JETEM

Journal of Economic Theory and Econometrics

The Journal of Korean Econometric Society

December 2007 · Volume 18 · Issue 4 · pp. 52-61

Correlation on belief and convergence to Nash equilibrium in repeated games

Jinwoo Kim*

Abstract This paper presents an example in which players never learn to play Nash equilibrium in the repeated games when one believes that others are correlating their strategies, though his belief satisfies `a grain of truth' condition. A sufficient condition on the learning procedure to restore the convergence is suggested and turns out to be quite demanding.

Keywords Repeated game, Bayesian learning, convergence to Nash equilibrium, correlation on belief.

JEL Classification C72, D83

* School of Economics, Yonsei University, Korea

1 Introduction

The concept of Nash equilibrium is being widely used in economics and other social sciences. While it can be readily described as a state where players are mutually best-responding to each other's strategies, there underlies a strong assumption that each player precisely predicts the opponents' equilibrium strategies. This assumption is quite demanding unless one introduces another step before the game in which players can exchange their information about strategies to be used in the subsequent game. It is more reasonable to assume that players start with (possibly) *inaccurate beliefs* about others' strategies and revise them to get more accurate prediction as information accumulates. This process is only made possible when players are interacting more than once and perhaps repeatedly. To formalize this idea, Fudenberg and Kreps (1988) and others have developed the learning theory to show that players' optimizing behavior coupled with various modes of learning can 'converge' to a Nash equilibrium.

In particular, the Bayesian learning theory assumes that players, who repeatedly play the same stage game every period, Bayes-update their subjective priors in light of the past history to form posterior beliefs, to which they are best-responding either myopically or non-myopically. The myopic players only maximize their per-period payoffs each period, thus behaving as if they play one-shot game, though they are actually playing in the repeated game. To overcome this problem, Kalai and Lehrer (1993, hereafter KL) assume that players maximize their long-run expected payoffs while considering how their current behavior will affect the learning process and thereby their payoffs in the future. KL explore under what conditions on the subjective beliefs players can eventually learn to play Nash equilibrium of the repeated game. They show that such convergence occurs when players have the beliefs over histories that are absolutely continuous with the probability distribution over histories induced by their actual strategy profile. They apply this result to the two player repeated game with incomplete information to obtain a similar convergence result.

As noted by KL, however, there is difficulty with extending the above convergence result to the repeated game with more than two players. It is due to the possibility that one might believe that others' strategies depend on some correlated signals, in which case KL claims that the convergence will fail. This paper confirms KL's claim. More precisely, I present an example in which the convergence to a Nash equilibrium fails with probability *one* when one of three players believes in others correlating their strategies, but puts a positive probability on their actual strategy profile, which is *not* correlated. Thus, if the player takes enough time to observe how opponents are playing, then his belief might well approximate others' actual strategies and enable him to play a corresponding optimal response eventually, which implies that the players' strategy profile would converge to a Nash equilibrium rather than a correlated equilibrium since the opponents' actual strategies are not correlated. The above example, however, exhibits that such convergence does not occur. This example is rather surprising, considering that the belief satisfies a "a grain of truth" condition, which is quite strong in that one player's belief puts a positive probability on opponents' actual strategies.

By investigating the example, I suggest a stronger assumption on learning than what is implied by a grain of truth condition, and provide a new convergence result to Nash equilibrium. Not surprisingly, the sufficient condition for the convergence turns out to be very demanding. Also, that condition is just intended to get rid of what has been the cause of convergence failure in the above example. In this sense, one can view the convergence result in this paper as a rather negative message about obtaining the convergence result without imposing the independence on players' beliefs. It still remains to be seen, though, what is necessary and sufficient condition if correlations are allowed in players' beliefs.

This paper is organized as follows. Section 2 introduces the formal framework and KL's results. Section 3 presents an aforementioned example. Section 4 establishes the convergence result along with an alternative learning assumption.

