ssdnm
środowiskowe
studia doktoranckie
z nauk matematycznych

Wanda Niemyska

Uniwersytet Śląski

Penney's game paradoxes - classic cases and general theory

Praca semestralna nr 1

(semestr letni 2011/12)

Opiekun pracy: dr Witold Bednorz

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

University of Silesia

Wanda Niemyska

Penney's game paradoxes - classic cases and general theory

Term paper
(June 30, 2012)

Under supervision of:
dr Witold Bednorz

1 Introduction

This paper was intended as an attempt to present Penney's game in a wide perspective - as an easy game of chance which is an interesting tool for promoting mathematics on the one hand and on the other hand as a subject, which might be a motivation for much more general studies, in which we can apply some results from pattern matching theory and obtain new interesting results, which possibly may find some applications for example in mathematical modeling of gene mutations or in game theory.

In this section we introduce contents of the paper and few notations which will be used throughout this paper. The next section is devoted to the study of classic Penney's game. This simple game of chance is called from Walter Penney, mathematician from Meryland, who has invented it in 1969. The classic version of Penney's game is intended for two players, each of them selects a sequence of Heads and Tails of a length 3, and then, subsequently, a fair coin is tossed until one of the player's sequence appears – the player whose sequence appears first is the winner. There are 8 different 3-bit long sequences of Heads and Tails, therefore we have 28 pairs of them.
Probably the simplest way to studying chances of winning the game by a player is to use stochastic graphs, which are pretty intuitive also for non mathematicians and that is strong point of them. In this section I will present analysis of one exemplary game, in which players have chosen sequences Head-Head-Tail and Tail-Head-Head and next I will introduce summary of the results of all classic games, emphasizing paradoxes, which appear among them.

More general version of Penney's game let players select sequences of any length (even they do not have to be of the same length) and instead of a fair coin we can use an unfair die with countable set of results. Stochastic graphs become really big and totally impractical, when chosen sequences are long. Very simple algorithm for computing the odds of winning for the competing patterns was discovered by Conway and is known as „Conway's Formula". Crucial quantities for this formula are some correlations of two patterns, which include information about overlaps between them. Throughout the paper we use the following notation.
Let X be an arbitrary but fixed discrete random variable and let Σ be the set of possible values of X. Set Σ is a countable - this is a representation of „an unfair die" which we have mentioned about. Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables having the same distribution as X.

Definition. Let $A = (a_1, a_2, \ldots, a_k)$ and $B = (b_1, b_2, \ldots, b_m)$ be two sequences over Σ. For $i \leqslant k$ let us denote first i elements of A by $A^{(i)}$ and last i elements of A by $A_{(i)}$.
For $i = 1, 2, ..., min(k, m)$ we define

$$\delta_i(A,B) = \begin{cases} \frac{1}{P(X=b_1)\cdot P(X=b_2)\cdots P(X=b_i)} & \text{if } A_{(i)} = B^{(i)}, \\ 0 & \text{otherwise.} \end{cases}$$

Then define

$$A : B = \delta_1(A, B) + \ldots + \delta_{min(k,m)}(A, B).$$

Conway's Formula says that the odds that sequence B wins are given by

$$\frac{A : A - A : B}{B : B - B : A}.$$

Conway's proof of this elegant formula was never published. The first widely disseminated proof is due to Collings [*S. Collings (1977). Impossible probabilities. Unpublished manuscript*].
In the third section we follow Jakubowski and Sztencel [3] to derive Conway's Formula using combinatorial methods - for the simplest case of Penney's game, when tossed coin is fair. Li [4] obtained other proof using martingales. His proof is more general and obtained formula includes the case of the game for many players who select sequences of any length and instead of a fair coin we use an unfair die with countable possible results. Li's proof can be found in Section 4.

One of the facts that can be deduced from Conway's Formula is: given a sequence of length at least three, there always exists another sequence of the same length that beats the given sequence more than half of the time. In the fifth section we discuss the proof of this fact, mainly based on Guibas and Odlyzko's article [5]. We also summarize without proofs other paradoxes, which are showed in the second section for classic Penney's game, and now we find out that they exist in much more general cases too. We follow Guibas and Odlyzko [5], as well as Chen and Zame [6].

The last section contains a brief summary of that paper. Some open problems are specified and some applications are indicated.

2 Classic Penney's game paradoxes

Definition. We call a *stochastic graph* an ordered pair (S, Q), where S is a finite set of vertices and Q is a function from $S \times S$ into $\mathbb{R}$ such that the following conditions are satisfied:

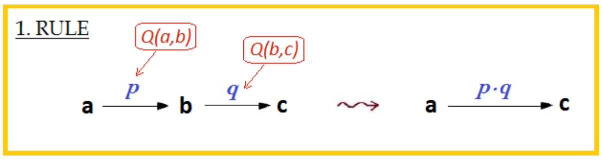
- $Q(i, j) \geqslant 0$ for all $i, j \in S$;

- $\sum_{j \in S} Q(i, j) = 1$ for all $i \in S$.

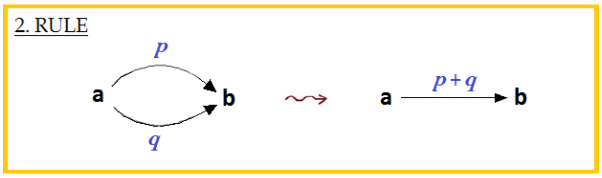
- $Q(i, j)$ we interpret as *the probability of transition* between states i and j in one step. We use the notation p_{ij}.
- A pair (i, j) is *an edge of the graph*, if $p_{ij} > 0$. Edges of the stochastic graphs are directed.

- If $p_{jj} = 1$, then j is *a boundary vertice* (such vertice that we know that if we come into it once we will stay here forever).
- The set B of all boundary vertices is called *the boundary of the graph.*

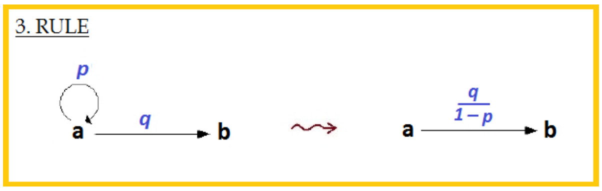
Lets look at a few rules of stochastic graph's reduction. We will use them in a moment.



Numbers p and q are the values of probabilities of transition in one step between next vertices. This rule results from the fact that the transitions – in our case it will be successive coin tosses - are independent events.



The second rule is also very intuitive and comes out of the formula for probability of the sum of mutually exclusive events.

