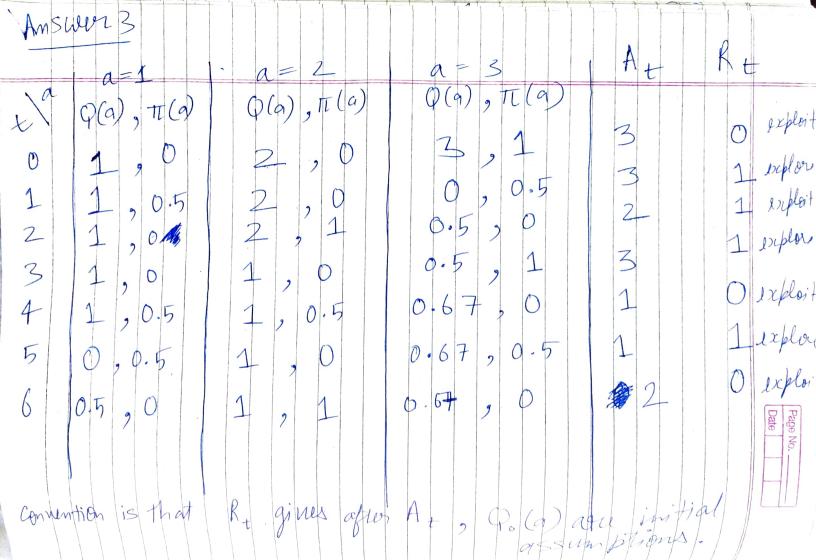
0JUS SIN PLOGNETAL
202009 Date

Answer 1: From the description, we can form the policy IT as: $\frac{2}{15}$, $a \in \{2,4,6,8\}$, a E (1,3,5,7,96 Now, $E[R_1+R_2+...+R_{10}] = \sum_{i=1}^{n} E[R_i]$ = 10 E/R1 (This is tyw Since the policy is constant) and E[Ri] is only dependent on policy) in this case =10 TL(a)where q(a) is the expected reward $=10\left(\frac{2(2+4+6+8+10)+1(1+3+5+7)}{15(1+3+5+7)}\right)$ = |56.67|

State of the state		Page No Date
	Anse	ver 2
		From the description we have
		$q(a) = \begin{cases} 0.5, a \in \{1,2,4,5,7,9\} \end{cases}$
_		0.46, da ae {3,6,8}
		A simple stochastic bolicy maximizing the expected reward is:
		$\pi(a) = \begin{cases} 1, & a = 1 \end{cases}$ $0, & a \in \{2,3,4,,9\}$
		Similarly, we can have 5 other optimal stochastic policies where 1 (a) = 1 where a = 2 or 4 or 5 or 7 or 9 and 0 otherwise.



Amsher 6 (X. Lington Sylvanos) Let $G_t = R_t + C + Y G_{t+1}$ In terms of Gt, Gt is: $G_{t} = \sum_{b=0}^{7} \begin{cases} k \\ k \end{cases} \begin{pmatrix} k \\ + k \end{pmatrix}$ $\frac{1}{G_t} = \frac{1}{C} \begin{cases} k + \sum_{k=0}^{\infty} k^k \\ k = 0 \end{cases}$ G = C + G - IPutting this in the Bellman Eg, $V_{\pi}(s) = E\left[G_{t} \mid S_{t} = s\right]$ = E[R+c+1G+1S=s]

= C + E [R++ Y G++s | S = 5] Vn (5) $= \frac{C}{1-Y} + V_{\pi}(S)$ Mence Vc = c 1-x Hence the signs of the rewards art.

	Moweuer, for episodic tasks,
	$\overline{C}_{t} = \sum_{k=0}^{T-t-1} \sum$
	$G_{+} = C \left(1 - Y^{T-+}\right) + G_{+}$ $1 - Y$
	Here the additional factor is.
	dependent on the current time-step, hence the constant can change the dynamics of the problem.
	The agent will try to increase the number of the Steps just to gain this constant reward, but we want the time-steps to be liss.
	Hence adding constants is not agricultant in episodics tasks.
Amsu	$\frac{\text{vr8}}{\text{v}*(S)} = \max_{a} \mathcal{Q}_*(S, a)$

Page No. _

Date