Problem of the Month
Problem 0: September 2022

(a) Consider the integers 392, 487, 638, and 791. For each of these integers, do the following.

   (i) Determine whether the integer is a multiple of 7.

   (ii) With the hundreds digit equal to A, the tens digit equal to B, and the units digit equal to C, compute $2A + 3B + C$.

   What do you notice?

(b) Suppose $n = ABC$ is a three-digit integer (A is the hundreds digit, B is the tens digit, and C is the units digit). Show that if ABC is a multiple of 7, then $2A + 3B + C$ is a multiple of 7.

(c) Show that if $2A + 3B + C$ is a multiple of 7, then the three-digit integer $n = ABC$ is a multiple of 7.

(d) Suppose $ABCDEF$ is a six-digit integer that has each of its digits different from 0. Show that $ABCDEF$ is a multiple of 7 if and only if $BCDEFA$ is a multiple of 7.

(e) Think of ways to generalize the fact in part (d).

Problem of the Month
Problem 0: September 2022

Hint

(b) What are the remainders when 100 and 10 are divided by 7?

(d) Try to find a condition on six-digit integers that detects divisibility by 7. There is a well-known general test for divisibility by 7, but it will be useful to find another one that is specific to six-digit integers and similar to the condition for three-digit integers implied by (b) and (c).

Problem of the Month
Solution to Problem 0: September 2022

(a) In the table below, the first column contains the four integers given in the problem, the middle column says whether that integer is a multiple of 7, and the third column contains the value of $2A + 3B + C$.

ABC	Multiple of 7?	$2A + 3B + C$
392	yes (7×56)	35
487	no	39
638	no	29
791	yes (7×113)	42

The two integers that are multiples of 7 have the property that $2A + 3B + C$ is also a multiple of 7, and the two integers that are not multiples of 7 have the property that $2A + 3B + C$ is not a multiple of 7.

(b) The integer ABC is equal to $100A+10B+C$. Observe that $100 = 7 \times 14 + 2$ and $10 = 7+3$, so we have that

$$ABC = 100A + 10B + C$$
$$= [7(14) + 2]A + (7 + 3)B + C$$
$$= 7(14A + B) + (2A + 3B + C)$$

which can then be rearranged to get $ABC - 7(14A + B) = 2A + 3B + C$.

We are assuming that ABC is a multiple of 7, so the expression $2A + 3B + C$ is equal to the difference of two multiples of 7, and so it must be a multiple of 7.

(c) Using the calculation from part (b), $ABC = 7(14A+B)+2A+3B+C$ for any three-digit positive integer ABC. If $2A + 3B + C$ is a multiple of 7, then ABC is the sum of two multiples of 7 and so must be a multiple of 7.

(d) The combined result of parts (b) and (c) is that a three-digit integer ABC is a multiple of 7 *if and only if* the integer $2A + 3B + C$ is a multiple of 7. To answer the question in this part, we will first establish a similar fact about six-digit integers.

To do this, we first divide each of 10, 10^2, 10^3, 10^4, and 10^5 by 7 to find the quotient and remainder.

$$10 = 7 + 3$$
$$10^2 = 7(14) + 2$$
$$10^3 = 7(142) + 6$$
$$10^4 = 7(1428) + 4$$
$$10^5 = 7(14285) + 5$$

Since $ABCDEF = A \times 10^5 + B \times 10^4 + C \times 10^3 + D \times 10^2 + E \times 10 + F$, we can substitute the equations above and rearrange to get that

$$ABCDEF = 7(14285A + 1428B + 142C + 14D + E) + (5A + 4B + 6C + 2D + 3E + F)$$

1

and since $7(14285A + 1428B + 142C + 14D + E)$ is a multiple of 7, a similar argument to those used in parts (b) and (c) imply that the integer $ABCDEF$ is a multiple of 7 if and only if the quantity $5A + 4B + 6C + 2D + 3E + F$ is a multiple of 7.

We will now prove that $ABCDEF$ is a multiple of 7 if and only if $BCDEFA$ is a multiple of 7. First, assume that $ABCDEF$ is a multiple of 7. By the fact established above, the integer $5A + 4B + 6C + 2D + 3E + F$ is a multiple of 7. Though it may seem like a strange observation, this implies that $3(5A + 4B + 6C + 2D + 3E + F)$ must be a multiple of 7. Expanding and grouping some terms, we have

$$
\begin{aligned}
3(5A + 4B + 6C + 2D + 3E + F) &= 15A + 12B + 18C + 6D + 9E + 3F \\
&= (14A + 7B + 14C + 7E) \\
&\quad + (A + 5B + 4C + 6D + 2E + 3F) \\
&= 7(2A + B + 2C + E) \\
&\quad + (5B + 4C + 6D + 2E + 3F + A)
\end{aligned}
$$

and so $7(2A + B + 2C + E) + (5B + 4C + 6D + 2E + 3F + A)$ is a multiple of 7. This implies that $5B + 4C + 6D + 2E + 3F + A$ is a multiple of 7, and by the fact established above, the six-digit integer $BCDEFA$ is a multiple of 7.

We now suppose that $BCDEFA$ is a multiple of 7. We have already shown above that if the digits are "cycled" to the left, then the integer obtained is also a multiple of 7. Applying this several times, we have that $CDEFAB$ is a multiple of 7, as are $DEFABC$, $EFABCD$, $FABCDE$, and finally $ABCDEF$.

As an example, you can verify for yourself that 314517 is a multiple of 7, and so are each of 145173, 451731, 517314, 173145, and 731451. Similarly, the integer 215739 is not a multiple of 7, and neither are any of 157392, 573921, 739215, 392157, and 921573.

Note: The assumption that none of the digits of $ABCDEF$ are zero implies that each of the integers obtained by cycling the digits is a six-digit integer. If we were to allow digits equal to 0, the claim remains true, but in practice we may need to include "leading zeros". For example, cycling the digits of the integer 300412 in this way would lead to 004123 (which is really a four-digit integer), then 041230, 412300, 123004, and so on.

(e) While we will not include proofs here, we will list a few ways that the result in (d) can be generalized.

- If the number of digits in an integer is any positive multiple of 6, then the result of part (d) still holds. That is, if n has $6k$ digits for some $k \geq 1$, then n is a multiple of 7 if and only if the integer obtained by cycling its digits is a multiple of 7.

- For any prime number p, other than 2 and 5, if an integer has $p - 1$ digits, then it is a multiple of p if and only if the integer obtained by cycling its digits is a multiple of p. For example, a 16-digit integer is a multiple of 17 if and only if the integer obtained by cycling its digits is a multiple of 17. The reason this fails for $p = 2$ and $p = 5$ is related to the fact that 2 and 5 are the prime factors of 10, which is the *base* of our number system.

- Suppose k is an integer. The Euler Totient Function of k, denoted by $\varphi(k)$, is defined to be the number of integers m with $1 \leq m \leq k$ such that $\gcd(m, k) = 1$. For example,

$\varphi(6) = 2$ since 1 and 5 are the only integers between 1 and 6 inclusive that have a gcd of 1 with 6. $\varphi(21) = 12$ since there are exactly 12 integers m from 1 to 21 inclusive with the property that $\gcd(m, 21) = 1$. They are $1, 2, 4, 5, 8, 10, 11, 13, 16, 17, 19, 20$.

If k is an integer that is neither a multiple of 2 nor a multiple of 5, then any $\varphi(k)$-digit integer is a multiple of k if and only if the integer obtained by rotating its digits is a multiple of k. For example, the 12-digit integer 439874621235 is a multiple of 21, and so are 398746212354, 987462123543, and so on.

One of the key observations for proving something like this is a result due to Euler that says that when a and k are nonzero integers with $\gcd(a, k) = 1$, the remainder when dividing $a^{\varphi(k)}$ by k is 1.

Problem of the Month
Problem 1: October 2022

In any triangle, there is a unique circle called its *incircle* that can be drawn in such a way that it is tangent to all three sides of the triangle. For a given triangle, the radius of its incircle is known as its *inradius* and is denoted by r.

For each side of the triangle (which is tangent to the incircle), another tangent to the incircle can be drawn in such a way that it is parallel to that side. The three sides as well as these three new tangents give a total of six tangents to the incircle. They uniquely determine a hexagon that we will call the *Seraj hexagon* of the triangle.

Finally, for a given triangle, we will denote by s its *semiperimeter*, which is defined to be half of its perimeter.

The diagram below is of a triangle showing its incircle and Seraj hexagon.
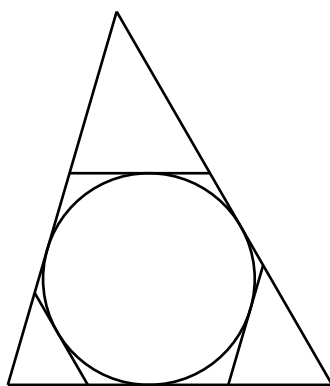
(a) Sketch the 3–4–5 triangle with its incircle and Seraj hexagon. Compute its inradius, semiperimeter, and the area of its Seraj hexagon.

(b) Find a general expression for the area of a triangle in terms only of its inradius and semiperimeter.

(c) Find a general expression for the area of the Seraj hexagon of a triangle in terms of its three side lengths, its semiperimeter, and its inradius.

(d) What is the largest possible value that can be obtained by dividing the area of a triangle's Seraj hexagon by the total area of the triangle?

Problem of the Month
Problem 1: October 2022

Hint

(a) There are several ways to compute the area of the Seraj hexagon. One is to subtract the areas of three smaller triangles from that of the full triangle.

(b) From the centre of the incircle, draw a radius to each point of tangency the circle has with the triangle.

(c) • If $\triangle ABC$ is similar to $\triangle DEF$, then $\dfrac{AB}{DE} = \dfrac{BC}{EF} = \dfrac{AC}{DF}$. It is a useful general fact that if we denote this common ratio by k, then any two corresponding altitudes of these two triangles also have a ratio of k. Can you compute the ratio of the areas of $\triangle ABC$ and $\triangle DEF$ in terms of k?

• One possible expression is

$$rs\left[1 - \left(1 - \frac{a}{s}\right)^2 - \left(1 - \frac{b}{s}\right)^2 - \left(1 - \frac{c}{s}\right)^2\right]$$

(d) Try to prove that $3(x^2 + y^2 + z^2) \geq (x + y + z)^2$ is true for all real numbers x, y, and z and determine a condition on x, y, and z that implies $3(x^2 + y^2 + z^2) = (x + y + z)^2$.

Problem of the Month
Solution to Problem 1: October 2022

Note: In this solution, we will denote by $|\triangle ABC|$ the area of $\triangle ABC$. We will also use the following facts about circles.

Fact 1: Suppose a circle has centre O and P is a point on its circumference. The tangent to the circle at point P is perpendicular to the radius from O to P.

Fact 2: For any point Q outside of a circle, there are exactly two tangents to the circle that pass through Q. If the points of tangency are A and B, then $AQ = BQ$.

(a) Below is a picture of the triangle. Its vertices are labelled by A, B, and C with the right angle at C, $AC = 3$, $BC = 4$, and $AB = 5$. The centre of the incircle is labelled by I, and AB, BC, and AC are tangent to the incircle at D, E, and F, respectively.
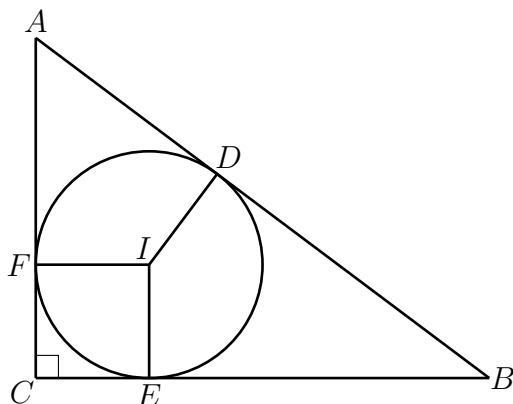
Since the perimeter is $3 + 4 + 5 = 12$, the semiperimeter is $s = \dfrac{12}{2} = 6$.

To compute r, we first use Fact 1 to get that $\angle IFC = \angle IEC = 90°$. We are assuming that $\angle FCE = 90°$ as well, and since the sum of the angles of a quadrilateral is always $360°$, $\angle FIE = 90°$. Therefore, $CFIE$ is a rectangle. Since IF and IE are radii of the incircle, $IF = IE = r$. Since $CFIE$ is a rectangle, $CF = CE = r$.

Using that $CF = CE = r$, we get $BE = BC - CE = 4 - r$ and $AF = AC - CF = 3 - r$. By Fact 2, $BD = BE$ and $AD = AF$, and since $AB = 5$,

$$
\begin{aligned}
5 &= AB \\
&= AD + BD \\
&= AF + BE \\
&= (3 - r) + (4 - r) \\
&= 7 - 2r
\end{aligned}
$$

This gives $5 = 7 - 2r$, which can be solved for r to get $r = 1$.

In the diagram below, the Seraj hexagon has been added. The side of the Seraj hexagon that is parallel to AB (and tangent to the incircle) intersects BC at L and AC at M. The

1

side of the Seraj hexagon that is parallel to BC intersects AC at G and AB at H. The side of the Seraj hexagon that is parallel to AC intersects AB at J and BC at K. The point of tangency of JK with the circle is labelled by N.
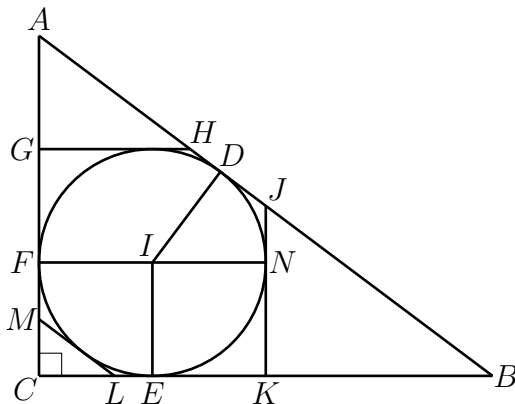
To compute the area of the Seraj hexagon, we will compute the areas of $\triangle MLC$, $\triangle AHG$, and $\triangle JBK$ and subtract their combined area from the area of $\triangle ABC$.

To compute the area of $\triangle MLC$, we first set $CM = x$ and $CL = y$. Since LM is parallel to AB, we have that $\angle CML = \angle CAB$ and $\angle CLM = \angle CBA$. Since the two share a right angle at C, $\triangle MLC$ is similar to $\triangle ABC$. Therefore, $\dfrac{y}{x} = \dfrac{CL}{CM} = \dfrac{BC}{AC} = \dfrac{4}{3}$, or $y = \dfrac{4}{3}x$.

From earlier, we have that $CF = CE = 1$, which means $FM = 1 - x$ and $EL = 1 - y$. By Fact 2, $LM = EL + FM = 2 - x - y$. By the Pythagorean theorem applied to $\triangle CML$ and using that $y = \dfrac{4}{3}x$, we have

$$CM^2 + CL^2 = LM^2$$
$$x^2 + y^2 = (2 - x - y)^2$$
$$x^2 + \left(\frac{4}{3}x\right)^2 = \left(2 - \frac{7}{3}x\right)^2$$
$$x^2 + \frac{16}{9}x^2 = 4 - \frac{28}{3}x + \frac{49}{9}x^2$$
$$0 = \frac{8}{3}x^2 - \frac{28}{3}x + 4$$
$$0 = 2x^2 - 7x + 3 \qquad \text{(multiply by } \tfrac{3}{4})$$
$$0 = (x - 3)(2x - 1)$$

If $x - 3 = 0$, then M is at A, but AB and LM are distinct parallel lines, so they have no points in common. Therefore, $2x - 1 = 0$ or $x = \dfrac{1}{2}$, so $y = \dfrac{4}{3} \times \dfrac{1}{2} = \dfrac{2}{3}$. We can now compute

$$|\triangle CML| = \frac{1}{2}xy = \frac{1}{2} \times \frac{1}{2} \times \frac{2}{3} = \frac{1}{6}$$

We now compute the area of $\triangle JBK$. By Fact 1, $\angle IEK = \angle INK = 90°$, and since JK is parallel to AC, $\angle EKN = 90°$ as well. By reasoning similar to earlier, we conclude that $INKE$ is a square of side-length r, which means $EK = r$. We also have that $CE = r$, so this means $BK = BC - CE - EK = 4 - 1 - 1 = 2$.

2

Since JK is parallel to AC, $\triangle JBK$ is similar to $\triangle ABC$ by reasoning similar to that which was used to show that $\triangle MLC$ was similar to $\triangle ABC$. This means $\dfrac{JK}{AC} = \dfrac{BK}{BC}$ or $JK = \dfrac{AC \times BK}{BC}$. Substituting $AC = 3$, $BK = 2$, and $BC = 4$ into this equation, we get $JK = \dfrac{3}{2}$. Therefore,

$$|\triangle JBK| = \frac{1}{2} \times BK \times JK = \frac{1}{2} \times 2 \times \frac{3}{2} = \frac{3}{2}$$

Using very similar reasoning, one can show that $|\triangle AHG| = \dfrac{2}{3}$. Therefore, the area of the Seraj hexagon is

$$|\triangle ABC| - |\triangle MLC| - |\triangle JBK| - |\triangle AHG| = \frac{1}{2} \times 4 \times 3 - \frac{1}{6} - \frac{3}{2} - \frac{2}{3} = \frac{11}{3}$$

(b) In the diagram below, a triangle has its incircle drawn. Radii are drawn from the centre of the incircle, I, to the points of tangency of AB, BC, and AC, which are labelled D, E, and F, respectively. As well, I is connected by line segments to each vertex of the triangle.
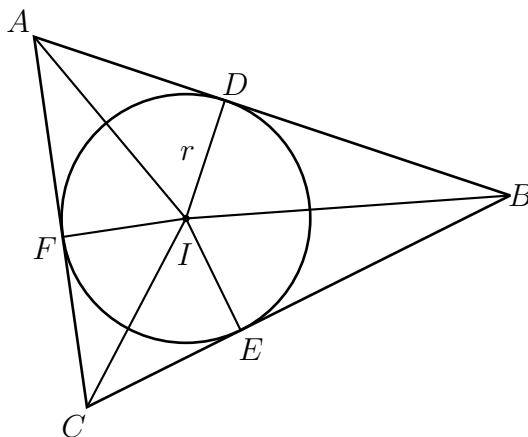

Set $AB = c$, $BC = a$, and $AC = b$. With this notation, the semiperiemter of $\triangle ABC$ is $s = \dfrac{a + b + c}{2}$. Since ID, IE, and IF are radii to points of tangency, they are perpendicular to AB, BC, and AC, respectively. Therefore, they are altitudes of $\triangle ABI$, $\triangle BCI$, and $\triangle CAI$, respectively. These three triangles compose the entirety of $\triangle ABC$ with no overlap, so the area of $\triangle ABC$ is equal to the sum of their areas. Therefore,

$$\begin{aligned}
|\triangle ABC| &= \frac{1}{2} \times AB \times ID + \frac{1}{2} \times BC \times IE + \frac{1}{2} \times AC \times IF \\
&= \frac{1}{2}cr + \frac{1}{2}ar + \frac{1}{2}br \\
&= r \times \frac{a + b + c}{2} \\
&= rs
\end{aligned}$$

As an example demonstrating this formula, consider $\triangle ABC$ from part (a). Using the usual area formula, $|\triangle ABC| = \dfrac{1}{2} \times 3 \times 4 = 6$. From part (a), its semiperimeter is $\dfrac{3 + 4 + 5}{2} = 6$ and inradius is $r = 1$, so $rs = 1 \times 6 = 6$, which is the correct area.

3

(c) In the diagram below, $\triangle ABC$ has its incircle and Seraj hexagon drawn. The side of the Seraj hexagon that is parallel to BC intersects AB at D and AC at E. The side parallel to AB intersects AC at F and BC at G. The side parallel to AC intersects BC at H and AB at J. The altitude of $\triangle ABC$ from A is also drawn and its points of intersection with BC and DE are labelled by K and L, respectively[1].
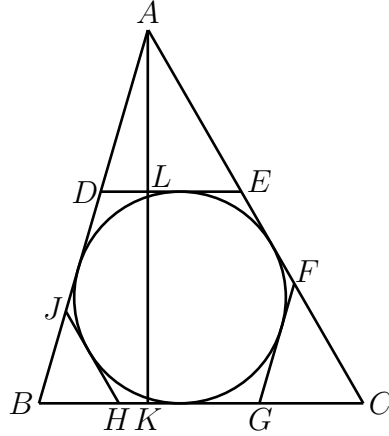
By construction, DE is parallel to BC. By reasoning similar to that which was used in earlier parts, this means $\triangle ABC$ is similar to $\triangle ADE$ and $\triangle ABK$ is similar to $\triangle ADL$. These two pairs of similar triangles imply that $\dfrac{DE}{BC} = \dfrac{AD}{AB}$ and $\dfrac{AD}{AB} = \dfrac{AL}{AK}$. We will call this common ratio k. The area of $\triangle ADE$ can be computed in terms of k and the area of $\triangle ABC$ as follows

$$
\begin{aligned}
|\triangle ADE| &= \frac{1}{2}(DE)(AL) \\
&= \frac{1}{2}\left(k(BC)\right)\left(k(AK)\right) \\
&= k^2 \left(\frac{1}{2} \times BC \times AK\right) \\
&= k^2 |\triangle ABC|
\end{aligned}
$$

We next examine the quantity k. Since DE and BC are different parallel tangents to the incircle and KL is perpendicular to both lines, the length of KL must be equal to the diameter of the incircle (you may want to think about why this is true). The diameter of the incircle is $2r$, so $KL = 2r$, which means $AL = AK - 2r$.

Therefore, we have $k = \dfrac{AL}{AK} = \dfrac{AK - 2r}{AK} = 1 - \dfrac{2r}{AK}$. From part (b), $|\triangle ABC| = rs$, but from the usual formula for the area of a triangle we also have $|\triangle ABC| = \dfrac{1}{2}(BC)(AK)$. This implies $2rs = (BC)(AK)$ or $AK = \dfrac{2rs}{BC}$. Substituting into the formula for k above, we have that

$$
k = 1 - \frac{2r}{AK} = 1 - \frac{2r(BC)}{2rs} = 1 - \frac{BC}{s}
$$

---

[1]If $\angle ABC$ or $\angle ACB$ is obtuse, then AK intersects some extensions DE and BC and not the line segments themselves. We leave it to the reader to verify that the argument that follows works even if AK does not pass through DE and BC.

If we set $AB = c$, $BC = a$, and $AC = b$, then we get $k = 1 - \dfrac{a}{s}$. Earlier, we showed that $|\triangle ADE| = k^2|\triangle ABC|$, so

$$|\triangle ADE| = \left(1 - \frac{a}{s}\right)^2 |\triangle ABC|$$

By similar reasoning, we also have

$$|\triangle JBH| = \left(1 - \frac{b}{s}\right)^2 |\triangle ABC|$$

$$|\triangle FGC| = \left(1 - \frac{c}{s}\right)^2 |\triangle ABC|$$

The area of the Seraj hexagon is equal to the area of $\triangle ABC$ minus the combined area of these three triangles, so using the formulas just above as well as $|\triangle ABC| = rs$, we have

$$|DEFGHJ| = |\triangle ABC| - |\triangle ADE| - |\triangle JBH| - |\triangle FGC|$$

$$= |\triangle ABC| \left[1 - \left(1 - \frac{a}{s}\right)^2 - \left(1 - \frac{b}{s}\right)^2 - \left(1 - \frac{c}{s}\right)^2\right]$$

$$= rs \left[1 - \left(1 - \frac{a}{s}\right)^2 - \left(1 - \frac{b}{s}\right)^2 - \left(1 - \frac{c}{s}\right)^2\right]$$

(d) As suggested in the hint, we will show that $3(x^2 + y^2 + z^2) \geq (x + y + z)^2$ for all real numbers x, y, and z. To see this, note that the inequality is equivalent to the inequality

$$3(x^2 + y^2 + z^2) - (x + y + z)^2 \geq 0$$

which, after expanding and rearranging the left side, is the same as

$$2(x^2 + y^2 + z^2) - 2(xy + yz + zx) \geq 0$$

After further manipulation, this is equivalent to

$$(x - y)^2 + (y - z)^2 + (z - x)^2 \geq 0$$

Therefore, the given inequality is true exactly when $(x - y)^2 + (y - z)^2 + (z - x)^2 \geq 0$. The sum of three squares is always non-negative, so this inequality is always true. Therefore, the given inequality is true for all real numbers x, y, and z, as claimed. Moreover, the only way that the quantity $(x - y)^2 + (y - z)^2 + (z - x)^2$ can be equal to 0 is when $x = y = z$, and hence, the only way that $3(x^2 + y^2 + z^2) = (x + y + z)^2$ is when $x = y = z$.