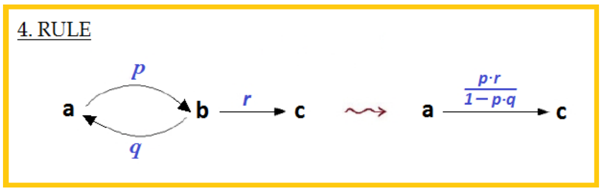


The third rule is a loop's reduction – when we are going from vertice a to vertice b first we can get through the loop from a to a as many times as we want. Thus possible ways from a to b are of such form:

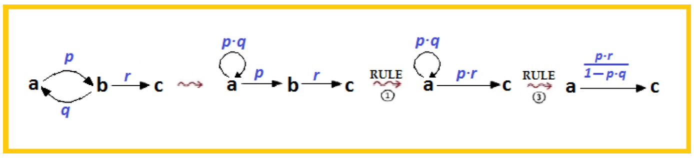
$$a \to b, \qquad a \to a \to b, \qquad a \to a \to a \to b, \quad \dots$$

Therefore the probabilities of going through successive ways from a to b form a geometrical progression

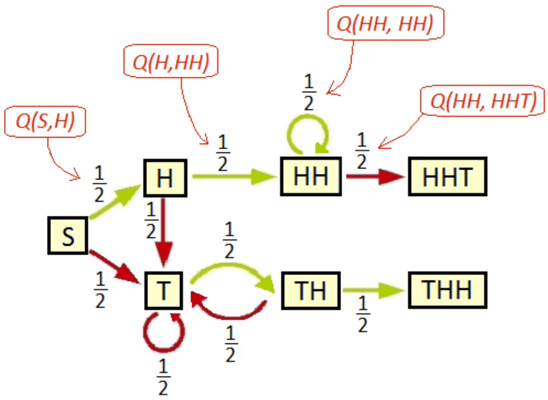
$$q + p \cdot q + p \cdot p \cdot q + \ldots = q \cdot \frac{1}{1-p}$$



The last rule – we can see that first we can repeat many times such sequence: transition from a to b and return from b to a, till finally we go to b and directly to c. So we can put the loop in the graph and then we apply prior rules

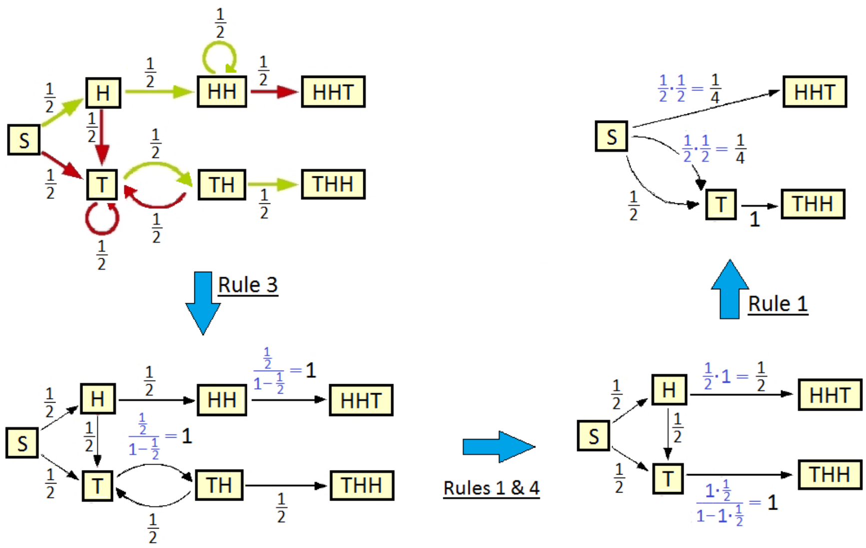


We are ready to calculate chances of winning for particular players. Let's look at the game, where players have chosen sequences HHT and THH. This is stochastic graph appropriate for this game

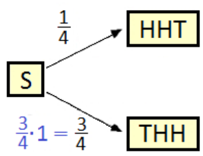


Chosen sequences HHT and THH are boundary vertices and the rest of vertices are prefixes of them. Our interpretation is that we are in a given vertice, if the last coin tosses have formed the sequence from this vertice. Green arrows correspond with tossing the Head and red ones - with tossing the Tail. Numbers over the edges mean the probabilities of transition between vertices –they are values of the function Q from graph's definition. We are tossing a fair coin, so all those numbers equal $\frac{1}{2}$.

Now we can reduce the graph bit by bit:



In the first step we reduce both loops applying third rule. Next we reduce vertices HH and TH using first and forth rules. Then we use first rule to replace the way $(S \to H \to HHT)$ by the direct edge from S to HHT and to replace the way $(S \to H \to T)$ by the direct edge from S to T and in this way we reduce vertice H. Finally we join two edges $(S \to T)$ using second rule and then we reduce vertice T.

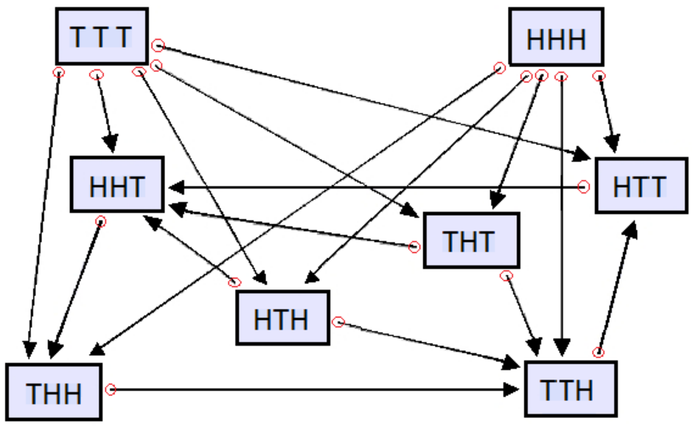


We have got that the sequence THH has as much as three times bigger chances of the winning than HHT. That is the first surprising result – each of eight 3-bit long sequences appears as an outcome of 3 successive coin tosses with the same probability equals $\frac{1}{8}$. So we can suppose that in the game for any two sequences both of them will have the same chances of the winning. However, it happens only in 10 times in 28.

Here is a table with collected chances of winning by a player in all classic Penney's games - the numbers in cells are odds that a sequence from a row wins with a sequence from a column. Ten fair games are emphasized with yellow background in a cell.

	HHH	*HHT*	*HTH*	*THH*	*HTT*	*THT*	*TTH*	*TTT*
HHH								
HHT	$\frac{1}{2}$							
HTH	$\frac{3}{5}$	$\frac{1}{3}$						
THH	$\frac{7}{8}$	$\frac{3}{4}$	$\frac{1}{2}$					
HTT	$\frac{3}{5}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$				
THT	$\frac{7}{12}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$			
TTH	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{2}{3}$		
TTT	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{5}{12}$	$\frac{2}{5}$	$\frac{1}{8}$	$\frac{2}{5}$	$\frac{1}{2}$	

And there are all relations between 3-bit long sequences on the next picture. If there is an arrow between two sequences, it means that the game with those sequences is not fair – they do not have the same chances of winning. The arrow is directed from „worse" sequence to „better" one.



