On the first occurrence of strings

Robert W. Chen
Alan Zame

Dept of Mathematics
University of Miami

Burton Rosenberg

Dept of Computer Science
University of Miami

Submitted: Feb 9, 2008; Accepted: Feb 6, 2009; Published: Feb 27, 2009
Mathematics Subject Classification: 65C50

Abstract

We consider a game in which players select strings over $\{0, 1\}$ and observe a series of fair coin tosses, interpreted as a string over $\{0, 1\}$. The winner of this game is the player whose string appears first. For two players public knowledge of the opponent's string leads to an advantage. In this paper, results for three players are presented. It is shown that given the choices of the first two players, a third string can always be chosen with probability of winning greater than 1/3. It is also shown that two players can chose strings such that the third player's probability of winning is strictly less than the greater of the other two player's probability of winning, and that whichever string is chosen, it will always have a disadvantage to one of the two other strings.

1 Introduction

We consider a game in which players select strings over $W = \{0, 1\}$ and observe a series of fair coin tosses, that is, a string $\sigma = s_1 s_2 \ldots$ where each s_i is chosen independently at random from $\{0, 1\}$, with equal probability of a 0 or 1 being chosen. The winner of this game is the player whose string appears first. This problem has been studied both in the context of games and as a pure probabilistic problem in Chen [1], [2], [3], Guibas et al [4], Li [5], Gerber et al [6] and Mori [7].

In Chen [3] it was proved that for two players, public knowledge of the opponent's string leads to an advantage.

Theorem 1 *For any string $\sigma \in W^*$, $|\sigma| \geq 3$, there exists a string $\tau \in W^*$, of the same length as σ, such that $P(T_\sigma > T_\tau) > 1/2$. That is, the first occurrence of τ is likely to be before that of σ.*

In this paper we establish results for three players.

In is quite natural to suspect that under some reasonable conditions we might have a positive answer to the following conjecture: given $k-1$ strings, $\sigma_1, \sigma_2, \ldots, \sigma_{k-1}$ all of length n, there always exists a distinct string σ_k, also of length n, which has the best chance of occurring first among the strings $\sigma_1, \ldots, \sigma_k$. However, the answer is negative.

In section 3 we show that if the third player chooses having knowledge of the choices of the first two players, a string can be chosen so that the probability of this string showing first is greater than 1/3. In section 4 we show that although the third player can have a greater than average result, his situation is not the most advantageous. For the other two players can choose strings such that the third player's probability of winning is strictly less than the greater of the other two player's probability of winning, and that whichever string he chooses, he will always have a disadvantage to one of the two other players.

We begin with some preliminaries, remarking that lemma 3 of these preliminaries is a very interesting result in it's own right. Given a string, the waiting time for the first occurrence of the string in a random sequence of characters depends on the structure of repetitions in the string. Lemma 3 states that the second occurrence has waiting time independent of the string, except for its length.

Some of our proofs require exhaustive testing of cases. In section 5 we provide the computer codes by which these checks were accomplished.

2 Preliminaries

Let Σ be a finite set. The set of all finite strings over Σ is denoted Σ^*. A string $\sigma \in \Sigma^*$ of length n can be written as $\sigma = s_1 s_2 \ldots s_n$ with each $s_i \in \Sigma$. Given two strings $\sigma, \tau \in \Sigma^*$, their concatenation is denoted $\sigma\tau$. The length of string σ is denoted $|\sigma|$. The empty string ϵ is the unique zero length string. Given a string σ, its prefixes $\pi(\sigma)$ are all strings π such that $\sigma = \pi\tau$, for some string τ; its suffixes $\lambda(\sigma)$ are all strings λ such that $\sigma = \tau\lambda$ for some string τ.

Let $\{X_i\}$ be a sequence of Σ valued random variables. The probability space Ω is such that the X_i are i.i.d. with $P(X_i = s_j) = p_j$ for all i and j. The space Ω can be identified with the space of semi-infinite strings over Σ by $\sigma = s_1 s_2 \ldots$ with $s_i = X_i(\omega)$. We extend the definition of the prefix operation $\pi(\omega)$ to apply to semi-infinite $\omega \in \Omega$ under this identification.

For each string $\sigma \in \Sigma^*$, let T_σ be the waiting time for the first occurrence of σ in a randomly chosen $\omega \in \Omega$,

$$T_\sigma(\omega) = \min\{|\tau| \,|\, \tau \in \pi(\omega) \text{ and } \sigma \in \lambda(\tau)\},$$

or $T_\sigma(\omega) = \infty$ if σ never appears in ω. For strings $\tau, \sigma \in \Sigma^*$ let $T_{\sigma|\tau}$ be the time of first occurrence of σ after the occurrence of the string τ. If $\sigma \in \lambda(\tau)$ then $T_{\sigma|\tau}(\omega) = 0$, otherwise,

$$T_{\sigma|\tau}(\omega) = \min\{|\rho| - |\tau| \mid \rho \in \pi(\tau\omega) \text{ and } \sigma \in \lambda(\rho)\}$$

or $T_{\sigma|\tau}(\omega) = \infty$ if σ never appears in $\tau\omega$.

For strings $\sigma = s_1 s_2 \ldots s_n$ we define $P(\sigma) = \prod_{i=1}^{n} P(X_i = s_i)$, that is, the probability that a randomly chosen $\omega \in \Omega$ begins with σ. For strings $\sigma, \tau \in \Sigma^*$ define,

$$\sigma \circ \tau = \sum_{\substack{\rho \in \lambda(\sigma) \cap \pi(\tau) \\ \rho \neq \epsilon}} P(\rho)^{-1}$$

This operation has great significance in the calculation of waiting times for the first occurrence of strings.

Lemma 1 *Suppose $\Sigma = \{s_1, \ldots, s_n\}$, and $\{X_j\}$ are i.i.d. random variables with $P(X_j = s_k) = p_k$. For any $\sigma \in \Sigma^*$ and any $i = 1, \ldots, n$,*

$$\sum_{j=1}^{n} p_j (\sigma\, s_j \circ \sigma\, s_i) = 1 + \sigma \circ \sigma.$$

Proof: For each $\tau \in \pi(\sigma) \cap \lambda(\sigma)$, the term $P(\tau)^{-1}$ appears on the right hand side of the equality. For this τ, $\tau\, s_j \in \pi(\sigma\, s_i)$, for exactly one j, and it contributes $(p_j P(\tau))^{-1}$ to the sum $(\sigma\, s_j \circ \sigma\, s_i)$ on the left hand side of the equality. In addition, the unique single character string $s_j \in \pi(\sigma\, s_i)$ contributes the term $1/p_j$ to the sum $(\sigma\, s_j \circ \sigma\, s_i)$ on the left hand side of the equality. This has no corresponding term in the sum $\sigma \circ \sigma$, but is balanced by the constant 1 on the right hand side of the equality.

Lemma 2 *Hypotheses as above, for any $\sigma \in \Sigma^*$, $E(T_\sigma) = \sigma \circ \sigma$; for any $\sigma, \tau \in \Sigma^*$, $E(T_{\sigma|\tau}) = \sigma \circ \sigma - \tau \circ \sigma$.*

Proof: It is sufficient to prove the case of conditional waiting times, since $T_\sigma = T_{\sigma|\epsilon}$ and $\epsilon \circ \sigma = 0$.

