Partial Core Fragment:
The Penney Ante Problem

Saleh Hindi

February 12, 2017

1 Introduction

Imagine a two player game where each player is assigned a sequence, for example THTH and HTHH, and a coin is flipped until either player sees their sequence. The first player to see their sequence appear wins. Given two sequences, which sequence is expected to come first in the sequence of coin flips? What is the probability of a certain player winning? Although these two questions sound similar, the result is that in our example, the expected number of turns for sequence A to appear is 20 and for B it is 18. However, surprisingly, the probability of A winning is 9/14 while the probability of B winning is 5/14. Additionally the game has the property that for any sequence A chooses, B can always find a sequence that has a higher probability of winning. These counterintuitive results are the core of the Penney Ante problem, discovered by Walter Penney [3]. My thesis will study this problem through three approaches based in Markov chains, martingales, and a counting approach.

2 Definitions and Notation

Let X_t be a random variable for $t \geq 1$, and let (X_t) be a stochastic sequence of letters chosen uniformly and randomly from a q-letter alphabet. The statespace of (X_t) will be denoted Ω with each state equal to an $x \in \Omega$. In the case of a coin, $q = 2$ and (X_t) represents the sequence of coin flips. Let $A = a_1a_2...a_n$ and $B = b_1b_2...b_n$ be sequences of n letters chosen from the q-letter alphabet. We say sequence A or B wins if it is the first sequence to appear within (X_t) or more precisely, A or B wins if it is the lowest i such that $A = a_1a_2...a_n = X_iX_{i+1}...X_n$ or $B = b_1b_2...b_n = X_iX_{i+1}...X_n$. Let τ_A and τ_B be random variables denoting the number of turns required for A or B to appear the first time. We denote the probability of A winning as $P(\tau_A < \tau_B)$ and we denote the expected time for sequence A and B to appear as $E(\tau_A)$ and $E(\tau_B)$. This notation will be used throughout of the paper.

As stated earlier, $E(\tau_A) < E(\tau_B)$ does not imply $P(\tau_A < \tau_B) > \frac{1}{2}$. To prove and understand this result, we will construct a Markov chain. A Markov chain is a stochastic sequence (X_t) that has the probability that each state of the chain occurs – ie each X_k – is based solely on the previous state, X_{k-1} for all $k > 1$. Markov chains allow the sequence to be represented by a graph with all the possible states of the sequence as vertices. Edges connecting vertices denote probabilities between that a state will occur after another state. More formally,

Definition 2.1 ([8]) *A* Markov chain *is a stochastic sequence such that*

$$P(X_{k+1} = x|X_1, X_2, ..., X_k) = P(X_{k+1} = x|X_k)$$

That is, the probability that each $x \in \Omega$ occurs given the entire history of previous states is only dependent on the most recent event. Denote the probability that $P(X_{k+1} = y|X_k = x)$ as $P(x,y)$. The matrix $P = P(x,y)$ for all $x,y \in \Omega$ is called the transition matrix of (X_t).

The Penney ante game can be represented as a Markov chain because each sequence of coin flips, for example $X_4 = HTHT$, depends only on the previous state X_3 and not the entire history of states. We can use Markov chains to find $E(\tau_A)$ and $P(\tau_A < \tau_B)$ by writing a system of equations for the probability and expected value of each state which will be done next section.

We can also study the Penney ante problem using another mathematical object called a martingale. A martingale is a stochastic sequence such that regardless of the time, the expected value of the random variable is the same. Call a martingale with an expected value of 0 a fair game. More precisely,

Definition 2.2 ([6]) *A* martingale *is a stochastic sequence (X_t) such that for any integer k and for any finite expected value $E(|X_k|)$,*

$$E(X_{k+1}|X_1, ..., X_k) = X_k$$

Let's define a martingale representing the Penney Ante game. Given a stochastic sequence of coin flips (X_t), and for each t, define W_t, assuming $W_0 = 0$,

$$W_{t+1} = \begin{cases} W_t + 2^t & \text{if } X_t = \text{some value that we are betting on} \\ 0 & \text{otherwise} \end{cases}$$

If W_t were to be imagined as some game at the Monte Carlo casino, imagine at each turn t, a player joins the game and bets 1 choosing to play double or nothing until they lose, on each letter of some sequence. Denote "betting AB" as the player betting on the the ith character of A and getting the ith character of b. On each subsequent turn, another player joins the game and bets on the same sequence. At each turn t, W_t is the earnings of all the players during that turn. The players gamble until one player gets a match between the random coin flips and their own sequence. At this point, one player will have fully matched the sequence and the rest of the players will have partial matches along with some winnings. (What is this value at the time of winning, ie a stopping time, represent?). The proof that W_t is a martingale will come later in this paper and at a later deadline.

