(1A). (a). As the prob. are not given, we can assume equal probabilities and hence arm 3 is most preferred and arm 1 is least preferred.

(b). $E[r_1] \cong 5.17$; $E[r_2] = 4.88$; $E[r_3] \cong 4.32$

(c). Arm 1 is best as it has highest expected reward.

(d). Write either ϵ-greedy | UCB discussed in class.

(2A).

(a).

State	Action	Next State	Prob.	Reward
R	M	R	0.6	-20+100
R	M	B	0.4	-20
R	NM	R	0.3	100
R	NM	B	0.7	0
B	Re	R	0.6	-40+100
B	Re	B	0.4	-40
B	RP	R	1	-150+100

Here, Running - R ; broken - B ; maintanence - M, no maintance - NM, repair - Re ; replace - RP.

(b).

$$v^*(R) = \max_{\{M, NM\}} \begin{cases} 0.6[80 + \gamma v^*(R)] + 0.4[-20 + \gamma v^*(B)], \\ 0.3[100 + \gamma v^*(R)] + 0.7[0 + \gamma v^*(B)] \end{cases}$$

$$v^*(B) = \max_{\{Re, RP\}} \begin{cases} 0.6[60 + \gamma v^*(R)] + 0.4[-40 + v^*(B)], \\ 1[-50 + \gamma v^*(R)] \end{cases}.$$

(3A). (a). $G_0 = 1 + 2\gamma + (2\gamma)^2 + (2\gamma)^3 + \cdots = \frac{1}{1-2\gamma} = 5.$

$G_1 = 2 + 4\gamma + 8\gamma^2 + 16\gamma^3 + \cdots$

$= 2(1 + 2\gamma + (2\gamma)^2 + (2\gamma)^3 + \cdots) = 2\left(\frac{1}{1-2\gamma}\right) = 10.$

(b). $G_0 = 2 + 4\gamma + 2\gamma^2 + 4\gamma^3 + 2\gamma^4 + \cdots$

$= 2(1 + \gamma^2 + \gamma^4 + \cdots) + 4(\gamma + \gamma^3 + \cdots)$

$= 2\left(\frac{1}{1-\gamma^2}\right) + 4\left(\frac{\gamma}{1-\gamma^2}\right) = \frac{5.2}{0.36} \approx 14.45$

$G_1 = 4 + 2\gamma + 4\gamma^2 + \cdots$

$= 4(1 + \gamma^2 + \cdots) + 2(\gamma + \gamma^3 + \cdots)$

$= 4\left(\frac{1}{1-\gamma^2}\right) + 2\left(\frac{\gamma}{1-\gamma^2}\right) = \frac{5.6}{0.36} \approx 15.56.$