The proof is by induction. The result follows from the definitions if σ, τ are the empty strings. Assume the result is true for all strings of length N or less.

Let σ' be string of length $N+1$ and τ' a string of length not more than $N+1$. Without loss of generality we can assume $\sigma' = \sigma\, s_1$. If $\tau' = \sigma'$ then $T_{\sigma'|\tau'} = T_{\sigma'|\sigma'} = 0$ and the result is trivial. Else if $|\tau'| = N+1$, we can write $\tau' = s_i\, \tau$ for some i, and noting $T_{\sigma'|\tau'} = T_{\sigma'|\tau}$ and $\tau' \circ \sigma' = \tau \circ \sigma'$, reduce to the case of $|\tau| \leq N$.

The expected waiting time for $\sigma\, s_1$ given τ is described recursively as the expected waiting time for σ given τ followed by the reception of one character, call it s_j, followed by the probability weighted sum of expected waiting times for $\sigma\, s_1$ given $\sigma\, s_j$ for each of the possible j, except if $s_j = s_1$,

$$\begin{aligned} E(T_{\sigma\, s_1|\tau}) &= E(T_{\sigma|\tau}) + 1 + \sum_{j=2}^{n} p_j E(T_{\sigma\, s_1|\sigma\, s_j}) \\ &= \sigma \circ \sigma - \tau \circ \sigma + 1 + S, \end{aligned}$$

where we have used the induction hypothesis and have let S stands for the summation. To evaluate the sum S, define strings σ_j by $\sigma\, s_j = s_i\, \sigma_j$, where s_i is the initial character of σ. Note that $E(T_{\sigma\, s_1 | \sigma\, s_j}) = E(T_{\sigma\, s_1 | \sigma_j})$ for $j \neq 1$. For $j = 2, \ldots, n$,

$$\begin{aligned} E(T_{\sigma\, s_1 | \sigma_j}) &= \sigma \circ \sigma - \sigma_j \circ \sigma + 1 + S \\ &= \sigma \circ \sigma - (s_i\, \sigma_j \circ \sigma\, s_1) + 1 + S. \\ &= \sigma \circ \sigma - (\sigma\, s_j \circ \sigma\, s_1) + 1 + S. \end{aligned}$$

Multiply each of these equations by p_j and sum over j from 2 to n,

$$\begin{aligned} S &= (1 - p_1)(\sigma \circ \sigma + 1 + S) - \sum_{j=2}^{n} p_j (\sigma\, s_j \circ \sigma\, s_1) \\ &= (1 - p_1)(\sigma \circ \sigma + 1 + S) - \sum_{j=1}^{n} p_j (\sigma\, s_j \circ \sigma\, s_1) + p_1 (\sigma\, s_1 \circ \sigma\, s_1) \\ &= (1 - p_1)(\sigma \circ \sigma + 1 + S) - (1 + \sigma \circ \sigma) + p_1 (\sigma\, s_1 \circ \sigma\, s_1) \\ &= (1 - p_1) S + p_1 (\sigma\, s_1 \circ \sigma\, s_1 - \sigma \circ \sigma - 1) \end{aligned}$$

using the previous lemma to reduce the sum, Therefore $S = \sigma\, s_1 \circ \sigma\, s_1 - \sigma \circ \sigma - 1$. Substituting,

$$\begin{aligned} E(T_{\sigma\, s_1 | \tau}) &= \sigma \circ \sigma - \tau \circ \sigma + 1 + \sigma\, s_1 \circ \sigma\, s_1 - \sigma \circ \sigma - 1 \\ &= \sigma\, s_1 \circ \sigma\, s_1 - \tau \circ \sigma \\ &= \sigma\, s_1 \circ \sigma\, s_1 - \tau \circ \sigma\, s_1, \end{aligned}$$

completing the induction.

The above lemma was proved in Chen [3] using the Renewal Theorem. The above proof is new. Note that the lemma also hold in the case of a countably infinite Σ provided $P(s) > 0$ for all $s \in \Sigma$.

Although the first occurrence of a string has a dependency on the repetition structure inside the string, an easy consequence of the previous lemma is that the following occurrences do not. This can also be derived by considering stopping times of an appropriate Markov chain, see for instance Levin et. al [8].

Lemma 3 *For $\sigma \in \Sigma^*$, define $T'_\sigma(\omega)$ to be the additional time to for the next occurrence of σ after its first occurrence in $\omega \in \Omega$. Then $E(T'_\sigma) = P(\sigma)^{-1}$.*

Proof: Since $T'_\sigma = T_{\sigma | \sigma'}$, where $\sigma = s\, \sigma'$ for the appropriate $s \in \Sigma$, we need to calculate $\sigma \circ \sigma - \sigma' \circ \sigma$. Note that all terms cancel except for the leading term $P(\sigma)^{-1}$.

Extend the prefix operator π to sets of strings S by $\pi(S) = \cup_{\sigma \in S} \pi(\sigma)$. A set of strings $\sigma_1, \sigma_2, \ldots, \sigma_k \in \Sigma^*$ is said to be *reduced* if no σ_i is a substring of σ_j, that is, $\sigma_i \notin \pi(\lambda(\sigma_j))$ for all distinct i, j . Define $N_k = \min(T_{\sigma_1}, T_{\sigma_2}, \ldots, T_{\sigma_k})$. If the set $\sigma_1, \sigma_2, \ldots, \sigma_k$ is reduced, and N_k is finite, there will be a unique i such that $N_k = T_{\sigma_i}$.

Lemma 4 *Hypotheses and notation as above, for each $i = 1, 2, \ldots k$,*

$$E(T_{\sigma_i}) = E(N_k) + \sum_{j=1}^{k} P(N_k = T_{\sigma_j}) E(T_{\sigma_i|\sigma_j}).$$

PROOF: For $i = 1, 2, \ldots, k$,

$$\begin{aligned} E(T_{\sigma_i}) &= E(N_k) + E(T_{\sigma_i} - N_k) \\ &= E(N_k) + E\big(E(T_{\sigma_i} - N_k \mid N_k = T_{\sigma_j})\big) \\ &= E(N_k) + \sum_{j=1}^{k} E(T_{\sigma_i} - N_k \mid N_k = T_{\sigma_j}) P(N_k = T_{\sigma_j}) \end{aligned}$$

Because the set of strings is reduced, the distribution of $T_{\sigma_i} - N_k$ conditioned on $N_k = T_{\sigma_j}$ is the same as that of $T_{\sigma_i|\sigma_j}$ and therefore $E(T_{\sigma_i} - N_k \mid N_k = T_{\sigma_j}) = E(T_{\sigma_i|\sigma_j})$. The result follows.

Lemma 5 *Hypotheses and notation as above. We have the following system of $k+1$ linear equations, where $q_i = P(T_{\sigma_i} = N_k)$, for $i = 1, 2, \ldots, k$,*

$$\begin{pmatrix} 0 & 1 & \ldots & 1 \\ 1 & & & \\ \vdots & (\sigma_i \circ \sigma_i & & \\ & & -\sigma_j \circ \sigma_i)_{i+1,j+1} & \\ 1 & & & \end{pmatrix} \begin{pmatrix} E(N_k) \\ q_1 \\ \vdots \\ q_k \end{pmatrix} = \begin{pmatrix} 1 \\ \sigma_1 \circ \sigma_1 \\ \vdots \\ \sigma_k \circ \sigma_k \end{pmatrix}$$

PROOF: Combine the previous two lemmas and the fact that $q_1 + q_2 + \ldots + q_k = 1$.

