Learning Nash Equilibria

Darong Dai

Department of Economics, School of Business, Nanjing University

3. May 2012

Learning Nash Equilibria

Darong Dai[1]

Department of Economics, School of Business, Nanjing University, Nanjing 210093, P. R. China

Abstract

In the paper, we re-investigate the long run behavior of an adaptive learning process driven by the stochastic replicator dynamics developed by Fudenberg and Harris (1992). It is demonstrated that the Nash equilibrium will be the robust limit of the adaptive learning process as long as it is *reachable* for the learning dynamics in almost surely finite time. Doob's martingale theory and Girsanov Theorem play very important roles in confirming the required assertion.

Keywords: Stochastic replicator dynamics; Adaptive learning; Nash equilibria; Global convergence; Robustness.

JEL classification: C72; C73.

[1] Corresponding author. E-mail: daidarong998@163.com.

1. INTRODUCTION

The major goal of the present exploration is to study the limiting characteristic of an adaptive learning process driven by the stochastic replicator dynamics pioneered by Fudenberg and Harris (1992), extended by Cabrales (2000), Imhof (2005), Hofbauer and Imhof (2009), and among others. First, it is argued that the adaptive learning process indeed provides us with a local-martingale process due to the Girsanov Theorem under certain conditions. That is, the adaptive learning process exhibits local-martingale property through transforming the original probability measure into a new equivalent probability measure under relatively weak assumptions, i.e., the well-known Novikov condition is fulfilled. And this naturally leads us to the employment and application of Doob's martingale theory. Second, it is proved that any given Nash equilibrium will be learned in the long run by the players equipped with adaptive learning mechanism as long as it is *reachable* for such kind of replicator dynamics in almost surely finite time. And it is believed that this assertion rather throws new insights into existing convergence studies on stochastic learning dynamics. Last but not least, the global convergence is ensured in the sense of uniform topology and the corresponding robustness is also demonstrated under reasonable assumptions.

1.1. *Related Literatures*

Many existing literatures are encouraged to study the limiting characteristic of the learning process depending on different requirements. For example, some literatures (see, Canning, 1992; Young, 1993; Kandori et al., 1993) proved the similar convergence essentially requiring that the errors or perturbations approach zeros. Specifically, Canning (1992) shows that, under certain regularity conditions, the stationary distribution of the perturbed process converges to a stationary distribution of the unperturbed one. Kandori et al. (1993) show that the stochastic evolutionary learning process will converge to a Nash equilibrium when the mistake probability is small. In his seminal paper, Young (1993) shows that the adaptive dynamics defined by random sampling will converge almost surely to a pure strategy Nash equilibrium when the likelihood of mistakes goes to zero. On the contrary, the present model emphasizes the importance of evolutionary drift (see, Binmore and Samuelson, 1999) and the stochastic perturbations can be arbitrarily different from zeros except that they are controlled in certain regions due to the Girsanov Theorem. And also, the corresponding limit can be either a pure strategy Nash equilibrium or a mixed strategy Nash equilibrium.

Moreover, it is broadly known that mixed equilibria might interpreted as the limits of some learning processes arising from fictitious play with randomly perturbed payoffs in the manner of Harsanyi's (1973) purification theorem (e.g., Fudenberg and Kreps, 1993; Kaniovski and Young, 1995; Benaïm and Hirsch, 1999; Ellison and Fudenberg, 2000; Hofbauer and Hopkins, 2005, and among others). Nonetheless, there exist some crucial problems that prevent the corresponding convergence (see, Jordan, 1993). Benaïm and Hirsch's (1999) study reveals that there are robust parameter values giving probability zero of convergence for Jordan's 3×2 matching game. Both Shapley (1964) and Gaunersdorfer and Hofbauer (1995) provide examples in which the stochastic fictitious play as well as the standard fictitious play fail to converge. However, it is worth emphasizing that the present assertion can be easily extended to include asymmetric games between heterogeneous groups of populations and the corresponding global convergence indeed does not depend on the choice of payoff structures of the games.