And finally, we will be using a combinatoric method for finding $E(\tau_A)$ based around finding the generating function for the number of $x \in \Omega$ which contain A as a substring. For a given sequence A, let $f(n)$ denote the number of sequences of length n that do not contain A as a subsequence. A generating function in general can be formally defined as,

Definition 2.3 ([9]) *Given any sequence (X_t), we define its generating function $F(z)$ to be*

$$F(z) = \sum_{n=0}^{\infty} f(n) z^{-n}$$

where z is an unknown variable.

The probability that A does not appear in the first n coin tosses is $f(n)2^{-n}$ [9]. The expected number of coin tosses can be derived from these facts and will involve a number called the correlation of two strings, denoted AB. The correlation AB will be better defined later when outlining Conway's algorithm. Conway devised a "magic algorithm" which let us compute AB which will be used to find the expected winning time and the probability of A winning. This operation AB has a careful connection to the AB defined in the betting strategy. The correlation will give us the expected number of coin tosses until A appears and it is also used to model the above gambling game. The connection between the two AB's will come in a later section.

An interesting property of $F(z)$ is that it is rational in this case so we can write it as $F(z) = \frac{P(z)}{Q(z)}$ for $Q(z) \neq 0$ and coprime $P(z)$, $Q(z)$. From $F(z)$ being rational, we can derive the main combinatoric result which will let us count the number subsequences of a sequence which are not equal to the sequence and thus the number of subsequences which are equal to the sequence.

Finally, we define nontransitivity. In this context, a game is nontransitive if for any sequence A, player B can choose a sequence B such that $P(\tau_A < \tau_B)$. In later sections we will see that the Penney ante game is nontransitive for $n \geq 4$. In general nontransitivity is....

3 Martingales

If we reformulate the Penney ante problem into the following game with W_t representing the total winnings of the game, we can see that W_t is a martingale. Using martingales we can find $E_{\varnothing}(\tau_B)$, $E_A(\tau_B)$, and $P(\tau_B < \tau_A)$. In this game imagine, given sequences A and B, at each turn t, a player joins the game and bets 1 choosing to play double or nothing until they lose, on each letter of some sequence. On a player's ith turn, the player bets on the ith letter of A appearing whereas the t-th letter of B actually appears. We will denote this "betting AB". If the player's bet matches the actual outcome, they double their money, otherwise, they lose all of their money. On each subsequent turn, another player joins the game and also bets AB. At each turn t, W_t is the earnings of all the players during that turn. The players gamble until one player gets a full match of sequence A. At this point, one player will have fully matched the sequence and the rest of the players will have partial matches along with some winnings.

So given a stochastic sequence of coin flips (X_t), and for each t, we can define W_t, assuming $W_0 = 0$, as,

$$W_{t+1} = \begin{cases} W_t + 2^t & \text{if } X_t \text{ = some value that we are betting on} \\ 0 & \text{otherwise} \end{cases}$$

Claim: W_t is a martingale.
Proof: Claim: From our earlier definition of a martingale we know that W_t is martingale if $E(W_{k+1}|W_1, W_2, ..., W_k) = W_k$ for all k. Informally, we can see at each time t, the value of each of the gambler's winnings double or go to zero with a .5 chance each way. That means the expected value of each gambler's winnings at time $t+1$ is the value of their winnings at time t. This seems to indicate that the expected value of W_{t+1} is just the

value of W_t. Thus we can see that W_t is a martingale. More formally... ■

4 Results

4.1 Markov Chains

Using Markov chains, we can construct a system of equations for the expected waiting time to win and probability of a sequence winning. Let p_x denote the probability sequence A of length n wins given that $x \in \Omega$ has occured. To find $p_\emptyset$, the probability of A winning given nothing has happened, we can construct a system of $n+1$ equations for each X_t composed of probabilities times of reaching x from y. Assuming $x - y$ means $P(x,y) > 0$,

$$p_x = \sum_{y-x} P(x,y) p_y$$

In this game where $q = 2$, each sum will have two terms since there will only be at most two states x that are reachable from each state y. Similarly, we can find the expected value with this method but with a small change. Let E_x denote the time to win given x has occurred. Then we can write a system of $n+1$ equations,

$$E_x = \sum_{y-x} P(x,y)(1+E_y)$$

4.2 Generating Functions

When thinking about the Penney ante problem in terms of martingales, we can use Conway's algorithm to find $P(\tau_A < \tau_B)$ and $E(\tau_A)$. Conway devised an algorithm for computing these two values which has been described as an algorithm that "cranks out the answer as if by magic" [3]. The algorithm is as follows,

