Classic Riddler

20 December 2019

Riddle:

I have 10 pairs of socks in a drawer. Each pair is distinct from another and consists of two matching socks. Alas, I'm negligent when it comes to folding my laundry, and so the socks are not folded into pairs. This morning, fumbling around in the dark, I pull the socks out of the drawer, randomly and one at a time, until I have a matching pair of socks among the ones I've removed from the drawer.

On average, how many socks will I pull out of the drawer in order to get my first matching pair?

(Note: This is different from asking how many socks I must pull out of the drawer to *guarantee* that I have a matching pair. The answer to *that* question, by the pigeonhole principle, is 11 socks. This question is instead asking about the *average*.)

Extra credit: Instead of 10 pairs of socks, what if I have some large number N pairs of socks?

Solution:

Of course, the answer to the question of how many socks must I grab to guarantee a match is always $N + 1$, or 11 in this case. But that is too easy, and not worthy of a Classic Riddler.

For this problem, I have to figure out the probability of drawing a matching sock exactly on the nth draw. I will call this probability $P(n)$. It is trivial that $P(1) = 0$, since a single sock cannot be a pair. For other n, I must draw a *new* sock type exactly $n - 1$ times, and only on the nth draw match any existing sock already drawn. For $n = 2$ the probability $P_{match}(2)$ of drawing a matching sock is $1/19$ because there are 19 socks left, of which only one will be a match. For $n = 3$, the probability $P_{new}(3)$ of drawing a new sock on the second draw is $18/19$, and the probability $P_{match}(3)$ of drawing a matching sock on the third draw is $2/18$.

Continuing this logic leads to the following formulae (for $n > 1$):

$$P_{new}(n) = \prod_{k=1}^{n-1} \frac{20 - 2*(k-1)}{20 - (k-1)}$$

$$P_{match}(n) = \frac{n-1}{20 - (n-1)}$$

To be more general, I note that $P_{new}(1) = 1$ and $P_{match}(1) = 0$. Now I simply have the total probability $P(n)$:

$$P(n) = P_{new}(n) * P_{match}(n)$$

The average number of draws for 10 pairs of socks $D(10)$ is given by:

$$D(10) = \sum_{n=1}^{11} n * P(n)$$

The sum stops at 11, because (as noted above), I am guaranteed a match after 11 draws, so no more draws are needed. Evaluating the sum gives $\boxed{D(10) = 5.6755}$.

For the general case of N pairs of socks, I substitute into the above equation:

$$D(N) = \sum_{n=1}^{N+1} n * \left(\prod_{k=1}^{n-1} \frac{2N - 2*(k-1)}{2N - (k-1)} \right) \frac{n-1}{2N - (n-1)}$$

This is not easy to deal with. Wolfram Alpha starts to time out around $N = 50$, and I'm not immediately seeing where it converges. For $N = 40$, Wolfram Alpha calculates $D(40) \approx 11.25$, which is less than my guesses of $N/2$ and N/e convergences. So who knows.