Rather, the basic idea behind the present framework is in line with the argument of Gale et al. (1995), Binmore et al. (1995), Börgers and Sarin (1997), Cabrales (2000) and Beggs (2002) that the adaptive or trial-and-error learning process can be reasonably approximated by replicator dynamics. And in existing studies, Imhof (2005) proves that if the population is in a state sufficiently near to a strict Nash equilibrium, then, with probability close to 1, that equilibrium will be actually selected by the adaptive learning process driven by the stochastic replicator dynamics in the long run. Theorem 4.1 of Imhof (2008) shows that the expected time average distance between the stochastic replicator dynamics and a Nash equilibrium may be small provided that the payoff matrix is strictly negative definite. And Imhof (2008) provides us with a sufficient condition, which strictly depends on the specification of the payoff matrix, under which the stochastic replicator dynamics converge to a strict Nash equilibrium almost surely. Moreover, Hofbauer and Imhof (2009) provides us with an averaging principle, i.e., the time average of the learning dynamics will converge to the unique interior Nash equilibrium of the randomly perturbed game almost surely and under certain assumptions. To summarize, one major innovation, when compared with the above explorations, is that a unified framework is supplied such that either pure strategy Nash equilibrium or mixed strategy Nash equilibrium can be selected in the long run under much weaker assumptions than that of existing literatures.

1.2. *Outlines*

The rest of the paper is organized as follows. Section 2 presents the adaptive learning dynamics driven by the stochastic replicator dynamics and also some necessary assumptions are supplied. Section 3 demonstrates the global convergence of the learning dynamics and the corresponding robustness is confirmed under relatively weak assumptions. There is a brief concluding section. All proofs, unless otherwise noted in the text, appear in the Appendix.

2. THE MODEL

Here, and throughout the current investigation, we will focus on an asymmetric two-player game, which is canonical in evolutionary game theory, with n pure strategies, and we denote the corresponding payoff matrix by $A=(a_{ij})_{n\times n}$. Naturally, for any two sampled players, they enjoy the same strategy space and for each of the two players, a_{ij} represents her payoff from using strategy i if her opponent employs strategy j. As usual, we will study the evolutionary game by employing the replicator dynamics, which approximately describes the law of motion of the proportions of strategies over any given population. And the population is assumed to be sufficiently large with every member programmed to play one of her pure strategies in each period. Now, let $Z_i(t)$ denote the number of $i-$strategy players at time t, and let $X_i(t) = Z_i(t) \big/ \sum_{j=1}^{n} Z_j(t)$ denote the corresponding proportion. Thus, $\{AX(t)\}_i$ will be the expected payoff to those individuals playing strategy i under the random matching mechanism and with $X(t) = (X_1(t),...,X_n(t))^T$. In what follows, following Fudenberg and Harris (1992), Imhof (2005), and Hofbauer and Imhof (2009), we will incorporate random perturbations, which is modeled by independent Gaussian white noises with intensities σ_1^2, … ,σ_n^2, into the game payoffs. Thus, we put,

$$dZ_i(t) = Z_i(t)\left[\{AX(t)\}_i\, dt + \sigma_i dW_i(t)\right], \ \ i = 1,2,...,n. \tag{1}$$

where,

$$X(t) = (X_1(t),...,X_n(t))^T = \frac{1}{\sum_{i=1}^{n} Z_i(t)} (Z_1(t),...,Z_n(t))^T .$$

and also $(W_1(t),...,W_n(t))^T = W(t)$ denotes an $n-$dimensional Brownian motion defined on the underlying stochastic basis $(\Omega, \mathcal{F}_T, \mathbb{P})$ for $\forall T > 0$. Hence, by applying the Itô's rule to (1), the evolution of the population state $X(t)$ is driven by the following stochastic replicator dynamics,

$$dX(t) = b(X(t))dt + C(X(t))dW(t), \ \ X(0) = x_0 . \tag{2}$$

where,

$$b(x) = \left[diag(x_1,...,x_n) - xx^T\right]\left[A - diag(\sigma_1^2,...,\sigma_n^2)\right]x .$$

and,

$$C(x) = \left[diag(x_1,...,x_n) - xx^T\right] diag(\sigma_1,...,\sigma_n) .$$

for $x \in \Delta = \left\{ y \in [0,1]^n ; \sum_{j=1}^{n} y_j = 1 \right\}$, and $x_0 \triangleq (x_{10},...,x_{n0})^T \in \text{int}(\Delta)$.

