DAWN PUTNAM SESSION

Putnam 1981 — Papers A and B

Putnam 120 Archive

Contents

Putnam 1981 — Paper A **5**
1981 A1 5
1981 A2 7
1981 A3 10
1981 A4 13
1981 A5 16
1981 A6 18

Putnam 1981 — Paper B **23**
1981 B1 23
1981 B2 25
1981 B3 28
1981 B4 31
1981 B5 33
1981 B6 35

Putnam 1981 — Paper A

1981 A1

1981 A1. Let $E(n)$ denote the largest integer k such that 5^k is an integral divisor of the product $1^1 2^2 3^3 \cdots n^n$. Calculate

$$\lim_{n\to\infty} \frac{E(n)}{n^2}.$$

Caveat. The OCR of the 1965–1984 volume drops the displayed limit expression after "Calculate" entirely (it reads only "Calculate" with nothing following). The statement above restores it from the official solution, which computes $\lim_{n\to\infty} E(n)/n^2 = 1/8$ and is internally consistent with the reconstructed problem; I have independently re-derived that limit below without relying on the OCR.

Fields: number theory (p-adic valuation), analysis (asymptotics of sums) **Position:** A1 — the opener **Answer:** $\lim_{n\to\infty} E(n)/n^2 = 1/8$. $E(n) = \sum_{i=1}^{n} i\, v_5(i)$, where $v_5(i)$ is the exponent of 5 in i; grouping by powers of 5 turns this into a geometric series in the limit.

Layer 1 THE WHAT — surface analysis

The natural attempt. You have seen "largest power of a prime dividing a product" before, and the reflex example is Legendre's formula for $v_p(n!) = \sum_{j\ge 1} \lfloor n/p^j \rfloor$. The product here, $1^1 2^2 \cdots n^n$, looks close enough to $n!$-flavoured that you reach for the same tool: write $E(n) = \sum_{j\ge 1} \lfloor n/5^j \rfloor$ and estimate $\lfloor n/5^j \rfloor \approx n/5^j$, giving $E(n) \approx n \sum_j 5^{-j} = n/4$, so $E(n)/n^2 \to 0$.

Check this against a small case before trusting it. Take $n = 5$: the product is $1^1 \cdot 2^2 \cdot 3^3 \cdot 4^4 \cdot 5^5$, and the only factor of 5 anywhere in it is 5^5 itself, contributing $v_5(5^5) = 5$ copies of the prime 5. So $E(5) = 5$. The Legendre-style guess gives $\lfloor 5/5 \rfloor = 1$. It is off by a factor of 5 — and 5 is exactly the exponent that 5 itself carries in 5^5. The guess counted *that* 5 *is a multiple of* 5; it forgot that in i^i, the factor $i = 5$ does not appear once, it appears $i = 5$ times. The natural attempt has correctly spotted the shape of the problem (a Legendre-type sum over powers of 5) but applied the wrong weight, because it borrowed the formula for $n!$ (each integer i appears exactly once) without checking that this product is different (each integer i appears i times).

The model chain. Write $v_5(i)$ for the exponent of 5 dividing i. Since i^i contributes $i \cdot v_5(i)$ copies of the prime 5,

$$E(n) = \sum_{i=1}^{n} i\, v_5(i).$$

Now use the standard indicator representation of a p-adic valuation: $v_5(i) = \sum_{j\ge 1} \mathbf{1}[5^j \mid i]$ (the number of powers of 5 that divide i). Substituting and swapping the order of summation,

$$E(n) = \sum_{i=1}^{n} i \sum_{j\ge 1} \mathbf{1}[5^j \mid i] = \sum_{j\ge 1} \sum_{\substack{1\le i\le n \\ 5^j \mid i}} i = \sum_{j\ge 1} 5^j \sum_{m=1}^{\lfloor n/5^j \rfloor} m = \sum_{j\ge 1} 5^j\, T\Big(\Big\lfloor \frac{n}{5^j} \Big\rfloor\Big),$$

where $T(m) = m(m+1)/2$ and the inner sum runs over $i = 5^j m$. Only $j \le h := \lfloor \log_5 n \rfloor$ contribute (for larger j, $\lfloor n/5^j \rfloor = 0$), so this is a finite sum of $h+1$ terms.

For each $j \le h$, write $x_j = n/5^j$ and $m_j = \lfloor x_j \rfloor$, so $x_j - 1 < m_j \le x_j$. Then $T(m_j) = m_j(m_j+1)/2 = x_j^2/2 + O(x_j)$, so

$$5^j T(m_j) = 5^j \cdot \frac{x_j^2}{2} + O(5^j x_j) = \frac{n^2}{2 \cdot 5^j} + O(n).$$

Summing over $j = 1, \dots, h$,

$$E(n) = \frac{n^2}{2} \sum_{j=1}^{h} \frac{1}{5^j} + O(nh) = \frac{n^2}{2} \sum_{j=1}^{h} \frac{1}{5^j} + O(n \log n),$$

using $h = \lfloor \log_5 n \rfloor = O(\log n)$. As $n \to \infty$, $h \to \infty$ and $\sum_{j=1}^{h} 5^{-j} \to \sum_{j=1}^{\infty} 5^{-j} = \dfrac{1/5}{1 - 1/5} = \dfrac{1}{4}$ (the tail beyond h is $O(5^{-h}) = O(1/n)$, negligible). Dividing by n^2,

$$\frac{E(n)}{n^2} = \frac{1}{2} \sum_{j=1}^{h} \frac{1}{5^j} + O\left(\frac{\log n}{n}\right) \longrightarrow \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}.$$

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths open the same way: "this is a Legendre-formula problem, sum over powers of 5." They split at the line "the product is $1^1 2^2 \cdots n^n$" being read carelessly versus read literally. The natural attempt treats the product as if it were $n!$ (each factor appearing once) and sums $\lfloor n/5^j \rfloor$; the model reads that i appears with multiplicity i, so the correct object is $\sum_i i\, v_5(i)$, not $\sum_i v_5(i)$.

> *This looks like Legendre's formula for $v_5(n!)$, but $n!$ is the product of the integers i themselves, each contributing $v_5(i)$ once. Here the factor is i^i, so i contributes $v_5(i)$ a total of i times over. I need to weight by i, not just count multiples of 5.*

> *$v_5(i)$ is itself a sum of indicators, $v_5(i) = \sum_j \mathbf{1}[5^j \mid i]$. If I substitute that into $\sum_i i\, v_5(i)$ and swap the order of summation, the inner sum over i divisible by 5^j becomes a plain arithmetic series $\sum m \cdot 5^j$ over $m = 1, \dots, \lfloor n/5^j \rfloor$ — something I can compute exactly with $T(m) = m(m+1)/2$.*

Layer 3 THE WHY

Why is the reflex to reach for Legendre's formula natural, and why is it also the trap? A p-adic valuation of a factorial-type product is a familiar enough object that pattern-matching to Legendre's formula is exactly the right instinct — the formula *is* the correct tool once you have identified what is being counted. The trap is not in choosing the tool, it is in skipping the step of checking what quantity the tool is meant to sum. Legendre's formula sums $v_p(i)$ once per integer i because $n! = \prod_{i=1}^{n} i$ gives each i exactly one factor. Here the definition of the product is $\prod_{i=1}^{n} i^i$, and $v_5(i^i) = i \cdot v_5(i)$ by the definition of v_5 as additive on products ($v_5(a^b) = b\, v_5(a)$). Reading the exponent i in i^i as data, not decoration, is what the small case $n = 5$ exposes: the factor 5 appears with multiplicity 5 inside 5^5, and no formula that treats every multiple of 5 identically can capture that.

The second structural fact is why swapping the order of summation is legitimate and not just convenient. $v_5(i) = \sum_{j \ge 1} \mathbf{1}[5^j \mid i]$ is exactly the definition of the 5-adic valuation restated as a counting function: it counts how many of the divisibility conditions $5 \mid i$, $25 \mid i$, $125 \mid i, \dots$ hold. A finite double sum $\sum_i \sum_j (\cdots)$ can always be swapped, so turning $\sum_i i \sum_j \mathbf{1}[5^j \mid i]$

into $\sum_j \sum_i (\cdots)$ costs nothing and buys everything: the inner sum, for fixed j, is now over an arithmetic progression ($i = 5^j, 2 \cdot 5^j, 3 \cdot 5^j, \ldots$) and collapses to the standard triangular-number formula. This is the general lesson behind "sum a valuation weighted by something": decompose the valuation into its defining indicators first, then swap, because the indicators are what make the inner sum an arithmetic series.

Finally, why does only the leading n^2 term survive, so that all the floor-function sloppiness washes out? Because there are only $O(\log n)$ nonzero terms in the sum over j (once $5^j > n$, the term vanishes), and each term's floor-function error is $O(n)$ rather than $O(n^2)$. An error of size $O(n \log n)$ divided by n^2 vanishes. This is the standard licence for replacing $\lfloor n/5^j \rfloor$ by $n/5^j$ in an asymptotic argument: it is not that the floor is unimportant, it is that its contribution is provably lower order, and checking that explicitly (rather than asserting it) is what turns a plausible limit into a proved one.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the exponent-on-exponent structure i^i inside a divisibility question about 5^k. Write down immediately: "$v_5(i^i) = i \cdot v_5(i)$, so $E(n) = \sum_i i\, v_5(i)$ — this is not the same sum as Legendre's formula for $n!$."

2. The method is to expand $v_5(i)$ as $\sum_j \mathbf{1}[5^j \mid i]$, swap the order of summation, and evaluate the resulting arithmetic series with $T(m) = m(m+1)/2$.

3. Hesitation occurs exactly at the reflex moment: seeing $5^k \mid$ (product) and jumping straight to $\sum \lfloor n/5^j \rfloor$ from memory. Reinforce it away by testing the recalled formula on the smallest instructive case ($n = 5$) before trusting it on the general sum.

4. The trap is applying a formula from a superficially similar problem (Legendre's formula for $n!$) without checking that the multiplicities match: $n!$ gives each integer one factor, $1^1 2^2 \cdots n^n$ gives integer i exactly i factors.

5. Whenever a problem gives $\prod_i i^{f(i)}$ and asks about v_p of the product, write down immediately that $v_p(\text{product}) = \sum_i f(i)\, v_p(i)$, and treat $f(i) = 1$ (plain factorial) as a special case, not the default.

6. Reusable rule: to sum $\sum_i w(i)\, v_p(i)$ for any weight w, expand $v_p(i) = \sum_j \mathbf{1}[p^j \mid i]$ and swap the order of summation, turning it into $\sum_j \sum_{p^j \mid i} w(i)$ — a sum of w over an arithmetic progression, which is almost always the tractable form.

▷ **Cue:** a divisibility question about v_p of a product $\prod_i i^{f(i)}$ $\longrightarrow$ **Move:** write $v_p(\text{product}) = \sum_i f(i)\, v_p(i)$; do not assume $f(i) = 1$ just because the shape resembles a factorial

▷ **Cue:** a sum $\sum_i w(i)\, v_p(i)$ $\longrightarrow$ **Move:** expand $v_p(i) = \sum_j \mathbf{1}[p^j \mid i]$ and swap summation order to get $\sum_j \sum_{p^j \mid i} w(i)$, a sum over an arithmetic progression

▷ **Cue:** an asymptotic limit built from a finite sum of $O(\log n)$ geometric-like terms, each with a floor-function error of lower order $\longrightarrow$ **Move:** replace floors by their real values, bound the total error by (number of terms)$\times$(per-term error), and check the error is o of the claimed leading order before trusting the limit

1981 A2

1981 A2. Two distinct squares of the 8 by 8 chessboard C are said to be *adjacent* if they have a vertex or side in common. Also, g is called a C-gap if for every numbering of the squares of C with all the integers $1, 2, \ldots, 64$ there exist two adjacent squares whose numbers differ by at least g. Determine the largest C-gap g.

Fields: combinatorics / graph theory (bandwidth of the king's graph), pigeonhole **Position:** A2
Answer: $g = 9$. Every numbering has an adjacent pair differing by at least 9 (king-move distance from corner to corner is 7, and a 63-unit change spread over 7 steps forces one step of at least 9); the row-major numbering achieves maximum adjacent difference exactly 9, so 9 is the largest such guaranteed gap.

Layer 1 THE WHAT — surface analysis

The natural attempt. You want a lower bound that holds for *every* numbering, so you look for an argument that does not depend on how the numbers are placed. The standard device for that is: find two squares whose numbers you control (namely 1 and 64, wherever they happen to sit), connect them by a chain of adjacent squares, and note that the numbers must change from 1 to 64 along that chain — a total change of 63 — so by pigeonhole some single step in the chain accounts for at least $63/(\text{chain length})$ of that change.

The question is how short a chain of adjacent squares can join any two squares of the board. Your first read of "adjacent" calls to mind squares sharing a full edge — the ordinary rook's-graph picture of a grid, where you move one row or one column at a time. Under that picture, going from one corner of the board to the opposite corner takes 7 steps across plus 7 steps up, i.e. 14 steps in the worst case (Manhattan distance). That gives a chain of length 14, hence a guaranteed gap of only $63/14 = 4.5$, so $g \geq 4$. This is a valid bound, and it is not wrong given that reading of "adjacent" — it is simply much weaker than the true answer, and no amount of cleverness inside the rook's-graph picture recovers the missing distance, because 14 really is the correct worst-case distance for that graph.

The model chain. Re-read the definition: two squares are adjacent if they share *a vertex or a side*. That is the king's-move graph, not the rook's-move graph: from a square at board coordinates (a, b) you may step to any (a', b') with $|a - a'| \leq 1$, $|b - b'| \leq 1$, not both zero. Under this metric, the distance between two squares (a, b) and (a', b') is $\max(|a - a'|, |b - b'|)$ (the Chebyshev distance), because a single step may change both coordinates at once. Since $1 \leq a, a' \leq 8$ and $1 \leq b, b' \leq 8$, the largest possible Chebyshev distance between any two squares of the board is 7 (from one corner to the opposite corner). So for *any* numbering, the square numbered 1 and the square numbered 64 are joined by a path of at most 7 adjacent steps.

Along that path the numbers change from 1 to 64, a total change of 63, distributed over at most 7 steps. By pigeonhole, some step has $|\text{difference}| \geq \lceil 63/7 \rceil = 9$. This holds for every numbering, so 9 is a C-gap.

To see 9 is the *largest* C-gap, exhibit a numbering with maximum adjacent difference exactly 9: number square (a, b) by $8(a - 1) + b$ (ordinary row-major order). For adjacent squares, $\Delta a, \Delta b \in \{-1, 0, 1\}$ (not both 0), and the numbering difference is $8\Delta a + \Delta b$, whose largest absolute value is $8 \cdot 1 + 1 = 9$, attained by diagonal neighbours. So this numbering has no adjacent pair differing by 10 or more, meaning 10 is not a C-gap. Hence the largest C-gap is exactly 9.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths agree on the strategy: chain squares 1 and 64 together and pigeonhole the total change of 63 over the chain length. They split at the reading of the word "adjacent" in the problem statement — specifically, at "if they have a vertex or side in common." The natural attempt's mental picture (shared side only) gives chain length 14; rereading the definition literally, including shared vertices, gives chain length 7.

