Equilibrium stochastic behaviors in repeated games

Arkady Kryazhimskiy

Mathematical Steklov Institute, Russian Academy of Sciences, Moscow, Russia

International Institute for Applied Systems Analysis, Laxenburg, Austria

1. Theory of repeated games essentially motivated by studies on biological adaptation and economic behavior concentrates on analyses of long-term dynamics in societies of repeatedly interacting agents (players), which follow their individual boundedly rational behavior strategies (see, e.g., [1] – [6]). If the players are allowed to choose their boundedly rational behavior strategies within given sets, their mutually acceptable choices are naturally associated with the behavior strategies forming a game-theoretic equilibrium with respect to the players' individual long-term performance criteria. Long-term equilibria in repeated games with non-specified deterministic behavior strategies were defined in [7].

In this paper, we focus on infinite repeated bimatrics games with non-specified stochastic behavior starategies, in which the expectations of the players' benefits averaged over the game rounds (the players' expected avaraged benefits) serve as the players' long-term performance criteria.

First, we provide conditions sufficient for a (Nash) equilibrium pair of the players' behavior strategies to exist within given sets of the players' admissible behavior strategies; the conditions require in particular that all the players' admissible behavior strategies are strictly randomized, and the sets of the players' admissible behavior strategies satisfy an appropriate convexity assumption.

Next, we consider a particular infinite repeated 2×2-bimatrix game with non-specified behaviour strategies. We depart from coupling the players' particular, 'traditional' deterministic boundedly rational behavior strategies – the 'best reply' ones. Then we allow the players to choose their stochastic behavior strategies in 'neigborhoods' of the 'traditional' ones and characterize the equilibrium behavior strategies. In particular, we find that the equilibrium behavior strategies differ from the 'traditional' deterministic ones and have necessarily non-trivial stochastic components.

The paper incorporates results of [8] – [10].

2. We start off with a bimatrix game given by benefit matrices $A = (a_{ij})_{i\in X_1, j\in X_2}$ and $B = (b_{ij})_{i\in X_1, j\in X_2}$ where $X_1 = \{1,\dots,n\}$, $X_2 = \{1,\dots,m\}$ with some natural n and m; here the row index, $i \in X_1$, and column index, $j \in X_2$, stand for pure strategies of player 1 and player 2, respectively; and a_{ij} and b_{ij} denote, respectively, the benefits player 1 and player 2 receive provided they use their pure strategies i and j, respectively. As usual, we understand mixed strategies of players 1 and 2 as probability measures on the sets of their pure strategies, X_1 and X_2, respectively.

We define *behavior strategies* of player 1 and player 2 as arbitrary maps of $X_1 \times X_2$ into the sets of all mixed strategies of players 1 and 2, respectively. Given

Author's email: kryazhim@iiasa.ac.at

an *initial* pair of the players' pure strategies, $(i_0, j_0) \in X_1 \times X_2$, every behavior strategy of player 1, $p : (i,j) \mapsto p_{ij}$, and every behavior strategy of player 2, $q : (i,j) \mapsto q_{ij}$, determine an *infinite repeated game*; the latter is defined as the infinite discrete-time Markov process on $X_1 \times X_2$, with times (also called *rounds*) $0, 1, \ldots$, the initial state (i_0, j_0) and the transition probability $(i,j) \mapsto p_{ij} \times q_{ij}$.

In the infinite repeated game corresponding to the players' behavior strategies p and q, the *expected average benefits* of players 1 and 2 on round $k \geq 1$ are, respectively, the expectations of the random variables $((i_0, j_0), (i_1, j_1), \ldots) \mapsto (a_{i_1 j_1} + \ldots + a_{i_k j_k})/k$ and $((i_0, j_0), (i_1, j_1), \ldots) \mapsto (b_{i_1 j_1} + \ldots + b_{i_k j_k})/k$ on the probability space of the repeated game's trajectories $((i_0, j_0), (i_1, j_1), \ldots)$. Using properties of finite-state Markov processes (see [1]), one can show that the expected average benefits of players 1 and 2 on round k converge to some limits, which we denote, respectively, $J_1(p,q)$ and $J_2(p,q)$, as $k \to \infty$; $J_1(p,q)$ and $J_2(p,q)$ represent the *expected average benefits* of, respectively, players 1 and 2 in the infinite repeated game corresponding to the players' behavior strategies p and q.

