帝京社会学第17号（2004年３月）

模倣ダイナミクスと試行錯誤ダイナミクス

Imitation Dynamics and Trial and Error Dynamics

大　浦　宏　邦
（Hirokuni Oura）

進化ゲーム理論は動学化されたゲーム理論で、遺伝ダイナミクスモデルと学習ダイナミクスモデルの二階建ての構造を持っている（石原、金井2002）。

学習ダイナミクスは、プレーヤーが戦略を何らかの方法で修正することによって生じるダイナミクスで、比較的短いタイムスケールの社会現象をモデル化するのに向いている。本論文では学習ダイナミクスのうち、模倣ダイナミクスと試行錯誤ダイナミクスについて、やや詳しく見てみることにしよう。

１　遺伝ダイナミクスと学習ダイナミクス

まず、遺伝ダイナミクスと学習ダイナミクスの違いについて確認する。

遺伝ダイナミクスモデルの代表的なものはレプリケーターダイナミクスモデルだが、これは遺伝や教育で戦略が親から子へ継承される場合に生じるダイナミクスである。このモデルではプレーヤーは生涯に渡って戦略を変えることはない。親は子供を残して死んでいくが、この出生死滅過程が

ダイナミクスを駆動するマイクロプロセスとなる。

これに対し、プレーヤーが途中で戦略を変えることを仮定するダイナミクスを考えることもできる。このタイプのダイナミクスを学習ダイナミクスと総称する。現実の社会ではプレーヤーは様々な方法で戦略を変えている。ある戦略で失敗すれば、それを変えるかもしれないし、他人がよさそうな戦略をとっていれば、それをまねするかも知れない。いろいろ考えた結果、現状では最善と判断する戦略をとることもあるだろう。

このように、戦略を変える手続きにはさまざまなものが考えられるので、学習ダイナミクスにも色々なタイプのものが考えられる。具体的には、試行錯誤ダイナミクス、模倣ダイナミクス、最適反応ダイナミクスなどが考えられている。

以下ではまず、レプリケーターダイナミクスと学習ダイナミクスの特徴をおおまかに比較してから、試行錯誤ダイナミクスと模倣ダイナミクスについて詳しく見ていくことにしよう。

１．１　ゲームの繰り返しと戦略修正

学習型のモデルの最大の特徴は、プレーヤーが戦略を変えることが想定される点である。プレーヤーはゲームを何度か繰り返して経験するうちに、戦略を変えたり変えなかったりする。戦略の変え方には、自分の手や他人の手を参照したり、自分の利得や他人の利得を参照したり、深く考えたり余り考えなかったり様々な方法がありうるが、いずれにせよ何らかの手続きに従ってプレーヤーは戦略を修正していく。この戦略を修正する手続きを戦略修正アルゴリズムという。学習ダイナミクスは、各プレーヤーが何らかの戦略修正アルゴリズムに従って戦略を変化させるときに生じるダイナミクスである。

戦略が修正されるということは、ゲームが繰り返して複数回行われることを意味している。逆にいうとゲームが１回切りしか行われない場合には学習ダイナミクスは生じない。ちなみに従来型のゲーム理論にも「繰り返しゲーム」というカテゴリーがあるが、従来の「繰り返しゲーム」は動学

的なプロセスを扱うものではない。プレーヤーは事前に「繰り返しゲーム」の中でどう振舞うかについての行動プラン（この行動のプランをスーパー戦略という）を持っていると想定され、この行動プラン（スーパー戦略）同士の間でのNash均衡を考えるという静学的なアプローチが取られてきた。「繰り返しゲーム」という用語はこの種の静学的な分析法を採るときの用語なので、ここでは「繰り返しゲーム」とはいわず、「繰り返しのあるゲーム」「ゲームの繰り返し」などという言い方を使うことにしよう。

１．２　現象のタイムスケール

学習ダイナミクスは、ゲームが複数回繰り返して行われる場合にプレーヤーが戦略を修正することによって生じるダイナミクスである。ダイナミクスの生じるタイムスケールは、ゲームが繰り返される頻度や、戦略の修正が起きる頻度に依存する。ゲームによっては一日に何回かプレーする機会がある場合もあるだろうし、一年に一回とか数年に一回しかプレーする機会がない場合もあるかもしれない。しかし、いずれの場合もプレーヤーが生きているうちに数回以上はゲームを行なう機会と戦略を変更する機会があると考えられるので、長くても１世代以内のタイムスケールで生じるダイナミクスであるといえる。実時間にして数年から数十年程度のタイムスケールであろう。

これに対し、レプリケーターダイナミクスの場合は出生死滅過程によってダイナミクスが駆動する。この場合に想定されるタイムスケールは数十世代から数百世代で、実時間にして数百年から数千年以上のタイムスケールになる。これらより、学習ダイナミクスによる分析は、短期的な現象を分析する場合に有用であり、レプリケーターダイナミクスによる分析は中長期的な現象を分析する場合に有用であるといえる。

