CHAPTER III: STOCHASTIC GAMES

1. Introduction

In a (two-person, zero-sum) stochastic game, the play proceeds by stages from state to state, according to transition probabilities controlled jointly by the two players. In accordance with Shapley [1953], we assume there are a finite number of states, $s \in \{0,1,\ldots,S\}$, where $s = 0$ is the state "stop the game," and for each $s \in \mathcal{S} = \{1,\ldots,S\}$, player 1 has a finite number of possible actions, $i = 1,\ldots,I_s$, and player 2 has a finite number of possible actions, $j = 1,\ldots,J_s$. We can assume, without loss of generality, that $I = I_s$ and $J = J_s$ for each $s \in \mathcal{S}$. If, when at a state $s \in \mathcal{S}$, the players choose actions i and j, respectively, then player 1 receives the payoff a^s_{ij} from player 2 and the probability that the next stage of the game is in state t is p^{st}_{ij}, where $p^{st}_{ij} \geq 0$ for $t = 0,1,\ldots S$, and $\sum_{t=0}^{S} p^{st}_{ij} = 1$. Payments accumulate throughout the course of the play.

1.1 Notation. For each state $s \in \mathcal{S}$, let $A^s = (a^s_{ij})_{\substack{i=1,\ldots,I \\ j=1,\ldots,J}}$ and let $p^s_{ij} = (p^{st}_{ij})_{t=0,1,\ldots S}$ for $i = 1,\ldots,I$, $j = 1,\ldots,J$.

1.2 Assumption. Assume that $p^{s0}_{ij} > 0$ for all $i \in \{1,\ldots,I\}$, $j \in \{1,\ldots,J\}$ and for all $s \in \mathcal{S}$, (i.e., assume that the probability of stopping after any state and any actions is always positive). Then $\Pi = \min_{s,i,j} p^{s0}_{ij} > 0$.

Since Π is positive, the game ends with probability 1 after a finite number of stages, because for any number n, the probability that the game has not stopped after n stages is no more than $(1 - \Pi)^n$, which converges to 0 as $n \to \infty$.

Let $M = \max_{s,i,j} |a_{ij}^s|$. Then the expected total gain or loss for the players is bounded by

$$M + (1 - \Pi) \cdot M + (1 - \Pi)^2 \cdot M + \ldots = \frac{M}{\Pi} .$$

By specifying a starting state $s \in S$, we obtain a particular game Γ^s, where the payoff is the expected gain or loss. The term "stochastic game" will refer to the collection $\Gamma = (\Gamma^1, \ldots, \Gamma^S)$.

2. Strategies

2.1 Definition. A pure strategy for player 1 is a sequence $(\sigma_n)_{n=1}^{\infty}$, $\sigma_n \colon H_n \times S \to \{1, \ldots, I\}$, where H_n is the set of all possible histories of the play prior to stage n, i.e., the set of all $H_n = [(i_k, j_k, s_k)]_{k=1}^{n-1}$, where i_k (respectively j_k) was the action of player 1 (respectively 2) at stage k and s_k was the state at stage k.

Pure strategies for player 2 can similarly be defined.

The full sets of pure and mixed strategies are rather cumbersome. It will be shown that it suffices to consider a subclass of the set of mixed strategies, known as the behaviour stationary strategies, being those strategies which employ an independent randomization at each stage, where the randomization depends only on the state s (and not on the stage n nor the history).

2.2 <u>Definition</u>. <u>A behaviour stationary strategy</u> for player 1 is a collection $x = [x^s]_{s \in S}$ where $x^s = (x_1^s, \ldots, x_I^s)$, $x_i^s \geq 0$, $\sum_{i=1}^{I} x_i^s = 1$, and x_i^s is the probability that player 1 chooses action i in state s. Similarly, a behaviour stationary strategy $y = [y^s]_{s \in S}$, can be defined for player 2.

3. Existence of a Value

3.1 <u>Definition</u>. For an $I \times J$ matrix $C = (c_{ij})_{\substack{i=1,\ldots,I \\ j=1,\ldots,J}}$, let val (C) be the value, with respect to mixed strategies of the two-person zero-sum game in which the payoff matrix is C, i.e., $\text{val}(C) = \min_{y \in Y} \max_{x \in X} x^T C y = \max_{x \in X} \min_{y \in Y} x^T C y$, where X is the (I - 1)-dimensional simplex, Y is the (J - 1)-dimensional simplex.