Dividing the expression for the area of the Seraj hexagon from part (c) by rs, we get that the ratio of the area of the Seraj hexagon to that of the triangle is

$$1 - \left(1 - \frac{a}{s}\right)^2 - \left(1 - \frac{b}{s}\right)^2 - \left(1 - \frac{c}{s}\right)^2$$

Using the inequality established above, we can multiply by $-\dfrac{1}{3}$ to get that for any real numbers x, y, and z, $-x^2 - y^2 - z^2 \leq -\dfrac{1}{3}(x + y + z)^2$. Applying this with $x = 1 - \dfrac{a}{s}$,

5

$y = 1 - \dfrac{b}{s}$, and $z = 1 - \dfrac{c}{s}$, we get that

$$1 - \left(1 - \frac{a}{s}\right)^2 - \left(1 - \frac{b}{s}\right)^2 - \left(1 - \frac{c}{s}\right)^2 = 1 - x^2 - y^2 - z^2$$

$$\leq 1 - \frac{1}{3}(x + y + z)^2$$

$$= 1 - \frac{1}{3}\left(1 - \frac{a}{s} + 1 - \frac{b}{s} + 1 - \frac{c}{s}\right)^2$$

$$= 1 - \frac{1}{3}\left(3 - \frac{a + b + c}{s}\right)^2$$

$$= 1 - \frac{1}{3}\left(3 - \frac{2s}{s}\right)^2$$

$$= 1 - \frac{1}{3}$$

$$= \frac{2}{3}$$

Therefore, the ratio is at most $\dfrac{2}{3}$ in every triangle. By the remark at the end of the proof of the inequality, the ratio equals $\dfrac{2}{3}$ exactly when $1 - \dfrac{a}{s} = 1 - \dfrac{b}{s} = 1 - \dfrac{c}{s}$. Since s is always nonzero, this is equivalent to $a = b = c$. Therefore, the ratio is maximized when the triangle is equilateral. It is not difficult to explicitly show that the area of the Seraj hexagon in an equilateral triangle is equal to two thirds of the area of the triangle.

Problem of the Month
Problem 2: November 2022

Let $\phi = \dfrac{1+\sqrt{5}}{2} \approx 1.61803$. For integers $d_k, d_{k-1}, \ldots, d_1, d_0, d_{-1}, \ldots, d_{-r}$, each equal to 0 or 1, the expression

$$(d_k d_{k-1} \cdots d_2 d_1 d_0 . d_{-1} d_{-2} \cdots d_{-r})_\phi$$

is called a *base ϕ expansion* and represents the real number

$$d_k \phi^k + d_{k-1} \phi^{k-1} + \cdots + d_1 \phi + d_0 + d_{-1} \phi^{-1} + d_{-2} \phi^{-2} + \cdots + d_{-r} \phi^{-r}$$

The integers d_k through d_{-r} are called the *digits* of the expansion. For example, the base ϕ expansion 1101.011_ϕ represents the real number

$$(1 \times \phi^3) + (1 \times \phi^2) + (0 \times \phi) + 1 + (0 \times \phi^{-1}) + (1 \times \phi^{-2}) + (1 \times \phi^{-3})$$

which can be simplified to get

$$
\begin{aligned}
\phi^3 + \phi^2 + 1 + \frac{1}{\phi^2} + \frac{1}{\phi^3} &= \left(\frac{1+\sqrt{5}}{2}\right)^3 + \left(\frac{1+\sqrt{5}}{2}\right)^2 + 1 + \left(\frac{2}{1+\sqrt{5}}\right)^2 + \left(\frac{2}{1+\sqrt{5}}\right)^3 \\
&= \frac{16+8\sqrt{5}}{8} + \frac{6+2\sqrt{5}}{4} + 1 + \frac{4}{6+2\sqrt{5}} + \frac{8}{16+8\sqrt{5}} \\
&= (2+\sqrt{5}) + \left(\frac{3}{2} + \frac{1}{2}\sqrt{5}\right) + 1 + \left(\frac{3}{2} - \frac{1}{2}\sqrt{5}\right) - (2 - \sqrt{5}) \\
&= 4 + 2\sqrt{5}
\end{aligned}
$$

and so $1101.011_\phi = 4 + 2\sqrt{5}$.

(a) What are the real numbers represented by 1011_ϕ and 10000_ϕ?

(b) Find a base ϕ expansion of the real number $4 + 3\sqrt{5}$.

(c) Show that $\phi^2 = \phi + 1$ and use this to deduce that $\phi^{n+1} = \phi^n + \phi^{n-1}$ for all integers n.

(d) Show that every positive integer has a base ϕ expansion and find a base ϕ expansion for each positive integer from 1 through 10. One approach is to prove and use the following two facts.

- If a real number n has a base ϕ expansion, then it has a base ϕ expansion that does not have two consecutive digits equal to 1.

- If a real number n has a base ϕ expansion, then it has a base ϕ expansion that has its units digit, d_0, equal to 0.

Problem of the Month
Problem 2: November 2022

Hint

(b) What is $\phi + \dfrac{1}{\phi}$?

(d) One way to prove the first fact is to show that if there are two consecutive digits equal to 1, then there is a base ϕ expansion with fewer non-zero digits. To prove the second fact, use the same idea as in the proof of the first fact, but in reverse.

Problem of the Month
Solution to Problem 2: November 2022

(a) Expanding and simplifying, we have

$$1011_\phi = \phi^3 + \phi + 1 = \left(\frac{1+\sqrt{5}}{2}\right)^3 + \frac{1+\sqrt{5}}{2} + 1$$

$$= \frac{16 + 8\sqrt{5}}{8} + \frac{1+\sqrt{5}}{2} + 1$$

$$= \frac{4 + 2\sqrt{5}}{2} + \frac{1+\sqrt{5}}{2} + \frac{2}{2}$$

$$= \frac{7 + 3\sqrt{5}}{2}$$

$$10000_\phi = \phi^4 = \left(\frac{1+\sqrt{5}}{2}\right)^4$$

$$= \left(\frac{1+\sqrt{5}}{2}\right)^2 \left(\frac{1+\sqrt{5}}{2}\right)^2$$

$$= \left(\frac{6 + 2\sqrt{5}}{4}\right) \left(\frac{6 + 2\sqrt{5}}{4}\right)$$

$$= \frac{56 + 24\sqrt{5}}{16}$$

$$= \frac{7 + 3\sqrt{5}}{2}$$

and so we see that $1011_\phi = 10000_\phi$

(b) There are several ways to approach this problem and we will demonstrate two of them. First, from the example given in the problem statement, we have that

$$4 + 2\sqrt{5} = \phi^3 + \phi^2 + 1 + \frac{1}{\phi^2} + \frac{1}{\phi^3}$$

and using the hint, we have that

$$\phi + \frac{1}{\phi} = \frac{1+\sqrt{5}}{2} + \frac{2}{1+\sqrt{5}} = \frac{1+\sqrt{5}}{2} + \frac{2(\sqrt{5}-1)}{4} = \sqrt{5}$$

and so

$$4 + 3\sqrt{5} = (4 + 2\sqrt{5}) + \sqrt{5}$$

$$= \left(\phi^3 + \phi^2 + 1 + \frac{1}{\phi^2} + \frac{1}{\phi^3}\right) + \left(\phi + \frac{1}{\phi}\right)$$

$$= 1111.111$$

1

Another approach is to use what might be called a "greedy algorithm". We first note that $4 + 3\sqrt{5} \approx 10.708204$ and then consider powers of ϕ and find the first one that does not exceed it. As it turns out, $\phi^4 \approx 6.854102$ and $\phi^5 \approx 11.090170$. It can be checked that $\phi^4 = \dfrac{7 + 3\sqrt{5}}{2}$, so if we let $\alpha = 4 + 3\sqrt{5}$, then

$$\alpha - \phi^4 = \frac{8 + 6\sqrt{5}}{2} - \frac{7 + 3\sqrt{5}}{2} = \frac{1 + 3\sqrt{5}}{2}$$

The approximate value of $\alpha - \phi^4$ is 3.854102. Now we check that $\phi^2 \approx 2.618034$ and $\phi^3 \approx 4.236068$, so the largest power of ϕ that does not exceed $\alpha - \phi^4$ is ϕ^2, and it can be checked that

$$\alpha - \phi^4 - \phi^2 = -1 + \sqrt{5} \approx 1.236068$$

Since $\phi^0 = 1$ and $\phi^1 = \phi \approx 1.618034$, the largest power of ϕ that does not exceed $\alpha - \phi^4 - \phi^2$ is $\phi^0 = 1$, and so we now consider the quantity

$$\alpha - \phi^4 - \phi^2 - 1 = -2 + \sqrt{5} \approx 0.236068$$

Considering ϕ taken to negative integer exponents, we find that $\phi^{-3} \approx 0.236068$, so we suspect that $-2 + \sqrt{5}$ is exactly equal to ϕ^{-3}. Indeed,

$$\left(\frac{2}{1 + \sqrt{5}}\right)^3 = \frac{8}{16 + 8\sqrt{5}} = \frac{1}{2 + \sqrt{5}} = -2 + \sqrt{5}$$

Therefore, $\alpha - \phi^4 - \phi^2 - 1 = \dfrac{1}{\phi^3}$ which can be rearranged to get $\alpha = \phi^4 + \phi^2 + \phi^0 + \phi^{-3}$, which means $4 + 3\sqrt{5} = 10101.001$. Notice that these two base ϕ expansions are different and both correct. A way to see that the expansions are equal is explained throughout the rest of the solution.

(c) Expanding ϕ^2 and simplifying, we have

$$\phi^2 = \left(\frac{1 + \sqrt{5}}{2}\right)^2$$
$$= \frac{6 + 2\sqrt{5}}{4}$$
$$= \frac{3 + \sqrt{5}}{2}$$
$$= \frac{1 + \sqrt{5}}{2} + \frac{2}{2}$$
$$= \phi + 1$$

Multiplying this equation by ϕ^{n-1} gives $\phi^{n-1} + \phi^n = \phi^{n+1}$. It will be useful in part (d) to note that this means that $011_\phi = 100_\phi$ and more generally, if 011 ever occurs in a base ϕ expansion, it can be replaced by 100, or vice versa, without changing the actual value of the number being expressed.

(d) As suggested in the problem, we will use the facts in the bullet points. Specifically, we will use Fact 1 and Fact 2 below:

2

Fact 1: If a real number n has a base ϕ expansion, then it has a base ϕ expansion that does not have two consecutive digits equal to 1.

Fact 2: If a real number n has a base ϕ expansion, then it has a base ϕ expansion that has its units digit d_0 equal to 0.

Proof of Fact 1. Let $n = d_k d_{k-1} \cdots d_2 d_1 d_0.d_{-1}d_{-2} \cdots d_{-r}$ be a base ϕ expansion of n. We will call the digit d_m *bad* if $d_m = 1$ and at least one of d_{m+1} and d_{m-1} is equal to 1. We want to show that n has a base ϕ expansion that has no bad digits.

Observe that if n has an expansion with every digit equal to 0, then n must itself be the real number 0. This is because every power of ϕ is positive.

Suppose the given expansion of n has at least one bad digit. Then we let m be the largest integer such that d_m is bad. By definition, $d_m = 1$ and either $d_{m+1} = 1$ or $d_{m-1} = 1$. However, it is not possible for $d_{m+1} = 1$ since this would imply that d_{m+1} is bad, but m is the largest integer such that d_m is bad.

Thus, we must have $d_{m+1} = 0$, $d_m = 1$, and $d_{m-1} = 1$, and so the given expansion is

$$n = d_k d_{k-1} \cdots d_{m+2} 011 d_{m-2} \cdots d_1 d_0.d_{-1}d_{-2} \cdots d_{-r}$$

By the remark at the end of part (c), we can replace the 011 by 100 to get that

$$n = d_k d_{k-1} \cdots d_{m+2} 100 d_{m-2} d_{m-2} \cdots d_1 d_0.d_{-1}d_{-2} \cdots d_{-r}$$

Notice that we have found a base ϕ expansion of n with fewer non-zero digits, under the assumption that the expansion had at least one bad digit. In other words, if a base ϕ expansion of n has at least one bad digit, then there is a base ϕ expansion of n that has fewer non-zero digits. Thus, as long as there is a bad digit, we can decrease the number of non-zero digits.

By the above remark, unless $n = 0$, the number of nonzero digits cannot decrease indefinitely since there were finitely many digits to begin with. Thus, we must eventually lose the ability to decrease the number of nonzero digits, and by the previous paragraph, this means we must eventually reach a base ϕ expansion that has no bad digits.

$\square$

Below is an example of the procedure explained in the proof of Fact 1. It is used to find a base ϕ expansion of $n = 1010101110011.1101_\phi$ that has no bad digits. There is a step where a leading 11 is replaced by 100. The digits are aligned in the equations to indicate where this happened. In every line except the last, the the digits that are changed in the

subsequent line are highlighted in red.

$$
\begin{aligned}
n &= 1010101110011.1101_\phi \\
&= 1010110010011.1101_\phi \\
&= 1011000010011.1101_\phi \\
&= 1100000010011.1101_\phi \\
&= 10000000010011.1101_\phi \\
&= 10000000010100.1101_\phi \\
&= 10000000010101.0001_\phi
\end{aligned}
$$

Proof of Fact 2. This can be shown by an argument rather similar to the proof of Fact 1. This time, we will introduce bad digits in order to remove a potential 1 from the units digit. Suppose n has a base ϕ expansion. By Fact 1, we can assume that the expansion $n = d_k d_{k-1} \cdots d_2 d_1 d_0 d_{-1} \cdots d_{-r}{}_\phi$ is a expansion without any bad digits. If $d_0 = 0$, then there is nothing to do, so we assume that $d_0 = 1$. We will now artificially add two "trailing" 0's as digits. That is, we also have that

$$
n = d_k d_{k-1} \cdots d_2 d_1 d_0 d_{-1} \cdots d_{-r} d_{-r-1} d_{-r-2}
$$

where $d_{-r-1} = d_{-r-2} = 0$. Let m be the largest integer with $0 > m$ and $d_m = d_{m-1} = 0$. Since we have added two trailing zeros, we know this must happen somewhere, so it is possible to choose m this way.

By the choice of m, we must have that $d_{m+1} = 1$, and since the expansion has no bad digits, $d_{m+2} = 0$. By the choice of m, it then follows that $d_{m+3} = 1$, then since there are no bad digits, we get $d_{m+4} = 0$, and so on. Since $d_0 = 1$, this alternating pattern must eventually reach $d_0 = 1$. In other words, the expansion takes the form

$$
n = d_k d_{k-1} \cdots d_3 d_2 d_1 1.010101 \cdots 010100 d_{m-2} d_{m-3} \cdots d_{-r} d_{-r-1} d_{-r-2}
$$

By part (c), we can change d_m and d_{m-1} to 1's and compensate by changing d_{m+1} to 0. The new expansion is

$$
n = d_k d_{k-1} \cdots d_3 d_2 d_1 1.010101 \cdots 010011 d_{m-2} d_{m-3} \cdots d_{-r} d_{-r-1} d_{-r-2}
$$

Repeating this process, we now change d_{m+1} and d_{m+2} to 1's and change d_{m+3} to a 0. This process terminates in the expansion

$$
n = d_k d_{k-1} \cdots d_3 d_2 d_1 0.111111 \cdots 111 d_{m-2} d_{m-3} \cdots d_{-r}
$$

which indeed has $d_0 = 0$. $\qquad\square$

The procedure in the proof of Fact 2 is used below to find a base ϕ expansion of the number represented as $n = 100101001.01010101001010010001_\phi$ that has $d_0 = 0$. As was done earlier, in each line but the last, the three digits highlighted in red are the ones that

change in the subsequent line.

$$n = 100101001.01010101001010010001_\phi$$
$$n = 100101001.01010100111010010001_\phi$$
$$n = 100101001.01010011111010010001_\phi$$
$$n = 100101001.01001111111010010001_\phi$$
$$n = 100101001.00111111111010010001_\phi$$
$$n = 100101000.11111111111010010001_\phi$$

Now we will apply the facts to derive base ϕ expansions for the first ten positive integers.

To start, we have that $1 = 1_\phi$ since $1 = \phi^0$. This means the integer 1 has a base ϕ expansion. Following the procedure outlined in the proof of Fact 2, $1 = 0.11_\phi$. It is easy to add 1 to a base ϕ expansion that has $d_0 = 0$ because we can simply change $d_0 = 1$. Thus, we have that $2 = 1.11_\phi$, and using the procedure from the proof of Fact 1 again, we get $2 = 10.01_\phi$.

Continuing in this way, if we have a base ϕ expansion of the integer n, then we can use the technique in the proof of Fact 1 to find a base ϕ expansion that does not have any bad digits. Then, if necessary, we can use the technique in the proof of Fact 2 to find a base ϕ expansion of n that has $d_0 = 0$. From this, we can find a base ϕ expansion of $n+1$ by taking the base ϕ expansion of n and switching d_0 from 0 to 1, which has the effect of adding 1. This process can be repeated to find a base ϕ expansion of $n+2$, then $n+3$, and so on. Starting with $n = 1$, this is demonstrated up to $n = 10$ below.

$$1 = 1_\phi$$
$$= 1.00_\phi$$
$$= 0.11_\phi$$

$$2 = 1.11_\phi$$
$$= 10.01_\phi$$

$$3 = 11.01_\phi$$
$$= 100.01_\phi$$

$$4 = 101.01_\phi$$
$$= 101.0100_\phi$$
$$= 101.0011_\phi$$
$$= 100.1111_\phi$$

$$5 = 101.1111_\phi$$
$$= 110.0111_\phi$$

$$6 = 111.0111_\phi$$
$$= 1001.1001_\phi$$
$$= 1010.0001_\phi$$

$$7 = 1011.0001_\phi$$
$$= 1100.0001_\phi$$

$$8 = 1101.0001_\phi$$
$$= 10000.1101_\phi$$

$$9 = 10001.1101_\phi$$
$$= 10010.0101_\phi$$

$$10 = 10011.0101_\phi$$
$$= 10100.0101_\phi$$

Problem of the Month
Problem 3: December 2022

This month's problem is an extension of Problem 6 from the November 2022 Canadian Senior Mathematics Contest. Here is the original problem.

> *A bag contains exactly 15 marbles of which 3 are red, 5 are blue, and 7 are green. The marbles are chosen at random and removed one at a time from the bag until all of the marbles are removed. One colour of marble is the first to have 0 remaining in the bag. What is the probability that this colour is red?*

Note: It might be useful to familiarize yourself with the notation of *binomial coefficients* before attempting this problem.

(a) Suppose there are r red marbles and b blue marbles. As in the original problem, the marbles are chosen at random and removed from the bag one at a time until all marbles are removed. One colour of marble is the first to have 0 marbles remaining in the bag. What is the probability that this colour is red?

(b) Suppose there are r red marbles, b blue marbles, and g green marbles. The marbles are chosen at random and removed one at a time until all marbles are removed. What is the probability that red is the colour of marble that is first to be completely removed from the bag?

(c) Suppose there are r red marbles, b blue marbles, and g green marbles with $r < b < g$. Let $p(r)$ be the probability that the red marbles are the first to be completely removed from the bag and define $p(b)$ and $p(g)$ similarly. Determine which of $p(r)$, $p(b)$, and $p(g)$ is the smallest and which is the largest. Does the result agree with your intuition?

(d) Show that the values of $p(r)$, $p(b)$, and $p(g)$ depend only on the proportions of r, b, and g to the total number of marbles. For example, if one bag has r red, b blue, and g green marbles and another has $7r$ red, $7b$ blue, and $7g$ green marbles, then the probability that the red are removed first is the same for both bags.

Problem of the Month
Problem 3: December 2022

Hint

(a) If the red marbles are all removed from the bag first, then what is the colour of the last marble to be removed from the bag?

(b) As in part (a), it is useful to think about the colour of the final marble to be removed. In part (a), the final colour alone determines which colour has completely removed first. In this part, it is a bit more complicated.

(c) Try to find a general expression for each of $p(r)$, $p(b)$, and $p(g)$, simplifying as much as possible. Once you have done this, try to determine the sign of $p(r) - p(b)$.

(d) One way to approach this is to compute the probabilities directly in terms of the proportions. Thinking about which colour is removed last will likely be helpful, as in earlier parts. Another approach is to use a general formula for the probabilities in terms of r, b, and g, and divide the numerator and denominator by $(r + b + g)^2$.

Problem of the Month
Solution to Problem 3: December 2022

(a) Suppose there are r red marbles and b blue marbles and set $n = r + b$. Upon removing the n marbles from the bag in random order, we can record the order in which they emerged from the bag by a sequence of R's and B's. For example, if the first two marbles were red and the third was blue, then the sequence would begin with RRB. If the fourth was red, then the sequence would continue $RRBR$, and so on.

An important observation in this part and later parts of the solution is that every possible sequence of n characters, r of which are R and b of which are B, is equally likely to occur through this process. To make this observation, we first imagine labelling the red marbles by R_1, R_2, and so on through to R_r, and the blue marbles by B_1, B_2, and so on through to B_b. With this labelling, we have made it so that the n marbles are distinct. There are exactly $n!$ orders in which the marbles can be removed, and each of them is equally likely. If we now imagine listing these $n!$ possible orderings of $R_1, \ldots, R_r, B_1, \ldots, B_b$ and removing the subscripts, each possible sequence of r R's and b B's will occur in the list exactly $(r!)(b!)$ times. Therefore, if we remove the (unindexed) marbles and write down the sequence of R's and B's, each possible sequence will occur with probability exactly $\dfrac{r!b!}{n!} = \dfrac{1}{\binom{n}{r}}$. Hence, every sequence occurs with the same probability. Notice that $\binom{n}{r}$ is equal to the number of sequences of r R's and b B's since any such sequence is completely determined by selecting r of the n possible positions and placing R in them.

Therefore, the probability that the red marbles are removed first is equal to the number of sequences of r R's and b B's with the property that there is at least one B to the right of the rightmost R, divided by $\binom{n}{r}$.

Since there are only two colours in this part of the problem, the statement "there is at least one B to the right of the rightmost R" is equivalent to "the final letter in the sequence is B". Put in a different way, the red marbles are completely removed first if and only if a blue marble is removed last. Therefore, the probability that we seek is simply the number of sequences of R's and B's that end in B, divided by $\binom{n}{r}$. The number of such sequences is equal to $\binom{n-1}{r}$ since, to construct all such sequences, we can place one B at the end, and then place the r R's in any of the other $n - 1$ positions. Therefore, the probability

that all the red marbles are removed first is

$$\frac{\binom{n-1}{r}}{\binom{n}{r}} = \frac{\dfrac{(n-1)!}{r!(n-r-1)!}}{\dfrac{n!}{r!(n-r)!}}$$

$$= \frac{(n-1)!r!(n-r)!}{r!(n-r-1)!n!}$$

$$= \frac{n-r}{n}$$

$$= \frac{b}{r+b}$$

Looking ahead to later parts, notice that this probability is equal to the proportion of the marbles that are blue.

(b) Consider the following two events:

- Event X: The final marble that is removed is blue and the final green marble is removed after the final red marble is removed.

- Event Y: The final marble that is removed is green and the final blue marble is removed after the final red marble is removed.

To say that the red marbles were removed first is the same as saying either Event X occurred or Event Y occurred. Since Event X has blue as the final marble and Event Y has green as the final marble, Event X and Event Y cannot occur simultaneously, so this means that the probability the red marbles are removed first is the sum of the probabilities of Event X and Event Y. Therefore, we can compute the desired probability by computing the probabilities of Events X and Y and adding the results together.

For both of these probabilities, we need to count the total number of ways that the marbles can be removed. As in part (a), we will think of an order that the marbles can be removed in as a sequence of r R's, b B's, and g G's. There are $\binom{r+b+g}{r}$ ways to place the r R's. After doing this, there will be $b+g$ empty positions, so there are $\binom{b+g}{b}$ ways to place the b B's. Once the R's and B's are placed, there is no choice of where to place the G's, so the total number of sequences is

$$\binom{r+b+g}{r}\binom{b+g}{b} = \frac{(r+b+g)!(b+g)!}{r!(b+g)!b!g!} = \frac{(r+b+g)!}{r!b!g!}$$

Note: This expression is an example of something called a *multinomial coefficient*, which is a generalization of a binomial coefficient. We will not use the notation of multinomial coefficients in this solution, but you might like to read about them if you have not before.

We will compute the number of sequences that end in B and have at least one G to the right of the rightmost R. This quantity divided by $\dfrac{(r+b+g)!}{r!b!g!}$ is the probability that Event X occurs.

2

There are $\binom{r+b+g-1}{b-1}$ ways to place the B's so that the final letter in the sequence is B. This is because we need to place one B in the final position, then place the $b-1$ remaining B's in the remaining $r+b+g-1$ positions.

