Reinforcement Learning	**Fall Semester, 1999/2000**

Homework 3: Dec. 16

Lecturer: Yishay Mansour

Homework number 3.

Programming Assignment

In this assignment you need to write a program that will learn to play the card game 21.

Game description: There is a deck of cards, total 52 cards. There are two players, the *house* and the *gambler*. The winner in the game is the player with the most number of points, which are less than (or equal to) 21.

When counting the points, each number card (2 to 10) has a value which is his number. Each face card (J, Q or K) is 10 points. The value of an ace A is 11 points (simplifying the rule of the real game where A can be either 1 or 11.)

At the beginning each player gets two cards, one is faced up (which you see) while the other is faced down (which you don't see). The gambler look at all its cards (and the house open card) and need to decide whether to ask for a another card (`hit`) or end (`stop`).

We fix the strategy of the house as follows. If the sum of the cards is 15 or less, the house perform `hit`, and if the sum is 16 or more it performs `stand`.

We model the game as an MDP whose states are labeled by the sum of the card of the gambler.

Task 1: Implement TD(0) and use it to compute the probability that the gambler wins, given that his policy is: If the sum is 18 or more then `stand` else `hit`.

Task 2: Implement either Q-Learning or Sarsa, and compute an optimal policy against the house policy we consider. (Give as an output of the optimal action in each state, and the probability of winning from that state.)

Task 3: Try to improve the results by modifying the structure of the MDP. (The strategy of the house remains unchanged.)

The homework is due in two weeks