APP MATH 3020 Stochastic Decision Theory
Assignment 4

Due: Wednesday, 17 October, 2018, 10 a.m. **Total marks: 33**

Question 1 2 marks

Make sure that in all your answers you

½ (a) use full and complete sentences.

½ (b) include units where necessary.

½ (c) use logical arguments in your answers and proofs.

½ (d) structure your answers and assignment clearly and precisely.

Question 2 10 marks

Each day, Squirrel the Singer is asked to take on a new singing gig. The gigs are independently distributed over 3 possible types; on a given day, the offered type is i with probability $\alpha_i \in (0,1)$ for $i = 1, \ldots, 3$. Upon completion, gigs of type i pay r_i dollars. Once Squirrel has accepted a gig, she may accept no other gigs until that gig is complete. The probability that a gig of type i takes k days is $(1-p_i)^{k-1}p_i$, for $k = 1, 2, \ldots$, where $p_i \in (0,1)$.

7 (a) Evaluate the average reward, g_1, of a stationary policy in which Squirrel accepts only gigs of type 1.

Solution: Let the state 0 be one in which Squirrel can select a new gig, otherwise let state 1 be one in which Squirrel is engaged in a gig of type 1. [1] Then, if we use the stationary policy that we choose only gigs of type 1, we have the following system of equations:

$$g_1 + \phi(0) = (1-\alpha_1)\phi(0) + \alpha_1\phi(1), \quad [1] \tag{1}$$

$$g_1 + \phi(1) = (1-p_1)\phi(1) + p_1\big(r_1 + \phi(0)\big), \quad [1] \tag{2}$$

where $\phi(i)$ is the value of starting in state i [1] , for $i = 0, 1$. Hence, we have two equations and three unknowns, so we set one unknown, $\phi(0)$, to 0. [1]

Subtracting Equation (1) from (2) gives

$$\begin{aligned} & \phi(1) = (1-p_1)\phi(1) + p_1 r_1 - a_1\phi(1) \\ \Leftrightarrow \quad & \phi(1) = \frac{p_1 r_1}{a_1 + p_1} \; [1] \\ \Leftrightarrow \quad & g_1 = \frac{a_1 p_1}{a_1 + p_1} r_1. \; [1] \end{aligned}$$

3 (b) Apply one step of the Policy Improvement Algorithm to determine an improved policy, clearly stating what the improved policy is.

Solution: Suppose we consider a different policy, where we accept only gigs of a certain type j.

There are two ways of approaching this. The first is to apply the Policy Improvement Algorithm *continuing* from Part (a). That is, evaluating the right-hand sides of (1) and (2) using the transition probabilities under *the new policy* and *the old values* of $\phi(i)$. However, it does not make sense to do that here, because—by our construction of the Markov chain in Part (a)—state 1 under the old policy (doing a gig of type 1) does not exist in the new policy. If we wanted to apply the PIA *continuing* from Part (a), then we would need to have a Markov chain with all possible states $0, 1, 2, 3$, representing *not working*, *working in type* 1, *working in type* 2, and *working in type* 3.

Alternatively, we can start the Policy Improvement Algorithm *again*, with a stationary policy where we choose gigs of type $j \neq 1$ only. Using a Markov chain with two states, 0 for *not working* and j for *working in type* j, and applying the same analysis in Part (a), we arrive at

$$g_j = \frac{a_j p_j}{a_j + p_j} r_j. \text{ [1]}$$

We accept this as an improved policy if $g_j > g_1$, [1] otherwise we retain our previous policy. [1]

Question 3 10 marks

Heather receives \$10 for every chess game that she wins. Playing costs her \$$c$ per hour. The total number of chess games that Heather can play is T. The probability of winning one game in the next hour is $\omega(r)$, where $\omega(r)$ is an increasing function of r, the remaining number of games. There is zero probability of winning more than one game in an hour. Heather wants to maximise her net expected profit. (There is zero probability of Heather ever losing or drawing.)

4 (a) Specify, with justification, Heather's stopping rule.

Solution: We use the one-step-look-ahead policy. The reward for stopping now is \$0, and for continuing for one time unit and then stop is $-c + 10\omega(x)$ [1] . Thus, the set $\mathcal{L}$ is

$$\mathcal{L} = \{r : 0 \geq -c + 10\omega(r)\}, \text{ [1]}$$

where r denotes the number of remaining games. Note that r is a decreasing function over time. On the other hand, we know that $\omega(r)$ is an increasing function in r. Thus, once we are in state r such that $10\omega(r) \leq c$ and therefore entering $\mathcal{L}$, we remain in $\mathcal{L}$ forever. This implies that the set $\mathcal{L}$ is closed. [1]

So the optimal policy is stop when $10\omega(x) \leq c$ [1] .

