# PS3: Practice problems: Revised

**Answer** **Last question is not finished. Final version out soon.** Answers are shown in green. These answers have not been proofread by anyone but me, so there is substantial chance of error. Please let me know if you have any doubts. -lpk

## 1 To Drop Or Not To Drop

A student is trying to decide whether to take Frobology from Professor X or Professor Y, or not to take it at all. There is no homework in either version of the course. They both have a final exam. The only difference between them is that Professor X gives a pass/fail midterm quiz, which is graded before the drop date. In either case, the ultimate outcome is a grade (let's just say, for simplicity, it's A, C, or F). The final grade distributions are the same in both versions. Remember also that it's not necessarily fun to take a quiz.

1. Draw the decision tree for a student faced with this decision. (There is no need to model the decision of whether to drop the course if you take it from Professor Y.)

   **Answer** Branch first on X, Y, or None. On the X branch, now you branch on whether you pass or fail the final. For each of these branches, branch on whether you drop the class or don't. Finally, for each branch in which you don't drop the class, branch on getting A, C, or F. On the Y branch, only branch on getting A, C, or F.

2. What probabilities and utilities would you need to know in order to compute the best course of action?

   **Answer** The easiest way to think about the probabilities here is that there is a fundamental condition (which grade you are going to get) and that the midterm is diagnostic of that. So, I'd specify the underlying distribution on A, C, or F (which is what we'll use for Y's class). Then, the probability of passing the midterm, given that you are an A student, a C student, or an F student. Using Bayes' rule, you can get the quantities you need to evaluate the X branch: which is the probability that you'll pass the midterm, and the conditional probabilities of the grades given that you pass the midterm and given that you fail the midterm.

   If you assume utilities for various aspects of the situation are additive, then you'd need utility for: getting no course credit (and no F), getting no course credit and an F, getting an A, getting a C, doing half of the work of the class, doing all of the work of the class, and for taking a midterm.

3. Describe, in English, a set of circumstances under which Professor X's course would be preferred.

   **Answer** If there's no cost to taking the midterm, then taking the class with Prof X is strictly no worse than taking the class with Prof Y. If the midterm is usefully diagnostic of your final grade, and if getting an F is significantly worse than dropping, then Prof X's class will be better.

**4.** Describe, in English, a set of circumstances under which Professor Y's course would be preferred.

**Answer** If the midterm doesn't give useful information about your final grade or you really hate taking tests then Prof X is not for you.

## 2 Short answers

**1.** Since belief space is real valued (and therefore infinite), why is the branching factor in a POMDP finite?

**Answer** There are a finite number of possible observations in the standard POMDP formulation.

**2.** In a POMDP, we can think of the reward of a belief state, $b$, as being $\sum_s b(s)R(s)$. Why doesn't this lead to self delusion?

**Answer** We are required to use correct Bayesian updating to compute $b$ from the history of actions and observations, so we can't persist in believing good things will happen in the face of evidence to the contrary.

**3.** Why is the finite-horizon optimal policy for an MDP non-stationary, while the infinite-horizon optimal policy is stationary?

**Answer** In the infinite horizon case, the future possibilities are always the same, so there is no reason to make decisions differently. In the finite horizon case, the actual opportunities available change with the number of remaining time steps.

**4.** Give an intuitive motivation for discounting.

**Answer** There may be uncertainty in the number of remaining steps of the game; or we might value a monetary payoff more in the short term because of investment opportunities.

**5.** Consider reinforcement learning. With the same amount of experience, would you expect Q-learning or adaptive dynamic programming (estimate the transition probabilities and rewards, and solve the model) to arrive at a better policy? Under what circumstances (if any) would you prefer the other method?

**Answer** Q learning is very slow to propagate reward through chains of states (it requires the agent to actually traverse them during learning), so ADP is always more efficient in terms of steps of interaction with the environment. We might prefer Q learning if the space is very large, so it's hard to represent and learn the model; in such cases, Q learning with function approximation might be a better strategy.

**6.** What makes it difficult to do Q learning in very large action spaces?

**Answer** Every choice of an action requires maximization over the space of actions; it also increases the size of the table to be stored, and the amount of experience required to learn good Q values.

