STOCHASTIC GAMES: EXISTENCE OF THE MINMAX

ABRAHAM NEYMAN
Hebrew University of Jerusalem
Jerusalem, Israel

1. Introduction

The existence of the value for stochastic games with finitely many states and actions, as well as for a class of stochastic games with infinitely many states and actions, is proved in [2]. Here we use essentially the same tools to derive the existence of the minmax and maxmin for n-player stochastic games with finitely many states and actions, as well as for a corresponding class of n-person stochastic games with infinitely many states and actions.

The set of states of the stochastic game Γ is denoted S, the set of actions of player $i \in I$ at state $z \in S$ is $A^i(z)$, and the stage payoff and the transition probability as a function of the state z and the action profile $a \in \times_{i \in I} A^i(z)$ are denoted $r(z,a)$ and $p\,(\cdot \mid z,a)$, respectively. The vector payoff at stage t, $r(z_t,a_t) = (r^i(z_t,a_t))_{i \in I}$, is denoted x_t; note that x_t can be viewed as a function that is defined on the measurable space of infinite plays $(z_1, a_1, \ldots, z_t, \ldots)$. If Γ is a two-player zero-sum stochastic game we write x_t for x_t^1. A strategy profile σ together with an initial state $z_1 \in S$ induces a probability distribution on the (measurable) space of plays. The expectation w.r.t. this probability distribution is denoted by $E_\sigma^{z_1}$ or E_σ for short.

1.1. DEFINITIONS OF THE VALUE AND THE MINMAX

First recall the definition of the value, minmax, and maxmin. We say that $v(z)$ is the *value* of a two-person zero-sum stochastic game Γ with initial state $z_1 = z$ if:
1) For every $\varepsilon > 0$ there is an ε-*optimal* strategy σ of player 1, i.e., a strategy σ of player 1, such that: there is a positive integer N $(N = N(\varepsilon, \sigma))$ such

that for every strategy τ of player 2 and every $n \geq N$ we have

$$E^{z_1}_{\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) \geq v(z_1) - \varepsilon$$

and

$$E^{z_1}_{\sigma,\tau}\left(\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} x_t\right) \geq v(z_1) - \varepsilon;$$

and
2) For every $\varepsilon > 0$ there is an ε-optimal strategy τ of player 2, i.e., a strategy τ of player 2 such that: there is a positive integer N such that for every strategy σ of player 1 and every $n \geq N$ we have

$$E^{z_1}_{\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) \leq v(z_1) + \varepsilon$$

and

$$E^{z_1}_{\sigma,\tau}\left(\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} x_t\right) \leq v(z_1) + \varepsilon.$$

We say that $\bar{v}^i(z)$ is the *minmax* of player i in the I-player stochastic game Γ with initial state $z_1 = z$ if:
1) For every $\varepsilon > 0$ there is an *ε-minimaxing $I \setminus \{i\}$* strategy profile σ^{-i} in Γ, i.e., an $I \setminus \{i\}$ strategy profile σ^{-i} such that: there is a positive integer N such that for every $n \geq N$ and every strategy σ^i of player i we have

$$E^{z_1}_{\sigma^i,\sigma^{-i}}\left(\frac{1}{n}\sum_{t=1}^{n} x^i_t\right) \leq \bar{v}^i(z_1) + \varepsilon \tag{1}$$

and

$$E^{z_1}_{\sigma^i,\sigma^{-i}}\left(\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} x^i_t\right) \leq \bar{v}^i(z_1) + \varepsilon; \tag{2}$$

and
2) For every $\varepsilon > 0$ there is a positive integer N such that for every $I \setminus \{i\}$ strategy profile σ^{-i} in Γ there is a (σ^{-i}-*ε-N-maximizing*) strategy σ^i of player i such that for every $n \geq N$ we have

$$E^{z_1}_{\sigma^i,\sigma^{-i}}\left(\frac{1}{n}\sum_{t=1}^{n} x^i_t\right) \geq \bar{v}^i(z_1) - \varepsilon \tag{3}$$

and

$$E^{z_1}_{\sigma^i,\sigma^{-i}}\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}x^i_t\right)\geq \bar{v}^i(z_1)-\varepsilon. \tag{4}$$

We say that $\underline{v}^i(z)$ is the *maxmin* of player i in the I-player stochastic game Γ with initial state $z_1=z$ if:
1) For every $\varepsilon>0$ there is a strategy σ^i of player i and a positive N such that for every $n\geq N$ and every strategy profile σ^{-i} of players $I\setminus\{i\}$ we have

$$E^{z_1}_{\sigma^i,\sigma^{-i}}\left(\frac{1}{n}\sum_{t=1}^{n}x^i_t\right)\geq \underline{v}^i(z_1)-\varepsilon$$

and

$$E^{z_1}_{\sigma^i,\sigma^{-i}}\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}x^i_t\right)\geq \underline{v}^i(z_1)-\varepsilon;$$

and
2) For every $\varepsilon>0$ there is a positive integer N such that for every strategy σ^i of player i in Γ there is an $I\setminus\{i\}$ strategy profile σ^{-i} such that for every $n\geq N$ we have

$$E^{z_1}_{\sigma^i,\sigma^{-i}}\left(\frac{1}{n}\sum_{t=1}^{n}x^i_t\right)\leq \underline{v}^i(z_1)+\varepsilon$$

and

$$E^{z_1}_{\sigma^i,\sigma^{-i}}\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}x^i_t\right)\leq \underline{v}^i(z_1)+\varepsilon.$$

In the above definitions of the value (minmax and maxmin, respectively) of the stochastic games with initial state z_1 the positive integer N may obviously depend on the state z_1. When N does not depend on the initial state we say that the stochastic game has a value (a minmax and a maxmin, respectively). Formally, the stochastic game has a value if there exists a function $v:S\to\mathbb{R}$ such that $\forall\varepsilon>0\ \exists\sigma_\varepsilon,\tau_\varepsilon\ \exists N$ s.t. $\forall z_1\in S\ \forall\sigma,\tau\ \forall n\geq N$ we have

$$\varepsilon+E^{z_1}_{\sigma_\varepsilon,\tau}\left(\frac{1}{n}\sum_{t=1}^{n}x_t\right)\geq v(z_1)\geq -\varepsilon+E^{z_1}_{\sigma,\tau_\varepsilon}\left(\frac{1}{n}\sum_{t=1}^{n}x_t\right)$$

and

$$\varepsilon+E^{z_1}_{\sigma_\varepsilon,\tau}\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}x_t\right)\geq v(z_1)\geq -\varepsilon+E^{z_1}_{\sigma,\tau_\varepsilon}\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}x_t\right)$$

where $\sigma, \sigma_\varepsilon$ (respectively, τ, τ_ε) stands for strategies of player 1 (respectively, strategies of player 2).

Similarly, the stochastic game has a minmax of player i if there is a function $\bar{v}^i : S \to \mathbb{R}$ such that:
1) $\forall \varepsilon > 0\ \exists \sigma^{-i}\ \exists N$ s.t. $\forall \sigma^i\ \forall n \geq N$

$$E^{z_1}_{\sigma^{-i},\sigma^i}\left(\frac{1}{n}\sum_{t=1}^{n} x^i_t\right) \leq \bar{v}^i(z_1) + \varepsilon$$

and

$$E^{z_1}_{\sigma^{-i},\sigma^i}\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} x^i_t\right) \leq \bar{v}^i(z_1) + \varepsilon;$$

and
2) $\forall \varepsilon > 0\ \exists \sigma\ \exists N$ s.t. $\forall \sigma^{-i}\ \forall n \geq N$

$$E^{z_1}_{\sigma^{-i},\sigma^i}\left(\frac{1}{n}\sum_{t=1}^{n} x^i_t\right) \leq \bar{v}^i(z_1) + \varepsilon$$

and

$$E^{z_1}_{\sigma^{-i},\sigma^i}\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} x^i_t\right) \geq \bar{v}^i(z_1) - \varepsilon.$$

There are several weaker concepts of value, minmax and maxmin.

