

Discounted Stochastic Games

EILON SOLAN[†]

October 26, 1998

Abstract

We give an alternative proof to a result of Mertens and Parthasarathy, stating that every n -player discounted stochastic game with general setup, and with a norm-continuous transition, has a subgame perfect equilibrium.

[†] Institute of Mathematics and Center for Rationality and Interactive Decision Theory, The Hebrew University, Givat Ram, 91904 Jerusalem, Israel.
e-mail: eilons@math.huji.ac.il

This paper is part of the Master thesis of the author done with Prof. Ehud Lehrer in Tel Aviv University. I would like to thank Prof. Lehrer for many discussions and ideas and for the continuous help he offered. My gratitude is also given to Abraham Neyman, Sergiu Hart and J.F. Mertens for the discussions we had on the subject. I deeply thank the associate editor and two anonymous referees for many comments which substantially improved the paper.

Key Words: Discounted stochastic games, subgame perfect equilibrium, uncountable state space.

1 Introduction

Discounted stochastic games were introduced by Shapley (1953), who proved that every two-player zero-sum discounted stochastic game with finite state space has a value and the players have stationary optimal strategies. His result was generalized by Fink (1964) to n -player games with countably many states, and by Rieder (1979) to games with countably many players.

A standard way to prove these results (see, e.g., Mertens, Sorin and Zamir (1994) p. 390) is to use an auxiliary “one-shot” game that has absorbing payoffs from the second stage on. It is proved that:

1. If there exists a vector function ϕ on the state space such that for every initial state s , the one-shot game with absorbing payoff ϕ has an equilibrium that yields the players an expected payoff $\phi(s)$, then ϕ is an equilibrium payoff in the discounted game.
2. There exists a function ϕ that satisfies the condition in 1.

Proving the first step is fairly easy. In order to prove the second step when the state space is finite, one works on the strategy space to get a fixed point. This approach extends to the countable case, but not to the uncountable case.

Then one looks for a fixed point in the space of measurable payoff functions defined on the state space. Since the correspondence that assigns for each absorbing function the set of all equilibrium payoff functions in the one-shot game is not convex valued, standard fixed point theorems cannot be applied. One can overcome this problem by introducing a correlation device (Nowak and Raghavan (1992)) or by using Lyapunov Theorem if the transition law is absolutely continuous w.r.t. some fixed probability distribution (Mertens and Parthasarathy (1991)).

Duffie et al. (1994) proved the existence of an equilibrium payoff (that is not supported by stationary strategy profile) when the transition probabilities are norm continuous and mutually absolutely continuous.

In order to prove existence of equilibrium in games when the transition law does not satisfy the absolute continuity condition, Mertens and Parthasarathy (1987) take a different approach.

Define for every $k \geq 0$ a correspondence $W_k : S \rightarrow \mathbf{R}^N$, where S is the state space and N is the set of players, as follows. $W_0(s) = [-R, R]^N$, where

R is the maximal payoff in the game (in absolute values) and $W_{k+1}(s)$ is the set of all equilibrium payoffs in a one shot game with initial state s and absorbing payoff function which is a selection of W_k .

Clearly $W_{k+1} \subseteq W_k$, nevertheless $\cap_k W_k$ may be empty. Therefore, Mertens and Parthasarathy consider a compactification W_k^* of W_k , and the correspondence $D = \cap_k W_k^*$ which has non-empty values.

It is proven that every $x \in D(s)$ is an equilibrium payoff in a one-shot game with initial state s and an absorbing payoff function which is a selection of D . Finally, it is proven that for every selection w of D one can construct an equilibrium strategy profile in the discounted game, whose corresponding equilibrium payoff is w .

Roughly speaking, the proof of Mertens and Parthasarathy can be divided into two steps:

1. If there exist a correspondence $D : S \rightarrow \mathbf{R}^N$, and for every s and $x \in D(s)$ a selection $\phi_{s,x}$ of D , such that x is an equilibrium payoff in the one-shot game with initial state s and absorbing payoff function $\phi_{s,x}$, then there exists an equilibrium in the discounted game.
2. There exists a correspondence D that satisfies the condition in 1.

In the present paper we give an alternative proof to the result of Mertens and Parthasarathy (1987), by proving directly these two steps.

To prove the first step, we construct an equilibrium strategy profile τ as follows. Let s_1 be the initial state of the game, and let $x_1 \in D(s_1)$. At the first stage, the profile τ indicates the players to play an equilibrium of the one-shot game with absorbing payoff function ϕ_{s_1,x_1} . Let s_2 be the state at the second stage and $x_2 \stackrel{\text{def}}{=} \phi_{s_1,x_1}(s_2)$. The profile τ indicates the players to play at the second stage an equilibrium in the one-shot game with absorbing payoff function ϕ_{s_2,x_2} . Note that $x_2 \in D(s_2)$, hence τ is well defined for the second stage. We define τ inductively in a similar way for every finite history.

The discounted nature of the game assures us that the expected payoff for the players from stage m on, after state s_m if they follow τ , is indeed x_m . Since at each stage the players play an equilibrium in a one-shot game, whose absorbing payoff function in every state is equal to the expected payoff for the players if the game moves to that state, it will follow that τ is a subgame perfect equilibrium.

The main difference between Mertens and Parthasarathy's proof and our's lies in the proof of the second step. Instead of working with compactifications of correspondences, we work with ϵ -equilibria of the discounted game, whose existence is proven rather easily.