**7.** Given that Q can be defined in terms of V and vice versa, why is it important to learn Q rather than V?

**Answer** If you have the Q function, you can easily find the optimal action for a given state; if you only have V, you also need the transition model to do one step of look-ahead to find the optimal action.

**8.** Consider a finite-horizon MDP for which we know the initial state. We know two strategies for finding the optimal k-step policy: (1) build a depth-k tree with state s at the root and (2) calculate the optimal k-step policy for all states. Under what circumstances is each approach best?

**Answer** Doing tree search is best if the size of the k-step tree is smaller than the size of the state space.

## 3 Slightly Less Naive Bayes

In the usual application of naive Bayes to supervised learning, we predict class 1 when $\Pr(Y = 1|X) > \Pr(Y = 0|X)$, where $X$ is the feature vector describing the instance.

What if we were in a situation in which one type of error was much more costly than another? For instance, if we're diagnosing a terrible disease, it might cost much more to miss a diagnosis than to make an erroneous one. So, our utilities might be described as in this table:

| Predicted Y | Actual Y | Utility |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | -10 |
| 1 | 0 | -1 |
| 1 | 1 | 0 |

If these are our utilities, then give a rule, in terms of $\Pr(Y = 1|X)$ and $\Pr(Y = 0|X)$, for making predictions that maximize expected utility.

**Answer**

$$\begin{aligned}
EU(\text{predict } 1|x) &= \Pr(Y=1 \mid X)U(Y=1, \text{predict } 1) + \Pr(Y=0 \mid X)U(Y=0, \text{predict } 1) \\
&= -\Pr(Y=0|X) \\
EU(\text{predict } 0|x) &= \Pr(Y=1 \mid X)U(Y=1, \text{predict } 0) + \Pr(Y=0 \mid X)U(Y=0, \text{predict } 0) \\
&= -10\Pr(Y=1|X)
\end{aligned}$$

So, we will predict 1 when

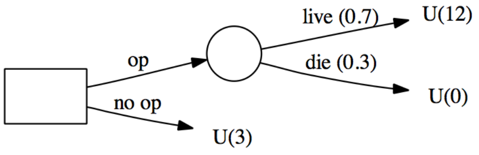
$$\begin{aligned}
EU(\text{predict } 1|x) &> EU(\text{predict } 0|x) \\
-\Pr(Y=0|X) &> -10\Pr(Y=1|X) \\
\Pr(Y=0|X) &< 10\Pr(Y=1|X) \\
\frac{\Pr(Y=1|X)}{\Pr(Y=0|X)} &> \frac{1}{10}
\end{aligned}$$

## 4 Decision Theory

Dr. No has a patient who is very sick. Without further treatment, this patient will die in about 3 months. The only treatment alternative is a risky operation. The patient is expected to live about 1 year if he survives the operation; however, the probability that the patient will not survive the operation is 0.3.

**1.** Draw a decision tree for this simple decision problem. Show all the probabilities and outcome values.

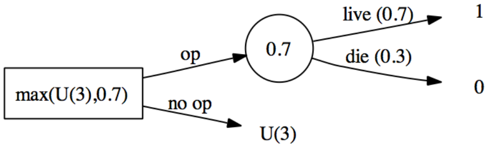
**Answer**



**2.** Let $U(x)$ denote the patient's utility function, where $x$ is the number of months to live. Assuming that $U(12) = 1.0$ and $U(0) = 0$, how low can the patient's utility for living 3 months be and still have the operation be preferred?

For the rest of the problem, assume that $U(3) = 0.8$.

**Answer** The operation would be preferred as long as $U(3) < 0.7$.



**3.** Dr. No finds out that there is a less risky test procedure that will provide uncertain information that predicts whether or not the patient will survive the operation. When this test is positive, the probability that the patient will survive the operation is increased. The test has the following characteristics:

- True-positive rate: The probability that the results of this test will be positive if the patient will survive the operation is 0.90.
- False-positive rate: The probability that the results of this test will be positive if the patient will not survive the operation is 0.10.

