedness) is preserved — usually by making the witnesses converge.

6. Decision rule: to beat a countable list with one object, pick the n-th witness against the n-th entry, and choose the witnesses to shrink towards a fixed target so that the assembled object keeps whatever limiting property (closed, convergent, bounded) the problem requires.

▷ **Cue:** "There is no sequence $K_0, K_1, \ldots$ such that every X is contained in (dominated by) some K_n" $\longrightarrow$ **Move:** Diagonalise: for each n choose a witness x_n that defeats K_n, and assemble all the witnesses into one X

▷ **Cue:** You must assemble countably many chosen points into a closed (or compact) set $\longrightarrow$ **Move:** Choose the points to converge to a single target you are allowed to include, and add the target

▷ **Cue:** A closed set is known to consist only of rationals (or to avoid some dense set) $\longrightarrow$ **Move:** It contains no interval, so it misses points of every neighbourhood — use those missed points as witnesses

▷ **Cue:** A claimed impossibility is false over $\mathbb{R}$ but asserted over $\mathbb{Q}$ $\longrightarrow$ **Move:** Locate the one step where density-with-holes is used; that step is where the proof lives