The basic idea is to define for every state s the set $D(s)$ as the set of all payoff vectors x that are a limit of ϵ -equilibrium payoffs for this state in the discounted game (i.e. there exists a sequence $\{\tau_n\}$ of ϵ_n -equilibrium profiles that yields the players a payoff x_n whenever the initial state is s , and satisfy $\epsilon_n \rightarrow 0$ and $x_n \rightarrow x$). Moreover, we choose for every pair (s, x) such that $x \in D(s)$ one accumulation point $\sigma_{s,x}$ of the sequence of mixed-action combination that the players play at the first stage according to such a sequence $\{\tau_n\}$. Finally, we prove that for every state s and $x \in D(s)$ there exists a selection $\phi_{s,x}$ of D such that $\sigma_{s,x}$ is an equilibrium payoff in the one-shot game with absorbing payoff function $\phi_{s,x}$, whenever the initial state is s , which yields the players an expected payoff x .

Using ϵ -equilibrium strategy profiles in the construction of the selections $\{\phi_{s,x}\}$ turns out to be insufficient. Since an ϵ -equilibrium strategy profile need not be an ϵ -equilibrium at the second stage, it might be impossible to find the selections $\{\phi_{s,x}\}$.

To overcome this difficulty, we use (m, ϵ) -equilibrium strategy profiles. A strategy profile τ is an (m, ϵ) -equilibrium if for every finite history h of length at most m , the induced strategy profile τ_h is an ϵ -equilibrium (where the induced strategy profile τ_h is defined by $\tau_h(h') = \tau(h, h')$). Thus, if $\{\tau(n)\}$ is a sequence of (m_n, ϵ_n) -equilibrium strategy profiles with $m_n \rightarrow \infty$ and $\epsilon_n \rightarrow 0$, then for every finite history h the sequence of induced strategy profiles $\{\tau_h(n)\}$ is a sequence of (m'_n, ϵ_n) -equilibrium strategy profiles with $m'_n \rightarrow \infty$.

The setup we work with is rather general, except for the transition law. We assume that the state and action spaces are measurable spaces, the payoff function is measurable, bounded and continuous over the actions for every fixed state, and the transition law is measurable and norm-continuous over the actions for every fixed state. The norm-continuity of the transition insures us that the one-shot game has an equilibrium.

The main tool that we use is selection theorems for measurable correspondences, and not fixed points theorems as in the finite case.

For measurability reasons we define D using only countably many (m, ϵ) -equilibrium strategy profiles, and not the set of all (m, ϵ) -equilibrium strategy

profiles.

The paper is arranged as follows. In section 2 we present the model of discounted stochastic games. In section 3 we define one-shot games, (m, ϵ) -equilibria and we give some preliminary results. In section 4 we give the results in measure theory that we use in the main proof, and in section 5 we prove that every n -player discounted stochastic game has a subgame perfect equilibrium.

2 The Model and the Main Result

A discounted stochastic game is given by $(N, S, A^*, (A^j, r^j)_{j \in N}, q, \beta)$ where

1. N is a set of players.
2. S is a state space.
3. A^* is the space of all actions that are available for the players.
4. For each player $j \in N$ $A^j : S \rightarrow A^*$ is a correspondence that assigns to each state the available actions of player j whenever the game is in that state. Denote $A(s) = \times_{j \in N} A^j(s)$ for every $s \in S$.
5. For each player $j \in N$, $r^j : \text{Gr}(A) \rightarrow \mathbf{R}$ is the stage payoff function, where $\text{Gr}(A)$ is the graph of A .
6. $q : \text{Gr}(A) \rightarrow \Delta(S)$ is the transition law, where $\Delta(S)$ is the space of all probability measures over S .
7. $0 < \beta < 1$ is the discount factor.

The game is played as follows. At the first stage the game is in an initial state $s_1 \in S$. At stage m the players are informed of the past history $(s_1, a_1, s_2, \dots, a_{m-1}, s_m)$, where s_k is the state of the game at stage k and a_k is the action combination the players played at that stage. Every player j chooses, independently of the others, an action $a_m^j \in A^j(s_m)$, receives a stage payoff $r^j(s_m, a_m)$, where $a_m = (a_m^j)_{j \in N}$, and the game moves to a new state s_{m+1} according to the probability measure $q(\cdot | s_m, a_m)$.

From now on we make the following assumptions:

A.1 N is a finite set.

A.2 (S, \mathcal{S}) is a measurable space.

A.3 (A^*, \mathcal{A}) is a Borel measurable separable metric space.

A.4 For every $j \in N$, $A^j : S \rightarrow A^*$ is a non-empty compact valued measurable correspondence.

A.5 For every $j \in N$, $r^j : \text{Gr}(A) \rightarrow \mathbf{R}$ is a measurable function, bounded by some $R \in \mathbf{R}_+$, and such that for every fixed $s \in S$, $r^j(s, \cdot)$ is continuous over $A(s)$.

A.6 For every $C \in \mathcal{S}$, the function $q(C|s, a)$ is measurable over $\text{Gr}(A)$, and norm continuous in a for every $s \in S$, i.e. for every $(s, a) \in \text{Gr}(A)$ and $\delta > 0$ there exists $\epsilon > 0$ such that whenever $d(a, a') < \delta$ we have $\|q(\cdot|s, a) - q(\cdot|s, a')\|_\infty < \epsilon$.