In the case of two strings, σ_1, σ_2, such that neither is a substring of the other, we provide for reference the solution to this matrix equation,

$$\begin{aligned} E(N_2) &= ((\sigma_1 \circ \sigma_1)(\sigma_2 \circ \sigma_2) - (\sigma_1 \circ \sigma_2)(\sigma_2 \circ \sigma_1))\Delta^{-1}, \\ q_1 &= (\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1)\Delta^{-1}, \\ q_2 &= (\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2)\Delta^{-1}, \end{aligned}$$

where $\Delta = \sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 - \sigma_2 \circ \sigma_1 + \sigma_2 \circ \sigma_2$. Therefore of two strings σ_1 and σ_2, σ_1 is strictly favorable to appear first exactly if $\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 > \sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2$.

3 Advantage of third player

In this section we establish as result for three players, where the third player choses having knowledge of the choices of the first two players. Given any two strings σ_1 and σ_2, both of length n, we exhibit a string τ, also of length n, such that the probability in a random series of coin tosses that τ appears first among the three is greater than $1/3$.

Theorem 2 (Main Theorem) *Let $n \geq 4$ and $\sigma_1, \sigma_2 \in \{0,1\}^*$ be any two distinct strings, both of length n. There exists a string τ distinct from σ_1 and σ_2 such that $P(T_\tau = N_3) > 1/3$.*

The proof constructs the string τ. There are two different constructions, depending on the form of σ_1 and σ_2. Throughout this section, and without loss of generality, we will assume $\sigma_1 \neq \sigma_2$ and,

$$\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 \leq \sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2.$$

We adopt a notation for the complement of a bit, $\bar{c} = c - 1$, for $c \in \{0,1\}$.

For a positive integer n, and two distinct strings $\sigma_1, \sigma_2 \in \{0,1\}^*$ both of length n, define,

$$L_n(\sigma_1, \sigma_2) = \max\{|\tau| \,|\, \tau \in \lambda(\sigma_1) \cap \pi(\sigma_2)\}.$$

For a single string, define,

$$L_n(\sigma) = \max\{|\tau| \,|\, \tau \in \lambda(\sigma) \cap \pi(\sigma) \setminus \{\sigma\}\}.$$

Note that in these definitions, the empty string is a possibility, so that $L_n \geq 0$.

For a string σ of length n let $l_n(\sigma) = n - L_n(\sigma)$. This is the number of characters dropped from the front of σ in the first non-trivial overlap of σ with itself. Similarly, for strings σ and σ' both of length n define $l_n(\sigma, \sigma') = n - L_n(\sigma, \sigma')$.

One construction takes care of the case that σ_2 is one of these four strings,

$$[0]^*, [0]^*1, [1]^*, [1]^*0,$$

where, for notational convenience, we write a repeating string such as $\sigma'\sigma' \ldots \sigma'$ as $[\sigma']^*$. Write $\sigma_2 = c_1\tau' c_2$ where $c_1, c_2 \in \{0,1\}$, and $\tau' \in \{0,1\}^*$. The winning string is then $\tau = \bar{c_1} c_1 \tau'$.

Else we construct the winning string $\beta_n(\sigma_1, \sigma_2)$, as follows. Write $\sigma_1 = \tau_1 c_1 \tau_2$ and $\sigma_2 = \tau_3 c_2$, where $c_1, c_2 \in \{0,1\}$ and $\tau_1, \tau_2, \tau_3 \in \{0,1\}^*$ and $|\tau_2| = L_n(\sigma_1, \sigma_2)$. Then $\beta_n(\sigma_1, \sigma_2) = \bar{c_1}\tau_3$.

Lemma 6 *Strings σ_1, σ_2 as above, $\beta_n(\sigma_1, \sigma_2)$ is distinct from σ_1.*

Proof: Recall that $\sigma_1 = \tau_1 c_1 \tau_2$ and $\sigma_2 = \tau_3 c_2$. If $|\tau_2| = |\tau_3|$ then $\sigma_1 = c_1\tau_2$ which is obviously not equal to $\beta_n(\sigma_1, \sigma_2) = \bar{c_1}\tau_3$. Else, by choice of τ_2, it must be that τ_3 is not a suffix of σ_1, and therefore σ_1 is not equal to $c\tau_3$ for any c.

Lemma 7 *Strings σ_1, σ_2 as above and $\tau = \beta_n(\sigma_1, \sigma_2)$, $L_n(\sigma_1, \tau) \leq L_n(\sigma_1, \sigma_2)$.*

Proof: Recalling again the construction of $\tau = \bar{c_1}\tau_3$, a prefix of τ overlapping a suffix of $\sigma_1 = \tau_1 c_1 \tau_2$ cannot match c_1 against $\bar{c_1}$, nor can $\bar{c_1}$ match against something in τ_1, as τ_2 is the maximum length suffix of σ_1 matching against a prefix of $\sigma_2 = \tau_3 c_2$.

Lemma 8 *Strings σ_1, σ_2 as above and $\tau = \beta_n(\sigma_1, \sigma_2)$, $\sigma_1 \circ \tau \leq \sigma_1 \circ \sigma_2$.*

Proof: Note $\sigma_1 \circ \sigma_2 = \tau_2 \circ \tau_2$. By the previous lemma, $\sigma_1 \circ \tau = \tau' \circ \tau'$ where τ' is a suffix of τ_2, and therefore $\tau' \circ \tau' \leq \tau_2 \circ \tau_2$.

Lemma 9 *Let $\sigma_1, \sigma_2, \sigma_3 \in \{0,1\}^*$ be three distinct strings, all of length $n \geq 6$. Suppose that $\sigma_1 \circ \sigma_3 \leq \sigma_1 \circ \sigma_2$ and that $(\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1) \leq (\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2)$. Let $p_i = P(T_{\sigma_i} = N_3)$ be the probability that σ_i appears first among the three. If either,*

$$\left(1 + \frac{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) \left(\frac{\sigma_3 \circ \sigma_3 - \sigma_3 \circ \sigma_2}{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_3}\right) + \left(\frac{\sigma_3 \circ \sigma_2 - \sigma_3 \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) < 2,$$

or,

$$2\left(\frac{\sigma_3 \circ \sigma_3 - \sigma_3 \circ \sigma_2}{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_3}\right) + \left(\frac{\sigma_3 \circ \sigma_2 - \sigma_3 \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) < 2,$$

then $p_3 > 1/3$.