3.2 <u>Lemma</u>. Let C, D be two $I \times J$ matrices. Then $|\text{val}(C) - \text{val}(D)| \leq \max_{i,j} |c_{ij} - d_{ij}|$.

<u>Proof</u>. Let $\alpha = \max_{i,j} |c_{ij} - d_{ij}|$. Then, $d_{ij} - \alpha \leq c_{ij} \leq d_{ij} + \alpha$. Considering the left-hand inequality, we see that each entry in C is greater than or equal to the corresponding entry in D minus α. It follows that $\text{val}(C) = \min_{y \in Y} \max_{x \in X} x^T C y \geq \min_{y \in Y} \max_{x \in X} x^T D y - \alpha = \text{val}(D) - \alpha$. Similarly from the right-hand inequality, $\text{val}(C) \leq \text{val}(D) + \alpha$. The desired result follows immediately. ∎

3.3 <u>Notation</u>. Let $\Gamma_{(n)}^s$ be the game starting at state s, which is stopped after stage n if a play reaches that stage. Let $\Gamma_{(n)} = (\Gamma_{(n)}^1, \ldots, \Gamma_{(n)}^S)$. Given a vector $w = (w^t)_{t \in S}$, let $B^s[w]$ denote the matrix with components $b^s_{ij}(w) = a^s_{ij} + \sum_{t\in S} p^{st}_{ij} w^t$. Let $\text{val}\ \Gamma^s_{(n)}$ denote the value (in mixed strategies) of the finite game $\Gamma^s_{(n)}$. Let $\text{val}\ \Gamma_{(n)} = (\text{val}\ \Gamma^s_{(n)})_{s\in S}$. It then follows by induction on n that for each $s \in S$

$$\text{val}\ \Gamma^s_{(1)} = \text{val}\ (A^s)$$

$$\text{val}\ \Gamma^s_{(n)} = \text{val}\ (B^s[\text{val}\ \Gamma_{(n-1)}]) \quad , \quad \text{for} \quad n = 2,3,\ldots \quad .$$

3.4 <u>Definition</u>. Let G be a two-person zero-sum game with mixed strategy sets X^1 and X^2, and expected payoff function $H: X^1 \times X^2 \to \mathbb{R}$. Then G <u>has a value</u> v if, for each $\varepsilon > 0$, there are mixed strategies $\bar{x}^1 \in X^1$ and $\bar{x}^2 \in X^2$ (depending on ε) such that

$$H(\bar{x}^1, x^2) \geq v - \varepsilon \quad \forall\, x^2 \in X^2 \quad \text{and}$$

$$H(x^1, \bar{x}^2) \leq v + \varepsilon \quad \forall\, x^1 \in X^1 \quad .$$

3.5 <u>Theorem</u>. For each $s \in S$, the game Γ^s has a value. Moreover, $\text{val}\ \Gamma^s = \lim_{n\to\infty} \text{val}\ \Gamma^s_{(n)}$.

<u>Proof</u>. Fix $s \in S$. In $\Gamma^s_{(n)}$, the expected payoff to player 1 is bounded in absolute value by

$$M + (1 - \Pi)M + \ldots + (1 - \Pi)^{n-1}M < \frac{M}{\Pi} \quad .$$

So $|\text{val}\ \Gamma^s_{(n)}| < M/\Pi$ for each n and so $\overline{\lim_{n\to\infty}}\ \text{val}\ \Gamma^s_{(n)}$

and $\underline{\lim}_{n\to\infty}$ val $\Gamma^s_{(n)}$ are finite. We shall prove that $\overline{\lim}_{n\to\infty}$ val $\Gamma^s_{(n)}$ $= \underline{\lim}_{n\to\infty}$ val $\Gamma^s_{(n)}$ and that this is a value for the game Γ^s. In a play of Γ^s, by playing an optimal strategy of $\Gamma^s_{(n)}$ for the first n stages and playing arbitrarily thereafter, the first player can guarantee his expected payoff will be at least

$$\text{val } \Gamma^s_{(n)} - \left(\sum_{m=n}^{\infty} (1 - \Pi)^m M\right) = \text{val } \Gamma^s_{(n)} - (1 - \Pi)^n \cdot \frac{M}{\Pi} .$$