ASSUMPTION 1: *The existence and uniqueness of the* (*strong*) *solution to the SDE given by* (2) *are ensured.*

We rewrite (2) as follows,

$$dX(t) = diag(X(t))\left(F(X(t))dt + \Sigma(X(t))dW(t)\right). \tag{3}$$

Then, we give,

ASSUMPTION 2: *To ensure that* Δ *is invariant, it is supposed that for each* $x \in \Delta$, *the drift vector* $diag(x)F(x)$ *and the columns* $S^1(x),...,S^n(x)$ *of the diffusion coefficient matrix* $S(x) \triangleq diag(x)\Sigma(x)$ *are elements of the tangent space* $T\Delta = \left\{u \in [0,1]^n; \sum_{j=1}^{n} u_j = 0\right\}$ *of* Δ.

Moreover, inspired by the well-known Girsanov Theorem, we directly give,

ASSUMPTION 3: *Here, and throughout the paper, suppose that there is an equivalent probability measure* $\mathbb{Q}$ *on* $\mathcal{F}_T$ *such that* $X(t)$ *defines a local-martingale w. r. t.* $\mathbb{Q}$.

3. GLOBAL CONVERGENCE

In the present section, we are encouraged to study the convergence property of the adaptive learning process driven by the stochastic replicator dynamics. Rather, we establish the following theorem,

THEOREM 1 (Global Convergence): *Based upon the above constructions and assumptions, it is demonstrated that for any given Nash equilibrium, denoted* $x^* \triangleq (x_1^*,...,x_n^*)^T$, *the adaptive learning process will strongly converge to* x^* *a.s. in the sense of uniform topology, that is,*

$$\lim_{T\to\infty} \sup_{0\le t\le T} \left\| X(t) - x^* \right\|_2 = 0 \quad a.s.$$

if we are provided that $\tau^*(\omega) \triangleq \inf\left\{t \ge 0; X(t) = x^*\right\} < \infty$ *a.s.*.

PROOF: It follows from Assumption 3 that $X_i(t) - x_i^*$ defines a local-martingale w. r. t. $\mathbb{Q}$. And by Assumption 2,

$$\tau_N(\omega) \triangleq \inf\left\{t \ge 0; X_i(t) = N\right\} \nearrow \infty \quad \text{a.s. as} \quad N \nearrow \infty. \tag{4}$$

So, $X_i(t\wedge\tau_N)-x_i^*$ defines a martingale w. r. t. $\mathbb{Q}$. Applying Doob's Martingale Inequality implies that,

$$\mathbb{Q}\left(\sup_{0\le t\le T}\left|X_i(t\wedge\tau_N)-x_i^*\right|\ge\frac{\varepsilon}{n}\right)\le\frac{n}{\varepsilon}\mathbb{E}_{\mathbb{Q}}\left[\left|X_i(T\wedge\tau_N)-x_i^*\right|\right],\quad \forall\varepsilon>0,\forall T>0. \qquad (5)$$

Based upon the assumption $\tau^*(\omega)\triangleq\inf\left\{t\ge0;X(t)=x^*\right\}<\infty$ a.s., put,

$$B_\beta\left(\tau^*(\omega)\right)\triangleq\left\{\tau(\omega)\ge0;\left|\tau(\omega)-\tau^*(\omega)\right|\le\beta\right\}.$$

for $\forall\beta>0$. Without loss of any generality, set up,

$$B_\beta\left(\tau^*(\omega)\right)\triangleq B_{2^{-k}}\left(\tau^*(\omega)\right)=\left\{\tau(\omega)\ge0;\left|\tau(\omega)-\tau^*(\omega)\right|\le2^{-k}\right\}.$$