Let players 1 and 2 be allowed to choose their behavior strategies within given sets of *admissible behavior strategies*, S_1 and S_2, respectively. In this situation each player is interested in choosing his/her admissible behavior strategy so as to maximize his/her expected average benefit. A game-theoretic interpretation of that is a *behavior game*, in which the actions (strategies) of players 1 and 2, p and q, vary within S_1 and S_2, respectively, and the benefit functions for players 1 and 2 are $(p,q) \mapsto J_1(p,q)$ and $(p,q) \mapsto J_2(p,q)$, respectively.

3. Consider the issue of the existence of a Nash equilibrium in the behavior game. Following a standard game-theoretical definition, we call a pair $(p^*, q^*) \in S_1 \times S_2$ to be a *Nash equilibrium* (in the behavior game) if p^* maximizes $p \mapsto J_1(p, q^*)$ on S_1 and q^* maximizes $q \mapsto J_2(p^*, q)$ on S_2.

Let us give several definitions. We shall say that S_1 (respectively, S_2) is *strictly randomized* if every $p \in S_1$ (respectively, every $q \in S_2$) takes values in the set of all strictly mixed strategies of player 1 (respectively, player 2).

We shall say that S_1 (respectively, S_2) is *parallepipedally convex* if for every $p^{(1)}, p^{(2)} \in P$ (respectively, $q^{(1)}, q^{(2)} \in Q$) and every family $(\lambda_{ij})_{(i,j) \in X_1 \times X_2}$ in $[0,1]$ the map $(i,j) \mapsto \lambda_{ij} p^{(1)}_{ij} + (1 - \lambda_{ij}) p^{(2)}_{ij}$ (respectively, $(i,j) \mapsto \lambda_{ij} q^{(1)}_{ij} + (1 - \lambda_{ij}) q^{(2)}_{ij}$) lies in P (respectively, in Q).

We shall say that S_1 (respectively, S_2) is *closed* if $\Pi_{(i,j) \in X_1 \times X_2} \{p_{ij} : p \in P\}$ is closed in $(R^n)^{nm}$ (respectively, $\Pi_{(i,j) \in X_1 \times X_2} \{q_{ij} : q \in Q\}$) is closed in $(R^m)^{nm}$).

Our existence theorem reads as follows.

Theorem 1. *Let S_1 and S_2 be strictly randomized, parallelepipedally convex and closed. Then the behavior game has a Nash equilibrium.*

4. Consider a situation where players 1 and 2 dominated by historically justified 'traditional' behavior paradigms explore if small 'innovations' in their 'traditional' behaviors can improve their performance in the long run.

Let each player have two pure strategies in the original bimatrix game, i.e., $X_1 =$

$X_2 = \{1, 2\}$, and the latter bimatrix game have the single mixed equilibrium; then, with no loss in generality (see [12]), we set $a_{11} > a_{21}$, $a_{22} > a_{12}$, $b_{12} > b_{11}$, $b_{21} > b_{22}$. Let the players' 'traditional' behavior be 'reply best to the opponent's latest action'. Therefore, traditionally, in each round each player chooses his/her pure strategy that brings him/her the maximal benefit provided the other player repeats his/her pure strategy used in the previous round. For player 1, that behavior is easily formalized as the deterministic behavior strategy p^0 such that $p^0_{i1} = (1, 0)$, $p^0_{i2} = (0, 1)$ $(i = 1, 2)$; and for player 2 as the deterministic behavior strategy q^0 such that $q^0_{1j} = (0, 1)$, $q^0_{2j} = (1, 0)$ $(j = 1, 2)$. We call p^0 and q^0 the *best reply* strategies of players 1 and 2, respectively. The infinite repeated game corresponding to the players' best reply strategies will be called the *best reply repeated game.*

Now let us allow each player to slightly deviate from his/her 'traditional' behavior, namely, to give a small probability for choosing, in each round, his pure strategy that does not reply best to the opponent's pure strategy realized in the previous round. We call a player's behavior strategy that describes such type of behavior the player's ε-best reply strategy. More specifically, given a small $\varepsilon > 0$, we define the ε-*best reply* strategy of player 1, p, by $p_{i1} = (1-\varepsilon_1, \varepsilon_1)$, $p_{i2} = (\varepsilon_2, 1-\varepsilon_2)$ $(i = 1, 2)$ with arbitrary nonnegative $\varepsilon_1, \varepsilon_2 \leq \varepsilon$; and we define the ε-*best reply* strategy of player 2, q, by $q_{1j} = (\varepsilon_1, 1-\varepsilon_1)$, $q_{2j} = (1-\varepsilon_2, \varepsilon_2)$ $(j = 1, 2)$ with arbitrary nonnegative $\varepsilon_1, \varepsilon_2 \leq \varepsilon$.