1.2. THE LIMITING AVERAGE VALUE

The *limiting average value* (of the stochastic game with initial state z_1) exists and equals $v_\infty(z_1)$ whenever $\forall \varepsilon > 0\ \exists \sigma_\varepsilon, \tau_\varepsilon$ s.t. $\forall \tau, \sigma$

$$\varepsilon + E^{z_1}_{\sigma_\varepsilon,\tau}\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} x_t\right) \geq v_\infty(z_1) \geq -\varepsilon + E^{z_1}_{\sigma,\tau_\varepsilon}\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} x_t\right).$$

A related but weaker concept of a value is the *limsup value*. The *limsup value* (of the stochastic game with initial state z_1) exists and equals $v_\ell(z_1)$ whenever $\forall \varepsilon > 0\ \exists \sigma_\varepsilon, \tau_\varepsilon$ s.t. $\forall \tau, \sigma$

$$\varepsilon + E^{z_1}_{\sigma_\varepsilon,\tau}\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} x_t\right) \geq v_\ell(z_1) \geq -\varepsilon + E^{z_1}_{\sigma,\tau_\varepsilon}\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} x_t\right).$$

1.3. THE UNIFORM VALUE

The *uniform value* of the stochastic game with initial state z_1 exists and equals $u(z_1)$ whenever $\forall\varepsilon > 0\ \exists\sigma_\varepsilon, \tau_\varepsilon\ \exists N$ s.t. $\forall\tau, \sigma\ \forall n \geq N$

$$\varepsilon + E^{z_1}_{\sigma_\varepsilon,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) \geq u(z_1) \geq -\varepsilon + E^{z_1}_{\sigma,\tau_\varepsilon}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right).$$

The stochastic game has a uniform value if there is a function $u : S \to \mathbb{R}$ such that $\forall\varepsilon > 0\ \exists\sigma_\varepsilon, \tau_\varepsilon\ \exists N$ s.t. $\forall z_1\ \forall\tau, \sigma\ \forall n \geq N$

$$\varepsilon + E^{z_1}_{\sigma_\varepsilon,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) \geq u(z_1) \geq -\varepsilon + E^{z_1}_{\sigma,\tau_\varepsilon}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right).$$

Analogous requirements define the limiting average minmax and maxmin, the limsup minmax and maxmin, and the uniform minmax and maxmin.

In a given two-player zero-sum stochastic game (1) existence of the value is equivalent to the existence of both the maxmin and the minmax, and their equality, (2) existence of the uniform value is equivalent to the existence of both the the uniform maxmin and the uniform minmax, and their equality, and (3) existence of the limiting average value is equivalent to the existence of both the limiting average maxmin and the limiting average minmax, and their equality.

1.4. EXAMPLES

The following example highlights the role of the set of inequalities used in the above definitions, and illustrates the differences of the various value concepts.

Consider the following example of a single-player stochastic game Γ with infinitely many states and finitely many actions: the state space S is the set of integers; at state 0 the player has two actions called $-$ and $+$ and in all other states the player has a single action (i.e., no choice); the payoff function depends only on the state. The payoff function r is given by: $r(k) = 1$ if either $(n-1)! \leq k < n!$ and $n > 1$ is even or $-n! < k \leq -(n-1)!$ and $n > 1$ is odd; in all other cases $r(k) = 0$. The transition is deterministic; $p\,(1 \mid 0, +) = 1 = p\,(-1 \mid 0, -)$, $p\,(k+1 \mid k) = 1$ if $k \geq 1$, and $p\,(k-1 \mid k) = 1$ if $k \leq -1$. The stochastic game Γ can be viewed as a two-player zero-sum game where player 2 has no choices.

Obviously,

$$\liminf_{n\to\infty} v_n(0) = 1/2 \quad \text{and} \quad \limsup_{n\to\infty} v_n(0) = 1$$

where v_n denotes the value of the normalized n-stage game. Therefore, the stochastic game Γ with the initial state $z_1 = 0$ does not have a uniform value.

For every strategy σ and every initial state z we have

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = 1.$$

Therefore, the stochastic game Γ with the initial state z_1 has a lim sup value ($= 1$). Since for every strategy σ and every initial state we have

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = 0$$

we deduce that (for any initial state) the limiting average value does not exist.

Consider the following modification of the above example. The payoff at state 0 is 1/2 and $p\,(\,1 \mid 0, *) = 1/2 = p\,(\,-1 \mid 0, *)$. All other data remains unchanged. The initial state is 0. Thus the payoff at stage 1 equals 1/2. For every $i > 1$ the payoff at stage i equals 1 with probability 1/2 and it equals 0 with probability 1/2. Therefore $E(\frac{1}{n}\sum_{i=1}^{n} x_i) = 1/2$ and therefore $v_n(0) = 1/2$. In particular, the stochastic game with initial state $z_1 = 0$ has a uniform value ($= 1/2$). However, since $\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = 0$ and $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = 1$ we deduce that the three value concepts—one based on the evaluation $\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i$ of a stream of payoffs $x_1, \ldots, x_i, \ldots$, one based on the valuation $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i$, and the other based on the payoff $\gamma(\sigma, \tau) = \lim_{n\to\infty} E_{\sigma,\tau}(\frac{1}{n}\sum_{i=1}^{n} x_i)$— give different results. However, such pathologies cannot arise in a game that has a value.

Section 2 discusses the candidate for the value. In Section 3 we present a basic probabilistic lemma which serves as the driving engine for the results to follow. Section 4 introduces constrained stochastic games and uses the basic probabilistic lemma to prove the existence of a value (of two-player zero-sum stochastic games) as well as the existence of the maxmin and the minmax (of n-player stochastic games).

2. The Candidate for the Value

Existence of the value v implies that the limit (as $n \to \infty$) of v_n, the (normalized) values of the n stage games, exists and equals v, and moreover the limit (as $\lambda \to 0+$) of v_λ, the (normalized) value of the λ-discounted games, exists and equals v.

Therefore, the only candidate for the value v is the limit of the values v_n as $n \to \infty$, which equals the limit of the values of the λ-discounted games v_λ as $\lambda \to 0+$.

Assume first that every stochastic game with finitely many states and actions indeed has a value (and thus in particular a uniform value). Denote by $v(z_1)$ the value as a function of the initial state z_1. Note that if σ is an ε-optimal strategy of player 1 in Γ_∞ then it must satisfy

$$E^{z_1}_{\sigma,\tau}(v(z_{n+1}) - v(z_1)) \geq -\varepsilon \tag{5}$$

for every strategy τ of player 2 and every positive integer n. Otherwise, there is a strategy τ' of player 2 and a positive integer n such that $E_{\sigma,\tau'}(v(z_n) - v(z_1)) < -\varepsilon$. Fix $\varepsilon_1 > 0$ sufficiently small such that $E_{\sigma,\tau'}(v(z_n) - v(z_1)) < -\varepsilon - 2\varepsilon_1$. Let τ'' be an ε_1-optimal strategy of player 2. Consider the strategy τ of player 2 that coincides with τ' in stages $1, \ldots, n$ and with τ'' thereafter, i.e., $\tau_i = \tau'_i$ if $i < n$ and $\tau_i(z_1, a_1, \ldots, z_i) = \tau''_{i-n+1}(z_n, a_n, \ldots, z_i)$ if $i \geq n$. It follows that for k sufficiently large $E_{\sigma,\tau}\left(\frac{1}{k}\sum_{i=1}^k x_i - v(z_1)\right) < -\varepsilon$, which contradicts the ε-optimality of σ.

The ε appearing in inequality (5) is essential. It is impossible to find for every two-person zero-sum stochastic game an ε-optimal strategy σ of player 1 such that for every n sufficiently large $E_{\sigma,\tau}(v(z_n) - v(z_1)) \geq 0$ for every strategy τ of player 2: the only such strategy σ in the big match is the one that always plays the non-absorbing action, and given such a strategy σ of player 1 there is a strategy τ of player 2 such that for $\varepsilon > 0$ sufficiently small and every n we have $E_{\sigma,\tau}\left(\frac{1}{n}\sum_{i=1}^n x_i\right) < v(z_1) - \varepsilon$.

