# Example 5.1 from Sutton’s book: Blackjack

The object of the popular casino card game of blackjack is to obtain cards the sum of whose numerical values is as great as possible without exceeding 21. We consider the version in which each player competes independently against the dealer (i.e., one player.)

* We consider one deck of cards with 52 cards.
* All face cards count as 10, and an ace can count either as 1 or as 11.
* Initial: The game begins with two cards dealt to both dealer and player. One of the dealer’s cards is face up and the other is face down. If the player has 21 immediately (an ace and a 10-card), it is called a natural. The player then wins unless the dealer also has a natural, in which case the game is a draw.
* If the player does not have a natural, then the player can request additional cards, one by one (hits), until either stops (sticks) or exceeds 21 (goes bust).
  + Your turn: your action is “hits” or “sticks” if you have not exceeded 21. The states depend on the player’s cards and the dealer’s showing card.
  + Dealer’s turn: If you go bust, you lose; if you stick, then it becomes the dealer’s turn. The dealer hits or sticks **according to a fixed strategy without choice: he sticks on any sum of 17 or greater (include facedown card value), and hits otherwise**.
  + End of game: If the dealer goes bust, then the player wins; otherwise, the outcome—win, lose, or draw—is determined by whose final sum is closer to 21.
  + **Terminal reward**: Playing blackjack is naturally formulated as an episodic finite MDP. Each game of blackjack is an episode. Rewards of +1, -1, and 0 are given for winning, losing, and drawing, respectively. You do not earn rewards within a game, except the terminal rewards you get at the end of the game (when you see the outcomes). We do not discount in such a setting.
  + Ace card: If the player holds an ace that he could count as 11 without going bust, then the ace is said to be *usable*. In this case it is always counted as 11 because counting it as 1 would make the sum 11 or less, in which case there is no decision to be made because, obviously, the player should always hit.
  + State: The player makes decisions on the basis of three variables: his current sum (12–21), the dealer’s one showing card (ace–10), and whether or not he holds a usable ace. This makes for a total of 10\*10\*2=200 states.

## Question:

Consider the policy that sticks if the player’s sum is 20 or 21, and otherwise hits.

1. To find the state-value function for this policy by a Monte Carlo approach, the first step is to simulate many blackjack games using this policy. Write your simulation program and report the first 10 simulate trajectories.
2. For the first 10 simulate trajectories, use the Monte Carlo approach to evaluate the value function associated with each state you see in these trajectories. Using the first-visit method and the every-visit method, respectively.
3. Report the value function for each state in the entire state space, after simulate 500,000 games.