MAS5

December 4, 2020

Leonard Vorbeck

No Group

Student Number : 2709813

[1]:
```
import pandas as pd
import numpy as np
from scipy.stats import pearsonr as r
import matplotlib.pyplot as plt
from scipy.stats import norm, uniform
%config InlineBackend.figure_format = 'retina'
plt.rcParams["figure.dpi"] = 93
plt.rcParams["figure.figsize"] = (12,7)
plt.style.use("default")
```

0.0.1 5.1 Bellman Equations

$$\mathbf{I} \ : \quad v_\pi(s) = \sum_a \pi(a|s) q(a,s) \mathbf{II} \ : \quad q_\pi(s,a) = \sum_{s'} p(s'|s,a)[r(s,a,s') + \delta v(s')]$$

For policy

$$\pi(a \mid s) = \begin{cases} 1 & \text{if } a = a_s \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{I}_\pi \ : \quad v_\pi(s) = \sum_{a_i \in A} \pi(a_i|s) q(a_i, s)$$

$$v_\pi(s) = \pi(a_1|s)q(a_1,s) + \pi(a_2|s)q(a_2,s) + \cdots + \pi(a_s|s)q(a_s,s) + \cdots + \pi(a_n|s)q(a_n,s)$$

$$v_\pi(s) = 0 + 0 + \cdots + q(a_s,s) + 0$$

$$v_\pi(s) = q(s,a_s)$$

and

$$\mathbf{II}_\pi \ : \quad q_\pi(s,a) = \sum_{s'} p(s'|s,a)[r(s,a,s') + \delta v(s')]$$

Since $v(s') = q(s', a_{s'})$:

$$\mathbf{II}_\pi \ : \quad q_\pi(s,a) = \sum_{s'} p(s'|s,a)[r(s,a,s') + \delta q(s', a_{s'})]$$

Now let

$$p\left(s' \mid s,a\right) = \begin{cases} 1 & \text{if } s' = s_a \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\mathbf{II}_\pi \ : \quad q_\pi(s,a) = \sum_{s'} p(s'|s,a)[r(s,a,s') + \delta v(s')]$$

$$q_\pi(s,a) = 0 + 0 + \cdots + p(s_a|s,a)[r(s,a,s_a) + \delta v(s_a)] + \cdots + 0$$

$$q_\pi(s,a) = r(s,a,s_a) + \delta v(s_a))$$

Since $v(s_a) = q(s_a, a_{s_a})$:

$$q_\pi(s,a) = r(s,a,s_a) + \delta q(s_a, a_{s_a}))$$

0.0.2 2.1 MDP 1

Let

$$A = \{a_+, a_-\} S = \{0, 1, 2, \ldots, n\}$$

and

$$v_\pi(s) = \sum_a \pi(a \mid s) \sum_{s'} p\left(s' \mid s,a\right) \left[r\left(s,a,s'\right) + \delta v_\pi\left(s'\right)\right]$$

Equivalently

$$v_\pi(s) = \delta \sum_{s'} P(s,s') v_\pi(s') + \sum_a R(s,a)\pi(a \mid s)$$

with

$$P(s,s') = \sum_a \pi(a \mid s) p\left(s' \mid s, a\right)$$

$$R(s,a) = \sum_{s'} p\left(s' \mid s, a\right) r\left(s, a, s'\right)$$

Now let

$$r(s) = \sum_a \{\sum_{s'} p\left(s' \mid s, a\right) r\left(s, a, s'\right)\} \ \pi(a \mid s)$$

$$r(s) = \sum_a R(s,a)\pi(a \mid s)$$

Such that

$$v_\pi(s) = \delta \sum_{s'} P(s,s') v_\pi(s') + r(s)$$

Now in matrix form:

$$\mathbf{v} = \delta \ \mathbf{P}\mathbf{v} + \mathbf{r}$$

with $\mathbf{v}, \mathbf{r} \in \mathcal{R}^n$ and $\mathbf{P} \in \mathcal{R}^{n \times n}$ for n states.

