CLASSIFICATION AND BASIC TOOLS

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

1. Basic Ingredients

A *stochastic game* is a multi-stage game played in discrete time where, at each stage, the stage game played depends upon a parameter called *state*. The value of the state evolves as a function of its current value and the actions of the players. Let I be the finite set of *players* and S be the set of states. For each state z in S, an I-player normal form game is specified by *action sets* $A^i(z)$ for each player i in I and *reward functions* $r^i(z,.)$, i in I, from the set of action profiles at z, $A(z) = \prod_{i\in I} A^i(z)$ to the reals, $\mathbb{R}$. In addition, for any pair consisting of a state z in S and an action profile a in $A(z)$, a probability $p(.|z,a)$ on S describes the random *transition*.

Comments. The *finite case* corresponds to the model where the state space S and any action set $A^i(z)$ are finite. One can then assume w.l.o.g. that $A^i(z)$ is independent of z. In the general case, S has a measurable structure and for any player i, $\{(z,a^i); z\in S,\ a^i\in A^i(z)\}$ is a measurable subset of $S\times\overline{A}^i$ where $\overline{A}^i$ is a measurable space.

2. Game Form

We describe here the traditional model of stochastic games ([19], [7]). Generalizations will be introduced later on in [4], [21].

A stochastic game is played in stages and a play of the multi-stage game evolves as follows. The initial state at stage $n = 1$, z_1, is known to the players (public knowledge). Each player i chooses an action a_1^i; this defines an action profile $a_1 = \{a_1^i\}_{i\in I}$ which is announced to all players. The stage reward is the vector $r_1 = \{r^i(z_1,a_1)\}_{i\in I}$ and the new state z_2 is selected according to the distribution $p(.|z_1,a_1)$ on S.

At stage n, knowing the *history* $h_n = (z_1,a_1,...,a_{n-1},z_n)$ (the sequence of states and actions up to that stage), each player i chooses an action

a_n^i. The state z_n and the profile $a_n = \{a_n^i\}_{i\in I}$ determine the stage reward $r_n = \{r^i(z_n, a_n)\}_{i\in I}$ and the distribution $p(.|z_n, a_n)$ of the new state z_{n+1}. Denote by H_n the set of histories at stage n and by $H = \cup_{n\geq 1} H_n$ the set of all histories. H_∞ is the set of *plays* defined as infinite sequences $(z_1, a_1, ..., a_{n-1}, z_n, ...)$; each play specifies a stream of payoffs $(r_1, ..., r_n, ...)$. Several games will be associated to specific evaluations of this sequence of payoffs, as in other multi-stage game forms with stage payoffs.

Remark. Stochastic games appear as the I-player extension of Stochastic Dynamic Programming or Markov Decision Processes (e.g., [3], [17]) that corresponds to the one-player case. Nonstationary models, where the transition is also a function of the stage, have also been studied.

3. Strategies

The next step is to introduce several classes of strategies, starting with the simpler finite case.

In this setup a *pure strategy* σ^i of player i is a mapping from histories to actions, i.e., from H to A^i. The restriction σ_n^i of σ^i to H_n describes the behavior at stage n. Let PS^i denote the set of pure strategies endowed with the natural product $\sigma-$algebra generated by the sets of strategies that coincide with some given strategy σ^i up to some stage n.

A *mixed strategy* is a probability distribution on PS^i: it is the random choice of a pure strategy. The set of mixed strategies is denoted MS^i.

A *behavioral strategy* is a mapping μ^i from histories to probabilities on actions, i.e., from H to $\Delta(A^i)$ (where, given a set C, $\Delta(C)$ denotes the set of probabilities on it). The restriction μ_n^i of μ^i to H_n describes the random behavior at stage n. The set of behavioral strategies is denoted by BS^i.

Consider now the general case where S and $\overline{A}^i$ are measurable spaces. Measurability requirements are needed in each class. Note that H_n, as a product space, has a measurable structure and endow H_∞ with the product σ-algebra. A pure strategy σ^i is a mapping from H to $\overline{A}^i$ such that σ_n^i is a measurable mapping from H_n to $\overline{A}^i$ that maps histories ending with z_n to elements of $A^i(z_n)$.

Similarly a behavioral strategy μ^i is a mapping from H to $\Delta(\overline{A}^i)$ such that μ_n^i is a measurable probability transition from H_n to $\overline{A}^i$ that maps histories ending with z_n to elements of $\Delta(A^i(z_n))$.