2 Kalai and Lehrer's convergence result

Consider a model of n-person infinitely repeated game with a finite stage game. We introduce an array of notations as follows:

$S := \prod_i S_i$	Product action space of the stage game, where S_i is player i's action space. $s \in S$ and $s_{-i} \in S_{-i} := \prod_{j \neq i} S_j$.
$x_i : S \to \mathbb{R}$	Payoff function of player i at the stage game
$H := S^\infty$	Set of all infinite play paths. $h \in H$.
$h(t)$	Finite truncation of infinite history h up to t^{th} period.
$\overline{H}$	Set of all finite histories.
$C(h(t)) \subset \overline{H}$	Set of all finite histories which belong to the subgame starting after a finite history $h(t)$.
Σ	Borel σ-field generated by $\overline{H}$
$M(\Sigma)$	Set of all probability measures on Σ.
$f_i : \overline{H} \to \Delta(S_i)$	Behavior strategy of player i. $f := (f_1, \cdots, f_n) \in F := \prod_i F_i$, where F_i is the set of player i's behavior strategies. $f_{-i} \in F_{-i} := \prod_{j \neq i} F_j$.
$f_{ih(t)}$	Behavior strategy of player i in the subgame $C(h(t))$ for a finite history $h(t)$, prescribed by f_i.
$m(f) \in M(\Sigma)$	Probability distribution on Σ induced by the behavior strategy profile.

For a strategy profile $f = (f_1, \cdots, f_n)$, the long-run expected payoff of player i with discount factor is given by

$$u_i(f) := \int_{h \in H} U_i(h) dm(f)(h),$$

where

$$U_i(h) := (1-\delta) \sum_{t=0}^{\infty} \delta^t x_i(f(h(t))).$$

If we endow each F_i with the pointwise convergence topology and $M(\Sigma)$ with the weak convergence topology, then $m(\cdot)$ and $u_i(\cdot)$ are continuous.[1] Now, $f = (f_1, \cdots, f_n)$ is a Nash equilibrium if for all i,

$$u_i(f_i, f_{-i}) \geq u_i(f_i', f_{-i}), \ \forall f_i' \in F_i.$$

Also, for $\epsilon > 0$, $f = (f_1, \cdots, f_n)$ is an $\epsilon-$Nash equilibrium if for all i,

$$u_i(f_i, f_{-i}) \geq u_i(f_i', f_{-i}) - \epsilon, \ \forall f_i' \in F_i.$$

KL establish the convergence result using the following notion of asymptotic equivalence between two strategy profiles:

DEFINITION 1. *Let $\epsilon > 0$ and let μ and $\tilde{\mu}$ be two probability measures defined on the same space. We say that μ is ϵ-close to $\tilde{\mu}$ if there is a measurable set Q satisfying:*

(i) $\mu(Q)$ and $\tilde{\mu}(Q)$ are greater than $1-\epsilon$.

(ii) for every measurable set $A \subset Q$,

$$(1-\epsilon)\tilde{\mu}(A) \leq \mu(A) \leq \tilde{\mu}(A).$$

DEFINITION 2. *A behavior strategy f plays ϵ-like g if $m(f)$ is ϵ-close to $m(g)$.*

Letting $f^i = (f_1^i, \cdots, f_n^i)$ represent player i's (subjective) belief about others' strategies, KL prove their main result:

THEOREM 1. (Kalai and Lehrer, 1993) *Letting $f = (f_1, \cdots, f_n)$ with $f_i = f_i^i$ be the actual strategy profile chosen by players, suppose that for each player i,*

(i) f_i is a best response to f_{-i}^i and

(ii) $m(f)$ is absolutely continuous with respect to $m(f^i)$.

Then, for every $\epsilon > 0$ and for almost all infinite histories h (with respect to $m(f)$), there is a time $T(\epsilon)$ such that for every $t > T(\epsilon)$, there exists an ϵ-Nash equilibrium $\hat{f}$ of the repeated game satisfying $f_{h(t)}$ plays ϵ-like $\hat{f}$.

[1] See, for instance, Jordan (1995).

As an application of this result, KL consider a two player infinitely repeated game with incomplete information, where types are represented by randomly generated payoff matrices $(A_i, B_j)_{(i,j)\in I\times J}$ with $I \times J$ being finite. Then, player 1 and 2 choose vectors of strategies, $(f_i)_{i\in I}$ and $(g_j)_{j\in J}$, respectively. In this setup, KL establish the following convergence result.