Thus, according to Doob's Optional Sampling Theorem, Assumption 2, the continuity of martingale w. r. t. time t for any given $\omega\in\Omega$, and based on the Lebesgue Dominated Convergence Theorem, we obtain for $\forall\,\tau^k(\omega)\in B_{2^{-k}}\left(\tau^*(\omega)\right)$,

$$\begin{aligned}&\limsup_{k\to\infty}\mathbb{Q}\left(\sup_{0\le t\le\tau^k(\omega)}\left|X_i(t\wedge\tau_N)-x_i^*\right|\ge\frac{\varepsilon}{n}\right)\\&\le\frac{n}{\varepsilon}\limsup_{k\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left|X_i(\tau^k(\omega)\wedge\tau_N)-x_i^*\right|\right]\\&=\frac{n}{\varepsilon}\mathbb{E}_{\mathbb{Q}}\left[\left|X_i(\tau^*(\omega)\wedge\tau_N)-x_i^*\right|\right],\quad\forall\varepsilon>0.\end{aligned}$$

by (5). It follows from Fatou's Lemma that,

$$\mathbb{Q}\left(\sup_{0\le t\le\tau^*(\omega)}\left|X_i(t\wedge\tau_N)-x_i^*\right|\ge\frac{\varepsilon}{n}\right)\le\frac{n}{\varepsilon}\mathbb{E}_{\mathbb{Q}}\left[\left|X_i(\tau^*(\omega)\wedge\tau_N)-x_i^*\right|\right],\quad\forall\varepsilon>0.$$

i.e.,

$$\begin{aligned}&\limsup_{N\to\infty}\mathbb{Q}\left(\sup_{0\le t\le\tau^*(\omega)}\left|X_i(t\wedge\tau_N)-x_i^*\right|\ge\frac{\varepsilon}{n}\right)\\&\le\frac{n}{\varepsilon}\limsup_{N\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left|X_i(\tau^*(\omega)\wedge\tau_N)-x_i^*\right|\right]=0,\quad\forall\varepsilon>0.\end{aligned}$$

by (4) and Lebesgue Dominated Convergence Theorem again. Moreover, applying Fatou's Lemma again shows,

$$\mathbb{Q}\left(\sup_{0\le t\le\tau^*(\omega)}\left|X_i(t)-x_i^*\right|\ge\frac{\varepsilon}{n}\right)=0,\quad\forall\varepsilon>0.$$

by (4). This gives rise to,

$$\mathbb{Q}\left(\sup_{0\le t\le\tau^*(\omega)}\left|X_i(t)-x_i^*\right|<\frac{\varepsilon}{n}\right)=1,\quad \forall\varepsilon>0.$$

which yields,

$$\sup_{0\le t\le\tau^*(\omega)}\left|X_i(t)-x_i^*\right|<\frac{\varepsilon}{n}\ \text{ a.s. }\ \forall\varepsilon>0. \tag{6}$$

Now, we define the supremum norm,

$$\left\|X(t)-x^*\right\|_\infty \triangleq \max_{i\in\{1,\ldots,n\}}\left|X_i(t)-x_i^*\right|.$$

Thus, one may easily obtain,

$$\begin{aligned}
&\sup_{0\le t\le\tau^*(\omega)}\left\|X(t)-x^*\right\|_2\\
&\le n\sup_{0\le t\le\tau^*(\omega)}\left\|X(t)-x^*\right\|_\infty\\
&= n\sup_{0\le t\le\tau^*(\omega)}\max_{i\in\{1,\ldots,n\}}\left|X_i(t)-x_i^*\right|\\
&= n\sup_{0\le t\le\tau^*(\omega)}\left|X_{\hat{i}}(t)-x_{\hat{i}}^*\right|\\
&\le n\frac{\varepsilon}{n}=\varepsilon,\quad \mathbb{Q}-\text{a.s.}
\end{aligned}$$

by (6) and for $\forall\varepsilon>0$. Notice the arbitrariness of ε, we get,

$$\lim_{\tau^*(\omega)\to\infty}\sup_{0\le t\le\tau^*(\omega)}\left\|X(t)-x^*\right\|_2=0,\quad \mathbb{Q}-\text{a.s.}$$

as required. And hence the proof is complete. ■

REMARK 3.1: (i) It is easily seen that the proof of Theorem 1 essentially depends on the assumption that $\tau^*(\omega)\triangleq\inf\left\{t\ge 0; X(t)=x^*\right\}<\infty$ a.s. by noting that the local-martingale condition is a natural property of the adaptive learning process thanks to Girsanov Theorem. Therefore, the economic intuition of Theorem 1 can be concluded as follows: one Nash equilibrium can be learned by the players in the long run as long as the equilibrium is *reachable* for the given learning dynamics in almost surely finite time.