Rewrite to

$$\mathbf{r} = \mathbf{v} - \delta \mathbf{P}\mathbf{v}$$

$$\mathbf{r} = (\mathbf{I} - \delta \mathbf{P})\mathbf{v}$$

$$\mathbf{v} = (\mathbf{I} - \delta \mathbf{P})^{-\mathbf{1}}\mathbf{r}$$

We can solve this system of l.eq. in the following way:

First, we define $\mathbf{P}$, whose elements are given by

$$P(s,s') = \sum_a \pi(a \mid s) p\left(s' \mid s, a\right)$$

First,

$$\mathbf{P_{s,s}} := \mathbf{P(s,s)} = \mathbf{0} \forall \ \mathbf{s}$$

For adjacent states we have the two cases

$$P(s, s+1) = \pi(a_+ \mid s) p(s+1 \mid s, a_+) + \pi(a_- \mid s) p(s+1 \mid s, a_-)$$

$$P(s, s+1) = \frac{1}{2}p(s+1 \mid s, a_+) + \frac{1}{2}p(s+1 \mid s, a_-)$$

$$P(s, s+1) = \frac{1}{2}1 + \frac{1}{2}0 = \frac{1}{2}$$

and

$$P(s, s-1) = \pi(a_+ \mid s)p(s-1 \mid s, a_+) + \pi(a_- \mid s)p(s-1 \mid s, a_-)$$
$$P(s, s-1) = \frac{1}{2}p(s-1 \mid s, a_+) + \frac{1}{2}p(s-1 \mid s, a_-)$$
$$P(s, s-1) = \frac{1}{2}0 + \frac{1}{2}1 = \frac{1}{2}$$

Such that we can define $\mathbf{P}$ as follows:

$$\mathbf{P_{s,s'}} = \mathbf{0} \text{ if } \mathbf{s} = \mathbf{0}$$
$$\mathbf{P_{s,s'}} = \mathbf{0} \text{ if } \mathbf{s'} = \mathbf{s}$$
$$\mathbf{P_{s,s'}} = \mathbf{0.5} \text{ if } \mathbf{s'} = \mathbf{s+1}$$
$$\mathbf{P_{s,s'}} = \mathbf{0.5} \text{ if } \mathbf{s'} = \mathbf{s-1}$$
$$\mathbf{P_{s,s'}} = \mathbf{0} \text{ if } \mathbf{s'} < \mathbf{s-1}$$
$$\mathbf{P_{s,s'}} = \mathbf{0} \text{ if } \mathbf{s'} > \mathbf{s+1}$$

Now we can define $\mathbf{K} = \mathbf{I} - \delta\mathbf{P}$. For $\delta = 1$ this will be:

$$\mathbf{K_{s,s'}} = -\mathbf{P_{s,s'}} \text{ if } \mathbf{s'} \neq \mathbf{s}$$
$$\mathbf{K_{s,s'}} = \mathbf{1} - \mathbf{P_{s,s'}} = \mathbf{1} \text{ if } \mathbf{s'} = \mathbf{s}$$

Now for the immediate expected reward $\mathbf{r}$, each element is calculated via:

$$r(s) = \sum_a \{\sum_{s'} p\left(s' \mid s, a\right) r\left(s, a, s'\right)\} \ \pi(a \mid s) = \sum_a R(s, a)\pi(a \mid s)$$

For example

$$r(1) = R(1, a_+)\pi(a_+ \mid s) + R(1, a_-)\pi(a_- \mid s)$$

with

$$R(1, a_+) = r(1, a_+, 2)p(2 \mid 1, a_+) + r(1, a_+, 0)p(0 \mid 1, a_+) = r(1, a_+, 2) = 0$$
$$R(1, a_-) = r(1, a_-, 2)p(2 \mid 1, a_-) + r(1, a_-, 0)p(0 \mid 1, a_-) = r(1, a_-, 0) = 10$$

s.t

$$r(1) = 0\frac{1}{2} + 10\frac{1}{2} = 5$$

which is anti-symmetric in a with $r(n)$, $R(n-1, a_+) = R(1, a_-)$, because of the symmetry in r:

$$r(1) = r(n-1) = 5r(s) = 0 \ \forall s \notin \{1, n\}$$

Finally

$$\mathbf{v}_\pi \leftarrow \mathbf{K}^{-1}\mathbf{r}$$

which yields $\mathbf{v}_\pi \in \mathcal{R}^{\mathbf{n+1}} \approx [\mathbf{0}, \mathbf{10}, \mathbf{10}, \ldots, \mathbf{10}]^{\mathbf{T}}$. Equivalently when omitting the terminal node 0 : $\mathbf{v}_\pi \in \mathcal{R}^{\mathbf{n}} \approx [\mathbf{10}, \mathbf{10}, \ldots, \mathbf{10}]^{\mathbf{T}}$