THEOREM 2. (Kalai and Lehrer, 1993) *For every ϵ and almost every play path z (relative to the distribution induced by (f_i, g_j)), we can find a time T such that for all $t > T$, there is an ϵ-Nash equilibrium of the realized repeated game (A_i, B_j), $(\hat{f}, \hat{g})$, with $(f_i, g_j)_{z(t)}$ playing ϵ-like $(\hat{f}, \hat{g})$.*

KL explain how this result can be extended to the repeated game of more than two players. As they explain, however, with more players, there arises a possibility that any given player could form a belief that opponents' strategies are correlated. KL claim that the convergence theorem would fail in such a case. In the next section, I presents an example which confirms the KL's claim.

3 Failure of convergence: An example

Consider the following stage game with 3 players where player 1, 2, and 3 choose among rows, columns, and matrices, respectively[2]:

	L_2	R_2
L_1	4	0
R_1	0	-4

L_3

	L_2	R_2
L_1	-4	0
R_1	0	4

M_3

	L_2	R_2
L_1	0	3
R_1	3	0

R_3

	L_2	R_2
L_1	0	1
R_1	1	0

A

The numbers in matrices represent the payoffs of player 3. The stage game payoffs of player 1 and 2 are assumed to be some constant numbers not depending on what action profile is chosen. So, we trivialize the strategic consideration of player 1 and 2 to focus on player 3 for whom the learning fails.

For later use, let us compute the minmax payoff for player 3: with p_1, p_2 denoting the probabilities with which player 1 and 2 use L_1 and L_2, respectively,

$$\min_{p_1,p_2}[\max\{4p_1p_2 - 4(1-p_1)(1-p_2), -4p_1p_2 + 4(1-p_1)(1-p_2),$$
$$3(1-p_1)p_2 + 3p_1(1-p_2), (1-p_1)p_2 + p_1(1-p_2)\}] = 1.5 \text{ when } p_1 = p_2 = 0.5.$$

Suppose now that the above stage game is infinitely repeated. Player 3 believes that player 1 and 2 use the following correlating device before the game starts: they

[2]This strategic form game is modified from the extensive form game in the Example 2 of Fudenberg and Levine (1993).

draw a number from the interval [0,1] using the uniform distribution and take its binary expansion up to T-th place. Let $w = (w_1, w_2, \cdots, w_T)$ represent such binary expansion, where $w_t = 0$ or 1 for all t. Thereafter, they play according to the strategies g_{1w}, g_{2w} defined as follows: for $i = 1, 2$

(i) $g_{iw}(h(t))(L_i) = 0.5$ if $t = 1$ or player 3 has chosen only A until period t,

(ii) $g_{iw}(h(t))(L_i) = w_t$ if player 3 has chosen other action than A at some period $t' < t$ for the first time and $t \le t' + T$,

(iii) $g_{iw}(h(t))(L_i) = 1$ if player 3 has chosen other action than A at some period $t' < t$ for the first time and $t > t' + T$.

Given this belief, at the beginning of any period t, the prediction of player 3 about the opponents' actions at that period, is given as follows:

(i) If player 3 has chosen only A until period t, it is

	L_2	R_2
L_1	0.25	0.25
R_1	0.25	0.25

(ii) If player 3 has chosen other action than A at some period $t' < t$ for the first time, then for period $t \le t' + T$, his prediction remains to be

	L_2	R_2
L_1	0.5	0
R_1	0	0.5

and for period $t > t' + T$, it becomes

	L_2	R_2
L_1	1	0
R_1	0	0

We claim that if T is sufficiently large, the best response of player 3 to his belief is to play A at every period. Clearly, it suffices to check if player 3 has no incentive to deviate at a history where he has so far played A only. If he does not deviate or keeps on playing A, his expected long-run payoff is

$$u_N = (1-\delta)\frac{0.5}{(1-\delta)} = 0.5.$$

If he deviates, the most he can get comes from choosing R_3 so

$$\begin{aligned} u_D &= (1-\delta)(1.5 + 0\delta + 0\delta^2 + \cdots + 0\delta^T + 4\delta^{T+1} + 4\delta^{T+2} + \cdots) \\ &= (1-\delta)(1.5 + \frac{4\delta^{T+1}}{(1-\delta)}) = 1.5(1-\delta) + 4\delta^{T+1}. \end{aligned}$$