What is the patient's probability of surviving the operation if the test is positive? **Answer**

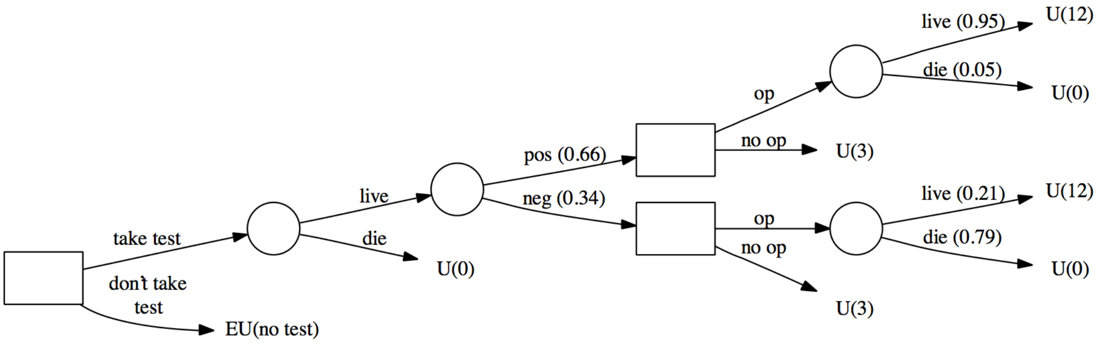
$$\Pr(\text{survive} \mid \text{pos}) = \frac{0.9 \cdot 0.7}{0.66} = 0.9545$$

**4.** Assuming the patient has the test done, at no cost, and the result is positive, should Dr. No perform the operation?

**Answer** Yes. $EU(\text{op}) > 0.8$

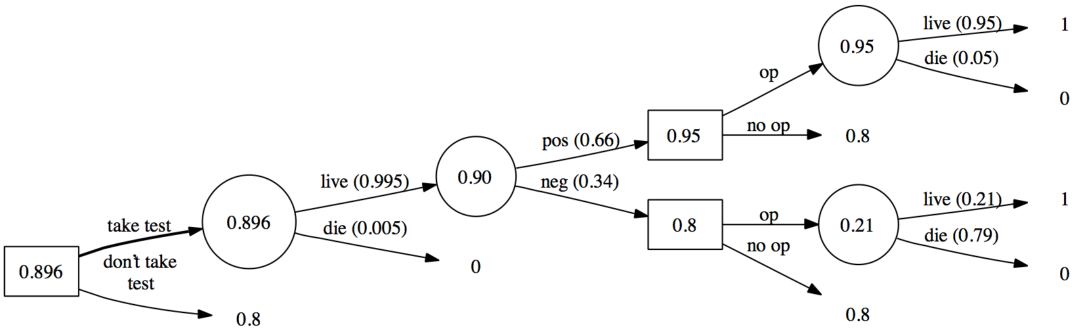
**5.** It turns out that the test may have some fatal complications, i.e., the patient may die during the test. Draw a decision tree showing all the options and consequences of Dr. No's problem.

**Answer** For this problem, we need to calculate $\Pr(\text{survive} \mid \text{neg}) = 0.2059$. Then we have the following decision tree:



**6.** Suppose that the probability of death during the test is 0.005 for the patient. Should Dr. No advise the patient to have the test prior to deciding on the operation?

**Answer** Yes, the test should be taken. The evaluated decision tree is:



## 5 Markov Decision Processes

You are playing a game at a carnival, in which you are trying to throw balls through a hoop. You are allowed to play this game for a total of $k$ steps. If, at the end of $k$ steps, you have gotten at least one ball through the hoop, then you win \$10. If not, you win nothing. On each step, you are allowed to buy, for \$1 each, as many balls as you would like, which you will try to throw simultaneously through the hoop. Each ball has a probability of $p$ of going through the hoop, and each ball's success is independent of the successes of the other balls and of the number of balls being thrown. After you've thrown one set of balls, you can observe whether or not any of them went through the hoop.

1. If you throw $n$ balls at once, what is the probability of getting at least one ball through the net? Write an expression in terms of $p$ and $n$. Call this quantity $f(p, n)$ in future parts of this problem.

   **Answer**

   $$f(p, n) = 1 - (1 - p)^n$$

2. If $k = 1$, that is, you can only play one round of this game, what is the optimal number, $n$, of balls to buy and throw? (You only need to write down an expression involving $n$ and $p$; don't worry about getting a closed form).