Theorem 4.1 *(Conway's Algorithm [3]) Given two sequences of length n, A and B, we find the correlation in base 2, AB_2, by the following algorithm:*

1. *loop through integers 1, 2, ..., n,*
2. *At every ith iteration we look at the ith through nth digits of A and compare it to the 1st through the (n-i)th digits of B.*
3. *If these subsequences are equal, the ith digit in the binary representation is 1 but 0 otherwise.*

The binary number of AB_2 is then converted to a decimal number, AB. Once we find AA, AB, BA, and BB, the probability that A precedes B is

$$P(\tau_A < \tau_B) = \frac{AA - AB}{(BB - BA) + (AA - AB)}$$

Furthermore,

$$E(\tau_A) = 2AA$$

Now that we have the correlation AB, we can also use the generating function approach the compute the desired values. The probability that A does not appear in the first n coin tosses is $f(n)2^{-n}$. The expected number of coin tosses until A appears is,

$$\sum_{n=0}^{\infty} f(n)2^{-n} = F(2)$$

From

$$F(z) = \frac{zAA_z}{1 + (z - q)AA_z}$$

we can get that $F(2) = 2AA$. [9] [3]

4.3 Martingales

The betting game described earlier creates a martingale process, (W_t). Notice how (X_t) and (W_t) derive from the same source of randomness, the independently and uniformly chosen coin flips. Although martingales do not directly give us an easy way to compute the expected waiting time and probability of a sequence winning, it provides a justification to Conway's algorithm. Counting all of the string overlaps of AB is checking the earnings of all the player for partial matches between sequences A and B. In binary, moving up each of the digits correspong to the player betting double or nothing. I will go into further detail about martingales and the connection between string enumeration and Conway's algorithm at a future date but know that for now, Conway's algorithm serves as our method of computing the expected waiting time and probability using the martingale approach.

As an example, let $A = THTH$ and $B = HTHH$. Although not all possible sequences have the property that, $E(\tau_A) < E(\tau_B)$ while $P(\tau_A < \tau_B) > \frac{1}{2}$, this particular example goes against our intuition so it is worthwhile to look at. As stated in the introduction, $E(\tau_A)$ is 20, $E(\tau_B)$ is 18, and $P(\tau_A > \tau_B) = \frac{9}{14}$. As a sanity check, we should check that the three methods described in the previous section give the same answer for $P(\tau_A > \tau_B)$ and $E(\tau_B)$.

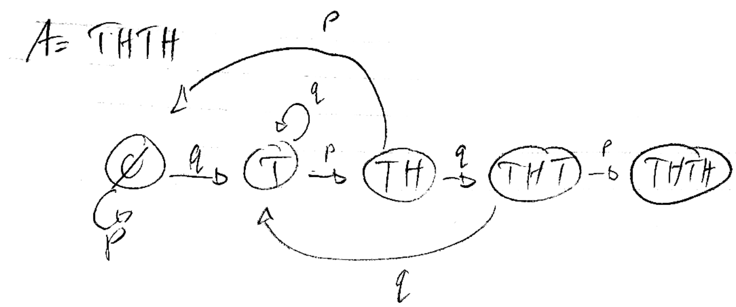

Figure 1: The graph of Ω_A

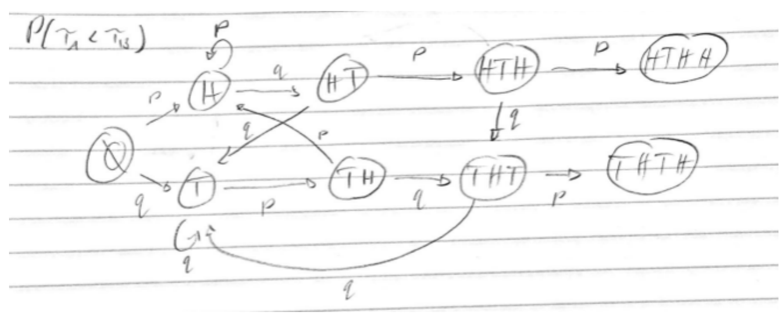

Figure 2: The graph of the game (What is the best way to draw graphs on a computer? This method looks very sloppy...)