When $\delta > \frac{2}{3}$, $u_N = 0.5 > \lim_{T\to\infty} u_D = 1.5(1-\delta)$. Thus, if $\delta > \frac{2}{3}$, then, for sufficiently large T, we have $u_N > u_D$, so it is better for player 3 not to deviate. Note that $u_N = 0.5$ is strictly less than player 3's minmax payoff, 1.5 as calculated above, which implies by the observation below, player 3's strategy never approaches a Nash or even an ϵ-Nash equilibrium of the repeated game.[3]

Observation Let $\underline{v}_i$ be the minimax payoff of player i. Player i's payoff is at least $\underline{v}_i$ in any Nash equilibrium of the repeated game, regardless of the level of the discount factor.

Given that player 1 and 2 choose one of the strategy profiles (g_{1w}, g_{2w}) described above, player 3's belief satisfies the property of 'a grain of truth', which means that his belief puts a positive weight on the actual strategy profile chosen by opponents.[4] Thus, our result implies that even a grain of truth condition is not enough to produce the convergence to a Nash equilibrium when players are allowed to believe that opponents are correlating their strategies. Indeed, the convergence fails with probability one.

It is worth noting that if player 3 did a small experiment by choosing some other action than A, then he could obtain the precise prediction about the future once T periods elapse. However, he never attempts to do so since he believes that trying such experiment will trigger the correlation on opponents' actions for the next T periods, which hurts player 3's payoff too badly to be compensated by the gain he can expect from $T+1^{th}$ period on.

4 A convergence result

This section aims to establish a convergence to Nash equilibrium analogous to the Theorem 2. Before doing so, let us back to the previous example to explore why a grain of truth condition was not sufficient to guarantee the convergence and investigate how it can be strengthened. Note first that if a player i's belief satisfies a grain of truth, then, *whatever* behavior strategy he adopts, the resulting (actual) probability distribution over the set of histories is absolutely continuous with the one induced by his belief, which implies by the Theorem 1 that the true probability distribution is eventually ϵ-close to the one induced by i's belief. Although this ensures that the convergence occurs on *almost every* play paths, it might fail along *some play paths in the support.*[5] This can be easily seen in our example. Consider, for instance, player 3's strategy that chooses each action with equal probability at every period. Given the belief in the example, at any history

[3] This is a basic result in the repeated game theory. For instance, see Fudenberg and Tirole (1995, p. 151).

[4] In fact, player 3 assigns the equal probability $(\frac{1}{2})^T$ to each strategy pair (g_{1w}, g_{2w}).

[5] As will be formally defined shortly, a play path belongs to the support if all of its finite truncations are assigned positive probabilities.

where player 3 has so far played A only, player 1 can never predict what will happen from the next period should some other action than A be chosen this period. So, the learning fails along the play path $h = (\cdot, \cdot, A)^\infty$, which, given the above strategy profile of player 3, belongs to the support of the actual probability distribution over histories but occurs with probability zero.

The above argument motivates us to impose a stronger condition than a grain of truth so that the learning occurs on every play path in the support, whether or not the prior beliefs held by players allow for the correlation among others' strategies. To do so, we represent player i's belief by a distribution π_i on F_{-i}. Then, his belief can be described by the function $f^i_{(-i)} : \overline{H} \to \Delta(S_{-i})$ satisfying for each $h(t) \in \overline{H}$ and s_{-i},

$$f^i_{(-i)}(h(t))(s_{-i}) = \int_{f_{-i}\in F_{-i}} \prod_{j\neq i} f_j(h(t))(s_j) d\pi_i(f_{-i}).$$