Proof: By Lemma 5, there is this system of equations,

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & \sigma_1 \circ \sigma_1 - \sigma_2 \circ \sigma_1 & \sigma_1 \circ \sigma_1 - \sigma_3 \circ \sigma_1 \\ 1 & \sigma_2 \circ \sigma_2 - \sigma_1 \circ \sigma_2 & 0 & \sigma_2 \circ \sigma_2 - \sigma_3 \circ \sigma_2 \\ 1 & \sigma_3 \circ \sigma_3 - \sigma_1 \circ \sigma_3 & \sigma_3 \circ \sigma_3 - \sigma_2 \circ \sigma_3 & 0 \end{pmatrix} \begin{pmatrix} e \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 1 \\ \sigma_1 \circ \sigma_1 \\ \sigma_2 \circ \sigma_2 \\ \sigma_3 \circ \sigma_3 \end{pmatrix}$$

where $e = E(N_3)$. From the two middle rows,

$$p_1(\sigma_2 \circ \sigma_2 - \sigma_1 \circ \sigma_2) - p_2(\sigma_1 \circ \sigma_1 - \sigma_2 \circ \sigma_1) + p_3(\sigma_2 \circ \sigma_2 - \sigma_3 \circ \sigma_2 - \sigma_1 \circ \sigma_1 + \sigma_3 \circ \sigma_1) = \sigma_1 \circ \sigma_1 - \sigma_2 \circ \sigma_2.$$

Since $p_1 = 1 - p_2 - p_3$ and $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 > 0$ this simplifies to,

$$1 = \left(1 + \frac{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) p_2 + \left(1 + \frac{\sigma_3 \circ \sigma_2 - \sigma_3 \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) p_3. \qquad (1)$$

From the third and fourth row of the matrix equality,

$$(\sigma_2 \circ \sigma_2 - \sigma_1 \circ \sigma_2 + \sigma_1 \circ \sigma_3 - \sigma_3 \circ \sigma_3)p_1 + (\sigma_2 \circ \sigma_3 - \sigma_3 \circ \sigma_3)p_2 + (\sigma_2 \circ \sigma_2 - \sigma_3 \circ \sigma_2)p_3 = \sigma_2 \circ \sigma_2 - \sigma_3 \circ \sigma_3.$$

Using that $p_1 = 1 - p_2 - p_3$, this implies,

$$(\sigma_3 \circ \sigma_3 - \sigma_3 \circ \sigma_2)p_3 - (\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_3)p_2 = (\sigma_1 \circ \sigma_2 - \sigma_1 \circ \sigma_3)p_1$$

Since we assumed $\sigma_1 \circ \sigma_3 \leq \sigma_1 \circ \sigma_3$, this value is non-negative, hence,

$$p_2 \leq \left(\frac{\sigma_3 \circ \sigma_3 - \sigma_3 \circ \sigma_2}{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_3}\right) p_3.$$

Combining this with equation (1):

$$1 \leq \left(\left(1 + \frac{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right)\left(\frac{\sigma_3 \circ \sigma_3 - \sigma_3 \circ \sigma_2}{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_3}\right) + 1 + \frac{\sigma_3 \circ \sigma_2 - \sigma_3 \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) p_3.$$

The assumptions of the lemma serve to bound the large expression within parenthesis by 3, hence the result.

Lemma 10 *Let σ_1, σ_2 be distinct strings in $\{0,1\}^*$, both of length $n \geq 6$, and $\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 \leq \sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2$, and $l_n(\tau') \geq 4$ where $\sigma_2 = \tau' c'$, $c' \in \{0,1\}$. Let $\tau = \beta_n(\sigma_1, \sigma_2) = c\tau'$. Then $P(T_\tau = N_3) > 1/3$.*

Proof: Since $l_n(\tau') \geq 4$, $\tau \neq \sigma_2$. By Lemma 8, $\sigma_1 \circ \tau \leq \sigma_1 \circ \sigma_2$, and recall that we have assumed $\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 \leq \sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2$. Therefore it is sufficient to show,

$$2\left(\frac{\tau \circ \tau - \tau \circ \sigma_2}{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \tau}\right) + \left(\frac{\tau \circ \sigma_2 - \tau \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) < 2$$

and invoke lemma 9. The inequality is a straightforward consequence of the following four inequalities,

$$\begin{aligned} \tau \circ \tau - \tau \circ \sigma_2 &\leq 2^{n-1} + 2^{n-4}, && (2)\\ \tau \circ \sigma_2 &\leq 2^{n-1} + 2^{n-4}, && (3)\\ 2^n - 2^{n-4} - 2^{n-5} &\leq \sigma_2 \circ \sigma_2 - \sigma_2 \circ \tau, && (4)\\ 2^n - 2^{n-2} &\leq \sigma_1 \circ \sigma_1 - \sigma_1 \sigma_2. && (5) \end{aligned}$$

We verify these inequalities directly for $n = 6$, and therefore assume $n \geq 7$.

Since $l_n(\tau') \geq 4$ (meaning that the first non-trivial overlap of τ' with itself drops at least four characters from the string) we have $\tau \circ \tau < 2^n + 2^{n+3}$ and $\sigma_2 \circ \tau < 2^{n-2}$. Suppose $\tau \circ \tau \geq 2^n + 2^{n-4}$. Then c is the fourth character of σ_2, the fifth character of σ_2 equals the first character of σ_2, the sixth character equals the second, and so forth. Therefore $\tau \circ \sigma_2 \geq 2^{n-1} + 2^{n-5}$ and so inequality 2 holds. Suppose $\tau \circ \tau \leq 2^n + 2^{n-4}$. Then since $\tau \circ \sigma_2 \geq 2^{n-1}$ inequality 2 holds.

Noticing that $\tau \circ \sigma_2 = \tau' \circ \tau'$, we conclude that inequality 3 holds.

Note also that for $2 \leq k \leq n-2$, if 2^k appears in $\sigma_2 \circ \tau$ then 2^{k-1} will appear in $\sigma_2 \circ \sigma_2$, and if 2^{n-3} appears in $\sigma_2 \circ \tau$ then 2^{n-4} will not appear in $\sigma_2 \circ \tau$ (since $l_n(\tau') \geq 4$). Therefore inequality 4 holds.

Suppose 5 does not hold. Since $\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 \leq \sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2$, then $\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 \leq 2^n - 2^{n-2}$ as well. Hence either 2^{n-1} or 2^{n-2} appears in $\sigma_1 \circ \sigma_2$, and either 2^{n-1} or 2^{n-2} appears in $\sigma_2 \circ \sigma_1$. If 2^{n-1} appears in either $\sigma_1 \circ \sigma_2$ or $\sigma_2 \circ \sigma_1$ we have a contradiction against the fact that $l_n(\tau') \geq 4$. If 2^{n-2} appears in $\sigma_1 \circ \sigma_2$ and $\sigma_2 \circ \sigma_1$, then 2^{n-4} appears in $\sigma_1 \circ \sigma_1$ and neither 2^{n-3} and 2^{n-4} will appear in either $\sigma_1 \circ \sigma_2$ or $\sigma_2 \circ \sigma_1$. This implies that inequality 5 holds, giving a contradiction.

Those all the cited inequalities hold, and the lemma is proven.

Lemma 11 *With all the hypothesis of the previous lemma except that $l_n(\tau') = 3$, $P(T_\tau = N_3) > 1/3$.*

Proof: By direct computation, the lemma is true when $n = 6$, therefore assume $n \geq 7$. Since $l_n(\tau') = 3$, σ_2 is of the form $[\sigma']^* \sigma'' c$ where $\sigma' \in \{001, 010, 011, 100, 101, 110\}$, σ'' is any proper prefix of the σ', including the empty string, and $c \in \{0,1\}$. This gives 36 cases for possible σ_2.

Let $\tau = \bar{c_1}[\sigma']^*\sigma''$, for some $c_1 \in \{0,1\}$. We give the proof only for the four cases arising from $\sigma' = 001$, $\sigma'' = 00$ by having $c, \bar{c_1} \in \{0,1\}$. The other many cases are similar.