With the Markov chain approach, we can directly construct the transition matrix P. The P for $A = THTH$ is,

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{1}$$

Where each ith column or row represents $\emptyset$, T, TH, THT, or $THTH$ accordingly which indicate how close A is to winning. For $B = HTHH$, the states would be $\emptyset$, H, HT, HTH, and $HTHH$.

Using this matrix, we construct the systems of equations for the probability as follows,

$$
\begin{aligned}
p_{\emptyset} &= \frac{1}{2} p_{THT} \\
p_{T} &= \frac{1}{2} p_{TH} + \frac{1}{2} p_{\emptyset} \\
p_{TH} &= \frac{1}{2} p_{THT} + \frac{1}{2} p_{T} \\
p_{THT} &= \frac{1}{2} p_{THTH} + \frac{1}{2} p_{TH} \\
p_{THTH} &= 1
\end{aligned} \tag{2}
$$

Using Mathematica (how should I include the code? Directly in the paper or as a footnote?) we find that $p_{THTH} = \frac{9}{14}$. (Insert Mathematica Code) For the expected value we have,

$$
\begin{aligned}
E_{\emptyset} &= \frac{1}{2}(1 + E_{THT}) \\
E_{T} &= \frac{1}{2}(1 + E_{TH}) + \frac{1}{2}(1 + E_{\emptyset}) \\
E_{TH} &= \frac{1}{2}(1 + E_{THT}) + \frac{1}{2}(1 + E_{THT}) \\
E_{THT} &= \frac{1}{2}(1 + E_{THTH}) + \frac{1}{2}(1 + E_{TH}) \\
E_{THTH} &= 1
\end{aligned} \tag{3}
$$

The calculation for B is omitted because it is essentially the same. Using the same Mathematica code we find that $E_{THTHT} = 20$. Thus concludes the calculation using the Markov chain approach. Conway's algorithm gives us the same result.

Consider $A = THTH$. The calculation AA is as follows,

```
1010
THTH
----
THTH
 THTH
  THTH
   THTH
```

The binary number 1010 in decimal is 9 so $E(\tau_{THTH}) = 18$ and $P(\tau_A < \tau_B) = \frac{9}{14}$.

5 Future Work

In the future I will prove the results stated in the results section. It will also be worthwhile to give a proof of the nontransitivity property which was not stated in the results.

References

[1] Boyer, Robert S., and J. Strother Moore. "A Fast String Searching Algorithm." Communications of the ACM 20.10 (1977): 762-72. Web.

[2] Breen, Stephen, Waterman Michael S., and Zhang Ning. "Renewal Theory for Several Patterns." Journal of Applied Probability 22.1 (1985): 228-34. Web.

[3] Gardner, Martin. "Mathematical Games: On the Paradoxical Situations That Arise from Nontransitive Relations." Scientific American 10 (1974): 120-25. Print.

[4] Grinstead, Charles M., and J. Laurie Snell. Introduction to Probability. Providence, RI: American Mathematical Society, 1997. Print.

[5] Guibas, L.j, and A.m Odlyzko. "String Overlaps, Pattern Matching, and Nontransitive Games." Journal of Combinatorial Theory, Series A 30.2 (1981): 183-208. Web.

[6] Li, Shuo-Yen Robert. "A Martingale Approach to the Study of Occurrence of Sequence Patterns in Repeated Experiments." The Annals of Probability 8.6 (1980): 1171-176. Web.

[7] Nickerson, R. S. "Penney Ante: Counterintuitive Probabilities in Coin Tossing." The UMAP Journal 28.4 (2007): 503-32. JSTOR. Web. 8 Sept. 2016.

[8] Levin, David Asher, Y. Peres, and Elizabeth L. Wilmer. Markov Chains and Mixing times. Providence, RI: American Mathematical Society, 2009. Print.

[9] Odlyzko, A. M. "Enumeration of Strings." Combinatorial Algorithms on Words (1985): 205-28. Web.