Let H^m be the space of all finite histories of length m , i.e. the set of all histories h of the form $h = (s_1, a_1, s_2, \dots, a_{m-1}, s_m)$ where $s_k \in S$ for every $1 \leq k \leq m$ and $a_k \in A(s_k)$ for every $1 \leq k \leq m-1$. H^m is a measurable space, with the product σ -algebra. Let $H = \cup_{m \in \mathbf{N}} H^m$ be the space of all finite histories. H is a measurable space, with the union σ -algebra. For every finite history $h = (s_1, a_1, s_2, \dots, a_{m-1}, s_m)$ denote by $L(h) = m$ its length and by $s_L(h) = s_m$ its last stage.

DEFINITION 2.1 *A strategy for player j is a measurable function $\tau^j : H \rightarrow \Delta(A^*)$ such that $\tau^j(h) \in \Delta(A^j(s_L(h)))$ for every $h \in H$. A strategy τ is called markovian if $\tau^j(h)$ depends only on $s_L(h)$ and $L(h)$ for every $h \in H$.*

A vector $\tau = (\tau^j)_{j \in N}$ of strategies is called a strategy profile. It is called markovian if each τ^j is markovian.

Every initial state s and strategy profile τ induce a probability measure over H . We denote expectation according to this probability measure by $\mathbf{E}_{s, \tau}$. The expected discounted payoff of player j if the initial state is s and the players follow the strategy profile τ is defined by:

$$v_\tau^j(s) = \sum_{m=1}^{\infty} \beta^{m-1} \mathbf{E}_{s, \tau} r^j(s_m, a_m).$$

Note that since τ is measurable, $\mathbf{E}_{s,\tau} r^j(s_m, a_m)$ is measurable over S for every m , and therefore v_τ^j is measurable over S .

DEFINITION 2.2 *A strategy profile τ is an ϵ -equilibrium if for every player j , every strategy τ'^j of player j and every initial state s ,*

$$v_\tau^j(s) \geq v_{\tau^{-j}, \tau'^j}^j(s) - \epsilon$$

where $\tau^{-j} = (\tau^k)_{k \neq j}$. An equilibrium is a 0-equilibrium.

For every strategy profile τ and finite history $h = (s_1, a_1, \dots, a_{m-1}, s_m) \in H$, the induced strategy profile τ_h is given by:

$$\tau_h(h') = \tau(s_1, a_1, \dots, a_{m-1}, s'_1, a'_1, \dots, a'_{k-1}, s'_k)$$

for every $h' = (s'_1, a'_1, \dots, a'_{k-1}, s'_k) \in H$.

DEFINITION 2.3 *A strategy profile τ is called a subgame perfect equilibrium if τ_h is an equilibrium strategy profile for every finite history h .*

We now state the main result of the paper:

THEOREM 2.4 *Every discounted stochastic game that satisfies assumptions **A.1** - **A.6** has a measurable subgame perfect equilibrium.*

3 Preliminaries

3.1 The One-Shot Game

For every measurable function $\phi : S \rightarrow \mathbf{R}^N$ and state $s \in S$ define the one-shot game $\Gamma_s(\phi)$ with initial state s and absorbing payoff function ϕ as follows:

- Every player j chooses an action $a^j \in A^j(s)$.
- A new state $t \in S$ is chosen according to $q(\cdot | s, a)$, where $a = (a^j)_{j \in N}$.
- Every player j receives the payoff $r^j(s, a) + \beta \phi^j(t)$.

Recall that every function from $\text{Gr}(A)$ to \mathbf{R}^N induces a game, and that $\Gamma_s(\phi)$ is strategically equivalent to the game with payoff function

$$r(s, a) + \beta \int_S \phi(t) dq(t|s, a).$$

We denote $\Gamma(\phi) = (\Gamma_s(\phi))_{s \in S}$.

DEFINITION 3.1 *For each player j , let $\sigma^j : S \rightarrow \Delta(A^*)$ be a measurable function such that $\sigma^j(s) \in \Delta(A^j(s))$ for every $s \in S$. The profile $\sigma = (\sigma^j)_{j \in N}$ is an equilibrium in $\Gamma(\phi)$ if for every $s \in S$, $\sigma(s)$ is an equilibrium in $\Gamma_s(\phi)$.*

By Theorem 3.1 in Rieder (1979) it follows that

THEOREM 3.2 *If the game satisfies assumptions **A.1** - **A.6** then for every measurable and bounded function $\phi : S \rightarrow \mathbf{R}^N$ there exists a measurable equilibrium in $\Gamma(\phi)$.*

3.2 On (m, ϵ) -Equilibria

DEFINITION 3.3 *Let $m \in \mathbf{N}$ and $\epsilon > 0$. A strategy profile τ is called an (m, ϵ) -equilibrium if for every history h of length at most m , the induced strategy profile τ_h is an ϵ -equilibrium.*

The following lemma follows from the definition:

LEMMA 3.4 *If τ is an (m, ϵ) -equilibrium, then for every $h \in H$ such that $L(h) < m$ and every $\epsilon' \geq \epsilon$, τ_h is an $(m - L(h), \epsilon')$ -equilibrium.*

LEMMA 3.5 *For every $m \in \mathbf{N}$ and $\epsilon > 0$ there exists a measurable markovian (m, ϵ) -equilibrium strategy profile in the game.*