The variable $v(z_n)$ represents the potential for payoffs starting at stage n. The above discussion shows that targeting the future potentials alone is necessary but insufficient; the player also has to reckon with the stream of payoffs $(x_n)_{n=1}^\infty$. Therefore, in addition to securing the future potential, player 1's ε-optimal strategy has to correlate the stream of payoffs $(x_n)_{n=1}^\infty$ to the stream of future potentials $(v(z_n))_{n=1}^\infty$.

The constructed ε-optimal strategies σ_ε of player 1 will thus guarantee in addition that for sufficiently large n,

$$E_{\sigma_\varepsilon,\tau}\left(\frac{1}{n}\sum_{i=1}^n (x_i - v(z_i))\right) \geq -\varepsilon \tag{6}$$

which together with inequality (5) guarantees that $E_{\sigma_\varepsilon,\tau}\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \geq v(z_1) - 2\varepsilon$.

The delicate point of the contraction of ε-optimal strategies is thus to find a strategy σ that guarantees both (5) and (6). We anchor the construction on the following inequality that holds for any behavioral strategy of player 1 that plays at stage i the optimal mixed action of the λ_i-discounted stochastic game:

$$E_{\sigma,\tau}\left(\lambda_i x_i + (1-\lambda_i) v_{\lambda_i}(z_{i+1}) \mid \mathcal{H}_i\right) \geq v_{\lambda_i}(z_i), \tag{7}$$

where $\mathcal{H}_i$ is the σ-algebra generated by the sequence $z_1, a_1, \ldots, z_i$ of states and actions up to the play at stage i. Moreover, player 1 can guarantee these inequalities to hold also for a sequence of discount rates λ_i that depends on the past history $(z_1, a_1, \ldots, z_i)$. The term $\lambda_i x_i + (1-\lambda_i) v_{\lambda_i}(z_{i+1})$ appearing in (7) is a weighted average of the stage payoff x_i and an approximation $v_{\lambda_i}(z_{i+1})$ of the future potential $v(z_{i+1})$. Note that inequality (7) states that the conditional expectations of this weighted average is larger than an approximation $v_{\lambda_i}(z_n)$ of the potential starting at stage n, $v(z_n)$.

In proving the existence of the minmax (of an I-player stochastic game with finitely many players, states and actions) we first define for every $\lambda > 0$ the functions $\bar{v}^i_\lambda : S \to \mathbb{R}$ by

$$\begin{aligned} \bar{v}^i_\lambda(z) &= \min_{\sigma^{-i}} \max_{\sigma^i} E_{\sigma^i,\sigma^{-i}}(\lambda\, r^i(z_1, a_1) + (1-\lambda)\bar{v}^i_\lambda(z_2) \mid z_1 = z) \\ &= \min_{y} \max_{x} \lambda\, r^i(z, x, y) + (1-\lambda) \sum_{z' \in S} p\,(z' \mid z, x, y)\, \bar{v}^i_\lambda(z') \end{aligned}$$

where the first min is over all $I \setminus \{i\}$-tuples of strategies $\sigma^{-i} = (\sigma^j)_{j \neq i}$ and the second min is over all $I \setminus \{i\}$-tuples $y = (y^j)_{j \neq i}$ of mixed actions $y^j \in \Delta(A^j(z))$; similarly, the first max is over all strategies σ^i of player i and the second max is over all mixed actions $x \in \Delta(A^i(z))$ of player i; $r(z, x, y)$ and $p\,(z' \mid z, x, y)$ are the multilinear extension of r and p respectively.

Next, we observe that the functions $\lambda \mapsto \bar{v}^i_\lambda(z)$ are bounded and semialgebraic and thus converge as $\lambda \to 0+$ to $\bar{v}^i(z)$, which will turn out to be the minmax of player i.

Next, we show that for every $\varepsilon > 0$ there is a positive integer $N = N(\varepsilon)$ and a sequence of discount rates $(\lambda_t)_{t=1}^\infty$ such that λ_t is measurable w.r.t. the algebras $\mathcal{H}_t$ and such that if σ is a strategy profile such that for every $t \geq 1$ we have

$$E_\sigma(\lambda_t\, r^i(z_t, a_t) + (1-\lambda_t)\, \bar{v}^i_{\lambda_t}(z_{t+1}) \mid \mathcal{H}_t) \leq \bar{v}^i_{\lambda_t}(z_t), \tag{8}$$

then for every $n \geq N$ inequalities (1) and (2) hold. Therefore an $I \setminus \{i\}$ strategy profile σ^{-i} such that for every strategy σ^i of player i the strategy profile (σ^{-i}, σ^i) obeys (8) is an ε-minimaxing $I \setminus \{i\}$ strategy profile.

Similarly, for every $\varepsilon > 0$ there is a positive integer $N = N(\varepsilon)$ and a sequence of discount rates $(\lambda_t)_{t=1}^\infty$ such that λ_t is measurable w.r.t. the $\mathcal{H}_t$ and such that if σ is a strategy profile such that for every $t \geq 1$ we have

$$E_\sigma(\lambda_t\, r^i(z_t, a_t) + (1-\lambda_t)\, \bar{v}^i_{\lambda_t}(z_{t+1}) \mid \mathcal{H}_t) \geq \bar{v}^i_{\lambda_t}(z_t), \tag{9}$$

then for every $n \geq N$ inequalities (3) and (4) hold. Given an $I \setminus \{i\}$ strategy profile $\sigma^{-i} = (\sigma^j)_{j \neq i}$, we can assume without loss of generality (using Kuhn's theorem) that σ^j is a behavioral strategy and therefore there exists a strategy σ^i of player i such that (9) holds and thus we conclude that $\bar{v}^i$ is indeed the maxmin of the stochastic game.

3. The Basic Lemma

The next lemma is stated as a lemma on stochastic processes. The statement of the lemma is essentially a reformulation of an implicit result in [2] and its proof is essentially identical to the proof there. Without needing to repeat and replicate the proof, the reformulation enables us to use an implicit result of [2] in various other applications, like (1) the present existence of the minmax in an n-player stochastic game, (2) the existence of the minmax of two-player stochastic games with imperfect monitoring [1], [4], [5], and (3) the existence of an extensive-form correlated equilibrium in n-player stochastic games [6].

We use symbols and notations that indicate its applicability to stochastic games. Let $(\Omega, \mathcal{H}_\infty)$ be a measurable space and $(\mathcal{H}_t)_{t=1}^\infty$ an increasing sequence of σ-fields with $\mathcal{H}_\infty$ the σ-field spanned by $\cup_{t=1}^\infty \mathcal{H}_t$. Assume that for every $0 < \lambda < 1$, $(r_{t,\lambda})_{t=1}^\infty$ is a sequence of (real-valued) random variables with values in [-1,1] such that $r_{t,\lambda}$ is measurable with respect to $\mathcal{H}_{t+1}$ and $(v_{t,\lambda})_{t=1}^\infty$ is a sequence of $[-1,1]$-valued functions such that $v_{t,\lambda}$ is measurable w.r.t. $\mathcal{H}_t$. In many applications to stochastic games, the measurable space Ω is the space of all infinite plays, and $\mathcal{H}_t$ is the σ-algebra generated by all finite histories $(z_1, a_1, \ldots, z_t)$. In stochastic games with imperfect monitoring the σ-algebra $\mathcal{H}_t$ may stand for (describe) the information available to a given player prior to his choosing an action at stage t; see [1], [4] and [5].

The random variable $v_{t,\lambda}$ may play the role of the value of the λ-discounted stochastic game as a function of the initial state z_t, or the minmax value of the λ-discounted stochastic game as a function of the initial state z_t. Note that in this case it is independent of t. More generally, $v_{t,\lambda}$ can stand for the solution of an auxiliary system of equations of the form $v_{t,\lambda} = \sup_x \inf_y f(x, y, t, \lambda)$ where the domain of x and y may depend on z_t, t and λ and the function f is measurable w.r.t. $\mathcal{H}_t$.