Several equivalent ways of defining mixed strategies are available, all corresponding to the idea of a random choice of a pure strategy: one can define a measurable structure on PS^i or consider "pure strategies" on $H_\infty \times \Omega$ where Ω is an auxiliary nonatomic probability space [1].

The initial state z_1 and a profile of strategies σ where each σ^i is in PS^i or MS^i or BS^i define (by Kolmogorov's extension theorem) a unique probability $\mathbf{P}_\sigma^{z_1}$ on the space of plays H_∞. $E_\sigma^{z_1}$ denotes the corresponding expectation.

Since the game has perfect recall (each player remembers what he did and what he knew), Kuhn's theorem [9] applies. For each player i and each strategy σ^i in MS^i (resp. μ^i in BS^i) there exists a strategy μ^i in BS^i (resp. σ^i in MS^i) such that for any $(I-1)$ profile τ^{-i} of any types of strategies of the other players, the induced probabilities on plays coincide:

$$\mathbf{P}_{\sigma^i,\tau^{-i}}^{z_1} = \mathbf{P}_{\mu^i,\tau^{-i}}^{z_1}.$$

This allows us to consider equivalently behavioral or mixed strategies.

The strategy σ^i is *Markov* (resp. *Markov stationary*, or stationary for short) if it is, at each stage n, a function α_n^i (resp. α^i) of the current state z_n and of the stage n (resp. of the current state z_n only). A stationary strategy α^i is thus a transition probability from S to $\overline{A}^i$ mapping z to a probability on $A^i(z)$.

Remark. Note that with the above definition the set of strategies is independent of the initial state. One can work with weaker measurability requirements, as long as the initial state and the profile of strategies define a probability $\mathbf{P}_\sigma^{z_1}$.

4. Payoffs and Solution Concepts

There are basically three different ways of evaluating the payoffs when dealing with games with a large number of stages.

4.1. ASYMPTOTIC STUDY

The first approach leads to the "compact case": under natural assumptions on the action spaces and on the reward function the mixed strategy spaces will be compact for a topology for which the payoff function will be continuous.

Two typical examples correspond to:
i) the *finite n-stage* game $\Gamma_n(z)$ with initial state z and payoff given by the average of the n first rewards:

$$\gamma_n^z(\sigma) = E_\sigma^z(\frac{1}{n}\sum\nolimits_{m=1}^{n} r_m).$$

In the finite case, this reduces to a game with finitely many pure strategies.
ii) the *λ-discounted game* $\Gamma_\lambda(z)$ with initial state z and payoff equal to the

discounted sum of the rewards:

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

In this setup the first task is to find conditions under which:
- in the two-person zero-sum case the value will exist; it will be denoted respectively by $v_n(z)$ and $v_\lambda(z)$;
- in the I-player case, equilibria will exist; the corresponding sets of equilibrium payoffs will be denoted by $E_n(z)$ and $E_\lambda(z)$. Similarly, one may consider correlated and communication equilibria.

A second aspect of interest is the nature of optimal (or ε-optimal) strategies: existence of Markov, stationary Markov optimal strategies, etc. A related issue is to design efficient algorithms to compute the value or optimal strategies.

Another consideration is the asymptotic behavior of the above objects (value, optimal strategies, equilibrium payoffs, equilibrium strategies) as n goes to ∞ or λ goes to 0 . This is the study of the "asymptotic game."

Remark. To any distribution μ on the positive integers (or any finite stopping time) one can associate a game with payoffs $E_\sigma^z(\sum \mu(m) r_m)$ and the value $v_\mu(z)$ will exist under natural conditions. Similarly, one can study the convergence along nets of such distributions as the weight on any finite set of stages goes to zero.

More generally, the asymptotic game in the compact case could be viewed as a game (in continuous time) played between 0 and 1. Both players know the time (and the length of the game) and the finite, discounted or other discrete-time versions correspond to constraints on the available strategies (basically they are piecewise constant).

In comparison with the MDP literature ([3], [17]), the focus is more on the asymptotics of the values than on the asymptotics of the strategies, which often are not sufficient: the limit of optimal strategies are not optimal.