Proof: Let $M \in \mathbf{N}$ be big enough such that $\frac{\beta^M}{1-\beta} R < \epsilon$. Define inductively the functions $\{\phi_i : S \rightarrow \mathbf{R}^N\}_{i=1}^{M+m+1}$ and $\{\sigma_i : S \rightarrow \Delta(A^*)\}_{i=1}^{M+m}$ as follows:

$$\phi_{M+m+1}(s) = 0.$$

For every $1 \leq i \leq M + m$ let σ_i be an equilibrium in $\Gamma(\phi_{i+1})$, and let ϕ_i be the corresponding equilibrium payoff, i.e.

$$\phi_i(s) = \int_{A(s)} \left(r(s, a) + \beta \int_S \phi_{i+1}(t) dq(t|s, a) \right) d\sigma_i(s)(a).$$

Define the strategy profile τ as follows:

- At each stage $1 \leq i \leq M + m$ the players play the mixed action combination that is indicated by σ_i .
- At each stage $i > M + m$, the players play randomly.

Let $x_l^j = r^j(s_l, a_l)$ be the payoff that player j receives at stage l . Let h be a history of length $k \leq m$, and consider the strategy profile τ_h . By the definition of $\{\sigma_i\}_{i=1}^{M+m}$, for every player j and every strategy τ^j of player j

$$\begin{aligned} \mathbf{E}_{s, \tau_h^{-j}, \tau^j} \left(\sum_{l \geq 1} \beta^{l-1} x_l^j \right) &\leq \mathbf{E}_{s, \tau_h^{-j}, \tau^j} \left(\sum_{1 \leq l \leq M+m-k} \beta^{l-1} x_l^j \right) + \frac{\beta^M}{1-\beta} R \\ &\leq \mathbf{E}_{s, \tau_h} \left(\sum_{1 \leq l \leq M+m-k} \beta^{l-1} x_l^j \right) + \epsilon \\ &\leq \mathbf{E}_{s, \tau_h} \left(\sum_{l \geq 1} \beta^{l-1} x_l^j \right) + 2\epsilon \end{aligned}$$

where the first and last inequalities follow by the choice of M . Hence player j cannot profit more than 2ϵ by deviating from τ_h^j , and the result follows. ■

3.3 On the Norm Continuity of the Transition

LEMMA 3.6 *For every fixed $s \in S$, the set $\{r(s, a) + \int f(t) dq(t|s, a)\}$ of functions from $A(s)$ to $[-2R, 2R]^N$, where f ranges over all measurable functions from S to $[-R, R]^N$, is equicontinuous, and its closure is compact.*

Proof: Fix $s \in S$, and let $\epsilon > 0$. $r(s, \cdot)$, as a continuous function over the compact space $A(s)$, is uniform continuous. Similarly, $q(\cdot|s, \cdot)$ is norm-uniform-continuous over $A(s)$. Therefore, there exists $\delta > 0$ such that if $a, a' \in A(s)$ and $d(a, a') < \delta$ then

$$\| r(s, a) - r(s, a') \|_\infty < \epsilon$$

and

$$\| q(\cdot|s, a) - q(\cdot|s, a') \|_\infty < \epsilon$$

Let $f : S \rightarrow [-R, R]^N$ be measurable. Then for every $a, a' \in A(s)$ such that $d(a, a') < \delta$ we have:

$$\| r(s, a) + \int_S f(t) dq(t|s, a) - r(s, a') - \int_S f(t) dq(t|s, a') \| \leq \epsilon + R\epsilon$$

and the first result follows.

By the Arzela-Ascoli Theorem the closure of this set is compact. ■

The norm continuity of the transition is used also in the proof of theorem 3.2.

4 On Measurability

In this section we give the results in measure theory that we use in the proof of the main result. Along the section, unless otherwise stated, (X, \mathcal{X}) is an arbitrary measurable space, and (Y, \mathcal{Y}) and (Z, \mathcal{Z}) are Borel measurable separable metric spaces. Whenever we consider a product of two measurable spaces, the σ -algebra is the product σ -algebra.

For every set $C \subset Y$, \overline{C} is the closure of C . We denote $B(y, \epsilon) = \{y' \in Y \mid d(y, y') \leq \epsilon\}$.

DEFINITION 4.1 *A correspondence $F : X \rightarrow Y$ is measurable if $F^{-1}(C) \in \mathcal{X}$ for every closed set $C \subset Y$, where*

$$F^{-1}(C) = \{x \in X \mid F(x) \cap C \neq \emptyset\}.$$

PROPOSITION 4.2 *Let $F : X \rightarrow Y$ be a measurable correspondence with compact values. Then*

- a) *The graph of F is a measurable set in $X \times Y$.*
- b) *If F has non-empty values then it has a measurable selection, i.e. there exists a measurable function $f : X \rightarrow Y$ such that $f(x) \in F(x)$ for every $x \in X$.*

Proof: (a) is proved in Theorem 3.5 in Himmelberg (1975), and (b) is proved in Corollary 1 in Kuratowski and Ryll-Nardzewski (1965). ■