> *I assumed adjacency meant sharing a side, the way I usually picture a grid graph, but the problem explicitly says "a vertex or side." That means diagonal neighbours count too, and a single step can change both coordinates at once. The distance between two squares is not the sum of the coordinate differences, it is the larger of the two.*

> *Since I can change row and column simultaneously each step, the worst-case distance across the board is* $\max(\Delta a, \Delta b)_{\max} = 7$*, not* $\Delta a_{\max} + \Delta b_{\max} = 14$*. That halves the chain length and should roughly double the guaranteed gap:* $63/7 = 9$ *instead of* $63/14 = 4.5$*.*

Layer 3 THE WHY

Why does the phrase "vertex or side in common" matter so much that it is worth pausing on? Because it silently redefines the metric on the board. A graph's distance between two vertices is the length of the shortest path connecting them, and that length depends entirely on which pairs are declared adjacent. Side-adjacency only (the rook's-graph reading) forces every step to change exactly one coordinate by exactly 1, so reaching (a', b') from (a, b) needs at least $|a - a'| + |b - b'|$ steps (Manhattan distance) — and this bound is achieved, so the diameter really is 14. Vertex-or-side adjacency (the king's-graph reading) allows a step to change both coordinates by 1 simultaneously, so the same journey needs only $\max(|a - a'|, |b - b'|)$ steps (Chebyshev distance) — and again this is achieved, with diameter 7. The problem-setter's choice of words is not stylistic; it is the entire content of the bound. Reading a graph-theory problem means first asking, explicitly, "what exactly is the adjacency relation, and what is the diameter it produces?" before reaching for pigeonhole.

The pigeonhole step itself is natural for a different, deeper reason: a numbering of a connected graph's vertices with 1 to N assigns real information to every path, because the number is a genuine function on the vertices and any path from the vertex labelled 1 to the vertex labelled N must, by the definition of a path (a sequence of adjacent vertices), realise the total change $N - 1$ as a sum of the changes along its edges. A sum of k real terms totalling 63 must contain a term of size at least $63/k$ — this is pigeonhole in its most literal form, applied not to objects in boxes but to a telescoping sum. The bound is only as strong as the shortest available path, which is precisely the graph's diameter; hence *diameter* is the single fact this entire problem turns on, and getting the adjacency relation right is getting the diameter right.

Finally, why does matching the bound with an explicit numbering matter, rather than stopping at "9 is a C-gap"? The problem asks for the *largest* C-gap, and a pigeonhole argument only ever produces a lower bound on what is guaranteed; it says nothing about whether some numbering does better. Exhibiting the row-major numbering, and checking directly that its worst adjacent difference is exactly 9 (via $8\Delta a + \Delta b$, maximised at the diagonal step), is what converts "at least 9" into "exactly 9." Every extremal combinatorics problem of this shape (find the best guaranteed value) needs both halves: a universal lower bound (works for every configuration) and a specific construction (achieves that bound), and neither half substitutes for the other.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the phrase "a vertex or side in common" defining adjacency. Write down immediately: "this is the king's-move graph, not the grid graph; distance is $\max(\Delta a, \Delta b)$, not $\Delta a + \Delta b$."

2. The method is pigeonhole on a shortest path between the squares numbered 1 and 64: total change 63 spread over (at most) the graph's diameter many steps forces one step of size at least 63/diameter.

3. Hesitation occurs at silently defaulting to the familiar grid-graph picture of "adjacent." Reinforce it away by rereading the adjacency clause word for word and drawing the 3×3 neighbourhood of a single square before doing anything else.

4. The trap is using the wrong graph metric (Manhattan instead of Chebyshev) from an unexamined mental default, which weakens the bound from 9 to 4 without any arithmetic error at all — the error is entirely in the setup, not the computation.

5. Whenever a grid problem defines adjacency, draw the neighbourhood of one interior square before doing anything else, and name the resulting distance metric (Manhattan for side-only, Chebyshev for vertex-or-side) so the diameter computation that follows is not silently wrong.

6. Reusable rule: a guaranteed-gap / bandwidth-type bound on a labelling of a graph's vertices with $1, \ldots, N$ is (label range)/(graph diameter), rounded up; compute the diameter from

the adjacency relation exactly as stated, and confirm the bound is tight with an explicit construction (a numbering achieving it) rather than trusting the lower bound alone.

- ▷ **Cue:** a problem defines "adjacent" on a grid via shared side, shared vertex, or both ⟶ **Move:** identify whether the resulting metric is Manhattan (side only, distance $= \sum |\Delta_i|$) or Chebyshev (vertex or side, distance $= \max |\Delta_i|$) before computing any diameter
- ▷ **Cue:** "for every labelling of a graph's vertices with $1, \ldots, N$, some adjacent pair differs by at least g; find the largest such g" ⟶ **Move:** $g = \lceil (N-1)/\operatorname{diam}(\text{graph}) \rceil$ by pigeonhole along a shortest path from the vertex labelled 1 to the vertex labelled N; confirm tightness with an explicit numbering
- ▷ **Cue:** a lower bound produced by pigeonhole on an extremal "largest guaranteed value" problem ⟶ **Move:** always pair it with an explicit construction achieving that bound — the lower bound alone never proves optimality

1981 A3

1981 A3. Find

$$\lim_{t\to\infty} e^{-t} \int_0^t \int_0^t \frac{e^x - e^y}{x-y}\, dx\, dy$$

or show that the limit does not exist.

Caveat. The OCR of the 1965–1984 volume renders this line as " [1'l'ex- eY] lim e-' —dx dy r oo 0 0 x-y" with the limits, the exponent t, and the integral signs all garbled together. The statement above is reconstructed from the official solution, which explicitly defines $G(t)$ as this double integral and computes $\lim_{t\to\infty} e^{-t}G(t)$; I have independently re-derived the divergence below, which matches the solution's conclusion, so I am confident in the reconstruction.

Fields: analysis (Leibniz rule for differentiating an integral, L'Hopital's rule, asymptotics of the exponential integral) **Position:** A3 **Answer:** The limit does not exist: $e^{-t}G(t) \to \infty$ as $t \to \infty$, where $G(t)$ is the double integral. The integrand $(e^x - e^y)/(x-y)$ (extended by continuity to e^x on the diagonal) forces $G(t)$ to grow strictly faster than e^t, by a factor of order $\log t$.

Layer 1 THE WHAT — surface analysis

The natural attempt. The integrand $(e^x - e^y)/(x-y)$ is a divided difference of the convex function e^x, and the Mean Value Theorem tells you immediately that it equals e^ξ for some ξ strictly between x and y — so it lies between $e^{\min(x,y)}$ and $e^{\max(x,y)}$. That gives you two computable double integrals to sandwich $G(t) = \int_0^t \int_0^t \frac{e^x - e^y}{x-y}\, dx\, dy$ between:

$$\int_0^t \int_0^t e^{\min(x,y)}\, dx\, dy \;\le\; G(t) \;\le\; \int_0^t \int_0^t e^{\max(x,y)}\, dx\, dy.$$

Both are tractable: splitting each square $[0,t]^2$ along the diagonal $x = y$ and using symmetry,

$$\int_0^t \int_0^t e^{\max(x,y)}\, dx\, dy = 2\int_0^t xe^x\, dx = 2[(t-1)e^t + 1], \qquad \int_0^t \int_0^t e^{\min(x,y)}\, dx\, dy = 2\int_0^t (t-x)e^x\, dx = 2(e^t - t - 1).$$

So $2(e^t - t - 1) \le G(t) \le 2[(t-1)e^t + 1]$, and dividing by e^t,

$$2 \;\lesssim\; e^{-t}G(t) \;\lesssim\; 2t.$$

This is real progress — it is a rigorous two-sided bound, obtained cheaply, using only the Mean Value Theorem. But look at the gap: the lower bound tends to a constant while the upper bound

grows without limit. The sandwich does not close as $t \to \infty$, so it cannot decide whether $e^{-t}G(t)$ converges, diverges, or oscillates. A pointwise bound on the integrand, however sharp-looking, has thrown away the information needed to find the exact growth rate; you need the derivative of G, not an envelope around it.

The model chain. Since $e^{-t}G(t)$ is a ∞/∞ quotient (both $G(t) \to \infty$ and $e^t \to \infty$, as the sandwich above already confirms), apply L'Hopital's rule: $\lim_{t\to\infty} \dfrac{G(t)}{e^t} = \lim_{t\to\infty} \dfrac{G'(t)}{e^t}$, provided the right-hand limit exists (finite or infinite).

Differentiate $G(t) = \int_0^t \int_0^t \frac{e^x - e^y}{x-y}\,dx\,dy$ with respect to its upper limit t, which appears in both the x- and y-ranges. By the Leibniz rule applied to each variable in turn (holding the other integral fixed, then adding the two boundary contributions),

$$G'(t) = \int_0^t \frac{e^x - e^t}{x-t}\,dx \;+\; \int_0^t \frac{e^y - e^t}{y-t}\,dy \;=\; 2\int_0^t \frac{e^x - e^t}{x-t}\,dx,$$

the two terms being equal by the manifest symmetry of the integrand under $x \leftrightarrow y$. Substitute $u = t - x > 0$, so $x = t - u$, $dx = -du$, and as x runs from 0 to t, u runs from t down to 0:

$$\frac{e^x - e^t}{x-t} = \frac{e^{t-u} - e^t}{-u} = e^t \cdot \frac{1 - e^{-u}}{u}, \qquad G'(t) = 2e^t \int_0^t \frac{1-e^{-u}}{u}\,du.$$

Hence

$$e^{-t}G'(t) = 2\int_0^t \frac{1-e^{-u}}{u}\,du.$$

The integrand $(1-e^{-u})/u$ is continuous and bounded (it tends to 1 as $u \to 0^+$), so the integral is well-behaved near 0; but for large u, $1 - e^{-u} \to 1$, so $(1-e^{-u})/u \sim 1/u$, and $\int^t 1/u\,du = \log t$ diverges. More precisely,

$$\int_0^t \frac{1-e^{-u}}{u}\,du = \int_0^1 \frac{1-e^{-u}}{u}\,du + \int_1^t \frac{1-e^{-u}}{u}\,du = C + \log t - \int_1^t \frac{e^{-u}}{u}\,du,$$

where $C = \int_0^1 (1-e^{-u})/u\,du$ is a finite constant and $\int_1^t e^{-u}/u\,du$ converges to a finite constant as $t \to \infty$ (since $e^{-u}/u \to 0$ exponentially fast). So $\int_0^t (1-e^{-u})/u\,du = \log t + O(1) \to \infty$. Therefore

$$e^{-t}G'(t) = 2\log t + O(1) \longrightarrow \infty.$$

By L'Hopital's rule (valid for an infinite limit of the derivative ratio), $\lim_{t\to\infty} e^{-t}G(t) = \lim_{t\to\infty} e^{-t}G'(t) = \infty$. **The limit does not exist** (it diverges to $+\infty$).

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths begin by trying to control the size of $G(t)$ relative to e^t. They split at the line "the sandwich bounds differ by a factor of t, and $t \to \infty$": the natural attempt keeps trying to tighten a pointwise inequality on the integrand; the model instead abandons bounding the integral itself and differentiates it exactly, converting an intractable size estimate into an exact, computable derivative.

> *My two bounds sandwich $e^{-t}G(t)$ between something tending to 2 and something tending to $2t$. That gap does not close as $t \to \infty$, so no amount of sharpening a pointwise inequality on the integrand can resolve whether the limit exists. I need the exact value or exact growth rate of $G(t)$, not an envelope around it — which means I should differentiate $G(t)/e^t$ and use L'Hopital, since both numerator and denominator go to infinity.*

> *$G(t)$ has the variable t appearing as the upper limit of both the x-integral and the y-integral, not inside a single-variable integrand. So differentiating it is a Leibniz-rule computation on each limit separately, and because the integrand is symmetric in x and y, the two resulting boundary terms must be equal — I only need to compute one of them and double it.*

> *After substituting $u = t - x$, the boundary integral becomes $e^t \int_0^t (1 - e^{-u})/u \, du$, and $(1 - e^{-u})/u$ behaves like $1/u$ for large u. An integral of $1/u$ over a growing range produces a logarithm, not a constant — so $e^{-t}G'(t)$ grows like $\log t$, and by L'Hopital, $e^{-t}G(t)$ must grow without bound too.*

Layer 3 THE WHY

Why does a pointwise sandwich fail here, structurally, rather than just by bad luck? A sandwich argument proves a limit exists only when the two bounding sequences converge to the *same* value; it is a tool for pinning down a number, not for measuring a growth rate. Here $e^{-t}G(t)$ turns out to grow like $2 \log t$ — neither a constant (so the tight lower bound of ≈ 2 is far too small) nor of order t (so the loose upper bound of $\approx 2t$ is far too large). No pointwise estimate on $(e^x - e^y)/(x - y)$, however sharp, can produce a bound of order $\log t$, because a pointwise bound throws away exactly the cancellation between the growing region (where the divided difference is close to $e^{\max}$) and the exponentially suppressed tail (where it is not) that produces the logarithm. When two honest bounds disagree in *order of growth*, that is the signal to stop bounding and start differentiating.

Why is differentiating under the integral sign, rather than bounding, the natural response to an ∞/∞ limit? L'Hopital's rule exists precisely because the ratio of two quantities that both diverge is controlled by the ratio of their instantaneous rates of change, and $G(t)$, being defined as an accumulating double integral with t as its own upper limit in two places, is exactly the kind of object whose rate of change is computable in closed form via the Fundamental Theorem of Calculus (in its two-variable, Leibniz-rule form) even when the accumulated total is not. The symmetry that halves the work — differentiating with respect to the x-limit and the y-limit gives identical contributions — is not a coincidence either: it follows from the integrand $(e^x - e^y)/(x - y)$ being invariant under swapping x and y (an odd function divided by an odd function), which is visible directly from the problem's own notation before any calculus is done.

Finally, why does the appearance of $\int (1 - e^{-u})/u \, du$ decide the whole problem, and why should a solver recognise its divergence on sight? The function $(1 - e^{-u})/u$ is a standard object (closely related to the exponential integral Ei), but the fact that matters here needs no special-function machinery: for large u, $e^{-u} \to 0$, so the integrand is asymptotically $1/u$, and $\int^t du/u = \log t \to \infty$ is the single most basic divergent integral in analysis. Recognising "integrand $\sim 1/u$ for large u" as an instant trigger for "the integral diverges like $\log t$" is the generalisable skill: it is the same comparison-test intuition used to decide convergence of $\sum 1/n$ versus $\sum 1/n^2$, applied to an integral tail instead of a series tail. Once that divergence is seen, L'Hopital converts it immediately into the divergence of the original limit, and the problem's instruction — "find the limit, or show it does not exist" — is answered by showing it diverges to $+\infty$.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is a divided difference $(e^x - e^y)/(x - y)$ of a convex function under a double integral, compared to e^t. Write down immediately: "by the Mean Value Theorem this lies between $e^{\min(x,y)}$ and $e^{\max(x,y)}$" — and separately, "t is the upper limit of the integral in two places, so $G(t)$ is differentiable via the Leibniz rule."
2. The method is L'Hopital's rule on the ∞/∞ quotient $G(t)/e^t$, computed via the Leibniz rule for differentiating a double integral with variable limits, followed by a substitution that isolates the exponential factor.

3. Hesitation occurs at the sandwich bound: it looks like real progress (two clean, correct inequalities), and it is tempting to keep refining it rather than admit it cannot close. Reinforce it away by checking the *order of growth* of both bounds explicitly (constant versus linear in t) rather than just their existence.

