TDT4171 - Exercise 4

Jens Waage

March 2020

1 Problem 1: Utility

(a) We classify the four people as follows:

Gabriel: risk-seeking. His utility function is convex, which means he will always have higher expected utility choosing an uncertain lottery over an option where he receives the expected value of the lottery for certain. This implies that he prefers the risk.

Gustav: risk-neutral. His utility function is linear, which means he'll always receive the same utility when choosing between an uncertain lottery and the expected value of the lottery for certain. In fact, his utility is constant no matter what he chooses.

Maria: risk-averse. Her utility function is concave, which indicates she always maximizes her expected utility by choosing to receive the expected value of a lottery for certain, rather than an uncertain lottery. This implies she prefers certainty over risk.

Sonja: risk-neutral. Same argument as Gustav, except her expected utility is not constant. She will receive the same expected utility from the lottery and the certain outcome, and so she is indifferent when choosing between risk and certainty.

(b) We set up the table

	Risk-seeking	Risk-averse
Lottery	[(0.5, $2), (0.5, $10)]	[(0.5, -$2), (0.5, -$10)]
Expected utility of lottery	504	-504
Utility of exepected monetary value	216	-216

We can see that $504 > 216$, so the lottery is preferred in the first case. Likewise, $-216 > -504$, so the certain outcome is preferred in the second case.

2 Problem 2: Decision Network

(a) The decision network is shown below.

(b) We compute the expected utilities:

$$
\begin{gathered}
EU(b) = U_1(b) + [P(p|b,m) + P(p|b,\neg m)] * U_2(p) + \\
\mathrm{P}(\neg p|b,m) + P(\neg p|b,\neg m) * U_2(\neg p) \\
EU(b) = -\$100 + [0.9 + 0.5] * \$2000 + [0.1 + 0.5] * \$0 \\
EU(b) = \$2700 \\
EU(\neg b) = U_1(\neg b) + [P(p|\neg b,m) + P(p|\neg b,\neg m)] * U_2(p) + \\
\mathrm{P}(\neg p|\neg b,m) + P(\neg p|\neg b,\neg m) * U_2(\neg p) \\
EU(\neg b) = -\$0 + [0.8 + 0.3] * \$2000 + [0.2 + 0.7] * \$0 \\
EU(b) = \$2200
\end{gathered}
$$

(c) Sam should buy the book. This is because taking the action of buying the book maximizes his expected utility, and so is the rational decision.

3 Problem 3: Markov Decision Process

(a) For each iteration i we go through each of the states and set the utility according to

$$U'[s] \leftarrow R(s) + \gamma \max_{a \in A(s)} \sum_{s'} P(s'|s,a)U[s'] \tag{1}$$

where

- s is the state we're moving from
- $A(s)$ is the set of all possible actions in state s
- $P(s'|s,a)$ is the transition probability of ending up in state s' given we're in state s and take action $a \in A(s)$
- $R(s)$ is the reward for being in state S
- γ is the discounting factor

Iteration 1:

- U[1] $= 0 + 0.5 * \max[(1/4 * 0 + 3/4 * 0), (3/4 * 0 + 1/4 * 0)] = 0$
- U[2] $= 1 + 0.5 * \max[(3/4 * 0 + 1/4 * 0), (1/4 * 0 + 3/4 * 0)] = 1$
- U[3] $= 0 + 0.5 * \max[(1/4 * 0 + 3/4 * 0), (3/4 * 0 + 1/4 * 0)] = 0$

Table looks as follows:

0	0	0
0	1	0

Iteration 2:

- $U[1] = 0 + 0.5 * \max[(1/4 * 1 + 3/4 * 0), (3/4 * 1 + 1/4 * 0)] = 0.375$
- $U[2] = 1 + 0.5 * \max[(3/4 * 0 + 1/4 * 0), (1/4 * 0 + 3/4 * 0)] = 1$
- $U[3] = 0 + 0.5 * \max[(1/4 * 0 + 3/4 * 1), (3/4 * 0 + 1/4 * 1)] = 0.375$

Table looks as follows:

0	0	0
0	1	0
0.375	1	0.375

- $U[1] = 0+0.5*\max[(1/4*1+3/4*0.375), (3/4*1+1/4*0.375)] = 0.375$
- $U[2] = 1 + 0.5 * \max[(3/4 * 0 + 1/4 * 0), (1/4 * 0 + 3/4 * 0)] = 1$
- $U[3] = 0 + 0.5 * \max[(1/4 * 0 + 3/4 * 1), (3/4 * 0 + 1/4 * 1)] = 0.375$

(b) To find the optimal action from state 1 we use

$$\pi^*(s) = \underset{a \in A(s)}{\arg\max} \sum_{s'} P(s'|s, a)U(s') \tag{2}$$

$\pi^*(1) = \arg\max[(1/4 * 1.25 + 3/4 * 0.5), (3/4 * 1.25 + 1/4 * 0.5)] = R$

4 4

See python code. Output:

```
E:\Jens\Anaconda3\python.exe "C:/Users/Jens/Documents/myWork/Projects/TDT4173-Exercises/Exercise 4/value_
Utilities:
[(0, 1.1339590752548871), (1, 0.475799246043496), (2, 1.4574163568804344), (3, 0.16403932694336357)]
[(4, 1.5126122588028232), (5, -9.999999999999995), (6, 2.0830984518216904), (7, -9.999999999999995)]
[(8, 3.300662715986901), (9, 5.784066128742564), (10, 5.53208118308568), (11, -9.999999999999995)]
[(12, -9.999999999999995), (13, 7.068221080313933), (14, 8.349326710415065), (15, 9.999999999999995)]

Actions:
[(0, 'down'), (1, 'right'), (2, 'down'), (3, 'left')]
[(4, 'down'), (5, None), (6, 'down'), (7, None)]
[(8, 'right'), (9, 'down'), (10, 'down'), (11, None)]
[(12, None), (13, 'right'), (14, 'right'), (15, 'left')]
```