Harsanyi-Shapley Values – Stochastic Games (n players)

Stochastic Games

Definition 6.2.1 (Stochastic game) *A* stochastic game *(also known as a* Markov game*) is a tuple* (Q, N, A, P, r)*, where:*

- Q *is a finite set of games;*
- N *is a finite set of* n *players;*
- $A = A_1 \times \cdots \times A_n$*, where* A_i *is a finite set of actions available to player* i*;*
- $P : Q \times A \times Q \mapsto [0, 1]$ *is the transition probability function;* $P(q, a, \hat{q})$ *is the probability of transitioning from state* q *to state* $\hat{q}$ *after action profile* a*; and*
- $R = r_1, \ldots, r_n$*, where* $r_i : Q \times A \mapsto \mathbb{R}$ *is a real-valued payoff function for player* i*.*

How to generalize the HS definition to stochastic games (with discounted payoff?)

Recall: For two players

$$\langle s, a, \overline{a}, r, \overline{r}, s' \rangle$$

$$Q'_s = Q_s + \alpha_{a,\overline{a}}(r + \gamma \, \mathrm{Coco}(Q_{s'}, \overline{Q}_{s'}) - Q_s);$$

$$\overline{Q}'_s = \overline{Q}_s + \alpha_{a,\overline{a}}(\overline{r} + \gamma \, \mathrm{Coco}(\overline{Q}_{s'}, Q_{s'}) - \overline{Q}_s);$$

Extend to n players: HS*

First, we define an n-player, normal-form game at each state $x \in X$. For joint action $\mathbf{a}$ of the players, the game assigns the expected utility for each player according to the reward from the transition and the current value of the resulting state. Formally, let $\mathbf{V} = (V_1, \ldots, V_n) : X \to \mathbb{R}^n$. Define the utility for player i for joint action $\mathbf{a}$: $U_i(x, \mathbf{a}, \mathbf{V}) = \sum_{y \in X} P(x, \mathbf{a}, y)[R_i(x, \mathbf{a}, y) + \gamma V_i(y)]$, and finally define $G_x(\mathbf{V}) = (U_1(x, \cdot, \mathbf{V}), \ldots, U_n(x, \cdot, \mathbf{V}))$.

Definition 3.3 (Operator T)**.** Let $\mathcal{B}$ be the space of functions from X to $\mathbb{R}^n$. Define operator $T : \mathcal{B} \to \mathcal{B}$ by

$$[T(\mathbf{V})](x) = \mathrm{HS}\left(G_x(\mathbf{V})\right).$$

Definition 3.4 (HS$*$ values for stochastic game)**.** The HS$*$ value $\mathbf{V}_{\mathrm{HS}*}$ is a solution of the following equations.

$$\mathbf{V}(x) = [T(\mathbf{V})](x),\ \forall x \in X. \tag{4}$$

Alternative approach

- Define two-player zero-sum stochastic game for each coalition *I* (by summing utilities of players in coalition and subtracting complement)
- Define operator $T_I(V)$:

$$[T_I(V)](x) = \text{maxmin}_I(H_x(V)).$$

- Since maxmin is a contraction, T_I has a unique fixed point: V_I^*
- Let $i \in N$. At state $x \in X$, define

$$V_{\text{HS},i}(x) = \frac{1}{n} \sum_{I \subseteq N : i \in I} \binom{n-1}{|I-1|}^{-1} V_I^*(x), \qquad (3)$$

and $\mathbf{V}_{\text{HS}} = (V_{\text{HS},i})_{i \in N}$.

Are the two definitions equal?

Proposition 3.6. *If $\mathcal{G}$ is a 2-player, general-sum stochastic game, $\mathbf{V}_{HS} = \mathbf{V}_{HS*}$. In general, $\mathbf{V}_{HS} \neq \mathbf{V}_{HS*}$.*

Proof. Let $\mathcal{G}$ be a 2-player stochastic game. Let $\mathbf{V} = (V_1, V_2) : X \to \mathbb{R}^2$ be a function. Define $W_1(x) = V_1(x) - V_2(x)$. Next, apply T to $\mathbf{V}$ to get $\widehat{\mathbf{V}} = T(\mathbf{V})$. Now, let $\widehat{W_1}(x) = \widehat{V_1}(x) - \widehat{V_2}(x)$. Then

$$\begin{aligned}
\widehat{W_1}(x) &= \widehat{V_1}(x) - \widehat{V_2}(x) \\
&= \frac{1}{2}[(\text{maxmin}_1(G_{x,1}(\mathbf{V})) + \text{maxmax}_1(G_{x,1,2}(\mathbf{V}))) \\
&\quad - (\text{maxmin}_2(G_{x,2}(\mathbf{V})) + \text{maxmax}_2(G_{x,1,2}(\mathbf{V}))) \\
&= \text{maxmin}_1(G_{x,1}(\mathbf{V})) \\
&= T_1(V_1 - V_2)(x) = T_1(W_1)(x),
\end{aligned}$$

Are the two definitions equal?

