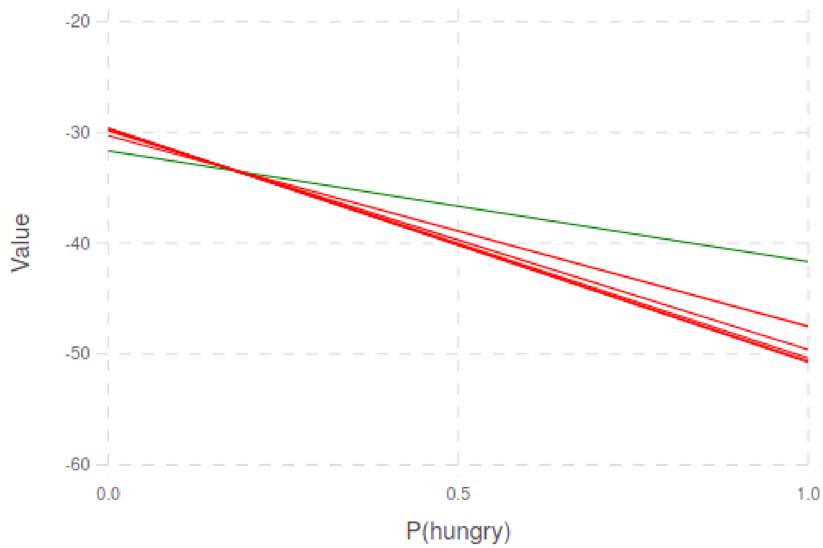
ASEN 6519-007 Decision Making under Uncertainty
Homework 5: Introduction to POMDPs

March 16, 2020

1 Conceptual Questions

Question 1. Consider a modified version of the Crying Baby POMDP discussed in class and in Section 6.1.1 of *Decision Making under Uncertainty* by Kochenderfer. This version is exactly as described in the book, except that the baby has a 20% chance of becoming hungry at the next time step if it is not currently hungry and only a 60% chance of crying when hungry. The alpha vectors for the optimal policy are shown below (green=`feed`, red=`don't feed`). The point at which the alpha vectors for feeding and not feeding intersect is at $P(\texttt{hungry}) = 0.19$. Draw an optimal policy graph if the initial belief is certainty that the baby is not hungry ($b_0(\texttt{hungry}) = 0$).



2 Exercises

Question 2. Consider the following POMDP that represents a personalized cancer monitoring plan[1]:

$$
\begin{aligned}
\mathcal{S} &= \{\texttt{healthy}, \texttt{in-situ-cancer}, \texttt{invasive-cancer}, \texttt{death}\} \\
\mathcal{A} &= \{\texttt{wait}, \texttt{test}, \texttt{treat}\} \\
\mathcal{O} &= \{\texttt{positive}, \texttt{negative}\} \\
\gamma &= 0.99 \\
s_0 &= \texttt{healthy}
\end{aligned}
$$

The **transition dynamics** are as follows:

- If the patient is `healthy`, at each timestep, they have a 2% chance of developing `in-situ-cancer`.
- If the patient has `in-situ-cancer` and they are `treat`ed, they have an 60% chance of becoming healthy at the next time step.
- If the patient has `in-situ-cancer` and they are not `treat`ed, they have a 10% chance of developing `invasive-cancer`.
- If the patient has `invasive-cancer` and they are `treat`ed, they have a 20% chance of recovering and a 20% chance of dying at the next time step.
- If the patient has `invasive-cancer` and they are not `treat`ed, they have a 60% chance of dying.
- In all other cases, the state remains the same as in the previous step.

The **observation** is determined as follows:

- If the action is `test` and the *new* state is `healthy`, then the observation will be (falsly) `positive` 5% of the time.
- If the action is `test` and the *new* state is `in-situ-cancer` then the observation will be `positive` 80% of the time.
- If the action is `test` and the *new* state is `invasive-cancer` then the observation will be `positive`.
- If the action is `treat` and the *new* state is `in-situ-cancer` or `invasive-cancer` then the observation will be `positive`.
- In all other cases, the observation is `negative`.

The **rewards** are defined as follows (one could interpret the reward as roughly quality years of life):

- $R(\texttt{death}, \text{any action}) = 0.0$ (i.e. `death` is a terminal state)
- $R(\text{any living state}, \texttt{wait}) = 1.0$
- $R(\text{any living state}, \texttt{test}) = 0.8$ (because of costs and anxiety about a positive result)
- $R(\text{any living state}, \texttt{treat}) = 0.1$

(a) Use Monte Carlo simulations to evaluate a policy that always **wait**s.

(b) Propose a better heuristic strategy based on the observation history or belief and evaluate it with Monte Carlo simulations. See if you can get an average discounted return of 75 or more.

[1] Note that the probabilities are not meant to be realistic. See `https://pubsonline.informs.org/doi/10.1287/opre.1110.1019` for an actual publication on this topic