DISCOUNTED STOCHASTIC GAMES: THE FINITE CASE

SYLVAIN SORIN
Université P. et M. Curie and École Polytechnique
Paris, France

Abstract. Recall that in the λ-discounted game $\Gamma_\lambda(z)$ with initial state $z_1 = z$ the payoff given a profile of strategies σ, $\gamma_\lambda^z(\sigma)$, is equal to the expectation, with respect to the distribution induced on plays by z and σ, of the discounted sum of the sequence of stage rewards $\{r_m\}$:

$$\gamma_\lambda^z(\sigma) = E_\sigma^z(\sum_{m=1}^{\infty} \lambda(1-\lambda)^{m-1} r_m).$$

This chapter considers the finite case where the state space S and each action space A^i, i in I, are finite.

1. Zero-Sum Case

1.1. THE AUXILIARY GAME AND THE SHAPLEY OPERATOR

As explained in [7], the basic tool is a family of one-shot games obtained by reducing the future of the game to a state-dependent payoff vector.

Given a real function f on S, $\Gamma(f)[z]$ is the two-person zero-sum game with strategy sets A and B and payoff function $L(f)(z,.,.)$ from $A \times B$ to $\mathbb{R}$ defined by $L(f)(z,a,b) = r(z,a,b) + \sum_{z'} f(z')p(z'|z,a,b)$. By von Neumann's minmax theorem this game has a value. This allows us to introduce the *Shapley operator* $\mathbf{\Psi} : f \mapsto \mathbf{\Psi}(f)$ from $\mathbb{R}^S$ to itself specified by the following relation:

$$\mathbf{\Psi}(f)[z] = \max_{x \in \Delta(A)} \min_{y \in \Delta(B)} \{\sum_{a,b} x(a)y(b)r(z,a,b) + \sum_{a,b,z'} x(a)y(b)p(z'|z,a,b)f(z')\}$$

$$= \max_{x \in \Delta(A)} \min_{y \in \Delta(B)} L(f)(z,x,y),$$

where $L(f)(z,x,y)$ is the bilinear extension of $L(f)(z,.,.)$ to $\Delta(A)\times\Delta(B)$, or, more concisely,

$$\mathbf{\Psi}(f)[z] = \mathtt{val}_{\Delta(A)\times\Delta(B)}\{r(z,\cdot)+E(f|z,\cdot)\}.$$

The main properties of $\mathbf{\Psi}$ are
- *monotonicity*: $f \le g$ implies $\mathbf{\Psi}(f) \le \mathbf{\Psi}(g)$
- *reduction of constants*: for any $t \ge 0$, $\mathbf{\Psi}(f+t) \le \mathbf{\Psi}(f)+t$.

These two properties imply that $\mathbf{\Psi}$ is *nonexpansive*:

$$\|\mathbf{\Psi}(f)-\mathbf{\Psi}(g)\|_\infty \le \|f-g\|_\infty.$$

1.2. THE CONTRACTING OPERATOR

In the framework of a discounted game the weight on the present is λ and on the future $(1-\lambda)$; hence it is natural to introduce the operator $\mathbf{\Phi}(\lambda,.)$ defined by

$$\mathbf{\Phi}(\lambda,f)[z] = \mathtt{val}_{\Delta(A)\times\Delta(B)}\{\lambda r(z,\cdot)+(1-\lambda)E(f|z,\cdot)\},$$

which corresponds to the value of a game $\mathcal{G}(\lambda,f)[z]$ played on $A\times B$ and with payoff $\lambda r(z,a,b)+(1-\lambda)\sum_{z'} f(z')p(z'|z,a,b)$. Both operators $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are related through the relation

$$\mathbf{\Phi}(\lambda,f) = \lambda\mathbf{\Psi}\Big(\frac{(1-\lambda)}{\lambda}f\Big),$$

hence in particular

$$\|\mathbf{\Phi}(\lambda,f)-\mathbf{\Phi}(\lambda,g)\|_\infty \le (1-\lambda)\|f-g\|_\infty,$$

so that $\mathbf{\Phi}(\lambda,\cdot)$ is contracting with constant $1-\lambda$. In particular, it has a unique fixed point in $\mathbb{R}^S$ denoted by w_λ.