Notice few new paradoxes on this picture:

- At least one arrow goes out from each vertice, it means that for each sequence there is better one. So there is no *the best* sequence in Penney's game and selection's priority is not a privilege in this game – it is more rewarding to wait for our opponent's choice.
- We see the loop $THH < TTH < HTT < HHT < THH$, where we denote by $A < B$ that sequence B is „betteer" than A. Therefore the property of *being better than* is not transitive in classic Penney's game.

Studying stochastic graphs is really interesting - it is kind of a puzzle which is possible to be solved and understood also by non mathematicians. And results of those puzzles often are surprising and it helps to understand that some of the results obtained with probabilistic methods are not consistent with our intuition.

3 Combinatorial derivation of the Conway's Formula for the game for two players and with a fair coin

Let $A = (a_1, a_2, \dots, a_k)$ and $B = (b_1, b_2, \dots, b_m)$ be two sequences over Σ.
In this section Σ will be of the form $\{H, T\}$ (which stand for Head and Tail) and we use a fair coin, therefore $P(X = H) = P(X = T) = \frac{1}{2}$.
Numbers $\delta_i(A, B)$ (for $i = 1, 2, ..., min(k, m)$) take such form:

$$\delta_i(A,B) = \begin{cases} 2^i & \text{if } A_{(i)} = B^{(i)}, \\ 0 & \text{otherwise.} \end{cases} \qquad = 2^i \cdot [A_{(i)} = B^{(i)}].$$

Square bracket equals 1 when the equation inside it is true.

For finite sequence s over Σ, set S of finite sequences over Σ, $s \notin S$, and $n > 0$ we use notation:

- $F(n, S)$ is the set of all n-bit long sequences, which do not include the patterns from the set S;
- $F_s(n, S)$ is the set of all n-bit long sequences, which include pattern s at the end and only there, and which do not include patterns from the set S.

First we derive the formula of mean time of waiting for any sequence.
Take the set $F(n, \{A\})$ of all n-bit long sequences which do not include the pattern A. Next we add the pattern A at the end of each sequence from $F(n, \{A\})$ and get a new set which we denote by G. The cardinalities of those two sets are the same

$$|F(n, \{A\})| = |G|.$$

Observe, that the set G includes the set $F_A(n + k, \emptyset)$ but might be bigger than it, because the pattern A may appear in sequences from G also between n-th and $(n + k)$-th position.
Lets look at this example:

Example 1. We are waiting for the pattern $A = HTH$. We have the sequence of length 4, $TTHT$, in which A has not appeared yet. Now we are adding pattern A at the end of it and lets notice that the sequence HTH appears first time on the fifth position (it is second appearance of HTH on the seventh position):

$$A = HTH, n = 4, \qquad TTHT \Rightarrow TT\underline{HT}\ \underline{H}TH$$

This is because the last element of pattern A, which we have needed, is the same as the first element of A, which we have added. We use such notation for this situation: $A_{(1)} = A^{(1)}$ – the last element of A is the same as the first element of A.

Thus we can divide set G into k classes with respect to the first time, when A appears. Then the cardinality of G maybe written as such sum:

$$|F(n,\{A\})| = |F_A(n+1,\emptyset)|\cdot[A_{(1)} = A^{(1)}]+\cdots+|F_A(n+k,\emptyset)|\cdot[A_{(k)} = A^{(k)}] \quad / : 2^n$$

$$\frac{|F(n,\{A\})|}{2^n} = \frac{|F_A(n+1,\emptyset)|}{2^{n+1}}\cdot\underbrace{2\cdot[A_{(1)} = A^{(1)}]}_{\delta_1(A,A)}+\cdots+\frac{|F_A(n+1,\emptyset)|}{2^{n+k}}\cdot\underbrace{2^k\cdot[A_{(k)} = A^{(k)}]}_{\delta_k(A,A)}.$$

Let us denote by T_A the random variable which stands for the time of waiting until pattern A appears first time.
Using this notation we can rewrite the above equation in such form

$$P(T_A > n) = P(T_A = n+1)\cdot\delta_1(A,A) + \cdots + P(T_A = n+k)\cdot\delta_k(A,A).$$

Now we sum up both sides of equation with respect to n

$$\underbrace{\sum_{n=0}^{\infty} P(T_A > n)}_{\mathbb{E}T_A} = \delta_1(A,A)\cdot\underbrace{\sum_{n=0}^{\infty} P(T_A = n+1)}_{1} + \cdots +$$
$$+\,\delta_k(A,A)\cdot\underbrace{\sum_{n=0}^{\infty} P(T_A = n+k)}_{1},$$

$$\mathbb{E}T_A = \delta_1(A,A) + \cdots + \delta_k(A,A) \quad = A : A.$$

We have proved such nice formula:

Theorem 2. *Mean time of waiting for any arbitrary sequence of Heads and Tails A, when we are tossing a fair coin, equals $A : A$:*

$$\mathbb{E}T_A = A : A.$$

□

Remark. Using last theorem we can easily calculate mean times of waiting for sequences. Note that the average time of waiting for a sequence may not be the same even for sequences of the same length - for overlapping sequences this time is longer. Observe another surprising fact - mean times of waiting for two sequences have little in common with their odds of winning:

- $\mathbb{E}T_{HHT} = \mathbb{E}T_{THH}$, but game between HHT and THH is not fair (as we have showed in Section 2, THH has as much as three times bigger chances of winning than HHT).
- HTH and THH have the same chances of winning though mean times of waiting for them differ ($10 = \mathbb{E}T_{HTH} \neq \mathbb{E}T_{THH} = 8$).
- If the patterns are $HTHT$ and $THTT$, then the odds are 9 to 5 in favor of the former, though it has expected waiting time 20 whereas the latter has only 18.

Now we use analogous methods to derive Conway's Formula - we compute the odds of winning for the competing two patterns of Heads and Tails in Penney's game, in which fair coin is used.
Let $A = (a_1, a_2, \dots, a_k)$ and $B = (b_1, b_2, \dots, b_m)$ be two sequences chosen by players. Denote by N the random variable which stands for the duration of the game - it means the time of waiting until pattern A or pattern B appear for the first time. We use the symbol p_A (p_B) to denote the probability that sequence A (B) appears first.

This time we take the set $F(n, \{A, B\})$ of all n-bit long sequences which do not include the patterns A and B. Next we add A at the end of each sequence from $F(n, \{A, B\})$ to get a new set G_A and we add B at the end of each sequence from $F(n, \{A, B\})$ to get a new set G_B.
The cardinalities of those three sets are the same

$$|F(n, \{A, B\})| = |G_A| = |G_B|$$

Observe, that the set G_A includes the set $F_A(n + k, \{B\})$ but might be bigger than it, because the pattern A or B may appear in sequences from G_A also between n-th and $(n + k)$-th position. In the same way the set G_B includes the set $F_B(n + m, \{A\})$.

