FiveThirtyEight's The Riddler
April 28, 2017: A painting puzzle
Solution by Tim Black

The puzzle:

> *From Dan Waterbury, a painting puzzle:*
>
> *You play a game with four balls: One ball is red, one is blue, one is green and one is yellow. They are placed in a box. You draw a ball out of the box at random and note its color. Without replacing the first ball, you draw a second ball and then paint it to match the color of the first. Replace both balls, and repeat the process. The game ends when all four balls have become the same color. What is the expected number of turns to finish the game?*
>
> *Extra credit: What if there are more balls and more colors?*

Answer: The expected number of turns is 9. Specifically, for $k = 1, 2, 3$, the expected number of turns on which exactly k balls match the final color is 3.

Extra credit answer: If there are n balls and they all start out a different color, then the expected number of turns is $(n-1)^2$. Specifically, for $k = 1, 2, \ldots, n-1$, the expected number of turns on which k balls match the final color is $n-1$.

Solution: Let's say that a color "wins" if eventually all n balls are that color.

Trying to keep track of n different colors seems like a big task, but we'll actually only have to keep track of a single color. Since the n colors are interchangeable, and exactly one color wins, we can just calculate the number of turns until all the balls are red, conditioned on the color red winning. So, let's just think about red for a bit.

Lemma 1. *Let $0 \le k \le n$. Given that k balls are currently red, the probability that red wins is k/n.*

Proof. For $0 \le k \le n$, let r_k be the probability that red wins given that k balls are currently red. We know that $r_0 = 0$ and $r_n = 1$.

Suppose that k balls are currently red, for some $1 \le k \le n-1$. On your next turn, if you draw a red ball then a non-red ball, the number of red balls increases by one. If you draw a non-red ball then a red ball, the number of red balls decreases by one. Otherwise, the number of red balls stays the same. So, it is equally likely that the number of red balls increases or decreases by one. This tells us that

$$r_k = \frac{1}{2} r_{k-1} + \frac{1}{2} r_{k+1},$$

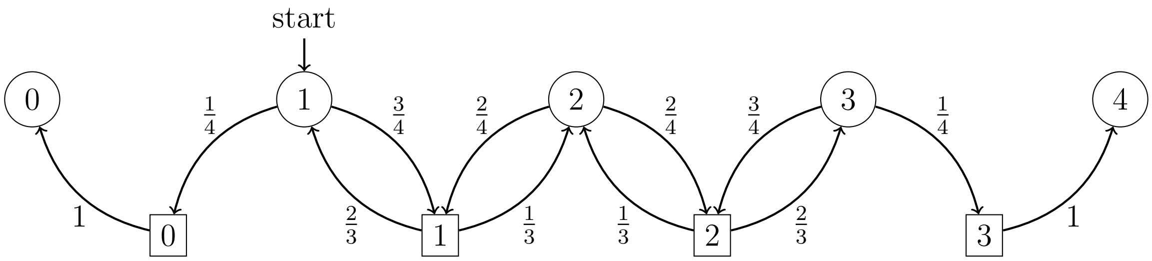
or equivalently, that

$$r_k - r_{k-1} = r_{k+1} - r_k.$$

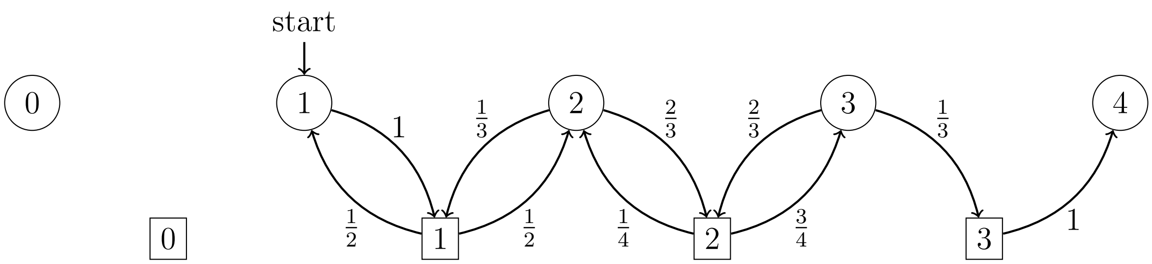
This means that $r_0, r_1, \ldots, r_n$ are evenly spaced; $r_k = \frac{k}{n}$. □