4. The trap is mistaking a valid but structurally too-weak method (pointwise bounding) for progress toward an exact answer; a sandwich that does not close as $t \to \infty$ is a dead end, not a step to be tightened indefinitely.

5. Whenever a limit has the form $F(t)/G(t)$ with both $F, G \to \infty$ and F or G is naturally an accumulating integral (its own upper limit is the variable), reach for L'Hopital and the Leibniz rule before reaching for pointwise bounds.

6. Reusable rule: a pointwise sandwich on an integrand can only certify a limit if the two resulting bounds converge to the *same order of growth*; if they disagree in order (constant vs. linear, here), stop bounding and differentiate the accumulating quantity exactly via L'Hopital plus the Leibniz rule.

▷ **Cue:** a divided difference $(f(x) - f(y))/(x - y)$ of a monotone function under an integral $\longrightarrow$ **Move:** apply the Mean Value Theorem to sandwich it between f' at the two endpoints of $[\min(x, y), \max(x, y)]$ — useful for a quick bound, but check the resulting sandwich actually closes before trusting it

▷ **Cue:** a limit of the form $F(t)/e^t$ (or any ∞/∞ ratio) where $F(t)$ is a double (or multiple) integral with t as a variable upper limit $\longrightarrow$ **Move:** differentiate F via the Leibniz rule (one boundary term per variable limit, using symmetry to identify equal terms) and apply L'Hopital's rule

▷ **Cue:** an integral $\int^t g(u)\,du$ where $g(u) \sim 1/u$ for large u $\longrightarrow$ **Move:** recognise the divergence $\int^t du/u = \log t \to \infty$ immediately, by the same comparison-test reflex used for the harmonic series

1981 A4

1981 A4. A point P moves inside a unit square in a straight line at unit speed. When it meets a corner it escapes. When it meets an edge its line of motion is reflected so that the angle of incidence equals the angle of reflection.

Let $N(T)$ be the number of starting directions from a fixed interior point P_0 for which P escapes within T units of time. Find the least constant a for which constants b and c exist such that

$$N(T) \leq aT^2 + bT + c$$

for all $T > 0$ and all initial points P_0.

Fields: geometry (billiards, unfolding), lattice point counting **Position:** A4 — mid-to-late paper; one idea, then a Gauss-circle estimate **Answer:** $a = \pi$. Unfolding the square turns escape paths into straight rays from P_0 to lattice points at distance at most T, so $N(T)$ is at most the number of lattice points in a disc of radius T, which is $\pi T^2 + O(T)$; for a point P_0 with an irrational coordinate the two counts are equal, so no $a < \pi$ works.

Layer 1 THE WHAT — surface analysis

The natural attempt. You draw the square $[0, 1]^2$, put P_0 inside, and follow a path. It hits the right edge, reflects, hits the top edge, reflects, and after three or four bounces the picture is a thicket of segments. You try to parametrise: start at $P_0 = (x_0, y_0)$ with direction $(\cos\theta, \sin\theta)$; the first wall hit depends on which of $(1 - x_0)/\cos\theta$ and $(1 - y_0)/\sin\theta$ is smaller; after the reflection one component of the direction changes sign and you recompute from the new point.

After k bounces the position is a case-split over 2^k sign patterns, and the condition "the path passes through a corner before time T" is a union of conditions over all these branches.

You can get somewhere with small cases: the four directions aimed straight at the corners escape within $\sqrt{2}$; directions aimed at the corners after one bounce ("aim at the mirror image of the corner") escape a bit later. That phrase — the mirror image of the corner — is the whole solution, but in the natural attempt it appears as a trick for one bounce, not as a principle for all of them. Counting how many directions escape by time T by chasing bounces is hopeless: the number of relevant bounce sequences grows with T, and you are asked for an asymptotic constant.

The model chain. *Unfolding.* Let $\varphi : \mathbb{R} \to [0,1]$ be the distance to the nearest even integer, so φ is the zigzag with $\varphi(2k) = 0$, $\varphi(2k+1) = 1$, linear with slope ± 1 in between. Put $\Phi(x,y) = (\varphi(x), \varphi(y))$, which folds the plane onto the square. Take a starting direction $u = (u_1, u_2)$, $|u| = 1$, and the straight ray $\ell(t) = P_0 + tu$.

Consider $\Phi(\ell(t))$. Its x-coordinate $\varphi(x_0 + tu_1)$ moves at rate $\pm u_1$, and the sign flips exactly when $x_0 + tu_1$ crosses an integer, which is exactly when the folded point reaches a vertical edge $x \in \{0,1\}$. Flipping the horizontal component of velocity at a vertical wall, keeping the vertical component, is the law of reflection. The same holds for y and horizontal walls. The speed stays $|u| = 1$. So $\Phi(\ell(t))$ is the billiard path, and it is at a corner precisely when both coordinates of $\ell(t)$ are integers, that is, when $\ell(t) \in \mathbb{Z}^2$. Hence:

$$P \text{ escapes within time } T \iff \text{the ray from } P_0 \text{ in direction } u \text{ meets a lattice point at distance} \le T.$$

Upper bound. Each escaping direction is determined by a lattice point at distance $\le T$ from P_0 (the first one on the ray), and different directions give different lattice points. So $N(T) \le L(T)$, the number of lattice points in the closed disc D of radius T about P_0. Place a unit square centred at each such lattice point. These squares have disjoint interiors, and each lies in the disc of radius $T + \frac{\sqrt{2}}{2}$ about P_0 (half the diagonal is $\frac{\sqrt{2}}{2}$). Comparing areas,

$$N(T) \le L(T) \le \pi\Big(T + \tfrac{\sqrt{2}}{2}\Big)^2 = \pi T^2 + \sqrt{2}\,\pi T + \tfrac{\pi}{2}$$

for every P_0. So $a = \pi$ works, with $b = \sqrt{2}\,\pi$, $c = \pi/2$.

Lower bound. Take $P_0 = (1/\sqrt{2},\, 1/2)$. If two lattice points $z \neq z'$ were on one ray from P_0, then P_0 would be on the line through them. That line is not horizontal (a horizontal line through lattice points has integer y, and $1/2 \notin \mathbb{Z}$), so it can be written $x = m + (y-n)\,\frac{m'-m}{n'-n}$ with integers m, n, m', n'; at $y = \frac{1}{2}$ this gives rational x, contradicting $1/\sqrt{2} \notin \mathbb{Q}$. So distinct lattice points give distinct directions and $N(T) = L(T)$. For $T \ge \frac{\sqrt{2}}{2}$, every point w of the disc of radius $T - \frac{\sqrt{2}}{2}$ about P_0 lies in the unit square centred at some lattice point z with $|w - z| \le \frac{\sqrt{2}}{2}$, hence $|z - P_0| \le T$. So those squares cover the smaller disc, and

$$N(T) = L(T) \ge \pi\Big(T - \tfrac{\sqrt{2}}{2}\Big)^2 = \pi T^2 - \sqrt{2}\,\pi T + \tfrac{\pi}{2}.$$

If $a < \pi$, then for any b, c the difference $\pi T^2 - \sqrt{2}\pi T + \frac{\pi}{2} - (aT^2 + bT + c)$ tends to $+\infty$, so the bound fails for large T at this P_0. The least constant is $a = \pi$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start by following the first reflection. At the first bounce the natural attempt updates the direction vector; the model keeps the direction and reflects the *square* instead. The split is the line "instead of bending the path, reflect the room".

> *A mirror does not bend the light for the person looking into it — they see a straight line into a reflected room. So I will not reflect the path; I will reflect the square across the wall and let the path go straight on. Do it at every wall and the square tiles the plane, the path is a straight ray, and the corners become the lattice points.*

> *Escape directions correspond to lattice points within distance T. That is a circle-counting problem, and the answer to "how many lattice points in a disc of radius T" is πT^2 plus a boundary error. The constant is π unless something forces directions to coincide.*

> *The inequality has to hold for* all P_0*, so the least a is set by the worst P_0. At the centre many lattice points line up behind each other and share a direction; I want a P_0 where that never happens — one with an irrational coordinate.*

Layer 3 THE WHY

Why is unfolding not a trick? Because of what reflection is. A reflection in a line is an isometry: it preserves lengths and angles. The law "angle of incidence equals angle of reflection" says precisely that the outgoing segment is the mirror image, in the wall, of the segment the particle would have travelled had the wall not been there. So the true path and the straight continuation differ by one reflection of the plane after each bounce. Apply those reflections to the square instead of to the path, and the path is straight by construction. The zigzag function φ is the bookkeeping of this: it is the composite of all the reflections in the lines $x = k$.

Once the path is straight, the question "which directions escape by time T" changes category. It was dynamics; it becomes counting. The corners of all the reflected copies of $[0,1]^2$ are exactly the points of $\mathbb{Z}^2$, because reflecting in $x = k$ sends integers to integers. So you are counting lattice points visible from P_0 in a disc. The count of lattice points in a large region is its area to first order — each lattice point "owns" one unit square, the squares tile the plane, and only the squares straddling the boundary are miscounted. There are $O(T)$ of those, which is why b and c exist and the problem asks only about a.

The MAA book's solution says "using the reflection properties, one can see that P escapes within T units of time if and only if the (infinite) ray from P_0 … goes through a lattice point within T units of distance". That sentence is the whole problem, and it deserves the paragraph above: the reflection law is the definition of an isometric unfolding.

The last point is the quantifier. "For all $T > 0$ and all initial points P_0" means a must work against the worst P_0, so a lower bound needs one bad P_0, and the bad case is where $N(T)$ is as large as the upper bound allows — where no two lattice points hide behind each other. Collinearity of P_0 with two lattice points is an algebraic coincidence (a rational linear relation), and an irrational coordinate kills it. If you test with the centre $(\frac{1}{2}, \frac{1}{2})$, you find far fewer directions (lattice points pair up along rays), and you might conclude $a < \pi$. That is the trap.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "reflected so that the angle of incidence equals the angle of reflection" inside a polygon that tiles the plane by reflections. Write down: unfold; path becomes a straight ray; corners become $\mathbb{Z}^2$.
2. Unfolding, followed by lattice-point counting by area: each lattice point owns a unit square, compare with discs of radius $T \pm \frac{\sqrt{2}}{2}$.
3. Hesitation comes at the quantifier: do I need the upper bound for every P_0 and the lower bound for some P_0? Yes; write the two halves separately. A second hesitation is whether several lattice points on one ray matter; they only lower $N(T)$, so the upper bound is safe, and an irrational coordinate removes them for the lower bound.

4. Two traps: chasing bounces (zombie maths with 2^k cases), and testing a symmetric P_0 such as the centre, where collinear lattice points reduce the count and suggest a smaller constant.

5. Any problem with straight-line motion and mirror reflections in a region that tiles the plane by reflections (rectangle, equilateral triangle, 45° and 30–60–90 triangles) is a straight-line problem in the unfolded tiling. Any "how many up to size T" question about lattice points has main term equal to an area.

6. Reusable rule: when a path is bent by reflections, reflect the space and keep the path straight. When a count of lattice points is asked for asymptotically, sandwich it between areas of the region shrunk and grown by the half-diagonal of the fundamental cell. When a bound must hold for all parameters, find the worst parameter by asking what coincidence reduces the count and choosing a parameter that avoids it.

▷ **Cue:** a particle reflecting off the walls of a rectangle (angle of incidence equals angle of reflection) $\longrightarrow$ **Move:** reflect the rectangle across walls to tile the plane; the path becomes a straight line; corners become lattice points

▷ **Cue:** counting lattice points in a disc or other region of size T $\longrightarrow$ **Move:** give each lattice point its unit square; compare areas of the region shrunk and enlarged by $\frac{\sqrt{2}}{2}$

▷ **Cue:** a bound required "for all initial points" $\longrightarrow$ **Move:** the least constant is set by the worst initial point; find it by avoiding coincidences that lower the count

▷ **Cue:** two lattice points possibly collinear with a given point $\longrightarrow$ **Move:** choose the point with an irrational coordinate; collinearity would force a rational relation

▷ **Cue:** asked for the least leading constant a in $aT^2 + bT + c$ $\longrightarrow$ **Move:** only the area term matters; prove a matching lower bound of the form $aT^2 - O(T)$ for one instance

1981 A5

1981 A5. Let $P(x)$ be a polynomial with real coefficients and form the polynomial

$$Q(x) = (x^2+1)P(x)P'(x) + x\big([P(x)]^2 + [P'(x)]^2\big).$$

Given that the equation $P(x) = 0$ has n distinct real roots exceeding 1, prove or disprove that the equation $Q(x) = 0$ has at least $2n-1$ distinct real roots.

Fields: algebra (polynomials), real analysis (Rolle's theorem) **Position:** A5 — late paper, but short once Q is factorised **Answer:** True. $Q = (P' + xP)(xP' + P)$; Rolle's theorem gives $n-1$ roots of the first factor and n roots of the second, and a common root $r > 1$ would force $(r^2-1)P(r) = 0$, which is impossible between consecutive roots of P. The bound is attained by $P(x) = x - 2$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You look at Q as given and try to count its roots directly. At a root a of P, $Q(a) = aP'(a)^2 \geq 0$ (as $a > 1$). So you try an intermediate value argument: find points where $Q < 0$ between consecutive roots $a_i < a_{i+1}$ of P. Near a_i, $P(x) \approx P'(a_i)(x - a_i)$, and $Q \approx aP'(a_i)^2 + \ldots$ — positive on both sides. At the critical point of P between a_i and a_{i+1}, $P' = 0$ and $Q = xP^2 > 0$. Positive again. Every natural test point gives $Q > 0$, and you can't find the sign changes you need.

The reason is that roots of Q may be where Q touches zero without changing sign, or its sign changes happen at points you have no handle on. Staring at Q as a quadratic form in (P, P'),

$$Q = x\,P^2 + (x^2+1)\,PP' + x\,P'^2,$$

you might compute its discriminant: $(x^2+1)^2 - 4x^2 = (x^2-1)^2 \geq 0$. It factorises. If you get here you are on the model path; if you don't, you are stuck with a polynomial of degree $2\deg P$ and nothing to grip.

The model chain. *Factorise.* As a quadratic form in P, P' the discriminant is $(x^2-1)^2$, a perfect square, and indeed

$$(P' + xP)(xP' + P) = xP'^2 + PP' + x^2PP' + xP^2 = Q.$$

Write $F = P' + xP$ and $G = xP' + P$. Each is a derivative in disguise:

$$F(x) = e^{-x^2/2}\,\frac{d}{dx}\big(e^{x^2/2}P(x)\big), \qquad G(x) = \frac{d}{dx}\big(xP(x)\big).$$

If $P \equiv 0$ then $Q \equiv 0$ has infinitely many roots and the claim holds, so assume $P \not\equiv 0$. Then P has finitely many roots; let $1 < a_1 < \cdots < a_m$ be *all* of its roots exceeding 1, with $m \geq n$.

Roots of F. $e^{x^2/2}P(x)$ vanishes at each a_i. By Rolle's theorem its derivative vanishes at some $b_i \in (a_i, a_{i+1})$, $1 \leq i \leq m-1$, and since $e^{-x^2/2} \neq 0$, $F(b_i) = 0$.

Roots of G. $xP(x)$ vanishes at $0, a_1, \ldots, a_m$. By Rolle, G vanishes at some $c_1 \in (0, a_1)$ and at $c_{i+1} \in (a_i, a_{i+1})$, $1 \leq i \leq m-1$.

