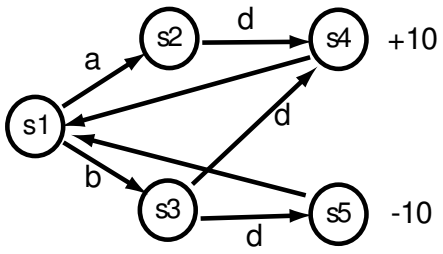
CSC 2534— Decision Making Under Uncertainty
Assignment 2

Craig Boutilier — Fall 2014

October 28, 2014
Due: November 11, 2014

1. Consider MDP diagrammed below. It consists of four states, s_1, s_2, s_3 and s_4; but the only action choice is at state s_1, where either action a or b can be taken. At all other states, there is a single action (call it d) that induces a specific stochastic transition function. s_4 and s_5 are the only states with associated rewards (10 and -10, respectively). Each action at each state has a 0.1 chance of inducing a self-transition, and a 0.9 chance of moving to the nominal successor state indicated in the diagram. There are two exceptions:

   - s_2, which is a "sticky state." It has a probability p_S (the stickiness factor) of self-transition (and probability $1 - p_S$ of moving to s_4.
   - s_3, which is a "risky state." It has the usual probability of self-transition (0.1), but has a probability p_R (the riskiness factor) of moving to s_5 and probability $0.9 - p_R$ of moving to s_4.



   Assume our utility is measured using total discounted reward over an infinite horizon with discount factor $\beta = 0.95$.

   (a) Fix the stickiness probability to be $p_S = 0.2$. Describe the optimal value function and policy for this MDP if the risky probability is $p_R = 0.01$. Describe the optimal value function and policy for this MDP if the risky probability is $p_R = 0.03$. For what value of p_R is the decision maker indifferent between doing a or b at state s_1. *Note: You may assume that the decision maker always starts out at s_1; you do not have to compute values at unreachable states if you like. Please give some brief justification or explanation for how you arrived at your conclusions.*

   (b) Fix the stickiness probability to be $p_S = 0.6$. Repeat the above questions for $p_R = 0.1$ and $p_R = 0.2$. And again compute the indifference level for p_R.

2. We're going to model the following problem as a POMDP. Patients arrives at a doctor's office, and either have disease X, disease Y or no disease N. These conditions are mutually exclusive. Depending on the patient characteristics, time of year, etc., the doctor's prior probability for each of these diseases may vary: so she wants you to develop a POMDP policy/value function to help diagnose and treat patients as a function of her prior beliefs (which will be modeled within a belief vector).

   The doctor has the following actions at her disposal, and rewards or costs associated with specific actions and patient states.

   - She can treat the patient with *one* of three medications, M_1, M_2, or M_3. Once treated with a medication, no further actions are possible by the doctor. Each medication is more or less effective depending on the actual disease and can be more or less harmful if applied to the wrong disease. We summarize the positive and negative effects using the following utilities for each medication-disease (or lack of disease) combination:

	X	Y	N
M_1	0	20	6
M_2	20	2	4
M_3	12	12	8

     In your POMDP, you should treat these as *rewards* associated with taking these treatment actions in states where those diseases are present.

     Each of these treatments can be applied only once. To model this, assume that once one of these actions is applied, the patient is considered "Treated". We model by assuming a "patient state" variable that can take one of four values, X, Y, N, Tr:

     - X means the patient has disease X and has *not* been treated.
     - Y means the patient has disease Y and has *not* been treated.
     - N means the patient has no disease and has *not* been treated.
     - Tr means the patient has been treated (this does not depend on the disease that was actually present when the patient was treated).

     Once any of the three treatments is applied, Tr becomes true with probability 1. If any of the three treatments is applied when Tr holds (i.e., a second treatment is applied), the action cost is -100.

   - She can run tests prior to prescribing a medication. Test T_1 returns either a High or Low result and and helps distinguish disease X from Y, but provides little information about the presense versus absense of a disease. Test T_2 returns a Yes or No result helps distinguish having *some* disease from having none. The following are the probabilities of each test result conditional on the underlying disease (or lack thereof):

	X	Y	N	Tr
$\Pr(T_1 = H)$	0.9	0.2	0.5	1.0
$\Pr(T_1 = L)$	0.1	0.8	0.5	0.0
$\Pr(T_2 = Y)$	0.7	0.8	0.1	1.0
$\Pr(T_2 = N)$	0.3	0.2	0.9	0.0

     Each test has an action cost of -2.

     Furthermore, once one of the tests is administered, it destroys the sensitivity of any other test, and may be harmful. To model this we assume a "test state" variable that can take on of two values: Ts (patient has been tested) or NTs (patient has not been tested).

Once any test is applied, Ts becomes true with probability 1. If any test is applied when Ts holds (i.e., a second test is applied), the action cost is -100. The result of either test is completely random (each result is equally likely) if a test is applied when Ts holds.

