Chapter-4

① $q_\pi(11, \text{down}) = -1 + V_\pi(T)$

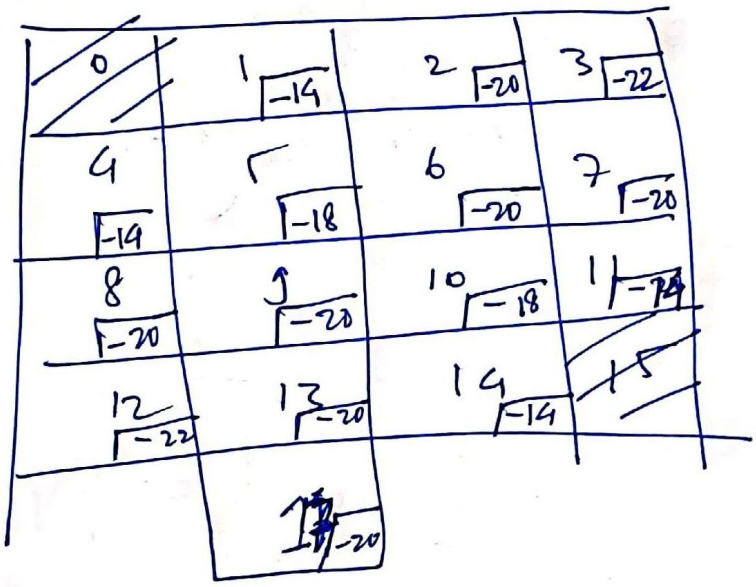
$= -1$

$q_\pi(7, \text{down}) = -1 + V_\pi(11)$

$= -1 - 14 = -15$

② Let the states look like this for consistency in code.

For 2nd case:



$V_\pi(15) = -20$

For 1st case:

$V_\pi(15) = -20$

(Please refer to code)

Ex, 4.3

~~$q_\pi(s) = E_\pi[R_{t+1}$~~

$$q_\pi(s,a) = \sum_{s'} p(s'|s,a)\left\{E[r|s',s,a] + \gamma V_\pi(s')\right]$$

$$= \sum_{s'} p(s'|s,a)\left\{E[r|s',s,a] + \gamma \sum_{a'} \pi(a'|s') q(s',a')\right\}$$

$$q_{k+1}(s,a) = \sum_{s'} p(s'|s,a)\left\{E[r|s',s,a] + \gamma \sum_{a'} \pi(a'|s') q_k(s',a')\right\}$$

Ex 4.8 Because betting \$1 is better when you have \$51. As if you loose then you could bet \$50. rather than betting all \$51 at once. (Note: Agent gets reward when it has total of \$100).

Ex 4.9 Please refer to the code

Ex 4.10

$$q_{k+1} = \sum p(s'|s,a)\left[r + \gamma \max_{a'} q_*(s',a')\right]$$

$$\boxed{q_{k+1} = \sum p(s'|s,a)\left[r + \gamma \max_{a'} q_k(s',a')\right]}$$