Note that $f^i_{(-i)}(h(t))(s_{-i})$ need not be a product measure belonging to $\prod_j \Delta(S_j)$ since π_i allows for a correlation. We say that an infinite play path h belongs to the *support* of a probability measure $m \in M(\Sigma)$, denoted $h \in \text{supp } m$, if $m(h(t)) > 0$ for every t. Then, the aforementioned learning assumption is stated as follow:

Assumption 1. *Let f_{-i} be the actual strategy chosen by player i's opponents. Then, player i's belief, $f^i_{(-i)}$, satisfies the followings: for any $f'_i \in F_i$ and $\epsilon > 0$,*

(i) if $h \in \text{supp } m(f'_i, f_{-i})$, then $h \in \text{supp } m(f'_i, f^i_{(-i)})$

(ii) there exist $T(\epsilon)$ such that for all $t > T(\epsilon)$,

$$\left|m(f'_{ih(t)}, f_{-ih(t)})(A) - m(f'_{ih(t)}, f^i_{(-i)h(t)})(A)\right| < \epsilon, \forall A \in \Sigma, \forall h \in \text{supp } m(f'_i, f_{-i}).$$

According to (i) of Assumption 1, an infinite history, which is in the support of true distribution, is also required to be in the support of the distribution induced by the player i's belief. It is required by (ii) that the learning occurs on every path in the support of the true distribution. In case agent i believes in a correlation among opponents' actions, this condition amounts to requiring that the correlation on each player i's belief will eventually unravel since the actual strategies of opponents, f_{-i}, are independently chosen and the probability measure corresponding to player i's belief, $m(f'_{ih(t)}, f^i_{(-i)h(t)})$, approximates $m(f'_{ih(t)}, f_{-ih(t)})$.

Note that the learning condition in Assumption 1 is very demanding and imposed on the learning process itself, not on the prior belief π_i. It begs the question what condition on the prior belief can yield the learning process satisfying Assumption 1. This paper does not attempt to answer that question since our aim here is to highlight the cause of convergence failure in the previous example by showing that the convergence can be restored by making an assumption, such as Assumption 1 above, that is intended to remove that cause.

It is now established that the convergence to a Nash equilibrium obtains under Assumption 1.

THEOREM 3. *Let $f^i_{(-i)}$ be the belief of each player i and f_i his best response to that. Suppose that $f^i_{(-i)}$ satisfies Assumption 1 for all i. Then, for any $h \in$ supp $m(f)$, every limit point of $f_{h(t)}$ is a Nash equilibrium.*[6]

PROOF: Fix $h \in \text{supp } m(f)$. Since, by (i) of the Assumption 1, $h \in \text{supp } m(f_i, f^i_{(-i)})$ and f_i is best response to $f^i_{(-i)}$, we have

$$u_i(f_{ih(t)}, f^i_{(-i)h(t)}) \geq u_i(f_i^{'}, f^i_{(-i)h(t)}), \forall f_i^{'} \in F_i.\forall t \tag{1}$$

Let $\bar{f}$ be a limit point of $f_{h(t)}$. Suppose to the contrary that $\bar{f}$ is not a Nash equilibrium. Then, there must be some i and $\hat{f}_i$ such that

$$u_i(\hat{f}_i, \bar{f}_{-i}) - u_i(\bar{f}_i, \bar{f}_{-i}) = \epsilon > 0. \tag{2}$$

We derive a contradiction.

Since $F_{(-i)}$ is compact in the topology of pointwise convergence, we can take a subsequence $t(k)$ such that $f^i_{(-i)h(t(k))}$ converges to some $\tilde{f}^i_{(-i)}$ as $k \to \infty$. By continuity of $m(\cdot)$ and (ii) of Assumption 1, it holds that $m(\bar{f})(A) = m(\bar{f}_i, \tilde{f}^i_{(-i)})(A)$ for all $A \in \Sigma$. Therefore,

$$u_i(\bar{f}) = u_i(\bar{f}_i, \tilde{f}^i_{(-i)}). \tag{3}$$

Also, the inequality (1) and the continuity of $u_i(\cdot)$ imply

$$u_i(\bar{f}_i, \tilde{f}^i_{(-i)}) \geq u_i(f_i', \tilde{f}^i_{(-i)}), \forall f_i' \in F_i. \tag{4}$$