By a similar argument as in Theorem 3 we can divide sets G_A and G_B into classes with respect to the first time, when A or B appear. Then the cardinalities of sets maybe written as

$$\begin{aligned}|F(n, \{A, B\})| &= |F_A(n+1, \{B\})| \cdot [A_{(1)} = A^{(1)}] + \dots + |F_A(n+k, \{B\})| \cdot [A_{(k)} = A^{(k)}] \quad + \\ &+ |F_B(n+1, \{A\})| \cdot [B_{(1)} = A^{(1)}] + \dots + |F_B(n+k, \{A\})| \cdot [B_{(k)} = A^{(k)}] = \\ &= |F_A(n+1, \{B\})| \cdot [A_{(1)} = B^{(1)}] + \dots + |F_A(n+m, \{B\})| \cdot [A_{(m)} = B^{(m)}] \quad + \\ &+ |F_B(n+1, \{A\})| \cdot [B_{(1)} = B^{(1)}] + \dots + |F_B(n+m, \{A\})| \cdot [B_{(m)} = B^{(m)}].\end{aligned}$$

We divide this equation by 2 raised to the n-th power

$$
\begin{aligned}
\frac{|F(n,\{A,B\})|)}{2^n} &= \frac{|F_A(n+1,\{B\})|}{2^{n+1}} \cdot \underbrace{2 \cdot [A_{(1)} = A^{(1)}]}_{\delta_1(A,A)} + \cdots + \frac{|F_A(n+k,\{B\})|}{2^{n+k}} \cdot \underbrace{2^k \cdot [A_{(k)} = A^{(k)}]}_{\delta_k(A,A)} + \\
&+ \frac{|F_B(n+1,\{A\})|}{2^{n+1}} \cdot \underbrace{2 \cdot [B_{(1)} = A^{(1)}]}_{\delta_1(B,A)} + \cdots + \frac{|F_B(n+k,\{A\})|}{2^{n+k}} \cdot \underbrace{2^k \cdot [B_{(k)} = A^{(k)}]}_{\delta_k(B,A)} \\
&= \frac{|F_A(n+1,\{B\})|}{2^{n+1}} \cdot \underbrace{2 \cdot [A_{(1)} = B^{(1)}]}_{\delta_1(A,B)} + \cdots + \frac{|F_A(n+m,\{B\})|}{2^{n+m}} \cdot \underbrace{2^m \cdot [A_{(m)} = B^{(m)}]}_{\delta_m(A,B)} + \\
&+ \frac{|F_B(n+1,\{A\})|}{2^{n+1}} \cdot \underbrace{2 \cdot [B_{(1)} = B^{(1)}]}_{\delta_1(B,B)} + \cdots + \frac{|F_B(n+m,\{A\})|}{2^{n+m}} \cdot \underbrace{2^m \cdot [B_{(m)} = B^{(m)}]}_{\delta_m(B,B)}.
\end{aligned}
$$

Using random variables N, T_A and T_B we can rewrite above equation as

$$
\begin{aligned}
P(N > n) &= P(N = T_A = n+1) \cdot \delta_1(A,A) + \cdots + P(N = T_A = n+k) \cdot \delta_k(A,A) + \\
&+ P(N = T_B = n+1) \cdot \delta_1(B,A) + \cdots + P(N = T_B = n+k) \cdot \delta_k(B,A) \\
&= P(N = T_A = n+1) \cdot \delta_1(A,B) + \cdots + P(N = T_A = n+m) \cdot \delta_m(A,B) + \\
&+ P(N = T_B = n+1) \cdot \delta_1(B,B) + \cdots + P(N = T_B = n+m) \cdot \delta_m(B,B).
\end{aligned}
$$

Now we sum up all sides of equation with respect to n and we obtain

$$
\begin{aligned}
\mathbb{E}N &= p_A \cdot \delta_1(A,A) + \cdots + p_A \cdot \delta_k(A,A) + p_B \cdot \delta_1(B,A) + \cdots + p_B \cdot \delta_k(B,A) \\
&= p_A \cdot \delta_1(A,B) + \cdots + p_A \cdot \delta_m(A,B) + p_B \cdot \delta_1(B,B) + \cdots + p_B \cdot \delta_m(B,B).
\end{aligned}
$$

Equivalently

$$
\mathbb{E}N = p_A \cdot (A : A) + p_B \cdot (B : A) = p_A \cdot (A : B) + p_B \cdot (B : B).
$$

This gives

$$
\frac{p_B}{p_A} = \frac{A : A - A : B}{B : B - B : A}.
$$

This way we have proved the following theorem

Theorem 3. *For any arbitrary sequences of Heads and Tails A and B the odds that B will win in Penney's game with a fair coin are given by*

$$
\frac{A : A - A : B}{B : B - B : A}.
$$

□

Corollary 4. *For $A = (a_1, \ldots, a_k)$ and $B = (b_1, \ldots, b_m)$ we have $T_A > T_B$, if $k > m$.*

PROOF Since $A_{(k)} = A = A^{(k)}$, we have $\delta_k(A, A) = 2^k$. Therefore

$$\mathbb{E}T_B = B : B = \delta_1(B, B) + \cdots + \delta_m(B, B) \leqslant 2^1 + \cdots + 2^m = 2^{m+1} - 2 < \\ < 2^{m+1} \leqslant 2^k = \delta_k(A, A) \leqslant A : A = \mathbb{E}T_A.$$

□

Remark. Last corollary says that on average we need to wait longer for each longer sequence than for shorter one. However we can use last theorem to prove that there is at least one example of game where longer sequence is „better" than shorter one.

For $A = \underbrace{H \dots H}_{2n-1}$ and $B = \underbrace{T \dots T}_{n}\underbrace{H \dots H}_{n}$ we have:

$$A : A = 2^1 + \ldots + 2^{2n} = 2^{2n} - 2, \qquad A : B = 0,$$
$$B : B = 2^{2n}, \qquad B : A = 2^1 + \ldots + 2^n = 2^{n+1} - 2,$$

Therefore

$$\frac{p_B}{p_A} = \frac{A : A - A : B}{B : B - B : A} = \frac{2^{2n} - 2 - 0}{2^{2n} - 2^{n+1} + 2} > 1$$
$$\Leftrightarrow \quad 2^{2n} - 2 > 2^{2n} - 2^{n+1} + 2$$
$$\Leftrightarrow \quad 2^{n+1} > 4 \quad \Leftrightarrow \quad n \geqslant 2$$

□

4 Conway's Formula derived with the use of martingales for Penney's game with many players and an unfair die

In probability theory, a martingale is a model of a fair game where no knowledge of past events can help to predict future winnings. Lets look at such example of a fair game:

Example 5. (from Li's article [4]) Let a die, which shows a, b and c with respective probabilities $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{6}$, be rolled repeatedly. Let A be the sequence (a, c, a). We want to compute $\mathbb{E}T_A$.
Imagine that a gambler bets 1 dollar on the sequence A according to the following rules of fair odds. At the first roll, if a appears, he receives 2 dollars (including his bet) and must parlay the 2 dollars on the occurrence of c at the second roll. In case he wins again, he receives 12 dollars and must parlay the whole amount of 12 dollars on the occurrence of a at the third roll. If he wins three times in a row, he receives the total amount of 24 dollars and the game is over. Lets notice - after each roll expected gambler's winnings equals 0.

