1.

(a) State Space: $\mathcal{X} = \{0, 1, 2, 3, 4, 11, 12, 22\}$, which means the money we have, e.g. if $x = 1$ then we have \$10.000. Then terminal states are $\mathcal{T} = \{x \leq 0 \text{ or } x \geq 3 \mid x \in \mathcal{X}\}$.

Control Space: $\mathcal{U} = \{1(r), 0(b)\} \times \{m\} \subseteq \mathbb{R}^2$, where 1 means betting on red and 0 means betting on black. m is integer and $m \geq 1$, this is the money we bet, e.g. if $m = 2$ then we bet \$20.000 this round.

Motion model: If $x_t \in \mathcal{X} \setminus \mathcal{T}$

$$x_{t+1} \mid x_t, u_t = (1, m) = \begin{cases} x_t + m, & \text{with prob. } 0.7 \\ x_t - m, & \text{with prob. } 0.3 \end{cases}$$

$$x_{t+1} \mid x_t, u_t = (0, m) = \begin{cases} x_t + 10m, & \text{with prob. } 0.2 \\ x_t - m, & \text{with prob. } 0.8 \end{cases}$$

If $x_t \in \mathcal{T}$,

$$x_{t+1} \mid x_t, u_t = x_t, \text{ i.e. } \begin{cases} p_f(x_t \mid x_t, u_t) = 1 \\ p_f(x_{t+1} \neq x_t \mid x_t, u_t) = 0 \end{cases}$$

Stage cost: $\ell(x, u, x') = -(x' - x)$

i.e. Once we get into terminal costs, we get $-x'$ reward.

Then $$\tilde{\ell}(x, u) = -\sum_{x'} p_f(x' \mid x, u)(x' - x)$$

Terminal cost: $q(x_T) = -x_T$

(b) $V^*(x) = \min_{u \in \mathcal{U}(x)} \tilde{\ell}(x,u) + \mathbb{E}_{x' \sim p_f(\cdot|u,x)}[V^*(x')]$

In this case, $\pi = u_t = (1, 1)$. then

$\mathcal{X} = \{0, 1, 2, 3\}$. terminal states $\mathcal{T} = \{0, 3\}$

Suppose at iteration $k=0$. $V_0^\pi(x) = 0$ for all $x \in \mathcal{X} \setminus \mathcal{T}$

$V_0^\pi(x_T) = -x_T$ for all $x_T \in \mathcal{T}$.

When $k=1$.

$$V^\pi(x) = \tilde{l}(x,u,x') + \sum_{x' \in \mathcal{X}} p_f(x'|x,u) V^\pi(x')$$

By program,

$\bar{V}^\pi = [0, -0.4, -2.5, -3, -4, -11, -12, -22]^T$

for $x = 0, 1, 2, 3, 4, 11, 12, 22$

Also, we get precise estimate

$\hat{V}^\pi = [0, -2.7215, -3.3165, -3, -4, -11, -12, -22]^T$

for $x = 0, 1, 2, 3, 4, 11, 12, 22$

(c) State space $\mathcal{X} = \{0, 1, 2, 3, 4, 11, 12, 22\}$, $\mathcal{T} = \{0, 4, 11, 12, 22\}$

Given $\bar{V}^\pi$, using computer we get

$\bar{\pi}'(1) = (0, 1)$, $\bar{\pi}'(2) = (0, 2)$, $\bar{\pi}'(x) = \{\text{go home}\}$, $x \in \mathcal{T}$.

which both means to bet on black and bet all money.

Given $\hat{V}^\pi$, using computer we get

$\hat{\pi}'(1) = (0, 1)$, $\hat{\pi}'(2) = (0, 2)$, $\hat{\pi}'(x) = \{\text{go home}\}$ $x \in \mathcal{T}$.

so if we have \$10,000, we bet on black and bet \$10,000;
if we have \$20,000, we bet on black and bet \$20,000.

$\bar{\pi}'$ and $\hat{\pi}'$ are the same. Compared to $\pi = (1, 1)$, the policy $\bar{\pi}'$, $\hat{\pi}'$ become bolder, trying to earn more money.

(d) Using Policy Iteration (PI), we'll have

$V^*(1) = -6.6364$, $\pi^*(1) = (1, 1) \longrightarrow$ bet red, \$10,000

$V^*(2) = -8.9091$, $\pi^*(2) = (0, 1) \longrightarrow$ bet black, \$10,000

$V^*(x) = x$, $\pi^*(x) = \{stay\}$ for $x \in 7$.

Therefore, we bet red, \$10,000 when we have \$10,000; and we bet black, \$10,000 when we have \$20,000.

Under this policy, I simulate the gambling for 50,000 times, and the average money I take home is \$38.614.

Celebration!