   **Answer**

   $$10f(p, n) - n$$

3. Let $s_1$ be the state of not having gotten a ball through the hoop and $s_2$ be the state of having gotten one through. Let $V^k(s)$ be the value of being in state $s$ with $k$ steps remaining in the game. Then $V^0(s_1) = 0$ and $V^0(s_2) = 10$. What is $V^k(s_2)$, assuming there is no discounting.

   **Answer** 10, because you're guaranteed to win 10 at this point and you don't need to buy any more balls.

4. Write an expression for $V^k(s_1)$, in terms of $V^{k-1}(s)$.

**Answer**

$$\begin{aligned} V^k(s_1) &= \max_n f(p,n)V^{k-1}(s_2) + (1 - f(p,n))V^{k-1}(s_1) \\ &= \max_n 10f(p,n) + (1 - f(p,n))V^{k-1}(s_1) \end{aligned}$$

# 6 Tiger hunting: Revised

Let's consider the tiger problem, augmented with two new actions: shoot the tiger on the left and shoot the tiger on the right. Shooting at the door where the tiger is means that, with probability 0.9, the tiger will be scared and move to the other room, and with probability 0.1, he'll stay where he is. Shooting at the door where there is no tiger will have no effect.

Assume that the reward of listening is -1, of shooting is -5, of opening any door without a live tiger is +10, and of opening a door with a live tiger is -50.

There are two possible observations: noise-left and noise-right. If there is a tiger on the right, then P(noise-right) = 0.85, and P(noise-left) = 0.15. The situation is symmetric when the tiger is on the left.

Assume a discount factor of 0.9, but that the game is over when a door is opened.

Draw the alpha vectors for the following short policy trees. Which one is best in which situations?

**Answer** I added some more cases here. The alpha vectors are all shown in a final figure.

**a.** Always listen.

**Answer** -10, no matter what the state.

**b.** Open left.

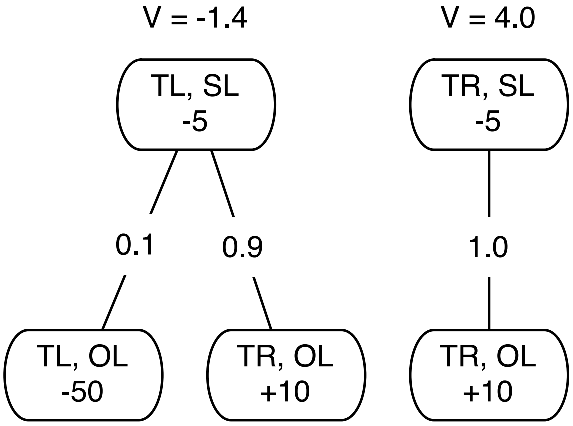
**Answer** -50 if TL; +10 if TR

**c.** Open right.

**Answer** +10 if TL; -50 if TR

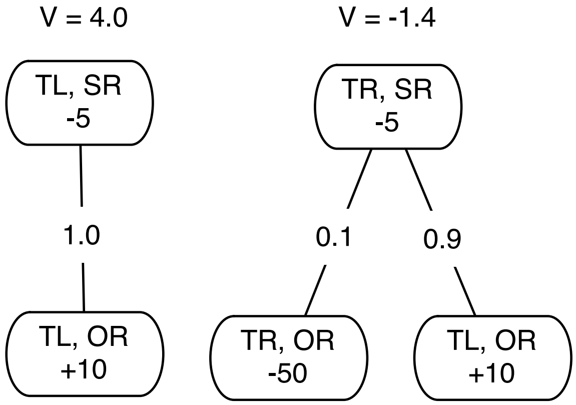
**d.** Shoot left, then open left.

**Answer** We have to evaluate at this starting in state TL and again starting in state TR. Here are the trees. Each node contains a world state, an action, and a cost. Arcs are labeled with probabilities. In this case, branches correspond to possible world transitions. The value is shown at the root of each tree.