From now on, it will be useful for us to think of each turn happening in two stages. In stage one, a ball is drawn at random from the box and discarded (so there are $n-1$ balls left in the box). In stage two, a ball is drawn from the box, and that ball is returned to the box along with a duplicate ball of the same color (so there are again n balls in the box). This process has the same result as the process described in the puzzle.

Here are the states of the system, and transition probabilities between them. For ease of presentation, the diagram only shows the case of $n = 4$ colors. A circled k represents the state where k out of n balls are red, while a boxed k represents the state where k out of $n-1$ balls are red (because a ball has been removed in stage 1 of the turn).



From Lemma 1, we know that the probability that red wins given that k out of n balls are red is k/n. It can similarly be shown that the probability that red wins given that k out of $n-1$ balls are red is $k/(n-1)$. Using this fact, the transition probabilities above, and Bayes' rule, we can calculate the transition probabilities condition on red winning. Again for ease of presentation, the diagram shows the case of $n = 4$ colors.



For $1 \leq k \leq n$, let c_k be the expected number of turns on which k out of n balls are red, conditioned on red winning. For $1 \leq k \leq n-1$, let d_k be the number of turns on which k out of $n-1$ balls are red, conditioned on red winning. We have that $c_n = 1$, because red wins, and the game ends. In the case of $n=4$, from the previous diagram we get the equations

$$\begin{aligned}
c_4 &= d_3, && \text{so, } d_3 = c_4 = 1;\\
d_3 &= \frac{1}{3}c_3, && \text{so, } c_3 = 3d_3 = 3;\\
c_3 &= \frac{3}{4}d_2, && \text{so, } d_2 = \frac{4}{3}c_3 = 4;\\
d_2 &= \frac{2}{3}c_3 + \frac{2}{3}c_2, && \text{so, } c_2 = \frac{3}{2}\left(d_2 - \frac{2}{3}c_3\right) = 3;\\
c_2 &= \frac{1}{4}d_2 + \frac{1}{2}d_1, && \text{so, } d_1 = 2\left(c_2 - \frac{1}{4}d_2\right) = 4;\\
d_1 &= \frac{1}{3}c_2 + c_1, && \text{so, } c_1 = d_1 - \frac{1}{3}c_2 = 3;\\
c_1 &= 1 + \frac{1}{2}d_1.
\end{aligned}$$

In particular, $c_1 = c_2 = c_3 = 3$. So, when $n=4$, for $k = 1,2,3$, the expected number of turns on which exactly k balls match the final color is 3. Summing these, the expected number of turns is 9.

For general n, we have the equations

$$\begin{aligned}
c_n &= d_{n-1} = 1,\\
d_{n-1} &= \frac{1}{n}c_{n-1},\\
c_k &= \frac{n-k-1}{n}d_k + \frac{k}{n}d_{k-1} && \text{for } 2 \le k \le n-1,\\
d_k &= \frac{k}{n-1}c_{k+1} + \frac{n-k}{n-1}c_k && \text{for } 1 \le k \le n-2,\\
c_1 &= 1 + \frac{n-2}{n}d_1.
\end{aligned}$$

The unique solution to this system of equations is $c_1 = c_2 = \cdots = c_{n-1} = n-1$, and $d_1 = d_2 = \cdots = d_{n-1} = n$, and $c_n = d_{n-1} = 1$. So,for $1 \le k \le n-1$, the expected number of turns on which exactly k balls match the final color is $n-1$. Summing these, the expected number of turns is $(n-1)^2$.