1.3. V_λ AND STATIONARY STRATEGIES

The next result proves that $w_\lambda(z)$ is actually the value of $\Gamma_\lambda(z)$. More precisely:

Theorem 1 $\Gamma_\lambda(z)$ *has a value* $v_\lambda(z)$ *and* $v_\lambda(z)=w_\lambda(z)$*; hence it is the only solution of*

$$\mathbf{\Phi}(\lambda,v_\lambda)=v_\lambda.$$

If for all z, x_z *is an* ε*-optimal strategy in* $\mathbf{\Phi}(\lambda,w_\lambda)[z]$*, then the induced stationary strategy* $\overline{x}=\{x_z\}$ *is* (ε/λ)*-optimal in* Γ_λ.

Proof. Denoting by $\mathcal{H}_n$ the algebra on plays generated by histories h_n of length n, one has, by the definition of $\overline{x}$,

$$E_{\overline{x},\tau}\{\lambda r(z_n,a_n,b_n)+(1-\lambda)\sum_{z'}p(z'|z_n,a_n,b_n)w_\lambda(z')|\mathcal{H}_n\} \geq w_\lambda(z_n)-\varepsilon \qquad \forall\tau.$$

This can be written as

$$E_{\overline{x},\tau}\{\lambda r_n + (1-\lambda)w_\lambda(z_{n+1})|\mathcal{H}_n\} \geq w_\lambda(z_n) - \varepsilon \qquad \forall\tau.$$

Multiplying by $(1-\lambda)^{n-1}$, taking expectation and summing over $n \geq 1$, one obtains

$$\sum_{n=1}^{\infty} E_{\overline{x},\tau}(\lambda(1-\lambda)^{n-1}r_n) = \gamma_\lambda^z(\overline{x},\tau) \geq w_\lambda(z) - \varepsilon/\lambda \qquad \forall\tau.$$

Similarly, if $\overline{y}$ is constructed from a family of ε-optimal strategies $\{y_z\}$ in $\mathcal{G}(\lambda, v_\lambda)[z]$, then

$$\gamma_\lambda^z(\sigma,\overline{y}) \leq w_\lambda(z) + \varepsilon/\lambda \qquad \forall\sigma,$$

which implies that $v_\lambda(z) = w_\lambda(z)$; hence the result. ∎

1.4. EXTENSIONS

Still in the finite framework (S, A and B finite), the Shapley operator allows us also to express the value of the n-stage repeated game $\Gamma_n(z)$. In fact, by induction one easily obtains

Proposition 1 *$\Gamma_n(z)$ has a value $v_n(z)$ that satisfies*

$$nv_n = \mathbf{\Psi}^n(0)$$

$$v_n = \mathbf{\Phi}(1/n, v_{n-1}).$$

The knowledge of the current state is sufficient to play optimally in the above "auxiliary one-shot game," which implies the existence of Markov optimal strategies in Γ_n.

The same tool applies for an evaluation of the stream of rewards using a stopping time θ for which $E_{\sigma,\tau}^z(\sum_{n=1}^{\theta} r_n)$ is finite.

The previous approach extends to the case of general action and state space. The aim is to look for a complete subset F of bounded functions on S such that:

1) the game $\Gamma(f)[z]$ has a value $\mathbf{\Psi}(f)[z]$ for all z and all f in F,
2) the function $\mathbf{\Psi}(f)$ belongs to F for all f in F,
3) ε-optimal "measurable" strategies exist (thus enabling us to define a payoff for $\overline{x}$).

In the finite state space case, 2) and 3) are immediate; hence one basically needs conditions to apply a minmax theorem like: A compact, $r(z,.,b)$ uppersemicontinuous and $p(z' \mid z,.,b)$ continuous on A.

For an uncountable state space this program is developed in [4].

2. Non-Zero-Sum Case

In the non-zero-sum case a similar approach through an auxiliary game can be used to study "subgame-perfect" equilibria. In the discounted case it will allow for a characterization of all stationary equilibria. The procedure is parallel to the previous one. One first constructs an operator and exhibits a fixed point. One then shows that it leads to an equilibrium. However, there is no monotonicity property here and we rely on Kakutani's fixed-point theorem on the strategy space, rather than on Picard's contraction principle on the payoff space.