The b_i lie in disjoint open intervals, so they are distinct; likewise the c_j. Also $c_1 < a_1 < b_i$ for all i. So the only possible coincidence is $b_i = c_{i+1} =: r$ for some i, with $a_i < r < a_{i+1}$. Then

$$P'(r) + rP(r) = 0, \qquad rP'(r) + P(r) = 0.$$

Multiply the first by r and subtract the second: $(r^2-1)P(r) = 0$. Since $r > a_i > 1$, $r^2 - 1 \neq 0$, so $P(r) = 0$. But r lies strictly between consecutive roots of P exceeding 1, and $r > 1$: contradiction. So the b_i and c_j are $2m - 1 \geq 2n - 1$ distinct roots of $Q = FG$.

Sharpness. For $P(x) = x - 2$ ($n = 1$): $F = 1 + x^2 - 2x = (x-1)^2$, $G = 2x - 2$, so $Q = 2(x-1)^3$ has exactly one real root. The bound $2n - 1$ cannot be improved in general.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths look at Q. The natural attempt evaluates it at test points; the model reads it as a quadratic form in P and P' and factorises. The split is the line $Q = (P' + xP)(xP' + P)$.

> *Q is symmetric in P and P' with coefficients x, $x^2 + 1$, x. Outer coefficients x and x, middle coefficient $x \cdot x + 1 \cdot 1$ — that is the shape of $(xA + B)(A + xB)$. Let me try to factor it before I try to count its roots.*

> *Rolle's theorem finds roots of derivatives. $P' + xP$ and $xP' + P$ are not derivatives, but each is one derivative multiplied by a nonvanishing function. $xP' + P = (xP)'$ by the product rule; for $P' + xP$ I need the integrating factor $e^{x^2/2}$.*

> *The two families of roots might coincide. If F and G share a root r, I have two linear equations in $P(r)$, $P'(r)$; their determinant is $1 - r^2$. So "exceeding 1" is exactly what stops that — that is why the problem says it.*

Layer 3 THE WHY

Why factorise first? Because counting roots of a product is counting roots of each factor, and a polynomial of degree $2 \deg P$ given as a messy sum offers nothing to count with. The problem gives Q as an expression in P and P' only, homogeneous of degree 2. Any homogeneous quadratic $\alpha u^2 + \beta uv + \gamma v^2$ splits into linear factors over the reals iff $\beta^2 - 4\alpha\gamma \geq 0$, and over polynomial coefficients it splits nicely iff that discriminant is a perfect square. Here it is $(x^2-1)^2$. The problem-setter chose $x^2 + 1$ as the middle coefficient because $x^2 + 1 = x \cdot x + 1 \cdot 1$, the cross term of $(xA + B)(A + xB)$.

Why does Rolle's theorem enter? The hypothesis is about roots of P; the conclusion is about roots of something built from P and P'. The one bridge between roots of a function and roots of its derivative is Rolle. So the factors must be turned into derivatives. $(xP)' = xP' + P$ is the product rule read backwards. For $P' + xP$, you want $(hP)' = h(P' + \frac{h'}{h}P)$ with $h'/h = x$, which forces $h = e^{x^2/2}$ — the integrating factor from first-order linear differential equations. Nothing is being pulled from a hat: "$P' + g(x)P$" always means "multiply by $e^{\int g}$".

Why "exceeding 1"? A student who only runs Rolle gets $n-1$ plus n roots and writes $2n-1$ without checking distinctness. That is the rigour gap, and the problem is built around it. A root shared by F and G satisfies the linear system

$$\begin{pmatrix} r & 1 \\ 1 & r \end{pmatrix} \begin{pmatrix} P(r) \\ P'(r) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

whose determinant is $r^2 - 1$. For $r > 1$ the only solution is $P(r) = P'(r) = 0$, and r is not a root of P because it sits strictly between consecutive roots. The sharpness example shows what happens without the condition: for $P = x - 2$ the Rolle roots of F and G both land at $x = 1$, where the determinant vanishes. The hypothesis is not decoration; it is the determinant.

A last point of rigour: the MAA solution says "we can assume P has exactly n zeros exceeding 1". What is used is that the a_i are *consecutive* roots, so that no root of P lies strictly between them; taking all roots exceeding 1 gives this and yields $2m - 1 \geq 2n - 1$.

Layer $+\alpha$ BACKTESTING

1. The first cue is that Q is a quadratic form in P and P' with coefficients x, $x^2 + 1$, x. Write down the discriminant $(x^2 - 1)^2$ and the factorisation $(P' + xP)(xP' + P)$.

2. Rolle's theorem applied to $e^{x^2/2}P$ and to xP, then a distinctness check.

3. Hesitation occurs at distinctness: the two lists of roots interleave the same intervals. Resolve it by writing the two root conditions as a 2×2 linear system and reading off its determinant.

4. The trap is to stop at "$n-1$ plus n" without checking that the roots are different. The other trap is sign-testing Q at convenient points, which all give $Q > 0$.

5. Any expression that is a quadratic form in f and f': compute the discriminant and factorise. Any $f' + gf$: it is $e^{-\int g}(e^{\int g} f)'$. Any "how many distinct roots" after Rolle: check overlaps with a determinant.

6. Reusable rule: to transfer root counts from f to an expression in f and f', rewrite each factor as (nonvanishing function) $\times$ (derivative of something that vanishes where f does), and apply Rolle between consecutive zeros. Then use any unexplained hypothesis to rule out coincidences — it is usually there for exactly that.

▷ **Cue:** an expression homogeneous and quadratic in P and P' $\longrightarrow$ **Move:** compute its discriminant as a quadratic form; factorise into linear factors in P, P'

▷ **Cue:** a linear combination $P' + g(x)P$ $\longrightarrow$ **Move:** write it as $e^{-G}(e^G P)'$ with $G' = g$ and apply Rolle to $e^G P$

▷ **Cue:** $xP' + P$ or $xP' + kP$ $\longrightarrow$ **Move:** recognise $(xP)'$, or $x^{1-k}(x^k P)'$

▷ **Cue:** roots from two Rolle arguments in the same intervals $\longrightarrow$ **Move:** a common root solves a 2×2 linear system in $P(r), P'(r)$; its determinant decides

▷ **Cue:** a hypothesis like "roots exceeding 1" that the main argument never used $\longrightarrow$ **Move:** find the step where it rules out a degenerate coincidence

1981 A6

1981 A6. Suppose that each of the vertices of $\triangle ABC$ is a lattice point in the (x, y)-plane and that there is exactly one lattice point P in the interior of the triangle. The line AP is extended to meet BC at E. Determine the largest possible value for the ratio of lengths of

segments

$$\frac{|AP|}{|PE|}.$$

[A lattice point is a point whose coordinates x and y are integers.]

Fields: geometry (lattice points), affine/barycentric coordinates **Position:** A6 — last problem of the morning; hard to find, short once the right points are manufactured **Answer:** 5. In barycentric coordinates $P = \alpha A + \beta B + \gamma C$, the ratio is $(1-\alpha)/\alpha$; the lattice points $2P - B$, $2P - C$ and $3P - B - C$ cannot be interior, which gives $\beta, \gamma \le \frac{1}{2}$ and $\min(\beta, \gamma) \le \frac{1}{3}$, so $\alpha \ge \frac{1}{6}$. Equality at $A = (0,0)$, $B = (0,2)$, $C = (3,0)$, $P = (1,1)$.

Layer 1 THE WHAT — surface analysis

The natural attempt. You experiment. Put A at the origin and try right-angled triangles with legs on the axes. $B = (2,0)$, $C = (0,2)$: an interior lattice point needs $x, y \ge 1$ and $x + y < 2$, so there is none. Enlarge to $B = (3,0)$, $C = (0,3)$: only $(1,1)$; $E = (\frac{3}{2}, \frac{3}{2})$ and the ratio is $\sqrt{2}/\frac{\sqrt{2}}{2} = 2$. Now $B = (3,0)$, $C = (0,2)$: interior points satisfy $x, y \ge 1$ and $\frac{x}{3} + \frac{y}{2} < 1$, so only $(1,1)$; E is on $y = x$ with $\frac{x}{3} + \frac{x}{2} = 1$, so $E = (\frac{6}{5}, \frac{6}{5})$ and the ratio is $1/\frac{1}{5} = 5$. Then $B = (4,0)$, $C = (0,2)$: again only $(1,1)$ (since $(2,1)$ gives $\frac{2}{4} + \frac{1}{2} = 1$, on the boundary), $E = (\frac{4}{3}, \frac{4}{3})$, ratio 3. Moving A to other vertices of these triangles gives smaller values in your trials.

So you conjecture 5. To prove it you reach for Pick's theorem: area $= I + \frac{B_\partial}{2} - 1 = \frac{B_\partial}{2}$ with $I = 1$. That bounds nothing directly, because the number B_∂ of boundary points is unbounded a priori (a long thin triangle has many), and you need to show that long thin triangles with one interior point either do not exist or give small ratios. The classification of lattice triangles with one interior point up to unimodular maps is finite, but proving that classification is a project, not an exam answer. The natural attempt finds the answer and cannot prove it.

The model chain. *Coordinates.* The triangle is nondegenerate (it has an interior point), so every point X of the plane has unique barycentric coordinates: $X = \lambda A + \mu B + \nu C$ with $\lambda + \mu + \nu = 1$, and X is in the interior iff $\lambda, \mu, \nu > 0$. Write $P = \alpha A + \beta B + \gamma C$ with $\alpha, \beta, \gamma > 0$.

The ratio. Put $E' = \dfrac{\beta B + \gamma C}{\beta + \gamma}$, a point of segment BC. Then $P = \alpha A + (1-\alpha)E'$, so E' lies on line AP and on BC: $E' = E$. Also $|AP| = (1-\alpha)|AE|$ and $|PE| = \alpha|AE|$, so

$$\frac{|AP|}{|PE|} = \frac{1-\alpha}{\alpha} = \frac{1}{\alpha} - 1.$$

Maximising the ratio means minimising α; the claim is $\alpha \ge \frac{1}{6}$.

Manufacturing lattice points. If $X_1, \dots, X_k$ are lattice points and $n_1, \dots, n_k$ are integers with $\sum n_i = 1$, then $\sum n_i X_i$ is a lattice point. So

$$2P - B = 2\alpha A + (2\beta - 1)B + 2\gamma C, \qquad 3P - B - C = 3\alpha A + (3\beta - 1)B + (3\gamma - 1)C$$

are lattice points, and the coefficients shown are their barycentric coordinates (each triple sums to 1). Neither equals P: $2P - B = P$ gives $P = B$, and $3P - B - C = P$ gives $P = \frac{B+C}{2}$, both on the boundary. Since P is the only interior lattice point, neither is interior, so each has a barycentric coordinate ≤ 0.

For $2P - B$: $2\alpha, 2\gamma > 0$, so $2\beta - 1 \le 0$, i.e. $\beta \le \frac{1}{2}$. By the same argument with $2P - C$, $\gamma \le \frac{1}{2}$.

For $3P - B - C$: $3\alpha > 0$, so $3\beta - 1 \le 0$ or $3\gamma - 1 \le 0$, i.e. $\min(\beta, \gamma) \le \frac{1}{3}$.

Hence $\beta + \gamma \le \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$ and $\alpha = 1 - \beta - \gamma \ge \frac{1}{6}$, so $|AP|/|PE| \le 5$.

Equality. $A = (0,0)$, $B = (0,2)$, $C = (3,0)$. A lattice point (x,y) is interior iff $x, y \ge 1$ and $\frac{x}{3} + \frac{y}{2} < 1$; only $(1,1)$ qualifies ($\frac{1}{3} + \frac{1}{2} = \frac{5}{6}$, while $(2,1)$ and $(1,2)$ fail). Here $(1,1) = \frac{1}{6}A + \frac{1}{2}B + \frac{1}{3}C$, so $\alpha = \frac{1}{6}$ and the ratio is 5. (Directly: $E = (\frac{6}{5}, \frac{6}{5})$, $|AP|/|PE| = 1/\frac{1}{5} = 5$.) The largest possible value is 5.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths reach the conjecture 5 by experiment. The natural attempt then tries to control the triangle's shape (Pick, classification); the model fixes the triangle and asks what the uniqueness of P forbids, by producing new lattice points from P and the vertices. The split is the line "$2P - B$ is a lattice point".

> *The only hypothesis with teeth is "exactly one lattice point inside". The only way to use "no other lattice point" is to exhibit another lattice point and conclude it must be outside. What lattice points can I build from A, B, C, P? Integer combinations with coefficients summing to* 1*: $2P - B$, $2P - A$, $3P - B - C$...*

> *The ratio $|AP|/|PE|$ is affine — it survives any linear map. So I should use coordinates that are affine too: write $P = \alpha A + \beta B + \gamma C$. Then the ratio is $(1 - \alpha)/\alpha$, and each manufactured lattice point being outside is a linear inequality on α, β, γ.*

> *$2P - B$ gives $\beta \leq \frac{1}{2}$ and $2P - C$ gives $\gamma \leq \frac{1}{2}$. That only yields $\alpha \geq 0$. I need a point whose coordinates involve β and γ together: subtract both B and C. For the coefficients to sum to* 1 *I need $kP - B - C$ with $k = 3$.*

Layer 3 THE WHY

Why barycentric coordinates? Look at what the problem measures and what it assumes. The conclusion is a ratio of lengths along one line, which is preserved by every affine map. The hypothesis — lattice points, exactly one inside — is preserved by every affine map sending $\mathbb{Z}^2$ to itself (integer matrices of determinant ± 1, plus integer translations). So the problem lives in affine geometry, and the natural coordinates are affine ones attached to the triangle: barycentrics. In them, "interior" is "all coordinates positive" and the ratio is $(1 - \alpha)/\alpha$. Cartesian coordinates carry lengths and angles that the problem does not care about, and that is why the experiments felt shapeless.

Why integer combinations with coefficients summing to 1? Because that is the precise set of operations that preserve lattice points without depending on where the origin is. $2P - B$ is "reflect B through P": the point on the far side of P from B at the same distance. If P is too close to A — which is what a large ratio means — then P is far from the side BC and these reflected points land inside the triangle. The uniqueness hypothesis forbids that. So the bound is a statement of the form "if P were further from BC, a second interior lattice point would exist", and the proof is to write that second point down.

The MAA solution sets $Q = 2P - B$ and $R = 3P - B - C$ and argues with images of medial triangles under the maps $X \mapsto 2X - B$ and $X \mapsto 3X - B - C$. That is the same argument dressed geometrically. Here is where $3P - B - C$ comes from: the target is a lower bound on $\alpha = 1 - \beta - \gamma$, so you need an upper bound on $\beta + \gamma$, so you need a lattice point whose B- and C-coordinates both decrease. $kP - B - C$ is the simplest, and "coefficients sum to 1" forces $k = 3$. Nothing else is chosen. The inequalities $\beta, \gamma \leq \frac{1}{2}$ and $\min(\beta, \gamma) \leq \frac{1}{3}$ then give $\beta + \gamma \leq \frac{5}{6}$ at once, and the extremal triangle has $(\beta, \gamma) = (\frac{1}{2}, \frac{1}{3})$, which is exactly the example.

Why is this an A6? Because the experiment gives you the answer and nothing about the proof, and the proof needs you to stop looking at triangles and start looking at what the one interior point forbids. That reversal is hard to find under time pressure. It is not a matter of talent; it is a matter of asking "how do I use this hypothesis?" literally.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is "exactly one lattice point P in the interior". Write down: every lattice point built from A, B, C, P other than P is not interior.