PROPOSITION 4.3 *Let $F : X \rightarrow Y$ be a measurable correspondence with non-empty compact values, and $g : X \rightarrow Y$ be a measurable function. There exists a measurable selection f of F such that for every $x \in X$*

$$d(g(x), f(x)) = d(g(x), F(x)).$$

Proof: The function $\rho : X \rightarrow Y$ that is defined by

$$\rho(x) = d(g(x), F(x))$$

is measurable by Theorem 3.3 in Himmelberg (1975). Hence the correspondence $G : X \rightarrow Y$ that is defined by

$$G(x) = B(g(x), \rho(x))$$

is measurable. The correspondence $F \cap G$ has non-empty and compact values, and by Theorem 4.1 in Himmelberg (1975) it is measurable. By Proposition 4.2.b it has a measurable selection f . Clearly f satisfies the conclusion of the lemma. ■

LEMMA 4.4 *Let $G : X \rightarrow Y \times Z$ be a measurable correspondence with compact values. Let $\hat{G} : X \times Y \rightarrow Z$ be defined by:*

$$\hat{G}(x, y) = \{z \mid (y, z) \in G(x)\}.$$

Then \hat{G} is measurable.

Proof: Define the correspondences $G_1, G_2 : X \times Y \rightarrow Y \times Z$ by:

$$G_1(x, y) = G(x)$$

and

$$G_2(x, y) = \{y\} \times Z.$$

Clearly G_1 and G_2 are measurable, and G_1 has compact values. By Theorem 4.1 in Himmelberg (1975) the correspondence $G_3 = G_1 \cap G_2$ is measurable. However, \hat{G} is the projection of G_3 over the second coordinate, and by Proposition 2.5 in Himmelberg (1975) it is measurable. ■

We shall need the following result:

THEOREM 4.5 For every $m \in \mathbf{N}$ and $1 \leq k < m$, let $g_m^k : X \rightarrow Y \times Z$ be a measurable function such that for every $x \in X$, $\{g_m^k(x)\}_{m,k}$ is included in some compact subset of $Y \times Z$. Let $f_m^k : X \rightarrow Y$ be the first coordinate of g_m^k . Define the correspondence $F : X \rightarrow Y$ by:

$$F(x) = \left\{ y \in Y \mid \exists (m_i, k_i)_{i=1}^{\infty} \text{ s.t. } m_i \rightarrow \infty, k_i/m_i \rightarrow 0, y = \lim f_{m_i}^{k_i}(x) \right\}.$$

Define the correspondence $G : \text{Gr}(F) \rightarrow Z$ by:

$$G(x, y) = \left\{ z \in Z \mid \exists (m_i, k_i)_{i=1}^{\infty} \text{ s.t. } m_i \rightarrow \infty, k_i/m_i \rightarrow 0, (y, z) = \lim g_{m_i}^{k_i}(x) \right\}.$$

Then F and G are measurable, and have non-empty and compact values.

Proof: Since $\{g_m^k(x)\}_{m,k}$ is included in some compact subset of $Y \times Z$ for every $x \in X$, it follows that F and G have non-empty and compact values.

Let us prove that F is measurable. For every $q \in \mathbf{Q}_{++} = \{q \in \mathbf{Q} \mid q > 0\}$ and $m \in \mathbf{N}$ define the correspondence $F_{m,q} : X \rightarrow Y$ by:

$$F_{m,q}(x) = \left\{ f_{m'}^k(x) \mid m' > m \text{ and } 1 \leq k \leq qm' \right\}.$$

Clearly the correspondence $\bar{F}_{m,q}$, which is defined by $\bar{F}_{m,q}(x) = \overline{F_{m,q}(x)}$ has compact values, and by Theorem 5.6 in Himmelberg (1975) it is measurable. Since $F(x) = \bigcap_{q \in \mathbf{Q}_{++}} \bigcap_{m \in \mathbf{N}} \bar{F}_{m,q}(x)$, it follows by Theorem 4.1 in Himmelberg (1975) that F is measurable.

By substituting f_m^k with g_m^k in the definition of F we get that the correspondence

$$x \rightarrow \left\{ (y, z) \in Y \times Z \mid \exists (m_i, k_i)_{i=1}^{\infty} \text{ s.t. } m_i \rightarrow \infty, k_i/m_i \rightarrow 0, (y, z) = \lim g_{m_i}^{k_i}(x) \right\}$$

is measurable. By Lemma 4.4 it follows that G is measurable. ■

Mertens (1987) proved the following measurable ‘‘Measurable Choice’’ Theorem:

THEOREM 4.6 Let (X, \mathcal{X}) and (Y, \mathcal{Y}) be measurable spaces. Let $q(\cdot|y)$ be a probability measure over (X, \mathcal{X}) for every $y \in Y$, such that for every $C \in$

\mathcal{X} , the function $q(C|\cdot)$ is \mathcal{Y} -measurable. Let $D : X \times Y \rightarrow [-R, R]^n$ be a measurable correspondence with non-empty compact values (where $R \in \mathbf{R}_+$ and $n \in \mathbf{N}$). Define the correspondence $\Omega : Y \rightarrow \mathbf{R}^n$ by

$$\Omega(y) = \left\{ \int_X f(x, y) dq(x|y) \mid f \text{ is an } \mathcal{X} \times \mathcal{Y}\text{-measurable selection of } D \right\}.$$

Then

- a) For every $y \in Y$, $\Omega(y)$ is a non-empty compact subset of \mathbf{R}^n .
- b) The correspondence Ω is measurable.
- c) There exists a measurable function $\phi : \text{Gr}(\Omega) \times X \rightarrow \mathbf{R}^n$ such that for every $(y, p) \in \text{Gr}(\Omega)$

$$p = \int_X \phi(y, p, x) dq(x|y)$$

and for every $x \in X$

$$\phi(y, p, x) \in D(x, y).$$

5 Existence of a Subgame Perfect Equilibrium

In this section we prove two lemmas, that correspond to the two steps which were described in the introduction.