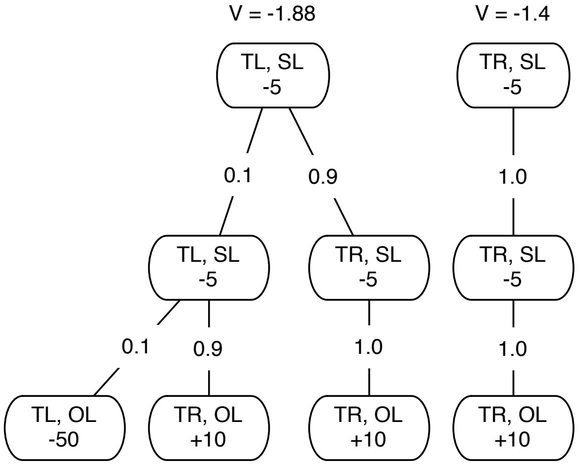
**e.** Shoot right, then open right.

**Answer**



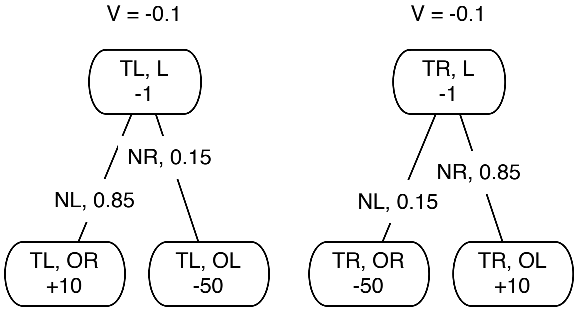
**f.** Shoot left, then shoot left, then open left.

**Answer**



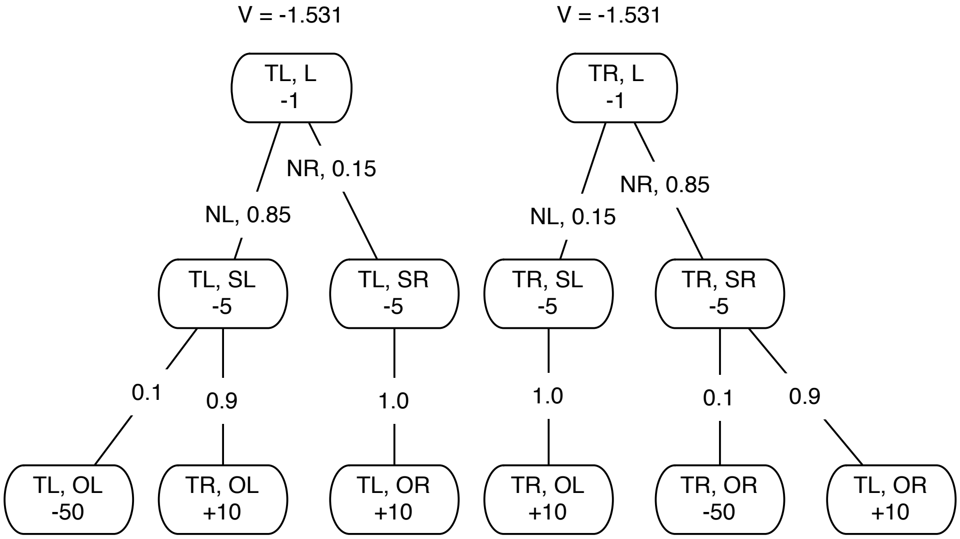
**g.** Listen; if noise-left, then open right; if noise-right, then open left.