Similarly, player 2 can guarantee that player 1's expected payoff will not exceed

$$\text{val } \Gamma^s_{(n)} + \sum_{m=n}^{\infty} (1 - \Pi)^m M = \text{val } \Gamma^s_{(n)} + (1 - \Pi)^n \frac{M}{\Pi} .$$

Let $\varepsilon > 0$ and let $m > 0$ be such that $(1 - \Pi)^m M/\Pi < \varepsilon/2$, so that $(1 - \Pi)^n M/\Pi < \varepsilon/2$ for all $n \geq m$. Let $k, \ell \geq m$ be such that val $\Gamma^s_{(k)} \geq \overline{\lim}_{n\to\infty}$ val $\Gamma^s_{(n)} - \varepsilon/2$ and val $\Gamma^s_{(\ell)} \leq \underline{\lim}_{n\to\infty}$ val $\Gamma^s_{(n)} + \varepsilon/2$. Then in a play of Γ^s, player 1 can guarantee his expected payoff will be at least

$$\underline{v}^{\varepsilon} = \text{val } \Gamma^s_{(k)} - (1 - \Pi)^k \frac{M}{\Pi} > \text{val } \Gamma^s_{(k)} - \frac{\varepsilon}{2} \geq \overline{\lim_{n\to\infty}} \text{ val } \Gamma^s_{(n)} - \varepsilon$$

and player 2 can guarantee that player 1's expected payoff will not exceed

$$\bar{v}^{\varepsilon} = \text{val } \Gamma^s_{(\ell)} + (1 - \Pi)^{\ell} \frac{M}{\Pi} < \text{val } \Gamma^s_{(\ell)} + \frac{\varepsilon}{2} \leq \underline{\lim}_{n\to\infty} \text{ val } \Gamma^s_{(n)} + \varepsilon$$

$$\leq \overline{\lim_{n\to\infty}} \text{ val } \Gamma^s_{(n)} + \varepsilon .$$

Hence $v^s = \overline{\lim_{n\to\infty}} \text{ val } \Gamma^s_{(n)}$ is a value for Γ^s, and since for each $\varepsilon > 0$,

$$\overline{\lim_{n\to\infty}} \text{ val } \Gamma^s_{(n)} - \varepsilon \le \underline{v}^\varepsilon \le \bar{v}^\varepsilon \le \underline{\lim}_{n\to\infty} \text{ val } \Gamma^s_{(n)} + \varepsilon \le \overline{\lim_{n\to\infty}} \text{ val } \Gamma^s_{(n)} + \varepsilon$$

it follows on letting $\varepsilon \to 0$ that $\overline{\lim_{n\to\infty}} \text{ val } \Gamma^s_{(n)} = \underline{\lim}_{n\to\infty} \text{ val } \Gamma^s_{(n)}$.
So $\text{val } \Gamma^s = v^s = \lim_{n\to\infty} \text{ val } \Gamma^s_{(n)}$. ∎

3.6 Notation. Let $T: \mathbb{R}^S \to \mathbb{R}^S$ be defined by $T^s w = \text{val } (B^s[w])$ for each $w = (w^1,\ldots,w^S) \in \mathbb{R}^S$ and each $s \in S$.

3.7 Definition. For $v = (v^1,\ldots,v^S) \in \mathbb{R}^S$, let $\|v\| = \max_{s\in S} |v^s|$.

3.8 Proposition. There is a unique $\bar{v} \in \mathbb{R}^S$ such that $\bar{v} = T\bar{v}$, and any sequence $(w_{(n)})_{n=0}^\infty$ defined by letting $w_{(0)}$ be arbitrary in $\mathbb{R}^S$ and letting $w_{(n)} = Tw_{(n-1)}$, $n = 1,2,\ldots$, converges to $\bar{v}$.