For every correspondence $D : S \rightarrow \mathbf{R}^N$ we denote $\mathcal{W}_D = \{(s, a, x, t) \in S \times A^* \times \mathbf{R}^N \times S \mid a \in A(s), x \in D(s)\}$. \mathcal{W}_D is the set of all tuples that consists of a current state, an available action combination, a ‘‘possible’’ payoff vector (according to D) and a possible state at the next stage.

LEMMA 5.1 *If there exists a measurable correspondence $D : S \rightarrow \mathbf{R}^N$ with non-empty compact values and two measurable functions $\psi_\star : \mathcal{W}_D \rightarrow \mathbf{R}^N$ and $\sigma_\star : \text{Gr}(D) \rightarrow (\Delta(A^*))^N$ that satisfy the following:*

- 1) $\psi_\star(s, a, x, t) \in D(t)$ for every $(s, a, x, t) \in \mathcal{W}_D$.

- 2) $\sigma_\star^j(s, x) \in \Delta(A^j(s))$ for each player j , state s and $x \in D(s)$.
- 3) For every $s \in S$ and $x \in D(s)$, $\sigma_\star(s, x)$ is an equilibrium in the game with payoff function $r(s, a) + \beta \int_S \psi_\star(s, a, x, t) dq(t|s, a)$ that yields the players an expected payoff x .

Then there exists a subgame perfect equilibrium in the discounted game.

Proof: Let $f : S \rightarrow \mathbf{R}^N$ be a measurable selection of D . By Proposition 4.2.b such a selection exists. We prove that there exists a subgame perfect equilibrium which yields the players an expected discounted payoff f .

Step 1: Constructing a parameter function π

Define a function $\pi : H \rightarrow \mathbf{R}^N$ inductively as follows:

$$\begin{aligned} \pi(s) &= f(s) & \forall s \in S \\ \pi(h, a, t) &= \psi_\star(s_L(h), a, \pi(h), t) & \forall h \in H, a \in A(s_L(h)), t \in S \end{aligned} \quad (1)$$

Note that $\pi(h) \in D(s_L(h))$ for every $h \in H$, and by induction on $L(h)$, π is well defined and measurable.

Step 2: Constructing a strategy profile τ

Define the strategy profile τ by:

$$\tau(h) = \sigma_\star(s_L(h), \pi(h)).$$

Note that τ , as a composition of measurable functions, is measurable.

Step 3: A connection between τ and π

We shall now prove that $\pi(h) = v_{\tau_h}(s_L(h))$ for every $h \in H$, i.e. for every $h \in H$, if the initial state is $s_L(h)$ and the players follow the strategy profile τ_h , then their expected payoff is $\pi(h)$. Equivalently, given the history $h \in H$ has occurred, the expected payoff for the players from stage $L(h)$ is $\pi(h)$.

Assume to the contrary that

$$c \stackrel{\text{def}}{=} \sup_{h \in H} |\pi(h) - v_{\tau_h}(s_L(h))| > 0. \quad (2)$$

Let $h = (s_1, a_1, \dots, a_{m-1}, s_m) \in H$ be such that $|\pi(h) - v_{\tau_h}(s_m)| > \beta c$. By the definition of τ

$$v_{\tau_h}(s_m) = \int_{A(s_m)} \left(r(s_m, a) + \beta \int_S v_{\tau_{h,a,t}}(t) dq(t|s_m, a) \right) d\sigma_\star(s_m, \pi(h)). \quad (3)$$

Since $\pi(h) \in D(s_m)$, it follows by assumption 3 that

$$\pi(h) = \int_{A(s_m)} \left(r(s_m, a) + \beta \int_S \psi_\star(s_m, a, \pi(h), t) dq(t|s_m, a) \right) d\sigma_\star(s_m, \pi(h)).$$

By (1), $\psi_\star(s_m, a, \pi(h), t) = \pi(h, a, t)$ and therefore

$$\pi(h) = \int_{A(s_m)} \left(r(s_m, a) + \beta \int_S \pi(h, a, t) dq(t|s_m, a) \right) d\sigma_\star(s_m, \pi(h)). \quad (4)$$

Subtracting (3) from (4), using (2) yields $|\pi(h) - v_{\tau_h}(s_m)| \leq \beta c$ which contradicts the choice of h .