Without using the potentially expensive inverse, the values can be derived explicitly :

$$q_\pi(s, a) = \sum_{s'} p(s' \mid s, a)[r(s, a, s') + \delta v_\pi(s')]$$

$$q_\pi(s, a) = p(s+1 \mid s, a)[r(s, a, s+1) + \delta v_\pi(s+1)] + p(s-1 \mid s, a)[r(s, a, s-1) + \delta v_\pi(s-1)]$$

Which is

$$q_\pi(s, a_+) = r(s, a, s+1) + \delta v_\pi(s+1)$$
$$q_\pi(s, a_-) = r(s, a, s-1) + \delta v_\pi(s-1)$$

We see that

$$q_\pi(s, a_+) = 0 + \delta v_\pi(s+1) \ \forall s \neq n$$
$$q_\pi(s, a_-) = 0 + \delta v_\pi(s-1) \ \forall s \neq 1$$
$$q_\pi(n, a_+) = 10 + 0$$
$$q_\pi(1, a_-) = 10 + 0$$

Now

$$v_\pi(s) = \sum_a \pi(a \mid s) q(s, a)$$

$$v_\pi(s) = \pi(a_+ \mid s) q(s, a_+) + \pi(a_- \mid s) q(s, a_-)$$

$$v_\pi(s) = \pi(a_+ \mid s)[r(s, a, s+1) + \delta v_\pi(s+1)] + \pi(a_- \mid s)[r(s, a, s-1) + \delta v_\pi(s-1)]$$

$$v_\pi(s) = \frac{1}{2}[r(s, a, s+1) + \delta v_\pi(s+1)] + \frac{1}{2}[r(s, a, s-1) + \delta v_\pi(s-1)]$$

E.g

$$v_\pi(1) = \frac{1}{2}[0+\delta v_\pi(2)]+\frac{1}{2}[10+0] v_\pi(2) = \frac{1}{2}[0+\delta v_\pi(3)]+\frac{1}{2}[0+\delta v_\pi(1)] \dots v_\pi(n) = \frac{1}{2}[10+0]+\frac{1}{2}[0+\delta v_\pi(n-1)]$$

which is the same system of linear equations.

0.0.3 5.1.2

There is no unique optimal policy since any policy will yield

$$E_\pi[v(s)] = 10 = v^* E_\pi[q(s)] = 10 = q^*$$

given $r_{NT} = 0$

0.0.4 5.1.3

If $r_{NT} = -1 $, there is an optimal policy. This policy needs to map from S to A in a way such that the states visited from s to 0 are minimized.

$$\pi^*(s) = a_- \text{ if } s \leq \frac{n}{2} \text{ else } a_+$$

with

$$v^*_\pi(s) = 10 - s \text{ if } s \leq \frac{n}{2} \text{ else } 10 - (n - s + 1)$$

This produces interesting q values. For tuples (s, a_+) we will see decreasing values until $\frac{n}{2}$, but then increasing again until $q(n,a_{+}) = q(1,a_{-}) = 10 $. This is inversely proportional to a_- of course.

0.0.5 5.1.4

For $r_{NT} = 0 $ and $\delta < 1$, the same policy is applied to get optimal results.

$$\pi^*(s) = a_- \text{ if } s \leq \frac{n}{2} \text{ else } a_+$$

with

$$v^*_\pi(s) = \delta^{s-1}10 \text{ if } s \leq \frac{n}{2} \text{ else } \delta^{n-s}10$$

$\widehat{}$ For example

$$v^*_\pi(3) = 0 + [\delta[0 + \delta[10 + \delta 0]]] = \delta^2 10$$

0.0.6 5.1.5

For $r_{NT} = -1 $ and an odd number of nodes the optimal policy would not change for all nodes except for the node in the middle of the circle. In this node the probability becomes $\frac{1}{2}$ for each action.