Proof. Let $u, w \in \mathbb{R}^S$. Then

$$\|Tu - Tw\| = \max_{s\in S} |T^s u - T^s w|$$

$$= \max_s |\text{val } (B^s[u]) - \text{val } (B^s[w])|$$

$$\le \max_s \max_{i,j} |b^s_{ij}(u) - b^s_{ij}(w)| \quad , \text{ by Lemma 3.2}$$

$$= \max_{s,i,j} \Big| \sum_{t\in S} p^{st}_{ij} u^t - \sum_{t\in S} p^{st}_{ij} w^t \Big|$$

$$\leq \max_{t\in S} |u^t - w^t| \cdot \max_{s,i,j} \sum_{t\in S} p_{ij}^{st}$$

$$= \|u - w\| \max_{s,i,j} (1 - p_{ij}^{s0})$$

$$= \|u - w\| \cdot (1 - \Pi) \quad .$$

So, $\|Tu - Tw\| \leq (1 - \Pi)\|u - w\|$.

Let $w_{(0)} \in \mathbb{R}^S$ be arbitrary. Define a sequence $(w_{(n)})_{n=0}^{\infty}$, inductively, by $w_{(n)} = Tw_{(n-1)}$, $n = 1,2,\ldots$. Then,

$$\begin{aligned}\|w_{(n)} - w_{(n-1)}\| &= \|Tw_{(n-1)} - Tw_{(n-2)}\| \\ &\leq (1 - \Pi)\|w_{(n-1)} - w_{(n-2)}\| \\ &\ \vdots \\ &\leq (1 - \Pi)^{n-1}\|w_{(1)} - w_{(0)}\| \quad .\end{aligned}$$

For $m > n$,

$$\begin{aligned}\|w_{(m)} - w_{(n)}\| &\leq \|w_{(n+1)} - w_{(n)}\| + \|w_{(n+2)} - w_{(n+1)}\| + \ldots + \|w_{(m)} - w_{(m-1)}\| \\ &\leq \{(1 - \Pi)^n + (1 - \Pi)^{n+1} + \ldots + (1 - \Pi)^m\}\|w_{(1)} - w_{(0)}\| \\ &< \frac{(1 - \Pi)^n}{\Pi}\| w_{(1)} - w_{(0)}\| \quad .\end{aligned}$$

Hence $(w_{(n)})_{n=0}^{\infty}$ is a Cauchy sequence. Since $\mathbb{R}^S$ is complete, this sequence has a limit, $\bar{v} \in \mathbb{R}^S$. Then

$$\|\bar{v} - T\bar{v}\| \le \|\bar{v} - w_{(n)}\| + \|w_{(n)} - T\bar{v}\|$$
$$= \|\bar{v} - w_{(n)}\| + \|Tw_{(n-1)} - T\bar{v}\|$$
$$\le \|\bar{v} - w_{(n)}\| + (1 - \Pi)\|w_{(n-1)} - \bar{v}\| \to 0 \quad \text{as} \quad n \to \infty \ .$$

Since the left hand side of the above inequality is independent of n, it follows that $\bar{v} = T\bar{v}$.

To prove that $\bar{v}$ is unique, suppose $\bar{w} \in \mathbb{R}^S$ such that $\bar{w} = T\bar{w}$. Then,

$$\|\bar{v} - \bar{w}\| = \|T\bar{v} - T\bar{w}\| \le (1 - \Pi)\|\bar{v} - \bar{w}\| \quad \text{and so}$$

$$0 \le \Pi\|\bar{v} - \bar{w}\| \le 0 \quad \text{and hence} \quad \bar{v} = \bar{w} \ .$$ ∎

3.9 Corollary. Let $\bar{v} = (\bar{v}^1, \ldots, \bar{v}^S)$ be the unique solution of $\bar{v} = T\bar{v}$. Then for each $s \in S$, $\bar{v}^s$ is the value of Γ^s.

