**Introduction**

Yahzee is a game that is regularly played by more than 100 million people in the world. We propose a simplified version of Yahzee as a benchmark for RL algorithms. We have already used it for this purpose, and an implementation is available.

**Description of the game**

The player scores points by rolling specific combinations of five 6-sided dice. The game consists of 13 rounds. In each round, the player rolls the 5 dice (up to 3 times) and then scores the roll in one of 13 categories. In the first 6 categories, the goal is to get as many 1s, 2s, 3s, 4s, 5s, 6s as possible, while the other categories are related to special combinations such as: three of a kind, four of a kind, full house, small straight, large straight, chance, and Yahzee. A Yahzee is rolled when all five dice show the same face. Each category can be scored only once. After 3rd roll, the player must choose his score; if he cannot enter a score, he must select a zero. If the sum of points scored in the first 6 categories is greater than 62, the player earns a bonus of 35 points. The object of the game is to maximize the total score. The game ends once all 13 categories have been scored.

In the simplified version of Yahzee we propose as a benchmark, the player cannot choose the dice to re-roll, while, in the original version, the player must throw all the dice on its first turn of each round, but he can select the dice to re-roll on its second and third turns. Therefore, the action space is reduced to 14 possible actions: the learner can choose either to fill one of the 13 categories with the obtained result, or reroll the dice. Furthermore, we have removed the possibility to obtain the mentioned bonus.

We propose a representation consisting of 14 variables. One is related to the number of rolls performed in the current round; the other 13 are associated to the 13 categories, and each one contains the score that the player would get by filling it with the current dice values. If the category has been already used, the corresponding position is set to -1. The size of this state space is about $0.51 \cdot 10^{11}$.

Given that in Yahzee the goal is to maximize the sum of the points scored in the thirteen categories, we propose the following reward function: the reward is equal to -100 when the player wants to fill a category which has been already filled, or the dice are rerolled for more than three times; each time a category is filled, the learner is rewarded with the related score. On rerolling the reward is 0.

**Motivations**

Yahzee is a game with very simple rules, and this has determined its great popularity. Among the reasons that make Yahzee a good benchmark for RL algorithms let us mention the following. The state space is large. The transition model is highly stochastic, but the reward function is deterministic. The game does not require any opponent, while other games, e.g., Backgammon, need it: the learning performance depends on how the opponent is implemented and this makes benchmarking difficult. The state space size can be reduced or enlarged by removing or adding categories; for instance, it would be possible to define a reduced version of the problem with only the first six categories in order to face it, without using function approximation, with standard tabular approaches, since the number of states reduces to about $0.35 \cdot 10^6$. Finally, it is easy to produce simple hand-coded policies that achieve good performances, while it is very challenging to find the optimal policy, even if it is known: recently, Woodward [1] has solved the game computing all of the more than 1 trillion possible outcomes and working out optimal playing strategies.

**References**