First triplet occurrence in coin toss

Consider tossing a fair coin (equal probability for head, H, and tail, T) for infinite times. Which sequence is more likely to appear first, HTT or HTH?

This is a question that got asked in a bioinformatics class. And you guessed it: I got it wrong the first time. I thought that they would have the same probability of appearing in a coin toss experiment. Let's see what turns out to be the right answer.

Let N_1 be the number of tosses needed to first observe HTH. H_i (or T_i) represents that the i-th toss is head (or tail).

Think about the outcome of the first toss.

$$E(N_1) = \frac{1}{2}E(N_1|\mathrm{H}_1) + \frac{1}{2}E(N_1|\mathrm{T}_1),$$
$$E(N_1|\mathrm{T}_1) = E(N_1) + 1.$$

For the second toss followed by H_1,

$$E(N_1|\mathrm{H}_1) = \frac{1}{2}E(N_1|\mathrm{H}_1\mathrm{T}_2) + \frac{1}{2}E(N_1|\mathrm{H}_1\mathrm{H}_2),$$
$$E(N_1|\mathrm{H}_1\mathrm{H}_2) = E(N_1|\mathrm{H}_1) + 1.$$

Continuing with $\mathrm{H}_1\mathrm{T}_2$,

$$E(N_1|\mathrm{H}_1\mathrm{T}_2) = \frac{1}{2}E(N_1|\mathrm{H}_1\mathrm{T}_2\mathrm{H}_3) + \frac{1}{2}E(N_1|\mathrm{H}_1\mathrm{T}_2\mathrm{H}_3),$$
$$E(N_1|\mathrm{H}_1\mathrm{T}_2\mathrm{H}_3) = 3, E(N_1|\mathrm{H}_1\mathrm{T}_2\mathrm{T}_3) = E(N_1) + 3.$$

By back substituting the equations, we get $E(N_1) = 10$.

Similarly, for the number of tosses needed to first observe HTT, N_2, the analysis on the first two tosses is exactly the same:

$$E(N_2) = \frac{1}{2}E(N_2|\mathrm{H}_1) + \frac{1}{2}E(N_2|\mathrm{T}_1),$$
$$E(N_2|\mathrm{T}_1) = E(N_2) + 1,$$
$$E(N_2|H_1) = \frac{1}{2}E(N_2|\mathrm{H}_1\mathrm{T}_2) + \frac{1}{2}E(N_2|\mathrm{H}_1\mathrm{H}_2),$$
$$E(N_2|\mathrm{H}_1\mathrm{H}_2) = E(N_2|\mathrm{H}_1) + 1.$$

Now, continuing with $\mathrm{H}_1\mathrm{T}_2$,

$$E(N_2|\mathrm{H}_1\mathrm{T}_2) = \frac{1}{2}E(N_2|\mathrm{H}_1\mathrm{T}_2\mathrm{H}_3) + \frac{1}{2}E(N_2|\mathrm{H}_1\mathrm{T}_2\mathrm{T}_3),$$
$$E(N_2|\mathrm{H}_1\mathrm{T}_2\mathrm{H}_3) = E(N_2|\mathrm{H}_1) + 2, E(N_2|\mathrm{H}_1\mathrm{T}_2\mathrm{T}_3) = 3.$$

By back substituting the equations, we get $E(N_2) = 8$.

HHH and TTT have the longest waiting time, 14 tosses in both cases. HTH and THT have a waiting time of 10, and all other triplets have a waiting time of 8.