Proof: Let $w_{(0)}$ be the zero vector in $\mathbb{R}^S$. Then the sequence $(w_{(n)})_{n=0}^{\infty}$ is such that for each $s \in S$,

$$w^s_{(1)} = T^s w_{(0)} = \text{val}\,(B^s[w_{(0)}]) = \text{val}\,(A^s) = \text{val}\,\Gamma^s_{(1)}$$
$$\vdots$$
$$w^s_{(n)} = T^s w_{(n-1)} = \text{val}\,(B^s[w_{(n-1)}]) = \text{val}\,(B^s[\text{val}\,\Gamma_{(n-1)}]) = \text{val}\,\Gamma^s_{(n)} \ .$$
$$\vdots$$

So $\bar{v}^s = \lim_{n\to\infty} w^s_{(n)}$

$= \lim_{n\to\infty} \text{val}\,\Gamma^s_{(n)}$

$= \text{val}\,\Gamma^s$, by Theorem 3.5 . ∎

3.10 Remark. Computation of the sequence $(w_{(n)})_{n=0}^{\infty}$ is an iterative procedure for computing $\bar{v}$. An upper bound for the maximum error when $w_{(0)} = 0$ is:

$$\|\bar{v} - w_{(n)}\| = \|T\bar{v} - Tw_{(n-1)}\| \leq (1 - \Pi)\|\bar{v} - w_{(n-1)}\| \leq \cdots \leq$$

$$\leq (1 - \Pi)^n \|\bar{v} - w_{(0)}\| \leq (1 - \Pi)^n \frac{M}{\Pi} \; .$$

3.11 Remark. We say that $\bar{v}$ is the value of the stochastic game Γ, and denote it by val Γ.

4. Optimal Strategies

4.1 Theorem. For each $s \in S$, let $\bar{x}^s$, $\bar{y}^s$ be optimal strategies for player 1, player 2, respectively, in the matrix game $B^s[\bar{v}]$, where $\bar{v} = \text{val}\,\Gamma$. Then, $\bar{x} = [\bar{x}^s]_{s\in S}$, $\bar{y} = [\bar{y}^s]_{s\in S}$ are optimal behaviour stationary strategies in Γ.

Proof: Let $\hat{\Gamma}_{(n)}$ be the same game as $\Gamma_{(n)}$, except that whenever a play reaches stage n, if the state at stage n is s, then the payoff at stage n is not a^s_{ij}, but $a^s_{ij} + \sum_{t\in S} p^{st}_{ij}\bar{v}^t$ (and the play stops as in $\Gamma_{(n)}$). Consider the game $\hat{\Gamma}_{(n)}$. If the state at stage n is s, then the payoff matrix for this stage is $B^s[\bar{v}]$. By playing $\bar{x}^s$ (respectively $\bar{y}^s$), player 1 (respectively player 2) guarantees that the expected payoff at stage n will be at least $\bar{v}^s$ (respectively at most $\bar{v}^s$). Consider stage $n - 1$. Since player 1 can guarantee (using $\bar{x}$) that the expected payoff at stage n will be at least $\bar{v}$, then if the state at stage $n - 1$ is t, player 1 can guarantee that the expected payoff matrix, starting at stage $n - 1$, is at least $B^t[\bar{v}]$. By playing $\bar{x}^t$ at stage $n - 1$ and $\bar{x}$ at stage n, player 1 guarantees that the expected payoff for the last two stages will be at least $\bar{v}^t$. Similarly, by playing $\bar{y}^t$ at stage $n - 1$ and $\bar{y}$ at stage n, player 2 guarantees that the expected payoff for the last two stages will be at most $\bar{v}^t$. Using backwards induction we see that by playing his behaviour stationary strategy at each stage, player 1 (respectively player 2) can guarantee that the expected payoff in $\hat{\Gamma}_{(n)}$ will be at least $\bar{v}$ (respectively, will be at most $\bar{v}$).