(ii) Moreover, it is especially worth emphasizing that if x^* stands for a pure strategy Nash equilibrium, then we do have $\tau^*(\omega)\triangleq\inf\left\{t\ge 0; X(t)=x^*\right\}<\infty$ a.s., which is demonstrated by Theorem 4.3 of Imhof (2005). That is to say, in this case, Theorem 1 is a natural conclusion internally implied by the current learning dynamics.

3.1. *Robustness*

Note that (3) can be rewritten as follows,

$$dX(t) = X(t) \circ F(X(t))dt + X(t) \circ \Sigma(X(t))dW(t). \tag{7}$$

where $\circ$ denotes the Hadamard product. Now, we introduce,

$$d\tilde{X}(t) = \tilde{X}(t) \circ \tilde{F}(\tilde{X}(t))dt + \tilde{X}(t) \circ \tilde{\Sigma}(\tilde{X}(t))dW(t). \tag{8}$$

where we have used,

ASSUMPTION 4: *For any* $\delta > 0$, *we suppose that,*

$$\sup_{x\in\Delta,\tilde{x}\in\tilde{\Delta}} \left\| F(x) - \tilde{F}(\tilde{x}) \right\|_2 \vee \sup_{x\in\Delta,\tilde{x}\in\tilde{\Delta}} \left\| \Sigma(x) - \tilde{\Sigma}(\tilde{x}) \right\|_2 \le \delta .$$

i.e., (8) *defines a* $\delta -$*perturbation of* (7), *and* $\tilde{\Delta} = \left\{ \tilde{y} \in [0,1]^n ; \sum_{j=1}^{n} \tilde{y}_j = 1 \right\}$.

Moreover, we put,

ASSUMPTION 5: *The existence and uniqueness of the* (*strong*) *solution to the SDE given by* (8) *are ensured throughout.*

ASSUMPTION 6: *To ensure that* $\tilde{\Delta}$ *is invariant, it is supposed that for each* $\tilde{x} \in \tilde{\Delta}$, *the drift vector* $diag(\tilde{x})\tilde{F}(\tilde{x})$ *and the columns* $\tilde{S}^1(\tilde{x}),...,\tilde{S}^n(\tilde{x})$ *of the diffusion coefficient matrix* $\tilde{S}(\tilde{x}) \triangleq diag(\tilde{x})\tilde{\Sigma}(\tilde{x})$ *are elements of the tangent space* $T\tilde{\Delta} = \left\{ \tilde{u} \in [0,1]^n ; \sum_{j=1}^{n} \tilde{u}_j = 0 \right\}$ *of* $\tilde{\Delta}$.

ASSUMPTION 7: *Suppose that there exists a constant* $K < \infty$, *sufficiently large, such that,*

$$\sup_{x\in\Delta} \left\| F(x) \right\|_2^2 \vee \sup_{x\in\Delta} \left\| \Sigma(x) \right\|_2^2 \le K .$$

Accordingly, the following proposition is established,

PROPOSITION 1: *Provided the above constructions and assumptions, and if* $X(0) = \tilde{X}(0) = x_0$ *, then we have,*

$$\mathbb{E}_{\mathbb{Q}}\left[\lim_{T\to\infty}\sup_{0\le t\le T}\left\|X(t)-\tilde{X}(t)\right\|_2^2\right]\to 0 \quad as \quad \delta\to 0 .$$

PROOF: In the Appendix. ■

Now, naturally, the following corollary is derived,

COROLLARY 1 (Robust Convergence): *It is clearly asserted that* $\tilde{X}(t)$ *converges to the Nash equilibrium* x^* *a.s. and in the sense of uniform topology*.

PROOF: Combining Theorem 1 with Proposition 1 easily confirms Corollary 1. And hence the proof is omitted. ■

4. CONCLUSION

In the current investigation, our major goal is to analyze the limiting behavior of the adaptive learning process driven by the stochastic replicator dynamics pioneered by Fudenberg and Harris (1992). First, a local-martingale process is implicitly and naturally implied by such kind of learning process under certain weak conditions thanks to the Girsanov Theorem. So, the corresponding martingale theory developed by Doob can be employed to prove the convergence assertion. Second, the main result of the paper asserts that one Nash equilibrium can be learned by the players equipped with local-martingale learning mechanism in the long run as long as the Nash equilibrium is *reachable* for the learning process in almost surely finite time. Finally, it is proved that the global convergence conclusion exhibits robustness under relatively weak assumptions.