4.2. INFINITE GAME

The second perspective considers games where the payoff $\gamma^z(\sigma)$ is defined as the expectation w.r.t. $\mathbf{P}_\sigma^z$ of an asymptotic evaluation on plays like: $\limsup r_n$, $\liminf r_n$ or $\limsup \frac{1}{n}\sum_{m=1}^n r_n$, $\liminf \frac{1}{n}\sum_{m=1}^n r_n$ (limiting average criterion). More generally, given a bounded measurable payoff function f on plays, one defines $\gamma^z(\sigma) = E_\sigma^z(f(h_\infty))$.

In this framework, the main difficulty consists in proving the existence of a value, called the *infinite value*, or of an equilibrium. Also of interest is the characterization of simple classes of ε-optimal strategies.

Results in this direction extend the work of Dubins and Savage [5] on gambling; see, e.g., [11].

Remark. In some cases (e.g., of positive reward) the sum of the stage rewards has a limit and the corresponding additive reward game has been studied. See, e.g., [16].

4.3. UNIFORM APPROACH

A third model approaches the infinite game by considering the whole family of "long games." It does not specify payoffs but requires uniformity properties on strategies to define concepts analogous to value or equilibrium [14].

In the zero-sum framework one introduces the following definitions. Player 1 can *guarantee* v, a real function on S, if $\forall z \in S, \forall \varepsilon > 0, \exists \sigma$ strategy of player 1, $\exists N$ such that $\forall n \geq N$, $\forall \tau$ strategy of player 2:

$$\gamma_n^z(\sigma, \tau) \geq v(z) - \varepsilon.$$

Similarly, player 2 can *guarantee* v if $\forall z \in S, \forall \varepsilon > 0, \exists \tau$ strategy of player 2, $\exists N$ such that $\forall n \geq N$, $\forall \tau$ strategy of player 1:

$$\gamma_n^z(\sigma, \tau) \leq v(z) + \varepsilon.$$

If both players can guarantee the same function, it is denoted by v_∞ and the game has a *uniform value*, v_∞. It follows from the above definitions that if player 1 can guarantee v, then both $\liminf_{n\to\infty} v_n(z)$ and $\liminf_{\lambda\to 0} v_\lambda(z)$ will be greater than $v(z)$. In particular, the existence of v_∞ implies

$$v_\infty(z) = \lim_{n\to\infty} v_n(z) = \lim_{\lambda\to 0} v_\lambda(z).$$

For the case where the uniform value does not exist, one defines $\underline{v}(z)$ to be the *maxmin* of the game starting at z if player 1 can *guarantee* $\underline{v}(z)$ and player 2 can *defend* $\underline{v}(z)$ in the sense that: $\forall \varepsilon > 0, \forall \sigma$ strategy of player 1, $\exists N$, $\exists \tau$ strategy of player 2 such that, $\forall n \geq N$:

$$\gamma_n^z(\sigma, \tau) \leq \underline{v}(z) + \varepsilon.$$

A dual definition holds for the *minmax* $\overline{v}(z)$.

In the non-zero-sum case one similarly defines equilibrium payoffs through approximate robust equilibria as follows. The set of *uniform equilibrium payoffs* starting from state z is $E_0^z = \cap_{\varepsilon>0} E_\varepsilon^z$ where E_ε^z is the set of I-vectors g of ε-equilibrium payoffs, namely satisfying: there exist a profile of strategies σ and a natural number N such that

$$\gamma_n^{z,i}(\underline{\sigma}^i, \sigma^{-i}) - \varepsilon \ \leq \ g^i \leq \gamma_n^{z,i}(\sigma) + \varepsilon$$

for all strategies $\underline{\sigma}^i$ of player i, for all i and for all $n \geq N$.

Note that these sets E_ε^z are decreasing as ε goes to 0. Heuristically, g belongs to E_0^z if for any positive ε, there exists a profile σ such that g is within ε of the asymptotic payoff induced by σ, and σ is an ε-equilibrium of any game $\Gamma_n(z)$ for n large enough or $\Gamma_\lambda(z)$ for λ small enough.

Within this approach the main problem is the existence of a uniform equilibrium payoff and eventually a characterization of the set of uniform equilibrium payoffs. Note that if the payoffs are bounded, E_0 is nonempty as soon as for any $\varepsilon > 0$, E_ε is nonempty.

Comments. In all previous cases one can in addition study ε-consistency, i.e., look for strategies that remain ε-optimal on any feasible path [10].