The difference in payoffs between $\hat{\Gamma}_{(n)}$ and $\Gamma_{(n)}$, and between $\Gamma_{(n)}$ and Γ, is no more that $(1 - \Pi)^n M/\Pi$. So the difference in payoffs between $\hat{\Gamma}_{(n)}$ and Γ is no more than

$$\alpha_n = 2(1 - \Pi)^n \frac{M}{\Pi} .$$

Let $\varepsilon > 0$ be arbitrary and let N be such that $\alpha_N < \varepsilon$. Then by playing the behaviour stationary strategy $\bar{x}$, player 1 guarantees that the expected payoff in $\hat{\Gamma}_{(N)}$ is at least $\bar{v}$ and hence the expected payoff in Γ is at least $\bar{v} - \alpha_N > \bar{v} - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, by playing $\bar{x}$, player 1 guarantees the expected payoff in Γ is at least $\bar{v}$. Similarly, by playing $\bar{y}$, player 2 guarantees that the expected payoff in Γ is no more than $\bar{v}$. Hence $(\bar{x},\bar{y})$ is an optimal strategy pair for Γ. ∎

<u>Remark</u>. $\bar{x}$, $\bar{y}$ actually guarantee $\bar{v}$ (not $\bar{v} \pm \varepsilon$ as in the general Definition 3.4).

5. Examples

5.1

s = 1

1; $\frac{1}{2}, 0$	0; $0, \frac{1}{2}$
0; $0, \frac{1}{2}$	3; $\frac{1}{2}, 0$

s = 2

1; $0, \frac{1}{2}$	1; $0, \frac{1}{2}$
1; $0, \frac{1}{2}$	1; $0, \frac{1}{2}$

The (i,j)-th cell contains the following information:

a_{ij}^{s}; $p_{ij}^{s1}, \ldots, p_{ij}^{sS}$

Note that $\Pi = \frac{1}{2}$.

Let $\bar{u} = \text{val}\ \Gamma^1$, $\bar{v} = \text{val}\ \Gamma^2$. Then

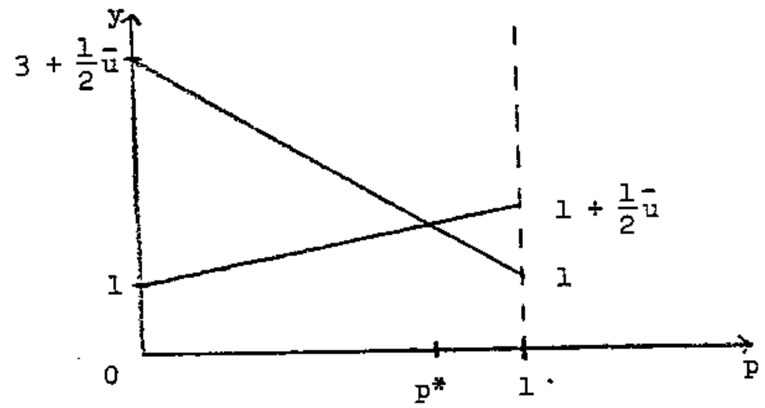
$$\bar{v} = \text{val} \begin{bmatrix} 1 + \frac{1}{2}\bar{v} & 1 + \frac{1}{2}\bar{v} \\ 1 + \frac{1}{2}\bar{v} & 1 + \frac{1}{2}\bar{v} \end{bmatrix},$$

and hence $\bar{v} = 1 + 1/2\ \bar{v}$, i.e., $\bar{v} = 2$.

$$\bar{u} = \text{val} \begin{bmatrix} 1 + \frac{1}{2}\bar{u} & 0 + \frac{1}{2}\bar{v} \\ 0 + \frac{1}{2}\bar{v} & 3 + \frac{1}{2}\bar{u} \end{bmatrix} = \text{val} \begin{bmatrix} 1 + \frac{1}{2}\bar{u} & 1 \\ 1 & 3 + \frac{1}{2}\bar{u} \end{bmatrix}.$$

So

$$\bar{u} = \max_{0 \leq p \leq 1} \min \{p(1 + \tfrac{1}{2}\bar{u}) + (1 - p) , p + (1 - p)(3 + \tfrac{1}{2}\bar{u})\} \quad .$$



Since $\bar{u} \geq 0$ (all entries ≥ 0), then $3 + 1/2\ \bar{u} \geq 1$ and $1 + 1/2\ \bar{u} \geq 1$, and so the lines intersect in $[0,1]$ at p^* satisfying

$$1 + \tfrac{1}{2}\bar{u}p^* = 3 + \tfrac{1}{2}\bar{u} - (2 + \tfrac{1}{2}\bar{u})p^* \Leftrightarrow p^* = \frac{2 + \frac{1}{2}\bar{u}}{2 + \bar{u}} \quad .$$