**Answer**



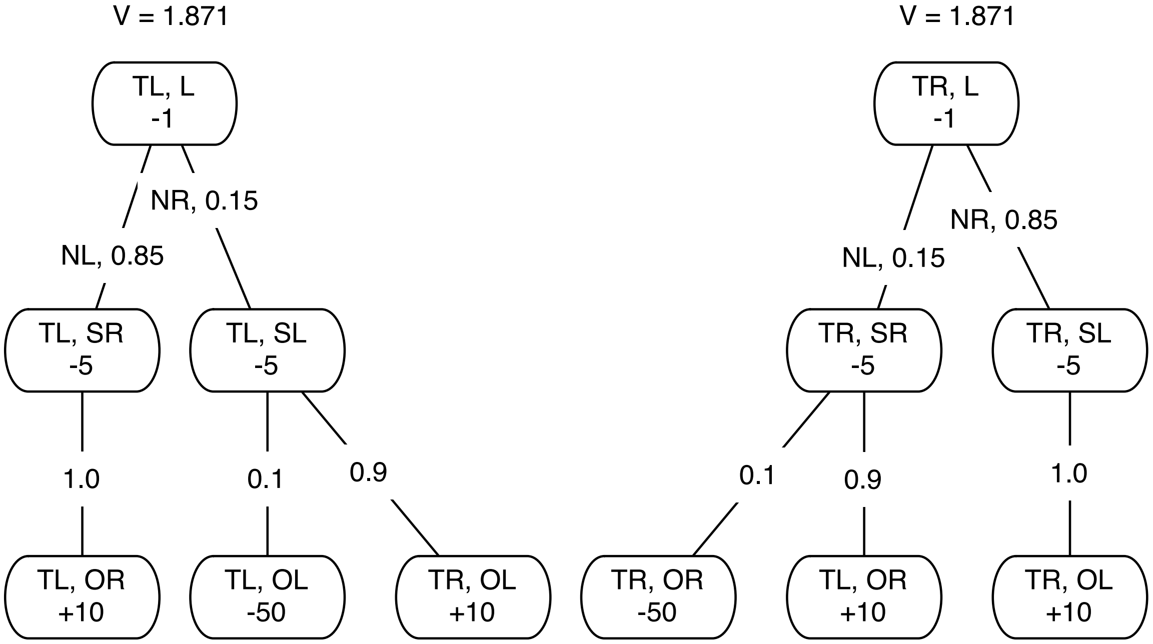
**h.** Listen; if noise-left, then shoot left, then open left; if noise-right, then shoot right, then open right.

**Answer**

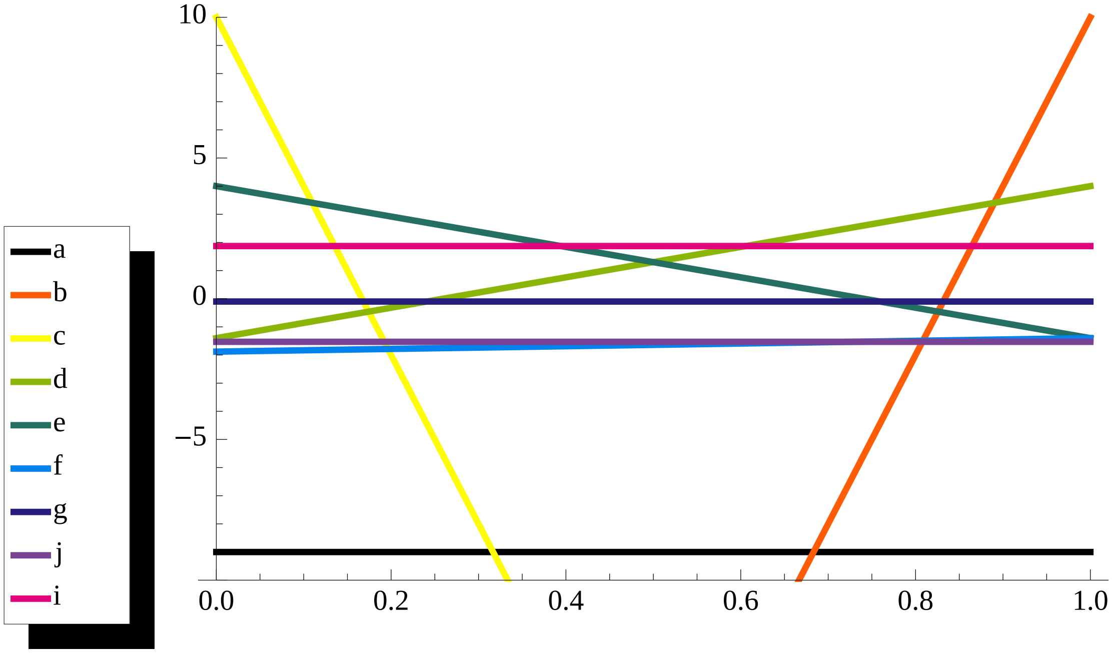


**i.** Listen; if noise-left, then shoot right, then open right; if noise-right, then shoot left, then open left.

**Answer**



**Answer** So, here is a plot of all of the associated $\alpha$-vectors.


10
5
0
−5
0.0
0.2
0.4
0.6
0.8
1.0
a
b
c
d
e
f
g
j
i