Science of Computer Programming 85 (2014) 166–203

Contents lists available at SciVerse ScienceDirect

Science of Computer Programming

www.elsevier.com/locate/scico

# First-past-the-post games

Roland Backhouse

*School of Computer Science, University of Nottingham, Nottingham, NG8 1BB, England, United Kingdom*

## HIGHLIGHTS

- Introduces the abstract notion of a first-past-the-post game.
- Constructs (non-linear) equations characterising the plays in a Penney-Ante game.
- Shows how to calculate probabilities of winning, expected lengths and generating functions.
- Generalises the construction to incomplete plays.
- Compares implementation techniques.

ARTICLE INFO

*Article history:*
Received 21 December 2012
Received in revised form 7 May 2013
Accepted 17 July 2013
Available online 26 July 2013

*Keywords:*
Algorithmic problem solving
Penney-Ante
Regular language
Probabilistic game
Generating function

ABSTRACT

Informally, a first-past-the-post game is a (probabilistic) game where the winner is the person who predicts the event that occurs first among a set of events. Examples of first-past-the-post games include so-called block and hidden patterns and the Penney-Ante game invented by Walter Penney. We formalise the abstract notion of a first-past-the-post game, and the process of extending a probability distribution on symbols of an alphabet to the plays of a game. We establish a number of properties of such games, for example, the property that an incomplete first-past-the-post game is also a first-past-the-post game.

Penney-Ante games are multi-player games characterised by a collection of regular, prefix-free languages. Analysis of such games is facilitated by a collection of simultaneous (non-linear) equations in languages. Essentially, the equations are due to Guibas and Odlyzko. However, they did not formulate them as equations in languages but as equations in generating functions detailing lengths of words. For such games, we show how to use the equations in languages to calculate the probability of winning and how to calculate the expected length of a game for a given outcome. We also exploit the properties of first-past-the-post games to show how to calculate the probability of winning in the course of a play of the game. In this way, we avoid the construction of a deterministic finite-state machine or the use of generating functions, the two methods traditionally used for the task.

We observe that Aho and Corasick's generalisation of the Knuth–Morris–Pratt pattern-matching algorithm can be used to construct the deterministic finite-state machine that recognises the language underlying a Penney-Ante game. The two methods of calculating the probabilities and expected values, one based on the finite-state machine and the other based on the non-linear equations in languages, have been implemented and verified to yield the same results.



A first-past-the-post game is a game where the winner is the person who predicts the event that occurs first among a set of events. There is no limit to the examples that can be invented. One example, which was published in the Communications of the ACM just before the final version of this paper was prepared [21], is the following:

*E-mail address:* roland.backhouse@nottingham.ac.uk.

0167-6423/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.scico.2013.07.007

"Alice and Bob roll a single die repeatedly. Alice is waiting until all six of the die's faces appear at least once. Bob is waiting for some face (any face) to appear four times. The winner is the one who gets his or her wish first; for example, if the successive rolls are 2, 5, 4, 5, 3, 6, 6, 5, 1, then Alice wins, since all numbers have appeared, none more than three times. If the successive rolls instead happen to be 4, 6, 3, 6, 6, 1, 2, 2, 6, then Bob wins because he has seen four 6s and no 5 (yet)."

Different versions of this game are obtained by varying the numbers six and four in the specification; see Example 6.

Perhaps the most widely studied example of a first-past-the-post game is the game that is now called Penney-Ante, a game with pennies named after its inventor, Walter Penney [18]. In the game, each player chooses a coin sequence (a sequence of "head"s or "tail"s); a coin is then repeatedly tossed until one of the sequences occurs. The winner is the player who chose the sequence that occurs first.

Typically in all such games, the interest is in determining the probability of winning for each of the players and the expected length of the games (assuming a fair roll of the die or toss of the coin).

The two-player Penney-Ante game has attracted much interest because it is non-transitive [10]; the game is also used to demonstrate the use of generating functions in the calculation of probability distributions [12,11]. Our interest in the game began as a simple (for us) introductory exercise in probability generating functions. It has turned out to be an exercise in applying the calculational method to the analysis of the game in the general case of an arbitrary number of players, with the added value of new insights in the formalisation of complex probabilistic events.

Analysis of the Penney-Ante game is facilitated by a collection of simultaneous (non-linear) equations between languages. In the literature, either the equations are stated without proof [11] or the equations are not given explicitly but translated directly into generating functions detailing lengths of words [12]. The primary contribution of this paper is to record a derivation of the equations and the associated probability distributions in which naming of word length and the use of generating functions is avoided.

Our derivation has several novel features. We introduce the abstract notion of a first-past-the-post game, and we formalise the process of extending a probability distribution on symbols of an alphabet to the plays of such a game (Section 2). (Multi-player) Penney-Ante games and so-called block and hidden patterns [9] are shown to be instances of first-past-the-post games. Such games are characterised by a collection of regular, prefix-free languages. We derive a collection of simultaneous non-linear equations in these languages and use these to show how to calculate the probability of winning (Section 4).

The equations are essentially the basis for the equations in generating functions derived by Guibas and Odlyzko [12]. The formula we derive generalises a formula attributed to Conway [10] for the original two-player Penney-Ante game. Another instance is the formula due to Solov'ev [19] for the expected number of coin tosses until a given (contiguous) pattern appears. Like Guibas and Odlyzko [12], we also consider the generalisation of Penney-Ante games to an arbitrary number of players; see Section 4.2. Section 4.3 briefly discusses how to obtain equations for the expected length of the game that ends with a given outcome.

We show in Section 4.4 that the equations in languages do not have a unique solution. Only when an additional requirement is added to the equations is the uniqueness property satisfied. This is surprising because the additional requirement is not explicitly used in constructing the generating functions or calculating probabilities.

Our work focuses on formalising and understanding the event space that underlies Penney-Ante games rather than the calculation of probabilities of winning and expected lengths of games. We think it is important to do so because then so-called "paradoxes" associated with Penney-Ante games can be explained. First-past-the-post games are characterised by prefix-free languages. We prove several different properties of such languages. For motivational purposes, these properties are interspersed throughout the text. Section 1 introduces some elementary properties in advance of the formal definition of a first-past-the-post game in Section 2. Section 2.2 helps to relate our use of prefix-free languages with other accounts of the games, and Section 2.3 exploits their properties to prove the fundamental property that the remainder of a game after some (typically incomplete) play of the game is a first-past-the-post game. This is used later in Section 5 where we generalise our results to such situations. Section 5.1 generalises the construction of the equations in languages and Section 5.2 then shows how the generating functions introduced by Guibas and Odlyzko [12] are constructed from these equations; in so doing, we observe an error in the statement of the central theorem of their paper.

Section 6 discusses the practical implementation of the calculations. The standard technique for calculating the probability of winning is based on the construction of a deterministic finite-state machine that recognises the language underlying a Penney-Ante game. We observe that this construction can be carried out most efficiently using an adaptation of Aho and Corasick's [1] generalisation of the Knuth–Morris–Pratt [14] pattern-matching algorithm, rather than the standard textbook method. The two methods of calculating the probabilities and expected values, one based on Aho and Corasick's algorithm and the other based on the non-linear equations in languages, have been implemented by Ngoc Do [15] and verified to yield the same results.

Several intriguing challenges remain, which are discussed in Section 7.