So

$$\bar{u} = 1 + \tfrac{1}{2}\bar{u}p^* = 1 + \frac{\frac{1}{2}\bar{u}(2 + \frac{1}{2}\bar{u})}{2 + \bar{u}} \Leftrightarrow \tfrac{3}{4}\bar{u}^2 = 2 \quad .$$

So $\bar{u} = \frac{2\sqrt{2}}{\sqrt{3}}$ (we choose the positive root since $\bar{u} \geq 0$), $p^* = \frac{2\sqrt{3} + \sqrt{2}}{2(\sqrt{3} + \sqrt{2})}$, $q^* = \frac{2\sqrt{3} + \sqrt{2}}{2(\sqrt{3} + \sqrt{2})}$.

The value for this stochastic game is $(\bar{u},\bar{v}) = (\frac{2\sqrt{2}}{\sqrt{3}}, 2)$, and the optimal behaviour stationary strategies are given by p^* and q^* in state $s = 1$, and any strategies are optimal for $s = 2$. Note that the value is not rational, even though the entries in the payoff matrices are integers, (for regular zero-sum games, if the entries are rational then the value is rational!)

5.2

s = 1

2 ; $\frac{1}{2}, 0$	-1 ; $0, \frac{1}{3}$
-1 ; $0, \frac{1}{3}$	2 ; $\frac{1}{2}, 0$

s = 2

0 ; $0, \frac{1}{2}$	6 ; $0, \frac{1}{2}$
2 ; $0, \frac{1}{2}$	0 ; $0, \frac{1}{2}$

Note that $\Pi = \frac{1}{2}$.

Let $\bar{u} = \text{val}\,\Gamma^1$, $\bar{v} = \text{val}\,\Gamma^2$.

(a) $\bar{u}$ and $\bar{v}$ are defined as the unique solutions of

$$\bar{u} = \text{val} \begin{array}{|c|c|} \hline 2 + \frac{1}{2}\bar{u} & -1 + \frac{1}{3}\bar{v} \\ \hline -1 + \frac{1}{3}\bar{v} & 2 + \frac{1}{2}\bar{u} \\ \hline \end{array} \;,$$

$$\bar{v} = \text{val} \begin{array}{|c|c|} \hline \frac{1}{2}\bar{v} & 6 + \frac{1}{2}\bar{v} \\ \hline 2 + \frac{1}{2}\bar{v} & \frac{1}{2}\bar{v} \\ \hline \end{array} = \text{val} \begin{array}{|c|c|} \hline 0 & 6 \\ \hline 2 & 0 \\ \hline \end{array} + \frac{1}{2}\bar{v} \;.$$

Since $\text{val} \begin{array}{|c|c|} \hline 0 & 6 \\ \hline 2 & 0 \\ \hline \end{array} = \frac{3}{2}$, $\bar{v} = 3$.

So $\bar{u} = \text{val} \begin{array}{|c|c|} \hline 2 + \frac{1}{2}\bar{u} & 0 \\ \hline 0 & 2 + \frac{1}{2}\bar{u} \\ \hline \end{array} = 1 + \frac{1}{4}\bar{u}$ and hence $\bar{u} = \frac{4}{3}$.

The optimal behaviour stationary strategies for Γ^1 and Γ^2 are the optimal strategies for

8/3	0
0	8/3

and

3/2	15/2
7/2	3/2

, respectively .

That is, $p^* = 1/2$, $q^* = 1/2$ for Γ^1 and $p^* = 1/4$, $q^* = 3/4$ for Γ^2.

(b) Starting with the initial approximations $u_{(0)} = 0$, $v_{(0)} = 0$, we calculate the first few terms of the sequences $(u_{(n)})$, $(v_{(n)})$ which converge to $\bar{u}$ and $\bar{v}$, respectively.