Step 4: τ is a subgame perfect equilibrium

Assume that an arbitrary finite history h has occurred. By (1) and step 3, for every $a \in A(s_L(h))$ we have $\int_S \psi_\star(s_L(h), a, \pi(h), t) dq(t|s_L(h), a) = \int_S \pi(h, a, t) dq(t|s_L(h), a) = \int_S v_{\tau_{h,a,t}}(t) dq(t|s_L(h), a)$. By assumption 3 it follows that $\sigma_\star(s_L(h), \pi(h))$ is an equilibrium strategy profile in the game with payoff function $r(s, a) + \beta \int_S v_{\tau_{h,a,t}}(t) dq(t|s, a)$ and therefore no player can profit by a deviation at a single stage. This fact implies that no player can profit by any deviation. \blacksquare

LEMMA 5.2 *There exist a measurable correspondence $D : S \rightarrow \mathbf{R}^N$ with non-empty and compact values, and two measurable functions $\psi_\star : \mathcal{W}_D \rightarrow \mathbf{R}^N$ and $\sigma_\star : \text{Gr}(D) \rightarrow (\Delta(A^\star))^N$ that satisfy the conditions of Lemma 5.1*

Proof: For every $m \in \mathbf{N}$, let $\tau_m = (\sigma_m^1, \sigma_m^2, \dots, \sigma_m^k, \dots)$ be a markovian $(m, \frac{1}{m})$ -equilibrium strategy profile of the discounted stochastic game. By Lemma 3.5 such strategy profiles exist. For every $1 \leq k < m$, let $\tau_m^k = (\sigma_m^k, \sigma_m^{k+1}, \dots)$ be the strategy profile τ_m truncated from stage k on, and let $w_m^k(s) = v_{\tau_m^k}(s)$ be the expected payoff for the players if τ_m^k is used and the initial state is s . Since τ_m is a markovian strategy profile, τ_m^k is well defined. By Lemma 3.4, τ_m^k is an $(m - k, \frac{1}{m-k})$ -equilibrium.

Define for every $m \in \mathbf{N}$ and $1 \leq k < m$ the function $u_m^k : \text{Gr}(A) \rightarrow \mathbf{R}^N$ by

$$u_m^k(s, a) = r(s, a) + \beta \int_S w_m^{k+1}(t) dq(t|s, a). \quad (5)$$

$u_m^k(s, a)$ is the expected payoff for the players in the discounted game if the initial state is s , the players play the action combination a at the first stage, and then follow the strategy profile τ_m^{k+1} .

Clearly $\sigma_m^k(s)$ is an $\frac{1}{m}$ -equilibrium in the game with payoff function $u_m^k(s, a)$, and $w_m^k(s)$ is the corresponding $\frac{1}{m}$ -equilibrium payoff.

Step 1: The correspondence D

For every $s \in S$ let

$$D(s) = \left\{ x \in \mathbf{R}^N \mid \exists (m_i, k_i)_{i=1}^{\infty} \text{ s.t. } m_i \rightarrow \infty, k_i/m_i \rightarrow 0, w_{m_i}^{k_i}(s) \rightarrow x \right\}.$$

Step 2: The functions σ_* and u_*

For every $m \in \mathbf{N}$ and $1 \leq k < m$ define the measurable function g_m^k on S by:

$$g_m^k(s) = (w_m^k(s), u_m^k(s, \cdot), \sigma_m^k(s)).$$

Let G be a correspondence on $S \times \mathbf{R}^N$ defined by:

$$G(s, x) = \left\{ (\sigma, u) \mid \exists (m_i, k_i)_{i=1}^{\infty} \text{ s.t. } m_i \rightarrow \infty, k_i/m_i \rightarrow 0, g_{m_i}^{k_i}(s) \rightarrow (x, u, \sigma) \right\}.$$

By Lemma 3.6 the closure of the set $\{u_m^k(s, \cdot)\}_{m,k}$ is compact in the supremum topology. By Alaoglu Theorem (see Dunford and Schwartz (1957) p. 424) $\Delta(A(s))$ is compact in the w^* -topology. It follows by Theorem 4.5 that D and G are measurable correspondences, and have non-empty and compact values.

By Proposition 4.2.b there exists a measurable selection (σ_*, u_*) of G .

Step 3: The function ψ_*

Consider the correspondence:

$$\Omega : (s, a) \mapsto \left\{ r(s, a) + \beta \int f(t) dq(t|s, a) \right\}$$

where f ranges over all measurable selections of $D(t)$.

By Theorem 4.6, Ω is measurable, has non-empty and compact values, and there exists a measurable function ψ on $\text{Gr}(\Omega) \times S$ such that for every $(s, a, p) \in \text{Gr}(\Omega)$

$$p = r(s, a) + \beta \int_S \psi(s, a, p, t) dq(t|s, a) \tag{6}$$

and for every $t \in S$

$$\psi(s, a, p, t) \in D(t).$$

Define the function ψ_\star from \mathcal{W}_D to \mathbf{R}^N by:

$$\psi_\star(s, a, x, t) = \psi(s, a, u_\star(s, x, a), t).$$

Note that if ψ_\star is well defined (or equivalently, if $u_\star(s, x, a) \in \Omega(s, a)$ for every $s \in S$, $a \in A(s)$ and $x \in D(s)$) then $\psi_\star(s, a, x, t) \in D(t)$, and by (6)

$$u_\star(s, x, a) = r(s, a) + \beta \int \psi_\star(s, a, x, t) dq(t|s, a). \quad (7)$$

Step 4: ψ_\star is well defined

Fix a state s , an action combination $a \in A(s)$, $x \in D(s)$ and $\epsilon > 0$. We show that there exists a measurable selection f_ϵ of D such that

$$\left| u_\star(s, x, a) - r(s, a) - \beta \int f_\epsilon(t) dq(t|s, a) \right| < (2R + 2)\epsilon$$

and conclude that $u_\star(s, x, a) \in \Omega(s, a)$ by the compactness of Ω .