Now suppose that before each roll a new gambler joins the game and start betting 1 dollar on the same sequence A. Say the roll turn out to be $(b, a, a, c, b, a, a, c, a)$. Gamblers $1, 2, 3, 4, 5, 6$ and 8 lose at first, third, fifth, fourth, fifth, seventh and eighth rolls, respectively. At the ninth roll the occurrence of A ends the game and Gambler 7 receives 24 dollars. The only other winner is Gambler 9, who receives 2 dollar. In general at the end of the game, the last gambler should receive 2 dollars and the third last gambler should receive 24 dollars. Thus, however the rolls turn out, the receipts of all the gamblers in the game total up to 26 dollars. But their total payment always equals T_A. So their net gain is $26 - T_A$. Since the odds are fair, the expected net gain should be 0. Therefore $\mathbb{E}T_A = 26$.
Now let $B = (b, a, a, c)$ be another sequence. Given a starting sequence B, we want to compute the expected waiting time for a sequence A, which we denote by $\mathbb{E}T_{BA}$. Suppose that the first four rolls yield the sequence B. Right before the fifth gambler joins the game, the total fortune of the gamblers amounts to 0+0+12+0 dollars. The total net gain of all the gamblers as the subsequent rolls will be $26 - T_{BA} - 12$. Again, since the odds are fair, we find that $\mathbb{E}T_{BA} = 14$.

In this section we introduce Li's [4] proof of Conway's Formula which uses martingales. First we give a brief exposition of martingales theory.

Definition.

- A *filtration* in a σ-field $\mathcal{F}$ is a sequence $(\mathcal{F}_t)_{t\in T}$ of sub-σ-fields of $\mathcal{F}$ such that $\mathcal{F}_t \subset \mathcal{F}_s$ for all $t < s$.
- A sequence of random variables $(Z_t)_{t\in T}$ is *adapted* to the filtration if Z_t is $\mathcal{F}_t$-measurable for all $t \in T$.
- Particularly a sequence of random variables $(Z_t)_{t\in T}$ is adapted to the *natural filtration* $(\mathcal{F}_t)_{t\in T}$, where $\mathcal{F}_t = \sigma(Z_s : s \leqslant t), \quad t \in T$.
- A random variable $\tau : \Omega \to T \cup \{+\infty\}$ is called a *stopping time* if $\{\tau \leqslant t\} \in \mathcal{F}_t$ for all $t \in T$ what is equivalent to condition that $\{\tau = t\} \in \mathcal{F}_t$ for all $t \in T$.
  For example moment of first visit in a set $B \in \mathcal{B}(\mathbb{R})$, which is defined as $\tau_B(\omega) = inf\{t \in T : Z_t(\omega) \in B\}$, is stopping time.

Definition.

- Let $(\mathcal{F}_t)_{t\in T}$ be a filtration in $\mathcal{F}$ and let $(Z_t)_{t\in T})$ be a sequence of random variables adapted to that filtration.
  Then $(Z_t)_{t\in T})$ is a *martingale* (with respect to $(\mathcal{F}_t)_{t\in T}$), if:
  1. $\mathbb{E}|Z_t| < \infty$
  2. $\mathbb{E}(Z_t|\mathcal{F}_s) = Z_s \quad$ for $s \leqslant t, \quad s, t \in T$.
- In a discrete case, $T = \mathbb{N}$, we obtain that $(Z_n)_{n>0}$ is a martingale, if

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = Z_n, \quad n = 1, 2, \ldots$$

  or

$$\mathbb{E}(Z_{n+1} - Z_n|\mathcal{F}_n) = 0, \quad n = 1, 2, \ldots$$

- If $(\mathcal{F}_n)_{n>0}$ is natural filtration for a sequence of random variables $(Z_n)_{n\geqslant 0}$, then martingale's condition 2. has such form:

$$\mathbb{E}(Z_{n+1}|Z_n, \dots Z_1) = Z_n$$

The last definition will be used further - if we write that a sequence is a martingale, it will mean discrete martingale with natural filtration.

The following lemma and theorem are well known (see, for instance, Theorem 9 on page 230 in [3]).

Lemma 1. *If a sequence of random variables $(Z_n)_{n>0}$ is a martingale and τ is a stopping time, then a stopped sequence $(Z_{n\wedge\tau})_{n\geqslant 0}$, where $\wedge$ stands for minimum, is also a martingale.*

Theorem 6. *(Doob) Let $(Z_n)_{n>0}$ be a martingale and N a stopping time. If $\mathbb{E}|Z_N| < \infty$ and*

$$\lim_{k\to\infty} \inf \int_{\{N>k\}} |Z_k| dP = 0,$$

then $\{Z_1, Z_N\}$ is a martingale and hence $\mathbb{E}Z_N = \mathbb{E}Z_1$.

Lets come back to Conway's Formula. We return to Σ which is any countable set of possible values of X in this section.
Firstly we will prove such lemma:

Lemma 2. *Given a starting sequence B, the expected waiting time for a sequence A is $T_{BA} = A : A - B : A$, provided that A is not a connected subsequence of B. In particular the expected waiting time for the sequence A (without a starting sequence) is $A : A$.*

PROOF

The scheme of this proof is the following:

1). For every nonnegative integer n, let ω_n denote the sequence $(X_1, \dots, X_n)$. Thus ω_n is a random sequence over Σ. Define the random variable

$$Z_n = B\bar{\omega}_n : A - n,$$

where $B\bar{\omega}_n$ is a sequence B followed by the sequence ω_n. Clearly:

$$Z_{T_{BA}} = A : A - T_{BA}. \tag{1}$$

2). We will prove that assumptions of Doob's Theorem are satisfied for the sequence $(Z_n)_{n\geqslant 0}$ and stopping time T_{BA}:

(A) the sequence $(Z_n)_{n\geqslant 0}$ is a martingale;

(B) $\mathbb{E}|Z_{T_{BA}}| < \infty$;

(C) $\lim_{n\to\infty} \inf \int_{\{N>n\}} |Z_n| dP = 0$.

3). Then we can apply Doob's Theorem and obtain that

$$\mathbb{E}Z_{T_{BA}} = \mathbb{E}Z_0 = B : A.$$

From (1), it follows that

$$\mathbb{E}T_{BA} = B : B - A : B.$$

This would prove the theorem.