For any such placement of the B's, in order to place the R's and G's so that at least one G occurs to the right of the rightmost R, we need to place a G in the rightmost empty position. Thus, for every such way to place the B's, we now have $\binom{r+g-1}{g-1}$ ways to place the G's. This is because we must place one G in the rightmost position that is not occupied by a B, and then the remaining $g-1$ G's can be placed in any of the remaining $r+g-1$ positions. Once the B's and G's are placed, there is no choice of where to place the R's, so the probability that Event X occurs is

$$\frac{r!b!g!}{(r+b+g)!}\binom{r+b+g-1}{b-1}\binom{r+g-1}{g-1} = \frac{r!b!g!}{(r+b+g)!} \times \frac{(r+b+g-1)!(r+g-1)!}{(b-1)!(r+g)!(g-1)!r!}$$
$$= \frac{bg}{(r+b+g)(r+g)}$$

By a very similar calculation, it can be shown that the probability that Event Y occurs is

$$\frac{bg}{(r+b+g)(r+b)}$$

and so the probability that all the red marbles are removed from the bag first is

$$\frac{bg}{(r+b+g)(r+g)} + \frac{bg}{(r+b+g)(r+b)} = \frac{bg}{r+b+g}\left(\frac{1}{r+g} + \frac{1}{r+b}\right)$$

(c) In part (b), we showed that $p(r) = \dfrac{bg}{r+b+g}\left(\dfrac{1}{r+g} + \dfrac{1}{r+b}\right)$. Following nearly identical reasoning, it can be shown that

$$p(b) = \frac{rg}{r+b+g}\left(\frac{1}{r+b} + \frac{1}{b+g}\right)$$
$$p(g) = \frac{rb}{r+b+g}\left(\frac{1}{r+g} + \frac{1}{b+g}\right)$$

Using the equations above, we will show that $p(b) - p(r)$ is negative. This result will show that $p(r)$ is greater than $p(b)$. Remember that we are assuming $r < b < g$ in this part.

$$p(b) - p(r) = \frac{rg}{r+b+g}\left(\frac{1}{r+b} + \frac{1}{b+g}\right) - \frac{bg}{r+b+g}\left(\frac{1}{r+g} + \frac{1}{r+b}\right)$$
$$= \frac{g}{r+b+g}\left(\frac{r}{r+b} + \frac{r}{b+g} - \frac{b}{r+g} - \frac{b}{r+b}\right)$$

Since r, b, and g are all positive, the quantity $\dfrac{g}{r+b+g}$ is positive, which means $p(b) - p(r)$ is negative if and only if $\dfrac{r}{r+b} + \dfrac{r}{b+g} - \dfrac{b}{r+g} - \dfrac{b}{r+b}$ is negative. Using a common denominator of $(r+b)(b+g)(r+g)$, we get

$$\frac{r(b+g)(r+g) + r(r+b)(r+g) - b(r+b)(b+g) - b(b+g)(r+g)}{(r+b)(b+g)(r+g)}$$

3

Noting that the denominator $(r+b)(b+g)(r+g)$ is also positive, we have that $p(b) - p(r)$ is negative if and only if the numerator of the fraction above is negative. Manipulating the numerator, we have

$$r(b+g)(r+g) + r(r+b)(r+g) - b(r+b)(b+g) - b(b+g)(r+g)$$
$$= r(r+g)(b+g+r+b) - b(b+g)(r+b+r+g)$$
$$= (r^2 + rg)(r+2b+g) - (b^2 + bg)(2r+b+g)$$
$$= r^3 + 2r^2b + r^2g + r^2g + 2rbg + rg^2 - 2rb^2 - b^3 - b^2g - 2rbg - b^2g - bg^2$$
$$= r^3 + 2r^2b + 2r^2g + rg^2 - 2rb^2 - b^3 - 2b^2g - bg^2$$

At first, it might seem that there is no hope of understanding the expression above. However, if we imagine the situation where $r = b$, then by symmetry we really should have that $p(r) = p(b)$ (in the problem, we are assuming that $r < b$, but the equation above is not aware that we plan to impose this restriction). This means the expression $p(b) - p(r)$ should evaluate to 0 when $r = b$, so it should have a factor of $r - b$. Indeed, if we group the terms that "look" alike, it is easier to notice the factor of $(r - b)$:

$$r^3 + 2r^2b + 2r^2g + rg^2 - 2rb^2 - b^3 - 2b^2g - bg^2$$
$$= (r^3 - b^3) + 2(r^2b - rb^2) + 2(r^2g - b^2g) + (rg^2 - bg^2)$$
$$= (r - b)(r^2 + rb + b^2) + (r - b)2rb + (r - b)2g(r + b) + (r - b)g^2$$
$$= (r - b)\left[(r^2 + rb + b^2) + 2rb + 2g(r + b) + g^2\right]$$

Since $r < b$ is given, we now have that $p(b) - p(r)$ is negative if and only if the quantity $(r^2 + rb + b^2) + 2rb + 2g(r + b) + g^2$ is positive, but each of r, b, and g is positive, so it is indeed positive!

Therefore, we have that $p(b) - p(r) < 0$, or $p(b) < p(r)$. Similar calculations can be used to show that $p(g) < p(b)$. Therefore, the marble colour with the smallest number of representatives is most likely to be removed first, and the marble colour with the largest number of representatives is least likely to be removed first.

(d) Let $f(r) = \dfrac{r}{r+b+g}$, $f(b) = \dfrac{b}{r+b+g}$, and $f(g) = \dfrac{g}{r+b+g}$. That is, $f(r)$ is the proportion of the marbles that are red, and similarly for $f(b)$ and $f(g)$.

From part (b), the probability that all the red marbles are removed first, $p(r)$, is

$$\frac{bg}{(r+b+g)(r+g)} + \frac{bg}{(r+b+g)(r+b)}$$

Dividing the numerator and denominator of each expression by $(r+b+g)^2$, the probability above is equal to

$$\frac{\dfrac{bg}{(r+b+g)^2}}{\dfrac{r+g}{r+b+g}} + \frac{\dfrac{bg}{(r+b+g)^2}}{\dfrac{r+b}{r+b+g}} = \frac{\dfrac{b}{r+b+g} \times \dfrac{g}{r+b+g}}{\dfrac{r}{r+b+g} + \dfrac{g}{r+b+g}} + \frac{\dfrac{b}{r+b+g} \times \dfrac{g}{r+b+g}}{\dfrac{r}{r+b+g} + \dfrac{b}{r+b+g}}$$
$$= \frac{f(b)f(g)}{f(r) + f(g)} + \frac{f(b)f(g)}{f(r) + f(b)}$$

So we have written the probability that all the red marbles are removed first in terms of the proportions of each of the three colours.

We will now try to explain why the formula above makes sense. Once again, we will think of a way of drawing the marbles randomly as a sequence of R's, B's, and G's. In any given position, the proportion of sequences with a B in that position is $f(b)$. In particular, $f(b)$ is the probability that the final letter in the sequence is B.

The quantity $\dfrac{f(g)}{f(r) + f(g)}$ is equal to the proportion of green marbles among those marbles that are either red or green. Similar to the reasoning above, this means that if we look at sequences of R's, B's, and G's the quantity $\dfrac{f(g)}{f(r) + f(g)}$ is the probability that G is the rightmost letter among R's and G's.

The overall rightmost letter being equal to B is independent of the probability that the rightmost letter among R's and G's is G, so the quantity

$$\frac{f(b)f(g)}{f(r) + f(g)} = f(b) \times \frac{f(g)}{f(r) + f(g)}$$

is equal to the probability that the rightmost letter in the sequence is a B and among R's and G's, the rightmost letter is G. Similarly, the quantity $\dfrac{f(b)f(g)}{f(r) + f(b)}$ is equal to the probability that the rightmost letter is G and among R's and B's, the rightmost letter is B. The R's all occur before the final G and before the final B if and only if one of these two events occurs, so this explains why the probability that all the red marbles are removed first is equal to

$$p(r) = \frac{f(b)f(g)}{f(r) + f(g)} + \frac{f(b)f(g)}{f(r) + f(b)}$$

This expression is in terms of the proportions of red, blue, and green marbles. Similar expressions can be computed for $p(b)$ and $p(g)$.

Problem of the Month
Problem 4: January 2023

For each positive integer k, define a function $p_k(n) = 1^k + 2^k + 3^k + \cdots + n^k$ for each integer n. That is, $p_k(n)$ is the sum of the first n perfect k^{th} powers. It is well known that $p_1(n) = \dfrac{n(n+1)}{2}$.

(a) Fix a positive integer n. Let S be the set of ordered triples (a, b, c) of integers between 1 and $n + 1$, inclusive, that also satisfy $a < c$ and $b < c$. Show that there are exactly $p_2(n)$ elements in the set S.

(b) With S as in part (a), show that there are $\dbinom{n+1}{2} + 2\dbinom{n+1}{3}$ elements in S and use this to show that
$$p_2(n) = \frac{n(n+1)(2n+1)}{6}$$

(c) For each k, show that there are constants $a_2, a_3, \ldots, a_k, a_{k+1}$ such that
$$p_k(n) = a_2 \binom{n+1}{2} + a_3 \binom{n+1}{3} + \cdots + a_k \binom{n+1}{k} + a_{k+1} \binom{n+1}{k+1}$$
for all n.

Note: Actually computing the constants gets more and more difficult as k gets larger. While you might want to compute them for some small k, in this problem we only intend that you argue that the constants always exist, not that you deduce exactly what they are.

(d) Use part (c) to show that $p_3(n) = \dfrac{n^2(n+1)^2}{4}$ and $p_4(n) = \dfrac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$.

(e) It follows from the fact in part (c) that $p_k(n)$ is a polynomial of degree $k + 1$. With $k = 5$, this means there are constants $c_0, c_1, c_2, c_3, c_4, c_5$, and c_6 such that
$$p_5(n) = c_0 + c_1 n + c_2 n^2 + c_3 n^3 + c_4 n^4 + c_5 n^5 + c_6 n^6$$

Use the fact that $p_5(1) = 1$ and $p_5(n) - p_5(n-1) = n^5$ for all $n \geq 2$ to determine c_0 through c_6, and hence, derive a formula for $p_5(n)$.

(f) Show that $n(n+1)$ is a factor of $p_k(n)$ for every positive integer k and that $2n+1$ is a factor of $p_k(n)$ for every even positive integer k.

Problem of the Month
Problem 4: January 2023

Hint

(a) What are the possible values of c?

(b) How many distinct integers can occur in a triple in S?

(c) Try to generalize the idea in part (c). The constants a_2 through a_{k+1} do not depend on n.

(d) For positive integers u and v with $u < v$, the usual convention is that $\binom{u}{v} = 0$. This convention makes sense for (at least) two reasons. First, there are zero ways to choose v objects from u distinct objects if $u < v$, so "u choose v" should be equal to 0. Second, the formula for $\binom{u}{v}$ given by

$$\binom{u}{v} = \frac{u(u-1)(u-2)\cdots(u-v+1)}{v!}$$

will have a factor of 0 in the numerator if $u < v$.

(e) Directly compute an expression for $p_5(n) - p_5(n-1)$. It should be a polynomial with coefficients depending on a_1 through a_6. By equating coefficients with the polynomial n^5, solve for a_1 through a_6. After these coefficients are known, a_0 can be computed from $p_5(1) = 1$.

(f) A polynomial with infinitely many roots must be the constant zero polynomial. Using this fact, show that $p_k(n) - p_k(n-1) = n^k$ for all real numbers, not just positive integers. This means you need to "extend" $p_k(n)$ to accept inputs that are not positive integers. Once this is done, determine the values of $p_k(0)$ and $p_k(-1)$. To show that $2n+1$ is a factor of $p_k(n)$ for even k, consider the values of $p_k(-n)$ when n is a positive integer.

Problem of the Month
Solution to Problem 4: January 2023

(a) Since a and b are positive integers and c is larger than both of them, the smallest value that c can take is $c = 2$. The possible values of c are therefore $c = 2$, $c = 3$, and so on up to $c = n + 1$.

The only triple in S with the property that $c = 2$ is $(1, 1, 2)$, so there is one triple with $c = 2$. The triples in S with $c = 3$ are $(1, 1, 3)$, $(1, 2, 3)$, $(2, 1, 3)$, and $(2, 2, 3)$, so there are four triples with $c = 3$.

In general, if $c = r$, then a and b can both be any integer from 1 through $r - 1$ inclusive. Thus, there are $(r - 1) \times (r - 1) = (r - 1)^2$ triples in S with $c = r$. Since r ranges from 2 through $n + 1$, there are exactly

$$1^2 + 2^2 + 3^2 + 4^2 + \cdots + (n - 1)^2 + n^2 = p_2(n)$$

triples in S.

(b) We will now count the triples in S by considering them according to the number of distinct integers that occur in the triple. For example, there are two distinct integers in $(1, 1, 2)$ and three distinct integers in $(1, 5, 7)$.

If (a, b, c) is in S, then there are at most three distinct integers in the triple (since there are only three integers in total). As well, since $a < c$, the integers cannot all be equal. Therefore, there are either two distinct integers or three distinct integers in (a, b, c).

Suppose x, y, and z are three distinct integers between 1 and $n + 1$ inclusive. Since they are distinct, one of them is the largest. Assuming z is the largest, the triples (x, y, z) and (y, x, z) are both in S, and these are the only triples in S that contain the integers x, y, and z. Therefore, for every way to choose three distinct integers between 1 and $n + 1$ inclusive, there are two triples in S. This observation means there are $2\binom{n + 1}{3}$ triples in S with three distinct integers in them.

Now suppose that x and y are two distinct integers with $x < y$. Then the only triple in S that contains exactly the integers x and y is (x, x, y). Thus, for every two distinct integers between 1 and $n + 1$ inclusive, there is exactly one triple in S that contains exactly those two integers. Therefore, there are $\binom{n + 1}{2}$ triples in S with two distinct integers in them.

Therefore, the number of elements in S is $\binom{n + 1}{2} + 2\binom{n + 1}{3}$.

From part (a),

$$p_2(n) = \binom{n+1}{2} + 2\binom{n+1}{3}$$

$$= \frac{(n+1)!}{2!(n-1)!} + 2\frac{(n+1)!}{3!(n-2)!}$$

$$= \frac{(n+1)n}{2} + \frac{2(n+1)n(n-1)}{6}$$

$$= n(n+1)\left(\frac{1}{2} + \frac{n-1}{3}\right)$$

$$= n(n+1)\left(\frac{3}{6} + \frac{2(n-1)}{6}\right)$$

$$= \frac{n(n+1)(3+2n-2)}{6}$$

$$= \frac{n(n+1)(2n+1)}{6}$$

(c) Fix a positive integer n. Using the idea from parts (a) and (b), we let S_k be the set of ordered lists of $k+1$ positive integers $(x_1, x_2, x_3, \ldots, x_k, x_{k+1})$ with the property that each integer x_i is between 1 and $n+1$ inclusive and x_{k+1} is larger than every other integer in the list. A list of the form $(x_1, x_2, x_3, \ldots, x_k, x_{k+1})$ is called a $(k+1)$-tuple.

As in the case with $k = 2$, it is not possible for a $(k+1)$-tuple in S_k to have $x_{k+1} = 1$. This is because the other integers are positive and x_{k+1} must be the largest integer in the list (it cannot be "tied" for the largest). The possible values of x_{k+1} are 2 through $n+1$.

For a fixed value of x_{k+1}, say $x_{k+1} = r$, if $(x_1, x_2, x_3, \ldots, x_k, x_{k+1})$ is in S_k, then x_1 through x_k can each be any positive integer less than r, of which there are $r - 1$. Therefore, the number of $(k+1)$-tuples in S_k with $x_{k+1} = r$ is $(r-1)^2$. The possible values of r are 2 through $n+1$, so we have that the number of elements in S_k is

$$1^k + 2^k + 3^k + \cdots + (n-1)^k + n^k = p_k(n)$$

Following the reasoning of part (b), we now want to count the elements in S_k a different way. Since x_{k+1} is the largest integer in the $(k+1)$-tuple, there are at least two distinct integers in every $(k+1)$-tuple in S_k. As well, there are at most $k+1$ distinct integers in every $(k+1)$-tuple. Suppose we have chosen r distinct positive integers between 1 and $n+1$ inclusive with $2 \leq r \leq k+1$. We would like to count the $(k+1)$-tuples in S_k that contain exactly those r integers. For example, suppose $k = 4$ and $r = 3$. We need to count the number of 5-tuples that use exactly three given integers with the largest integer occurring only in the rightmost position. Suppose the largest integer is C and the other two are A and B. The 5-tuples in S_k are

(A, A, A, B, C), (A, A, B, A, C), (A, B, A, A, C), (B, A, A, A, C), (B, B, B, A, C),
(B, B, A, B, C), (B, A, B, B, C), (A, B, B, B, C), (A, A, B, B, C), (A, B, A, B, C),
(A, B, B, A, C), (B, A, A, B, C), (B, A, B, A, C), (B, B, A, A, C)

and so there are 14 5-tuples in S_k with exactly those three integers. This means that there are $14\binom{n+1}{3}$ 5-tuples in S_k that have exactly three distinct integers in them. The important thing to notice here is that the constant 14 does not depend on n. We could do a similar count to see how many 9-tuples there are in S_8 with exactly five distinct integers. The combinatorics might be a bit tedious, but in the end, we would find that the number of 9-tuples in S_8 with exactly five distinct integers is some multiple of $\binom{n+1}{5}$ where the multiplier does not depend on n.

Putting this together, there must be some constants $a_2, \ldots, a_{k+1}$ so that

$$p_k(n) = a_2\binom{n+1}{2} + a_3\binom{n+1}{3} + \cdots + a_k\binom{n+1}{k} + a_{k+1}\binom{n+1}{k+1}$$

where each coefficient a_r is equal to the number of $(k+1)$-tuples in S_k that contain exactly a fixed set of r distinct integers.

Put differently, suppose R is a set of r distinct positive integers. Then a_r is the number of $(k+1)$-tuples that satisfy the following three conditions: every integer in the $(k+1)$-tuple comes from R, every integer in R occurs at least once in the $(k+1)$-tuple, and the largest integer in R occurs exactly once and is in the rightmost position.

Before moving on, we make one final observation. With the convention that $\binom{u}{v} = 0$ when $u < v$, the equation above makes sense even when $n+1$ is smaller than the bottom number in the binomial coefficient. For example, if $k = 5$ and $n = 2$, then the term $a_4\binom{n+1}{4} = a_4\binom{3}{4}$ is counting the number of 6-tuples consisting of four distinct integers between 1 and $n+1 = 3$ inclusive. There is no way to choose four distinct integers from the list 1, 2, 3, so there should be zero such 6-tuples, and the fact that $\binom{3}{4} = 0$ records this observation.

(d) From part (c), we have that

$$p_3(n) = a_2\binom{n+1}{2} + a_3\binom{n+1}{3} + a_4\binom{n+1}{4}$$

for some constants a_2, a_3, and a_4. Using that $p_3(1) = 1^3 = 1$, we have that

$$
\begin{aligned}
1 &= p_3(1) \\
&= a_2\binom{1+1}{2} + a_3\binom{1+1}{3} + a_4\binom{1+1}{4} \\
&= a_2\binom{2}{2} + a_3\binom{2}{3} + a_4\binom{2}{4} \\
&= a_2 + 0 + 0
\end{aligned}
$$

3

and so $a_2 = 1$. With $n = 2$, we have $p_3(2) = 1^3 + 2^3 = 9$, so

$$9 = p_2(2)$$
$$= \binom{2+1}{2} + a_3 \binom{2+1}{3} + a_4 \binom{2+1}{4}$$
$$= \binom{3}{2} + a_3 \binom{3}{3} + a_4 \binom{3}{4}$$
$$= 3 + a_3 + 0$$

and so $a_3 = 6$. Finally, using $n = 3$, we have $p_3(3) = 1^3 + 2^3 + 3^3 = 36$, so

$$36 = p_3(3)$$
$$= \binom{4}{2} + 6 \binom{4}{3} + a_4 \binom{4}{4}$$
$$= 6 + 6 \times 4 + a_4$$

from which it follows that $a_4 = 6$. Recall that a_4 is equal to the number of four-tuples in S_3 consisting of a fixed set of four distinct integers. Indeed, the largest integer must go in the rightmost position, and there are 6 ways to arrange the other three integers, so $a_4 = 6$ makes sense from a combinatorial perspective.

We now have shown that

$$p_3(n) = \binom{n+1}{2} + 6 \binom{n+1}{3} + 6 \binom{n+1}{4}$$
$$= \frac{(n+1)n}{2} + \frac{6(n+1)n(n-1)}{6} + \frac{6(n+1)n(n-1)(n-2)}{24}$$

and after some simplification, we get that

$$p_3(n) = \frac{n^2(n+1)^2}{4}$$

We can approach $p_4(n)$ similar to how $p_3(n)$ was approached above. We start with

$$p_4(n) = a_2 \binom{n+1}{2} + a_3 \binom{n+1}{3} + a_4 \binom{n+1}{4} + a_5 \binom{n+1}{5}$$

and then use that $p_4(1) = 1$, $p_4(2) = 1^4 + 2^4 = 17$, $p_4(3) = 1^4 + 2^4 + 3^4 = 98$, and $p_4(4) = 1^4 + 2^4 + 3^4 + 4^4 = 354$ to solve for a_2, a_3, a_4, and a_5. After doing this, we find that $a_2 = 1$, $a_3 = 14$, $a_4 = 36$, and $a_5 = 24$. Therefore, after some simplification, we get

$$p_4(n) = \binom{n+1}{2} + 14 \binom{n+1}{3} + 36 \binom{n+1}{4} + 24 \binom{n+1}{5}$$
$$= \frac{n(n+1)(2n+1)(3n^2 + 3n - 1)}{30}$$

4

(e) Using that $p_5(n) = c_0 + c_1 n + c_2 n^2 + c_3 n^3 + c_4 n^4 + c_5 n^5 + c_6 n^6$, we can find a general expression for $p_5(n) - p_5(n-1)$.

$$
\begin{aligned}
n^5 &= p_5(n) - p_5(n-1) \\
&= c_0 + c_1 n + c_2 n^2 + c_3 n^3 + c_4 n^4 + c_5 n^5 + c_6 n^6 \\
&\quad - \left(c_0 + c_1(n-1) + c_2(n-1)^2 + c_3(n-1)^3 + c_4(n-1)^4 + c_5(n-1)^5 + c_6(n-1)^6 \right) \\
&= c_0 + c_1 n + c_2 n^2 + c_3 n^3 + c_4 n^4 + c_5 n^5 + c_6 n^6 \\
&\quad - c_0 - c_1(n-1) - c_2(n^2 - 2n + 1) - c_3(n^3 - 3n^2 + 3n - 1) \\
&\quad - c_4(n^4 - 4n^3 + 6n^2 - 4n + 1) - c_5(n^5 - 5n^4 + 10n^3 - 10n^2 + 5n - 1) \\
&\quad - c_6(n^6 - 6n^5 + 15n^4 - 20n^3 + 15n^2 - 6n + 1)
\end{aligned}
$$

and after collecting like terms, we get

$$
\begin{aligned}
n^5 &= (c_1 - c_2 + c_3 - c_4 + c_5 - c_6) + (2c_2 - 3c_3 + 4c_4 - 5c_5 + 6c_6)n \\
&\quad + (3c_3 - 6c_4 + 10c_5 - 15c_6)n^2 + (4c_4 - 10c_5 + 20c_6)n^3 \\
&\quad + (5c_5 - 15c_6)n^4 + 6c_6 n^5
\end{aligned}
$$

We can now equate coefficients to get the system of equations

$$
\begin{aligned}
6c_6 &= 1 \\
5c_5 - 15c_6 &= 0 \\
4c_4 - 10c_5 + 20c_6 &= 0 \\
3c_3 - 6c_4 + 10c_5 - 15c_6 &= 0 \\
2c_2 - 3c_3 + 4c_4 - 5c_5 + 6c_6 &= 0 \\
c_1 - c_2 + c_3 - c_4 + c_5 - c_6 &= 0
\end{aligned}
$$

From the first equation, we get $c_6 = \dfrac{1}{6}$. Substituting this into the second equation gives $5c_5 - 15 \times \dfrac{1}{6} = 0$ so $c_5 = \dfrac{1}{2}$. Continuing this way, we get that $c_4 = \dfrac{5}{12}$, $c_3 = 0$, $c_2 = -\dfrac{1}{12}$, and $c_1 = 0$. Therefore

$$
p_5(n) = c_0 - \frac{1}{12}n^2 + \frac{5}{12}n^4 + \frac{1}{2}n^5 + \frac{1}{6}n^6
$$

To solve for c_0, we can use that $p_5(1) = 1$ to get the equation

$$
1 = c_0 - \frac{1}{12} + \frac{5}{12} + \frac{1}{2} + \frac{1}{6} = c_0 + 1
$$

which means $c_0 = 0$. Finally, we can rearrange $p_5(n)$ into a nicer form

$$
\begin{aligned}
p_5(n) &= -\frac{1}{12}n^2 + \frac{5}{12}n^4 + \frac{1}{2}n^5 + \frac{1}{6}n^6 \\
&= \frac{-n^2 + 5n^4 + 6n^5 + 2n^6}{12} \\
&= \frac{n^2(n+1)^2(2n^2 + 2n - 1)}{12}
\end{aligned}
$$

(f) Fix a positive integer k and consider the function $p_k(n)$. We observed earlier that $p_k(n)$ is a polynomial in n. While $p_k(n)$ is designed to output $1^k + 2^k + \cdots + n^k$ when n is a positive integer, there is nothing to stop us from "symbolically" evaluating $p_k(n)$ when n is not an integer. For example, even though $p_1\left(\dfrac{1}{4}\right)$ does not have the same meaning as $p_1(n)$ when n is a positive integer (we cannot "add together the first $\dfrac{1}{4}$ positive integers"), we can still substitute $\dfrac{1}{4}$ into the formula for p_1 to get $p_1\left(\dfrac{1}{4}\right) = \dfrac{\frac{1}{4} \times \frac{5}{4}}{2} = \dfrac{5}{32}$.

Now consider the function $f_k(n) = p_k(n) - p_k(n-1) - n^k$ where n is allowed to be any real number. The function $f_k(n)$ is the sum/difference of three polynomials, so it is itself a polynomial. Since $p_k(n) - p_k(n-1) = n^k$ for all positive integers n, n is a root of $f_k(n)$ for all positive integers n. By the fact in the hint, a polynomial with infinitely many roots must be the constant zero function, so $p_k(n) - p_k(n-1) - n^k = 0$ for all real numbers n.

With $n = 1$, we get that $p_k(1) - p_k(0) = 1^k = 1$, but since $p_k(1) = 1$, we have that $1 - p_k(0) = 1$, so $p_k(0) = 0$. Similarly, by considering $n = 0$, we get that $p_k(0) - p_k(-1) = 0^k$, and since $p_k(0) = 0$, we have $0 - p_k(-1) = 0$, so $p_k(-1) = 0$ as well. We have shown that 0 and -1 are roots of the polynomial $p_k(n)$, which shows that n and $n+1$ are factors of $p_k(n)$. This means $n(n+1)$ is a factor of $p_k(n)$.