Consider the case when $c = \bar{c_1} = 0$, i.e. $\sigma_2 = [001]^*000$ and $\tau = 0[001]^*00$. Note that $\tau \neq \sigma_2$, and that,

$$\begin{aligned}
\sigma_2 \circ \sigma_2 &= \tau \circ \tau = 2^n + 6, \\
\sigma_2 \circ \tau &= 14, \\
\tau \circ \sigma_2 &= 2^{n-1} + 2^{n-4} + \ldots + 2^5 + 6.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&2\left(\frac{\tau \circ \tau - \tau \circ \sigma_2}{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \tau}\right) + \left(\frac{\tau \circ \sigma_2 - \tau \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) \\
&\quad \leq 2\left(\frac{2^{n-1} - 2^{n-4} - \ldots - 2^5}{2^n - 8}\right) + \left(\frac{\tau \circ \sigma_2}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2}\right) \\
&\quad \leq 1 + \left(\frac{2^{n-1} + 2^{n-4} + \ldots + 2^5 + 6}{2^n - \sigma_1 \circ \sigma_2}\right)
\end{aligned}$$

We show that the second term in the above inequality is strictly less than 1 so that we can invoke lemma 9. Suppose otherwise, $2^{n-1} + 2^{n-4} + \ldots + 6 \geq 2^n - \sigma_1 \circ \sigma_2$. Then $\sigma_1 \circ \sigma_2 > 2^{n-2} + 2^{n-3}$, and therefore $l_n(\sigma_1, \sigma_2) \leq 2$. If $l_n(\sigma_1, \sigma_2) = 1$ then in the construction of $\tau = \bar{c_1}\tau_3$ we have $\sigma_1 = c_1\tau_3$, and therefore $\sigma_1 = 1[001]^*00$. We then calculate that $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 < \sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1$, contradicting an hypothesis of our construction.

Suppose instead that $l_n(\sigma_1, \sigma_2) = 2$. Then $\sigma_1 = c'c_1\tau_3$ and $\sigma_2 = \tau_3 c'' c_2$, that is, $\sigma_1 = c'1[001]^*0$. If $c' = 0$ then $\sigma_1 \circ \sigma_2 < 2^{n-2} + 2^{n-3}$, and we have our contradiction. If $c' = 1$ then $\sigma_1 \circ \sigma_2 - \sigma_1 \circ \sigma_2 < \sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1$, contradicting an hypothesis of our construction.

Consider the case when $c = 1$ and $\bar{c_1} = 0$, i.e. $\sigma_2 = [001]^*001$ and $\tau = 0[001]^*00$. Note that $\tau \neq \sigma_2$ and,

$$\begin{aligned}
\sigma_2 \circ \sigma_2 &= 2^n + 2^{n-3} + \ldots + 2^3, \\
\sigma_2 \circ \tau &= 0, \\
\tau \circ \tau &= 2^n + 6, \\
\tau \circ \sigma_2 &= 2^{n-1} + 2^{n-4} + \ldots + 2^2 + 2.
\end{aligned}$$

Thus $2(\tau \circ \tau - \tau \circ \sigma_2)/(\sigma_2 \circ \sigma_2 - \sigma_2 \circ \tau) < 1$, and we need only show $(\tau \circ \sigma_2 - \tau \circ \sigma_1)/(\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2) \leq 1$. If inequality is not satisfied, then $l_n(\sigma_1, \sigma_2) \leq 2$, and the i-th letter in σ_1 is 1, for $i = l_n(\sigma_1, \sigma_2)$.

As in the previous case, we argue contradictions for $l_n(\sigma_1, \sigma_2) = 1$ and $l_n(\sigma_1, \sigma_2) = 2$ individually by considering possible values of σ_1.

Consider the case when $c = 0$ and $\bar{c_1} = 1$, i.e. $\sigma_2 = [001]^*000$ and $\tau = 1[001]^*00$. Note that $\tau \neq \sigma_2$ and,

$$\begin{aligned}
\sigma_2 \circ \sigma_2 &= 2^n + 6, \\
\sigma_2 \circ \tau &= 0, \\
\tau \circ \tau &= 2^n + 2^{n-3} + \ldots + 2^3, \\
\tau \circ \sigma_2 &= 2^{n-1} + 2^{n-4} + \ldots + 2^2 + 2.
\end{aligned}$$

By lemma 9 it is sufficient to show,

$$\frac{2^n + 2^{n-3} + \ldots + 2^3 - 4}{2^n + 6} + \frac{2^{n-1} + 2^{n-4} + \ldots + 2^2 + 2}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2} \quad < \quad 2.$$

If $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 \geq 2^n - 2^{n-2}$, then the above inequality is satisfied. On the other hand, if $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 < 2^n - 2^{n-2}$, then $\sigma_1 \circ \sigma_2 > 2^{n-2}$, since $\sigma_1 \circ \sigma_1 \geq 2^n$, and this implies $l_n(\sigma_1, \sigma_2) \leq 2$ and the i-th letter in σ_1 is 0, for $i = l_n(\sigma_1, \sigma_2)$.

As in the previous case, we argue contradictions for $l_n(\sigma_1, \sigma_2) = 1$ and $l_n(\sigma_1, \sigma_2) = 2$ individually by considering possible values of σ_1.

Finally, consider the case when $c = \bar{c_1} = 1$, i.e. $\sigma_2 = [001]^*001$ and $\tau = 1[001]^*00$. Note that $\tau \neq \sigma_2$ and,

$$\begin{aligned}
\sigma_2 \circ \sigma_2 &= \tau \circ \tau = 2^n + 2^{n-3} + \ldots + 2^3, \\
\sigma_2 \circ \tau &= 2^{n-2} + 2^{n-5} + \ldots + 2^4 + 2, \\
\tau \circ \sigma_2 &= 2n - 1 + 2^{n-2} + \ldots + 2^2 + 2.
\end{aligned}$$

By lemma 9 it is sufficient to show,

$$\frac{2^n + 2^{n-3} + \ldots + 2^3 - 4}{2^n - 2^{n-3} - \ldots - 2^3 - 2} + \frac{2^{n-1} + 2^{n-4} + \ldots + 2^2 + 2}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2} \quad < \quad 2.$$

The first term in the sum on the left hand side, above, is strictly less than 4/3, in any case. Hence it is sufficient to show that the second term on the left hand side is not more than 2/3. Assuming otherwise, we have $\sigma_1 \circ \sigma_2 > 2^{n-3}$, so $l_n(\sigma_1, \sigma_2) \leq 3$ and the i-th letter in σ_1 is 0, for $i = l_n(\sigma_1, \sigma_2)$. Given these facts, we have the contradiction $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 < \sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1$.

This completes consideration of all cases, and the proof of the lemma.

Lemma 12 *With all the hypothesis of the previous lemma except that* $l_n(\tau') = 2$, $P(T_\tau = N_3) > 1/3$.

Proof: By direct computation, we verify the lemma for $n = 6$, therefore assume $n \geq 7$. Since $l_n(\tau') = 2$, σ_2 is of the form $[\sigma']^*\sigma''c$ where $\sigma' \in \{01, 10\}$, σ'' is any proper prefix of σ', including the empty string, and $c \in \{0, 1\}$. That gives 8 different cases for possible σ_2.

