CHAPTER 3: FINITE MARKOV DECISION PROCESS

Exercise 3.11.

$$\begin{aligned}
\mathbb{E}(R_{t+1}|S_t) &= \sum_{r_{t+1}} r_{t+1} p(r_{t+1}|S_t) \\
&= \sum_{r_{t+1}} r_{t+1} \sum_{S_{t+1}} \sum_{A_t} p(r_{t+1}, S_{t+1}, A_t|S_t) \\
&= \sum_{r_{t+1}} \sum_{s'} \sum_{a} r_{t+1} p(r_{t+1}, S_{t+1} = s'|S_t = s, A_t = a)\pi(a|s)
\end{aligned}$$

Exercise 3.12.

$$\begin{aligned}
\nu_\pi(s) &= \mathbb{E}(G_t|S_t = s) \\
&= \sum_{g_t} g_t p(g_t|S_t = s) \\
&= \sum_{g_t} g_t \sum_{a} p(g_t, A_t = a|S_t = s) \\
&= \sum_{g_t} g_t \sum_{a} p(g_t, |S_t = s, A_t = a)p(A_t = a|S_t = s) \\
&= \sum_{a} \left(\sum_{g_t} g_t p(g_t, |S_t = s, A_t = a) \right) \pi(a|s) \\
&= \sum_{a} q_\pi(s, a)\pi(a|s)
\end{aligned}$$

Exercise 3.13. Give an equation for q_π in terms of v_π and the four-argument p.

$$
\begin{aligned}
q_\pi(s,a) &= \mathbb{E}(G_t|S_t=s,A_t=a) \\
&= \mathbb{E}(R_t+\gamma G_{t+1}|S_t=s,A_t=a) \\
&= \mathbb{E}(R_t|s,a)+\gamma\mathbb{E}(G_{t+1}|s,a) \\
&= \sum_r rp(r|s,a)+\gamma\sum_{g_{t+1}} g_{t+1}p(g_{t+1}|s,a) \\
&= \sum_r\sum_{s'} rp(r,s'|s,a)+\gamma\sum_{g_{t+1}}\sum_{s'} g_{t+1}p(g_{t+1},s'|s,a) \\
&= \sum_r\sum_{s'} rp(r,s'|s,a)+\gamma\sum_{g_{t+1}}\sum_{s'} g_{t+1}p(g_{t+1}|s',s,a)p(s'|s,a) \\
&= \sum_r\sum_{s'} rp(r,s'|s,a)+\gamma\sum_{g_{t+1}}\sum_{s'} g_{t+1}p(g_{t+1}|s')\sum_r p(r,s'|s,a) \\
&= \sum_r\sum_{s'} \left(r+\gamma\nu_\pi(s')\right)p(r,s'|s,a)
\end{aligned}
$$

Exercise 3.17. Bellman equation for action values.

$$
\begin{aligned}
q_\pi(s,a) &= \mathbb{E}(G_t|S_t=s,A_t=a) \\
&= \mathbb{E}(r+\gamma G_{t+1}|s,a) \\
&= \sum_{s'}\sum_r p(r,s'|s,a)\left(r+\gamma\mathbb{E}(G_{t+1}|r,s',s,a)\right) \\
&= \sum_{s'}\sum_r p(r,s'|s,a)\left(r+\gamma\mathbb{E}(G_{t+1}|s')\right) \\
&= \sum_{s'}\sum_r p(r,s'|s,a)\left(r+\gamma\sum_{a'}p(a'|s')\mathbb{E}(G_{t+1}|s',a')\right) \\
&= \sum_{s'}\sum_r p(r,s'|s,a)\left(r+\gamma\sum_{a'}\pi(a'|s')q_\pi(s',a')\right)
\end{aligned}
$$