2. Barycentric coordinates plus integer affine combinations: $2P - B$, $2P - C$, $3P - B - C$, each giving a linear inequality.
3. Hesitation arises after $2P - B$ and $2P - C$ give only $\alpha \geq 0$. Reinforce with the target: I want a lower bound on α, so I need upper bounds on $\beta + \gamma$ jointly, so I subtract both B and C.
4. The trap is Pick's theorem and shape classification: they point at the right objects but lead to a long case analysis. A second trap is forgetting that the manufactured point might coincide with P or lie on the boundary; check both.
5. Any problem where lattice polygons contain "no" or "exactly one" interior lattice point: build new lattice points by integer affine combinations of the given ones (reflections $2P - X$, sums $X + Y - Z$) and use that they are not interior.
6. Reusable rule: to use a hypothesis of the form "there is no other object of type T", manufacture objects of type T from the ones you have and read off constraints. When the conclusion is affine-invariant, work in affine (barycentric) coordinates so that each constraint is linear.

▷ **Cue:** a lattice triangle with exactly one (or no) interior lattice point $\longrightarrow$ **Move:** form $2P - B$, $2P - C$, $3P - B - C$ (integer coefficients summing to 1); they are lattice points and cannot be interior

▷ **Cue:** a ratio of lengths along a line through a vertex and an interior point $\longrightarrow$ **Move:** use barycentric coordinates; $|AP|/|PE| = (1 - \alpha)/\alpha$

▷ **Cue:** a hypothesis "there is no other point of this kind" $\longrightarrow$ **Move:** construct another point of that kind and conclude where it cannot be

▷ **Cue:** you need a bound on a sum of two coordinates, and single-point constructions bound each separately $\longrightarrow$ **Move:** combine: subtract both vertices at once, $kP - B - C$ with k fixed by the sum-to-one condition

▷ **Cue:** a small example attains the conjectured extremal value $\longrightarrow$ **Move:** read its barycentric coordinates; they identify which inequalities are tight

Putnam 1981 — Paper B

1981 B1

1981 B1. Find

$$\lim_{n\to\infty} \frac{1}{n^5} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(5i^4 - 18i^2j^2 + 5j^4\right).$$

Caveat. The source scan (MAA, *Putnam 1965–1984*) prints only the word "Find" for this problem; the display equation itself did not survive the OCR pass. The statement above is reconstructed from the printed official solution, which works with $S_2(n) = \sum_{k=1}^{n} k^2$, $S_4(n) = \sum_{k=1}^{n} k^4$ and the combination $10nS_4(n) - 18S_2(n)^2$, and asserts this equals $-n^5 + \cdots$. Writing $10nS_4(n) = 5\sum_{i,j} i^4 + 5\sum_{i,j} j^4$ (by relabelling) and $18S_2(n)^2 = 18\sum_{i,j} i^2j^2$ recovers exactly the double sum above, and I have independently expanded $S_2(n) = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}$ and $S_4(n) = \frac{n^5}{5} + \frac{n^4}{2} + \frac{n^3}{3} - \frac{n}{30}$ and confirmed $10nS_4(n) - 18S_2(n)^2 = -n^5 - \frac{19}{6}n^4 - 3n^3 - \frac{5}{6}n^2$ by hand, matching the printed leading term and the printed answer -1. The reconstruction is therefore checked against the mathematics, not merely copied.

Fields: analysis (finite sums, asymptotics of power sums) **Position:** B1 – opening problem of the second session **Answer:** The limit is -1.

Layer 1 THE WHAT — surface analysis

The natural attempt. The expression is symmetric in i, j and quartic, so your first instinct is to treat the double sum as a scaled Riemann sum. Set $i = nx$, $j = ny$ with $x, y \in (0, 1]$ on a grid of spacing $1/n$; then

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(5i^4 - 18i^2j^2 + 5j^4) \approx n^2 \int_0^1 \int_0^1 \left(5(nx)^4 - 18(nx)^2(ny)^2 + 5(ny)^4\right) dx\, dy,$$

which factors as $n^6 \int_0^1 \int_0^1 (5x^4 - 18x^2y^2 + 5y^4)\, dx\, dy$. You compute the integral: $\int_0^1 5x^4\, dx = 1$ against $\int_0^1 dy = 1$, contributing 1; the y^4 term contributes 1 by symmetry; the cross term is $18 \int_0^1 x^2\, dx \int_0^1 y^2\, dy = 18 \cdot \frac{1}{3} \cdot \frac{1}{3} = 2$. So the bracket is $1 + 1 - 2 = 0$. The leading n^6 coefficient of the continuum approximation is zero — which matches the fact that the true sum has no n^6 term at all — but the Riemann sum then tells you nothing about the n^5 term you actually need, because dividing by n^5 and taking $n \to \infty$ of an approximation that is only accurate to within its own $O(n^5)$ error term is circular. You have proved the answer is not ∞; you have not found it.

The model chain. Do not approximate the sums; compute them exactly. The two power-sum formulas (provable by induction or by the standard telescoping trick $\sum\left[(k+1)^{m+1} - k^{m+1}\right] = (n+1)^{m+1} - 1$) are

$$S_2(n) = \sum_{k=1}^{n} k^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}, \qquad S_4(n) = \sum_{k=1}^{n} k^4 = \frac{n^5}{5} + \frac{n^4}{2} + \frac{n^3}{3} - \frac{n}{30}.$$

Now rewrite the target sum in terms of these. Since j ranges over $1, \ldots, n$ independently of i, $\sum_{i,j} i^4 = n\, S_4(n)$, and likewise $\sum_{i,j} j^4 = n\, S_4(n)$; and $\sum_{i,j} i^2 j^2 = S_2(n)^2$. Hence the double sum equals $10nS_4(n) - 18S_2(n)^2$. Substituting the two formulas and expanding:

$$10nS_4(n) = 2n^6 + 5n^5 + \tfrac{10}{3}n^4 - \tfrac{1}{3}n^2,$$
$$18S_2(n)^2 = 2n^6 + 6n^5 + \tfrac{13}{2}n^4 + 3n^3 + \tfrac{1}{2}n^2.$$

Subtracting, the n^6 terms cancel exactly, leaving

$$10nS_4(n) - 18S_2(n)^2 = -n^5 - \tfrac{19}{6}n^4 - 3n^3 - \tfrac{5}{6}n^2.$$

Dividing by n^5 and letting $n \to \infty$, every term but the first vanishes, and the limit is -1.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both approaches agree that the sum is quartic in i, j over an $n \times n$ grid, and both correctly predict that the natural degree-6 scaling collapses. They split at the moment you decide *how* to see that collapse. The natural attempt integrates the continuum limit of the summand; the model expands the exact discrete sum. The line is: "the integral of the bracket over the unit square is 0, so the n^6 coefficient vanishes" — true, but computed in a representation (the continuum) that is blind to everything below its own error term.

> *The problem divides by n^5, not n^6. That tells me the designer has already arranged for the natural n^6 leading term to cancel, and the entire content of the answer lives in the term one power of n below it. A continuum approximation to a sum of $N = n^2$ terms each of size $O(n^4)$ carries an error of the same order as that very n^5 term (Euler–Maclaurin says the Riemann-sum error is one order below the sum itself). So the tool the naive attempt reaches for is exactly precise enough to answer a question this problem is not asking, and exactly too crude to answer the one it is.*

> *Since I need the n^5 coefficient and not just the n^6 one, I cannot afford to drop any term of $S_2(n)$ or $S_4(n)$ below the leading power – I need the full polynomial, not its leading behaviour, because $S_2(n)^2$ turns the second term of $S_2(n)$ into a contribution at degree n^5 once squared.*

Layer 3 THE WHY

Why does the exact power-sum formula succeed where the Riemann sum fails, and why is that not an accident of this particular problem? A Riemann sum with mesh $1/n$ approximates $\int_0^1 g(x)\,dx$ to within $O(1/n)$ relative error for a smooth g (Euler–Maclaurin), which for a sum of total size $\Theta(n^6)$ becomes an absolute error of $\Theta(n^5)$. The problem's divisor of n^5 is precisely the scale of that discarded error term. In other words, the problem was constructed (the coefficients $5, -18, 5$ are exactly the ones that annihilate the n^6 integral) so that the "obvious" continuum method throws away the only surviving digit. This is not a quirk of quartic sums; it is a structural fact about any problem where dividing by n^k leaves a k that is one less than the natural scaling degree of the sum – the designer has told you, via the exponent on n, exactly which order of approximation you are not allowed to use. The correct response is always the same: replace the sum by its exact closed form (Faulhaber's formulas, or induction) and only then take the limit; a closed form has no error term to hide the answer in.

There is a second reason the exact route is natural once you see it: $S_2(n)$ and $S_4(n)$ are themselves *polynomials* in n (this is the content of Faulhaber's theorem – $\sum k^m$ is a polynomial of degree $m+1$ in n), so $10nS_4(n) - 18S_2(n)^2$ is a polynomial in n, and a polynomial's behaviour under $n \to \infty$ after dividing by n^5 is read off instantly from its coefficients. The moment you recognise "sum of a fixed power of consecutive integers" you should recognise "polynomial in n", and a polynomial invites exact algebra, not integral estimates.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the mismatch between the quartic terms (naturally $\Theta(n^6)$ when summed twice) and the n^5 divisor. On seeing it, write down: "the leading term must cancel; the answer lives one order down."

2. That cue suggests exact Faulhaber formulas for $S_2(n)$ and $S_4(n)$, substituted whole (not truncated to leading order) into the target combination.

3. Hesitation occurs right after noticing the continuum integral of the bracket is 0 — it feels like the problem is solved ("the coefficient vanishes, so the limit is 0"). It is reinforced away by asking what order of accuracy a Riemann sum actually carries, and comparing that order to the order you are asked for.

4. The trap is answering "0" by treating the vanishing continuum integral as the whole story, when it only confirms the (correct but insufficient) vanishing of the n^6 term.

5. Whenever a sum over $\{1,\dots,n\}$ is divided by a power of n that is *one less* than its naive scaling degree, expect an exact cancellation at top order and reach immediately for closed-form power-sum formulas rather than an integral estimate.

6. Reusable rule: a limit of $(\text{sum})/n^k$ is safe to estimate by an integral only when k matches the sum's natural scaling degree; when k is one power lower, the leading term is designed to cancel and only an exact polynomial expression for the sum will expose the surviving coefficient.

▷ **Cue:** a double sum of a fixed power of i, j over $\{1,\dots,n\}^2$, divided by a power of n one less than the natural scaling degree $\longrightarrow$ **Move:** expect the leading term to cancel by design; expand the exact Faulhaber polynomials for the relevant power sums rather than estimating by an integral

▷ **Cue:** deciding whether a Riemann-sum approximation is strong enough for a given limit $\longrightarrow$ **Move:** compare the order of the Euler–Maclaurin error (one power of n below the sum's own size) to the order at which you are dividing; if they match, the approximation is too crude

▷ **Cue:** $\sum_{k=1}^{n} k^m$ appearing inside a larger expression whose limit is wanted $\longrightarrow$ **Move:** replace it by its exact polynomial-in-n form (Faulhaber), since a polynomial's asymptotics are read off exactly from its coefficients, with no discarded remainder

1981 B2

1981 B2. Determine the minimum value of

$$(r-1)^2+\left(\frac{s}{r}-1\right)^2+\left(\frac{t}{s}-1\right)^2+\left(\frac{4}{t}-1\right)^2$$

for all real numbers r, s, t with $1 \le r \le s \le t \le 4$.

Fields: algebra/analysis (optimisation by pairwise reduction), maximum & minimum problems **Position:** B2 **Answer:** The minimum value is $4(\sqrt{2}-1)^2 = 12-8\sqrt{2}$, attained at $r=\sqrt{2},\ s=2,\ t=2\sqrt{2}$.

Layer 1 THE WHAT — surface analysis

The natural attempt. Four variables (well, three free ones) and a chain constraint $1 \le r \le s \le t \le 4$ scream Lagrange multipliers. You set up $F(r,s,t) = (r-1)^2 + (s/r-1)^2 + (t/s-1)^2 +$

$(4/t-1)^2$ and differentiate:

$$\partial_r F = 2(r-1) - \frac{2s}{r^2}\left(\frac{s}{r}-1\right) = 0,$$
$$\partial_s F = \frac{2}{r}\left(\frac{s}{r}-1\right) - \frac{2t}{s^2}\left(\frac{t}{s}-1\right) = 0,$$
$$\partial_t F = \frac{2}{s}\left(\frac{t}{s}-1\right) - \frac{8}{t^2}\left(\frac{4}{t}-1\right) = 0.$$

Each equation couples three of the four "ratio" quantities $z_1 = r$, $z_2 = s/r$, $z_3 = t/s$, $z_4 = 4/t$ (note $z_1z_2z_3z_4 = 4$ always) in a nonlinear, asymmetric-looking way, because the derivative of $(z_i-1)^2$ with respect to the shared variable picks up a factor of the *next* ratio's denominator. You can grind through the algebra, but it is easy to lose the thread: the system does not visibly reduce to "all four ratios are equal", even though the answer, once known, is exactly that symmetric. You end up trusting a guess (equal ratios) and verifying it, rather than being led to it – which is backwards for a "determine the minimum" problem that wants you to *find* the configuration, not confirm someone else's.

The model chain. Rename the four numbers $1 = x_0 \le x_1 \le x_2 \le x_3 \le x_4 = 4$ (so $x_1 = r, x_2 = s, x_3 = t$) and the four ratios $z_i = x_i/x_{i-1} \ge 1$ for $i = 1,\dots,4$, so that $F = \sum_{i=1}^4 (z_i-1)^2$ subject only to the single constraint $z_1z_2z_3z_4 = x_4/x_0 = 4$ (with each $z_i \ge 1$).

Reduce two variables to one. First solve the sub-problem: for fixed $0 < a < b$, minimise $f(x) = \left(\frac{x}{a}-1\right)^2 + \left(\frac{b}{x}-1\right)^2$ over $a \le x \le b$. Substituting $z = x/a$, $c = b/a > 1$, this is $g(z) = (z-1)^2 + (c/z-1)^2$ for $1 \le z \le c$. Then

$$g'(z) = 2(z-1) - \frac{2c}{z^2}\left(\frac{c}{z}-1\right) = \frac{2}{z^3}\left(z^4 - z^3 + cz - c^2\right) = \frac{2}{z^3}(z^2-c)(z^2-z+c).$$

Since $c > 1 > 1/4$, the second factor $z^2 - z + c$ has negative discriminant $1 - 4c < 0$ and never vanishes, so the only critical point in range is $z = \sqrt{c}$. Comparing endpoints, $g(1) = g(c) = (c-1)^2 = (\sqrt{c}-1)^2(\sqrt{c}+1)^2$, and since $c > 1$ gives $\sqrt{c} > 1$ so $(\sqrt{c}+1)^2 > 4 > 2$, this exceeds $2(\sqrt{c}-1)^2 = g(\sqrt{c})$. So the minimum on $[1,c]$ is at $z = \sqrt{c}$, i.e. $x = \sqrt{ab}$, the geometric mean.

