Special Topics

1 Streaks

Was the table of H's and T's below generated by flipping a fair coin 100 times, or by someone tapping the H and T keys in a what felt like a random way?

$$\begin{array}{c} HTTTHTHTTHTTHTHTHTHT \\ TTTHHTHHTHTHTTHHHTHT \\ HHTHHTTTHHHTHTTHHHHT \\ THTTHHTHTHTHTHTHTTHTHH \\ HTTHHHTHTHHHTHTHHHTH \end{array}$$

There is no way to be sure. However, this sequence has a distinctive feature that is common in "random" human-generated sequences and unusual in truly random sequences: namely, there is no long streak of H's or T's. In fact, no symbol appears above more than four times in a row. How likely is that? If we flip a fair coin 100 times, what is the probability that we never get five heads in a row?

1.1 From a Probability Problem to a Counting Problem

The sample space for this experiment is $\{H, T\}^{100}$; that is, the set of all length-100 sequences of H's and T's. If the coin tosses are fair and independent, then all 2^{100} such sequences are equally likely. Therefore, we need only count the number of sequences with no streak of five heads; given that, the probability that a random length-100 sequence contains no such streak is:

$$\Pr\left(\text{sequence has no } HHHHH\right) = \frac{\text{\# sequences with no } HHHHH}{2^{100}}$$

This is a common situation. We have reduced a probability problem to a counting problem. Unfortunatley, we have no hope of solving the counting problem by direct computation. No computer can consider all 2^{100} sequences of H's and T's, keeping track of how many lack a streak of five heads. But, on the bright side, there is a big bag of mathematical tricks for solving counting problems. In this case, we'll use a *recurrence equation*. The recurrence equation approach involves two steps:

1. Solve some small problems.
2. Solve the n-th problem using preceding solutions.

Let's see how this approach plays out in the analysis of streaks.

1.2 Step 1: Solve Small Instances

Let S_n be the set of length-n sequences of H's and T's that do not contain a streak of five heads. Our eventual goal is to compute $|S_{100}|$. But for now, let's just compute $|S_n|$ for some very small values of n:

$$\begin{aligned} |S_1| &= 2 \quad (H \text{ and } T) \\ |S_2| &= 4 \quad (HH,\ HT,\ TH, \text{ and } TT) \\ |S_3| &= 8 \\ |S_4| &= 16 \\ |S_5| &= 31 \quad (HHHHH \text{ is excluded!}) \end{aligned}$$

These are called *base cases*.

1.3 Step 2: Solve the n-th Problem Using Preceding Solutions

We can classify the sequences in S_n into five groups:

1. Sequences that end with a T.
2. Sequences that end with TH.
3. Sequences that end with THH.
4. Sequences that end with $THHH$.
5. Sequences that end with $THHHH$.

Every sequence in S_n falls into exactly one of these groups. Thus, the size of S_n is the sum of the sizes of these five groups.

How many sequences are there in the first group? That is, how many sequences in S_n end with T? The preceding $n-1$ symbols in such a sequence can not contain a streak of five heads. Therefore, those $n-1$ symbols form a sequence in S_{n-1}. On the other hand, putting a T at then end of any sequence in S_{n-1} gives a sequence in S_n. Therefore, the number of sequences in the first group is exactly equal to $|S_{n-1}|$.

How many sequences are there in the second group? Arguing as before, the preceding $n-2$ symbols in each such sequence must form a sequence in S_{n-2}. On the other hand, appending TH to any sequence in S_{n-2} gives a sequence in the second group. Therefore, there are exactly $|S_{n-2}|$ sequences in the second group.

By similar reasoning, the number of sequences in the third group is $|S_{n-3}|$, the number in the fourth is $|S_{n-4}|$, and the number in the fifth is $|S_{n-5}|$. Therefore, we have:

$$|S_n| \quad = \quad |S_{n-1}| + |S_{n-2}| + |S_{n-3}| + |S_{n-4}| + |S_{n-5}| \quad (\text{for } n > 5)$$

This *recurrence equation* expresses the solution to a large problem ($|S_n|$) in terms of the solutions to smaller problems (S_{n-1}, S_{n-2}, ...).