The only point remaining to prove is point 2). - that assumptions of Doob's Theorem are satisfied.

(A) Claim that the sequence $(Z_n)_{n\geqslant 0}$ is a martingale. As before we write $A = a_1 \dots a_k$ and $B = b_1 \dots b_m$. For integers $n \geqslant 0$ and $j \geqslant 1-m$, define

$$M_n^{(j)} = \begin{cases} 0, & \text{if } n < j, \\ \frac{1}{P(X=a_1)\cdots P(X=a_{n-j+1})} - 1 & \text{if } \bar{B\omega}_{n(n-j-1)} = A^{(n-j-1)}, \\ -1, & \text{otherwise.} \end{cases}$$

Hence for $n \geqslant j$ we have $M_n^{(j)} = \delta_{n-j+1}(\bar{B\omega}_n, A)$.
When $n < T_{BA}$, we may interprete the quantity $M_n^{(j)}$ as the gain of the j-th gambler at the time n in the game of fair odds described at the end of Example 5. This shows that, for every fixed j, the sequence $\{M_{n\wedge T_{BA}}^{(j)}\}_{n=0,1,\dots}$ is a martingale (because all the time expected value of the payment for fixed gambler in one round. $M_{n+1\wedge T_{BA}}^{(j)} - M_{n\wedge T_{BA}}^{(j)}$ in the game of fair odds equals 0). Since

$$\Sigma_{j=1-m}^{\infty} M_n^{(j)} = \Sigma_{j=1-m}^{n} M_n^{(j)} = \bar{B\omega}_n : A - (n+m) = Z_n - m,$$

for $n \geqslant T_{BA}$, we see that Z_n is also a martingale.

(B) Claim that $\mathbb{E}|Z_{T_{BA}}|$ is finite. Since the random variable X assumes every value in Σ with a positive probability, T_{BA} is dominated by a geometric random variable. Therefore $\mathbb{E}T_{BA}$ is finite.
From (1), it follows that

$$\mathbb{E}Z_{T_{BA}} = \mathbb{E}(A : A - T_{BA}) = \mathbb{E}|A : A - T_{BA}| \leqslant \mathbb{E}|\underbrace{A : A}_{number}| + \underbrace{\mathbb{E}|T_{BA}|}_{finite} < \infty.$$

(C) On the set $\{T_{BA} > n\}$, we have

$$|Z_n| = |\bar{B\omega}_n : A - n| \leqslant |\bar{B\omega}_n : A| + n \leqslant A : A + T_{BA}.$$

This implies that

$$\lim_{n\to\infty} \int_{\{T_{BA}>n\}} |Z_n| dP \leqslant \lim_{n\to\infty} \int_{\{T_{BA}>n\}} (A : A + T_{BA}) dP = 0.$$

□

Now we can turn to the main theorem - Conway's formula for the game, in which many players take part choosing sequences of any length and instead of a fair coin an unfair die is used.

Let $A_1, A_2, \ldots, A_n$ be a sequence over Σ. For each i, we want to calculate the probability that A_i precedes all the remaining $n-1$ sequences in a realization of the process $X_1, X_2 \ldots$. Naturally we assume that none of the sequences contain any other as a connected subsequence. As before we consider the situation when a sequence B is given at the beginning of the process. Write T_i for T_{BA_i}. Let T be the minimum among $T_1, \ldots, T_n$. We want to compute $P(T = T_j)$ for each j (we denote it with p_j).

Theorem 7. *Let $X, X_1, X_2, \ldots$ be discrete i.i.d. random variables and $A_1, \ldots, A_n$ be finite sequence of possible values of X not containing one another. Let B be another such sequence not containing any A_i. Given the starting sequence B, let p_i be the probability that A_i precedes the remaining $n-1$ sequences in a realization of the process $X_1, X_2 \ldots$. Then for every i,*

$$\Sigma_{j=1}^{n} p_j A_j : A_i = \mathbb{E}T + B : A_i,$$

where T is the stopping time when any A_i occurs. In particular when B is void, we have, for every i,

$$\Sigma_{j=1}^{n} p_j A_j : A_i = \mathbb{E}T.$$

With $p_1 + \ldots + p_n = 1$ we obtain system of $n+1$ linear equations for $p_1, \ldots, p_n$ and T.

PROOF

We have

$$\mathbb{E}T_i = \mathbb{E}T + \mathbb{E}(T_i - T) = \mathbb{E}T + \Sigma_{j=1}^{n} p_j \mathbb{E}(T_i - T | T = T_j). \tag{2}$$

From Lemma 2 we also know that

$$\mathbb{E}T_i = A_i : A_i - B : A_i \tag{3}$$

and

$$\mathbb{E}(T_i - T | T = T_j) = A_i : A_i - A_j : A_i \tag{4}$$

Substituting (3) and (5) in (2), we have

$$A_i : A_i - B : A_i = \mathbb{E}T + \Sigma_{j=1}^{n} p_j (A_i : A_i - A_j : A_i) = \mathbb{E}T + A_i : A_i - \Sigma_{j=1}^{n} p_j A_i : A_i \tag{5}$$

This proves the theorem.

□

The nonsingularity of the n by n submatrix $(A_j : A_i - B : A_i)$ was proved by Gerber. Thats why we are sure that system of linear equations from the theorem has solution.

5 General version of Penney's game paradoxes

In the second section we have given a brief exposition of paradoxes in classic Penney's game which is intended for two players who select sequences of Heads and Tails of a length 3 and fair coin is tossed. In particular we have pointed that for each sequence of a length 3 there is better one of the same length. So there is no best sequence in classic Penney's game and selection's priority isn't a privilege in this game – it's more rewarding to wait for our opponent's choice. Furthermore we have indicated one „cycle" among 3-bit long sequences, therefore property of „being better" isn't transitive in classic Penney's game.
In this section we will discuss the case of the game played with an unbiased q-sided die, in which two players select sequences of any but the same length. Thus we have $|\Sigma| = q$ and the random variable X assumes every value in Σ with the same probability equal $\frac{1}{q}$.

Theorem 8. *For any sequence $A = a_1 \dots a_k$ over Σ, for $k \geqslant 3$, there is a sequence B of the same length as A, which is „better" than A, it means it has bigger chances of winning than its opponent.*

PROOF

The probability that A (B) has bigger chances of winning we denote by p_A (p_B). So we want to prove that there exist the sequence B, for which $p_B > p_A$ which is equivalent to

$$\frac{p_B}{p_A} > 1 \quad \Leftrightarrow \quad \frac{A:A - A:B}{B:B - B:A} > 1 \quad \Leftrightarrow \quad A:A - A:B > B:B - B:A.$$

Thus we are looking for such sequence B that $A : B$ and $B : B$ are small, and $B : A$ is relatively large.

