Patterns in coin tosses

Assume that you repeatedly toss a coin, with heads outcome represented by 1 and tails represented by 0. On any toss 1 occurs with probability p. Assume also that you have a *pattern* of outcomes, say 1011101. What is the expected number of tosses needed to obtain this pattern? It should be about $2^7 = 128$ when $p = 1/2$, but what is it exactly? One could compare two patterns by this *waiting game*, saying that the one with smaller expected value wins.

Another way to compare two patterns is the *horse race*: you and your adversary each choose a pattern, say 1001 and 0100, and the person whose pattern appears first wins.

Here are the natural questions. How do we compute the expectations in the waiting game, and the probabilities in the horse race? Is the pattern that wins in the waiting game more likely to win in the horce race? There are several ways how to solve this problem (a particularly elegant one uses the so called Optional Stopping Theorem for martingales), but we will be using Markov chains.

The markov chain X_n we will use has the state space all patterns of length ℓ. Each time, the chain transitions into the pattern obtained by appending 1 (with probability p) or 0 (with probability $1-p$) at the right end of the current pattern, and deleting the symbol at the left end of the current pattern. That is, the chain simply keeps track of the last ℓ symbols in the sequence of tosses.

There is a slight problem before we have ℓ tosses. For now, just assume that the chain starts with some particular sequence of ℓ tosses, chosen in some way.

The one thing that we can figure out immediately for this chain is its invariant distribution. At any time $n \geq 2\ell$, and any pattern A with k 1's and $\ell - k$ 0's,

$$P(X_n = A) = p^k (1-p)^{\ell-k}$$

as the chain is generated by independent coin tosses! Therefore, the invariant distribution of X_n assigns A the probability

$$\pi_A = p^k (1-p)^{\ell-k}.$$

Now if you have two patterns B and A, denote by $N_{B\to A}$ the expected number of *additional* tosses you need to get A provided that the first tosses ended in B. Here, if A is a subpattern of B, this *does not count*, you have to actually make A in the additional tosses, although you can use *a part* of B. For example, if $B = 111001$ and $A = 110$, and next tosses are 10 then $N_{B\to A} = 2$, and if the next tosses are 001110 then $N_{B\to A} = 6$.

Also denote

$$E(B \to A) = E(N_{B\to A}).$$

Our initial example can therefore be formulated as follows: compute

$$E(\emptyset \to 1011101).$$

The convergence theorem for Markov chains guarantees that, for every A

$$E(A \to A) = \frac{1}{\pi_A}.$$

The hard part of our problem is over. We now show how to analyze the waiting game on the example.

We know that
$$E(1011101 \to 1011101) = \frac{1}{\pi_{1011101}}.$$

However, starting with 1011101, we can only use the *overlap* 101 to help us get back to 1011101, so that

$$E(1011101 \to 1011101) = E(101 \to 1011101).$$

Now to get from $\emptyset$ to 1011101 we have to get first to 101, and then from there to 1011101, so that

$$E(\emptyset \to 1011101) = E(\emptyset \to 101) + E(101 \to 1011101).$$

Now we have reduced the problem to 101 and we iterate our method:

$$\begin{aligned} E(\emptyset \to 101) &= E(\emptyset \to 1) + E(1 \to 101) \\ &= E(\emptyset \to 1) + E(101 \to 101) \\ &= E(1 \to 1) + E(101 \to 101) \\ &= \frac{1}{\pi_1} + \frac{1}{\pi_{101}}. \end{aligned}$$

The final result is

$$\begin{aligned} E(\emptyset \to 1011101) &= \frac{1}{\pi_{1011101}} + \frac{1}{\pi_{101}} + \frac{1}{\pi_1} \\ &= \frac{1}{p^5(1-p)^2} + \frac{1}{p^2(1-p)} + \frac{1}{p}, \end{aligned}$$

which is equal to $2^7 + 2^3 + 2 = 138$ when $p = 1/2$.

In general, the expected time $E(\emptyset \to A)$ can be computed by adding to $1/\pi_A$ all the overlaps between A and its shifts, that is, all the patterns by which A begins and ends. In the example, the overlaps are 101 and 1. The more overlaps A has, the larger $E(\emptyset \to A)$ is. Accordingly, for $p = 1/2$, of all patterns of length ℓ, the largest expectation is $2^\ell + 2^{\ell-1} + \cdots + 1 = 2^{\ell+1} - 1$ (for constant patterns $11\ldots1$ and $00\ldots0$) and the smallest 2^ℓ when there is no overlap at all (for example, for $100\ldots0$).

Now that we know how to compute expectations in the waiting game, we will look at the horse race. Fix two patterns A and B and let $p_A = P(A \text{ wins})$ and $p_B = P(B$ wins). The trick is to consider the time N, the first time *one* of the two appears. Then we can write

$$N_{\emptyset\to A} = N + I_{\{B \text{ appears before } A\}} N_{B\to A}$$

In words, to get to A you either stop at N or go further starting from B, but the second case only occurs when B occurs before A. Taking expectations,

$$\begin{aligned}
&E(\emptyset \to A) = E(N) + p_B \cdot E(B \to A),\\
&E(\emptyset \to B) = E(N) + p_A \cdot E(A \to B),\\
&p_A + p_B = 1.
\end{aligned}$$

We aleady know how to compute $E(\emptyset \to A)$, $E(\emptyset \to B)$, $E(B \to A)$, and $E(A \to B)$ so this is a system of three equations with three unknowns: p_A, p_B and N.

For our initial example $A = 1001$ and $B = 0100$, and $p = 1/2$, we can compute $E(\emptyset \to A) = 16 + 2 = 18$ and $E(\emptyset \to B) = 16 + 2 = 18$.

How do we compute $E(B \to A) = E(0100 \to 1001)$? First we note $E(0100 \to 1001) = E(100 \to 1001)$ and then $E(\emptyset \to 1001) = E(\emptyset \to 100) + E(100 \to 1001)$, so that $E(0100 \to 1001) = E(\emptyset \to 1001) - E(\emptyset \to 100) = 18 - 8 = 10$. Similarly, $E(A \to B) = 18 - 4 = 14$, and then the above three equations for three unknowns give $p_A = 5/12$, $p_B = 7/12$, $E(N) = 73/6$.

Two remarks are in order, each somewhat paradoxical (thus illuminating about probability).

The first remark concerns sequences $A = 1010$ and $B = 0100$. As an exercise, you should verify that $E(\emptyset \to A) = 20$, $E(\emptyset \to B) = 18$, while $p_A = 9/14$. So A loses in the waiting game but wins in the horce race! What is going on? Simply, when A loses in the horce race, it loses by a lot, thereby tipping the waiting game towards B.

The second remark concerns horce race only. Consider the relation $\geq$ given by $A \geq B$ if $P(A \text{ beats } B) \geq 0.5$. Naively, one would expect that this relation is transitive, but this is not true! The simplest example are triples $011 \geq 100 \geq 001 \geq 011$, with probabilities $1/2$, $3/4$ and $2/3$.