APPENDIX: Proof of Proposition 1

It follows from Assumption 2 and 6 that for $\forall\ 2<p<\infty$, $\forall T>0$, we have,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t)\right\|_2^p\right]\vee\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|\tilde{X}(t)\right\|_2^p\right]\le 1, \tag{A.1}$$

where by Assumption 1 and 5,

$$X(t) = x_0 + \int_0^t X(s) \circ F\left(X(s)\right) ds + \int_0^t X(s) \circ \Sigma\left(X(s)\right) dW(s).$$

$$\tilde{X}(t) = x_0 + \int_0^t \tilde{X}(s) \circ \tilde{F}\left(\tilde{X}(s)\right) ds + \int_0^t \tilde{X}(s) \circ \tilde{\Sigma}\left(\tilde{X}(s)\right) dW(s).$$

Moreover, suppose that $\|X(t)\|_2 \vee \|\tilde{X}(t)\|_2 \le E$ for $\forall t \ge 0$ and $E < \infty$. Indeed, one just need let $E \ge 1$. In what follows, we first define the following stopping times,

$$\tau_E \triangleq \inf\left\{t \ge 0; \|X(t)\|_2 \ge E\right\},\ \ \tilde{\tau}_E \triangleq \inf\left\{t \ge 0; \|\tilde{X}(t)\|_2 \ge E\right\},\ \ \tau_E^0 \triangleq \tau_E \wedge \tilde{\tau}_E.$$

By the Young Inequality (see, Higham et al., 2003) and for any $S > 0$,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t) - \tilde{X}(t)\right\|_2^2\right]$$

$$= \mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t) - \tilde{X}(t)\right\|_2^2 1_{\{\tau_E > T, \tilde{\tau}_E > T\}}\right] + \mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t) - \tilde{X}(t)\right\|_2^2 1_{\{\tau_E \le T, or \tilde{\tau}_E \le T\}}\right]$$

$$\le \mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X\left(t\wedge\tau_E^0\right) - \tilde{X}\left(t\wedge\tau_E^0\right)\right\|_2^2 1_{\left\{\tau_E^0 > T\right\}}\right] + \frac{2S}{p}\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t) - \tilde{X}(t)\right\|_2^p\right]$$

$$+\frac{1-\frac{2}{p}}{S^{\frac{2}{p-2}}}\mathbb{Q}\left(\tau_E \le T, or \tilde{\tau}_E \le T\right), \qquad \text{(A.2)}$$

It follows from (A.1) that,

$$\mathbb{Q}\left(\tau_E \le T\right) = \mathbb{E}_{\mathbb{Q}}\left[1_{\{\tau_E \le T\}}\frac{\left\|X\left(\tau_E\right)\right\|_2^p}{E^p}\right] \le \frac{1}{E^p}\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t)\right\|_2^p\right] \le \frac{1}{E^p}.$$

Similarly, $\mathbb{Q}\left(\tilde{\tau}_E \le T\right) \le 1/E^p$. So,

$$\mathbb{Q}\left(\tau_E \le T, or \tilde{\tau}_E \le T\right) \le \mathbb{Q}\left(\tau_E \le T\right) + \mathbb{Q}\left(\tilde{\tau}_E \le T\right) \le \frac{2}{E^p}.$$

Moreover, we obtain by (A.1),

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t) - \tilde{X}(t)\right\|_2^p\right] \le 2^{p-1}\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left(\left\|X(t)\right\|_2^p + \left\|\tilde{X}(t)\right\|_2^p\right)\right] \le 2^p.$$

Hence, (A.2) becomes,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t) - \tilde{X}(t)\right\|_2^2\right]$$

$$\leq \mathbb{E}_{\mathbb{Q}}\left[\sup_{0\leq t\leq T}\left\|X\left(t\wedge\tau_E^0\right)-\tilde{X}\left(t\wedge\tau_E^0\right)\right\|_2^2\right]+\frac{2^{p+1}S}{p}+\frac{2(p-2)}{pS^{\frac{2}{p-2}}E^p}. \tag{A.3}$$