Let $A' = a_1, \dots, a_{k-1}$ and let r be the *basic period* of A' (that is, the smallest nonzero shift that causes A' to overlap itself, it means the smallest number of terms which we need to add at the end of A' to have it repeated), so that $a_{r+1} \dots a_{k-1} = a_1 \dots a_{k-r-1}$, what we denote by $A'_{(k-r-1)} = A'^{(k-r-1)}$. So the smallest r corresponds with the largest $(k-r-1)$ such that $\delta_{k-r-1}(A', A') > 0$. (If A' has the trivial autocorrelation, $A_{(i)} \neq A^{(i)}$ for all $i < k-1$, then we take $r = k-1$).
Choose b so that $b \neq a_r$ and set $B = bA' = ba_1 \dots a_{k-1}$. This choice ensures that B has only periods $> r$:

* r would be a period of $B \quad \Leftrightarrow \quad a_r \dots a_{k-1} = ba_1 \dots a_{k-r-1} \quad \Rightarrow \quad a_r = b$, which is a contradiction;

* if B had a period $< r$, then it would be also a period of A', which contradicts the minimality of r.

Let t be the basic period of B. We have $t > r$. Thus the reason for the particular choice of B is that it makes the basic period of B very large, and this way - $B : B$ is small. The rest of the proof falls naturally into three parts:

1). combinatorial proof of the fact that $t \geqslant \lfloor \frac{k+1}{2} \rfloor + 1$;

2). corollary that then

$$A:B \leqslant \frac{q^{\lfloor \frac{k}{2} \rfloor}-1}{q-1}, \tag{6}$$

$$B:B-B:A \leqslant q^{k-1}-q^{k-2}+q^{\lfloor \frac{k}{2} \rfloor -2}, \tag{7}$$

$$A:A \geqslant q^k; \tag{8}$$

3). finally application (6), (7) and (8) to an inequality $(A:A-A:B)>(B:B-B:A)$, which completes the proof.

Follow above steps.

Ad. 1). We have $t>r$. First we study only the case $t \leqslant k-2$.

* Any period of B is a period of A':

$$s \text{ is a period of } B \quad \Leftrightarrow \quad a_s a_{s+1}\dots a_{k-1} = b a_1 \dots a_{k-s-1} \quad \Rightarrow$$
$$a_{s+1}\dots a_{k-1} = a_1 \dots a_{k-s-1} \quad \Leftrightarrow \quad s \text{ is a period of } A'$$

On the other hand, no period of B that is $\leqslant k-2$ can be a multiple of r, the basic period of A':

if $A^* = a_1 \dots a_r$, then as r is the basic period of A' we have $A' = A^* \dots A^* A^+$, where A^+ is a prefix of A^* (which may be the empty string), and since $b \neq a_r$, we have $b \neq a_{mr} = a_r$ for all $m \geqslant 1$, which is a desired conclusion (if mr was a period of B, b would be equal as a_{mr}).

* The fact that t, the basic period of B, is a period of A' which is not a multiple of the basic period r implies that $t+r \geqslant k$:

suppose therefore that $t+r \leqslant k-1 \quad \Rightarrow \quad t \leqslant k-1-r$.
As all periods of B, t is also a period of A' which is not a multiple of r. Let s be the smallest integer which is a period of A', but s is not a multiple of r ($s \leqslant t \leqslant k-1-r$). We obtain a contradiction by proving that $s-r$ is a period of A'. Thus the procedure is to prove that $a_{s-r+1}\dots a_{k-1} = a_1 \dots a_{k-s-1+r}$
Since s is a period of A', $a_{s+1}a_{s+2}\dots a_{k-1} = a_1 \dots a_{k-s-1}$. Since $k-s-1 \geqslant r$, we can take prefixes of both sequences from last equation of a length r and obtain that $a_{s+1}\dots a_{s+r} = a_1 \dots a_r$. But r is a period, so $a_{s+1-r}\dots a_s = a_{s+1}\dots a_{s+r} = a_1 \dots a_r$.
Applying above equations we obtain $a_1 \dots a_{k-s-1+r} = a_1 \dots a_r a_{r+1} \dots a_{k-s-1+r} = a_1 \dots a_r a_1 \dots a_{k-s-1} = a_{s+1-r}\dots a_s a_{s+1} \dots a_{k-1}$, and thus $s-r$ is a period of A'.

* We have shown that $t \geqslant r+1$ and $t+r \geqslant k$. Adding those inequalities we obtain $t \geqslant \frac{k+1}{2}$. We will show next that we cannot have exactly $t = \frac{k+1}{2}$. For this equality to hold, we would need to have k odd, and both inequalities should be equations, thus $r = t-1 = \frac{k-1}{2}$, to that $A' = A^* A^*$, with $A^* = a_1 \dots a_r$. But since $t = r+1$ is also a period of A', we must have $a_2 \dots a_r = a_1 \dots a_{r-1}$, which implies that $a_1 = a_2 =$

$\ldots = a_r$, and so 1 is a period of A'. Therefore $r = 1$ (as r is the basic period of A') and $r = \frac{k-1}{2} \Rightarrow k = 3$. But we assumed that $t \leqslant k-2$, which in this case equals 1, which contradicts the fact that $t > r$. Thus we cannot have $t = \frac{k+1}{2}$, and so $t \geqslant \frac{k+2}{2}$.

We have proved that if the basic period t of B satisfies $t \leqslant k-2$, then $t \geqslant \frac{k+2}{2}$. We next wish to show that $t \geqslant \frac{k+2}{2}$ in all cases. If $t = k$ then $t \geqslant \frac{k+2}{2}$ for all $k \geqslant 2$, so what is left is the case $t = k-1$. But $k-1 \geqslant \frac{k+2}{2}$ for $k \geqslant 4$, so we only have to consider $k = 3$. Then $t = k-1 = 2$ and $1 \leqslant r < t = 2$, what implies $r = 1$, and so $A' = aa, B = baa$ for some $b \neq a$, ant then $t = 3$, a contradiction.

We obtain that $t \geqslant \frac{k+2}{2}$ for any basic period t of B. But t is integer, so we can write $t \geqslant \lfloor \frac{k+1}{2} \rfloor + 1$.

Ad. 2). Now we find upper or lower bounds on $A : B$, $B : B - B : A$ and $A : A$. The result from first step of the proof is crucial.

