

Problem 1. Assume that you play a chess match with a friend. If you play timid your probability of making a draw is $p = 0.9$, the probability to win is 0 and the probability to lose is 0.1. If you play bold you either win with probability $q = 0.45$, or you lose. Each win brings one point to the score of the winner. The match consists of 5 games. If the score is a tie after the fifth game, then a “sudden death” rule is adopted; that is, whoever wins the next game is a winner of the match; if it is a draw, then the game is repeated with the same rule.

Formulate a Markov decision problem to determine the optimal strategy of your play (to maximize the probability of winning the match) and solve it. Clearly describe the state space, control space, transition probabilities, and the reward function.

Time $t=1,2,\dots,N=6$

State space $X = \{0; 0.5; 1; 1.5; 2; 2.5; W\}$ - x is current points . If $x \geq 3$ we mark this state as W =”we won” . $x_1 = 0$, we have 0 points at the beginning

Control space $U = \{1; 2\}$, 1=”play timid”, 2=”play bold”

Transition probabilities

$$P_t(y|x_t, 1) = \begin{cases} 0.9 & , y = x + 0.5 \\ 0.1 & , y = x \\ 0 & , \text{for else } y \end{cases} \quad x, y \in X \quad P_t(W|W, 1) = 1$$

, i.e. if $x < 3$, $u=1$, then $P(y=x)=0.9$, $P(y=x+0.5)=0.1$

$$P_t(y|x_t, 2) = \begin{cases} 0.45 & , y = x + 1 \\ 0.55 & , y = x \\ 0 & , \text{for else } y \end{cases} \quad x, y \in X \quad P_t(W|W, 2) = 1$$

, i.e. if $x < 3$, $u=2$, then $P(y=x)=0.55$, $P(y=x+1)=0.45$

Reward function $c_t(x_t) = 0$ if $t < 6$, if $x_6 < 2.5$, $c_6(x_6) = 0$ (lose match), $c_6(W) = 1$. If $x_6 = 2.5$, $c_6(2.5) = 0.45$ - probability to win 6th game with $u=2$ (if $u=1$, we can't win)

We use columns $V_t \in R^7$

$$V_7 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.45 \\ 1 \end{pmatrix}, \quad V_t = \begin{pmatrix} v_t(0) \\ v_t(0.5) \\ v_t(1) \\ v_t(1.5) \\ v_t(2) \\ v_t(2.5) \\ v_t(W) \end{pmatrix}$$

and Markov Matrices

$$U^{(1)} = \begin{pmatrix} 0.1 & 0.9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0.9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad U^{(2)} = \begin{pmatrix} 0.55 & 0 & 0.45 & 0 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0.45 & 0 & 0 & 0 \\ 0 & 0 & 0.55 & 0 & 0.45 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 & 0.45 & 0 \\ 0 & 0 & 0 & 0 & 0.55 & 0 & 0.45 \\ 0 & 0 & 0 & 0 & 0 & 0.55 & 0.45 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$v_t^*(x_t) = \max_{u_t} \left\{ \sum_{y \in X} P_t(y|x_t, u_t) v_{t+1}(y) \right\}$$

,so

$$V_t = \max_{u_t} \{ U^{(u_t)} V_{t+1}^* \}$$

We use Excel for multiplying matrices

$$V_5^* = \max \left\{ \begin{pmatrix} 0.1 & 0.9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0.9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.45 \\ 1 \end{pmatrix} \right\},$$

$$, \begin{pmatrix} 0.55 & 0 & 0.45 & 0 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0.45 & 0 & 0 & 0 \\ 0 & 0 & 0.55 & 0 & 0.45 & 0 & 0 \\ 0 & 0 & 0 & 0.55 & 0 & 0.45 & 0 \\ 0 & 0 & 0 & 0 & 0.55 & 0 & 0.45 \\ 0 & 0 & 0 & 0 & 0 & 0.55 & 0.45 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.45 \\ 1 \end{pmatrix} \} =$$

$$= \max \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0.405 \\ 0.945 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.203 \\ 0.45 \\ 0.698 \\ 1 \end{pmatrix} \right\} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.203 \\ 0.45 \\ 0.945 \\ 1 \end{pmatrix}, \quad u_5^*(x_5) = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \end{pmatrix}$$

, we remember $u_5^*(x_5)$ - a list of the best strategies at $t=5$

$$\begin{aligned}
V_4^* &= \max\left\{ \begin{pmatrix} 0 \\ 0 \\ 0.182 \\ 0.425 \\ 0.896 \\ 0.995 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0.091 \\ 0.203 \\ 0.537 \\ 0.698 \\ 0.97 \\ 1 \end{pmatrix} \right\} = \begin{pmatrix} 0 \\ 0.091 \\ 0.203 \\ 0.537 \\ 0.896 \\ 0.995 \\ 1 \end{pmatrix}, & u_4^*(x_4) &= \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\
V_3^* &= \max\left\{ \begin{pmatrix} 0.082 \\ 0.191 \\ 0.503 \\ 0.86 \\ 0.985 \\ 0.999 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.091 \\ 0.292 \\ 0.514 \\ 0.743 \\ 0.943 \\ 0.997 \\ 1 \end{pmatrix} \right\} = \begin{pmatrix} 0.091 \\ 0.292 \\ 0.514 \\ 0.86 \\ 0.985 \\ 0.999 \\ 1 \end{pmatrix}, & u_3^*(x_3) &= \begin{pmatrix} 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\
V_2^* &= \max\left\{ \begin{pmatrix} 0.272 \\ 0.492 \\ 0.825 \\ 0.972 \\ 0.998 \\ 0.9999 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.282 \\ 0.547 \\ 0.726 \\ 0.923 \\ 0.992 \\ 0.9997 \\ 1 \end{pmatrix} \right\} = \begin{pmatrix} 0.282 \\ 0.547 \\ 0.825 \\ 0.972 \\ 0.998 \\ 0.9999 \\ 1 \end{pmatrix}, & u_2^*(x_2) &= \begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\
v_1^*(0) &= \max\{0.520643, 0.526156\} = 0.526156, & u_1^*(0) &= 2
\end{aligned}$$

We have

$$u_t(x) = \begin{cases} 2 & , x \leq \frac{t}{2} \\ 1 & , x > \frac{t}{2} \end{cases}$$

-we use timid strategy if and only if we lead in this match. Expectation of the reward is $v_1^*(0) = 0.526156$