By combining the base cases and the recurrence equation, we can compute $|S_6|$, $|S_7|$, $|S_8|$, and so forth until we reach $|S_{100}|$.

$$\begin{aligned}
|S_6| &= |S_5| + |S_4| + |S_3| + |S_2| + |S_1| \\
&= 31 + 16 + 8 + 4 + 2 + 1 \\
&= 62 \\
|S_7| &= |S_6| + |S_5| + |S_4| + |S_3| + |S_2| \\
&= 62 + 31 + 16 + 8 + 4 + 2 \\
&= 123 \\
|S_8| &= \ldots
\end{aligned}$$

Computing $|S_{100}|$ still requires about 500 additions, so a computer helps. Plugging the value of $|S_{100}|$ into our earlier probability formula, we find:

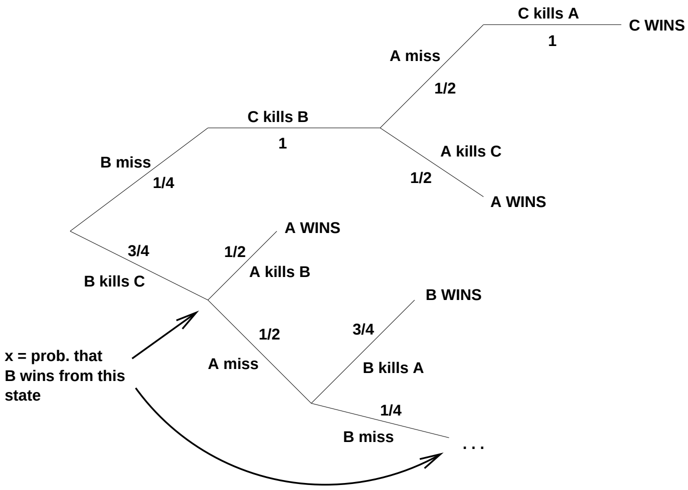
$$\Pr(\text{sequence has no } HHHHH) \quad = \quad 0.193\ldots$$

Thus, four out of five sequences of 100 coin tosses contain a streak of five heads. By symmetry, we also know that four out of five sequences contain a streak of five tails. If we suppose that these two events are nearly independent, then only about one random sequence in twenty-five contains no streak of five heads or five tails. This is the situation for the sequence given at the start of this section. Sure enough, I made up that sequence up by “randomly” tapping keys!

2 The Truel

Three gunfighters meet for a *truel*, a three-person duel. Gunfighter A hits his target 50% of the time, gunfighter B hits 75% of the time, and gunfighter C hits 100% of the time. The gunfighters take turns shooting in the order A, B, C, A, B, C, etc. Of course, a dead gunfighter misses his turn. The last one standing is the winner.

What is A's best strategy? If A kills C, then B will probably kill A on the next shot. On the other hand, if A kills B, then C will certainly kill A on the next shot. This does not look good. But there is a another possibility: A could intentionally miss, let B and C shoot it out, and then try to kill the winner! Let's evaluate that strategy, assuming that B and C actually try to hit each other.



From the tree diagram, we have:

$$\begin{aligned}
\Pr(\text{C wins}) &= \frac{1}{4} \cdot 1 \cdot \frac{1}{2} \cdot 1 \\
&= \frac{1}{8} \\
&= 12.5\%
\end{aligned}$$

Now let x be the probability that B eventually wins in the situation where C is dead and A has the next shot. This situation arises at two different points in our tree diagram. We can exploit that fact to obtain an equation expressing x in terms of itself:

$$x = \frac{1}{2} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{4} \cdot x$$

Solving this equation, we find that $x = 3/7$. The probability that B wins overall is:

$$\begin{aligned}\Pr(\text{B wins}) &= \frac{3}{4}\cdot x \\ &= \frac{9}{28} \\ &\approx 32.1\%\end{aligned}$$

Finally, we have:

$$\begin{aligned}\Pr(\text{A wins}) &= 1 - \Pr(\text{B wins}) - \Pr(\text{C wins}) \\ &\approx 55.4\%\end{aligned}$$

Amazingly, the worst shooter has the best chance of winning, and the best shooter has the worst chance of winning!

Of course, an explicit assumption in this analysis was that B and C are both shooting to kill, unlike A in the first round. If B and C have no such requirement, then the problem is underspecified; there is no definite mathematical solution. Every gunfighter might reason that he is better off not shooting and the whole lot might go toast smores over a campfire.