The random variable $r_{t,\lambda}$ may play the role of the t-th stage payoff to player i, i.e., $r^i(z_t, a_t)$. Note that in this case it is independent of λ. More generally, $r_{t,\lambda}$ can stand for a payoff of an auxiliary one-stage game that depends on the state z_t as well as on the discount parameter λ, in which case it does depend on λ .

Lemma 1 *Assume that for every $\delta > 0$ there exist two functions, $L(s)$ and $\lambda(s)$, of the real variable s, and a positive constant $M > 0$ such that λ is strictly decreasing with $0 < \lambda(s) < 1$ and L is integer-valued with $L(s) > 0$, and such that, for every $s \geq M$, $|\theta| \leq 3$ and $\omega \in \Omega$,*

$$4L(s) \leq \delta s \tag{10}$$

$$|\lambda(s + \theta L(s)) - \lambda(s)| \leq \delta\lambda(s) \tag{11}$$

$$|v_{n,\,\lambda(s+\theta L(s))}(\omega) - v_{n,\,\lambda(s)}(\omega)| \leq 4\delta L(s)\lambda(s) \tag{12}$$

$$\int_M^\infty \lambda(s)\, ds \leq \delta. \tag{13}$$

Then, a) the limit $\lim_{\lambda \to 0+} v_{t,\lambda}$ exists and is denoted $v_{t,\infty}$, and b) for every $\varepsilon > 0$ and $\lambda_0 > 0$ there is n_0 sufficiently large and a sequence $(\lambda_t)_{t=1}^\infty$ with $0 < \lambda_t < \lambda_0$ and λ_t measurable w.r.t. $\mathcal{H}_t$ such that for every probability P on $(\Omega, \mathcal{H}_\infty)$ with

$$E_P(\lambda_t r_{t,\lambda_t} + (1 - \lambda_t)v_{\lambda_t,t+1}) \mid \mathcal{H}_t) \geq v_{\lambda_t,t} - \varepsilon\lambda_t,$$

we have

$$E_P\left(\frac{1}{n}\sum_{t=1}^n r_{t,\lambda_t}\right) \geq v_{1,\infty} - 5\varepsilon \qquad \forall n \geq n_0 \tag{14}$$

$$E_P\left(\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^n r_{t,\lambda_t}\right) \geq v_{1,\infty} - 5\varepsilon \tag{15}$$

$$E_P(\sum_{t\geq 1} \lambda_t) < \infty. \tag{16}$$

In the case that $v_{t,\lambda}(\omega)$ is either the λ-discounted value of a two-player zero-sum stochastic game or the minmax (or maxmin) of player i of the λ-discounted stochastic game it is actually a function of the two variables λ and z_t (which depends obviously on ω and t). Whenever the stochastic game has finitely many states and actions, each one of the (finitely many) functions $\lambda \mapsto v_{t,\lambda}(\omega)$ is a bounded semialgebraic function. Therefore, the set of functions $\lambda \mapsto v_{t,\lambda}(\omega)$, where t and ω range over all positive integers t and all points $\omega \in \Omega$, is a finite set of bounded real-valued semialgebraic functions. In that case, the assumption and conclusion (a) of Lemma 1 hold. Indeed, it follows (see, e.g., [3]) that there is a constant $0 < \theta < 1$ and finitely many functions $f_j :]0, \theta] \to \mathbb{R}$, $j \in J$, which have a convergent expansion in fractional powers of λ: $f_j(\lambda) = \sum_{i=1}^\infty a_{i,j}\lambda^{i/m}$ where m is a positive integer, such that for every t and ω there is $j \in J$ such that for $0 < \lambda \leq \theta$, $v_{t,\lambda}(\omega) = f_j(\lambda)$. Therefore, one could take $L(s) = 1$, $\lambda(s) =$

$s^{-1-\frac{1}{M}}$ where M is sufficiently large. Alternatively, one can choose $L(s)=1$, $\lambda(s)=1/(s\ln^2 s)$, and M sufficiently large. Conclusion (a) holds since the limit (as $\lambda \to 0+$) of a bounded semialgebraic function exists.

Proof of Lemma 1. We assume w.l.o.g. that $\delta < 1/4$. We first note that conditions (10), (11), (12) and (13) on the positive constant M, the strictly decreasing function $\lambda : [M,\infty) \to (0,1)$ and the integer-valued function $L : [M,\infty) \to \mathbb{N}$ imply that the limit $\lim_{\lambda\to 0+} v_{n,\lambda}(\omega)$ exists for every ω and n, and denoting this limit by $v_{n,\infty}(\omega)$ we have $v_{n,\lambda(s)}(\omega) \to_{s\to\infty} v_{n,\infty}(\omega)$ and

$$|v_{n,\lambda(s)}(\omega) - v_{n,\infty}(\omega)| \le \delta. \tag{17}$$

Indeed, define inductively $q_1 = M$ and $q_{k+1} = q_k + 3L(q_k)$. It follows from (11) that $\lambda(s) \ge (1-\delta)\lambda(q_k)$ for every $q_k \le s \le q_{k+1}$ and thus $\int_{q_k}^{q_{k+1}} \lambda(s)ds \ge 3L(q_k)(1-\delta)\lambda(q_k)$. Therefore,

$$\sum_{k=1}^{\infty} 4\delta L(q_k)\lambda(q_k) \le \frac{4\delta}{3(1-\delta)} \int_M^\infty \lambda(s)ds$$

which by (13) (and using the inequality $\delta < 1/4$) is $\le \delta/2$.

Using (12), the sequence $(v_{\lambda(q_k),\,n})_{k=1}^{\infty}$ is a Cauchy sequence and thus it converges to a limit, $v_{n,\,\infty}$, and

$$|v_{n,\lambda(q_k)} - v_{n,\,\infty}| \le \sum_{k=1}^{\infty} 4\delta L(q_k)\lambda(q_k) \le \frac{4\delta}{3(1-\delta)} \int_M^\infty \lambda(s)ds \le \delta/2.$$

Given $s \ge M$, let k be the largest positive integer such that $q_k \le s$. It follows that $s = q_k + \theta L(q_k)$ with $0 \le \theta \le 3$, and thus, using (12), $|v_{n,\lambda(s)} - v_{n,\lambda(q_k)}| \le 4\delta L(q_k)\lambda(q_k) \to_{k\to\infty} 0$. Therefore $v_{n,\lambda(s)} \to_{s\to\infty} v_{n,\infty}$ (moreover, the convergence is uniform) and $|v_{n,\lambda(s)} - v_{n,\infty}| \le \delta$ for every $s \ge M$.

Recall that the above step is redundant in the special case where the set of functions $\lambda \mapsto v_{n,\,\lambda}(\omega)$, $0 < \lambda \le 1$, where n and ω range over all positive integers and all points $\omega \in \Omega$, constitute a finite set of bounded semialgebraic functions.