Let $\tau = \bar{c_1}[\sigma']^*\sigma''$, for some $c_1 \in \{0, 1\}$. We give the proof only for the four cases arising from $\sigma' = 01$, $\sigma'' = 0$ and $c, \bar{c_1} \in \{0, 1\}$. The many other cases are similar.

Consider the case when $c = \bar{c_1} = 0$, i.e. $\sigma_2 = [01]^*00$ and $\tau = 0[01]^*0$. Then $\sigma_2 \neq \tau$ and,

$$\begin{aligned}\sigma_2 \circ \sigma_2 &= \tau \circ \tau = 2^n + 2,\\ \sigma_2 \circ \tau &= 6,\\ \tau \circ \sigma_2 &= 2^{n-1} + 2^{n-3} + \ldots + 2^3 + 2.\end{aligned}$$

By lemma 9 it is sufficient to show,

$$2\frac{\tau \circ \tau - \tau \circ \sigma_2}{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \tau} + \frac{\tau \circ \sigma_2 - \tau \circ \sigma_1}{\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2} < 2.$$

By the values determined, the first term of the sum on the left hand side can be shown to be strictly less than 1. It is therefore sufficient to show that the second term is less than 1.

In the case where 2^{n-1} appears in $\sigma_1 \circ \sigma_2$, it must be that $\sigma_1 = c'[01]^*$. If $c' = 0$ then $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 < \sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1$, which is a contradiction. If $c' = 1$ then 2^{n-2} and 2^{n-4} will appear in $\sigma_1 \circ \sigma_1$ and $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 > 2^n - 2^{n-1}$. Also 2^{n-2} will appear in $\sigma_1 \circ \tau$ and $\tau \circ \sigma_2 - \tau \circ \sigma_1 < 2^{n-1}$. This will make the second term less than 1, as required.

Next consider the case $2^{n-2} \leq \sigma_1 \circ \sigma_2 < 2^{n-1}$. Then $l_n(\sigma_1, \sigma_2) = 2$ and the second character of σ_1 is a 1. That is, $\sigma_1 = c'1[01]^*$. If $c' = 1$ then $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 < \sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1$, which is a contradiction. If $c' = 0$ then $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 = 2^n$, which makes the second term less than 1, as required.

If $\sigma_1 \circ \sigma_2 < 2^{n-2}$, then second term is also less than 1.

Consider the case when $c = 1$ and $\bar{c_1} = 0$, i.e. $\sigma_2 = [01]^*00$ and $\tau = 0[01]^*0$. Given that $\sigma_1 \neq \sigma_2$, that $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 \geq \sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1$, and the character at the $l_n(\sigma_1, \sigma_2)$ location of σ_1 is a 1, we deduce that the possible values for σ_1 are either $[01]^*1$ or $[1]^*$. In either case, it is possible to check that the second inequality of lemma 9 holds.

Consider the case when $c = 0$ and $\bar{c_1} = 1$, i.e. $\sigma_2 = [01]^*00$ and $\tau = 1[01]^*0$. Then,

$$\begin{aligned}\sigma_2 \circ \sigma_2 &= 2^n + 2,\\ \sigma_2 \circ \tau &= 0,\\ \tau \circ \tau &= 2^n + 2^{n-2} + \ldots + 2^4 + 2^2,\\ \tau \circ \sigma_2 &= 2^{n-1} + 2^{n-3} + \ldots + 2^3 + 2.\end{aligned}$$

To apply the first inequality of lemma 9, we first note that these values imply that,

$$\frac{\tau \circ \tau - \tau \circ \sigma_2}{\sigma_2 \circ \sigma_2 - \sigma_2 \circ \tau} < \frac{2}{3}.$$

After using this bound in the first inequality of lemma 9, multiplying through by $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2$, we have that it is sufficient for the lemma to establish that,

$$\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 + (3/2)(\tau \circ \sigma_2 - \tau \circ \sigma_2) \leq 2(\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2).$$

Notice that $(3/2)\tau \circ \sigma_2 = 2^n - 1$. So it is sufficient to show,

$$2^{n+1} + 1 - \sigma_2 \circ \sigma_1 - (3/2)\tau \circ \sigma_1 \quad \leq \quad 2(\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2). \tag{6}$$

If $\sigma_1 \circ \sigma_2 = 0$ or $\sigma_2 \circ \sigma_1 \leq 2$, then $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 \geq 2^n$ and also either $\sigma_2 \circ \sigma_1$ or $\tau \circ \sigma_1$ will be at least 2. Therefore we can assume for the remainder of the proof that $\sigma_1 \circ \sigma_2 > 0$ and $\sigma_2 \circ \sigma_1 > 2$.

Recall the notation $L_n(\sigma, \sigma')$, the number of characters in the maximum overlap of a suffix of σ with a prefix of σ', where the strings σ and σ' have common length n; and the notation $L_n(\sigma)$ for $L_n(\sigma, \sigma)$, disallowing for the trivial overlap of length n. If $L_n(\sigma_1) \geq L_n(\sigma_1, \sigma_2)$ then $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 \geq 2^n$ and inequality 6 is satisfied. We continue are arguments under the assumption that $L_n(\sigma_1) < L_n(\sigma_1, \sigma_2)$.

If $L_n(\sigma_2, \sigma_1) < L_n(\sigma_1, \sigma_2)$ since, $L_n(\sigma_2, \sigma_1) \geq 2$,

$$\begin{aligned}
\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 \quad &\geq \quad 2^2 - 2^{L_n(\sigma_2, \sigma_1)} \\
&> \quad 2^n + 2^{L_n(\sigma_1)} + 2^{L_n(\sigma_1) - 2} + \ldots - 2^{L_n(\sigma_1, \sigma_2)} - 2^{L_n(\sigma_1, \sigma_2) - 2} - \ldots \\
&\geq \quad \sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2
\end{aligned}$$

we have a contradiction.

If $L_n(\sigma_2, \sigma_1) \geq L_n(\sigma_1, \sigma_2) + 2$,

$$\begin{aligned}
2(\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2) \quad &\geq \quad 2^{n+1} - 2^{L_n(\sigma_1, \sigma_2) + 2} \\
&\geq \quad 2^{n+1} - 2^{L_n(\sigma_2, \sigma_1)} - (3/2)\, \tau \circ \sigma_1 \\
&> \quad 2^{n+1} - \sigma_2 \circ \sigma_1 - (3/2)\, \tau \circ \sigma_1
\end{aligned}$$

establishing inequality 6.

If $L_n(\sigma_2, \sigma_1) = L_n(\sigma_1, \sigma_2)$ and even, the $L_n(\sigma_1) = L_n(\sigma_2, \sigma_1) - 2$ and,

$$\begin{aligned}
\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 \quad &= \quad 2^n - 2^{L_n(\sigma_2, \sigma_1)}, \\
\sigma_2 \circ \sigma_1 \quad &= \quad 2^{L_n(\sigma_2, \sigma_1)} + 2, \\
\tau \circ \sigma_1 \quad &= \quad (2/3)\,(2^{L_n(\sigma_2, \sigma_1)} - 1),
\end{aligned}$$

establishing inequality 6.