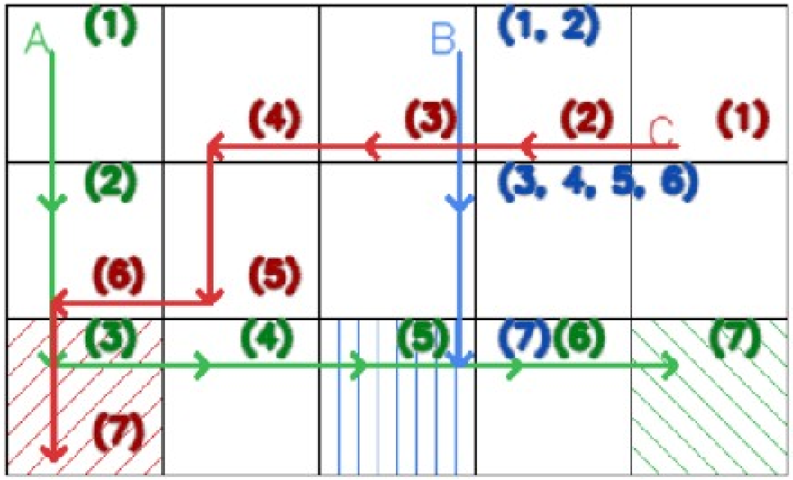
Proposition 3.6. *If $\mathcal{G}$ is a 2-player, general-sum stochastic game,* $\mathbf{V}_{HS} = \mathbf{V}_{HS*}$. *In general,* $\mathbf{V}_{HS} \neq \mathbf{V}_{HS*}$.

where T_1 is the operator defined in Section 3.1. Therefore, as T is applied repeatedly, W_1 converges to V_1^* as defined in Section 3.1.

Similarly, if $W_2 = V_2 - V_1$; and $W_{1,2} = V_1 + V_2$, we have that $W_2 \to V_2^*$ and $W_{1,2} \to V_{1,2}^*$ as T is iteratively applied. Therefore, $V_1 = \frac{1}{2}(W_{1,2} + W_1)$ and $V_2 = \frac{1}{2}(W_{1,2} + W_2)$ both converge, precisely to $(V_{\mathrm{HS},1}, V_{\mathrm{HS},2}) = \mathbf{V}_{\mathrm{HS}}$, which completes the proof.

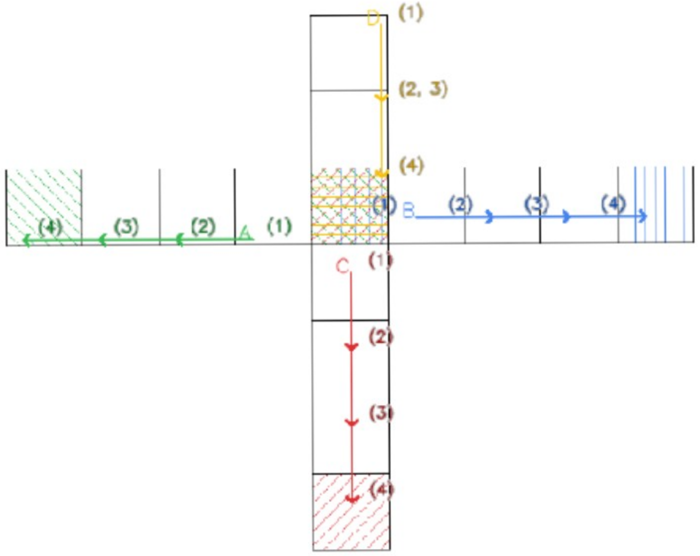
Examples that show $\mathbf{V}_{\mathrm{HS}} \neq \mathbf{V}_{\mathrm{HS}*}$ when $n > 2$ are provided in Section 4. □

Example



State	HS	HS*
1	(-0.1, -0.2, 0.3)	(-7.6, 15.2, -7.6)
2	(0.2, -0.3, 0.1)	(-5.0, -7.9, 12.9)
3	(-16.3, 32.1, -15.8)	(-30.4, 52.2, -21.8)
4	(-16.2, 32.7, -16.6)	(-13.7, 54.0, -40.3)
5	(0.5, -0.4, -0.0)	(12.4, -25.1, 12.7)
6	(0.0, -0.0, -0.0)	(0.0, 0.0, 0.0)
V	(62.0, 161.6, 62.0)	(49.8, 185.9, 49.8)
SP	(-32.0, 64.0, -32.0)	(-44.2, 88.3, -44.2)

Example



State	HS
1	(116.3, 116.3, 116.3, -349.0)
2	(-332.8, -332.8, -332.8, 998.4)
3	(0.0, 0.0, 0.0, 0.0)
V	(780.6, 780.6, 780.6, 747.2)
SP	(-216.4, -216.4, -216.4, 649.3)

State	HS*
1	(236.8, 236.8, 236.8, -710.5)
2	(-332.8, -332.8, -332.8, 998.4)
3	(0.0, 0.0, 0.0, 0.0)
V	(901.0, 901.0, 901.0, 385.8)
SP	(-96.0, -96.0, -96.0, 287.9)