0.0.7 5.2.1

Let

$$A = \{a_+, a_-\}$$
$$S = \{1, 2, 3, 4, 5, 6\}$$

$$q_\pi(s, a) = \sum_{s'} p(s' \mid s, a)[r(s, a, s') + \delta v_\pi(s')]$$
$$q_\pi(s, a) = p(s+1 \mid s, a)[r(s, a, s+1) + \delta v_\pi(s+1)] + p(s-1 \mid s, a)[r(s, a, s-1) + \delta v_\pi(s-1)]$$

Which is

$$q_\pi(s, a_+) = r(s, a, s+1) + \delta v_\pi(s+1)$$
$$q_\pi(s, a_-) = r(s, a, s-1) + \delta v_\pi(s-1)$$

Given the symmetry of the problem, it is sufficient to evaluate the policy for $s \in \{1, 2, 3\}$.

$$q_\pi(1, a_-) = 20$$
$$q_\pi(1, a_+) = 0 + \delta v_\pi(2)$$
$$q_\pi(2, a_-) = 0 + \delta v_\pi(1)$$
$$q_\pi(2, a_+) = 0 + \delta v_\pi(3)$$
$$q_\pi(3, a_-) = 0 + \delta v_\pi(2)$$
$$q_\pi(3, a_+) = 0$$

Now

$$v_\pi(s) = \sum_a \pi(a \mid s) q(s, a)$$
$$v_\pi(s) = \pi(a_- \mid s) q(s, a_-) + \pi(a_+ \mid s) q(s, a_+)$$

yields

$$v_\pi(1) = \frac{1}{2} 20 + \frac{1}{2} \delta v_\pi(2)$$

$$v_\pi(2) = \frac{1}{2}\delta v_\pi(1) + \frac{1}{2}\delta v_\pi(3)$$
$$v_\pi(3) = \frac{1}{2}\delta v_\pi(2)$$

For $\delta = 1$:

$$v_1 = 10 + \frac{1}{2}v_2 v_2 = \frac{1}{2}v_1 + \frac{1}{2}S_3 v_3 = \frac{1}{2}v_2$$

yields $\mathbf{v}_\pi = [\mathbf{15}, \mathbf{10}, \mathbf{5}]$

0.0.8 5.2.2

The optimal policy is unqiue with

$$\pi^*(s) = a_- \text{ if } s \in \{1, 2, 3\} \text{ else } a_+$$

Again, solving for $\{1, 2, 3\}$ is sufficient.

$$q_{\pi^*}(s, a_+) = r(s, a, s+1) + \delta v_{\pi^*}(s+1)$$
$$q_{\pi^*}(s, a_-) = r(s, a, s-1) + \delta v_{\pi^*}(s-1)$$

Actually, we only need those:

$$q_{\pi^*}(1, a_-) = 20 + 0$$
$$q_{\pi^*}(2, a_-) = 0 + \delta v_\pi(1)$$
$$q_{\pi^*}(3, a_-) = 0 + \delta v_\pi(2)$$

$$v_{\pi^*}(s) = \sum_a \pi^*(a \mid s) q(s, a)$$
$$v_{\pi^*}(s) = q(s, a_-)$$

yields

$$v_{\pi^*}(1) = 20 + 0$$
$$v_{\pi^*}(2) = 0 + \delta v_\pi(1)$$
$$v_{\pi^*}(3) = 0 + \delta v_\pi(2)$$

yields $\mathbf{v}_\pi = [\mathbf{20}, \mathbf{20}, \mathbf{20}]$ for $\delta = 1$.

0.0.9 5.2.3

Let $r_{NT} = -1$.

Again unique:

$$\pi^*(s) = a_- \text{ if } s \in \{1, 2, 3\} \text{ else } a_+$$

maximizes the expected reward (i.e is optimal)

Now

Again, solving for $\{1, 2, 3\}$ is sufficient.

$$q_{\pi^*}(s, a_+) = r(s, a, s+1) + \delta v_{\pi^*}(s+1)$$
$$q_{\pi^*}(s, a_-) = r(s, a, s-1) + \delta v_{\pi^*}(s-1)$$

$$q_{\pi^*}(1, a_-) = 20 + 0$$
$$q_{\pi^*}(2, a_-) = -1 + \delta v_{\pi}(1)$$
$$q_{\pi^*}(3, a_-) = -1 + \delta v_{\pi}(2)$$

$$v_{\pi^*}(s) = \sum_a \pi^*(a \mid s) q(s, a)$$
$$v_{\pi^*}(s) = q(s, a_-)$$

yields

$$v_{\pi^*}(1) = 20 + 0$$
$$v_{\pi^*}(2) = -1 + \delta v_{\pi}(1)$$
$$v_{\pi^*}(3) = -1 + \delta v_{\pi}(2)$$

yields $\mathbf{v}_\pi = [\mathbf{20}, \mathbf{19}, \mathbf{18}]$ for $\delta = 1$.

0.0.10 5.2.4

Now let $r_{NT} = -10$.