By the Cauchy-Schwarz Inequality, we get,

$$\left\|X\left(t\wedge\tau_E^0\right)-\tilde{X}\left(t\wedge\tau_E^0\right)\right\|_2^2$$

$$=\left\|\int_0^{t\wedge\tau_E^0}\left[X(s)\circ F\left(X(s)\right)-\tilde{X}(s)\circ\tilde{F}\left(\tilde{X}(s)\right)\right]ds+\right.$$

$$\left.\int_0^{t\wedge\tau_E^0}\left[X(s)\circ\Sigma\left(X(s)\right)-\tilde{X}(s)\circ\tilde{\Sigma}\left(\tilde{X}(s)\right)\right]dW(s)\right\|_2^2$$

$$\leq 2\left\{T\int_0^{t\wedge\tau_E^0}\left\|X(s)\circ F\left(X(s)\right)-\tilde{X}(s)\circ\tilde{F}\left(\tilde{X}(s)\right)\right\|_2^2 ds+\right.$$

$$\left.\left\|\int_0^{t\wedge\tau_E^0}\left[X(s)\circ\Sigma\left(X(s)\right)-\tilde{X}(s)\circ\tilde{\Sigma}\left(\tilde{X}(s)\right)\right]dW(s)\right\|_2^2\right\}.$$

Taking expectations on both sides, and using Itô's Isometry, we have for any $\tau\leq T$,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\leq t\leq\tau}\left\|X\left(t\wedge\tau_E^0\right)-\tilde{X}\left(t\wedge\tau_E^0\right)\right\|_2^2\right]$$

$$\leq 4\left\{T\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|X(s)\circ F\left(X(s)\right)-\tilde{X}(s)\circ F\left(X(s)\right)\right\|_2^2 ds\right]+\right.$$

$$T\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|\tilde{X}(s)\circ F\left(X(s)\right)-\tilde{X}(s)\circ\tilde{F}\left(\tilde{X}(s)\right)\right\|_2^2 ds\right]+$$

$$\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|X(s)\circ\Sigma\left(X(s)\right)-\tilde{X}(s)\circ\Sigma\left(X(s)\right)\right\|_2^2 ds\right]+$$

$$\left.\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|\tilde{X}(s)\circ\Sigma\left(X(s)\right)-\tilde{X}(s)\circ\tilde{\Sigma}\left(\tilde{X}(s)\right)\right\|_2^2 ds\right]\right\}$$

$$\leq 4\left\{TK\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|X(s)-\tilde{X}(s)\right\|_2^2 ds\right]+T\delta^2\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|\tilde{X}(s)\right\|_2^2 ds\right]+\right.$$

$$\left.K\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|X(s)-\tilde{X}(s)\right\|_2^2 ds\right]+\delta^2\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|\tilde{X}(s)\right\|_2^2 ds\right]\right\}$$

$$\le 4\left\{(T+1)K\mathbb{E}_{\mathbb{Q}}\left[\int_0^{t\wedge\tau_E^0}\left\|X(s)-\tilde{X}(s)\right\|_2^2 ds\right]+T(T+1)\delta^2\right\}$$

$$\le 4(T+1)K\int_0^{\tau}\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t_0\le s}\left\|X\left(t_0\wedge\tau_E^0\right)-\tilde{X}\left(t_0\wedge\tau_E^0\right)\right\|_2^2\right]ds+4T(T+1)\delta^2 .$$

where we have used Assumption 2, 4, 6 and 7. Hence, applying Gronwall's Inequality (see, Higham et al., 2003) implies that,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le\tau}\left\|X\left(t\wedge\tau_E^0\right)-\tilde{X}\left(t\wedge\tau_E^0\right)\right\|_2^2\right]\le 4T(T+1)\exp\left\{4(T+1)K\right\}\delta^2 .$$

Inserting this into (A.3) leads us to,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t)-\tilde{X}(t)\right\|_2^2\right]\le 4T(T+1)\exp\left\{4(T+1)K\right\}\delta^2+\frac{2^{p+1}S}{p}+\frac{2(p-2)}{pS^{\frac{2}{p-2}}E^p} .$$

Hence, for $\forall\varepsilon>0$, we can shoose S and E such that,

$$\frac{2^{p+1}S}{p}\le\frac{\varepsilon}{3}\quad\text{and}\quad\frac{2(p-2)}{pS^{\frac{2}{p-2}}E^p}\le\frac{\varepsilon}{3} .$$