We now continue with the proof. Fix $\varepsilon > 0$ sufficiently small ($\varepsilon < 1/2$) and set $\delta = \varepsilon/12$. As λ is strictly decreasing and integrable by (13), $\lim_{s\to\infty} s\lambda(s) = 0$; hence by (10) it follows that $\lim_{s\to\infty} \lambda(s)L(s) = 0$ and therefore by choosing M sufficiently large

$$\lambda(s)L(s) \le \delta \quad \text{for} \quad s \ge M. \tag{18}$$

Define inductively, starting with $s_0 \ge M$:

$$L_k = L(s_k), \quad B_{k+1} = B_k + L_k, \quad B_0 = 1,$$

$$s_{k+1} = \max\{\, M,\, s_k + \sum_{B_k \le\, i \,<\, B_{k+1}} (r_{i,\, \lambda(s_k)} - v_{B_{k+1},\, \lambda(s_k)} + 2\varepsilon)\}.$$

Define $\lambda_i = \lambda(s_k)$ for $B_k \le i < B_{k+1}$. Let P be a distribution on Ω such that for every $j \ge 1$,

$$E_P(\lambda_j\, r_{j,\lambda_j} + (1-\lambda_j)\, v_{j+1,\lambda_j} \mid \mathcal{H}_j) \ge v_{j,\lambda_j} - \varepsilon\lambda_j = v_{j,\lambda_j} - 12\delta\lambda_j. \quad (19)$$

In order to simplify the notation in the computations below we denote $\alpha_k = \lambda_{B_k}$, $w_k = v_{B_k,\alpha_k}$, $\mathcal{F}_k = \mathcal{H}_{B_k}$ and $t_k = \int_{s_k}^{\infty} \lambda(s)\, ds$. Define

$$Y_k = v_{B_k,\alpha_k} - t_k.$$

We will prove that $(Y_k)_{k=0}^{\infty}$ is a submartingale adapted to the increasing sequence of σ-fields $(\mathcal{F}_k)_{k=0}^{\infty}$, and moreover that

$$E_P(Y_{k+1} - Y_k \mid \mathcal{F}_k) \ge 3\delta L_k \alpha_k. \quad (20)$$

Note that $Y_{k+1} - Y_k = v_{B_{k+1},\alpha_{k+1}} - v_{B_k,\alpha_k} + \int_{s_k}^{s_{k+1}} \lambda(s)ds$. In the computations that follow and prove (20) we replace $v_{B_{k+1},\alpha_{k+1}}$ by v_{B_{k+1},α_k} and the resulting error term is $\le 4\delta L_k\alpha_k$, and we replace the term $\int_{s_k}^{s_{k+1}} \lambda(s)ds$ by $\alpha_k(s_{k+1} - s_k)$ and the resulting error term is bounded by $3\delta\alpha_k L_k$. The definition of s_{k+1} implies that $s_{k+1} - s_k \ge 24\delta L_k + \sum_{B_k \le i < B_{k+1}} (r_{i,\alpha_k} - v_{B_{k+1},\alpha_k})$ and we bound the sum $\sum_{B_k \le i < B_{k+1}} (r_{i,\alpha_k} - v_{B_{k+1},\alpha_k})$ from below, using the inequalities $1 \ge (1-\alpha_k)^j \ge 1 - \alpha_k L_k$ for $1 \le j \le L_k$, with

$$\left(\sum_{B_k \le i < B_{k+1}} (1-\alpha_k)^{i-B_k} (r_{i,\alpha_k} - v_{B_{k+1},\alpha_k}) \right) - 2\alpha_k L_k^2.$$

The assumption that the functions $r_{i,\lambda}$ and $v_{n,\lambda}$ are $[-1,1]$-valued and that $\varepsilon < 1/2$ imply that $|r_{i,\lambda}| + |v_{n,\lambda}| + 2\varepsilon < 3$. Therefore, for every k, there is $|\theta| \le 3$ such that $s_{k+1} - s_k = \theta L_k$. Therefore,

$$|s_{k+1} - s_k| \le 3L_k \quad (21)$$

and it follows from (12) that

$$|v_{B_{k+1},\alpha_k} - v_{B_{k+1},\alpha_{k+1}}| \le 4\delta L_k \alpha_k. \quad (22)$$

Fix $k \ge 1$ and let $g_i = r_{B_k+i,\alpha_k}$ and $u_i = v_{B_k+i,\alpha_k}$ for $0 \le i \le L_k$. Taking conditional expectations (with respect to $\mathcal{F}_k$) of the inequalities (19) for $B_k \le j = B_k + i < B_{k+1}$, $0 \le i < L_k$, and multiplying the resulting inequality by $(1-\alpha_k)^i$ we have for every $0 \le i < L_k$,

$$E_P\left(\alpha_k(1-\alpha_k)^i g_i + (1-\alpha_k)^{i+1} u_{i+1} - (1-\alpha_k)^i u_i \mid \mathcal{F}_k\right) \ge -\varepsilon\alpha_k = -12\delta\alpha_k.$$

Summing the above inequalities over $0 \le i < L_k$ we have

$$E_P\left(\alpha_k \sum_{i=0}^{L_k-1} (1-\alpha_k)^i g_i + (1-\alpha_k)^{L_k} u_{L_k} - u_0 \mid \mathcal{F}_k\right) \ge -12\delta L_k \alpha_k,$$

or, as $1 - \lambda \sum_{0 \le i < L} (1-\lambda)^i = (1-\lambda)^L$,

$$E_P\left(u_{L_k} - u_0 + \alpha_k \sum_{0 \le i < L_k} (1-\alpha_k)^i (g_i - u_{L_k}) \mid \mathcal{F}_k\right) \ge -12\delta L_k \alpha_k.$$

The inequalities $1-\lambda L \le (1-\lambda)^i \le 1$, $0 \le i \le L$, imply that $\sum_{0 \le i < L_k}(g_i - u_{L_k}) \ge \sum_{0 \le i < L_k}(1-\alpha_k)^i(g_i - u_{L_k}) - 2L_k L_k \alpha_k$. By (18), $L_k L_k \alpha_k < \delta L_k$, and therefore we have

$$E_P(u_{L_k} - u_0 + \alpha_k \sum_{0 \le i < L_k} (g_i - u_{L_k}) \mid \mathcal{F}_k) \ge -14\delta L_k \alpha_k.$$

Hence, using $s_{k+1} - s_k \ge \sum_{0 \le i < L_k}(g_i - u_{L_k} + 2\varepsilon)$ by the definition of s_{k+1}, and $v_{B_{k+1},\alpha_k} \le v_{B_{k+1},\alpha_{k+1}} + 4\delta L_k \alpha_k$ by (22), we deduce that

$$E(w_{k+1} - w_k + \alpha_k(s_{k+1} - s_k) \mid \mathcal{F}_k) \ge (-14-4)\delta L_k \alpha_k + 2\varepsilon L_k \alpha_k = 6\delta L_k \alpha_k.$$

Finally, as $\alpha_k(s_{k+1} - s_k) \le \int_{s_k}^{s_{k+1}} \lambda(s)ds + 3\delta \alpha_k L_k$ (using (21) and (11)), we deduce that

$$E_P(w_{k+1} - w_k + \int_{s_k}^{s_{k+1}} \lambda(s)ds \mid \mathcal{F}_k) \ge 3\delta L_k \lambda_k,$$

i.e.,

$$E_P(Y_{k+1} - Y_k \mid \mathcal{F}_k) \ge 3\delta L_k \alpha_k,$$

which proves (20).

The random variables $Y_k - Y_j$ are bounded by 3. Therefore,

$$3 \ge E(Y_k - Y_0) \ge 3\delta E_P\left(\sum_{i<k} L_i \alpha_i\right).$$

Hence, by the monotone convergence theorem,

$$E_P\left(\sum_k L_k \alpha_k\right) \le 1/\delta.$$

As $L_k\alpha_k \geq L(M)\lambda(M)I_{s_k=M} \geq \lambda(M)I_{s_k=M}$, it follows that

$$E_P\left(\sum_k \lambda(M)I_{s_k=M}\right) \leq 1/\delta.$$

Therefore,

$$E_P\left(\sum_k I_{s_k=M}\right) \leq \frac{1}{\delta\lambda(M)}\ . \tag{23}$$

Also, as $\sum_k L_k\alpha_k = \sum_t \lambda_t$, (16) follows. Let $k(i)$ be the smallest integer such that B_k is greater than i. For every i, $k(i)$ is a random variable which is measurable w.r.t. $\mathcal{H}_i$; and let $\ell_i = v_{B_{k(i)},\alpha_{k(i)-1}}$. Note that for $B_k \leq i < B_{k+1}$ we have $v_{B_{k+1},\alpha_k} = \ell_i$. The definition of s_k implies that if $s_{k+1} \neq M$ then $s_{k+1} - s_k = \left(\sum_{B_k\leq i<B_{k+1}}(r_{i,\alpha_k} - \ell_i)\right) + 24\delta L_k$, and that if $s_{k+1} = M$ then $s_{k+1} - s_k \leq 0 \leq \sum_{B_k\leq i<B_{k+1}}(r_{i,\alpha_k} - \ell_i) + 24\delta L_k + 2L_k$. Hence,

$$s_{k+1} - s_k \leq \left(\sum_{B_k\leq i<B_{k+1}} (r_{i,\alpha_k} - \ell_i)\right) + 24\delta L_k + 2L_k I_{s_{k+1}=M}.$$