Smoothing to all four at once. This pairwise fact is an exchange argument: whenever two *adjacent* ratios z_i, z_{i+1} (sharing the middle point x_i) are unequal, replacing x_i by $\sqrt{x_{i-1}x_{i+1}}$ makes $z_i = z_{i+1} = \sqrt{z_iz_{i+1}}$, strictly decreases $(z_i-1)^2 + (z_{i+1}-1)^2$ by the sub-problem above, leaves the product z_iz_{i+1} (hence the overall product $z_1z_2z_3z_4$) unchanged, and keeps every ratio ≥ 1 (the geometric mean of two numbers ≥ 1 is ≥ 1). So any configuration with two unequal adjacent ratios is not minimal. Repeating this smoothing (or, more directly, applying Lagrange's condition to $\sum(z_i-1)^2$ subject to $\prod z_i = 4$, which forces every z_i to satisfy the same quadratic $z_i^2 - z_i - 2\lambda = 0$ and hence take at most two values, and then re-running the pairwise argument on any two that differ) drives every ratio to the common value ρ with $\rho^4 = 4$, i.e. $\rho = \sqrt{2}$. This gives $r = \sqrt{2}$, $s = 2$, $t = 2\sqrt{2}$, which indeed satisfies $1 \le \sqrt{2} \le 2 \le 2\sqrt{2} \le 4$. The minimum value is $4(\sqrt{2}-1)^2 = 4(3-2\sqrt{2}) = 12 - 8\sqrt{2}$.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths correctly identify that the four terms are $(z_i-1)^2$ for ratios $z_i = x_i/x_{i-1}$ with $\prod z_i = 4$ fixed. They split at the moment of choosing *which variables* to differentiate with respect to. The natural attempt differentiates with respect to r, s, t directly, so each partial derivative entangles two adjacent ratio-terms through a shared variable and never presents the symmetric structure. The model first solves the two-term version of the same problem in the variable that actually decouples it (the ratio z, not the raw endpoint x), gets a clean closed form for the optimal middle point (the geometric mean), and only then promotes that local fact to a global smoothing argument.

> *Four terms of the same shape, chained end to end with a fixed total product of ratios – I should not differentiate in the original variables r, s, t, because the constraint $\prod z_i = 4$ is multiplicative and the objective is additive in the z_i; I should substitute so that each term depends on exactly one new variable, and optimise in that variable.*

> *If exactly two adjacent ratios are unequal while their product is fixed, replacing both by their common geometric mean cannot make things worse and (by the strict inequality in the two-variable sub-problem) makes things strictly better unless they were already equal – so no minimal configuration can have two unequal adjacent ratios, and the minimum must be at all ratios equal.*

Layer 3 THE WHY

Why is the geometric mean, and not the arithmetic mean or some other combination, the right point to insert? Because the constraint linking z_i and z_{i+1} is $z_i z_{i+1} =$ constant – it is their *product*, not their sum, that is fixed (this is forced by the telescoping structure $x_{i-1} \to x_i \to x_{i+1}$: the ratios multiply to give x_{i+1}/x_{i-1}). Whenever a smoothing argument holds one quantity's *product* fixed while trying to move its factors towards equality, the invariant point is the geometric mean, because that is the unique value m with $m \cdot m = z_i z_{i+1}$. This is exactly analogous to the AM–GM inequality's equality case, and in fact the sub-problem's conclusion is a disguised instance of convexity: $(z-1)^2$ is a strictly convex function of z, but the constraint is convex in $\log z$ (since $\log z_i + \log z_{i+1}$ is what is actually fixed), so the true smoothing variable is $w = \log z$, and $\varphi(w) = (e^w - 1)^2$ being smoothed toward equal w's at fixed $\sum w$ is precisely equalising the z_i on a log scale – i.e. setting them all equal to the geometric mean of the extremes. Once you see the constraint as multiplicative, "take logs, then it's just two numbers averaging to reduce a strictly convex cost" is the natural move, not a trick.

The second thing worth extracting: this is the general $n = 4$ case of a chain $1 = x_0 \le x_1 \le \cdots \le x_n = N$ split into ratios; the same smoothing argument gives, for any n, that the minimum of $\sum_{i=1}^{n} (z_i - 1)^2$ subject to $\prod z_i = N$ is $n(N^{1/n} - 1)^2$, attained at the geometric progression $x_i = N^{i/n}$. The problem's particular numbers ($n = 4$, $N = 4$) were chosen so that $N^{1/n} = \sqrt{2}$ comes out clean, but the mechanism generalises to any chain-of-ratios optimisation whatsoever.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is the telescoping chain of ratios $r/1$, s/r, t/s, $4/t$ whose product is fixed at 4 regardless of r, s, t. On seeing it, write down the product identity explicitly before touching a derivative.

2. That cue suggests optimisation under a multiplicative constraint, which in turn suggests substituting ratio variables (or taking logs) so the constraint becomes additive and the objective stays separable term by term.

3. Hesitation occurs when you try to differentiate $F(r, s, t)$ directly and the three equations refuse to decouple. It is reinforced away by asking "what is the natural variable in which this constraint is simple?" rather than pushing harder on the algebra in the given variables.

4. The trap is guessing the symmetric answer (equal ratios) from the shape of the problem and merely verifying it, instead of having a method that derives it – which fails to generalise and gives no proof that it is a minimum rather than another critical point.

5. Any time a sum of same-shape terms is constrained by a fixed *product* of associated quantities, substitute the ratio (or its log) as the variable, and expect the geometric mean (equivalently, equal ratios) to be the extremiser by a pairwise smoothing or AM–GM-type argument.

6. Reusable rule: for $\sum \phi(z_i)$ with ϕ strictly convex and $\prod z_i$ fixed, the minimum is at all z_i equal to the common geometric-mean value; find that value first from the constraint alone, then verify feasibility (here: does the resulting equal-ratio point respect the ordering $r \le s \le t$?) before committing to it as the answer.

- ▷ **Cue:** a chain of ratios (or a telescoping product) with a fixed total product $\longrightarrow$ **Move:** substitute ratio variables $z_i = x_i/x_{i-1}$ (or their logs) to turn the multiplicative constraint into an additive one, then optimise term by term
- ▷ **Cue:** minimising $\sum(z_i - 1)^2$ (or any $\sum \phi(z_i)$ with ϕ strictly convex) subject to $\prod z_i = N$ $\longrightarrow$ **Move:** expect the minimiser at all z_i equal to $N^{1/n}$, proved by a pairwise smoothing argument: if $z_i \neq z_{i+1}$, replacing both by their geometric mean strictly decreases the sum while preserving the product
- ▷ **Cue:** a two-term optimisation $(x/a - 1)^2 + (b/x - 1)^2$ on $a \leq x \leq b$ $\longrightarrow$ **Move:** substitute $z = x/a$, $c = b/a$; the unique interior critical point is $z = \sqrt{c}$, i.e. $x = \sqrt{ab}$, and it beats both endpoints

1981 B3

1981 B3. Prove that there are infinitely many positive integers n with the property that if p is a prime divisor of $n^2 + 3$, then p is also a divisor of $k^2 + 3$ for some integer k with $k^2 < n$.

Fields: number theory (norm-form factorisation, induction) **Position:** B3 **Answer:** There are infinitely many such n; one explicit infinite family is given by $a_1 = 5$ and $a_{i+1} = (a_i - 1)a_i + 3$ for $i \geq 1$.

Caveat. The printed solution (MAA, *Putnam 1965–1984*) begins from the same key identity used below but proceeds via an n defined as $(m^2 + m + 2)(m^2 + m + 3) + 3$; the OCR breaks off after "This is easily proved" and the construction is never finished on the page. I have not tried to recover the printed author's exact continuation. Instead I built and fully verified my own infinite family from the same identity, checked every inequality below by direct algebra, and confirmed the base case numerically ($a_1 = 5$: $5^2 + 3 = 28 = 2^2 \cdot 7$, and indeed $2 \mid 1^2 + 3 = 4$, $7 \mid 2^2 + 3 = 7$, with $1^2 < 5$ and $2^2 < 5$). Nothing below is asserted without having been checked.

Layer 1 THE WHAT — surface analysis

The natural attempt. The statement is about primes dividing $n^2 + 3$, so the reflex is number-theoretic: factor $n^2 + 3$ for small n and look for a pattern in which primes appear. $n = 1$: $4 = 2^2$. $n = 2$: 7. $n = 3$: $12 = 2^2 \cdot 3$. $n = 4$: 19. $n = 5$: $28 = 2^2 \cdot 7$. $n = 7$: $52 = 2^2 \cdot 13$. You notice primes $2, 7, 13, \ldots$ keep recurring and start hunting for a congruence condition (values of n mod some modulus) that controls which primes divide $n^2 + 3$ – a perfectly reasonable first instinct, since $n^2 + 3 \equiv 0 \pmod{p}$ iff -3 is a quadratic residue mod p, which happens precisely for $p \equiv 1 \pmod 3$ (plus $p = 3$). But this line of attack answers a different question: it classifies which primes *can ever* divide some $n^2 + 3$, not which n have the *descent* property demanded here (every prime factor of $n^2 + 3$ already shows up for a strictly smaller witness k with $k^2 < n$). Quadratic-residue theory alone gives you no mechanism to compare the size of n against the size of the witnessing k – it is blind to magnitude entirely – so this path, however natural, cannot reach the finish line unassisted.

The model chain. Work with $f(x) = x^2 + 3$ and look for an *algebraic* factorisation of $f(n)$ into two smaller values of f, since that is the only way to control both "which prime" and "how large the witness is" at once. Compute $f(x)f(x+1) = (x^2 + 3)(x^2 + 2x + 4) = x^4 + 2x^3 + 7x^2 + 6x + 12$, and compare with $f(x^2 + x + 3) = (x^2 + x + 3)^2 + 3 = x^4 + 2x^3 + 7x^2 + 6x + 12$. They match exactly:

$$f(x)\, f(x+1) = f\big(x^2 + x + 3\big) \qquad \text{for every integer } x. \tag{$\star$}$$

This says: the value of f at $x^2 + x + 3$ factors into $f(x)$ and $f(x+1)$, so any prime dividing $f(x^2 + x + 3)$ divides one of those two smaller values.

Now build the family by induction, choosing at each step *which* of the two factors to use as witness so that the size inequality $k^2 < n$ always holds. Define $a_1 = 5$ (from $x = 1$ in $(\star)$: $f(1)f(2) = 4 \cdot 7 = 28 = f(5)$, since $1^2 + 1 + 3 = 5$), and for $i \geq 1$,

$$a_{i+1} = (a_i - 1)\, a_i + 3.$$

This is $(\star)$ applied with $x = a_i - 1$ (so $x + 1 = a_i$): $f(a_i - 1)f(a_i) = f\big((a_i - 1)^2 + (a_i - 1) + 3\big) = f(a_{i+1})$, since $(a_i - 1)^2 + (a_i - 1) + 3 = a_i^2 - a_i + 3 = (a_i - 1)a_i + 3 = a_{i+1}$.

Claim. For every $i \geq 1$, every prime factor p of $f(a_i)$ divides $f(k)$ for some integer k with $k^2 < a_i$.

Base case $i = 1$: $f(5) = 28 = 2^2 \cdot 7$. We have $2 \mid f(1) = 4$ and $7 \mid f(2) = 7$, with $1^2 = 1 < 5$ and $2^2 = 4 < 5$. Both prime factors are witnessed. (Equivalently: apply the general argument below with $a_0 = 1$.)

Inductive step: assume the claim for a_i. Since $f(a_{i+1}) = f(a_i - 1)f(a_i)$, a prime $p \mid f(a_{i+1})$ divides $f(a_i - 1)$ or $f(a_i)$.

- If $p \mid f(a_i - 1)$: take $k = a_i - 1$. Then $a_{i+1} - (a_i - 1)^2 = \big[(a_i - 1)a_i + 3\big] - (a_i - 1)^2 = (a_i - 1)\big[a_i - (a_i - 1)\big] + 3 = (a_i - 1) + 3 = a_i + 2 > 0$, so $k^2 = (a_i - 1)^2 < a_{i+1}$.
- If $p \mid f(a_i)$: by the inductive hypothesis, $p \mid f(k)$ for some k with $k^2 < a_i$. Since $a_{i+1} - a_i = a_i^2 - 2a_i + 3 = (a_i - 1)^2 + 2 > 0$, we have $a_i < a_{i+1}$, so $k^2 < a_i < a_{i+1}$.

In either case $p \mid f(k)$ for some k with $k^2 < a_{i+1}$, proving the claim for $i + 1$.

By induction the claim holds for all $i \geq 1$; since $a_{i+1} > a_i$ for every i (just shown), the sequence $a_1 < a_2 < a_3 < \cdots$ is strictly increasing, giving infinitely many $n = a_i$ with the required property.

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both paths start from the same object, $f(x) = x^2 + 3$, and the same question: which primes divide it. They split at the point where the natural attempt asks "for which n does a given prime divide $f(n)$" (a residue question, symmetric in all valid n, with no notion of size), while the model asks "can I write f of one number as a *product* of f of two strictly smaller numbers" (a factorisation question, which by construction hands you both the prime's source and a size bound in the same stroke).

> *The property I need to prove links two different things about n: which primes divide $f(n)$, and how big n is compared to a witness k. A pure congruence classification of "which primes divide some $f(n)$" can never produce a size comparison, because congruences don't see magnitude – I need an identity that literally factors f of a large input into f of smaller inputs, so the size bound comes for free from the factorisation.*

> *Once I have $f(x)f(x+1) = f(x^2 + x + 3)$, applying it directly at each step reuses the* larger *of the two witnesses, $x + 1$, whose square can exceed the new value; I should instead shift the recursion by one – feed in $x = a_i - 1$ so that $x + 1 = a_i$ is the term I already control by induction, and only the* other *factor $x = a_i - 1$ needs a fresh, easy size check.*

Layer 3 THE WHY

Why does an identity like $(\star)$ exist at all, and why is it not a lucky coincidence? The form $x^2 + 3$ is the norm from $\mathbb{Z}[\sqrt{-3}]$ (or, up to a factor of 4, from the ring of integers of $\mathbb{Q}(\sqrt{-3})$) restricted to elements of a special shape, and norms are always multiplicative: $N(\alpha)N(\beta) = N(\alpha\beta)$. Concretely, $(a^2 + 3)(b^2 + 3) = (ab - 3)^2 + 3(a + b)^2 = (ab + 3)^2 + 3(a - b)^2$ for any a, b (both are the Brahmagupta–Fibonacci identity for the form $x^2 + 3y^2$). The identity $(\star)$ is

the special case $b = a + 1$ of the second form, where $a - b = -1$ makes the "$3(a-b)^2$" term collapse to exactly $3 \cdot 1^2 = 3$, landing back on the *same* form $x^2 + 3$ (with $y = 1$) rather than the more general $x^2 + 3y^2$. So the identity is not an accident: it is the one place where multiplying two values of $x^2 + 3$ by consecutive integers keeps you inside the same one-parameter family instead of spilling into the two-parameter family $x^2 + 3y^2$. Recognising $n^2 + 3$ as a norm, and recalling that norms multiply, is what makes you look for a product identity in the first place, rather than stumbling on $(\star)$ by polynomial-matching alone.

Why must the recursion be shifted by one (feeding $a_i - 1$ rather than a_i into $(\star)$), rather than iterating $(\star)$ directly on a_i? Because $(\star)$ applied to $x = a_i$ produces two factors, $f(a_i)$ and $f(a_i + 1)$, and only one of them ($f(a_i)$) is already covered by the induction; the other, $f(a_i + 1)$, is a brand-new value with no size guarantee yet, and checking $(a_i + 1)^2 < a_{i+2}$ directly turns out to fail for $a_i \geq 3$ (a short computation shows $a_{i+2} - (a_i + 1)^2 = 2 - a_i$). Shifting the recursion so that the *known* term a_i is the *larger* of the two factors, and the *new* term $a_i - 1$ is the smaller one, converts the one unchecked factor into an elementary inequality ($a_i + 2 > 0$) instead of a size race you can lose. This is the general lesson of building an infinite family by a self-referential identity: always arrange the recursion so that the previously-verified quantity plays the role that is automatically safe, and isolate the fresh computation into the one piece you can check directly.

Layer $+\alpha$ BACKTESTING

1. The first useful cue is that the conclusion couples *which primes* divide $n^2 + 3$ with *how large n* is relative to a witness. On seeing a divisibility claim entangled with a size inequality, write down: "I need an identity that factors the big value into smaller values of the same form, not just a congruence classification."

2. That cue suggests testing whether $x^2 + 3$ (or the given quadratic form) is multiplicative – i.e., whether $f(x)f(y)$ can be rewritten as f of some combination of x, y – motivated by recognising the form as a number-theoretic norm.

3. Hesitation occurs right after finding $(\star)$, when the first instinct is to iterate it naively ($a_{i+1} = a_i^2 + a_i + 3$) and hope both factors $f(a_i)$, $f(a_i + 1)$ stay small enough. It is reinforced away by explicitly checking the size inequality for the *new* factor rather than assuming the pattern that worked for the old one continues to work.

4. The trap is treating the two factors produced by $(\star)$ symmetrically; only one of them is already known-good by induction, and forcing the new one to also satisfy the size bound without adjusting the recursion produces a false step for larger indices.

5. Whenever a quadratic (or higher) form's value needs to be shown to factor through smaller values of the same form, check whether the form is a norm of some quadratic ring, since norms are multiplicative by construction and immediately suggest the identity to search for.

6. Reusable rule: to build an infinite family by a self-referential factorisation identity $f(\text{combination of } x, y) = f(x)f(y)$, always route the recursion so the term already validated by induction is the *larger* of the two factors and the fresh term is the *smaller*; then the one unchecked size inequality is elementary rather than a race that can fail.

▷ **Cue:** a divisibility claim about $n^2 + c$ entangled with a size comparison to a smaller witness $\longrightarrow$ **Move:** do not reach for a congruence/quadratic-residue classification (it is blind to size); search instead for a multiplicative identity that factors f of a large value into f of smaller values

▷ **Cue:** recognising a quadratic form $x^2 + c$ (for constant $c > 0$) as a candidate norm form $\longrightarrow$ **Move:** test Brahmagupta–Fibonacci-style identities $(a^2 + c)(b^2 + c) = (ab \mp c)^2 + c(a \pm b)^2$ for a specialisation (e.g. $b = a + 1$) that collapses the extra term back onto the same one-parameter family

▷ **Cue:** building an infinite family by iterating a factorisation identity $f(\text{combination}) = f(x)f(y)$ where only one of x, y is already validated $\longrightarrow$ **Move:** shift the recursion so the

validated term is always the larger factor and the new term is the smaller one, turning the outstanding size check into an elementary inequality

1981 B4

1981 B4. Let V be a set of 5 by 7 matrices, with real entries and with the property that $rA + sB \in V$ whenever $A, B \in V$ and r and s are scalars (i.e., real numbers). Prove or disprove the following assertion: If V contains matrices of ranks 0, 1, 2, 4, and 5, then it also contains a matrix of rank 3.

[The rank of a nonzero matrix M is the largest k such that the entries of some k rows and some k columns form a k by k matrix with a nonzero determinant.]

Fields: linear algebra; rank; counterexample construction **Position:** B4 — mid-to-late in the paper; nothing technical, but the wording pulls you towards a proof that does not exist **Answer:** The assertion is false. The space $V = \{M(a,b,c)\}$ with $M = aE_{11} + c(E_{12} + E_{21}) + b(E_{22} + E_{33} + E_{44} + E_{55})$ contains matrices of ranks $0, 1, 2, 4, 5$ and none of rank 3.

Layer 1 THE WHAT — surface analysis

The natural attempt. The list "$0, 1, 2, 4, 5$" with 3 conspicuously missing reads like an intermediate value theorem: rank ought to move one step at a time, so a subspace that reaches both 2 and 4 must pass through 3. So you set out to prove it.

Take $A \in V$ of rank 2 and $B \in V$ of rank 4, and walk along the line $C(t) = (1-t)A + tB$, which stays in V. Every $k \times k$ minor of $C(t)$ is a polynomial in t. Now look at what that tells you. If some 4×4 minor of $B = C(1)$ is nonzero, that polynomial is not identically zero, so it vanishes at only finitely many t: $\operatorname{rank} C(t) \geq 4$ for all but finitely many t. Meanwhile at $t = 0$ the rank is 2. The rank does not slide from 2 to 4; it sits at a generic value ≥ 4 and drops, at isolated points, by whatever amount it likes.

Second try: use the fact that adding a rank-one matrix changes the rank by at most one. If V were spanned by rank-one matrices you could step from rank 2 to rank 4 one unit at a time and must land on 3. But nothing in the hypothesis says V has any rank-one matrices beyond the one you are given. That is the wall, and a wall in a "prove or disprove" problem is information: the proof is failing for a reason, and the reason is a recipe for a counterexample.

The model chain. Write E_{ij} for the 5×7 matrix with 1 in position (i, j) and 0 elsewhere. For real a, b, c put

$$M(a,b,c) = aE_{11} + c\,(E_{12} + E_{21}) + b\,(E_{22} + E_{33} + E_{44} + E_{55}),$$

so the top-left 5×5 block is

$$\begin{pmatrix} a & c & 0 & 0 & 0 \\ c & b & 0 & 0 & 0 \\ 0 & 0 & b & 0 & 0 \\ 0 & 0 & 0 & b & 0 \\ 0 & 0 & 0 & 0 & b \end{pmatrix}$$

and columns 6, 7 are zero. Let $V = \{M(a,b,c) : a, b, c \in \mathbb{R}\}$. Since M is linear in (a, b, c), $rM(a,b,c) + sM(a',b',c') = M(ra + sa', rb + sb', rc + sc') \in V$.

Case $b \neq 0$. Rows $2, 3, 4, 5$ contain the 4×4 minor on columns $2, 3, 4, 5$, which is $\det \operatorname{diag}(b, b, b, b) = b^4 \neq 0$. So the rank is at least 4, and there are only five rows, so it is 4 or 5.

Case $b = 0$. Only rows 1 and 2 can be nonzero, so the rank is at most 2.

Hence no element of V has rank 3. The required ranks occur: $M(0,0,0)$ has rank 0; $M(1,0,0) = E_{11}$ has rank 1; $M(0,0,1) = E_{12} + E_{21}$ has the 2×2 minor $\det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -1$,

rank 2; $M(0,1,0)$ has rank 4 (row 1 is zero, so not 5); $M(1,1,0)$ has the 5×5 identity minor, rank 5.

So V is closed under linear combinations and contains matrices of ranks $0, 1, 2, 4, 5$ but none of rank 3. The assertion is false. □

(The official solution uses $c\,(E_{16} + E_{21})$ in place of $c\,(E_{12} + E_{21})$; the case split on b is identical. All ranks above were also checked by machine on random (a, b, c).)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains write down the line $C(t) = (1-t)A + tB$ inside V and see that rank is governed by minors. The natural attempt reads that as a difficulty to be overcome and keeps trying to prove. The model reads the same line as the answer: rank on a line is generic-with-drops, so nothing forces the value 3, and the job becomes *design a space whose ranks split into "at least 4" and "at most 2".*

> *Rank is not continuous. On a line it takes one generic value and drops at finitely many points by any amount. So there is no intermediate value theorem here, and I should stop proving and start building.*

> *I want every matrix in V to have rank ≥ 4 or ≤ 2. Rank ≥ 4 is easy to force: one parameter b that fills four rows. Rank ≤ 2 is easy to force: when $b = 0$ everything else lives in two rows. Then put the rank-1 and rank-2 pieces inside the same two rows so they cannot add up to 3.*

- ▷ **Cue:** "prove or disprove" with a list of values that skips one ⟶ **Move:** test the intermediate-value intuition first; if the quantity is not continuous (rank, degree, dimension of a fibre), look for a counterexample
- ▷ **Cue:** you need a linear family avoiding one rank ⟶ **Move:** split by one parameter: a block that forces rank $\geq k+1$ when switched on, and everything else confined to $k-1$ rows when it is off
- ▷ **Cue:** two rank pieces in a direct sum give ranks that add ⟶ **Move:** overlap their supports instead, so the combination cannot exceed the larger

Layer 3 THE WHY

Why is rank the wrong thing to apply an intermediate value argument to? Go back to the definition the problem hands you in brackets: rank $\geq k$ means *some* $k \times k$ minor is nonzero. That is an open condition — a nonzero determinant stays nonzero under small perturbation — so rank can only jump *up* near a point, never down. Rank is lower semicontinuous, and that is all. On a line $C(t)$ the minors are polynomials, so "rank $\geq k$" fails at either every t or finitely many t. The picture is a constant function with isolated dips, and a dip from 5 to 2 passes through nothing.

The one situation where rank does move in unit steps is when you add rank-one matrices, because $|\operatorname{rank}(X+Y) - \operatorname{rank} X| \leq \operatorname{rank} Y$. That identifies exactly what a counterexample must avoid: V may contain only very few rank-one matrices. In the model, the only rank-one matrices are the multiples of E_{11} (rank 1 forces $b = 0$, and then $\det \begin{pmatrix} a & c \\ c & 0 \end{pmatrix} = -c^2$ forces $c = 0$), and adding them to anything moves you around inside the "≤ 2" world or the "≥ 4" world, never across.

Now why does the construction look the way it does? The first idea is direct sums: put blocks of ranks 1, 2, 4 on disjoint rows and columns with independent scalars. Then the ranks in V are all subset sums of $\{1, 2, 4\}$, which include $1 + 2 = 3$. That failure is instructive: the rank-1 and rank-2 generators must *not* add. So place them in the same two rows. The 2×2 block $\begin{pmatrix} a & c \\ c & 0 \end{pmatrix}$ has rank 2 whenever $c \neq 0$ and rank ≤ 1 when $c = 0$ — it never exceeds 2 because

it only has two rows. Then add a parameter b whose presence alone forces rank 4 (four rows, diagonal). With five rows total, $b \neq 0$ gives 4 or 5 and $b = 0$ gives at most 2. The gap at 3 is not a coincidence of arithmetic; it is the pigeonhole on rows.

Is three parameters necessary? Yes. If $\dim V = 2$, the rank-1 matrix A and rank-2 matrix B given by hypothesis are linearly independent (a multiple of A has rank ≤ 1), so they span V, and every element has rank $\leq \operatorname{rank} A + \operatorname{rank} B = 3$ — no rank 4 or 5 at all. So the smallest counterexample is three-dimensional, as both the official solution and the model are.

Layer $+\alpha$ BACKTESTING

1. The cue is the phrase "prove or disprove" next to a list $0, 1, 2, 4, 5$ that skips 3. Write down: *what makes rank move one step at a time, and can I stop it?*

2. Counterexample by construction: a small parametrised family of matrices with a case split on one parameter.

3. Hesitation comes after the line $C(t)$ fails to force rank 3: students assume they need a better proof. Reinforce with the definition of rank as "some minor nonzero" — an open condition — which tells you rank can dip arbitrarily.

4. The trap is the intermediate value intuition, and the secondary trap is a direct-sum construction whose ranks add to 3.

5. Whenever a discrete invariant (rank, number of roots, dimension) is asked to take an "in-between" value, ask whether it is continuous. Semicontinuous invariants jump.

6. Reusable decision rule: *to make a linear family miss a value k, let one parameter force the invariant above k and confine everything else so it stays below k; overlap the small pieces so they cannot combine to k.*

1981 B5

1981 B5. Let $B(n)$ be the number of ones in the base two expression for the positive integer n. For example, $B(6) = B(110_2) = 2$ and $B(15) = B(1111_2) = 4$. Determine whether or not

$$\exp\Big(\sum_{n=1}^{\infty} \frac{B(n)}{n(n+1)}\Big)$$

is a rational number. Here $\exp(x)$ denotes e^x.

Fields: infinite series; binary representation; self-similar sums **Position:** B5 — late in the paper, but short once the right structural move is made **Answer:** Yes. $\sum_{n\geq 1} B(n)/(n(n+1)) = 2\ln 2 = \ln 4$, so the expression equals 4.

Layer 1 THE WHAT — surface analysis

The natural attempt. Call the sum S. The question is really "is S the logarithm of a rational?", and the only natural source of logarithms of rationals from a series like this is $\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \cdots$. So you expect S to be a rational multiple of $\ln 2$ (or a sum of logs), and you set out to compute it.

The standard move with a digit-counting function is to write it as a sum over digits. With $n = \sum_i n_i 2^i$, $n_i \in \{0, 1\}$, we have $B(n) = \sum_i n_i$, so

$$S = \sum_{i\geq 0} \sum_{n:\, n_i = 1} \frac{1}{n(n+1)}.$$

Which n have $n_i = 1$? Those in the blocks $[\,2^i(2j+1),\ 2^i(2j+2) - 1\,]$ for $j = 0, 1, 2, \ldots$. Over a block of consecutive integers, $\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$ telescopes to $\frac{1}{2^i(2j+1)} - \frac{1}{2^i(2j+2)}$. Summing over j,

$$\sum_{n:\, n_i = 1} \frac{1}{n(n+1)} = \frac{1}{2^i}\Big(1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \cdots\Big) = \frac{\ln 2}{2^i},$$

and $S = \sum_i \ln 2/2^i = 2\ln 2$.

This is correct and is the official book's first solution. It costs: you must characterise the blocks where bit i is set, telescope inside each block, recognise the alternating harmonic series, and justify swapping the order of summation. None of it is hard, but there are four places to slip, and the block description is where most people do.

The model chain. *Convergence.* If n has d binary digits then $2^{d-1} \le n$, so $B(n) \le d \le 1 + \log_2 n$. Hence $0 \le B(n)/(n(n+1)) \le (1+\log_2 n)/n^2$, which is summable. So $S = \sum_{n\ge 1} B(n)/(n(n+1))$ is a finite real number, and since all terms are nonnegative we may split and regroup the series freely.

Self-similarity. Appending a binary digit to m gives $2m$ (digit 0) or $2m+1$ (digit 1), so

$$B(2m) = B(m)\ (m \ge 1), \qquad B(2m+1) = B(m) + 1\ (m \ge 0),$$

with the convention $B(0) = 0$. Split S by the parity of n:

$$S = \sum_{m\ge 0} \frac{1 + B(m)}{(2m+1)(2m+2)} + \sum_{m\ge 1} \frac{B(m)}{2m(2m+1)}.$$

The constant part. $\sum_{m\ge 0} \frac{1}{(2m+1)(2m+2)} = \sum_{m\ge 0}\left(\frac{1}{2m+1} - \frac{1}{2m+2}\right) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \ln 2$.

The $B(m)$ part. Collecting the coefficient of $B(m)$ for $m \ge 1$ (the $m = 0$ term has $B(0) = 0$):

$$\frac{1}{(2m+1)(2m+2)} + \frac{1}{2m(2m+1)} = \frac{2m + (2m+2)}{2m(2m+1)(2m+2)} = \frac{2(2m+1)}{2m(2m+1)(2m+2)} = \frac{1}{2}\cdot\frac{1}{m(m+1)}.$$

So the $B(m)$ part is $\frac{1}{2}\sum_{m\ge 1} B(m)/(m(m+1)) = S/2$.

Solve. $S = \ln 2 + S/2$. Because S is finite, subtract: $S = 2\ln 2 = \ln 4$, and

$$\exp(S) = \boxed{4}, \text{ a rational number.} \qquad \square$$

(Numerically, the partial sum to $n = 2 \times 10^6$ is $1.386288\ldots$ against $\ln 4 = 1.386294\ldots$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains start from S and a structural description of B. They split at the choice of description. (N) takes the *static* definition, $B(n) = \sum_i n_i$, and sums digit by digit. (M) takes the *recursive* definition, $B(2m) = B(m)$, $B(2m+1) = B(m) + 1$, splits by parity, and watches S reappear inside itself.

> *B is defined by what happens when I strip the last binary digit. So split the sum by the last digit, even n and odd n, and see whether the sum I am computing comes back.*

> *It came back with a factor $\frac{1}{2}$. An equation $S = c + S/2$ is a closed form, provided I have already shown S is finite.*

- ▷ **Cue:** a sum of a digit function (binary digit sum, number of trailing zeros, base-b anything) against a weight $\longrightarrow$ **Move:** split by the last digit using $f(2m)$, $f(2m+1)$ in terms of $f(m)$, and look for the original sum reappearing
- ▷ **Cue:** a question "is exp(series) rational?" $\longrightarrow$ **Move:** the series is almost certainly a rational combination of logs; aim for $\ln 2$ via $1 - \frac{1}{2} + \frac{1}{3} - \cdots$
- ▷ **Cue:** you are about to solve $S = c + \lambda S$ $\longrightarrow$ **Move:** prove S finite first; otherwise you may be subtracting infinity

Layer 3 THE WHY

The official book gives both routes and puts the digit-by-digit one first. Here is why the second is the natural one.

Start from the definition. "The number of ones in the base two expression" is not really a formula in n; it is a statement about a string, and strings are built by appending characters. The map $m \mapsto 2m$ appends 0, the map $m \mapsto 2m+1$ appends 1, and every positive integer arises from a smaller one in exactly one of these two ways. So the definition of B *is* the pair $B(2m) = B(m)$, $B(2m+1) = B(m)+1$; the digit-sum formula is the unrolled version. When the object is defined recursively, sums over it want to be evaluated recursively — that is what the parity split does.

Why does S come back so cleanly? Because the weight $\frac{1}{n(n+1)}$ is itself nearly scale-invariant: $\frac{1}{n(n+1)} \approx \frac{1}{n^2}$, and halving n multiplies $\frac{1}{n^2}$ by 4, while two values of n ($2m$ and $2m+1$) feed each m. Roughly $2 \cdot \frac{1}{4} = \frac{1}{2}$, which is the factor in $S = \ln 2 + S/2$. The exact identity $\frac{1}{2m(2m+1)} + \frac{1}{(2m+1)(2m+2)} = \frac{1}{2m(m+1)}$ is the precise form of that heuristic: it is a partial-fractions check you can predict before doing it. And the constant $\ln 2$ comes from the "+1" in $B(2m+1)$, weighted over odd positions only — odd-minus-even reciprocals, the alternating harmonic series.

The digit-by-digit route computes the same thing in a different order: it asks "how much does bit i contribute?" and answers $\ln 2/2^i$. The factor 2^{-i} there is the same self-similarity (bit i of n is bit 0 of $\lfloor n/2^i \rfloor$), recomputed separately for each i. The recursive route uses it once.

One more thing the recursive route demands: finiteness of S before you subtract. The official solution states $B(n) \le 1 + \log_2 n$ at the outset for exactly this reason. Without it, $S = \ln 2 + S/2$ is also satisfied by $S = \infty$.

Layer $+\alpha$ BACKTESTING

1. The first cue is "number of ones in the base two expression". Write down $B(2m) = B(m)$, $B(2m+1) = B(m)+1$ immediately.
2. Split the series by parity and derive a functional equation for the sum.
3. Hesitation comes at the coefficient of $B(m)$: two fractions that do not obviously combine. Reinforce with the scaling heuristic: you expect $\frac{1}{2} \cdot \frac{1}{m(m+1)}$, so check it.
4. The trap is solving $S = \ln 2 + S/2$ without having shown $S < \infty$. A second trap is misdescribing the set of n with bit i set in the digit-by-digit route.
5. Any quantity defined digit by digit (binary weight, Thue–Morse sign, base-3 Cantor digits) has a two- or b-term recursion under $n \mapsto bn + r$; sums and integrals against it usually satisfy a self-similar equation.
6. Reusable decision rule: *when a function is defined by a recursion, evaluate sums of it by the same recursion; look for the unknown sum to reappear with a factor less than* 1*, and prove finiteness before solving.*

1981 B6

1981 B6. Let C be a fixed unit circle in the Cartesian plane. For any convex polygon P each of whose sides is tangent to C, let $N(P,h,k)$ be the number of points common to P and the unit circle with center at (h,k). Let $H(P)$ be the region of all points (x,y) for which $N(P,x,y) \ge 1$ and $F(P)$ be the area of $H(P)$. Find the smallest number u with

$$\frac{1}{F(P)} \iint N(P,x,y)\,dx\,dy \le u$$

for all polygons P, where the double integral is taken over $H(P)$.

Caveat. The source is book OCR and renders the final inequality sign as "<". The standard statement has $\le$. The answer is $8/3$ under either reading, because the ratio is

always strictly less than $8/3$ and approaches it; the analysis below covers both.

Fields: integral geometry; double integrals; extremal problem **Position:** B6 — the last problem, hard to start because the integrand N looks hopelessly piecewise; short once the count is split over the sides **Answer:** $u = 8/3$. For every such P with perimeter L, $F(P) = \pi + \frac{3}{2}L$ and $\iint N = 4L$, so the ratio is $8/(3 + 2\pi/L) < 8/3$, and it tends to $8/3$ as $L \to \infty$.

Layer 1 THE WHAT — surface analysis

The natural attempt. First decode. N counts *points*, so P here means the boundary of the polygon, a closed broken line. Take a concrete case, the square of side 2 circumscribing C, and try to find the region where $N = 1, 2, 3, 4, \ldots$, then integrate.

Far from the square, $N = 0$. A unit circle whose centre is near the middle of one side, just outside, crosses that side twice: $N = 2$. A centre near a corner, outside, gives a circle that can cross both sides twice: $N = 4$. A centre inside near a corner likewise: $N = 4$. Near the centre of the square, the circle of radius 1 touches all four sides: $N = 4$ at the single point O, and nearby it crosses each side twice, so $N = 8$ on a small curvilinear region around O. Each of these regions is bounded by arcs of unit circles centred at the vertices and by lines parallel to the sides at distance 1. You can compute their areas — lenses, sectors, curved squares — and sum $\sum_k k \cdot \text{area}\{N = k\}$.

This is where it dies. The bookkeeping is heavy for a square and has no chance for a general tangential polygon, and the problem asks for a supremum over *all* of them. Even if you finished the square, you would learn one number and nothing about how the ratio depends on the shape.

The model chain. Let O be the centre of C, let the sides of P be segments $S_1, \ldots, S_m$ of lengths $\ell_1, \ldots, \ell_m$, and $L = \sum \ell_i$. Each S_i touches C and P is convex, so C is the incircle of P. Write Γ for the boundary of P (so $N(P, X) = \#(\Gamma \cap \text{circle}(X, 1))$), $\overline{P}$ for the closed polygonal region, and $d(X)$ for the distance from X to Γ.

Step 1: $H(P) = \{X : d(X) \le 1\}$. The function $Q \mapsto |QX|$ is continuous on the connected set Γ, so the unit circle about X meets Γ iff $\min_\Gamma |QX| \le 1 \le \max_\Gamma |QX|$. The maximum of the convex function $|QX|$ over $\overline{P}$ is attained at a vertex, which is on Γ; and $\overline{P}$ contains the unit disc about O, which contains a point at distance $|OX| + 1 \ge 1$ from X. So the right-hand inequality always holds, and $N \ge 1$ iff $d(X) \le 1$.

Step 2: $F(P) = \pi + \frac{3}{2}L$. Joining O to the vertices cuts $\overline{P}$ into triangles of height 1, so $\text{area}(\overline{P}) = \frac{1}{2}L$. For X inside P let d_i be its distance to the line of S_i; the same dissection from X gives $\frac{1}{2}\sum \ell_i d_i = \frac{1}{2}L$, so $\min_i d_i \le 1$. Walking from X perpendicular to the line of S_i you reach that line after distance d_i, and $\overline{P}$ lies on one side of it, so you cross Γ on the way: $d(X) \le \min d_i \le 1$. Hence all of $\overline{P}$ lies in H. Outside P, the points within distance 1 form rectangles $\ell_i \times 1$ on the sides together with circular sectors at the vertices whose angles are the exterior angles, summing to 2π: area $L + \pi$. Total $F(P) = \frac{1}{2}L + L + \pi$.

Step 3: $\iint N = 4L$. The centres X whose circle passes through a vertex lie on finitely many circles, a set of area zero. For all other X, every point of $\Gamma \cap \text{circle}(X, 1)$ lies on exactly one side, so $N = \sum_i N_i$ with $N_i(X) = \#(S_i \cap \text{circle}(X, 1))$. Fix i and choose coordinates with $S_i = [0, \ell_i] \times \{0\}$ and $X = (h, k)$. If $|k| > 1$, $N_i = 0$. If $|k| < 1$, put $s = \sqrt{1 - k^2} > 0$; the circle meets the x-axis at $h - s$ and $h + s$, so

$$N_i = \mathbf{1}[\,0 \le h + s \le \ell_i\,] + \mathbf{1}[\,0 \le h - s \le \ell_i\,], \qquad \int_{-\infty}^{\infty} N_i \, dh = \ell_i + \ell_i = 2\ell_i.$$

Integrating over $k \in (-1, 1)$ gives $\iint N_i = 4\ell_i$. Since $N = 0$ off H, $\iint_H N = \sum_i 4\ell_i = 4L$.

Step 4: the extremum.

$$\frac{1}{F(P)} \iint_H N = \frac{4L}{\pi + \frac{3}{2}L} = \frac{8}{3 + 2\pi/L} < \frac{8}{3}.$$

This increases with L. And L is unbounded: an isosceles triangle circumscribing C with apex angle 2θ has its apex at distance $1/\sin\theta$ from O and tangent lengths $\cot\theta$ from the apex, so $L > 2\cot\theta \to \infty$ as $\theta \to 0^+$. So the supremum of the ratio is $8/3$, not attained. Every $u \geq 8/3$ works (with $\leq$ or with $<$), and no $u < 8/3$ does. Hence

$$u = \boxed{\tfrac{8}{3}}. \qquad \square$$

(Monte Carlo check on the square, $L = 8$: $\iint N \approx 31.97$ against 32, $F \approx 15.135$ against $\pi + 12 = 15.142$.)

Layer 2 THE HOW — the bifurcation

Bifurcating point. Both chains agree on what N counts. They split at the first attempt to integrate it. (N) partitions the plane by the *value* of N and integrates region by region. (M) partitions the *curve* into its sides, writes $N = \sum_i N_i$, and integrates one side at a time, where Fubini along the side gives $4\ell_i$ in two lines.

> *N is a count of intersection points, and counts add over pieces of the curve. I do not care where the regions $N = 2$ and $N = 4$ are; I integrate each side's count separately and add.*

> *For one segment on the x-axis, slide the centre horizontally: each of the two crossing points $h \pm s$ sits on the segment for a stretch of length exactly ℓ. So the horizontal integral is 2ℓ whatever the height, and the height ranges over an interval of length 2.*

> *Both F and $\iint N$ turned out to depend only on the perimeter. So the supremum over all polygons is a one-variable question in L.*

- ▷ **Cue:** an integral over positions of the number of intersections of a moving curve with a fixed one $\longrightarrow$ **Move:** write the count as a sum over pieces of the fixed curve (linearity), then integrate each piece by Fubini along it
- ▷ **Cue:** a polygon all of whose sides touch a fixed circle of radius r $\longrightarrow$ **Move:** area $= \frac{1}{2}rL$; distances to sides average to r in the weighted sense $\sum \ell_i d_i = rL$
- ▷ **Cue:** area of all points within distance 1 outside a convex polygon $\longrightarrow$ **Move:** rectangles on the sides plus sectors at the vertices: $L + \pi$

Layer 3 THE WHY

The official solution computes each side's contribution by a picture (two semicircles and two rectangles) and Cavalieri's principle. Here is where the side-by-side decomposition itself comes from.

Look at the definition: N is "the number of points common to P and the unit circle". A number of points is a sum of indicator functions, one per point, and the points live on the pieces of P. So N is additive over any decomposition of P into pieces meeting at finitely many points — and the integral is additive too. That is not a trick; it is the only structure N has. Partitioning the plane by the value of N throws that structure away and replaces it with the geometry of where circles overlap, which is exactly what makes the natural attempt collapse.

Once you integrate one side at a time, the answer 4ℓ is forced by translation invariance. Translating the centre along the segment's direction moves the crossing points rigidly; each one spends a length ℓ on the segment. This is the same mechanism as Buffon's needle: the expected number of crossings of a randomly placed curve with a fixed one is proportional to the product of their lengths (Poincaré's formula $\int n\,dg = 4L_1L_2$ over rigid motions; here the moving curve is a circle of length 2π, which is rotation invariant, so dividing by 2π for the rotation leaves $4L$).

You need none of that theory to solve the problem, but it tells you in advance that the answer will be a constant times the perimeter.

Why does F also depend only on L? Because of the tangency hypothesis. The incircle gives area $= \frac{1}{2}L$, and the same identity from an arbitrary interior point shows every interior point is within 1 of the boundary, so the inside of H is the whole polygon. The outside strip is the Steiner formula for convex sets, $L \cdot 1 + \pi \cdot 1^2$. The tangency condition is in the problem precisely so that both quantities collapse to functions of L.

After that, the extremal problem is a monotone function of one variable, and the hard part is recognising that the supremum is approached, not attained — which is why the problem asks for the smallest u rather than a maximum.

Layer $+\alpha$ BACKTESTING

1. The cue is "the number of points common to P and the unit circle" inside a double integral. Write down: $N = \sum_{\text{sides}} N_i$, integrate each N_i.

2. Linearity of the integral over a decomposition of the curve, then Fubini along each side; then reduce the extremum to the single parameter L.

3. Hesitation comes when drawing the square and seeing regions with $N = 2, 4, 8$. Reinforce away by asking what N is made of — points on sides — rather than where its level sets lie.

4. The traps: integrating over level sets of N; forgetting the interior of P belongs to H (it does, because the inradius is 1); and claiming a maximum is attained when it is only a supremum.

5. Any integral or average of an intersection count — random lines crossing a curve, lattice translates hitting a set — is additive over pieces and usually proportional to length. Split first, compute one piece.

6. Reusable decision rule: *integrate a count by summing over the objects being counted, not by partitioning into regions of constant count.*