* First of all lets notice that if $A_{(m)} = B^{(m)}$, then $B_{(m-1)} = B^{(m-1)}$ or $m = 1$:
  $a_{k-m+1} \ldots a_{k-1} a_k = b a_1 \ldots a_{m-2} a_{m-1} \quad \Rightarrow \quad a_{k-m+1} \ldots a_{k-1} = b a_1 \ldots a_{m-2}$

  The basic period of B equals t, so the largest m such that $B_{(m)} = B^{(m)}$ equals $k-t$, thus the largest m such that $A_{(m)} = B^{(m)}$ equals $(k-t)+1$. Therefore

  $$A : B = \delta_1(A, B) + \ldots + \delta_{k-t+1}(A, B) + 0 + \ldots + 0 \leqslant q + \ldots + q^{k-t+1}.$$

  We have shown that $t \geqslant \lfloor \frac{k+1}{2} \rfloor + 1$, thus

  $$A : B \leqslant q + \ldots + q^{k - (\lfloor \frac{k+1}{2} \rfloor + 1) + 1} = q + \ldots + q^{k - \lfloor \frac{k+1}{2} \rfloor}.$$

  A few times we will use such simple equation

  $$k + 1 = \lfloor \frac{k+1}{2} \rfloor + \lceil \frac{k+1}{2} \rceil \quad \Rightarrow \quad k - \lfloor \frac{k+1}{2} \rfloor = \lceil \frac{k+1}{2} \rceil - 1. \tag{9}$$

  Therefore

  $$A : B \leqslant q + \ldots + q^{\lceil \frac{k+1}{2} \rceil - 1} = q \cdot \left(\frac{q^{\lceil \frac{k+1}{2} \rceil - 1} - 1}{q - 1} \right).$$

* Next since $B_{(m)} = B^{(m)}$ implies $B_{(m-1)} = A^{(m-1)}$ or $m = 1$, then $(B : B - B : A)$ equals $q^k - q^{k-1}$ plus a sum of terms of the form $q^m - q^{m-1}$, where $B_{(m)} = B^{(m)}, 1 < m < k$, minus possibly some q^{j-1}, where $B_{(j-1)} = A^{(j-1)}$, but $B_{(j)} \neq B^{(j)}$, plus possibly 1. Since the largest m, where $B_{(m)} = B^{(m)}$ for which $m < k$ satisfies $m \leqslant k - t \leqslant k - (\lfloor \frac{k+1}{2} \rfloor + 1) = \lceil \frac{k+1}{2} \rceil - 2$, we obtain

  $$(B : B - B : A) \leqslant q^k - q^{k-1} + q^{\lceil \frac{k+1}{2} \rceil - 2}.$$

* At last $A_{(k)} = A^{(k)}$, so obviously $A : A \geqslant q^k$.

Ad. 3). Finally we want to prove that

$$(A : A - A : B) \geqslant q^k - q\cdot(\frac{q^{\lceil \frac{k+1}{2} \rceil -1} - 1}{q-1}) \geqslant q^k - q^{k-1} + q^{\lceil \frac{k+1}{2} \rceil -2} \geqslant (B : B - B : A),$$

which completes the proof. First and third inequality is just application of (6), (7) and (8). The middle inequality is equivalent to

$$q^{k-1} \geqslant q \cdot (\frac{q^{\lceil \frac{k+1}{2} \rceil -1} - 1}{q-1}) + q^{\lceil \frac{k+1}{2} \rceil -2} \quad | \cdot q(q-1) \quad \Leftrightarrow$$

$$q^k(q-1) \geqslant q^{\lceil \frac{k+1}{2} \rceil -1}(q^2+q-1) - q^2 = q^{\lceil \frac{k+1}{2} \rceil -1}(q^2+q) - q^{\lceil \frac{k+1}{2} \rceil -1} - q^2$$

It suffices to prove that

$$q^k(q-1) \geqslant q^{\lceil \frac{k+1}{2} \rceil -1} q(q-1) \quad \Leftrightarrow$$

$$q^{\lfloor \frac{k+1}{2} \rfloor + \lceil \frac{k+1}{2} \rceil -1} \geqslant q^{\lceil \frac{k+1}{2} \rceil -1} q \quad \Leftrightarrow$$

$$q^{\lfloor \frac{k+1}{2} \rfloor} \geqslant q \quad \Leftrightarrow$$

$$k \geqslant 1.$$

□

Guibas and Odlyzko [5] have proved even more - that for any sequence A we know less or more which sequence B has the biggest chances of winning with A, it means for which B proportion $\frac{p_B}{p_A}$ is the largest (last theorem ensures that this proportion is > 1):

Theorem 9. *If $A = a_1 \dots a_k$ is a choice of Player I, then all the choices for Player II which maximize his probability of winning are of the form $B = ba_1 \dots a_{k-1}$, for a suitable b.*

Chen and Zame [6] have a different proof of the fact that for each sequence A there is a sequence B of the same length as A which is better than A. They divided sequences with respect to their form and found the form of better sequence separately for each group. Then they applied those results to the proof of next paradox:

Theorem 10. *For each integer $k \geqslant 4$ and each pair of patterns $A = (a_1 \dots a_k)$ and $B = (b_1 \dots b_k)$, there exists a finite sequence of patterns $\{C_1, C_2, \dots, C_m\}$ such that $A < C_1 < C_2 < \dots < C_m < B$ (where $A_1 < A_2$ means that $\frac{p_{A_2}}{p_{A_1}} > 1$).*

The simple corollary of this theorem is that the property of „being better" isn't transitive also in the games with sequences of any length:
we can take any two sequences A and B, for which the game is not fair, so one of them is better than the other one. There is no loss of generality in assuming $A > B$. However from Theorem 10 there is a cycle:

$$B < A < C_1 < C_2 < \dots < C_m < B$$

6 Conclusion

Throughout the whole paper we assumed that X is a random variable on Σ - a finite set of values of X. We can ask ourselves a question what happens if X is an arbitrary random variable, perhaps continuous. Does an analogous Conway's formula exist in such a case?

To summarize this paper let us answer the question why Penney's games can be interesting for us?

First, thanks to its simplicity and surprising paradoxes Penney's game can be an interesting and useful tool for promoting mathematics.
But we can also prove more general properties of the game. It is possible that this general theory will find some applications for example in mathematical modeling of gene mutations or game theory, especially on the stock.

Literatura

[1] W. Penney: *Problem: penney-ante*, Journal of Recreational Mathematics, 2 (1969), 241.

[2] W. Niemyska: *O pewnych zagadnieniach związanych z grami Penneya* (in Polish), Warsaw Uniwersity, Masters Thesis (2010)

[3] J. Jakubowski, R. Sztencel: *Wstęp do teorii prawdopodobieństwa* (in Polish), (2000), 273-274, 438-439.

[4] Shuo-Yen Robert Li: *A Martingale Approach to the Study of Occurence of Sequence Patterns In Repeated Experiments*, Annals of Probability, 8 (1980), 1171-1176

[5] Leonidas J. Guibas, Andrew M. Odlyzko: *String Overlaps, Pattern Matching, and Nontransitive Games*, Journal of Combinatorial Theory, ser. A, vol. 30 (1981), 183-208

[6] Robert Chen, Alan Zame: *On Fair Coin-Tossing Games*, Journal of Multivariate Analysis, 9 (1979), 150-156