Summing the above inequalities over $k' \leq k$ and rearranging the terms we have

$$\sum_{i<B_k} r_{t,\lambda_t} \geq s_k - s_0 + \sum_{t<B_k} \ell_t - 24\delta B_k - \delta M \sum_{k=0}^{\infty} I_{s_{k+1}=M}.$$

Hence, for any n we have

$$\sum_{i=1}^{n} r_{i,\lambda_i} \geq \sum_{i=1}^{n} \ell_i - 3(B_{k(n)} - n) - 24\delta n - s_0 - \delta M \sum_{k=0}^{\infty} I_{s_{k+1}=M}. \tag{24}$$

Note that $s_k \leq s_0 + 3B_k$ for every k, and thus

$$B_{k(n)} - n \leq L(s_{k(n)-1}) \leq \delta s_{k(n)-1}/4 \leq (\delta/4)(s_0 + 3n). \tag{25}$$

Therefore, using also (23) and the bound $B_{k(n)} - n \leq (\delta/4)s_0 + \delta n$, we have

$$E(\frac{1}{n}\sum_{t=1}^{n} r_{t,\lambda_t}) \geq E(\frac{1}{n}\sum_{t=1}^{n} \ell_t) - 2\varepsilon - \frac{4}{n}(\delta n) - \frac{2s_0}{n} - \frac{M}{n\lambda(M)}.$$

Since for sufficiently large k we have $3\delta + \frac{4}{n}(\delta n) + \frac{2s_0}{B_k} + \frac{M}{B_k\lambda(M)} < \varepsilon$ and $E_P(\ell_i) \geq v_{1,\infty} - 2\varepsilon$, inequality (14) follows.

Next, as (Y_k) is a bounded submartingale we deduce that P a.e. $Y_k \to_{k\to\infty} Y_\infty$ and that $E_P(Y_\infty \mid \mathcal{F}_1) \geq Y_1$. It follows from (16) (which has already been proved) that $\lambda_t \to 0$ a.e. (w.r.t. P), and therefore P a.e. $t_k \to 0$ and thus $\ell_i \to Y_\infty$ a.e. implying that $\frac{1}{n}\sum_{t=1}^{n} \ell_t \to Y_\infty$ a.e. As $\sum_{k=1}^{\infty} I_{s_k=M}$ is integrable by (23), it is finite a.e. and therefore we have

$$\frac{1}{n}\sum_{k=1}^{\infty} I_{s_k=M} \to_{n\to\infty} 0 \quad P \text{ a.e.}$$

Obviously, $\frac{s_0}{n} \to_{n\to\infty} 0$. Therefore, we deduce that

$$E_P(\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} r_{t,\lambda_t}) \geq v_{1,\infty} - 5\varepsilon,$$

which proves (15).

The next lemma provides conditions on the random variable $v_{n,\lambda}$ which imply the assumption of the previous lemma.

Lemma 2 *Assume that 1) the random variables $v_{t,\lambda}/\lambda$ are uniformly Lipschitz as a function of $1/\lambda$ and that 2) for every $\alpha > 0$ there exists a sequence λ_i $(0 < \lambda_i < 1)$ such that $\lambda_{i+1} \geq \alpha\lambda_i$, $\lim_{i\to\infty}\lambda_i = 0$ and $\sum_{i=1}^{\infty} \|v_{\lambda_i,\cdot} - v_{\lambda_{i+1},\cdot}\| < \infty$ where $\|\cdot\|$ is the supremum (over $n \geq 1$ and $\omega \in \Omega$) norm. Then the assumption of Lemma 1 holds.*

Whenever the random variables $v_{n,\lambda}$ are the values or the minmax or the maxmin of the λ-discounted stochastic game (with uniformly bounded stage payoff), assumption (1) of Lemma 2 holds.

The proof of Lemma 2 can be found in [2]. The assumption that the variables $v_{n,\lambda}/\lambda$ are uniformly Lipschitz as a function of $1/\lambda$ is a corollary of the assumption there that $v_{n,\lambda}$ are the values of the λ-discounted stochastic game with uniformly bounded stage payoff.

4. Existence of the Minmax

Let Γ be a two-player zero-sum stochastic game with finitely many states and actions. For every state $z \in S$ and player $i = 1, 2$, let $X^i(z)$ be a nonempty subset of $\Delta(A^i(z))$, the mixed actions available to player i at state z. Set $X^i = (X^i(z))_{z\in S}$ and $X = (X^i)_{i=1,2}$. An X^i-*constrained strategy* of player i is a behavioral strategy σ such that for every finite history $z_1, a_1, \ldots, z_t$ we have $\sigma^i(z_1, a_1, \ldots, z_t) \in X^i(z_t)$.

The λ-discounted minmax (of player 1) in the X-constrained stochastic game, $\bar{w}^1_\lambda \in \mathbb{R}^S$, is defined by

$$\bar{w}^1_\lambda(z) = \inf_\tau \sup_\sigma E^z_{\sigma,\tau}\left(\lambda \sum_{t=1}^{\infty} (1-\lambda)^{t-1} r(z_t, a_t)\right)$$

where the supremum is over all X^1-constrained strategies σ of player 1 and the infimum is over all X^2-constrained strategies τ of player 2.

Similarly, the λ-discounted maxmin (of player 1) in the X-constrained stochastic game, $\underline{w}^1_\lambda \in \mathbb{R}^S$, is defined by

$$\underline{w}^1_\lambda(z) = \sup_\sigma \inf_\tau E^z_{\sigma,\tau}\left(\lambda \sum_{t=1}^{\infty}(1-\lambda)^{t-1} r(z_t, a_t)\right)$$

where the supremum is over all X^1-constrained strategies σ of player 1 and the infimum is over all X^2-constrained strategies τ of player 2.

For every $0 < \lambda < 1$ we consider the system of equations

$$w(z) = \sup_x \inf_y \left(\lambda\, r(z,x,y) + (1-\lambda)\sum_{z' \in S} p\,(z' \mid z,x,y)\, w(z')\right) \tag{26}$$

in the variable $w(z)$, $z \in S$, and where the sup is over all $x \in X^1(z)$ and the inf is over all $y \in X^2(z)$. The system of equations depends on the data $\langle S, A, r, p\rangle$. As we show below, its solution $w_\lambda \in \mathbb{R}^S$ turns out to be the λ-discounted maxmin of player 1.

We use the classical contraction argument to show that the system (26) has a unique solution: for every $0 < \lambda < 1$ the map $T : \mathbb{R}^S \to \mathbb{R}^S$ where

$$[Tw](z) = \sup_{x \in X^1(z)} \inf_{y \in X^2(z)} \left(\lambda\, r(z,x,y) + (1-\lambda)\sum_{z' \in S} p\,(z' \mid z,x,y) w(z')\right)$$

is a strict contraction, and thus has a unique fixed point $w_\lambda \in \mathbb{R}^S$. Let w_λ be the unique fixed point of the contraction map T. Finally, a point $w \in \mathbb{R}^S$ is a solution of (26) if and only if it is a fixed point of T.

The definition of w_λ enables us to construct, as a function of any $(0,1)$-valued function λ defined on all finite histories (equivalently, a sequence of $(0,1)$-valued functions λ_t defined on all histories $z_1, a_1, \ldots, z_t$ of length t): a) an X^1-constrained strategy of player 1, and b) for every X^1-constrained strategy σ of player 1 an X^2-constrained strategy $\tau = \tau(\sigma)$, such that a proper system of inequalities holds. The details follow.