The optimal policy is non-unique ($r_{NT} < -9$).

To be formally correct, we now first exclude the left side of the game to avoid confusion because of the notation.

$$S \leftarrow \{1, 2, 3\}$$

$$\pi^{*1}(s) = a_- \text{ if } s \neq 3 \text{ else } (\frac{1}{2})$$
$$\pi^{*2}(s) = a_- \text{ if } s \neq 3 \text{ else } a_+$$
$$\pi^{*3}(s) = a_-$$

They will all give optimality.

Assume $\pi^*(s) = a_-$

$$q_{\pi^*}(s, a_+) = r(s, a, s+1) + \delta v_{\pi^*}(s+1)$$
$$q_{\pi^*}(s, a_-) = r(s, a, s-1) + \delta v_{\pi^*}(s-1)$$

$$q_{\pi^*}(1, a_-) = 20 + 0$$
$$q_{\pi^*}(2, a_-) = -10 + \delta v_{\pi}(1)$$
$$q_{\pi^*}(3, a_-) = -10 + \delta v_{\pi}(2)$$

$$v_{\pi^*}(s) = \sum_a \pi^*(a \mid s) q(s, a)$$
$$v_{\pi^*}(s) = q(s, a_-)$$

yields

$$v_{\pi^*}(1) = 20 + 0$$
$$v_{\pi^*}(2) = -10 + \delta v_{\pi}(1)$$
$$v_{\pi^*}(3) = -10 + \delta v_{\pi}(2)$$

yields $\mathbf{v}_\pi = [\mathbf{20}, \mathbf{10}, \mathbf{0}]$ for $\delta = 1$.

0.0.11 5.3.1

state(s)	action(a)	$\pi(a \mid s)$	reward(r)
1	L	1/4	0
1	R	3/4	-2
2	L	1/2	-2
2	R	1/2	-2
3	L	3/4	-2
3	R	1/4	20

Let

$$S = \{1, 2, 3\} A = \{L, R\}$$

Now

$$q_\pi(s,a) = \sum_{s'} p(s' \mid s,a)[r(s,a,s') + \delta v_\pi(s')]$$
$$q_\pi(s,a) = p(s+1 \mid s,a)[r(s,a,s+1) + \delta v_\pi(s+1)] + p(s-1 \mid s,a)[r(s,a,s-1) + \delta v_\pi(s-1)]$$

Which is

$$q_\pi(s,L) = r(s,a,s-1) + \delta v_\pi(s-1)$$
$$q_\pi(s,R) = r(s,a,s+1) + \delta v_\pi(s+1)$$

Now

$$q_\pi(1,L) = 0 + 0$$
$$q_\pi(1,R) = -2 + \delta v_\pi(2)$$
$$q_\pi(2,L) = -2 + \delta v_\pi(1)$$
$$q_\pi(2,R) = -2 + \delta v_\pi(3)$$
$$q_\pi(3,L) = -2 + \delta v_\pi(2)$$
$$q_\pi(3,R) = 20 + 0$$

Now

$$v_\pi(s) = \sum_a \pi(a \mid s) q(s,a) v_\pi(s) = \pi(L \mid s) q(s,L) + \pi(R \mid s) q(s,R)$$

yields

$$v_\pi(1) = \frac{1}{4}[0+0] + \frac{3}{4}[-2+\delta v_\pi(2)] v_\pi(2) = \frac{1}{2}[-2+\delta v_\pi(1)] + \frac{1}{2}[-2+\delta v_\pi(3)] v_\pi(3) = \frac{3}{4}[-2+\delta v_\pi(2)] + \frac{1}{4}[20+0]$$

yields $\mathbf{v}_\pi = [-\frac{\mathbf{9}}{\mathbf{2}}, -\mathbf{4}, \frac{\mathbf{1}}{\mathbf{2}}]^{\mathbf{T}}$ for $\delta = 1$.

0.0.12 5.3.2

$$q_\pi(1,L) = 0 + 0$$
$$q_\pi(1,R) = -2 - 4$$
$$q_\pi(2,L) = -2 - \frac{9}{2}$$
$$q_\pi(2,R) = -2 + \frac{1}{2}$$
$$q_\pi(3,L) = -2 - \frac{9}{2}$$
$$q_\pi(3,R) = 20 + 0$$

Horrible policy

0.0.13 5.3.3

For this game with $r_{NT} = -2$, choosing R at any point is unique and optimal..

[ ]: