1 Probabilistic Reasoning

2 Making Decisions

2.1 Utility Theory (1)

The lottery over A and B that is indifferent to C must be of the form [p, A; 1-p, B]. It must also satisfy this equation to be indifferent to C: $p * U(A) + (1-p) * U(B) = U(C)$
That can be resolved as $p * 455 + (1-p) * (-150) = 50$ and finally $p = \frac{200}{605} = 0,331$. The lottery over A and B is therefore [0,331, A; 0,669, B].

2.2 Utility Theory (2)

The given lotteries are [0,5, B; 0,5, C] and [1, A] = A. The expected Utilities of these are: $U([0,5,B;0,5,C]) = 0,5 * U(B) + 0,5 * U(C) = 0,5 * 20 + 0,5 * 0 = 10$ and $U(A) = 5$. Therefore the lottery [0,5, B; 0,5, C] is preferred to the lottery [1, A].

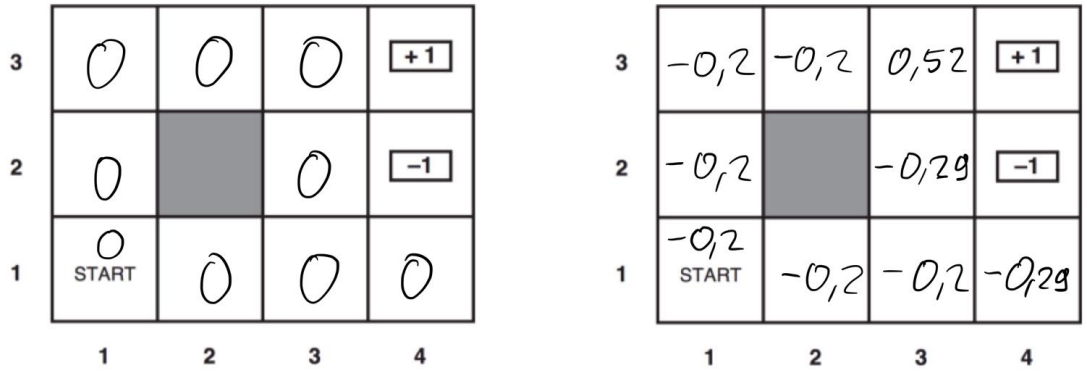
2.3 Markov Decision Processes (MDPs)

2.3.1 Computing discount factor γ

Since only continuing in state 4 yields a reward, the optimal policy is to continue in every state until state 5 is reached. The optimal value of state 1 is therefore: $U^*(s_1) = R(s_1, a_C) + \gamma R(s_2, a_C) + \gamma^2 R(s_3, a_C) + \gamma^3 R(s_4, a_C) + \gamma^4 R(s_5, a_C) = 1$
Adding the rewards results in: $U^*(s_1) = 0 + 0 + 0 + \gamma^3 * 10 + 0 = 10 * \gamma^3 = 1$.
γ is therefore $\sqrt[3]{\frac{1}{10}} \approx 0,464$.

2.3.2 Markov Decision Process - Value Iteration

The utility values of the grid world are initialized as 0 for every state, except for the terminal states (3,4) and (2,4) which have fixed utilities of +1 and -1 respectively. This is shown in the figure below on the left side. After the first iteration these values change as shown in the figure below on the right side. They are calculated as follows:
$U_1((3,3)) = \max_{a \in A} \sum_{s'} P(s'|(3,3),a) * [R((3,3),a,s') + \gamma * U_0(s')] = 0,8 * (-0,2 + 0,9 * 1) + 0,1 * (-0,2 + 0,9 * 0) + 0,1 * (-0,2 + 0,9 * 0) = 0,52$
$U_1((2,3)) = \max_{a \in A} \sum_{s'} P(s'|(2,3),a) * [R((2,3),a,s') + \gamma * U_0(s')] = 0,8 * (-0,2 + 0,9 * 0) + 0,1 * (-0,2 + 0,9 * -1) + 0,1 * (-0,2 + 0,9 * 0) = -0,29$
$U_1((1,4))$ is calculated analogously to $U_1((2,3))$ and results in -0,29.
$U_1((1,1)) = \max_{a \in A} \sum_{s'} P(s'|(1,1),a) * [R((1,1),a,s') + \gamma * U_0(s')] = 0,8 * (-0,2 + 0,9 * 0) + 0,1 * (-0,2 + 0,9 * 0) + 0,1 * (-0,2 + 0,9 * 0) = -0,2$
All other states are calculated analogously to $U_1((1,1))$ and result to -0,2.



The utility value of state (3,3) for the second iteration can now be calculated as follows:
$U_2((3,3)) = \max_{a \in A} \sum_{s'} P(s'|(3,3),a) * [R((3,3),a,s') + \gamma * U_1(s')] = 0,8 * (-0,2 + 0,9 * 1) + 0,1 * (-0,2 + 0,9 * 0,52) + 0,1 * (-0,2 + 0,9 * -0,29) = 0,56 + 0,0268 - 0,0461 = 0,5407$