The definition of w_λ implies that:
1) for every sequence $(\lambda_t)_{t=1}^{\infty}$ with $0 < \lambda_t < 1$ and λ_t measurable w.r.t. $\mathcal{H}_t$, there is an X^1-constrained strategy σ of player 1 such that for every history $h = (z_1, \ldots, z_t)$ and every mixed action $y \in X^2(z_t)$,

$$\lambda_t r(z_t, \sigma(h), y) + (1-\lambda_t)\sum_{z' \in S} p\,(z' \mid z_t, \sigma(h), y) w_{\lambda_t}(z')) \geq w_{\lambda_t}(z_t) - \varepsilon\lambda_t$$

where λ_t also stands for $\lambda_t(h)$ for short, and thus for every X^2-constrained strategy τ of player 2 we have

$$E_{\sigma,\tau}(\lambda_t\, x_t + (1-\lambda_t)\, w_{\lambda_t}(z_{t+1}) \mid \mathcal{H}_t) \geq w_{\lambda_t}(z_t) - \varepsilon\lambda_t \tag{27}$$

where x_t stands for $r(z_t, a_t)$ for short, and
2) for every sequence $(\lambda_t)_{t=1}^{\infty}$ with $0 < \lambda_t < 1$ and λ_t measurable w.r.t. $\mathcal{H}_t$, and every X^1-constrained strategy σ of player 1 there is an X^2-constrained strategy τ of player 2 such that for every history $h = (z_1, \ldots, z_t)$,

$$\lambda_t r(z_t, \sigma(h), \tau(h)) + (1-\lambda_t) \sum_{z' \in S} p\,(z' \mid z_t, \sigma(h), \tau(h)) w_{\lambda_t}(z')) \leq w_{\lambda_t}(z_t) + \varepsilon\lambda_t$$

and thus

$$E_{\sigma,\tau}(\lambda_t\, x_t + (1-\lambda_t)\, w_{\lambda_t}(z_{t+1}) \mid \mathcal{H}_t) \leq w_{\lambda_t}(z_t) + \varepsilon\lambda_t. \tag{28}$$

In particular, if $\lambda_t = \lambda$ is a constant discount rate, it follows (by multiplying the inequalities (27) by $(1-\lambda)^{t-1}$ and summing over all $t \geq 1$) that there is an X^1-constrained strategy σ of player 1 such that for every X^2-constrained strategy τ of player 2, $E_{\sigma,\tau}\left(\sum_{t=1}^{\infty} \lambda(1-\lambda)^{t-1} x_t\right) \geq w_\lambda(z_1) - \varepsilon$, and for any X^1-constrained strategy σ of player 1 there is an X^2-constrained strategy τ of player 2 such that $E_{\sigma,\tau}\left(\sum_{t=1}^{\infty} \lambda(1-\lambda)^{t-1} x_t\right) \leq w_\lambda(z_1) + \varepsilon$. Therefore, w_λ is the λ-discounted maxmin of the constrained stochastic game.

We prove the existence of the minmax and maxmin under the additional assumption that the constrained sets $X^i(z)$ are semialgebraic subsets of $\Delta(A^i(z))$, which is thus assumed in the sequel.

The map $\lambda \mapsto w_\lambda$ is bounded and semialgebraic [3] and therefore the function $\lambda \mapsto w_\lambda$ is of bounded variation on $]0,1]$ and thus in particular it has a limit w_∞ as $\lambda \to 0+$.

Theorem 1 *(a) For every $\varepsilon > 0$ there is n_0 sufficiently large and an X^1-constrained strategy σ of player 1 such that for every X^2-constrained strategy τ of player 2, the following inequalities hold:*

$$E_{\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} r_t(z_t, a_t)\right) \geq w_\infty(z_1) - \varepsilon \quad \forall n \geq n_0$$

and

$$E_{\sigma,\tau}\left(\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} r_t(z_t, a_t)\right) \geq w_\infty(z_1) - \varepsilon.$$

(b) For every $\varepsilon > 0$ there is n_0 sufficiently large such that for every X^1-constrained strategy σ of player 1, there is an X^2-constrained strategy τ of player 2 such that:

$$E_{\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} r_t(z_t, a_t)\right) \leq w_\infty(z_1) + \varepsilon \quad \forall n \geq n_0$$

and

$$E_{\sigma,\tau}\left(\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} r_t(z_t, a_t)\right) \leq w_\infty(z_1) + \varepsilon.$$

Moreover, these "maximizing" and "minimizing" constrained strategies, σ in part (a) and $\tau = \tau(\sigma)$ in part (b), can be chosen as arbitrary strategies that satisfy a proper list of inequalities. The following parts (a) and (b*) are generalizations of parts (a) and (b) respectively.*

(a) For every $\varepsilon > 0$ and $\lambda_0 > 0$ there is n_0 sufficiently large and a sequence $(\lambda_t)_{t=1}^{\infty}$ with $0 < \lambda_t < \lambda_0$ and λ_t measurable w.r.t. $\mathcal{H}_t$ such that for every strategy σ of player 1 such that for every history $h = (z_1, \ldots, z_t)$ and every mixed action $y \in X^2(z_t)$,*

$$\lambda_t r(z_t, \sigma(h), y) + (1 - \lambda_t) \sum_{z' \in S} p\left(z' \mid z_t, \sigma(h), y\right) w_{\lambda_t}(z') \geq w_{\lambda_t}(z_t) - \varepsilon\lambda_t,$$

the following inequalities hold:
for every $(X^2(z))_{z\in S}$-constrained strategy τ of player 2,

$$E_{\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} r_t(z_t, a_t)\right) \geq w_\infty(z_1) - 5\varepsilon \quad \forall n \geq n_0$$

and

$$E_{\sigma,\tau}\left(\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} r_t(z_t, a_t)\right) \geq w_\infty(z_1) - 5\varepsilon.$$

(b) For every $\varepsilon > 0$ and $\lambda_0 > 0$ there is n_0 sufficiently large and a sequence $(\lambda_t)_{t=1}^{\infty}$ with $0 < \lambda_t < \lambda_0$ and λ_t measurable w.r.t. $\mathcal{H}_t$ such that for every X^1-constrained strategy σ of player 1, and every strategy τ of player 2 such that for every history $h = (z_1, \ldots, z_t)$,*

$$\lambda_t r(z_t, \sigma(h), \tau(h)) + (1 - \lambda_t) p\left(z_t, \sigma(h), \tau(h)\right) \cdot w_{\lambda_t} \leq w_{\lambda_t}(z_t) + \varepsilon\lambda_t,$$

the following inequalities hold:

$$E_{\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} r_t(z_t, a_t)\right) \leq w_\infty(z_1) + 5\varepsilon \quad \forall n \geq n_0$$

and

$$E_{\sigma,\tau}\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} r_t(z_t,a_t)\right) \le w_\infty(z_1) + 5\varepsilon.$$

Proof. W.l.o.g. we assume that the payoff function r has values in $[-1,1]$. It follows that the solution of the system of equations (26) is also $[-1,1]$-valued. Let Ω be the measurable space of all plays of the stochastic game, and $\mathcal{H}_t$ the σ-field spanned by all histories $z_1, a_1 \dots, z_t$. Setting $v_{n,\lambda}(\omega) = w_\lambda(z_n)$ we deduce that the assumption of Lemma 1 holds, i.e., for every $\delta > 0$ there exist two functions, $L(s)$ and $\lambda(s)$, of the real variable s, and a positive constant $M > 0$ such that λ is strictly decreasing with $0 < \lambda(s) < 1$ and L is integer-valued with $L(s) > 0$, and such that, for $s \ge M$, $|\theta| \le 3$ and every $\omega \in \Omega$ inequalities (10), (11), (12), and (13) hold.

Fix $\varepsilon > 0$ sufficiently small (e.g., $\varepsilon < 3$) and set $\delta = \varepsilon/12$. Set $r_{n,\lambda} = r(z_n, a_n)$.