6 (b) If $T = 12$, $\omega(r) = 1 - e^{-r/5}$ and $c = \$0.5$, determine Heather's expected profit and detail the stopping rule.

Solution: Solving $10(1 - e^{-r/5}) \leq 0.5$ for r, we have

$$r \leq -5 \log(0.95) \approx 0.2565,$$

thus we stop once there are 0 games remaining! [1]

The probability of winning one game per unit time is $1 - e^{-r/5}$, so the expected time for winning a game is $1/(1 - e^{-r/5})$ when there are r games remaining. [1]
Thus, the expected time to win all 12 is

$$\sum_{r=1}^{12} \frac{1}{1 - e^{-r/5}} \approx 22.7544. \text{ [1]}$$

The return for winning 12 games \$120. [1]
The expected cost for winning 12 games is

$$c \sum_{r=1}^{12} \frac{1}{1 - e^{-r/5}} \approx \$11.3772. \text{ [1]}$$

Consequently, Heather's expected profit under the optimal stopping rule is

$$120 - 11.3772 = \$108.6228. \text{ [1]}$$

Question 4 $\boxed{\text{11 marks}}$

You are moving overseas soon! Suppose you need to sell your car (a twenty-year-old aqua Mirage) and have 10 weeks in which to advertise and sell it. You receive one offer per week; these offers are independent with a value of j dollars with probability p_j, for $j = 1, \dots, 10$. Any offer not immediately accepted, can be accepted at a later date. Every week that the Mirage remains unsold, it costs you c dollars per week.

The state space is $\mathcal{S} = \{1, 2, \dots, 10\}$, where state i corresponds to the highest offer to date. There are only two actions you might take when in state i, to either accept the best offer to date with value i or not accept the best offer to date and continue with costs c.

$\boxed{6}$ (a) Give the transition probabilities when continuing on and not accepting the best offer to date, with justification.

Solution: The transition probabilities [3] are

$$p_{ij} = \begin{cases} p_j & \text{for} \quad j > i, \\ 1 - \displaystyle\sum_{k=i+1}^{10} p_k & \text{for} \quad j = i, \\ 0 & \text{for} \quad j < i. \end{cases}$$

The transition probabilities rely on the current offer j; if it is higher than the previous highest offer i, the state changes to the current offer j, the highest offer to date, with probability p_j for any $j > i$ [1] .
The state can never reduce as the process changes only if the offer increases [1] .
Hence, there is probability $\sum_{k=i+1}^{10} p_k$ of changing to a higher offer; otherwise, the state remains the same with the residual probability [1] .

$\boxed{5}$ (b) What is the optimal policy for selling your car?

Solution: We look for a closed set $\mathcal{L}$, where the first time we enter that set, we accept the highest offer i so far. At each step, if we accept the highest offer i so far, we obtain a value of i [1] . On the other hand, if we wait for one more step and then accept the highest offer i so far, we incur a cost of c, and then sell the house at the highest offer so far at the expected value

$$\sum_{j>i} jp_j + i\left(1 - \sum_{j>i} p_j\right).$$

This is because we either get a higher offer j with probability p_j (for each of those $j > i$), or get the same or a lower offer with probability $1 - \sum_{k>i} p_k$ and so the highest offer remains at i. [1]

Hence,

$$\mathcal{L} = \left\{i : i \geq -c + \sum_{j>i} jp_j + i\left(1 - \sum_{j>i} p_j\right)\right\}. \text{ [1]}$$

We can also write

$$\mathcal{L} = \left\{i : c \geq \sum_{j>i}(j-i)p_j\right\}, \tag{3}$$

and note that the expression on the RHS of (3) is monotonically decreasing in i. So this is a closed set. [1 mark for justification of $\mathcal{L}$ being a closed set]

If we let

$$i_0 := \min\left\{i : c \geq \sum_{j>i}(j-i)p_j\right\},$$

the policy is to accept the first offer that is at least i_0 or greater. (Of course if we are at the end of Week 10 then we should accept whatever that is the best offer at the time.) [1 mark for statement of policy]