Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 183 (2007) 805–811
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

O.R. Applications

Dynamic programming analysis of the TV game "Who wants to be a millionaire?"

Federico Perea *, Justo Puerto

University of Seville, Department of Statistics and Operations Research, Faculty of Mathematics, c/Tarfia sn, 41012 Seville, Spain

Received 10 February 2006; accepted 5 October 2006
Available online 15 December 2006

Abstract

This paper uses dynamic programming to investigate when contestants should use lifelines or when they should just stop answering in the TV quiz show 'Who wants to be a millionaire?'. It obtains the optimal strategies to maximize the expected reward and to maximize the probability of winning a given amount of money.

Keywords: Dynamic programming; Markov processes; Recreation

1. Introduction

'Who wants to be a millionaire?' is a successful television game show in many countries. One contestant addresses 15 multiple-choice questions in a row. In each step, a question and four possible answers are shown. After being shown the question, the contestant can decide to stop playing and keep the money accumulated up till then, and the game is over, or to answer the question. If they decide to stay in the game, they can use up to three lifelines to answer the question. Each lifeline may only be used once during a contestant's entire game. These lifelines are:

- Lifeline 1 or the *50:50 option*: two of the three incorrect answers are removed.
- Lifeline 2 or *phone a friend*: the contestants may speak to a friend or relative on the phone for 30 s to discuss the question.
- Lifeline 3 or *ask the audience*: the audience votes with their keypads on their choice of answer. The result of this poll is listed in percentages and shown to the contestant.

There are two stages ("guarantee points") where the money is banked and cannot be lost even if the candidate gives an incorrect response. Those questions are the 5th one and the 10th one.

The decision of when to stop playing or when to use the lifelines should be treated rationally, although contestants rarely seem to make such rational decisions. In this paper, we address the problem of when to stop playing and when to use the lifelines as a dynamic programming (DP for short) problem and the optimal strategies are identified. The probabilities of correctly answering are based on observation of the Spanish game and the

* Corresponding author. Tel.: +34 654111231.
E-mail addresses: perea@us.es (F. Perea), puerto@us.es (J. Puerto).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2006.10.041

empirical model. There have been some approaches to the mathematical analysis of the game using simplified versions and as an educational resource in classrooms, for instance Cochran (2001) and Rump (2001). For other examples on the use of DP to analyze other contests see Thomas (2003), who analyzes 'The weakest link' or Sniedovich (2005) and Smith (2007) for all sorts of board games. The interested reader is also referred to the analysis of the HI-LO game in Freeman (2001) and of Cricket in Clarke and Norman (2003) and the references there. Our formulation of 'Who wants to be a millionaire?' works for all existing tables of prices of the game. We give the results for the Spanish version in 2003, where the monetary values of the questions were 150, 300, 450, 900, 1800, 2100, 2700, 3600, 4500, 9000, 18 000, 36 000, 72 000, 144 000 and 300 000 Euros, respectively.

The rest of the paper is organized as follows: Section 2 shows the general mathematical model (states, feasible actions, rewards, transition function, probabilities of answering correctly and their estimation). We present in Section 3 the description of the two particular models to be studied in this paper, in which we want to maximize the expected reward and the probability of reaching and correctly answering a given question respectively. To finish, Section 4 presents some concluding remarks based on simulations of how to play in a dynamic way.

2. Basic ideas

In the game, the contestant makes a decision each time a question and four possible answers are shown. The planning horizon is finite, there are $N = 16$ stages, where the 16th stage stands for the situation after answering question number 15 correctly. To make a decision, contestants have to know the index of the question they are facing and the lifelines they have already used. The history of the game is summarized in this information. Let $\mathscr{S}$ be the set of state vectors, whose elements are of the form $s = (k, l_1, l_2, l_3)$. Variable k is the index of the current question and

$$l_i = \begin{cases} 1 & \text{if lifeline } i \text{ may be used,} \\ 0 & \text{if lifeline } i \text{ was already used.} \end{cases} \tag{1}$$

Let $\mathscr{A}(s)$ denote the set of feasible actions in state s. $\mathscr{A}(s)$ depends on the question index and the lifelines left. If $k = 16$, the game is over and there are no feasible actions. If $k \leqslant 15$, the contestant has several possibilities:

- To answer the question without using lifelines.
- To answer the question employing one or more lifelines. In this case, contestants must also specify the lifelines they are going to use, remembering that each lifeline may only be used once during the whole game.
- To stop and quit the game.

If the player decides to stop, the immediate reward is the monetary value of the last question answered. If the candidate decides to answer, the immediate reward is a random variable and depends on the probability of answering correctly. If the candidate fails, the immediate reward is the last guarantee point reached before failing. If the candidate chooses the correct answer, there is no immediate reward and he or she goes on to the next question, and the reward is the expected (final) reward from the resulting state.

Denote r_k the immediate reward if the candidate decides to quit the game after answering question k correctly, and denote r_k^* the immediate reward if the candidate fails in question $k+1$. The values of r_k and r_k^* are shown in Table 1.

After a decision is made, the process proceeds to a new state.

- If the contestant decides to stop at a question or answers it incorrectly, the game is over.
- If the contestant decides to play and chooses the correct answer, there is a transition to another

Table 1
Immediate versus assured rewards

r_0	0	r_0^*	0
r_1	150	r_1^*	0
r_2	300	r_2^*	0
r_3	450	r_3^*	0
r_4	900	r_4^*	0
r_5	1800	r_5^*	1800
r_6	2100	r_6^*	1800
r_7	2700	r_7^*	1800
r_8	3600	r_8^*	1800
r_9	4500	r_9^*	1800
r_{10}	9000	r_{10}^*	9000
r_{11}	18 000	r_{11}^*	9000
r_{12}	36 000	r_{12}^*	9000
r_{13}	72 000	r_{13}^*	9000
r_{14}	144 000	r_{14}^*	9000
r_{15}	300 000	r_{15}^*	300 000