If $L_n(\sigma_2, \sigma_1) = L_n(\sigma_1, \sigma_2)$ and odd, the $L_n(\sigma_1) = L_n(\sigma_2, \sigma_1) - 1$ and,

$$\begin{aligned}
\sigma_2 \circ \sigma_1 \quad &= \quad 2^{L_n(\sigma_2, \sigma_1)}, \\
\tau \circ \sigma_1 \quad &= \quad 2^{L_n(\sigma_2, \sigma_1) - 1} + 2^{L_n(\sigma_2, \sigma_1) - 3} + \ldots + 2^2, \\
\sigma_1 \circ \sigma_1 \quad &= \quad 2^n + 2^{L_n(\sigma_2, \sigma_1) - 1} + 2^{L_n(\sigma_2, \sigma_1) - 3} + \ldots + 2^2, \\
\sigma_1 \circ \sigma_2 \quad &= \quad 2^{L_n(\sigma_2, \sigma_1)} + 2^{L_n(\sigma_2, \sigma_1) - 2} + \ldots + 2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
2^{n+1} - \sigma_2 \circ \sigma_1 - (3/2)\, \tau \circ \sigma_1 \quad &= \quad 2^{n+1} - 2^{L_n(\sigma_2, \sigma_1) + 1} + 2 \\
&< \quad 2^{n+1} - 2^{L_n(\sigma_2, \sigma_1)} - 2^{L_n(\sigma_2, \sigma_1) - 1} \\
&< \quad 2(\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2),
\end{aligned}$$

establishing inequality 6.

If $L_n(\sigma_2, \sigma_1) = L_n(\sigma_1, \sigma_2) + 1$ and $L_n(\sigma_2, \sigma_1)$ is odd then,

$$L_n(\sigma_1) = L_n(\sigma_1, \sigma_2) - 1 = L_n(\sigma_2, \sigma_1) - 2$$

so

$$\begin{aligned} 2^{n+1} - \sigma_2 \circ \sigma_1 - (3/2)\, \tau \circ \sigma_1 &= 2^{n+1} - 2^{L_n(\sigma_2,\sigma_1)+1} + 2 \\ &< 2^{n+1} - 2^{L_n(\sigma_2,\sigma_1)-1} - 2^{L_n(\sigma_2,\sigma_1)-2} \\ &< 2(\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2), \end{aligned}$$

establishing inequality 6.

If $L_n(\sigma_2, \sigma_1) = L_n(\sigma_1, \sigma_2) + 1$ and $L_n(\sigma_2, \sigma_1)$ is even then,

$$L_n(\sigma_1) = L_n(\sigma_2, \sigma_1) - 1 = L_n(\sigma_1, \sigma_2),$$

contradicting the assumption $L_n(\sigma_1) < L_n(\sigma_1, \sigma_2)$.

Consider the case when $c = \bar{c_1} = 1$, i.e. $\sigma_2 = [01]^*01$ and $\tau = 1[01]^*0$. Since $\sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2 \geq \sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1$, and the $l_n(\sigma_1, \sigma_2)$ character of σ_1 is 0, the possible values of σ_1 are either $[0]^*$ or $[01]^*00$. For these two situations we compute p_3 directly, showing it is strictly greater than 1/3.

We have shown the lemma for the four cases under consideration. The many other cases can be shown in a similar manner.

Lemma 13 *With all the hypothesis of the previous lemma except that $l_n(\tau') = 1$. In which case, we use an alternative construction, for which $P(T_\tau = N_3) > 1/3$.*

PROOF: Since $l_n(\tau') = 1$, σ_2 is one of these four strings, $[0]^*, [0]^*1, [1]^*, [1]^*0$. Write $\sigma_2 = c_1 \tau' c_2$ where $c_1, c_2 \in \{0, 1\}$, and $\tau' \in \{0, 1\}^*$. The winning string is then $\tau = \bar{c_1} c_1 \tau'$. We give the proof for $\sigma_2 = [1]^*$ and $[1]^*0$, the other two cases are similar.

Suppose $\sigma_2 = [1]^*$. Then $\tau = 0[1]^*$ and, since $\sigma_2 \circ \sigma_2 - \sigma_2 \circ \sigma_1 \leq \sigma_1 \circ \sigma_1 - \sigma_1 \circ \sigma_2$, $\sigma_1 = [0]^*$ or $[1]^*0$. From this we can directly compute that $p_3 > 1/3$.

Suppose $\sigma_2 = [1]^*0$. Then $\tau = 0[1]^*$ and we can directly compute,

$$\begin{aligned} \sigma_2 \circ \sigma_2 &= \tau \circ \tau = 2^n, \\ \sigma_2 \circ \tau &= 2, \\ \tau \circ \sigma_2 &= 2^n - 2, \\ \sigma_1 \circ \tau - \sigma_1 \circ \sigma_2 &= \tau \circ \tau - \tau \circ \sigma_2 = 2, \\ \sigma_2 \circ \sigma_2 - \sigma_2 \circ \tau &= 2^n. \end{aligned}$$

Using these values in the equation,

$$\big((\sigma_1 \circ \tau - \sigma_1 \circ \sigma_2) + (\sigma_2 \circ \sigma_2 - \sigma_2 \circ \tau)\big) p_2 + \big((\sigma_1 \circ \tau - \sigma_1 \circ \sigma_2) - (\tau \circ \tau - \tau \circ \sigma_2)\big) p_3 = \sigma_1 \circ \tau - \sigma_1 \circ \sigma_2$$

we have that $p_2 = 2/2^n$.

Using the equation,

$$\big((\sigma_1\circ\sigma_1-\sigma_1\circ\sigma_2)+(\sigma_2\circ\sigma_2-\sigma_2\circ\sigma_1)\big)p_2+\big((\sigma_1\circ\sigma_1-\sigma_1\circ\sigma_2)+(\tau\circ\sigma_2-\tau\circ\sigma_1)\big)p_3 = \sigma_1\circ\sigma_1-\sigma_1\circ\sigma_2$$

it is sufficient for $p_3 > 1/3$ that,

$$\frac{\tau\circ\sigma_2-\tau\circ\sigma_1}{\sigma_1\circ\sigma_1-\sigma_1\circ\sigma_2} < 2-\frac{12}{2^n}.$$

This is verified by noticing that since $L_n(\sigma_1,\sigma_2)\geq 3$ then $\sigma_1\circ\sigma_1-\sigma_1\circ\sigma_2\geq 2^n-2^{n-2}$ and $\tau\circ\sigma_2-\tau\circ\sigma_1\leq 2^n$.

We can now complete the proof of the Main Theorem. For $n=4$ and $n=5$, the proof is demonstrated by direct computation. We provide Mathematica code which solves the matrix equation for the p_i. For $n\geq 6$, we use the above lemmas, remarking that we have exhausted all possible values of $l_n(\tau')$.

4 Advantage of coalition of two of three players

Although in a two-person game, it is possible for one player to react to the other in order to pick a favorable string, in a three-person game, two players can collude to attain an advantage.

