1. Policy & Value Function

① Value Function:

Probability transition matrix

$$v = l + v P(v)$$

value vector

$$= l + lP + lP^2 + \cdots$$

expected under this iteration

$\Leftrightarrow$

expected under j-th step

$$v(j) = l(j) + \sum_{i=1}^{n} v(i)\, p(s_{t+1} = i \mid s_t = j, v)$$

value at j-th step — next step value — probability for $j \to i$

② Policy

- In terms of value:

uncontrolled transition matrix (restriction) — sensitivity

$$P(v)_{ij} = \frac{P(0)_{ij} \exp(\lambda v(i))}{\sum_k P(0)_{kj} \exp(\lambda v(k))}$$

- In terms of attraction: $A_{ij} = \frac{1}{\lambda} v(i) + \ln P(0)_{ij}$

$$P(A)_{ij} = \frac{\exp(\lambda A_{ij})}{\sum_k \exp(\lambda A_{kj})}$$

Intuition: $v(i) \uparrow$, more attractive, easier to be selected

③ The problem is:

$$v_{\tau+1} = l + v_\tau P(v_\tau)$$

$$\text{s.t. } v(I - P(v)) = l$$

2. Multiple agents

① Joint-state: (two agents)

$$S = S_1 \times S_2$$

② Policy:

$$P_1(V_1)_{ij} = \frac{\overset{I \otimes P_1(0)}{\downarrow}\atop{\Pi_1(0)_{ij} \exp(\vec{V_1}(i))}}{\sum_k \Pi_1(0)_{kj} \exp(\vec{V_1}(k))}, \quad P_2(V_2)_{ij} = \frac{\overset{P_2(0) \otimes I}{\downarrow}\atop{\Pi_2(0)_{ij} \exp(\vec{V_2}(i))}}{\sum_k \Pi_2(0)_{kj} \exp(\vec{V_2}(k))}$$

$$P(V_1, V_2) = P_2(V_2) P_1(V_1)$$

③ Value:

$$v_1^{(1)} = \vec{l_1} + \vec{v_1}^{(1)} P(v_1^{(1)}, 0)$$

$$v_2^{(1)} = \vec{l_2} + \vec{v_2}^{(1)} P(0, v_2^{(1)})$$

$$\vdots$$

$$v_1^{(i)} = \vec{l_1} + \vec{v_1}^{(i)} P(v_1^{(i)}, v_2^{(i-1)})$$

$$v_2^{(i)} = \vec{l_2} + \vec{v_2}^{(i)} P(v_1^{(i-1)}, v_2^{(i)})$$

until $v_k^{(i)} \approx v_k^{(i+1)}$

3. Stag-Hunt Game

① Process: stag → subject → computer

② Update: $P(A|B,C) = \dfrac{P(B|A,C)P(A|C)}{P(B|C)}$

$$\underbrace{p(k_{com}(T)|y,k_{sub})}_{\text{object: posterior of T step}} \propto \underbrace{p(y(1,\cdots,T)|k_{sub}(1,\cdots,T),k_{com})}_{\text{likelihood}}\overbrace{p(k_{com})}^{\text{prior}}$$

likelihood: Assume kcom is the opponent's strategy, the probability of generating trajectory $y(1,\cdots,T)$

$$= \prod_{t=1}^{T-1} \underbrace{k^{T-1}}_{\text{forget discount}} p(s_{t+1}|s_t, k_{sub}(t), k_{com})$$

③ We should solve:

$$\hat{k}_{com}(t) = \underset{k_{com}\in\{1,\cdots,k_{sub}\}}{\arg\max}\, p(k_{com}(t)|y,k_{sub}),$$

$$k_{sub}(t+1) = \hat{k}_{com}(t)+1$$