We now suppose that k is even, which means $n^k = (-n)^k$ for all real numbers n. From above, we have that $p_k(0) = p_k(-1) = 0$. Considering $p_k(n) - p_k(n-1) = n^k$ at $n = -1$, we have that $p_k(-1) - p_k(-2) = (-1)^k$, and since k is even and $p_k(-1) = 0$, we have $-p_k(-2) = 1^k$ or $p_k(-2) = -1^k$.

Next, we use $p_k(n) - p_k(n-1) = n^k$ with $n = -2$ to get $p_k(-2) - p_k(-3) = (-2)^k = 2^k$, and since $p_k(-2) = -1^k$, this means $-p_k(-3) = 1^k + 2^k$ or $p_k(-3) = -(1^k + 2^k)$.

Continuing this way, we have $p_k(-3) - p_k(-4) = (-3)^k = 3^k$, so $-1^k - 2^k - p_k(-4) = 3^k$, and so $p_k(-4) = -(1^k + 2^k + 3^k)$. We have now shown that $p_k(-n) = -p_k(n-1)$ for the positive integers $n = 0$, $n = 1$, $n = 2$, $n = 3$, and $n = 4$. This pattern continues. It can be proven using mathematical induction that $p_k(-n) = -p_k(n-1)$ for all positive integers n.

This means that $p_k(-n) + p_k(n-1) = 0$ for every non-negative integer. Therefore, the polynomial $p_k(-n) + p_k(n-1)$ has infinitely many roots, and so must be equal to 0 for every real number n.

Taking $n = \dfrac{1}{2}$, we get $p_k\left(-\dfrac{1}{2}\right) + p_k\left(\dfrac{1}{2} - 1\right) = 0$ which simplifies to $2p_k\left(-\dfrac{1}{2}\right) = 0$. Therefore, $-\dfrac{1}{2}$ is a root of $p_k(n)$ when k is even. Thus, $p_k(n)$ has a factor of $n + \dfrac{1}{2}$, which is equivalent to having a factor of $2n + 1 = 2\left(n + \dfrac{1}{2}\right)$.

Problem of the Month
Problem 5: February 2023

The sequence $(1, 3, 5, 2, 1, 2, 1)$ has the property that every integer in the sequence is a divisor of the sum of the integers adjacent to it. That is, 1 is a divisor of 3, 3 is a divisor of $1 + 5 = 6$, 5 is a divisor of $3 + 2 = 5$, and so on.

For $n \geq 3$, the sequence $(a_1, a_2, a_3, \ldots, a_n)$ of positive integers is called a *splendid sequence of length n* if it satisfies conditions S1, S2, and S3 found below.

  S1. a_1 is a divisor of a_2 and a_n is a divisor of a_{n-1}.

  S2. a_i is a divisor of $a_{i-1} + a_{i+1}$ for each i from 2 through $n - 1$ inclusive.

  S3. There is no prime number that is a divisor of every integer in the sequence.

For example, $(1, 3, 5, 2, 1, 2, 1)$ is a splendid sequence because it satisfies S1, S2, and S3. The sequence $(2, 4, 6, 2)$ satisfies S1 and S2, but it is not a splendid sequence because it fails S3 since 2 is a divisor of every integer in the sequence.

For $n = 2$, (a_1, a_2) is a *splendid sequence of length* 2 if it satisfies S1 and S3. Mostly for notational convenience, we also define a *splendid sequence of length* 1 to be the "sequence" (1). That is, the only splendid sequence of length 1 consists of a single integer equal to 1.

  (a) Show that there is only one splendid sequence of length 2.

  (b) Show that there is at least one splendid sequence of every possible length $n \geq 1$.

  (c) Suppose $(a_1, a_2, a_3, \ldots, a_n)$ is a splendid sequence of length $n \geq 2$. Show that for every integer i with $1 \leq i \leq n - 1$ there is a positive integer c so that $(a_1, a_2, \ldots, a_i, c, a_{i+1}, \ldots, a_n)$ is a splendid sequence of length $n + 1$.

  (d) Suppose $(a_1, a_2, a_3, \ldots, a_n)$ is a splendid sequence of length n. Show that $a_1 = a_n = 1$.

  (e) For each $n \geq 1$, show that there are only finitely many splendid sequences of length n.

  (f) Find a closed form for the number of splendid sequences of length n. Your answer should be an expression in terms of n.

Note: In part (f), we are asking you to find a closed form for the number of splendid sequences, the existence of which immediately implies that there are only finitely many. Hence, one way to answer part (e) is to answer part (f). With that said, we decided to include part (e) because it can be done without part (f) and (at least as far as we can tell) it is quite a bit easier than part (f). Part (f) is very challenging, so the hint will have more detail than usual.

Problem of the Month
Problem 5: February 2023

Hint

(a) Remember the condition that no prime number can divide every integer in a splendid sequence.

(b) Think of an integer that is a divisor of every integer.

(c) Try the positive integer $c = a_i + a_{i+1}$.

(d) Prove that if the divisibility conditions (S1 and S2 in the problem statement) are satisfied by $(a_1, a_2, a_3, \ldots, a_n)$, then a_1 and a_n must both divide every integer in the sequence.

(e) Try to write down a few splendid sequences of length at least 5. With the hint for (c) in mind, what do you notice about the largest integer in a splendid sequence? Show that, for each n, there is a largest possible value that an integer in any splendid sequence of length n can take. For instance, it can be shown that in a splendid sequence of length $n \geq 2$, every integer must be less than 2^{n-2}.

(f) There is a famous sequence called the *Catalan numbers* where the nth Catalan number equal to $\dfrac{1}{n+1}\dbinom{2n}{n}$. The Catalan numbers arise in many interesting ways in mathematics. One such way is that the nth Catalan number is equal to the number of sequences $(b_1, b_2, \ldots, b_n)$ of length n with the following properties.

- Each b_k is a positive integer.

- $b_1 = 1$.

- For each k satisfying $1 \leq k \leq n-1$, $b_{k+1} \leq b_k + 1$. That is, b_{k+1} is no more than one more than b_k (it is allowed to be less than or equal to b_k).

In the solution, such sequences will be called *tame sequences*. One way to answer this question is to show that the number of tame sequences of length $n-1$ is equal to the number of splendid sequences of length n and use the closed form for the number of tame sequences. To do this, we suggest trying to devise a way to use a tame sequence of length $n-1$ as "instructions" to construct a splendid sequence of length n. The idea from part (c) will probably be important.

Note: In the solution, we will provide a proof that the number of tame sequences of length n is the nth Catalan number. You might want to try to prove this yourself, but we recommend taking it for granted when trying to solve part (f).

Problem of the Month
Solution to Problem 5: February 2023

Definition 1: For integers a and b, we say that a *divides* b if there is an integer c with the property that $ac = b$. In this case, we write $a \mid b$.

The phrases "a *is a divisor of* b" and "b *is a multiple of* a" both have the exact same meaning as "a divides b". Notice that, by this definition, every integer is a divisor of 0, but the only divisor of 0 is 0 itself. We give a few facts that will be used in this solution. Their proofs are not included.

Fact 1: If a and b are positive integers such that $a \mid b$, then $a \leq b$.

Fact 2: If a, b, and c are integers such that $a \mid b$ and $a \mid c$, then $a \mid (b - c)$ and $a \mid (b + c)$.

Fact 3: If a, b, and c are integers such that $a \mid b$ and $b \mid c$, then $a \mid c$.

(a) Suppose (a, b) is a splendid sequence. By definition, this means $a \mid b$ and $b \mid a$. Since the integers in a splendid sequence must be positive, Fact 1 implies that $a \leq b$ and $b \leq a$, which implies $a = b$. If p is a prime number such that $p \mid a$, then $p \mid b$ by Fact 3. However, no prime number can divide both a and b beacuse (a, b) is splendid. Therefore, no prime number divides a. Similarly, no prime number divides b, and so $a = b = 1$.

Therefore, the only splendid sequence of length 2 is $(1, 1)$.

(b) There are several "generic" sequences that one might find. The simplest is probably the sequence $(1, 1, 1, \ldots, 1)$. That is, the sequence $(a_1, a_2, a_3, \ldots, a_n)$ with $a_i = 1$ for all i is always a splendid sequence. This is because no prime number divides 1, and 1 divides every integer. Another splendid sequence is $(1, 2, 3, 4, \ldots, n - 1, 1)$. For each integer k in this sequence, other than the 1's on the end and $n - 1$, the integers next to it are $k - 1$ and $k + 1$, so their sum is $(k - 1) + (k + 1) = 2k$, which is a multiple of k. The integers next to $n - 1$ are $n - 2$ and 1, which have a sum of $n - 1$.

In the remaining parts of the solution as well as in the Appendix, we will often denote a sequence by a bold letter. For example, we might refer to the sequence $(a_1, a_2, \ldots, a_n)$ by $\mathbf{x}$.

(c) Assume that $\mathbf{x} = (a_1, a_2, \ldots, a_n)$ is a splendid sequence. We will show that

$$\mathbf{y} = (a_1, a_2, \ldots, a_i, c, a_{i+1}, \ldots, a_n)$$

is a splendid sequence when $c = a_i + a_{i+1}$.

If some prime number p divides every integer in $\mathbf{y}$, then it also divides every integer in $\mathbf{x}$. Since $\mathbf{x}$ is a splendid sequence, there is no such prime number, so no prime number divides all of the integers in $\mathbf{y}$.

If $i = 1$, then $\mathbf{y} = (a_1, c, a_2, a_3, \ldots, a_n)$ and $c = a_1 + a_2$. Since $\mathbf{x}$ is a splendid sequence, $a_1 \mid a_2$. We also have that $a_1 \mid a_1$, so $a_1 \mid (a_1 + a_2)$ or $a_1 \mid c$ by Fact 2. Since every integer divides itself, $c \mid (a_1 + a_2)$. To see that $a_2 \mid (c + a_3)$, we note that $a_2 \mid (a_1 + a_3)$ because $\mathbf{x}$ is splendid and $a_2 \mid a_2$ because every integer divides itself. By Fact 2, $a_2 \mid (a_1 + a_2 + a_3)$ so $a_2 \mid (c + a_3)$. The integers a_3 through a_n all have exactly the same neighbours in $\mathbf{y}$ as

1

they do in $\mathbf{x}$, which is a splendid sequence, so $\mathbf{y}$ satisfies all other divisibility conditions required for it to be a splendid sequence.

If $i = n-1$, then $\mathbf{y}$ is splendid by a similar argument to the one in the previous paragraph.

If $1 < i < n-1$, then $\mathbf{y} = (a_1, a_2, \ldots, a_{i-1}, a_i, c, a_{i+1}, a_{i+2}, \ldots, a_n)$. When $k < i$ and when $k > i+1$, a_k divides the sum of its neighbours because it has the exact same neighbours as it did in $\mathbf{x}$. In $\mathbf{y}$, the neighbours of the integer a_i are a_{i-1} and c. Since $\mathbf{x}$ is a splendid sequence, $a_i \mid (a_{i-1}+a_{i+1})$. We also have that $a_i \mid a_i$, and so by Fact 2, $a_i \mid (a_i+a_{i-1}+a_{i+1})$ which means $a_i \mid (a_{i-1}+c)$. By similar reasoning, $a_{i+1} \mid (c+a_{i+2})$. The integer c is equal to the sum of its neighbours in $\mathbf{y}$ by definition, so it also divides the sum of its neighbours. Therefore, every integer in $\mathbf{y}$ divides the sum of its neighbours. We already argued that no prime number divides every integer in $\mathbf{y}$, so $\mathbf{y}$ is a splendid sequence.

(d) Suppose $\mathbf{x} = (a_1, a_2, a_3, \ldots, a_n)$ is a splendid sequence. If $n = 2$, then the solution to part (a) implies that $a_1 = a_n = 1$.

Suppose $n \geq 3$. By definition, we have that $a_n \mid a_{n-1}$ and that $a_{n-1} \mid (a_{n-2} + a_n)$. By Fact 3, $a_n \mid (a_{n-2}+a_n)$. Since $a_n \mid a_n$, we can apply Fact 2 to get that $a_n \mid (a_{n-2}+a_n - a_n)$ which implies that $a_n \mid a_{n-2}$.

Since $\mathbf{x}$ is splendid, we also have that $a_{n-2} \mid (a_{n-3} + a_{n-1})$. We have just shown that a_n divides a_{n-2}, so by Fact 3, $a_n \mid (a_{n-3} + a_{n-1})$. We also have that $a_n \mid a_{n-1}$, so Fact 2 implies $a_n \mid (a_{n-3} + a_{n-1} - a_{n-1})$ or $a_n \mid a_{n-3}$. Continuing in this way, we can show that $a_n \mid a_i$ for all i with $1 \leq i \leq n$. By the condition that no prime number can divide every integer in $\mathbf{x}$, we conclude that $a_n = 1$. Essentially the same argument shows that $a_1 = 1$.

(e) Most of the work is to prove these two claims.

Claim 1: If $(a_1, a_2, \ldots, a_n)$ is a splendid sequence of length $n \geq 3$ that contains at least one integer that is greater than 1, then there is some i with $1 < n$ such that $a_i = a_{i-1}+a_{i+1}$ and $(a_1, a_2, \ldots, a_{i-1}, a_{i+1}, a_{i+2}, \ldots, a_n)$ is a splendid sequence. That is, there is an integer in the sequence that is equal to the sum of its neighbours, and if it is removed, the remaining shorter sequence is a splendid sequence.

Claim 2: If $(a_1, a_2, \ldots, a_n)$ is a splendid sequence of length $n \geq 2$, then $a_i \leq 2^{n-2}$ for every i.

Proof of Claim 1. To prove Claim 1, suppose $\mathbf{x} = (a_1, a_2, \ldots, a_n)$ is a splendid sequence with at least one integer greater than 1 and let i be such that a_i is the largest integer in the sequence, choosing the rightmost occurrence if there is a "tie". More precisely, i is the largest integer with the property that $a_i \geq a_j$ for all $1 \leq j \leq n$.

By part (d), $a_n = a_1 = 1$, and since there is at least one integer in $\mathbf{x}$ that is greater than 1, neither a_1 nor a_n can be the largest integer in $\mathbf{x}$, which means $1 < i < n$. The choice of i ensures that $a_{i-1} \leq a_i$ and $a_{i+1} \leq a_i$. If $a_i = a_{i+1}$, then since a_i is the largest integer in the sequence, this would imply that a_i is *not* the rightmost occurrence of the largest integer in the sequence. Therefore, we actually have that $a_{i+1} < a_i$.

The inequalities $a_{i-1} \leq a_i$ and $a_{i+1} < a_i$ imply that $a_{i-1} + a_{i+1} < 2a_i$ and since $\mathbf{x}$ is splendid, $a_i \mid (a_{i-1} + a_{i+1})$. As well, all integers in a splendid sequence are positive, which means that $a_{i-1} + a_{i+1}$ is a positive multiple of a_i that is less than $2a_i$. The only such multiple is a_i itself, and so $a_{i-1} + a_{i+1} = a_i$.

2

We have shown that one of the integers in $\mathbf{x}$ is equal to the sum of its neighbours. To finish proving the claim, we need to show that $\mathbf{y} = (a_1, a_2, \ldots, a_{i-1}, a_{i+1}, a_{i+2}, \ldots, a_n)$ is a splendid sequence. In $\mathbf{y}$, only a_{i-1} and a_{i+1} have different neighbours than they did in $\mathbf{x}$, so the divisibility conditions we need to verify are that $a_{i-1} \mid (a_{i-2} + a_{i+1})$ and $a_{i+1} \mid (a_{i-1} + a_{i+2})$

We know that $a_{i-1} \mid (a_{i-2} + a_i)$ and we have just shown that $a_i = a_{i-1} + a_{i+1}$. Substituting, we get that $a_{i-1} \mid (a_{i-2} + a_{i-1} + a_{i+1})$. By Fact 2, $a_{i-1} \mid (a_{i-2} + a_{i-1} + a_{i+1} - a_{i-1})$ or $a_{i-1} \mid (a_{i-2} + a_{i+1})$. A nearly identical argument shows that $a_{i+1} \mid (a_{i-1} + a_{i+2})$. Therefore, $\mathbf{y}$ satisfies the divisibility conditions.

If a prime number p divides every integer in $\mathbf{y}$, then $p \mid a_{i-1}$ and $p \mid a_{i+1}$, so $p \mid a_i$ by Fact 2 since $a_i = a_{i-1} + a_{i+1}$. This would mean p divides every integer in $\mathbf{x}$, which is not the case since $\mathbf{x}$ is splendid. $\qquad\square$

We will now prove Claim 2 by mathematical induction. The essence of the proof is that, by Claim 1, the largest integer in a splendid sequence must be the sum of two integers in a shorter splendid sequence. Therefore, the maximum size of an integer in a splendid sequence of length $n + 1$ is at most twice the maximum size of an integer in a splendid sequence of length n. This means that there is always a fixed upper bound on the size of integers in a splendid sequence of a fixed length.

Proof of Claim 2. To get an idea of how the induction will work, we first prove this for $n = 2$, $n = 3$, and $n = 4$. For $n = 2$, we showed in the solution to part (a) that the only splendid sequence of length 2 is $(1, 1)$. The largest element in this sequence is $1 = 2^0 = 2^{2-2} = 2^{n-2}$, so the claim holds for $n = 2$.

Now suppose (a_1, a_2, a_3) is a splendid sequence of length 3. If $a_1 = a_2 = a_3 = 1$, then every integer in the sequence is less than $2^{3-2} = 2$. Otherwise, since $a_1 = a_3 = 1$ by part (d), Claim 1 implies that $a_2 = a_1 + a_3 = 1 + 1 = 2$ is the largest integer in the sequence, so all integers in the sequence are at most $2 = 2^{3-2}$.

Continuing to $n = 4$, suppose (a_1, a_2, a_3, a_4) is a splendid sequence. Again, if the sequence consists entirely of 1's, then $a_i \leq 2^{4-2} = 4$ for all i. Otherwise, either (a_1, a_3, a_4) is a splendid sequence of length 3 and $a_2 = a_1 + a_3$, or (a_1, a_2, a_4) is a splendid sequence of length 3 and $a_3 = a_2 + a_4$. Either way, three of the four integers are in a splendid sequence of length 3, and the fourth is the sum of two integers in a splendid sequence of length 3. We just showed that an integer in a splendid sequence of length 3 is at most 2^{3-2}, so an integer in a splendid sequence of length 4 is at most $2^{3-2} + 2^{3-2} = 2 \times 2^{3-2} = 2^{4-2}$. Therefore, no integer in the sequence (a_1, a_2, a_3, a_4) can exceed 2^{4-2}.

Now for the inductive step. Suppose, for some $n \geq 2$, that every integer in every splendid sequence of length n is at most 2^{n-2}. This is our *inductive hypothesis*. Consider a splendid sequence $(a_1, a_2, \ldots, a_n, a_{n+1})$ of length $n + 1$. If $a_i = 1$ for all i, then $a_i \leq 2^{n-2}$ for all i. Otherwise, Claim 1 implies that there is some i so that $a_i = a_{i-1} + a_{i+1}$ and $(a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n, a_{n+1})$ is a splendid sequence of length n. By the inductive hypothesis, this means each of a_1, a_2, a_3, $\ldots$, a_{i-1}, a_{i+1}, $\ldots$, a_n, and a_{n+1} is at most 2^{n-2}, and since $2^{n-2} < 2^{(n+1)-2}$, each of these integers is at most $2^{(n+1)-2}$. For a_i, we have $a_i = a_{i-1} + a_{i+1}$ and since $a_{i-1} \leq 2^{n-2}$ and $a_{i+1} \leq 2^{n-2}$, we conclude that $a_i \leq 2^{n-2} + 2^{n-2} = 2(2^{n-2}) = 2^{(n+1)-2}$ as well. We have shown that every integer in a

3

splendid sequence of length $n + 1$ is at most $2^{(n+1)-1}$. This completes the induction and the proof. $\qquad\square$

By Claim 2, every integer in a splendid sequence of length n is at most 2^{n-2}. There are only finitely many sequences of length n consisting of positive integers less than or equal to 2^{n-2}, regardless of whether they are splendid. Therefore, for fixed n, there are only finitely many splendid sequences of length n.

The proof os part (f) will use some language about sets. Specifically, we will use the language of injective, surjective, and bijective functions. If you have seen this before, you should be ready to read the solution to part (f). Otherwise, we recommend reading Appendix 1 first.

(f) As pointed out in the hint, the number of splendid sequences of length n is equal to the $(n-1)^{\text{st}}$ *Catalan number*. The nth Catalan number is equal to $\dfrac{1}{n+1}\dbinom{2n}{n}$, so the number of splendid sequences of length n is $\dfrac{1}{n}\dbinom{2n-2}{n-1}$. The sequence of Catalan numbers shows up in many contexts. A useful example for this solution is given in Definition 2.

Definition 2: For each positive integer $n \geq 1$, we call a sequence $(b_1, b_2, \ldots, b_n)$ of positive integers a *tame sequence* if $b_1 = 1$ and $b_{k+1} \leq b_k + 1$ for every integer k with $1 \leq k < n$. In other words, $b_1 = 1$ and every other integer in the sequence is at most 1 more than the previous integer.

The name *tame sequence* was made up for the purpose of this solution, but it is known that the number of tame sequences of length n is equal to the nth Catalan number. There are proofs of this in various places in the literature. For completeness, we have included a proof in Appendix 2. It is stated as Claim 5.

Let T_n denote the set of tame sequences of length n and S_n denote the set of splendid sequences of length n. We will show, for $n \geq 2$, that there is a bijection with domain T_{n-1} and codomain S_n. By the discussion in Appendix 1, this will show that the number of splendid sequences of length n is the $(n-1)^{\text{st}}$ Catalan number.

Recall from part (c) that if $(a_1, a_2, \ldots, a_n)$ is a splendid sequence of length n, then

$$(a_1, a_2, a_3, \ldots, a_i, a_i + a_{i+1}, a_{i+1}, \ldots, a_n)$$

is a splendid sequence of length $n + 1$.

From this point on, it will be notationally useful to prepend a zero at the beginning of splendid sequences. For example, the sequence $(0, 1)$ will now be the unique splendid sequence of length 1. The sequence $(0, 1, 2, 5, 3, 1)$ is a splendid sequence of length 5. This means that a splendid sequence of length n now has $n + 1$ integers, the first of which is 0. For instance, $a_1 = 0$, $a_2 = 1$, $a_3 = 2$, $a_4 = 5$, $a_5 = 3$, and $a_6 = 1$ is how we would index the sequence $(0, 1, 2, 5, 3, 1)$ going forward. Notice that the observation from part (c) mentioned above also works if we insert the sum between the first and second integers, 0 and 1. You should convince yourself of this before moving on.

For each $n \geq 2$, we define a function, f_n, with domain T_{n-1} and codomain S_n. The way f_n works is to use a tame sequence as a list of instructions to build a splendid sequence. Consider a tame sequence $\mathbf{x} = (b_1, b_2, \ldots, b_{n-1})$ of length $n - 1$. Starting with $(0, 1)$, the

4

unique splendid sequence of length 1, we read $\mathbf{x}$ from left to right and each integer in the tame sequence tells us where to insert a sum to get a longer splendid sequence. Specifically, in the kth step, the integer b_k tells us that we should insert a sum between a_{b_k} and $a_{b_{k+1}}$ to get a longer splendid sequence.

For example, suppose $n = 6$ and the tame sequence is $\mathbf{x} = (1, 2, 3, 2, 1, 1)$. Start with $(0, 1)$, which has $a_1 = 0$ and $a_1 = 1$. The first integer in $\mathbf{x}$ is 1, so in the first step we insert the sum between a_1 and a_2. This means we go from $(0, 1)$ to $(0, 0 + 1, 1) = (0, 1, 1)$. To start the second step, we reindex to $a_1 = 0$, $a_2 = 1$, and $a_3 = 1$. The next integer in $\mathbf{x}$ is 2, so we insert the sum between a_2 and a_3 to get $(0, 1, 1 + 1, 1) = (0, 1, 2, 1)$. The next integer in $\mathbf{x}$ is 3, so we insert the sum between the third and fourth integers in the current splendid sequence to get $(0, 1, 2, 2 + 1, 1) = (0, 1, 2, 3, 1)$. Continuing, we get $(0, 1, 1 + 2, 2, 3, 1) = (0, 1, 3, 2, 3, 1)$, followed by $(0, 0 + 1, 1, 3, 2, 3, 1) = (0, 1, 1, 3, 2, 3, 1)$, and finally $(0, 0 + 1, 1, 1, 3, 2, 3, 1) = (0, 1, 1, 1, 3, 2, 3, 1)$. Thus, $f_7(\mathbf{x}) = (0, 1, 1, 1, 3, 2, 3, 1)$.

By part (c), if $\mathbf{x}$ is a tame sequence of length $n - 1$, then $f_n(\mathbf{x})$ is a splendid sequence of length n. Also note that at the start of the kth step, the splendid sequence has $k + 1$ integers in it. Because of the way tame sequences are defined, the kth integer in a tame sequence is at most k, so at each step, the splendid sequence is always long enough for the instruction to makes sense.

For each $n \geq 2$, we will show that the function f_n is a bijection. The proof will be by induction, but we first need a definition and then a useful fact.

Definition 3: For a splendid sequence $(a_1, a_2, a_3, \ldots, a_{n+1})$ of length n (remember that $a_1 = 0$), we say that a_i is a *peak* if $a_i = a_{i-1} + a_{i+1}$.

By Claim 1 (see the solution to part (e)), every splendid sequence with an integer greater than 1 has at least one peak. Moreover, if that peak is "removed", the resulting shorter sequence is splendid. As well, with our new notation, if there is no integer greater than 1, then the sequence is of the form $(0, 1, 1, 1, \ldots, 1)$ and the first (leftmost) 1 is its only peak. If it is removed, the resulting shorter sequence is also splendid. Indeed, the reason for introducing the 0 at the beginning was to avoid having to treat the sequence of all 1's separately in this part of the argument.

Now for the useful fact:

Claim 3: Suppose $\mathbf{y} = (a_1, a_2, \ldots, a_{n+1})$ is a splendid sequence of length n and that $\mathbf{x} = (b_1, b_2, \ldots, b_{n-1})$ is a tame sequence of length $n - 1$ such that $f_n(\mathbf{x}) = \mathbf{y}$. If we let $b_{n-1} = m$, then a_{m+1} is the leftmost peak of $\mathbf{y}$.

A proof of Claim 3 is given at the end. We will now prove by induction that f_n is a bijection for all $n \geq 2$.

It was observed in part (a) that the only splendid sequence of length 2 is $\mathbf{y} = (0, 1, 1)$. As well, the only tame sequence of length $2 - 1 = 1$ is $\mathbf{x} = (1)$. It is easily checked that $f_2(\mathbf{x}) = \mathbf{y}$. The sets T_1 and S_2 each have only one element. There is only one function between two sets with one element, and it is always a bijection (convincing yourself of this is a good exercise in understanding definitions). Therefore, f_2 is a bijection.

For the inductive hypothesis, we assume for some $n \geq 2$ that f_n is a bijection.

We will show that f_{n+1} is a bijection, which means we need to show that it is injective and

surjective. To show that it is surjective, we assume that $\mathbf{y} = (a_1, a_2, \ldots, a_n, a_{n+1}, a_{n+2})$ is in S_{n+1} and let a_k be its leftmost peak. By Claim 1 from the solution to part (e), the sequence $\mathbf{z} = (a_1, a_2, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n, a_{n+1}, a_{n+2})$ is in S_n. Since f_n is a bijection by the inductive hypothesis, it is surjective, so there is some tame sequence $\mathbf{w} = (b_1, b_2, \ldots, b_{n-1})$ in T_{n-1} with $f_{n-1}(\mathbf{w}) = \mathbf{z}$. For convenience, let $m = b_{n-1}$.

Recall that the leftmost peak of $\mathbf{x}$ is a_k. Note that $k \neq 1$ since $a_1 = 0$ can never be a peak. Therefore, $k \geq 2$ and so $k - 1 \geq 1$.

Suppose a_r is the leftmost peak of $\mathbf{z}$. If $r \leq k - 2$, then a_r is a peak of $\mathbf{y}$ because a_r has the same neighbours in $\mathbf{y}$ and $\mathbf{z}$ when $r \leq k - 2$. However, a_k was chosen to be the leftmost peak of $\mathbf{y}$, so we cannot have $r \leq k - 2$. This means $r \geq k - 1$, but by Claim 3, $r = m + 1$, so we get $k - 1 \leq m + 1$. Combining with $1 \leq k - 1$, we have that $1 \leq k - 1 \leq m + 1 = b_{n-1} + 1$, which means the sequence $(b_1, b_2, \ldots, b_{n-1}, k - 1)$ is a tame sequence. We will call this tame sequence $\mathbf{x}$.

To recap, the tame sequence $\mathbf{x}$ is obtained by appending $k - 1$ to $\mathbf{w}$, the splendid sequence $\mathbf{y}$ is obtained by inserting the sum between the $(k-1)^{\text{st}}$ and k^{th} integer in $\mathbf{z}$, and $f_n(\mathbf{w}) = \mathbf{z}$. It follows that $f_{n+1}(\mathbf{x}) = \mathbf{y}$. We have found $\mathbf{x} \in T_n$ such that $f_{n+1}(\mathbf{x}) = \mathbf{y}$. Since $\mathbf{y}$ was an arbitrary element of S_{n+1}, this concludes the proof that f_{n+1} is surjective.

We will now show that f_{n+1} is injective. To do this, we suppose $\mathbf{x} = (b_1, b_2, \ldots, b_n)$ and $\mathbf{w} = (c_1, c_2, \ldots, c_n)$ are in T_n with $f_{n+1}(\mathbf{x}) = f_{n+1}(\mathbf{w})$. We will show that $\mathbf{x} = \mathbf{w}$.

Let $\mathbf{y} = (a_1, a_2, \ldots, a_{n+1}, a_{n+2})$ be such that $\mathbf{y} = f_{n+1}(\mathbf{x}) = f_{n+1}(\mathbf{w})$. Suppose a_k is the leftmost peak of $\mathbf{y}$. By Claim 3, both $c_n = k - 1$ and $b_n = k - 1$. This shows that $c_n = b_n$. As well, if we set $\mathbf{u} = (b_1, b_2, \ldots, b_{n-1})$ and $\mathbf{v} = (c_1, c_2, \ldots, c_{n-1})$ and $\mathbf{z} = (a_1, a_2, \ldots, a_{k-1}, a_{k+1}, \ldots, a_{n+1}, a_{n+2})$, then $f_n(\mathbf{u}) = f_n(\mathbf{v}) = \mathbf{z}$. By the inductive hypothesis, f_n is bijective, and hence, it is injective, so $\mathbf{u} = \mathbf{v}$. This shows that $\mathbf{x}$ and $\mathbf{w}$ have the same first $n - 1$ integers, and since $b_n = c_n$ as well, we have that $\mathbf{x} = \mathbf{w}$, which concludes the proof that f_{n+1} is injective.

We have now shown that f_{n+1} is bijective, which proves that T_{n-1} and S_n have the same number of elements when $n \geq 2$. Therefore, the number of splendid sequences of length n is

$$\frac{1}{n}\binom{2n-2}{n-1}$$

as claimed earlier.

Proof of Claim 3. The proof is by induction on n. It was noted earlier that the only tame sequence of length $2 - 1 = 1$ is $\mathbf{x} = (1)$ ($b_1 = 1$), the only splendid sequence of length 2 is $\mathbf{y} = (0, 1, 1)$ ($a_1 = 0$, $a_2 = 1$, $a_3 = 1$), and that $f_2(\mathbf{x}) = \mathbf{y}$. The leftmost peak of $\mathbf{y}$ is a_2 and $2 = b_1 + 1$. This shows that Claim 3 is true when $n = 2$.

For the inductive hypothesis, we suppose, for some $n \geq 2$, that if $\mathbf{y} = (a_1, a_2, \ldots, a_n, a_{n+1})$ is a splendid sequence of length n and $\mathbf{x} = (b_1, b_2, \ldots, b_{n-1})$ is a tame sequence of length $n - 1$ such that $f_n(\mathbf{x}) = \mathbf{y}$, then the leftmost peak of the $\mathbf{y}$ is a_{m+1} where $m = b_{n-1}$.

Suppose $\mathbf{y} = (a_1, a_2, \ldots, a_n, a_{n+1}, a_{n+2})$ is a splendid sequence of length $n + 1$ and that $\mathbf{x} = (b_1, b_2, \ldots, b_n)$ is a tame sequence of length n with $f_{n+1}(\mathbf{x}) = \mathbf{y}$. Because of how f_{n+1} is applied, a_{m+1} is a peak of $\mathbf{y}$. As well, $\mathbf{z} = (a_1, a_2, \ldots, a_m, a_{m+2}, \ldots, a_{n+1}, a_{n+2})$ is

a splendid sequence of length n such that $f_n(\mathbf{w}) = \mathbf{z}$ where $\mathbf{w} = (b_1, b_2, \ldots, b_{n-1})$. For convenience, set $b_n = m$ and $b_{n-1} = k$. Suppose the leftmost peak of $\mathbf{y}$ is a_r for some r. For now, assume $r < m + 1$. It is not difficult to show that it is impossible for a splendid sequence to have two consecutive peaks. This means we must have $r \leq m - 1$ since r must be at least two less than $m + 1$. If this happens, a_r is also a peak of $\mathbf{z}$ because a_{m-1} has exactly the same neighbours in $\mathbf{z}$ and $\mathbf{y}$. By the inductive hypothesis, a_{k+1} is the leftmost peak of $\mathbf{z}$, so $r = k + 1$ and we get $k + 1 \leq m - 1$. Since $\mathbf{z}$ is a tame sequence, $b_n \leq b_{n-1} + 1$ or $m \leq k + 1$. This implies that $m \leq k + 1 \leq m - 1$ so $m \leq m - 1$, which is impossible. Therefore, we cannot have $r < m + 1$, which means a_{m+1} is indeed the leftmost peak of $\mathbf{y}$. This completes the induction and the proof. $\square$

Appendix 1

Suppose X and Y are finite sets, where by "set" we mean an unordered collection of objects. Suppose there is a "rule" that, for every element in the set X, produces an element in the set Y. For instance, if the sets were $X = \{(1, 2), (5, 3), (1, -2)\}$ and $Y = \{(4, 1), (5, 2), (9, 5)\}$ (both sets of three ordered pairs), the "rule" might be "square the second entry, then reverse the order". With this rule, $(1, 2)$ becomes $(4, 1)$, $(5, 3)$ becomes $(9, 5)$, and $(1, -2)$ becomes $(4, 1)$, so every element of X is transformed into an element of Y. Such a rule is called a function. The set X is called its "domain" and Y is called its "codomain". If the function is named f, we would use $f(x)$ to denote the function applied to an element x in the domain. You have probably seen functions before where the domain and codomain are all or part of the set of real numbers, but the notion of a function applies in a much broader context. Below are three important properties that functions may (or may not) have.

Injectivity: A function f with domain X and codomain Y is called *injective* if for every two elements of the domain, x_1 and x_2, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$. In other words, a function is injective if its application to two different elements of the domain always gives two different results. Note that when trying to prove that a function is injective, we typically assume that $f(x_1) = f(x_2)$ and deduce that $x_1 = x_2$. You might want to think about this logic.

Surjectivity: A function f with domain X and codomain Y is called *surjective* if for every $y \in Y$ there exists an $x \in X$ so that $f(x) = y$. In other words, a function is surjective if every element of the codomain is the result of applying f to some element in the domain. [We might also say that the *range equals the codomain* to describe surjectivity.]

Bijectivity: A function f is with domain X and codomain Y is called *bijective* or *is a bijection* if it is both injective and surjective.

There is a lot to be said about injective, surjective, and bijective functions, but for us, the useful observation will be that if X and Y are finite sets and there is a bijective function f with domain X and codomain Y, then X and Y have the same number of elements. Indeed, if X has m elements and Y has n elements, then being injective implies that $m \leq n$ and being surjective implies that $m \geq n$. Thus, being bijective implies that $m \leq n \leq m$, so $m = n$.

Observe that the example given at the beginning of this appendix is neither injective nor surjective, so it is not bijective. However, X and Y do have the same number of elements. It is important to keep in mind that we are only claiming that if there is a bijection from X to Y, then they have the same number of elements. There are six bijective functions from X to Y, the example we gave just happens to not be one of them.

We will now show that the number of tame sequences of length n is equal to the nth Catalan number, $\frac{1}{n+1}\binom{2n}{n}$. The proof relies on the material from Appendix 1. As well, the results here are well known and proofs of them can be found in the literature.

Definition 4: For each positive integer n, a sequence of $2n$ integers is called a *jagged* sequence of length $2n$ if properties $P1$ and $P2$ hold:

   $P1$ Exactly n of the integers are equal to 1 and exactly n of the integers are equal to -1.

   $P2$ For each integer k with $1 \leq k \leq 2n$, the sum of the first k integers in the sequence is non-negative.

Claim 4: There are $\frac{1}{n+1}\binom{2n}{n}$ jagged sequences of length $2n$.

Proof of Claim 4. Fix a positive integer n. Let X be the set of sequences of $2n$ integers that satisfy $P1$ and fail $P2$. Also, let Y be the set of sequences of $2n$ integers, $n+1$ of which equal -1 and $n-1$ of which equal 1. We will show that X and Y have the same number of elements.

Suppose $\mathbf{x} = (a_1, a_2, \ldots, a_{2n})$ is in X. Since $\mathbf{x}$ fails $P2$, there must be some k with $1 \leq k \leq 2n$ and the property that the sum of the first k entries is negative. Let k be the smallest such position in the sequence. If $a_1 = -1$, then $k = 1$. This means n of the integers in the list $a_2, a_3, \ldots, a_{2n}$ are equal to 1, and $n-1$ of them are equal to -1. Therefore, the sequence

$$(a_1, -a_2, -a_3, \ldots, -a_{2n})$$

has $n+1$ integers equal to -1 and $n-1$ integers equal to 1, which means it is in Y.

If $k \neq 1$, then $a_1 \geq 0$, $a_1+a_2 \geq 0$, and so on up to $a_1+a_2+\cdots+a_{k-1} \geq 0$, but $a_1+a_2+\cdots+a_k < 0$. Since $a_1+a_2+\cdots+a_{k-1} \geq 0$ but $a_1+a_2+\cdots+a_{k-1}+a_k < 0$, we must have that a_k is negative, but $a_k = \pm 1$, so $a_k = -1$. As well, each of the a_i are integers, so the two sums above are integers, which means $a_1+a_2+\cdots+a_{k-1} = 0$ and $a_1+a_2+\cdots+a_{k-1}+a_k = -1$ (there is no other way to add -1 to a non negative integer and get a negative integer). The fact that $a_1+a_2+\cdots+a_{k-1} = 0$ implies that exactly half of the integers in the list $a_1, \ldots, a_{k-1}$ are equal to -1, and so the number of -1's in $(a_1, a_2, \ldots, a_k)$ is one more than the number of 1's. Since the number of -1's and 1's is equal in $\mathbf{x}$, this means the number of 1's in $(a_{k+1}, \ldots, a_{2n})$ is one more than the number of -1's. All of this implies that

$$(a_1, a_2, \ldots, a_k, -a_{k+1}, -a_{k+2}, \ldots, -a_{2n})$$

has two more -1's than 1's. Two numbers that differ by 2 and have a sum of $2n$ must be $n-1$ and $n+1$, so the sequence above is in Y.

The above work defines a function, that we will call f, with domain X and codomain Y. Specifically, if $\mathbf{x} = (a_1, a_2, \ldots, a_{2n})$ in X and k the smallest integer such that $a_1 + a_2 + \cdots + a_k < 0$, $f(\mathbf{x}) = (a_1, a_2, \ldots, a_k, -a_{k+1}, \ldots, -a_{2n})$. That is, $f(\mathbf{x})$ is the sequence obtained by negating every integer from a_{k+1} to the end of the sequence. We will show that f is a bijection.

To see that f is injective, suppose $\mathbf{w} = (a_1, a_2, \ldots, a_{2n})$ and $\mathbf{x} = (b_1, b_2, \ldots, b_{2n})$ are in X with $f(\mathbf{w}) = f(\mathbf{x})$. We suppose that k is the smallest such that $a_1 + a_2 + \cdots + a_k < 0$

and m is the smallest such that $b_1 + b_2 + \cdots + b_m < 0$. We might as well assume that $k \leq m$. Our assumption says that the sequences $(a_1, a_2, \ldots, a_k, -a_{k+1}, -a_{k+1}, \ldots, -a_{2n})$ and $(b_1, b_2, \ldots, b_k, -b_{m+1}, -b_{m+1}, \ldots, -b_{2n})$ are equal. Since $k \leq m$, this means $a_i = b_i$ when $1 \leq i \leq k$ and when $m + 1 \leq i \leq 2n$. Observe that $a_1 + a_2 + \cdots + a_k = b_1 + b_2 + \cdots + b_k$, and since $a_1 + a_2 + \cdots + a_k < 0$ by assumption, we get that $b_1 + b_2 + \cdots + b_k < 0$. This means $m \leq k$ as well and so $k = m$. Since $a_i = b_i$ when $1 \leq i \leq k$ and $k + 1 \leq i \leq 2n$, we have that $a_i = b_i$ for all i. In other words, $\mathbf{w} = \mathbf{x}$, so f is injective.

Now suppose $\mathbf{y} = (c_1, c_2, c_3, \ldots, c_{2n})$ is a sequence is in Y. Because $\mathbf{y} \in Y$, exactly $n + 1$ of the integers in $\mathbf{y}$ are equal to -1 and $n - 1$ of them are equal to 1. Consider the list of sums

$$c_1$$
$$c_1 + c_2$$
$$c_1 + c_2 + c_3$$
$$\vdots$$
$$c_1 + c_2 + c_3 + \cdots + c_{2n}$$

The first "sum", c_1, is either -1 or 1. The final sum is $(n - 1) - (n + 1) = -2$. As we move from one sum to the next in the list above, we add c_i for some i, which means the sums either increase or decrease by 1 as we move down the list. Therefore, there is at least one sum that equals -1 (it could be the first). Suppose k is the smallest such that $c_1 + c_2 + c_3 + \cdots + c_k = -1$. Then the sequence

$$\mathbf{x} = (c_1, c_2, \ldots, c_k, -c_{k+1}, -c_{k+2}, \ldots, c_{2n})$$

is in X and $f(\mathbf{x}) = \mathbf{y}$. To see that $\mathbf{x} \in S$, we have that $c_1 + c_2 + \cdots + c_k = -1$ and $c_1 + c_2 + \cdots + c_{2n} = -2$, ad so it must be that $c_{k+1} + c_{k+1} + \cdots + c_{2n} = -1$. Therefore, $c_1 + c_2 + c_3 + \cdots + c_k + (-c_{k+1}) + (-c_{k+2}) + \cdots + (-c_{2n}) = 0$. This means $\mathbf{x}$ has the same number of 1's and -1's, which means there are n of each. As well, the sequence fails $P2$ because the first k integers have a negative sum. This shows that $\mathbf{x} \in X$, and that $f(\mathbf{x}) = \mathbf{y}$ is essentially by the definition of $\mathbf{x}$. Therefore, f is surjective, which completes the proof that it is a bijection.

The number of sequences in Y is $\binom{2n}{n+1}$. This is because we can choose where to put the -1's in $\binom{2n}{n+1}$ ways, and then there is no choice of where to place the 1's. By what we have shown, we now know that there are $\binom{2n}{n+1}$ sequences in X as well.

We can now compute the number of jagged sequences. The number of sequences of $2n$ integers that satisfy $P1$ is $\binom{2n}{n}$ by reasoning similar to that in the previous paragraph. To get the number of jagged sequences of length $2n$, we need to subtract from $\binom{2n}{n}$ the number of sequences that satisfy $P1$ but fail $P2$, which is exactly the number of sequences in X. Therefore, the number

of jagged sequences is

$$\binom{2n}{n} - \binom{2n}{n+1} = \frac{(2n)!}{n!n!} - \frac{(2n)!}{(n+1)!(n-1)!}$$
$$= \frac{(2n)!}{n!(n-1)!}\left(\frac{1}{n} - \frac{1}{n+1}\right)$$
$$= \frac{(2n)!}{n!(n-1)!} \times \frac{1}{n(n+1)}$$
$$= \frac{1}{n+1} \times \frac{(2n)!}{n!n!}$$
$$= \frac{1}{n+1}\binom{2n}{n}$$

$\square$

Claim 5: The number of tame sequences of length n is $\dfrac{1}{n+1}\dbinom{2n}{n}$.

Proof of Claim 5. We will find a bijection from the set of jagged sequences of length $2n$ to the number of tame sequences of length n.

Suppose $\mathbf{x} = (a_1, a_2, \ldots, a_{2n})$ is jagged and that $i_1 < i_2 < \cdots < i_n$ are the *indices* where the 1's occur. That is, $a_{i_1} = a_{i_2} = \cdots = a_{i_n} = 1$ and all other integers in $\mathbf{x}$ are equal to -1. We will now define a sequence $\mathbf{y} = (b_1, b_2, b_3, \ldots, b_n)$ so that b_k is the sum of the integers in $\mathbf{x}$ from a_1 up to and including the kth integer equal to 1. In symbols, $b_k = a_1 + a_2 + \cdots + a_{i_k}$.

Of the first i_k integers in $\mathbf{x}$, exactly k of them are equal to 1 and the other $i_k - k$ of them are equal to -1. Therefore, their sum (which is b_k by definition), is $b_k = k - (i_k - k) = 2k - i_k$. Thus, for a jagged sequence $\mathbf{x} = (a_1, a_2, \ldots, a_{2n})$, we define $f(\mathbf{x})$ to be the sequence $(b_1, b_2, \ldots, b_n)$ where $b_k = 2k - i_k$.

We will show that f is a bijection, but we first need to confirm that $f(\mathbf{x})$ is always tame, which means we need to show that $b_1 = 1$, $b_k \geq 1$ for all k, and that $b_{k+1} \leq b_k + 1$ for all $k < n$. We know that $a_1 = 1$ by $P2$, so this means $i_1 = 1$ and $b_1 = 2(1) - i_1 = 2 - 1 = 1$. Suppose $i_k \geq 2k$. In $\mathbf{x}$, there are only k 1's up to and including a_{i_k}, which means that among the first i_k integers in $\mathbf{x}$, there are at least as many -1's as 1's. This means the sum of the first i_k integers cannot be positive. Therefore, we must have that $i_k < 2k$, and since both are integers, $i_k + 1 \leq 2k$, which rearranges to $1 \leq 2k - i_k$, which gives $b_k \geq 1$. To see that $b_{k+1} \leq b_k + 1$, note that $i_k + 1 \leq i_{k+1}$, so $i_k - i_{k+1} \leq -1$. Then

$$b_{k+1} - b_k = 2(k+1) - i_{k+1} - (2k - i_k)$$
$$= 2k + 2 - i_{k+1} - 2k + i_k$$
$$= 2 + i_k - i_{k+1}$$
$$\leq 2 - 1$$
$$= 1$$

and so $b_{k+1} - b_k \leq 1$ or $b_{k+1} \leq b_k + 1$. This completes the proof that $f(\mathbf{x})$ is tame.

To see that f is injective, notice that $b_k = 2k - i_k$ can be rearranged to get $i_k = 2k - b_k$. In other words, if $f(\mathbf{x}) = \mathbf{y}$, then i_k is uniquely determined from b_k. This means that from $\mathbf{y}$, the

positions of the 1's in $\mathbf{x}$ are uniquely determined, so the entirety of $\mathbf{x}$ is uniquely determined by $\mathbf{y}$. This means there is only one $\mathbf{x}$ with the property that $f(\mathbf{x}) = \mathbf{y}$, so f is injective.

To see that f is surjective, suppose $\mathbf{y} = (b_1, b_2, \ldots, b_n)$ is a tame sequence. For each k from 1 through n, define $i_k = 2k - b_k$, then define $\mathbf{x} = (a_1, a_2, \ldots, a_{2n})$ so that $a_j = 1$ if $j = i_k$ for some k, and $a_j = -1$ otherwise.

That $f(\mathbf{x}) = \mathbf{y}$ follows by rearranging $i_k = 2k - b_k$ to get $b_k = 2k - i_k$. However, to conclude that f is surjective, we need to verify that $\mathbf{x}$ is indeed a jagged sequence.

By one of the conditions of tameness, $b_1 = 1$, so so we have that $i_1 = 2(1) - 1 = 1$. Using the assumption that $b_{k+1} \leq b_k + 1$ which can be rearranged to $b_k - b_{k+1} \geq -1$, we get that

$$\begin{aligned} i_{k+1} - i_k &= 2(k+1) - b_{k+1} - (2k - b_k) \\ &= 2k + 2 - b_{k+1} - 2k + b_k \\ &= 2 + b_k - b_{k+1} \\ &\geq 2 + (-1) \\ &= 1 \end{aligned}$$

which means $i_{k+1} - i_k \geq 1$ and it follows that $i_{k+1} > i_k$. Finally, since b_n is positive, $i_n = 2n - b_n < 2n$. We have shown that $1 = i_1 < i_2 < \cdots < i_n < 2n$. This shows that all of the i_k are distinct, so $\mathbf{x}$ satisfies $P1$.

Rearranging $i_k = 2k - b_k$, we get $b_k = k - (i_k - k)$, and by the reasoning from earlier, this means the sum of the first i_k integers in $\mathbf{x}$ (always ending with the kth 1) is equal to b_k, which is positive because $\mathbf{y}$ is tame. Now consider the sum $a_1 + a_2 + \cdots + a_m$ for some an arbitrary m with $1 \leq m \leq 2n$. If $m = i_k$ for some k, then the sum is positive by the reasoning just given. Otherwise, there is some k for which $i_k < m < i_{k+1}$. Since every integer in $\mathbf{x}$ strictly between a_{i_k} and $a_{i_{k+1}}$ equals -1 by construction, we must have that

$$a_1 + a_2 + \cdots + a_m \geq a_1 + a_2 + \cdots + a_m + a_{m+1} + \cdots + a_{i_{k+1}-1}$$

because the latter is obtained from the former by adding some (possibly zero) -1's. We know that $a_{i_{k+1}} = 1$ and that

$$a_1 + a_2 + \cdots + a_m + a_{m+1} + \cdots + a_{i_{k+1}-1} + a_{i_{k+1}}$$

is positive, so this means $a_1 + a_2 + \cdots + a_m \geq 0$.

We have shown that $\mathbf{x}$ satisfies $P2$ as well, so $\mathbf{x}$ is a jagged sequence with $f(\mathbf{x}) = \mathbf{y}$. Therefore, f is surjective, which completes the proof that it is bijective. Therefore, the number of tame sequences of length n is $\dfrac{1}{n+1}\dbinom{2n}{n}$.

$\square$

Problem of the Month
Additional Information about Problem 5: February 2023

Splendid sequences are part of a much more general curious combinatorial context. Here are a few definitions:

1. A *graph* is a collection of vertices (denoted in this document by circles) and edges (lines connecting some of these circles).