Theorem 3 *For $n\geq 3$, let σ_1,σ_2 and σ_3 be three distinct strings of length n in $\{0,1\}^*$, where $\sigma_1=[1]^*0$, $\sigma_2=[0]^*1$ and σ_3 is arbitrary. Let $p_i=P(T_{\sigma_i}=N_3)$ be the probability that σ_i appears first among the three. Then $p_3<\max(p_1,p_2)$.*

Proof: The set of strings $\{\sigma_1,\sigma_2,\sigma_3\}$ is reduced. Note that $\sigma_1\circ\sigma_1=\sigma_2\circ\sigma_2=2^n$ and $\sigma_1\circ\sigma_2=\sigma_2\circ\sigma_1=2$. Putting these values into the system of linear equations of lemma 5,

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2^n-2 & 2^n-\sigma_3\circ\sigma_1 \\ 1 & 2^n-2 & 0 & 2^n-\sigma_3\circ\sigma_2 \\ 1 & \sigma_3\circ\sigma_3-\sigma_1\circ\sigma_3 & \sigma_3\circ\sigma_3-\sigma_2\circ\sigma_3 & 0 \end{pmatrix}\begin{pmatrix} E(N_3) \\ p_1 \\ p_2 \\ p_3\end{pmatrix}=\begin{pmatrix} 1 \\ 2^n \\ 2^n \\ \sigma_3\circ\sigma_3\end{pmatrix}$$

Without loss of generality, σ_3 begins with a 1. The argument proceeds by considering four cases.

Case 1. Suppose $\sigma_3=1^n$, a sequence of n ones. Note that $\sigma_3\circ\sigma_1=2^n-2$ and $\sigma_3\circ\sigma_2=0$. Subtracting the second row from the third, and using these values, $(2^n-2)p_1-(2^n-2)p_2+(2^n-2)p_3=0$. Therefore $p_1-p_2+p_3=0$ and $p_1+p_2+p_3=1$, implying $p_2=1/2$. Intuitively, σ_1 and σ_3 must be equally likely to occur first (or continue to formally solve this system of equations) hence $p_1=p_3=1/4$. Therefore $p_2>p_3$.

Case 2. Suppose $\sigma_3=1^i0^j$, a sequence of i ones followed by a sequence of j zeros, $1<i,j<n-1$, distinct from σ_1 and σ_2. Note that $\sigma_3\circ\sigma_1=0,\sigma_1\circ\sigma_3=2^{i+1},\sigma_2\circ\sigma_3=2$

and $\sigma_3 \circ \sigma_3 = 2^n$. Subtracting the second row from the fourth, and using these values, $(2^n - 2^{i+1})p_1 - 2^n p_3 = 0$. Therefore $p_1 > p_3$.

Case 3. Suppose $\sigma_3 = 1^i \tau 1^j$, a sequence of i ones, followed by the string τ, followed by a sequence of j ones, $0 < i, j < n-1$, where either $\tau = 0$ or $\tau = 0\tau' 0$ for any string $\tau' \in \{0,1\}^*$. Note that $\sigma_3 \circ \sigma_3 = 2^n + \alpha$, where $\alpha > 0$. Also $\sigma_3 \circ \sigma_2 = 0, \sigma_1 \circ \sigma_3 = 2^{i+1}$ and $\sigma_2 \circ \sigma_3 = 2$. Subtracting the third row from the fourth, and using these values, $(\alpha - 2^{i+1} + 2)p_1 + (2^n + \alpha - 2)p_2 - 2^n p_3 = \alpha$. Collecting terms in α on the right hand side, and using that $\alpha(1 - p_1 - p_2) > 0$, then $p_2 > p_3$.

Case 4. Suppose $\sigma_3 = 1^i 0 \tau 1 0^j$, a sequence of i ones and a zero, followed by the string τ, followed by a one and a sequence of j zeros, $0 < i, j < n-2$, where $\tau \in \{0,1\}^*$. Note that $\sigma_3 \circ \sigma_3 = 2^n + \alpha$, where $\alpha \geq 0$. Also, $\sigma_3 \circ \sigma_1 = 0, \sigma_1 \circ \sigma_3 = 2^{i+1}$ and $\sigma_2 \circ \sigma_3 = 2.$. Subtracting the second row from the fourth, and using these values, $(2^n + \alpha - 2^{i+1})p_1 + \alpha p_2 - 2^n p_3 = \alpha$. Collecting terms in α on the right hand side, and using that $\alpha(1 - p_1 - p_2) \geq 0$, we have $p_1 > p_3$.

Since any string starting with a 1 follows one of these cases, and strings starting with a 0 must follow similar results by symmetry, the theorem is proved.

Theorem 4 *For $n \geq 3$, let σ_1, σ_2 and σ_3 be three distinct strings of length n in $\{0,1\}^*$, where $\sigma_1 = 11\ldots10$, $\sigma_2 = 00\ldots01$ and σ_3 is arbitrary. Then either $P(T_{\sigma_3} < T_{\sigma_1}) < 1/2$ or $P(T_{\sigma_3} < T_{\sigma_2}) < 1/2$.*

Proof: The proof is the same as the previous theorem, and we will omit it.

5 Computer Codes

```
LeadNumList[s_, t_] :=
    Table[
        Take[s, -i] == Take[t, i],
        {i, Min[Length[s], Length[t]]}]

LeadNumAux[l_, p_] :=
    If[(Length[l] ==0),
        0,
        If[First[l],
            p + LeadNumAux[Rest[l], 2*p],
            LeadNumAux[Rest[l], 2*p]]]

LeadingNumber[s_,t_] := LeadNumAux[ LeadNumList[s,t], 2]

LeadNumMatrix[sl_] :=
    Table[ If[i==0,
```

```
              If[j==0, 0, 1],
              If[j==0, 1,
                 LeadingNumber[sl[[i]],sl[[i]]]
                    -LeadingNumber[sl[[j]],sl[[i]]]
              ]],
            {i,0,Length[sl]},{j,0,Length[sl]}]

LeadNumVector[sl_] :=
    Table[ If[i==0, 1,
             LeadingNumber[sl[[i]],sl[[i]]]
           ],
           {i,0,Length[sl]}]

SolveLeadNum[sl_] :=
    Inverse[LeadNumMatrix[sl]] . LeadNumVector[sl]
```

References

[1] Chen, R., *A circular property of the occurrence of sequence patterns in the fair coin-tossing process,* Adv. Appl. Prob., Vol. 21, 1989. pp. 938–940.

[2] Chen, R. and Lin, H. E., *On the fair coin-tossing processes,* J. Multivariate Anal., Vol. 15, 1984. pp. 222–227.

[3] Chen, R. and Zame, A, *On the fair coin-tossing games,* J. Multivariate Anal., Vol. 9, 1979. pp. 150–157.

[4] Guibas, L. and Odlyzko, A. M., *String overlaps, pattern matching, and nontransitive games,* J. Combin. Theory Ser. A., Vol 30., 1981. pp. 183–208.

[5] Li, S. Y. R, *A martingale approach to the study of occurrence of sequence patterns in repeated experiments,* Ann. Prob., Vol. 8, 1980. pp. 1171–1176.

[6] Gerber, H. V. and Li. S. Y. R., *The occurrence of sequence patterns in repeated experiments and hitting times in a Markov Chain,* Stoch. Processes and their Appli., Vol. 11, 1981. pp. 101–108.

[7] Mori, Tamas F., *On the waiting times till each of some given patterns occurs as a run,* Probab. Th. Rel. Fields, **87**, 1991. pp. 313–323.

[8] Levin, David A, Peres, Yuval, and Wilmer, Elizabeth A., Markov Chains and Mixing Times, American Mathematical Society, 2008.