Let the players' admissible behavior strategies be his/her ε-best reply strategies; in this way we define S_1 and S_2 – see section 2. We call the above defined behavior game the 2×2 ε-*best reply* one.

Note that S_1 and S_2 include the deterministic best reply strategies; therefore, S_1 and S_2 are not strictly randomized (see section 3). Consequently, for the 2×2 ε_0-best reply behavior game, the conditions sufficient for the existence of a Nash equilibrium, given in Theorem 1, do not hold. The next theorem states the existence and structure of the Nash equilibrium in the 2×2 ε_0-best reply behavior game.

Theorem 2. *Let $a_{12} \neq a_{21}$, $b_{11} \neq b_{22}$ and ε be sufficiently small. Then the following statements hold true:*

(i) the 2×2 ε-best reply behavior game has the single Nash equilibrium (p^*, q^*);

(ii) $p^*_{i1} = (1, 0)$, $p^*_{i2} = (\varepsilon, 1-\varepsilon)$ if $a_{12} > a_{21}$, and $p^*_{i1} = (1-\varepsilon, \varepsilon)$, $p^*_{i2} = (0, 1)$ $(i = 1, 2)$ if $a_{12} < a_{21}$ $(i = 1, 2)$;

(iii) $q^*_{1j} = (\varepsilon, 1-\varepsilon)$, $q^*_{2j} = (1, 0)$ if $b_{11} > b_{22}$, and $q^*_{1j} = (0, 1)$, $q^*_{2j} = (1-\varepsilon, \varepsilon)$ if $b_{11} < b_{22}$ $(j = 1, 2)$.

Thus, the equilibrium behavior strategies differ from the 'traditional' deterministic ones and have necessarily non-trivial stochastic components. A substantial interpretation can be the following: as soon as the players realize that they are allowed to introduce stochastic perturbations in their 'traditional' deterministic best reply behaviors, they get a motivation to change their 'traditional' deterministic behaviors to the equilibrium stochastic ones that are more favorable for both of them.

Acknowledgements. This work was partially supported by RFBR, project 11-01-12112-ofi-m-2011.

References

[1] J. Hofbauer, K. Sigmund (1988). *The Theory of Evolution and Dynamical Systems.* Cambridge University Press.

[2] J. Weibull (1995). *Evolutionary Game Theory.* The M.I.T. Press.

[3] D. Fudenberg, D. M. Kreps (1993). Learning mixed equilibria. *Games and Economic Behavior* 5: 320-367.

[4] M. Nowak, K. Sigmund (1994). The alternating Prisoner's Dilemma. *J. Theor. Biol.* 168: 219-226.

[5] G. Van der Laan, X. Tieman (1998). Evolutionary game theory and the modeling of economic behavior *De Economist* 146: 59-89.

[6] Yu. M. Kaniovski, A.V. Kryazhimskiy, H.P. Young (2000). Learning equilibria in games played by heterogeneous populations. *Games and Economic Behavior* 31: 50-96.

[7] A.V. Kryazhimskiy, Yu.S. Osipov (1995). On evolutionary-differential games. *Proc. Steklov Math. Inst.* 211: 257-287.

[8] A.V. Kryazhimskiy, A.V.Raigorodskaya (2012). On equilibrium behavior strategies in infinite repeated games. *Problems of Dynamic Control. Proceedings of the Faculty of Computational Math. Cybern., Moscow State Univ.* Maskpress. Issue 6 (in press).

[9] A.V.Raigorodskaya (2011). Equilibrium behaviors of players in an infinite repeated game of ε-best replies of dimension 2×2. *Proc. Inst. Math. Mechan., Ural Branch, Russian Adac. Sci.* 17: 201-216.

[10] A.V.Raigorodskaya (2012). To choosing the equilibrium behavior in an infinite repeated game of dimension 2×2: the case of innovative players and conservative players. *Vestnik Moscow State Univ. Series 15. Computational Math. Cybern.*, No 1: 15-22.

[11] A.N. Shiryaev (2004). *Probability – 2.* MCNMO.

[12] N.N. Vorobiev (1985). *Game Theory for Economists-Cyberneticians.* Nauka.