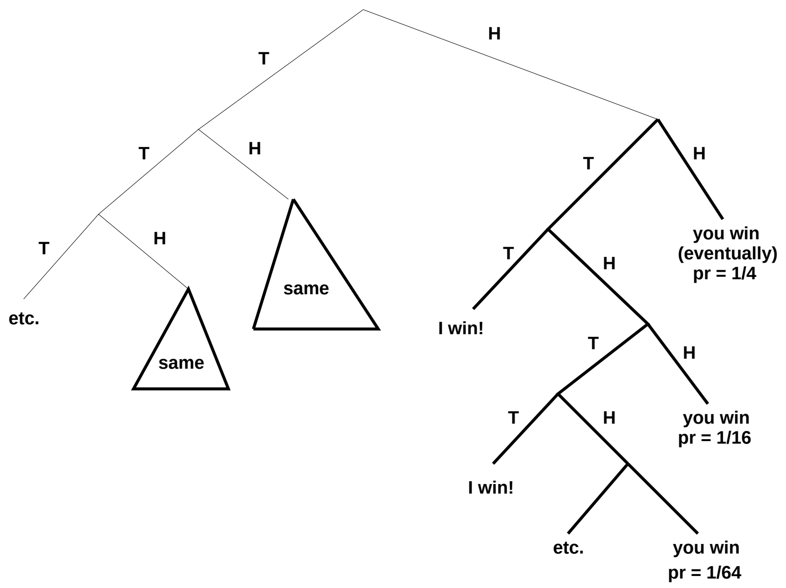
3 Penney-Ante

Let's play a game! We repeatedly flip a fair coin. You have the sequence HHT, and I have the sequence HTT. If your sequence comes up first, then you win. If my sequence comes up first, then I win. For example, if the sequence of tosses is:

$$TTHTHT\underline{HHT}$$

then you win. This problem is tricky, because the game could go on for an arbitrarily long time. Draw enough of the tree diagram to see a pattern, and then sum up the probabilities of the (infinitely many) outcomes in which you win.

A partial tree diagram is shown below. All edge probabilities are $1/2$.



Let's first focus on the subtree shown in bold. Note that if two heads are flipped in a row, then you are guaranteed to win eventually. The sum of the probabilities of all your winning outcomes in this subtree is:

$$\begin{aligned} \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \ldots &= \frac{1}{4} \cdot \frac{1}{1 - 1/4} \\ &= \frac{1}{3} \end{aligned}$$

The upppermost subtree marked **same** is the identical to the one shown in bold, except that each outcome probability is reduced by $1/2$, because it is one edge farther from the root. Thus, the sum of your winning outcomes in this subtree is $1/6$. Similarly, the sum of your winning outcomes in the next subtree marked **same** is $1/12$, and so forth. Overall, your probability of winning is:

$$\begin{aligned} \frac{1}{3} + \frac{1}{6} + \frac{1}{12} + \ldots &= \frac{1}{3} \cdot \frac{1}{1 - 1/2} \\ &= \frac{2}{3} \end{aligned}$$

In fact, as long as you pick a sequence first, and I pick a sequence second, I can always have at least a 2/3 probability of winning the game.

How is this possible? We know that each sequence of length three is equally likely, so therefore there should be no advantage to one sequence over another, and certainly no disadvantage to picking a sequence first. However, as we've already seen, this does not appear to be true.

One way to observe that I can always win with probability of at least 2/3 is if you first select the sequence HHH. I then select the sequence THH. I claim that in this situation, my probability of winning is 7/8.

Why is this true? In the case where the first three tosses are HHH, with probability 1/8, you win immediately. The remaining 7/8 of the time, the game continues.

Suppose at some point later on, say starting at toss 30, the sequence HHH finally appears for the first time. This means that nowhere earlier than toss 30 do we observe HHH. So, toss 29 must have been a T. However, this means that the sequence THH appeared before the sequence HHH, and I won! Using this logic, no matter when in the series of flips HHH appeared, THH must have appeared first, except when HHH is the first three coin flips.

So, 7/8 of the time THH occurs before HHH, and I win! Pretty good odds.

Similar logic can be used for each sequence of length three selected by my opponent. The key is that I must pick my sequence after my opponent selects his sequence in order to guarantee that I have the advantage. Table 1 contains the relevant information to ensure you continued success.

Table 1: In Penny-ante, your best response to your opponent's choice, and the corresponding chance that your pattern appears first

Opponent's Choice	Your Choice	Chance of Winning
HHH	THH	7/8
HHT	THH	3/4
HTH	HHT	2/3
THH	TTH	2/3
HTT	HHT	2/3
THT	TTH	2/3
TTH	HTT	3/4
TTT	HTT	7/8