2. Two vertices in a graph connected by an edge are called *adjacent.*

3. A *splendid numbering* of a graph is an assignment of a positive integer to each vertex so that

   - Each number divides the sum of the integers in the adjacent vertices, and

   - the greatest common divisor of all integers is 1.

Below are some examples of splendid numberings of some graphs:
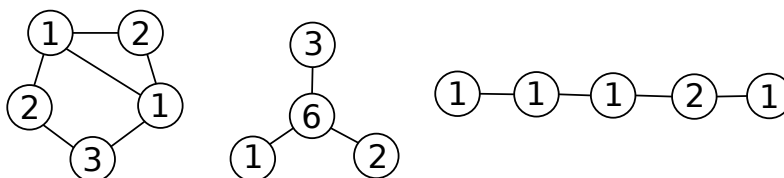
You may notice that a splendid sequence of length n is simply a splendid numbering of the graph that is a *path of length n* (the rightmost example above is a path of length 5).

So, given a graph, we can ask the same question as in the problem of the month: How many splendid numberings are there? Just like splendid sequences, the number of splendid numberings for any fixed graph is finite! (see part (e) of the problem).

The exact number of splendid numberings is known for paths of length n. This was essentially the content of part (f) of the problem.

Two other families of graph for which splendid numbers have been studied are n *cycles* and n-*pointed stars*. Rather than defining these generally, the image below has a 6 *cycle* on the left and a 6 *pointed star* on the right. We leave it to the reader to guess the general definition.
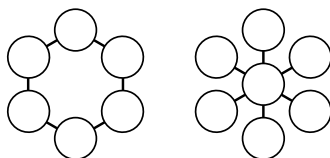
A 6 cycle (left) and a 6-pointed star (right)

The number of splendid numberings of an n cycle is known to be $\binom{2n-1}{n-1}$. This was computed in a paper from 2018 (see reference (1)), which is quite recent!

For an n-pointed star, the number of splendid numberings is closely related to the number of ways to decompose the number 1 as a sum of the form

$$1 = \frac{1}{k_1} + \frac{1}{k_2} + \cdots + \frac{1}{k_n}$$

where each k_i is a positive integer, which is a very hard problem. In fact, it is currently unknown how many splendid numberings there are for the 9-pointed star.

This is almost all of what is known at the moment, and counting the number of splendid numberings (called arithmetical structures in the research community) is an active area of research. In 2020, for example, the number of splendid numberings for a particular family of graphs called "bidents" was computed (see reference (2)). It created a new entry in The On-Line Encyclopedia of Integer Sequences (OEIS)! The new entry is entry number A335676. If you've never played around with the OEIS, it's definitely worth it. Just make sure you have a few procrastination hours to burn.

As a final note, we should mention that there are many resources out there about Catalan numbers. For example, Richard P. Stanley's 2015 book "Catalan Numbers" contains a very long list of contexts where the Catalan numbers arise.

References

(1) Benjamin Braun et al. "Counting arithmetical structures on paths and cycles". In: *Discrete Math.* 341.10 (2018), pp. 2949-2963

(2) Kassie Archer et al. "Arithmetical structures on bidents". In: *Discrete Math.* 343.7 (2020), pp. 111850, 23

Problem of the Month
Problem 6: March 2023

For a non-negative integer n, define $f(n)$ to be the first digit after the decimal point in the decimal expansion of $\sqrt{n}$. For example, $\sqrt{10} = 3.162277\ldots$ and so $f(10) = 1$. Note that $f(0) = 0$ and that $f(n) = 0$ when n is a perfect square. You will likely want a calculator that can compute square roots for this question.

(a) Compute $f(n)$ for every integer n strictly between 1 and 4 as well as every integer n strictly between 36 and 49.

(b) Compute $f(n)$ for every integer n strictly between 4 and 9 as well as every integer n strictly between 49 and 64.

(c) Show that if n is a positive multiple of 5, then each digit from 0 through 9 appears in the list
$$f(n^2 + 1), f(n^2 + 2), f(n^2 + 3), \ldots, f(n^2 + 2n - 1), f(n^2 + 2n)$$
the same number of times.

(d) For each digit d from 0 through 9, determine how many times d occurs in the list
$$f(1), f(2), f(3), \ldots, f(10^4)$$

(e) Here are a couple of other things that you might like to think about. No solution will be provided for either of these questions, but as always, we would love to hear about any observations you make!

- How are the digits 0 through 9 distributed among the infinite list
$$f(1), f(2), f(3), f(4), \ldots$$

For example, in the long run, are the ten digits distributed roughly "uniformly"? One way to make sense of this question is to think about the frequency of each digit among the list $f(1), f(2), f(3), \ldots, f(n)$ for very large n.

- Are there similar patterns to those in the earlier parts of this problem if we consider the first two digits after the decimal place? What if we consider three or more digits?

Problem of the Month
Problem 6: March 2023

Hint

(a)/(b) There is no hint given for these parts, but it might be useful in later parts to see if you notice any patterns in the distribution of the possible outputs of the function f.

(c) If $n^2 < m < (n+1)^2$, then $f(m) = d$ is equivalent to $n + \dfrac{d}{10} < \sqrt{m} < n + \dfrac{d+1}{10}$.

(d) Similar to the result in (c), if n is one more than a multiple of 5, then in the list

$$f(n^2 + 1), f(n^2 + 2), \ldots, f(n^2 + 2n)$$

every possible value from 0 through 9 appears exactly $\dfrac{n-1}{5}$ times, with the exception of 4 and 7 which appear $\dfrac{n-1}{5} + 1$ times each. Try to find and prove other similar results.

Problem of the Month
Solution to Problem 6: March 2023

Note: In this solution, we will take for granted that if n is a positive integer that is *not* a perfect square, then $\sqrt{n}$ is an irrational number.

(a) Since $\sqrt{2} \approx 1.4142$ and $\sqrt{3} \approx 1.7320$, we have that $f(2) = 4$ and $f(3) = 7$.

In the table below, the first column contains the integers n from 37 through 48, the second column contains $\sqrt{n}$ truncated to four digits past the decimal point, and the third column contains $f(n)$.

n	$\sqrt{n}$	$f(n)$
37	6.0827	0
38	6.1644	1
39	6.2449	2
40	6.3245	3
41	6.4031	4
42	6.4807	4
43	6.5574	5
44	6.6332	6
45	6.7082	7
46	6.7823	7
47	6.8556	8
48	6.9282	9

Among the values of $f(n)$ for n from 37 through 48, each integer from 0 through 9 appears exactly once except for 4 and 7, which appear exactly twice each. Notice that 4 and 7 are the values of $f(2)$ and $f(3)$, respectively.

(b) Since $\sqrt{5} \approx 2.2360$, $\sqrt{6} \approx 2.4494$, $\sqrt{7} \approx 2.6457$, and $\sqrt{8} \approx 2.8284$, we have that $f(5) = 2$, $f(6) = 4$, $f(7) = 6$, and $f(8) = 8$.

In the table below, the first column contains the integers n from 50 through 63, the second column contains $\sqrt{n}$ truncated to four digits past the decimal point, and the third column contains $f(n)$.

n	$\sqrt{n}$	$f(n)$
50	7.0710	0
51	7.1414	1
52	7.2111	2
53	7.2801	2
54	7.3484	3
55	7.4161	4
56	7.4833	4
57	7.5498	5
58	7.6157	6
59	7.6811	6
60	7.7459	7
61	7.8102	8
62	7.8740	8
63	7.9372	9

Similar to part (a), observe that among the values of $f(n)$ for n between 50 and 63 inclusive, every integer from 0 through 9 appears either once or twice, and the integers that appear twice are exactly $f(5) = 2$, $f(6) = 4$, $f(7) = 6$, and $f(8) = 8$.

(c) Suppose m is a positive integer in the list $n^2 + 1, n^2 + 2, \ldots, n^2 + 2n$. This is equivalent to $n^2 < m < (n+1)^2$, which is equivalent to $n < \sqrt{m} < n+1$ since m, n, and $n+1$ are non-negative.

Fix an integer d satisfying $0 \le d \le 9$ and assume that m is an integer satisfying both $n^2 < m < (n+1)^2$ and $f(m) = d$. That $f(m) = d$ means d is the first digit past the decimal point in the decimal expansion of $\sqrt{m}$. That $\sqrt{m}$ is between n and $n+1$ means the integer part of $\sqrt{m}$ is n. Therefore, $\sqrt{m}$ is at least $\dfrac{d}{10}$ more than n and at most $\dfrac{d+1}{10}$ more than n. Since $\sqrt{m}$ is between two consecutive perfect squares, it is not a perfect square, which means $\sqrt{m}$ is irrational. Putting this together, we get that

$$n + \frac{d}{10} < \sqrt{m} < n + \frac{d+1}{10}$$

where the strict inequalities are justified by the irrationality of $\sqrt{m}$. We are also assuming that n is a positive multiple of 5, which means there is some positive integer k such that $n = 5k$.

The quantities in the chain of inequalities above are all positive, which means

$$\left(n + \frac{d}{10}\right)^2 < \sqrt{m}^2 < \left(n + \frac{d+1}{10}\right)^2$$

Expanding and substituting $n = 5k$, we get

$$25k^2 + dk + \frac{d^2}{100} < m < 25k^2 + (d+1)k + \frac{(d+1)^2}{100}$$

We will now count the number of integers that satisfy the inequalities above.

Since the integer d satisfies $0 \leq d \leq 9$, we must have that $0 \leq \dfrac{d^2}{100} < 1$. Therefore, $25k^2 + dk + \dfrac{d^2}{100}$ is between the consecutive integers $25k^2 + dk$ and $25k^2 + dk + 1$, possibly equal to the smaller of the two but not equal to the larger of the two. Since m is an integer that is greater than $25k^2 + dk + \dfrac{d^2}{100}$, we conclude that

$$25k^2 + dk + 1 \leq m$$

Similarly, $0 \leq d \leq 9$ implies that $0 < \dfrac{(d+1)^2}{100} \leq 1$. This means $25k^2 + (d+1)k + \dfrac{(d+1)^2}{100}$ is between the consecutive integers $25k^2 + (d+1)k$ and $25k^2 + (d+1)k + 1$, possibly equal to the larger of the two but not equal to the smaller of the two. Since m is an integer that is smaller than $25k^2 + (d+1)k + \dfrac{(d+1)^2}{100}$, we can conclude that $m \leq 25k^2 + (d+1)k$. Combining this inequality with the inequality from earlier, we now have that

$$25k^2 + dk + 1 \leq m \leq 25k^2 + (d+1)k$$

The number of integers m that satisfy this inequality is

$$(25k^2 + (d+1)k) - (25k^2 + dk) = k$$

We have now shown the following: For each integer d satisfying $0 \leq d \leq 9$, there are at most k integers m in the list $n^2 + 1, n^2 + 2, \ldots, n^2 + 2n$ with the property that $f(m) = d$. There are $2n$ integers in the list above and only ten different values that the function f can output. Since $10k = 2n$, there is no possibility other than that f takes every possible value exactly $k = \dfrac{n}{5}$ times.

(d) To get an idea of how to proceed, we first discuss the results of parts (a), (b), and (c).

In part (a), we saw that for values of m between 1^2 and 2^2, the function f takes on every possible value the same number of times (zero) with the exception of 4 and 7, which occur one extra time each. For values of m between 6^2 and 7^2, the function f takes on every possible value the same number of times (one) with the exception of 4 and 7, which occur one extra time each.

Similarly, in part (b) we observed that f takes on every value the same number of times, with the exceptions of 2, 4, 6, and 8 when m ranges between 2^2 and 3^2 as well as between 7^2 and 8^2.

The result of part (c) was that if n is a multiple of 5, then $f(m)$ takes on every possible value the same number of times (with no exceptions) as m ranges over the integers strictly between n^2 and $(n+1)^2$.

The patterns in part (a) and (b) continue, and there are similar patterns for other ranges between perfect squares. A bit more precisely, if n is a positive integer, then as m ranges over the integers between n^2 and $(n+1)^2$, $f(m)$ takes on every possible value the same number of times, possibly with some exceptions that occur one extra time each. These exceptional values depend only on the remainder when n is divided by 5.

Even more precisely, the following five claims are true. The claims are numbered to correspond to remainders after division by 5.

3

Claim 0: Suppose $n = 5k$ for some integer $k \geq 0$. As m ranges over the integers strictly between n^2 and $(n+1)^2$, $f(m)$ takes every value from 0 through 9 exactly k times.

Claim 1: Suppose $n = 5k + 1$ for some integer $k \geq 0$. As m ranges over the integers strictly between n^2 and $(n+1)^2$, $f(m)$ takes the values 4 and 7 a total of $k+1$ times each and takes every other value k times.

Claim 2: Suppose $n = 5k + 2$ for some integer $k \geq 0$. As m ranges over the integers strictly between n^2 and $(n+1)^2$, $f(m)$ takes the values 2, 4, 6, and 8 a total of $k+1$ times each and takes every other value k times.

Claim 3: Suppose $n = 5k + 3$ for some integer $k \geq 0$. As m ranges over the integers strictly between n^2 and $(n+1)^2$, $f(m)$ takes the values 1, 3, 4, 6, 7, and 8 a total of $k+1$ times each and takes every other value k times.

Claim 4: Suppose $n = 5k + 4$ for some integer $k \geq 0$. As m ranges over the integers strictly between n^2 and $(n+1)^2$, $f(m)$ takes the values 1, 2, 3, 4, 5, 6, 7, and 8 a total of $k+1$ times each and takes the values 0 and 9 a total of k times each.

In part (c), we showed that Claim 0 is true. Before proving Claims 1, 2, 3, and 4, we will answer the actual question which was to determine how many times each digit occurs in the list

$$f(1), f(2), f(3), \ldots, f(10^4)$$

Because 10^4 is a perfect square, $f(0) = f(10^4) = 0$, so we can instead count the number of times each digit occurs in the slightly different list

$$f(0), f(1), f(2), \ldots, f(10^4 - 1)$$

There are 100 perfect squares among the integers from 0 through $10^4 - 1$, so there are 100 occurrences of 0 in the list that come from perfect squares.

For every other integer m in the list $0, 1, 2, 3, 4, \ldots, 10^4 - 1$, there are unique integers k and r with $0 \leq k \leq 19$ and $0 \leq r \leq 4$ with the property that $(5k + r)^2 < m < (5k + r + 1)^2$.

Suppose d is a non-zero digit. For fixed k and r, as we let m take the values strictly between $(5k+r)^2$ and $(5k+r+1)^2$, Claim r tells us whether the digit d occurs either k times or $k+1$ times in that range. We are considering values of m satisfying $(5k+r)^2 < m < (5k+r+1)^2$ for every pair (k, r) with $0 \leq k \leq 19$ and $0 \leq r \leq 4$. For every pair of the form $(k, 0)$, we get k occurrences of d (see Claim 0). For every pair of the form $(k, 1)$, we always get k occurrences or we always get $k + 1$ occurrences of d, depending on what Claim 1 says about the digit d. Continuing with this reasoning, the number of occurrences of d in the list $f(0), f(1), f(2), \ldots, f(10^4 - 1)$ can be expressed in the form

$$a(0 + 1 + 2 + \cdots + 18 + 19) + b(1 + 2 + 3 + \cdots + 19 + 20) = 190a + 210b$$

where a is the number of values of r for which Claims 0 through 4 say there are k occurrences of d, and b is the number of values of r for which Claims 0 through 4 say there are $k + 1$ occurrences of d.

For example, for $d = 1$, there are k occurrences when $r = 0$, $r = 1$, and $r = 2$, and there are $k + 1$ occurrences when $r = 3$ and $r = 4$. Therefore, with $d = 1$, we have $a = 3$ and $b = 2$, so the number of times that $d = 1$ occurs in the list is $3 \times 190 + 2 \times 210 = 990$.

4

Using Claims 0 through 4, the table below summarizes the count for the remaining non-zero digits. In the first column, the digit d appears, in the second column the value of a appears, in the third column, the value of b occurs, and in the fourth column, $190a + 210b$ appears, which is the number of times d appears in the list.

d	a	b	$190a + 210b$
1	3	2	990
2	3	2	990
3	3	2	990
4	1	4	1030
5	4	1	970
6	2	3	1010
7	2	3	1010
8	2	3	1010
9	5	0	950

Finally, for $d = 0$, we already have 100 occurrences of 0 in the list coming from the perfect squares. Otherwise, we can use the exact same technique as above to count the number of 0s in the list coming from integers that are not perfect squares. For $d = 0$, $a = 5$ and $b = 0$, so there are $5(190) + 0(210) + 100 = 1050$ occurrences of the digit 0 in the list $f(0), f(1), f(2), \ldots, f(10^4 - 1)$.

We will now prove Claims 1 through 4.

Suppose k is a non-negative integer, r is 0, 1, 2, 3, or 4, and that d is a digit between 0 and 9 inclusive. We wish to count the number of integers m such that $f(m) = d$ and $(5k + r)^2 < m < (5k + r + 1)^2$. Specifically, we want to show that this number is either k or $k + 1$ depending on r and d in the way delineated in Claims 0 through 4.

The condition $(5k + r)^2 < m < (5k + r + 1)^2$ is equivalent to $5k + r < \sqrt{m} < 5k + r + 1$, and this combined with $f(m) = d$ (note that m is not a perfect square and that $5k + r$ and $5k + r + 1$ are consecutive integers) is equivalent to

$$5k + r + \frac{d}{10} < \sqrt{m} < 5k + r + \frac{(d+1)}{10}.$$

Since everything is non-negative, everything can be squared to get the equivalent chain of inequalities

$$(5k + r)^2 + \frac{2(5k + r)d}{10} + \frac{d^2}{100} < m < (5k + r)^2 + \frac{2(5k + r)(d + 1)}{10} + \frac{(d + 1)^2}{100}.$$

After some expansion, we have

$$(5k + r)^2 + kd + \frac{2rd}{10} + \frac{d^2}{100} < m < (5k + r)^2 + kd + k + \frac{2r(d + 1)}{10} + \frac{(d + 1)^2}{100}.$$

We are interested in how many integers m satisfy the chain of inequalities above. Since $(5k + r)^2 + kd$ is an integer, the number of integers m satisfying the inequalities above is the same as the number of integers m' that satisfy

$$\frac{2rd}{10} + \frac{d^2}{100} < m' < k + \frac{2r(d + 1)}{10} + \frac{(d + 1)^2}{100}$$

5

and after combining fractions on the left and right, we get

$$\frac{20rd + d^2}{100} < m' < k + \frac{20r(d+1) + (d+1)^2}{100} \tag{$*$}$$

Suppose $\dfrac{20rd + d^2}{100}$ and $\dfrac{20r(d+1) + (d+1)^2}{100}$ are both strictly between 0 and 1. Then the integers m' satisfying $(*)$ are precisely the integers 1 through k. More generally, we will show that whether there are k or $k+1$ integers depends on whether or not the quantities $\dfrac{20rd + d^2}{100}$ and $\dfrac{20r(d+1) + (d+1)^2}{100}$ are between the same pair of consecutive integers.

To give an idea of how the argument will proceed, consider the situation when $r = 1$ and $d = 4$. Then $\dfrac{20rd + d^2}{100} = \dfrac{96}{100}$, but $\dfrac{20r(d+1) + (d+1)^2}{100} = \dfrac{125}{100}$. Thus, $(*)$ becomes

$$\frac{96}{100} < m' < k + \frac{125}{100}$$

The integers $m' = 1$ through $m' = k + 1$ satisfy these inequalities, for a total of $k + 1$ integers.

Observe that

$$\begin{aligned}
\frac{20r(d+1) + (d+1)^2}{100} - \frac{20rd + d^2}{100} &= \frac{(20rd + 20r + d^2 + 2d + 1) - (20rd + d^2)}{100} \\
&= \frac{20r + 2d + 1}{100} \\
&\leq \frac{20(4) + 2(9) + 1}{100} \qquad \text{(since } r \leq 4,\ d \leq 19\text{)} \\
&= \frac{99}{100}
\end{aligned}$$

This means the difference between $\dfrac{20rd + d^2}{100}$ and $\dfrac{20r(d+1) + (d+1)^2}{100}$ is less than 1, which leads to the following two possibilities.

<u>Possibility 1</u>: There is an integer a so that

$$a \leq \frac{20rd + d^2}{100} < \frac{20r(d+1) + (d+1)^2}{100} \leq a + 1$$

In this case, the integers satisfying $(*)$ are $a + 1$, $a + 2$, $a + 3$, and so on through $a + k$ for a total of k integers.

<u>Possibility 2</u>: There is some integer a with the property that

$$a < \frac{20rd + d^2}{100} < a + 1 < \frac{20r(d+1) + (d+1)^2}{100} < a + 2$$

In this case, the integers satisfying $(*)$ are $a + 1$ through $a + 1 + k$, for a total of $k + 1$ integers.

We have now shown that there will be $k + 1$ integers satisfying $(*)$ exactly when there is an integer strictly between $\dfrac{20rd + d^2}{100}$ and $\dfrac{20r(d+1) + (d+1)^2}{100}$. Multiplying through

6

by 100, this is equivalent to there being a multiple of 100 strictly between $20rd + d^2$ and $20r(d+1) + (d+1)^2$

The table below contains the values of $20rd + d^2$ for every $0 \le r \le 4$ and $0 \le d \le 9$. The columns after the first are indexed by a value of d from 0 through 9 from left to right and the rows after the first are indexed by values of r from 0 through 4 from top to bottom. The cell in the row corresponding to r and the column corresponding to d contains $20rd + d^2$. Cells are highlighted if there is a multiple of 100 between the value in the cell and the value in the cell immediately to its right.

r \ d	0	1	2	3	4	5	6	7	8	9	10
0	0	1	4	9	16	25	36	49	64	81	100
1	0	21	44	69	96	125	156	189	224	261	300
2	0	41	84	129	176	225	276	329	384	441	500
3	0	61	124	189	256	325	396	469	544	621	700
4	0	81	164	249	336	425	516	609	704	801	900

The highlighted cells correspond to pairs (r, d) for which $20rd + d^2$ is less than a multiple of 100 and $20r(d+1) + (d+1)^2$ is greater than that multiple of 100. As noted above, these correspond to the pairs (r, d) for which $k + 1$ integers satisfy $(*)$. For $r = 0$, we see no values of d, which gives another proof of Claim 0. For $r = 1$, the values of d for which there are $k + 1$ integers are $d = 4$ and $d = 7$, which proves Claim 1. Continuing, the data in the table above verifies Claims 0 through 4.

Problem of the Month
Problem 7: April 2023

Let A_n denote the set of all n-tuples of 0s and 1s. For example, A_2 is the set of all ordered pairs of 0s and 1s or $A_2 = \{(0,0),(0,1),(1,0),(1,1)\}$. To improve readability, we will omit the commas and parentheses from the elements of A_n. For example, the elements of A_3 will be denoted by 000, 001, 010, 011, 100, 101, 110, and 111.

Variables referring to elements of A_n will be bold lowercase letters. For example, we might refer to elements in A_2 as $\boldsymbol{a}$, or $\boldsymbol{b}$, and so on. To refer to the coordinates of elements in A_n, we will use square brackets. For example, if $\boldsymbol{a} = 11010$, an element in A_5, then $\boldsymbol{a}[1] = 1$, $\boldsymbol{a}[2] = 1$, $\boldsymbol{a}[3] = 0$, $\boldsymbol{a}[4] = 1$, and $\boldsymbol{a}[5] = 0$.

For two elements $\boldsymbol{a}$ and $\boldsymbol{b}$ in A_n, the *Hamming distance*, denoted $d(\boldsymbol{a},\boldsymbol{b})$ between $\boldsymbol{a}$ and $\boldsymbol{b}$ is equal to the number of coordinates where they differ. For example, if $\boldsymbol{a} = 11010$ and $\boldsymbol{b} = 01110$, then $d(\boldsymbol{a},\boldsymbol{b}) = 2$ because $\boldsymbol{a}[1] \neq \boldsymbol{b}[1]$ and $\boldsymbol{a}[3] \neq \boldsymbol{b}[3]$, but $\boldsymbol{a}[i] = \boldsymbol{b}[i]$ for $i = 2$, $i = 4$, and $i = 5$. It is important to convince yourself that $d(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{b},\boldsymbol{a})$ for any $\boldsymbol{a}$ and $\boldsymbol{b}$.

The notion of a *graph* was defined in the extra information about the February 2023 problem, but there are also plenty of places online that have definitions. We will keep things simple here and define a graph to be a collection of *vertices*, some of which are connected to each other by *edges*. When we draw a graph, a vertex will be represented by a circle and an edge will be represented by a line segment from one vertex to another. Two examples of graphs are depicted below. The one on the left has four vertices and the one on the right has eight. Note that two edges intersecting does not necessarily imply the presence of a vertex.
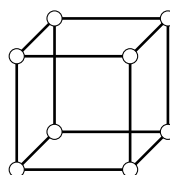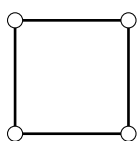

For each n, we define a graph with 2^n vertices called the *natural graph* of A_n. In the natural graph of A^n, it is possible to label every vertex by exactly one element of A_n such that there is an edge between two vertices exactly when their Hamming distance is 1. The two examples above are the natural graphs of A_2 and A_3. They are shown again below with their vertices labelled.
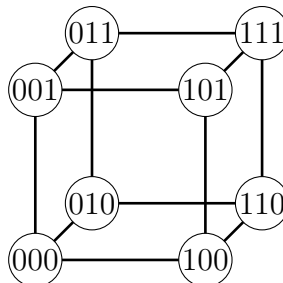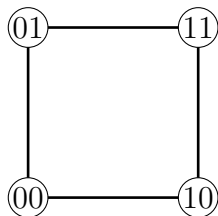
A *walk* in a graph from vertex v to vertex w is a sequence of vertices starting at v and ending at w with the property that there is an edge connecting every pair of consecutive vertices in the sequence. The *length* of a walk is the number of edges it uses. For example, let v, w, x, and y be the vertices labelled by 000, 110, 100, and 010 in the natural graph of A_3, respectively. Then v, x, w and v, y, w are walks of length 2 from v to w. The *distance* between v and w in a graph is equal to the shortest possible length of a walk from v to w. In the example above, v and w have a distance of 2 because there are walks of length 2, but there are no shorter walks from v to w.