And for any given $T>0$, we put δ such that,

$$4T(T+1)\exp\left\{4(T+1)K\right\}\delta^2\le\frac{\varepsilon}{3} .$$

Thus, for $\forall\varepsilon>0$, we obtain,

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\le t\le T}\left\|X(t)-\tilde{X}(t)\right\|_2^2\right]\le\frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\varepsilon .$$

Notice the arbitrariness of ε, and employ the well-known Levi Lemma gives the desired result. And this completes the whole proof. ■

REFERENCES

Beggs, A., 2002. Stochastic Evolution with Slow Learning. *Economic Theory* **19**, 379-405.

Benaïm, M. and M. W. Hirsch, 1999. Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games. *Games and Economic Behavior* **29**, 36-72.

Binmore, K., L. Samuelson and R. Vaughan, 1995. Musical Chairs: Modeling Noisy Evolution. *Games and Economic Behavior* **11**, 1-35.

Binmore, K. and L. Samuelson, 1999. Evolutionary Drift and Equilibrium Selection. *Review of Economic Studies* **66**, 363-393.

Börgers, T. and R. Sarin, 1997. Learning Through Reinforcement and Replicator Dynamics. *Journal of Economic Theory* **77**, 1-14.

Cabrales, A., 2000. Stochastic Replicator Dynamics. *International Economic Review* **41**, 451-481.

Canning, D., 1992. Average Behavior in Learning Models. *Journal of Economic Theory* **57**, 442-472.

Ellison, G. and D. Fudenberg, 2000. Learning Purified Mixed Equilibria. *Journal of Economic Theory* **90**, 84-115.

Fudenberg, D. and D. Kreps, 1993. Learning Mixed Equilibria. *Games and Economic Behavior* **5**, 320-367.

Fudenberg, D. and C. Harris, 1992. Evolutionary Dynamics with Aggregate Shocks. *Journal of Economic Theory* **57**, 420-441.

Gale, J., K. Binmore and L. Samuelson, 1995. Learning to be Imperfect: The Ultimatum Game. *Games and Economic Behavior* **8**, 56-90.

Gaunersdorfer, A. and J. Hofbauer, 1995. Fictitious Play, Shapley Polygons, and the Replicator Equation. *Games and Economic Behavior* **11**, 279-303.

Harsanyi, J., 1973. Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points. *International Journal of Game Theory* **2**, 1-23.

Higham, D. J., X. R. Mao and A. M. Stuart, 2003. Strong Convergence of Euler-Type Methods for Nonlinear Stochastic Differential Equations. *SIAM Journal on Numerical Analysis* **40**, 1041-1063.

Hofbauer, J. and E. Hopkins, 2005. Learning in Perturbed Asymmetric Games. *Games and Economic Behavior* **52**, 133-152.

Hofbauer, J. and L. A. Imhof, 2009. Time Averages, Recurrence and Transience in the Stochastic Replicator Dynamics. *Annals of Applied Probability* **19**, 1347-1368.

Imhof, L. A., 2005. The Long-Run Behavior of the Stochastic Replicator Dynamics. *Annals of Applied Probability* **15**, 1019-1045.

Imhof, L. A., 2008. Multiple-Trial Conflicts and Stochastic Evolutionary Game Dynamics. *Advances in Applied Probability* **40**, 1174-1197.

Jordan, J., 1993. Three Problems in Learning Mixed-Strategy Nash Equilibria. *Games and Economic Behavior* **5**, 368-386.

Kandori, M., G. Mailath and R. Rob, 1993. Learning, Mutation, and Long-run Equilibria in Games. *Econometrica* **61**, 29-56.

Kaniovski, Y. M. and H. P. Young, 1995. Learning Dynamics in Games with Stochastic Perturbations. *Games and Economic Behavior* **11**, 330-363.

Shapley, L. S., 1964. Some Topics in Two-Person Games. In *Advances in Game Theory*, ed. by M. Drescher, L. S. Shapley, and A. W Tucker. Annals of Mathematics Study 52. Princeton: Princeton University Press, 1-28.

Young, P., 1993. The Evolution of Conventions. *Econometrica* **61**, 57-84.