Given f from S to $\mathbb{R}^I$, one introduces, for each z in S, an auxiliary one-shot game $\mathcal{G}(\lambda, f)[z]$ with strategy sets A^i and payoff $\lambda r(z,.) + (1-\lambda)E(f|z,.)$. Define X as $\prod_i \Delta(A^i)$ and, given x in X^S, considered as a stationary strategy, let $\gamma_\lambda(x)[z]$ be the induced payoff in the discounted stochastic game $\Gamma_\lambda(z)$. Let T be a correspondence from X^S to itself defined by

$$T(x) = \{y \in X^S : y^i[z] \text{ is a best reply of player } i \text{ to } x[z] \text{ in the game } \mathcal{G}(\lambda, \gamma_\lambda(x))[z], \forall z\}.$$

Proposition 2 *The correspondence T has a fixed point.*

Proof. We verify that T satisfies the condition of Kakutani's theorem. It is defined on a compact convex set with nonempty compact convex values. Since $\gamma_\lambda(x)$ is continuous in x, the uppersemicontinuous property of T follows. ∎

Note that if x is a fixed point of T, the corresponding equilibrium payoff profile in $\mathcal{G}(\lambda, \gamma_\lambda(x))[z]$ is $\gamma_\lambda(x)[z]$.

Proposition 3 *If $x \in X^S$ defines, for each z in S, an equilibrium of $\mathcal{G}(\lambda, f)[z]$ with payoff $f(z)$, then the induced stationary strategy is an equilibrium in Γ_λ with payoff f.*

Proof. We first notice that $f(z) = \gamma_\lambda(x)[z]$, which is the payoff if x is played in $\Gamma_\lambda(z)$. By the property of x one has, for any σ^i, with $z_1 = z$,

$$E^z_{\sigma^i, x^{-i}}(\lambda r^i_1 + (1-\lambda) f^i(z_2)) \leq f^i(z_1),$$

and similarly at each stage $n \geq 1$

$$E^z_{\sigma^i, x^{-i}}(\lambda r^i_n + (1-\lambda) f^i(z_{n+1}) | \mathcal{H}_n) \leq f^i(z_n),$$

and one multiplies by $(1-\lambda)^{n-1}$, takes expectation and summation to obtain

$$\gamma^i_\lambda(\sigma^i, x^{-i})[z] \leq f^i(z) = \gamma^i_\lambda(x)[z].$$

∎

The two previous propositions now imply

Theorem 2 *Any finite discounted stochastic game has an equilibrium in stationary strategies.*

The same proof extends to compact action spaces when payoff and transition functions are jointly continuous in actions.

Also, one can handle the case of a countable state space by successive approximations. If $S = \{s_m\}$, $\Gamma(n)$ is the game with $n+1$ states where all the states with rank $> n$ are replaced by a single absorbing state with payoff 0. Let $x(n)$ be a corresponding equilibrium profile. Then $\overline{x}$ obtained by the diagonal procedure is an equilibrium of the original game.

For the general state case see [2] and [5].

References

1. Fink, A.M. (1964) Equilibrium in a stochastic n-person game, *Journal of Science of the Hiroshima University*, Series A-I, **28**, 89–93.
2. Mertens, J.-F. and Parthasarathy, T.E.S. (2003) Equilibria for discounted 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 10, pp. 131–172.
3. Mertens, J.-F., Sorin, S. and Zamir, S. (1994) Repeated games, CORE Discussion Papers 9420, 9421, 9422, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
4. Nowak, A.S. (2003) Zero-sum stochastic games with Borel state spaces, 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 7, pp. 77–91.
5. Nowak, A.S. (2003) N–person stochastic games: Extensions of the finite state space case and correlation, 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 8, pp. 93–106.
6. Shapley, L.S. (1953) Stochastic games, *Proceedings of the National Academy of Sciences of the U.S.A.* **39**, 1095–1100 (Chapter 1 in this volume).
7. Sorin, S. (2003) Classification and basic tools, 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 3, pp. 27–35.
8. Takahashi, M. (1964) Equilibrium points of stochastic non-cooperative n-person games, *Journal of Science of the Hiroshima University*, Series A-I, **28**, 95–99.