$$u_{(1)} = \text{val}\begin{bmatrix} 2+\frac{1}{2}u_{(0)} & -1+\frac{1}{3}v_{(0)} \\ -1+\frac{1}{3}v_{(0)} & 2+\frac{1}{2}u_{(0)} \end{bmatrix} = \text{val}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \frac{1}{2}$$

$$v_{(1)} = \text{val}\begin{bmatrix} 0+\frac{1}{2}v_{(0)} & 6+\frac{1}{2}v_{(0)} \\ 2+\frac{1}{2}v_{(0)} & 0+\frac{1}{2}v_{(0)} \end{bmatrix} = \text{val}\begin{bmatrix} 0 & 6 \\ 2 & 0 \end{bmatrix} = \frac{3}{2}$$

$$u_{(2)} = \text{val}\begin{bmatrix} 2+\frac{1}{2}u_{(1)} & -1+\frac{1}{3}v_{(1)} \\ -1+\frac{1}{3}v_{(1)} & 2+\frac{1}{2}u_{(1)} \end{bmatrix} = \text{val}\begin{bmatrix} 9/4 & -\frac{1}{2} \\ -\frac{1}{2} & 9/4 \end{bmatrix} = \frac{7}{8}$$

$$v_{(2)} = \text{val}\begin{bmatrix} 0+\frac{1}{2}v_{(1)} & 6+\frac{1}{2}v_{(1)} \\ 2+\frac{1}{2}v_{(1)} & 0+\frac{1}{2}v_{(1)} \end{bmatrix} = \text{val}\begin{bmatrix} 0 & 6 \\ 2 & 0 \end{bmatrix} + \frac{1}{2}v_{(1)} = \frac{9}{4}$$

$$u_{(3)} = \text{val}\begin{bmatrix} 39/16 & -\frac{1}{4} \\ -\frac{1}{4} & 39/16 \end{bmatrix} = 35/32$$

$$v_{(3)} = \text{val}\begin{bmatrix} 0 & 6 \\ 2 & 0 \end{bmatrix} + \frac{1}{2}v_{(2)} = \frac{21}{8}$$

$$u_{(4)} = \text{val}\begin{bmatrix} 163/64 & -\frac{1}{8} \\ -\frac{1}{8} & 163/64 \end{bmatrix} = 155/128$$

$$v_{(4)} = \text{val}\begin{bmatrix} 0 & 6 \\ 2 & 0 \end{bmatrix} + \frac{1}{2}v_{(3)} = \frac{45}{16} .$$

6. Discount Games

Consider a stochastic game with zero stop probability at each state, that is $p_{ij}^{s0} = 0$ for all s, i, j. Then if this game is modified so that the payoff at stage n, in state s, is $(1 - r)^{n-1} a_{ij}^{s}$, for some fixed r in $(0,1)$, then the resulting game is called the r-discount game. It is easily seen that the r-discount game is equivalent to a stochastic game with minimum stop probability $\Pi = r$, and with transition probabilities $q_{ij}^{s0} = r$, $q_{ij}^{st} = (1 - r)p_{ij}^{st}$ for $s, t \in S$, for all i and j, where p_{ij}^{st} are the given transition probabilities for the r-discount game.

References

Shapley, L. S. [1953], "Stochastic Games," Proceedings of the National Academy of Sciences of the U.S.A., 39, 1095-1100.

Further References

Bewley, T. and E. Kohlberg [1976], "The Asymptotic Theory of Stochastic Games," Mathematics of Operations Research, 1, 197-208.

Bewley, T. and E. Kohlberg [1976], "The Asymptotic Solution of a Recursion Equation Occuring in Stochastic Games," Mathematics of Operations Research, 1, 321-336.

Bewley, T. and E. Kohlberg [1978], "On Stochastic Games with Stationary Optimal Strategies," Mathematics of Operations Research, 3, 104-125.

Gilette, D [1957], "Stochastic Games with Zero Stop Probabilities, Contributions to the Theory of Games," Vol. III, Annals of Mathematics Studies, 39, Princeton University Press, Princeton, N.J., 179-188.

Luce, R. and H. Raiffa [1957], Games and Decisions, New York: John Wiley and Sons, Inc., Appendix A8.2.

Owen, G. [1968], Game Theory, Philadelphia: W. B. Saunders.

Sobel, M. J. [1971], "Noncooperative Stochastic Games," Annals of Mathematics Statistics, 42, 1930-1935.

Sobel, M. J. [1973], "Continuous Stochastic Games," Journal of Applied Probability, 10, 597-604.