For the comparison of the different approaches (finite, discounted and uniform) and their interpretation we refer to the illuminating comments of [2], Chapter 2, postscripts c, f, g, h.

5. Recursive Structure and Functional Equation

The fact that in a stochastic game the current state is public knowledge among the players allows for a simple recursive structure. A crucial role is played by the following class of one-stage games. Given a profile of functions $f = \{f^i\}$, where each f^i belongs to the set $\mathcal{F}$ of bounded measurable functions on S, define $\Gamma(f)(z)$, the auxiliary game associated to f at z, as the I-player strategic game with strategy set $A^i(z)$ and payoff function $r^i(z,.) + E(f^i|z,.)$, where $E(f^i|z,a) = \int_S f^i(z')p(dz'|z,a)$, for all i in I.

Consider first the finite zero-sum case. Denote by A and B the action sets of the players and by $\mathtt{val}$ the value operator. Assuming that the game $\Gamma(f)(z)$ has a value for all z, the *Shapley operator* is defined by $\Psi : f \mapsto \Psi(f)$ that maps the function f to the values of the family, indexed by S, of auxiliary games associated to f. Ψ is specified on (a complete subset of) $\mathcal{F}$ by the following relation:

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

or explicitly

$$\begin{aligned}\Psi(f)(z) &= \max_{x\in\Delta(A)} \min_{y\in\Delta(B)} \Big(\sum_{a\in A, b\in B} x(a)y(b)r(z,a,b) \\ &\quad + \sum_{a\in A, b\in B, z'\in S} x(a)y(b)p(z'|z,a,b)f(z') \Big) \\ &= \min_{y\in\Delta(B)} \max_{x\in\Delta(A)} \Big(\sum_{a\in A, b\in B} x(a)y(b)r(z,a,b) \\ &\quad + \sum_{a\in A, b\in B, z'\in S} x(a)y(b)p(z'|z,a,b)f(z') \Big).\end{aligned}$$

$\Psi(f)(z)$ expresses the value of the game $\Gamma(f)(z)$ where starting from state z, the stochastic game is played once and there is an additional payoff determined by f at the new state. This corresponds to the usual Bellman operator in the one-player case (MDP).

Notice that Ψ as an operator from $\mathcal{F}$ to itself has two properties: *monotonicity* and *translation of constants*; hence it is *non-expansive*.

Let $0 < \alpha < 1$. Giving a relative weight α on the current reward and $(1-\alpha)$ on f evaluated at the next state, one obtains the *discounted Shapley operator* $\Phi(\alpha, .) : f \mapsto \Phi(\alpha, f)$ defined by

$$\Phi(\alpha, f)(z) = \mathtt{val}\Big(\alpha r(z, .) + (1-\alpha)E(f|z, .)\Big).$$

Both operators Ψ and Φ are related through the equation

$$\Phi(\alpha, f) = \alpha\Psi\Big(\frac{(1-\alpha)}{\alpha}f\Big).$$

These tools allow us to obtain inductively the functions v_n

$$v_{n+1} = \Phi(\frac{1}{n+1}, v_n)$$

and to express v_λ as the unique fixed point of a contracting operator

$$v_\lambda = \Phi(\lambda, v_\lambda)$$

(this is one advantage of the discounted case: the model itself is stationary). The knowledge of the current state is sufficient to play optimally in the above "auxiliary one-shot game" which will imply Markov properties of optimal strategies.

The natural approach, which extends to the general action and state space, is thus to look for a class of functions f such that all corresponding games $\Gamma(f)(z)$ have a value, and are in the same class; and moreover to have enough regularity w.r.t. z to exhibit ε-optimal strategies.

The operator $\Phi(0, .)$ will appear naturally in the asymptotic analysis [22] as well as in the infinite game [12], [13].

In the non-zero sum case a similar approach can be used to study "subgame perfect" equilibria. In the discounted case it will allow us to characterize stationary equilibria [20].

6. Special Classes and Extensions

The basic classification of stochastic games specifies whether the action spaces are finite or compact; the state space can be finite, uncountable or measurable.