(a) Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be elements of A_n. Show that $d(\boldsymbol{a}, \boldsymbol{b})$ is equal to the distance between $\boldsymbol{a}$ and $\boldsymbol{b}$ in the natural graph of A_n.

(b) For each k with $1 \le k \le n$, determine the number of two-element subsets $\{\boldsymbol{a}, \boldsymbol{b}\}$ of A_n that satisfy $d(\boldsymbol{a}, \boldsymbol{b}) = k$.

(c) Suppose we were to relabel the vertices of the natural graph of A_n by permuting the labels. That is, we keep the graph the same but use the elements of A_n to label a vertex of the graph in some other way. For example, in A_2, we might leave 00 and 10 where they are and swap the positions of 01 and 11, as shown.
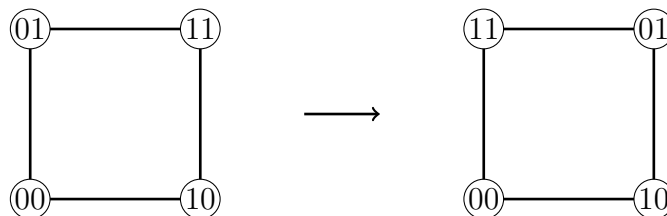
When this is done, the distance in the new graph between elements of $\boldsymbol{a}$ and $\boldsymbol{b}$ is not necessarily equal to $d(\boldsymbol{a}, \boldsymbol{b})$ any more. The table below has the two-element subsets of A_2 in the first column, $d(\boldsymbol{a}, \boldsymbol{b})$ in the second column, and their distance in the relabelled graph in the third column.

$\{\boldsymbol{a}, \boldsymbol{b})$	$d(\boldsymbol{a}, \boldsymbol{b})$	new distance
$\{00, 01\}$	1	2
$\{00, 10\}$	1	1
$\{00, 11\}$	2	1
$\{01, 10\}$	2	1
$\{01, 11\}$	1	1
$\{10, 11\}$	1	2

Among the four subsets $\{\boldsymbol{a}, \boldsymbol{b}\}$ with $d(\boldsymbol{a}, \boldsymbol{b}) = 1$, there are two that have a distance in the relabelled graph of 1 and two that have a distance in the relabelled graph of 2.

Now for the question: For each n, find a way to permute the elements of A_n so that the following happens: Among all two-element subsets $\{\boldsymbol{a}, \boldsymbol{b}\}$ of A_n with $d(\boldsymbol{a}, \boldsymbol{b}) = 1$, there are the same number with each possible distance in the relabelled graph.

Problem of the Month
Problem 7: April 2023

Hint

(a) Suppose $d(\boldsymbol{a}, \boldsymbol{b}) = k$ for some k. Try to construct a path of length k in the natural graph from the vertex labelled $\boldsymbol{a}$ to the vertex labelled $\boldsymbol{b}$.

(b) For fixed $\boldsymbol{a} \in A_n$, how many $\boldsymbol{b} \in A_n$ have the property that $d(\boldsymbol{a}, \boldsymbol{b}) = k$?

(c) Find a function that works for $n = 2$ and use this to build one for $n = 3$. It might be useful to think of the natural graph of A_2 as a square and the natural graph of A_3 as a cube. As well, a cube can be thought of as two squares on top of each other with vertical edges connecting corresponding vertices in the top and bottom faces.

Problem of the Month
Solution to Problem 7: April 2023

(a) Suppose $\boldsymbol{a}$ and $\boldsymbol{b}$ are elements in A_n and that $d(\boldsymbol{a}, \boldsymbol{b}) = k$ for some k. If $k = 0$, then $\boldsymbol{a} = \boldsymbol{b}$, so their distance in the graph is also 0.

Otherwise, $\boldsymbol{a}$ and $\boldsymbol{b}$ differ at exactly k coordinates $i_1, i_2, \ldots, i_k$ where $i_1 < i_2 < \cdots < i_k$. Using the notation introduced in the problem statement, we mean that $\boldsymbol{a}[i] \neq \boldsymbol{b}[i]$ if i is in the list $i_1, i_2, \ldots, i_k$ and $\boldsymbol{a}[i] = \boldsymbol{b}[i]$ otherwise. Notice that the function $g(x) = 1 - x$ has the property that $g(0) = 1$ and $g(1) = 0$, so g switches 1 and 0. We will now define a sequence $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ of elements in A_n. Informally, $\boldsymbol{a}_1$ is obtained from $\boldsymbol{a}$ by leaving all coordinates alone except $\boldsymbol{a}[i_1]$, which gets changed from 0 to 1 or 1 to 0 as appropriate. Continuing, $\boldsymbol{a}_2$ is obtained from $\boldsymbol{a}_1$ by leaving all coordinates alone except $\boldsymbol{a}_1[i_2]$, which gets switched, and this continues for $\boldsymbol{a}_3$, $\boldsymbol{a}_4$, and so on. More formally, for each $m \geq 1$ with $1 \leq m \leq k$ we define $\boldsymbol{a}_m$ as follows.

- $\boldsymbol{a}_m[i] = \boldsymbol{a}[i]$ if i is not in the list $i_1, i_2, \ldots, i_k$.

- $\boldsymbol{a}_m[i] = g(\boldsymbol{a}[i])$ for each i in the list $i_1, i_2, \ldots, i_m$.

- $\boldsymbol{a}_m[i] = \boldsymbol{a}[i]$ for each i in the list $i_{m+1}, i_{m+2}, \ldots, i_k$.

By construction, the list $\boldsymbol{a}, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{k-1}, \boldsymbol{a}_k$ is a list in which every pair of elements differ at exactly one coordinate. Moreover, the list is that which is generated by changing the coordinates of $\boldsymbol{a}$ that differ from those of $\boldsymbol{b}$ one at a time, from leftmost to rightmost. This means $\boldsymbol{b} = \boldsymbol{a}_k$, and the above is a walk from $\boldsymbol{a}$ to $\boldsymbol{b}$. There are $k + 1$ vertices in this walk, so there are k edges.

We have constructed a walk from $\boldsymbol{a}$ to $\boldsymbol{b}$ in the natural graph of A_n that has length k, which means the distance from $\boldsymbol{a}$ to $\boldsymbol{b}$ in the natural graph is at most k. To see that it is at least k, we suppose $\boldsymbol{a}, \boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_{m-1}, \boldsymbol{b}$ is a walk in the natural graph of A_n of length m for some m. Since there are m vertices in this walk in addition to $\boldsymbol{a}$ and two vertices have an edge between them exactly when they differ at exactly one coordinate, the total number of coordinates at which $\boldsymbol{a}$ and $\boldsymbol{b}$ (the ends of the walk) can differ is at most m. Since we know that they differ at exactly k coordinates, we must have that $m \geq k$. This means that any walk from $\boldsymbol{a}$ to $\boldsymbol{b}$ in the natural graph of A_n has at least k edges.

We have shown that the distance in the natural graph between $\boldsymbol{a}$ and $\boldsymbol{b}$ is at least k and at most k, which means it is exactly k.

(b) For convenience, in the solution to this part and the solution to part (c), we will refer to a two element subset as a *pair*. Since $d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{b}, \boldsymbol{a})$ for any elements $\boldsymbol{a}, \boldsymbol{b} \in A_n$, we will say that $d(\boldsymbol{a}, \boldsymbol{b})$ is *the Hamming distance of the pair* $\{\boldsymbol{a}, \boldsymbol{b}\}$ or *the pair* $\{\boldsymbol{a}, \boldsymbol{b}\}$ *has Hamming distance* $d(\boldsymbol{a}, \boldsymbol{b})$), and possibly other similar things depending on the grammar in that particular sentence. Similarly, we might say that $\{\boldsymbol{a}, \boldsymbol{b}\}$ *has distance k in a graph* to mean that the distance between the vertices labelled by $\boldsymbol{a}$ and $\boldsymbol{b}$ is k in that graph.

For a fixed element $\boldsymbol{a} \in A_n$, an element $\boldsymbol{b} \in A_n$ satisfies $d(\boldsymbol{a}, \boldsymbol{b}) = k$ exactly when it differs from $\boldsymbol{a}$ at exactly k coordinates. There are n coordinates in total, so there are

$\binom{n}{k}$ possible choices of k coordinates that could be different from $\boldsymbol{a}$. Therefore, for a fixed element $\boldsymbol{a} \in A_n$, there are exactly $\binom{n}{k}$ elements $\boldsymbol{b} \in A_n$ with the property that $d(\boldsymbol{a}, \boldsymbol{b}) = k$. There are 2^n elements in A_n and each $\boldsymbol{a} \in A$ belongs to exactly $\binom{n}{k}$ pairs with a Hamming distance of k. This gives a total of $2^n \times \binom{n}{k}$ pairs. However, this total counts every pair twice, once for each of its two elements. Thus, the number of pairs in A_n with a Hamming distance of k is $\frac{1}{2} 2^n \binom{n}{k} = 2^{n-1} \binom{n}{k}$.

(c) Denote by E_n the set of pairs $\{\boldsymbol{a}, \boldsymbol{b}\}$ from A_n satisfying $d(\boldsymbol{a}, \boldsymbol{b}) = 1$. From part (b), there are $2^{n-1} \binom{n}{1} = n2^{n-1}$ pairs in E_n. In the relabelled natural graph of A_n, we want the distances of the pairs in E_n to be equally distributed among all possible distances in the graph. There are n possible distances between distinct vertices in the graph, so the fact that $n2^{n-1}$ is a multiple of n is a good sign.

We want to permute the elements of A_n in such a way that for each k from 1 through n, exactly $\dfrac{n2^{n-1}}{n} = 2^{n-1}$ pairs in E_n have a distance of k in the relabelled graph.

There are many ways to do this. The approach given here is inductive, starting by examining A_2. Consider the example from the problem statement. In that example, 01 and 11 were switched and 00 and 10 stayed the same
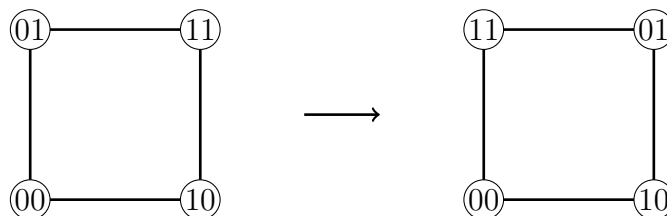
Although it is not very interesting in A_2, there is an observation we can make that will generalize. The vertices that are connected by horizontal edges in the diagram of the natural graph of A_2 remain connected by an edge after permuting. The vertices in the top change order but their distance apart does not change. Meanwhile, the vertices that are connected by a vertical edge are moved to occupy opposite corners of the square so their distance goes from 1 to 2. While this is only an increase of 1, it will be useful going forward to think of the vertices connected by vertical edges as having gone from as close together as possible (connected by an edge) to as far apart as possible.

Now consider the natural graph of A_3, which is pictured below.
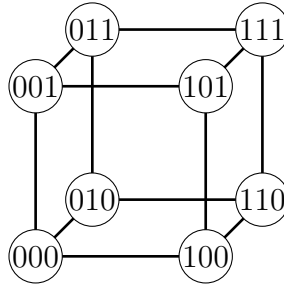
There are 12 edges in the graph. After permuting the labels of the graph, we want $\dfrac{12}{3} = 4$ pairs of adjacent vertices to be sent to adjacent vertices, 4 pairs of adjacent vertices to be sent to vertices with a distance of 2, and 4 pairs of adjacent vertices to be sent to vertices with a distance of 3.

Looking at the diagram, we can think of the natural graph of A_3 as being composed of two copies of the natural graph of A_2 laid horizontally on top of each other. The labelling also has some coherence with the labelling of A_2. First, label the bottom and top square as if they were copies of the natural graph of A_2, making sure to label vertically-adjacent vertices in the same way. Next, append a 0 to the right of every label in the bottom layer and append a 1 to the right of every label in the top layer.

To permute the labels in the way we want, we will first perform the same permutation on each layer as we did in A_2. In each layer, this moves two pairs of labels from adjacent vertices so that they are at a distance of 2. There are also two pairs of adjacent vertices in each layer that remain adjacent after permuting. The four pairs of vertically-adjacent vertices will remain vertically adjacent because we will have performed the exact same permutation in each layer. At this point, the labels on four pairs of adjacent vertices have been sent to vertices that are 2 apart and the other 8 pairs of labels remain on adjacent vertices. The diagram below shows what we have done so far:
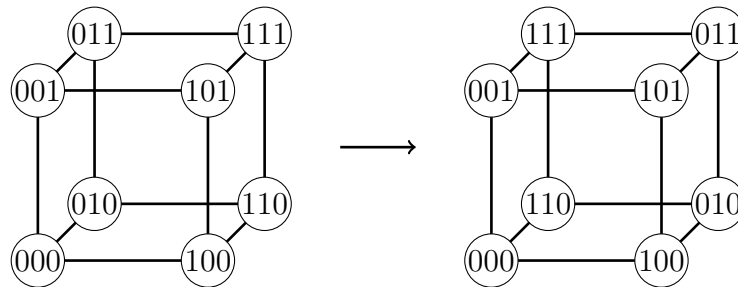
The second and final step is to swap the corners in the top layer. This will have the effect of moving the labels on vertically-adjacent vertices to be as far apart as possible. In a "cube", this means they will end up at opposite ends of a "space diagonal". Swapping the corners in a layer preserves the distance between all pairs of vertices in that layer. This means the net effect of the second step is to move four pairs of adjacent vertices so that they are at a distance of 3. The overall effect is shown below:

3
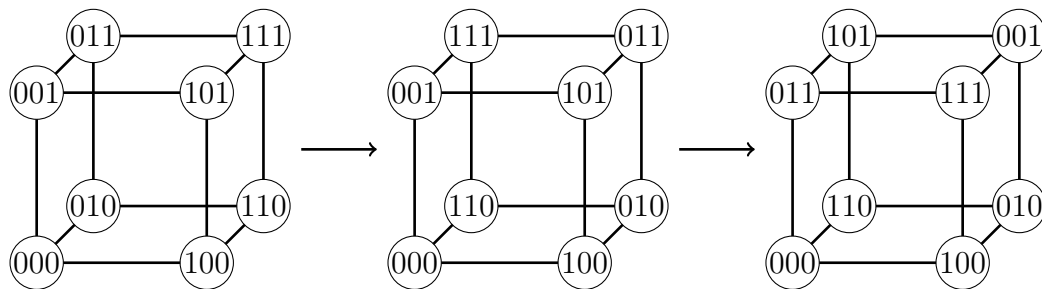
In the table below, the first column has the 12 pairs from E_3 and the second column has the distance of the corresponding pair in the relabelled graph.

$\{a, b\}$	new distance
$\{000, 001\}$	3
$\{000, 010\}$	2
$\{000, 100\}$	1
$\{001, 011\}$	2
$\{001, 101\}$	1
$\{010, 011\}$	3
$\{010, 110\}$	1
$\{011, 111\}$	1
$\{100, 101\}$	3
$\{100, 110\}$	2
$\{101, 111\}$	2
$\{110, 111\}$	3

As n grows, the natural graph of A_n gets harder and harder to draw in a useful way, so we need some notation to help translate the geometric idea into symbols. First, we will clarify what we actually want.

Suppose f is a function with *domain* A_n and *codomain* A_n. This means f is a function that takes elements of A_n as input and also outputs elements of A_n. When we talk about a permutation of A_n, we really mean a function f with domain and codomain both equal to A_n that is a *bijection*. For a brief discussion about what a bijection is, you can consult Appendix 1 from the solution to the February 2023 problem. In the context of this solution, a function from A_n to A_n is a permutation if every possible output is attained by exactly one input. For example, the function f with domain and codomain A_2 given by

$$f(00) = 11$$
$$f(01) = 01$$
$$f(10) = 00$$
$$f(11) = 10$$

is a bijection from A_2 to A_2. Every possible output is attained (the four elements of A_2 appear on the right side of the displayed equations above) and no output is attained more than once. If you think about it, every way to order the elements of A_2 (a permutation) corresponds to exactly one such function: choose an order of the elements, then write them in that order in the second column above. It will not be important for this solution, but it might help you to understand this connection if you convince yourself that there are exactly $4! = 24$ bijections from A_2 to itself.

A rearrangement of the labels in the natural graph of A_n can be thought of as a bijection from A_n to itself. If f is such a function, then $f(\boldsymbol{a})$ is equal to the original label of the vertex to which the label $\boldsymbol{a}$ is sent by the permutation. For example, the permutation for A_3 corresponding to the relabelling from earlier (shown below)
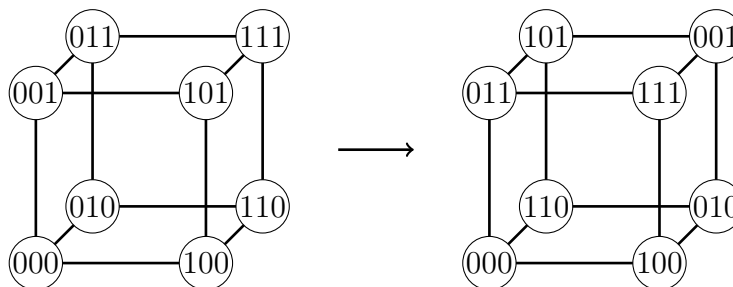
is given by

$$f(000) = 000$$
$$f(001) = 111$$
$$f(010) = 110$$
$$f(011) = 001$$
$$f(100) = 100$$
$$f(101) = 011$$
$$f(110) = 010$$
$$f(111) = 101$$

For example, the fourth of the displayed equations above is $f(011) = 001$ because in the second diagram 011 appears where 001 originally appeared.

This is an important observation because we can now recognize distance in the relabelled graph as a Hamming distance. The distance between $\boldsymbol{a}$ and $\boldsymbol{b}$ in the relabelled graph is equal to the distance between $f(\boldsymbol{a})$ and $f(\boldsymbol{b})$ in the original graph. By part (a), this means the distance between $\boldsymbol{a}$ and $\boldsymbol{b}$ in the relabelled graph is equal to $d(f(\boldsymbol{a}), f(\boldsymbol{b}))$.

We can now formally articulate what we seek. For each n, we would like a function f_n with domain and codomain both equal to A_n with the following properties.

- f_n is a permutation of A_n (a bijection).

- Among the $n2^{n-1}$ pairs $\{\boldsymbol{a}, \boldsymbol{b}\}$ from E_n, $d(f_n(\boldsymbol{a}), f_n(\boldsymbol{b}))$ takes on each value from 1 through n exactly 2^{n-1} times.

It may be a good idea to digest what has been said so far, possibly going back to see how this applies to A_2 and A_3.

We can now define f_n for each n, but the definition will be recursive, so we need a bit more notation. For an element $\boldsymbol{a}$ in A_n where $n \geq 1$, we will write $\boldsymbol{a}|0$ to mean the element of A_{n+1} that is obtained by appending a 0 to the right end of $\boldsymbol{a}$. For example, $00110|0 = 001100$. Similarly, $\boldsymbol{a}|1$ is the element of A_{n+1} obtained by appending 1 to the right end of $\boldsymbol{a}$. Also, we will denote by $\overline{\boldsymbol{a}}$ the element of A_n obtained by changing every

coordinate of $\boldsymbol{a}$ from 0 to 1 or 1 to 0, as appropriate. For example, if $\boldsymbol{a} = 0010111$, then $\overline{\boldsymbol{a}} = 1101000$.

Starting with $n = 1$ (which we have not addressed up to this point) we will let $f_1(\boldsymbol{a}) = \boldsymbol{a}$. That is, $f_1(\boldsymbol{a})$ is the *identity* function. The elements of A_1 are 0 and 1, and $f_1(0) = 0$ and $f_1(1) = 1$. We now continue recursively. For $n \geq 1$, we define f_{n+1} from f_n as follows.

- If $\boldsymbol{a} \in A_{n+1}$ is of the form $\boldsymbol{b}|0$ for some $\boldsymbol{b} \in A_n$, then $f_{n+1}(\boldsymbol{a}) = f_n(\boldsymbol{b})|0$.

- If $\boldsymbol{a} \in A_{n+1}$ is of the form $\boldsymbol{b}|1$ for some $\boldsymbol{b} \in A_n$, then $f_{n+1}(\boldsymbol{a}) = \overline{f_n(\boldsymbol{b})}|1$.

Notice that the above instructions indeed explain how to evaluate f_{n+1} at every element of A_{n+1} because each element of A_{n+1} can be constructed in exactly one way by appending either a 0 or a 1 to the right of an element from A_n. As an example, we will determine exactly what f_2 does to each element in A_2. For 00, we have to use the rule in the first bullet point because the rightmost digit is 0. $f_1(0) = 0$, so we have that $f_2(00) = 00$. Since the second digit of 10 is also 0 and $f_1(1) = 1$, we get that $f_2(10) = 10$. For 01, we have to use the rule in the second bullet point. This means $f_2(01) = \overline{f_1(0)}|1 = \overline{0}|1 = 11$. Finally, $f_2(11) = \overline{f_1(1)}|1 = \overline{1}|1 = 01$. This is exactly the function from A_2 to itself discussed earlier. We leave it as an exercise to verify that f_3 is exactly the permutation of A_3 discussed earlier.

We can now prove by induction that f_n does what we want for every n. Before doing that, we will discuss how this function corresponds to the geometric idea from earlier. The elements in A_{n+1} can be obtained by taking each element of A_n and appending a 0 to the right and a 1 to the right, in a way getting two elements in A_{n+1} from every element in A_n. By this reasoning, A_{n+1} can be thought of as two copies of A_n: elements ending in 0 and elements ending in 1. If you look at the natural graph of A_3 above, these two copies are exactly the "bottom" and the "top" squares. The way f_{n+1} is defined is to operate differently on the two copies of A_n, since how $f_{n+1}(\boldsymbol{a})$ is computed depends on the rightmost digit of $\boldsymbol{a}$. In other words, it depends on which copy of A_n $\boldsymbol{a}$ belongs to. If $\boldsymbol{a}$ has a rightmost digit of 0, then f_{n+1} essentially does what f_n did. This corresponds to the bottom face of the cube being permuted exactly as the square was. If the rightmost digit of $\boldsymbol{a}$ is 1, then $\boldsymbol{a}$ is in the other copy, corresponding to the top face of the cube in the case of A_3. In this situation, we apply f_n to the element of A_n obtained by removing the last digit, just as we would if the rightmost digit were 0. However, we then switch every digit, and this corresponds to "swapping the diagonals" in A_3. Finally, a 1 is appended to the right of the result, which corresponds to making sure that elements in the top of the cube stay in the top of the cube during the permutation.

For $n = 1$, it is an exercise in understanding definitions to see that f_n satisfies the given conditions. We have already established this for $n = 2$, and $n = 3$ can be verified by confirming that f_3 is exactly the permutation from earlier that we verified worked.

We now assume, for some $n \geq 1$, that f_n is a permutation of A_n with the property that among pairs $\{\boldsymbol{a}, \boldsymbol{b}\}$ from E_n, $d(f_n(\boldsymbol{a}), f_n(\boldsymbol{b}))$ takes on every value from 1 through n exactly 2^{n-1} times. We will show that f_{n+1} is a permutation of A_{n+1} with the property that among pairs $\{\boldsymbol{a}, \boldsymbol{b}\}$ from E_{n+1}, $d(f_{n+1}(\boldsymbol{a}), f_{n+1}(\boldsymbol{b}))$ takes on every value from 1 through $n + 1$ exactly 2^n times.

To see that f_{n+1} is a permutation, let $\boldsymbol{a} \in A_n$ be arbitrary. Since f_n is a permutation,

there is a unique element $\boldsymbol{b} \in A_n$ such that $f_n(\boldsymbol{b}) = \boldsymbol{a}$ and a unique element $\boldsymbol{c} \in A_n$ such that $f_n(\boldsymbol{c}) = \overline{\boldsymbol{a}}$. Using the definition of f_{n+1}, we have $f_{n+1}(\boldsymbol{b}|0) = f_n(\boldsymbol{b})|0 = \boldsymbol{a}|0$ and $f_{n+1}(\boldsymbol{c}|1) = \overline{f_n(\boldsymbol{c})}|1 = \overline{\overline{\boldsymbol{a}}}|1 = \boldsymbol{a}|1$. Since every element in A_{n+1} is of the form $\boldsymbol{a}|0$ or $\boldsymbol{a}|1$ for some $\boldsymbol{a} \in A_n$, we have shown that every element of A_{n+1} is in the range of f_{n+1}. Now suppose $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \ldots, \boldsymbol{a}_{2^{n+1}}$ are the elements of A_{n+1} (in some order). We have shown that every element of A_{n+1} appears in the list

$$f_{n+1}(\boldsymbol{a}_1), f_{n+1}(\boldsymbol{a}_2), \ldots, f_{n+1}(\boldsymbol{a}_{2^{n+1}})$$

at least once. There are 2^{n+1} elements in the list above and 2^{n+1} elements in A_{n+1}, so every element of A_{n+1} must appear in the list above exactly once. In other words, f_{n+1} is a permutation of A_{n+1}.