Indeed, let $(m_i, k_i)_{i=1}^\infty$ satisfy that $m_i \rightarrow \infty$, $k_i/m_i \rightarrow 0$ and $u_{m_i}^{k_i}(s, \cdot) \rightarrow u_\star(s, x, \cdot)$. Since $\frac{k_i+1}{m_i} \rightarrow 0$, it follows that for every $t \in S$, $d(w_{m_i}^{k_i+1}(t), D(t)) \rightarrow 0$. Hence, there exists i_0 sufficiently large such that the set

$$C \stackrel{\text{def}}{=} \left\{ t \in S \mid d(w_{m_i}^{k_i+1}(t), D(t)) \leq \epsilon \quad \forall i \geq i_0 \right\}$$

satisfies $q(C|s, a) > 1 - \epsilon$. Assume i_0 is sufficiently large to satisfy also

$$\| u_\star(s, x, \cdot) - u_{m_{i_0}}^{k_{i_0}}(s, \cdot) \|_\infty < \epsilon.$$

By Proposition 4.3 there exists a measurable selection f_ϵ of D such that for every $t \in S$

$$d(w_{m_{i_0}}^{k_{i_0}+1}(t), D(t)) = d(w_{m_{i_0}}^{k_{i_0}+1}(t), f_\epsilon(t)).$$

Hence, by (5),

$$\begin{aligned}
& \left| u_*(s, x, a) - r(s, a) - \beta \int f_\epsilon(t) dq(t|s, a) \right| \leq \\
& \leq \left| u_*(s, x, a) - u_{m_{i_0}}^{k_{i_0}}(s, a) \right| + \beta \int_C |f_\epsilon(t) - w_{m_{i_0}}^{k_{i_0}+1}(t)| dq(t|s, a) \\
& \quad + \beta \int_{C^c} |f_\epsilon(t) - w_{m_{i_0}}^{k_{i_0}+1}(t)| dq(t|s, a) \\
& \leq \epsilon + \beta\epsilon + 2\beta\epsilon R
\end{aligned}$$

as desired.

Step 5: $\sigma_*(s, x)$ is an equilibrium in the game with payoff function $r(s, a) + \beta \int_S \psi_*(s, a, x, t) dq(t|s, a)$

Let $(m_i, k_i)_{i=1}^\infty$ be a sequence such that $(\sigma_{m_i}^{k_i}(s), u_{m_i}^{k_i}(s, \cdot)) \rightarrow (\sigma_*(s, x), u_*(s, x, \cdot))$. For every i , $\sigma_{m_i}^{k_i}(s)$ is an $\frac{1}{m}$ -equilibrium in the game with payoff function $u_{m_i}^{k_i}(s, \cdot)$. Since $u_{m_i}^{k_i}(s, \cdot) \rightarrow u_*(s, x, \cdot)$ uniformly, and $\sigma_{m_i}^{k_i}(s) \rightarrow \sigma_*(s, x)$ in the w^* -topology, it follows that $\sigma_*(s, x)$ is an equilibrium in the game with payoff function $u_*(s, x, \cdot)$, and by (7) the result follows. ■

Clearly Lemmas 5.1 and 5.2 prove Theorem 2.4.

Remark: The proof can be generalized for the case that each player has a different discount factor, as well as for the case that the discount factor, the transition law and the payoff functions depend on the whole history. All that is needed is the existence of (m, ϵ) -equilibrium strategy profiles. For further generalizations one can refer to Mertens and Parthasarathy (1987).

References

- [1] Duffie D., J. Geanakoplos, A. Mas-Colell, A. McLennan (1994) Stationary Markov Equilibria, *Econometrica*, **62**, 745-781
- [2] Dunford N., J.T. Schwartz (1957) *Linear Operators Part I*, Interscience Publishers, New York
- [3] Fink A.M. (1964) Equilibrium in a Stochastic n -Person Game, *J. Sci. Hiroshima Univ.*, **28**, 89-93
- [4] Himmelberg C.J. (1975) Measurable Relations, *Fund. Math.*, **87**, 53-72
- [5] Kuratowski K., C. Ryll-Nardzewski (1965) A General Theorem on Selectors, *Bull. Acad. Polon. Sci.*, **13**, 397-403
- [6] Mertens J.F. (1987) A Measurable “Measurable Choice” Theorem, CORE Discussion Paper No. 8749
- [7] Mertens J.F., T. Parthasarathy (1987) Equilibria for Discounted Stochastic Games, CORE Discussion Paper No. 8750
- [8] Mertens J.F., T. Parthasarathy (1991) Non Zero Sum Stochastic Games, In *Stochastic Games and Related Topics*, Raghavan T.E.S. et al. (eds.), Kluwer Academic Publishers, 145-148
- [9] Mertens J.F., S. Sorin, S. Zamir (1994) Repeated Games, CORE Discussion Paper 9421
- [10] Nowak A.S., T.E.S. Raghavan (1992) Existence of Stationary Correlated Equilibria with Symmetric Information for Discounted Stochastic Games, *Math. Oper. Res.*, **17**, 519-526
- [11] Rieder U. (1979) Equilibrium Plans for Non-Zero Sum Markov Games, *Game Theory and Related Topics*, Moeshlin O. and Pallaschke D. (eds.) North-Holland Publishing Company, 91-101
- [12] Shapley L.S. (1953) Stochastic Games, *Proc. Nat. Acad. Sci. U.S.A.*, **39**, 1095-1100