By the basic probabilistic lemma there is for every $\lambda_0 > 0$ a sufficiently large positive integer n_0 and a sequence λ_t with $0 < \lambda_t < \lambda_0$ such that λ_t is measurable w.r.t. $\mathcal{H}_t$ and such that for every probability P on $(\Omega, \mathcal{H}_\infty)$ with

$$E_P(\lambda_t r(z_t,a_t) + (1-\lambda_t)w_{\lambda_t}(z_{t+1})) \mid \mathcal{H}_t) \ge w_{\lambda_t} - \varepsilon\lambda_t,$$

inequalities (14), (15), and (16) hold; i.e., we have

$$E_P\left(\frac{1}{n}\sum_{t=1}^{n} r(z_t,a_t)\right) \ge w_\infty - 5\varepsilon \quad \forall n \ge n_0 \tag{29}$$

$$E_P\left(\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} r(z_t,a_t)\right) \ge w_\infty - 5\varepsilon \tag{30}$$

$$E_P(\sum_{i\ge 1}\lambda_i) < \infty. \tag{31}$$

Fix such a sequence $(\lambda_t)_{t=1}^{\infty}$. By the definition of w_λ there is a strategy σ of player 1 such that for every history $h = (z_1, \dots, z_t)$ and every mixed action $y \in X^2(z_t)$,

$$\lambda_t E^{z_t}_{\sigma(h),y}(r(z_t,a_t)) + (1-\lambda_t)E^{z_t}_{\sigma(h),y}(w_{\lambda_t}(z_{t+1})) \ge w_{\lambda_t}(z_t) - \varepsilon\lambda_t.$$

Therefore, for every X^2-constrained strategy τ of player 2 we have

$$E_{\sigma,\tau}(\lambda_t r(z_t,a_t) + (1-\lambda_t)w_{\lambda_t}(z_{t+1}) \mid \mathcal{H}_t) \ge w_{\lambda_t} - \varepsilon\lambda_t$$

and thus inequalities (29), (30), and (31) hold.

Similarly, by the basic probabilistic lemma there is for every $\lambda_0 > 0$ a sufficiently large positive integer n_0 and a sequence $(\lambda_t)_{t=1}^{\infty}$ with $0 < \lambda_t < \lambda_0$ such that λ_t is measurable w.r.t. $\mathcal{H}_t$ and such that for every probability P on $(\Omega, \mathcal{H}_\infty)$ with

$$E_P(\lambda_t r(z_t, a_t) + (1 - \lambda_t) w_{\lambda_t}(z_{t+1}) \mid \mathcal{H}_t) \le w_{\lambda_t} + \varepsilon \lambda_t,$$

we have

$$E_P\left(\frac{1}{n}\sum_{t=1}^{n} r(z_t, a_t)\right) \le w_\infty + 5\varepsilon \qquad \forall n \ge n_0 \tag{32}$$

$$E_P\left(\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} r(z_t, a_t)\right) \le w_\infty + 5\varepsilon \tag{33}$$

$$E_P(\sum_{i\ge 1} \lambda_i) < \infty. \tag{34}$$

Fix such a sequence $(\lambda_t)_{t=1}^{\infty}$. It follows from the definition of w_λ that for any X^1-constrained strategy of player 1, there is an X^2-constrained strategy τ of player 2 such that

$$E_{\sigma,\tau}\left(\lambda_t r(z_t, a_t) + (1 - \lambda_t) w_{\lambda_t}(z_{t+1}) \mid \mathcal{H}_t\right) \le w_{\lambda_t} + \varepsilon \lambda_t$$

and thus inequalities (32), (33), and (34) hold. ■

Theorem 1 establishes the existence of the maxmin of player 1 in a two-player constrained stochastic game with semialgebraic constraints and finitely many states and actions. By duality, the minmax exists.

Consider an I-player stochastic game with finitely many states and actions and standard signaling (perfect monitoring). Every $I \setminus \{i\}$ strategy profile is equivalent by Kuhn's theorem to a strategy profile $\sigma^{-i} = (\sigma^j)_{j \ne i}$ of behavioral strategies. Therefore, the study of player i's minmax in the I-player stochastic game is equivalent to the study of the minmax of player 1 in a two-player constrained stochastic game where player 1 (represents player i and) has action sets $A^i(z)$ and is not constrained, i.e., $X^1(z) = \Delta(A^1(z))$, and player 2 (represents the set $I \setminus \{i\}$ of players and) has action sets $\times_{j\ne i} A^i(z)$ and with constraint sets $X^2(z) = \times_{j \ne i} \Delta(A^j(z))$. The set $X^2(z)$ is a semialgebraic subset of $\Delta(\times_{j\ne i} A^j(z))$. Thus, the existence of the minmax of player i is a direct corollary of Theorem 1.

Before stating the corollary, we recall that the existence of the uniform minmax of player i, $\bar{v}^i(z_1)$, in an I-player stochastic game with initial state z_1 and standard signaling, implies that the λ-discounted minmax of player i, $\bar{v}^i_\lambda(z_1)$, converges as $\lambda \downarrow 0$ to $\bar{v}^i(z_1)$. Moreover, if the stochastic game has a uniform minmax of player i the convergence is uniform in $z = z_1$.

Corollary 1 *Fix an I-player stochastic game with finitely many states and actions.*
a) The minmax $\bar{v}^i : S \to \mathbb{R}$ of player i exists.
Moreover,
b) for every $0 < \lambda_0 < 1$ and $\varepsilon > 0$ there is a sequence of discount rates $0 < \lambda_t < \lambda_0$, where λ_t is measurable w.r.t. $\mathcal{H}_t$, such that every $I \setminus \{i\}$ strategy profile σ^{-i} that satisfies: for every strategy σ^i of player i and every $t \geq 1$ we have

$$E^z_{\sigma^{-i},\sigma^i}\left(\lambda_t r^i(z_t, a_t) + (1-\lambda_t)v_{l_t}(z_{t+1}) \mid \mathcal{H}_t\right) \leq v_{\lambda_t}(z_t) + \varepsilon\lambda_t/5,$$

is an ε-minimaxing $I \setminus \{i\}$ strategy profile, and
c) for every $0 < \lambda_0 < 1$ and $\varepsilon > 0$ there is a positive integer N and a sequence of discount rates $0 < \lambda_t < \lambda_0$, where λ_t is measurable w.r.t. $\mathcal{H}_t$, such that for every $I \setminus \{i\}$ strategy profile σ^{-i}, if the strategy $\sigma^i = \sigma^i(\sigma^{-i})$ of player i satisfies for every $t \geq 1$ the inequality

$$E^z_{\sigma^{-i},\sigma^i}\left(\lambda_t r^i(z_t, a_t) + (1-\lambda_t)v_{l_t}(z_{t+1}) \mid \mathcal{H}_t\right) \geq v_{\lambda_t}(z_t) - \varepsilon\lambda_t/5,$$

then σ^i is a σ^{-i}-ε-N-maximizing strategy.

The conclusions of Corollary 1 and Theorem 1 also apply to stochastic games with infinitely many states and actions whenever the payoffs are bounded and the solutions w_λ of (26) obey assumption (2) of Lemma 2.

It should be pointed out that throughout this paper we have stressed in addition to the main conclusion a structural property of the established minimaxing (or optimal) strategies. The advantage of the additional structural property is that it can be used to derive various results concerning, e.g., the existence of stationary minimaxing strategies when additional structure, e.g., irreducibility, of the stochastic game is given.

References

1. Coulomb, J.M. (2002) Stochastic games without perfect monitoring, mimeo.
2. Mertens, J.-F. and Neyman, A. (1981) Stochastic games, *International Journal of Game Theory* **10**, 53–66.
3. Neyman, A. (2003) Real algebraic tools in stochastic games, in A. Neyman and S. Sorin (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 6, pp. 57–75.
4. Rosenberg, D., Solan, E. and Vieille, N. (2001) On the maxmin value of stochastic games with imperfect monitoring, Discussion Paper 1337, Center for Mathematical Studies in Economics and Management Science, Northwestern University.
5. Rosenberg, D., Solan, E. and Vieille, N. (2002) Stochastic games with a single controller and incomplete information, Discussion Paper 1341, Center for Mathematical Studies in Economics and Management Science, Northwestern University.
6. Solan, E. and Vieille, N. (2002) Correlated equilibrium in stochastic games, *Games and Economic Behavior* **38**, 362–399.