To prove the other fact about f_{n+1}, we will use the fact that $d(\boldsymbol{a}, \boldsymbol{b}) = d(\overline{\boldsymbol{a}}, \overline{\boldsymbol{b}})$ for any $\boldsymbol{a}$ and $\boldsymbol{b}$. It is left as an exercise to verify this.

Suppose k is a positive integer such that $1 \leq k \leq n$. By induction, there are exactly 2^{n-1} pairs $\{\boldsymbol{a}, \boldsymbol{b}\}$ in E_n with $d(f_n(\boldsymbol{a}), f_n(\boldsymbol{b})) = k$. If $\{\boldsymbol{a}, \boldsymbol{b}\}$ is one of these 2^{n-1} pairs, we have

$$\begin{aligned}
d(f_{n+1}(\boldsymbol{a}|0), f_{n+1}(\boldsymbol{b}|0)) &= d(f_n(\boldsymbol{a})|0, f_n(\boldsymbol{b})|0) \\
&= d(f_n(\boldsymbol{a}), f_n(\boldsymbol{b})) \\
&= k
\end{aligned}$$

where the second equality is because the elements in question agree in the last coordinate, so their Hamming distance is equal to the Hamming distance between the elements obtained by removing the rightmost elements. We similarly have that

$$\begin{aligned}
d(f_{n+1}(\boldsymbol{a}|1), f_{n+1}(\boldsymbol{b}|1)) &= d(\overline{f_n(\boldsymbol{a})}|1, \overline{f_n(\boldsymbol{b})}|1) \\
&= d(\overline{f_n(\boldsymbol{a})}, \overline{f_n(\boldsymbol{b})}) \\
&= d(f_n(\boldsymbol{a}), f_n(\boldsymbol{b})) \\
&= k
\end{aligned}$$

Therefore, there are $2 \times 2^{n-1} = 2^n$ pairs $\{\boldsymbol{a}, \boldsymbol{b}\}$ from E_{n+1} satisfying $d(f_{n+1}(\boldsymbol{a}), f_{n+1}(\boldsymbol{b})) = k$ for each $1 \leq k \leq n$.

Next, for any $\boldsymbol{a} \in A_n$, we have

$$\begin{aligned}
d(f_{n+1}(\boldsymbol{a}|0), f_{n+1}(\boldsymbol{a}|1)) &= d(f_n(\boldsymbol{a})|0, \overline{f_n(\boldsymbol{a})}|1) \\
&= d(f_n(\boldsymbol{a}), \overline{f_n(\boldsymbol{a})}) + 1 \\
&= n + 1
\end{aligned}$$

where the second equality is because we know that the two elements being handed to d have different rightmost coordinates, so their Hamming distance is one more than the Hamming distance between the elements obtained by removing the rightmost coordinates. The third equality is because $\overline{f_n(\boldsymbol{a})}$ and $f_n(\boldsymbol{a})$ differ in all n coordaintes by definition.

There are 2^n elements of A_n, and each pair $\{\boldsymbol{a}|0, \boldsymbol{a}|1\}$ is in E_{n+1}, so we get 2^n pairs from E_{n+1} such that $d(f_{n+1}(\boldsymbol{a}), f_{n+1}(\boldsymbol{b})) = n + 1$. For each k from 1 through $n + 1$, we have found 2^n pairs $\{\boldsymbol{a}, \boldsymbol{b}\}$ from E_{n+1} with the property that $d(f_{n+1}(\boldsymbol{a}), f_{n+1}(\boldsymbol{b})) = k$. This completes the proof.

Problem of the Month
Problem 8: May 2023

This month's problem is based on Problem 6 part (b) of the 2023 Euclid contest. Here is a modified and rephrased version of that problem:

A square is drawn in the plane with vertices at $(0,0), (1,0), (1,1),$ and $(0,1)$. Two blue lines are drawn with slope 3, one passing through $(0,0)$ and the other through $(\frac{1}{3}, 0)$. Two red lines are drawn with slope $-\frac{1}{3}$, one passing through $(0,1)$ and the other through $(0, \frac{2}{3})$. What is the area of the square bounded by the red and blue lines?
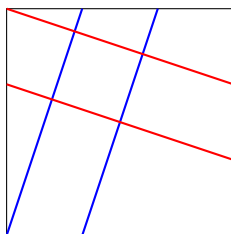
The answer to this question is $\frac{1}{10}$. We suggest convincing yourself of this before attempting the rest of the problem. The fact that this small square has an area exactly one tenth of the larger square suggests that there is a way to answer this question by showing that exactly 10 squares the size of the small one should fit into the large square. Let's explore!

Here is some terminology that we will use in this problem.

Definition: A *lattice point* is a point (x, y) for which x and y are both integers.

Definition: A 1×1 square whose vertices are at lattice points is called a *unit lattice square*. We will denote by T the unit lattice square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$.

Definition: $L_{q,p}$ denotes the line segment connecting $(0,0)$ to the point (q, p). Note that this line has slope $\frac{p}{q}$.

Definition: An $m-$lattice line is a line with slope m that passes through at least one lattice point.

The next two definitions are more complicated. There are examples given after they are stated.

Definition: The *tricky unit square*, $\mathbb{T}$, is a modified version of T (see above) with the property that if a line reaches an edge of $\mathbb{T}$, it "jumps" to the opposite side and continues with the same slope. For example, if a line reaches the top edge of $\mathbb{T}$, it continues with the same slope from the bottom edge directly below where it reached the top edge. If a line reaches a vertex of $\mathbb{T}$, then it has simultaneously reached two edges. In this situation, it continues with the same slope from the opposite vertex of $\mathbb{T}$.

Definition: Let $\frac{p}{q}$ be a rational number written in lowest terms. The $\frac{p}{q}-loop$ on $\mathbb{T}$ is the line passing through $(0,0)$ with slope $\frac{p}{q}$.

Below are diagrams, from left to right, of the $\frac{1}{2}$–loop, the $\frac{-2}{1}$–loop, and the $\frac{2}{5}$–loop on $\mathbb{T}$. In each diagram, equal letters mark places where the line jumps from one side of the square to the opposite side.
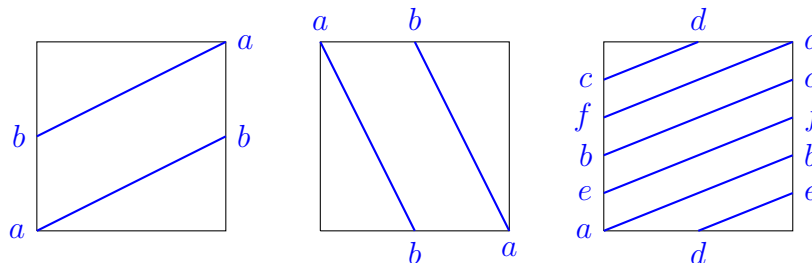
Each of the loops above eventually come back to their starting point and repeat. This happens because $\frac{p}{q}$ is rational (think about this!). Notice that in the image of the $\frac{-2}{1}$–loop (the middle image), the loop starts at $(0, 1)$ instead of $(0, 0)$. This is because a line of negative slope starting at $(0, 0)$ (on the bottom edge) immediately jumps to the top edge and continues from $(0, 1)$. It is worth thinking about how all four vertices of $\mathbb{T}$ really represent the same point.

Below is an image of $\mathbb{T}$ with the $\frac{3}{1}$–loop in blue and the $\frac{-1}{3}$–loop in red. These two loops divide $\mathbb{T}$ into 10 smaller squares. The squares numbered 1, 2, 3, 4, 5, and 6 are split in two pieces each across edges of $\mathbb{T}$. Notice that both the loops pass thorough the point $(0, 0)$ since the four vertices of the square are the same point. This means, if we count the four vertices as one intersection point, the two loops intersect exactly 10 times (count them!). Think about how this compares to the Euclid problem mentioned earlier.
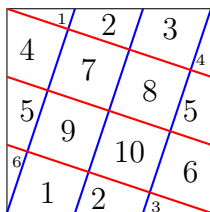
(a) For each pair of loops below, draw $\mathbb{T}$ with that pair of loops, count the number of times the loops intersect, and count the number of squares into which the loops divide $\mathbb{T}$.

   (i) The $\frac{4}{1}$–loop and the $\frac{-1}{4}$–loop.

   (ii) The $\frac{2}{3}$–loop and the $\frac{-3}{2}$–loop.

   (iii) The $\frac{4}{3}$–loop and the $\frac{-3}{4}$–loop.

   (iv) The $\frac{1}{1}$–loop and the $\frac{-1}{1}$–loop.

In the remaining problems, p and q are positive integers that have no positive divisors in common other than 1.

(b) How many line segments make up the $\frac{p}{q}$–loop?

(c) How many $\frac{p}{q}$–lattice lines intersect T?

(d) Explain why the number of points of intersection of the $\frac{p}{q}$–loop and the $-\frac{q}{p}$–loop on $\mathbb{T}$ is equal to the number of small squares into which the loops divide $\mathbb{T}$. Remember that the four vertices of $\mathbb{T}$ represent the same point.

(e) How many $-\frac{q}{p}$–lattice lines intersect $L_{q,p}$?

(f) Compute the area of the small squares in $\mathbb{T}$ created by the $\frac{p}{q}$–loop and the $-\frac{q}{p}$–loop. Our solution will take for granted that these small squares all have the same area, but you might like to think about how to prove this.

Problem of the Month
Problem 8: May 2023

Hint

(a) When counting, remember that some squares are split into pieces. As well, the four vertices of $\mathbb{T}$ count together as one intersection point.

(b) Relate the number of segments to the number of unit lattice squares through which $L_{q,p}$ passes.

(c) Such a line has an equation of the form $y = \dfrac{p}{q}x + b$. What can you say about $f(0)$ and $f(1)$ based on the fact that the line intersects T? do not forget about the condition that the line must pass through at least one lattice point!

(d) Each square has a leftmost vertex.

(e) Similar to the hint for part (c), such a line must have equation $y = -\dfrac{q}{p}x + b$. Try to deduce restrictions on the value of b from the fact that this line intersects $L_{q,p}$.

(f) Since the squares have the same size (though some might be broken into pieces) and $\mathbb{T}$ has area 1, the area of the squares can be computed by computing the number of them. Our solution will put together the ideas from (b) through (e) to count the squares. Specifically, the count in part (e) can be related to the number of squares. Remember to be careful with the four corners of $\mathbb{T}$, which count as one point.

Problem of the Month
Solution to Problem 8: May 2023

The following fact will come in handy later in the solution.

Fact 1: Let p and q be integers with $q \neq 0$ such that p and q have no positive divisors in common other than 1. If A and B are distinct lattice points such that the line segment joining them has slope $\frac{p}{q}$, then the distance between A and B is at least $\sqrt{p^2 + q^2}$.

Proof of Fact 1. Suppose A is the point (a, b) and B is the point $(a + s, b + t)$ where a, b, s, and t are integers so that the line segment joining A and B has slope $\frac{p}{q}$. Then the line segment joining A and B has slope $\frac{t}{s} = \frac{p}{q}$. By the assumption on p and q, the fraction $\frac{p}{q}$ is written in lowest terms, so we must have $|t| \geq |p|$ and $|s| \geq |q|$ implying $t^2 \geq p^2$ and $s^2 \geq q^2$. The distance between A and B is $\sqrt{t^2 + s^2} \geq \sqrt{p^2 + q^2}$, which completes the proof. $\square$

(a) (i) Below is a picture of the $\frac{4}{1}$–loop (in blue) and the $\frac{-1}{4}$–loop (in red) on $\mathbb{T}$.
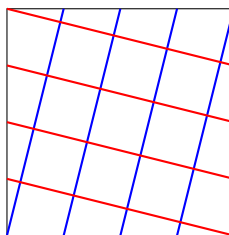
There are 17 intersection points. Remember that the two loops intersect at $(0, 0)$, which is the same point as the other three vertices of $\mathbb{T}$.
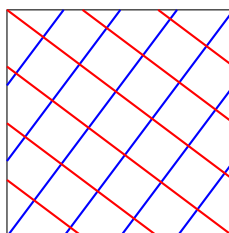
There are 17 small squares. Of these 17, 9 are whole squares in the middle of $\mathbb{T}$, 4 are split into two pieces over the horizontal edges of $\mathbb{T}$, and 4 are split over the vertical edges of $\mathbb{T}$.

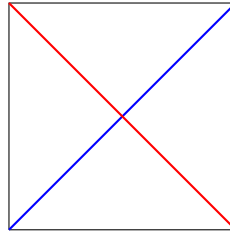(ii) Here is a picture of a $\frac{2}{3}$–loop (in blue) and a $\frac{-3}{2}$–loop (in red) on $\mathbb{T}$.
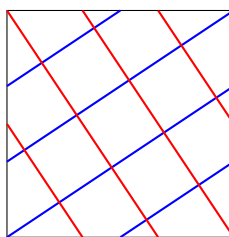
There are 13 intersection points and 13 small squares.

(iii) Here is a picture of a $\frac{4}{3}$–loop (in blue) and a $\frac{-3}{4}$–loop (in red) on $\mathbb{T}$.

There are 25 intersection points and 25 small squares.

(iv) Here is a picture of a $\frac{1}{1}$–loop (in blue) and a $\frac{-1}{1}$–loop (in red) on $\mathbb{T}$.

There are 2 intersection points and 2 small squares. Here the squares are harder to see, since both of them pass through the edges of $\mathbb{T}$, passing over to the other side.

It is not a coincidence that the number of small squares is equal to the number of intersection points in all four cases.

(b) Consider the line of slope $\frac{p}{q}$ through the origin. By Fact 1, (q, p) is the lattice point in the first quadrant that is closest to the origin. Therefore, there are no lattice points on $L_{q,p}$ other than its endpoints. To illustrate the idea in this solution, the diagram below shows $L_{5,11}$ along with every unit lattice square through which it passes.
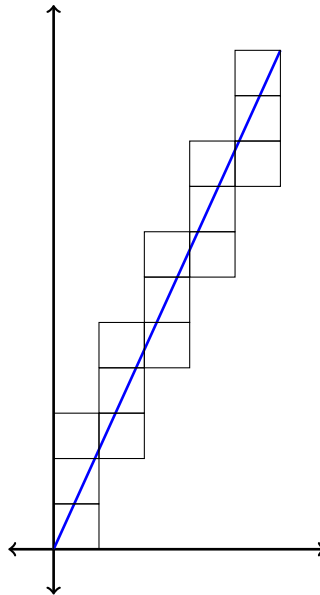
The $\frac{11}{5}$–loop continues to wrap around $\mathbb{T}$ until it reaches a corner again. Because of how $\mathbb{T}$ behaves, we can think of the $\frac{11}{5}$–loop as the result of overlaying all of the unit lattice squares in the diagram above on top of each other. Indeed, if the blue line reaches the top of the square, it jumps to the bottom with the same x-coordinate and continues with the same slope. This means that after it jumps, what the line does in $\mathbb{T}$ is the same as what it does as it continues into the adjacent unit lattice square.

This is true in general. So the number of line segments in the $\frac{p}{q}$–loop is equal to the number of unit lattice squares that $L_{q,p}$ intersects. By "intersects" we mean that the line passes through some part of the *interior* of the square and not just a vertex.

Therefore, to count the line segments in the $\frac{p}{q}$–loop, we can count the unit lattice squares through which $L_{q,p}$ passes. Line segment $L_{q,p}$ begins in T and passes into a new unit lattice

2

square exactly when it crosses either a vertical line with equation $x = a$ for some integer a or a horizontal line with equation $y = b$ for some integer b. Since $L_{q,p}$ passes through no lattice points other than its endpoints, it will never cross such a vertical and a horizontal line at the same time. Since $L_{q,p}$ starts at $(0,0)$ and ends at (q,p), it must cross $q - 1$ such vertical lines and $p - 1$ such horizontal lines. Adding one to account for the unit square from which it originates, this means there are $(p-1) + (q-1) + 1 = p + q - 1$ line segments in the $\frac{p}{q}$–loop. For the $\frac{11}{5}$–loop, this gives $11 + 5 - 1 = 15$ line segments. By counting, you can verify that there are indeed 15 unit lattice squares in the diagram above. If you carefully draw the $\frac{11}{5}$–loop, you will see that there are 15 segments.

(c) Since p and q are positive, $\frac{p}{q}$ is positive. Suppose $f(x) = \frac{p}{q}x + b$ is a $\frac{p}{q}$–lattice line that intersects T. Then we must have $b \leq 1$, otherwise T would be entirely below the graph of $f(x)$ (remember, its slope is positive). Additionally, we must have that $f(1) \geq 0$, otherwise T would be completely above the graph of $f(x)$.

The inequality $f(1) \geq 0$ is equivalent to $\frac{p}{q} + b \geq 0$, or $-\frac{p}{q} \leq b$. Combining with $b \leq 1$, we get $-\frac{p}{q} \leq b \leq 1$, which we will write as $-\frac{p}{q} \leq b \leq \frac{q}{q}$.

So far, we have not used the fact that $f(x)$ contains at least one lattice point. Suppose $f(x)$ passes through some lattice point (u, v). That is, u and v are integers and $f(u) = v$. Then $v = \frac{up}{q} + b$ which can be rearranged to get $b = \frac{vq - up}{q}$. This implies that there is some integer a for which $b = \frac{a}{q}$. Substituting into $-\frac{p}{q} \leq b \leq \frac{q}{q}$, we get

$$-\frac{p}{q} \leq \frac{a}{q} \leq \frac{q}{q}$$

which is equivalent to $-p \leq a \leq q$. There are exactly $p + q + 1$ integers a that satisfy these inequalities, so there are $p + q + 1$ $\frac{p}{q}$–lattice lines that intersect through T.

Before moving on, we make the observation that two of the $\frac{p}{q}$–lattice lines that intersect T only pass through the vertices $(0,1)$ and $(1,0)$. This means there are actually $p + q - 1$ $\frac{p}{q}$–lattice lines that pass through the interior of T.

(d) Since $p \neq 0$ and $q \neq 0$, the loops are neither horizontal nor vertical. Therefore each of the small squares has a unique left-most vertex, which is the vertex with the smallest x-coordinate. Each of the small squares has exactly one left-most vertex, and every point of intersection is the left-most vertex of exactly one square. This means that the number of intersection points is equal to the number of squares.

Note that this lends some credibility to counting the four vertices as one intersection point. See, for example, the $\frac{3}{1}$ and $\frac{-1}{3}$ loops from the problem statement. In that example, the square labelled by 1 is split into two pieces, and its leftmost vertex is represented by all four vertices of T. It might be useful to think about this for a moment before moving on.

(e) Since p and q are positive, $-\frac{q}{p}$ is negative. Suppose $f(x) = -\frac{q}{p}x + b$ is a $-\frac{q}{p}$–lattice line that intersects $L_{q,p}$. Since the slope of $f(x)$ is negative, we must have $b \geq 0$. Otherwise, $L_{q,p}$ would be entirely above the graph of $f(x)$.

For similar reasoning, we also require that $f(q) \leq p$, which means $-\frac{q^2}{p} + b \leq p$. This can be rearranged to $b \leq \frac{p^2 + q^2}{p}$. Combining with $b \geq 0$ gives $0 \leq b \leq \frac{p^2 + q^2}{p}$. By essentially the

3

same argument that was used in the solution to part (c), b must take the form $\frac{a}{p}$ for some integer a. Therefore, we have

$$\frac{0}{p} \leq \frac{a}{p} \leq \frac{p^2 + q^2}{q}$$

and there are exactly $p^2 + q^2 + 1$ integers a with this property. These lines are all different and all intersect $L_{q,p}$, so the answer is $p^2 + q^2 + 1$.

Before moving on, we note that the lines of slope $-\frac{q}{p}$ that pass through $(0,0)$ and (q,p) are $-\frac{q}{p}$–lattice lines that intersect $L_{q,p}$. Therefore, there are $p^2 + q^2 + 1 - 2 = p^2 + q^2 - 1$ points of intersection of $-\frac{q}{p}$–lattice lines with $L_{q,p}$, excluding the endpoints. We will refer to this set of $p^2 + q^2 - 1$ points several times in the solution to part (f), so we will denote it by X for convenience.

(f) Let Y be the set of intersection points of the $\frac{p}{q}$–loop with the $\frac{-q}{p}$–loop, other than the point represented by the four courners of $\mathbb{T}$. We will show that every point in X corresponds to exactly one point in Y, thereby showing that X and Y have the same number of elements. Since there are $p^2 + q^2 - 1$ points in X, if we include the point in $\mathbb{T}$ represented by the corners, this will show that the loops have exactly $p^2 + q^2$ intersection points in $\mathbb{T}$. By part (d), this will show that the loops divide $\mathbb{T}$ into $p^2 + q^2$ small squares.

The rest of the solution is devoted to showing that X and Y have the same number of elements. We will use the following fact several times. Its proof is not included.

Fact 2: If an m–lattice line is translated a units to the left and b units to the right for some integers a and b, then the line obtained is also an m–lattice line.

We first observe that every line segment in the $\frac{p}{q}$–loop is part of some $\frac{p}{q}$–lattice line. To see why this is true, consider the solution to part (c). It was discussed there that the line segments of the $\frac{p}{q}$–loop can be obtained by overlaying on T the unit lattice squares through which $L_{q,p}$ passes. For each such unit lattice square, there are integers a and b so that the square can be translated a units to the left and b units down in order to land on T. Therefore, if we translate $L_{q,p}$ a units to the left and b units down, it will land exactly on the segment of the $\frac{p}{q}$–loop in question. $L_{q,p}$ is part of a $\frac{p}{q}$–lattice line, so by Fact 2 above, we have shown that every line segment in the $\frac{p}{q}$–loop is part of some $\frac{p}{q}$–lattice line.

By part (b), there are $p + q - 1$ segments in the $\frac{p}{q}$–loop. By the remark at the end of the solution to part (c), there are $p + q - 1$ $\frac{p}{q}$–lattice lines that pass through the interior of T. Each of the segments in the $\frac{p}{q}$–loop is in the interior of the square. We conclude that the segments of the $\frac{p}{q}$–loop are exactly the parts of $\frac{p}{q}$–lattice lines that pass through the interior of T.

By very similar reasoning, we can also conclude that the $\frac{-q}{p}$–loop is made precisely of the segments of $-\frac{q}{p}$–lattice lines that pass through the interior of T.

Consider a point (u, v) in X and suppose it is inside the unit lattice square with vertices (a, b), $(a+1, b)$, $(a+1, b+1)$, and $(a, b+1)$. Let $f(x)$ be the $-\frac{q}{p}$–lattice line that intersects $L_{q,p}$ at (u, v). The point $(u - a, v - b)$ is in T. If we translate $f(x)$ and $L_{q,p}$ to the left by a units and down by b units, the resulting lines will intersect at $(u - a, v - b)$. By Fact 2, translating $f(x)$ in this way results in a $-\frac{q}{p}$–lattice line and translating $L_{q,p}$ results in a $\frac{p}{q}$–lattice line. By the previous argument, these will be segments of the $\frac{-q}{p}$ and $\frac{p}{q}$ loops,

4

respectively. Therefore, $(u - a, v - b)$ is a point of intersection of the two loops. This gives a natural way to translate points of X to points in Y. In symbols, we send (u, v) to $(u - \lfloor u \rfloor, v - \lfloor v \rfloor)$.

Suppose that (u, v) and (x, y) are two different points in X. If these points are in different unit lattice squares, then when they are translated to T, they will end up on different line segments of the $\frac{p}{q}$–loop, so they cannot possibly be translated to the same point in Y. If they are in the same unit lattice square, then they will be translated to the left and down by exactly the same amount, so they will end up in different places in T. Either way, we can conclude that different points in X are sent to different points in Y by the rule described in the previous paragraph.

Suppose (u, v) is a point in Y. This point is on some line segment of the $\frac{p}{q}$–loop, and we have argued earlier that these segments come from translating unit lattice squares through which $L_{q,p}$ passes (and taking the parts of $L_{q,p}$ along for the ride). Suppose the segment on which (u, v) lies comes from the unit lattice square with vertices (a, b), $(a+1, b)$, $(a+1, b+1)$, and $(a, b+1)$ and call this square S. The point $(a + u, b + v)$ is in S and lies on $L_{q,p}$.

We showed earlier that every line segment on the $\frac{-q}{p}$–loop is part of some $-\frac{q}{p}$–lattice line. Let $f(x)$ be the $-\frac{q}{p}$–lattice line on which (u, v) lies. If we translate $f(x)$ to the right by a units and up by b units, the result will be a new $-\frac{q}{p}$–lattice line by Fact 2. Moreover, this line will pass through $(u + a, v + b)$, which also lies on $L_{q,p}$. Therefore, $(u + a, v + b)$ is a point in X that is inside the square S. If we translate $(u + a, v + b)$ to the left by a and down by b, it will be sent to (u, v) by construction.

We have shown that each of the $p^2 + q^2 - 1$ points in X corresponds to exactly one point in Y. As explained earlier, if we consider the four corners of $\mathbb{T}$ as one intersection point, then we get a total of $p^2 + q^2$ intersection points of the two loops.

As mentioned in the problem, we are taking for granted that the squares all have the same size. This means that each of them has an area of exactly $\frac{1}{p^2+q^2}$. Referring back to the Euclid problem from the beginning, we were dealing with the situation where $p = 3$ and $q = 1$, so the area of the squares is $\frac{1}{3^2+1^2} = \frac{1}{10}$.