- The doctor has a Null action (in which she does nothing). It has no cost/reward and has no impact the patient state or test state.
- Once the consultation with the doctor ends, there is a terminal reward that depends on the patient state at the end of the process: $r(X) = 3; r(Y) = 3; r(N) = 10; r(Tr) = 0$. Note that the reward for treating a patient (more or less) appropriately is encoded in the action rewards, not the terminal reward. (The terminal reward encodes the relative value of ending the process without having treated a patient as a function of his disease state.)

All action costs, action rewards, and terminal rewards are additive. We model this problem as an undiscounted, finite-horizon POMDP with two stages. The POMDP has eight states, which are numbered/labeled as follows:

- $s_1 : X, NTs$ $s_2 : Y, NTs$ $s_3 : N, NTs$ $s_4 : Tr, NTs$;
- $s_5 : X$, Ts $s_6 : Y$, Ts $s_7 : N$, Ts $s_8 : Tr$, Ts.

Please use this exact numbering in your answer to any questions below that refer to states.

(a) Give an intuitive justification why we should model this problem using *two* stages (i.e., why one stage or three or more stages are either insufficient or unnecessary).

(b) List all 1-stage-to-go conditional plans for this problem. Which of these are useful and which are pointwise dominated by some other 1-stage-to-go conditional plan? You can justify your response qualitatively (no need to describe their α-vectors).

(c) Consider the following 10 2-stage-to-go conditional plans:

- $P1$: Do test T_1. If H, do M_2, if L, do M_1.
- $P2$: Do test T_1. If H, do M_2, if L, do $Null$.
- $P3$: Do test T_1. If H, do M_2, if L, do M_3.
- $P4$: Do test T_2. If Y, do M_1, if N, do $Null$.
- $P5$: Do test T_2. If Y, do M_2, if N, do $Null$.
- $P6$: Do test T_2. If Y, do M_3, if N, do $Null$.
- $P7$: Do test T_1. If H, do M_1, if L, do M_2.
- $P8$: Do test T_1. If H, do M_1, if L, do M_1.
- $P9$: Null. Do M_1.
- $P10$: Null. Do M_3.

For each of these ten plans, give their α-vectors *restricted to states* s_1, s_2, s_3, s_4. Give each α-vector in the form: $[v(s_1), v(s_2), v(s_3), v(s_4)]$ in *that order*. You may list them in a table, using column or row vectors, etc., but be sure that however you do it, the components are listed in the proper order. Be sure to label your α-vectors consistent with the conditional plans, i.e., α_1, α_2, etc.

Tips: Don't try to compute these by hand: a simple script or even spreadsheet table will allow your to compute these values quickly, since they use many of the same structural ingredients.

(d) Three of the conditional plans in the set above are pointwise dominated by another conditional plan in the set (hence these three plans are useless no matter what the doctor's belief state is). Identify them and, for each, state which other plan pointwise dominates them. (As above, belief states give probability zero to any of $s_5, \ldots, s_8$, so you may ignore these.)

(e) For each of the seven conditional plans that is not pointwise dominated, describe a belief state (over the four reachable states) for which that plan is better than any of the others in this set. *Hint: you need only consider belief states that assign probability 0 to s_4.)* Justify your answer by stating the expected value of all seven nondominated conditional plans for the belief state chosen: use a script or spreadsheet to produce a table or matrix showing the values for the seven computed belief states for each of the seven plans.

(f) None of the plans above is optimal for state s_4. What 2-stage-to-go conditional plan *is* optimal for s_4? Why?

3. Suppose you are given an MDP $M = \langle S, A, \Pr, R \rangle$ and asked to determine the optimal value $V^*(s)$ for a specific state s, assuming a discounted, infinite-horizon optimality criterion, with discount rate β. We assume the reward function is non-negative, that $R^+ = \max\{R(s) : s \in S\}$, and that $R^- = \min\{R(s) : s \in S\}$. You are asked to build a search tree rooted at s in order to compute an estimate of this value $V^*(s)$. In the following you are only allowed to build the search tree and back up values *once* (i.e., you cannot build a search tree to a certain depth, compute an estimated value, and then build a deeper tree).

(a) To what depth will you need to build the tree in order to ensure that the backup up value at the root is within ε of $V^*(s)$? (In other words, give an (as tight as possible) *a priori* bound on the required depth.) Prove your bound.

(b) Suppose you are given a heuristic function $\widetilde{V}$ that can be used to evaluate the states at the leaves of the search tree. You are told that $\widetilde{V}(t)$ is within δ of $V^*(t)$ for all states t. To what *a priori* depth must you build the tree, using $\widetilde{V}$ at the leaves, to ensure and ε-accurate estimate of $V^*(s)$? Prove your bound.

(c) Suggest ways in which the heuristic function $\widetilde{V}$ might be used to prune the search tree and speed up the computation of $V^*(s)$.