In addition, specific assumptions on the transitions have interesting consequences:
- *Irreducible* games are such that a.s. every state will be visited infinitely many times on any play; they are usually much simpler to study. More generally, *unichain* games possess a unique ergodic class $S' \subset S$, for any profile of stationary strategies.
- An *absorbing* state is a state z that one cannot leave, i.e., such that: $p(z|z,a) = 1$, for all profile a. An *absorbing* game [8] is a game with a single nonabsorbing state.
- A *recursive* game [6] is a game where the reward function in any nonabsorbing state is identically 0. Note that in this class $\lim r_n$ and $\lim \frac{1}{n}\sum_{m=1}^{n} r_m$ exist and coincide on any play.
- Other examples will be found in [23], [18].

The principal extensions of the model are obtained by relaxing the hypotheses on the information of the players. Two main streams of research have been considered up to now:
- The first does not assume the "standard signalling" (perfect monitoring) hypothesis that the previous actions are observed. In some cases, the existence of an ε-optimal Markov or stationary strategy indicates that much less than the knowledge of the past history is needed; only the current state and the date are enough. On the other hand, in the zero-sum "uniform approach" framework, the fact that the players know the payoff is sufficient, but also necessary (see [15]. More generally, it is interesting to study the general case of signalling functions: after each stage n, rather than knowing a_n, each player i receives a signal according to some distribution $q^i(.|z_n, a_n)$. However, the current state is assumed to be known. See [4].
- The second allows for incomplete information on the current state for at least one player. Hence the state is no longer public knowledge and the usual recursive structure is no longer available. However, it is usually assumed that standard signalling on the moves holds. See [21].

References

1. Aumann, R.J. (1964) Mixed and behavior strategies in infinite extensive games, in M. Dresher, L.S. Shapley and A.W. Tucker (eds.), *Advances in Game Theory*, Annals of Mathematics Studies 52, Princeton University Press, Princeton, NJ, pp. 627–650.
2. Aumann, R.J. and Maschler, M. (1995) *Repeated Games with Incomplete Information* (with the collaboration of R. E. Stearns), MIT Press, Cambridge, MA.
3. Blackwell, D. (1962) Discrete dynamic programming, *Annals of Mathematical Statistics* **33** 719–726.
4. Coulomb, J.-M. (2003) Absorbing games with a signalling structure, 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 22, pp. 335–355.

5. Dubins, L.E. and Savage, L.J. (1976) *Inequalities for Stochastic Processes: How to Gamble if You Must*, Dover, Mineola, NY.
6. Everett, H. (1957) Recursive games, in M. Dresher, A.W. Tucker and P. Wolfe (eds.), *Contributions to the Theory of Games, Vol. III*, Annals of Mathematics Studies 39, Princeton University Press, Princeton, NJ, pp. 47–78.
7. Filar, J. and Vrieze, O.J. (1997) *Competitive Markov Decision Processes*, Springer-Verlag, Berlin.
8. Kohlberg, E. (1974) Repeated games with absorbing states, *Annals of Statistics* **2**, 724–738.
9. Kuhn, H.W. (1953) Extensive games and the problem of information, in H.W. Kuhn and A.W. Tucker (eds.), *Contributions to the Theory of Games, Vol. II*, Annals of Mathematics Studies 28, Princeton University Press, Princeton, NJ, pp. 193–216.
10. Lehrer, E. and Sorin, S. (1998) ε-consistent equilibrium in repeated games, *International Journal of Game Theory* **27**, 231–244.
11. Maitra, A. and Sudderth, W. (1996) *Discrete Gambling and Stochastic Games*, Springer, Berlin.
12. Maitra, A. and Sudderth, W. (2003) Stochastic games with lim sup payoff, 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 23, pp. 357–366.
13. Maitra, A. and Sudderth, W. (2003) Stochastic games with Borel payoffs, 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 24, pp. 367–373.
14. 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.
15. Neyman, A. (2003) Stochastic games: Existence of the minmax, 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 11, pp. 173–193.
16. 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.
17. Puterman, M. (1994) *Markov Decision Processes*, Wiley, New York.
18. Raghavan, T.E.S. (2003) Finite-step algorithms for single-controller and perfect information 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 15, pp. 227–251.
19. 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).
20. Sorin, S. (2003) Discounted stochastic games: The finite case, 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 5, pp. 51–55.
21. Sorin, S. (2003) Stochastic games with incomplete information, 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 25, pp. 375–395.
22. Sorin, S. (2003) The operator approach to zero-sum 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 27, pp. 417–426.
23. Vrieze, O.J. (2003) Stochastic games, practical motivation and the orderfield prop-

erty for special classes, 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 14, pp. 215–225.