Now, we construct a strategy $g_i \in F_i$ such that for each t and $\overline{h} \in \overline{H}$,

$$g_i(\overline{h}) = \epsilon_t f_i(\overline{h}) + (1-\epsilon_t)\hat{f}_i(\overline{h}) \text{ if } \overline{h} \in C(h(t)) - C(h(t+1)),$$

where each $\epsilon_t > 0$ and $\epsilon_t \to 0$ as $t \to \infty$. By construction of g_i, $g_{ih(t)}$ converges to $\hat{f}_i$ and $h \in \text{supp } m(g_i, f^i_{-i})$. Thus, (ii) of the Assumption 1 with $f_i' = g_i$ and the continuity of $u(\cdot)$ imply that for sufficiently large k,

$$(5) = \left|u_i(g_{ih(t(k))}, f_{-ih(t(k))}) - u_i(g_{ih(t(k))}, f^i_{(-i)h(t(k))})\right| < \frac{\epsilon}{3}.$$

Also, since the profiles $(g_{ih(t(k))}, f_{-ih(t(k))})$ and $(g_{ih(t(k))}, f^i_{(-i)h(t(k))})$ converges to $(\hat{f}_i, \bar{f}_{-i})$ and $(\hat{f}_i, \tilde{f}^i_{(-i)})$, respectively, taking sufficiently large k yields

$$\begin{aligned} (6) &= \left|u_i(g_{ih(t(k))}, f_{-ih(t(k))}) - u_i(\hat{f}_i, \bar{f}_{-i})\right| < \frac{\epsilon}{3} \\ (7) &= \left|u_i(g_{ih(t(k))}, f^i_{(-i)h(t(k))}) - u_i(\hat{f}_i, \tilde{f}^i_{(-i)})\right| < \frac{\epsilon}{3}, \end{aligned}$$

[6] As mentioned in Section 2, the relevant topology is that of pointwise convergence.

again, by the continuity of $u_i(\cdot)$.

Therefore,

$$\left|u_i(\hat{f}_i, \bar{f}_{-i}) - u_i(\hat{f}_i, \tilde{f}^i_{(-i)})\right| \leq (5) + (6) + (7) < \epsilon,$$

which implies together with (2) and (3) that

$$u_i(\hat{f}_i, \tilde{f}^i_{(-i)}) - u_i(\bar{f}_i, \tilde{f}^i_{(-i)}) = u_i(\hat{f}_i, \tilde{f}_{(-i)}) - u_i(\hat{f}_i, \bar{f}^i_{-i}) + \epsilon > 0,$$

contradicting with (4). ▮

5 Concluding remark

This paper presents an example in which players in repeated games fail to learn Nash equilibrium if one player believes opponents are correlating their strategies, even though they are actually choosing strategies independently of each other. The example also shows that the conditions of Kalai and Lehrer (1993)–a grain of truth or, a fortiori, absolute continuity–do not guarantee the convergence to Nash equilibrium if a correlation is allowed on players' beliefs. An alternative approach would be to consider a convergence to correlated equilibrium and see what conditions are needed for the convergence.[7] Though interesting, this would be better left for future research.

Acknowledgments

The author is grateful to In-Koo Cho for his encouragement and comments.

References

Fudenberg, D. and D. Kreps, 1988. A Theory of Learning, Experimentation and Equilibrium in Games. mimeo, Stanford University.

Fudenberg, D. and D. Levine, 1993. Self-confirming Equilibria. Econometrica 61, 523-45.

Fudenberg, D. and J. Tirole, 1995. Game Theory. MIT Press.

Jordan, J., 1995. Bayesian Learning in Repeated Games. Games and Economic Behavior 9, 8-20.

Kalai, E. and E. Lehrer, 1993. Rational Learning Leads to Nash Equilibrium. Econometrica 61, 1019-45.

[7] For doing so, we may first have to obtain an appropriate definition of *correlated equilibrium in repeated games*, which has yet to be done to the best of this author's knowledge.