生物学的な現象の場合は、数千年から数万年というタイムスケールで起こることも多いが、社会科学の分野では数年から数十年程度の現象がほとんどであろう。したがって、多くの社会科学的な現象の分析には学習ダイナミクスが用いられることになる。しかし、社会科学的な現象でも社会の

起源や協力の進化といった問題にかかわる現象は、長いタイムスケールで起きる現象なのでレプリケーターダイナミクスによる分析が有用であり必要となる。

以下では、具体的な学習ダイナミクスとして、模倣ダイナミクスと試行錯誤ダイナミクスについて見てみよう。

2　模倣ダイナミクス

模倣ダイナミクスは、プレーヤーが他者の戦略をまねることによって生じるダイナミクスである。ここでは２人２戦略の対称ゲーム（自分と相手の立場を入れ替えても戦略や利得が変わらないゲーム）の場合について模倣ダイナミクスのモデルを紹介する。

2．1　模倣と同調

模倣とは、簡単にいうと他のプレーヤーの戦略をまねすることである。他人の服装や仕草をまねして取り入れることは良くあるし、ものの考え方や生き方をまねるということもあるだろう。

模倣と良く似た現象に同調がある。他者をまねるという点で同調と模倣は共通している。相違点は、同調が多くの人が行っている行動を「みんなしてるから」という理由でまねるのに対し、模倣では他者の行動などを「いいな」と思ってまねる点にある。同調では周りと同じかどうかが大事であって、その行動そのものが良いかどうかはあまり問題にならないのに対し、模倣ではその行動が良いか悪いかが問題で、良ければまねするし悪ければまねしない（その結果相手と違う行動となってもかまわない）点が異なっているといえる。

もちろん実際の行動はこれらの要素が混在しているが、理念型としては模倣と同調を上のように区別することができる。同調によるダイナミクスは「多い方の戦略が増える」という単純なものとなりやすい。以下では、他人の行動などを「いいな」と思ってまねをする場合のダイナミクスを考

えることにする。

２．２　仮定すべき事柄

模倣によってプレーヤーの戦略は変化し、それにつれて集団中の戦略シェアも変化していく。このプロセスをモデル化する上で、考慮すべき点を確認しておこう。

まず、一口に「他人の行動」をまねするといっても、具体的に誰の行動をまねすると想定すればよいのであろうか。模倣対象についての仮定が必要である。次に「いいな」と思う場合の「良い悪い」を判断する基準はなんなのであろうか。評価基準についての仮定が必要である。さらに、他人の何をまねするのか（模倣事項)、どういうときにまねするのか（模倣タイミング）についての仮定も必要であろう。

模倣対象については、集団中からランダムに選ぶ（たまたま道で会った人を対象とするなど）と想定することもできるし、羽振りのよさそうな人を対象にすると想定することもできる。

良い悪いの評価基準としては、効用関数を想定することが一般的である。効用の高い行動を良い、低い行動を悪いと評価するのである。ただしこの場合、模倣しようとする人の効用関数か、模倣される人の効用関数か、どちらが妥当かということが問題になる。誰かが気にいって着ている服でも、他の人にはちっとも良くないということがありうる。この場合、良くないと思っている人はその服装をまねしないであろう。したがって、評価の基準は「模倣しようとする人の効用関数」とすることが妥当である。

他人の何をまねするのか（模倣事項）については経験的な研究が必要であるが、次の事柄が指摘できるであろう。まず、他人から見えにくい事柄より見えやすい事柄の方が模倣の対象となりやすいと考えられる。例えば下着や寝巻きよりも上着やコートの方が模倣されやすいし、流行の対象にもなりやすいであろう。また、複雑な事柄よりも単純な事柄の方が模倣されやすいであろう。複雑な事柄が模倣されるときには、単純化されて模倣される可能性も考えられる。その意味で、混合戦略（複数の戦略を確率的

に組み合わせる）やスーパー戦略（相手の出方に応じた行動プラン）は模倣されにくく、模倣される場合もそのままではなく純粋戦略として模倣されることが多いと予想される。ここでは、単純な純粋戦略が模倣の対象となる場合を考える。

模倣のタイミングについては、すべてのプレーヤーが同じ頻度で他人の行動を参照しようとすると想定することもできるし、特に現状に不満をもっている人が高い頻度で他人の行動をまねしようとすることもあるであろう。この点については、両方のタイプのモデルを立てることができる。

２．３　街角模倣モデル

ここではまず、もっとも単純なモデルとして次の事柄を仮定するモデルを考えよう。

［仮定］

・どのプレーヤーも同じ頻度で戦略の見直しをする。
・効用の比較対象は集団中からランダムに選ぶ。
・各プレーヤーは微小時間dtの間にｒdtの確率で戦略の見直しをする。
・プレーヤー集団全体の人数はＮ人で一定。

効用の比較対象を集団中からランダムに選ぶということは、街角で偶然出会った人の戦略をまねるかまねないか意思決定するようなモデルなので、このタイプのモデルを街角模倣モデルと呼ぶ。

各プレーヤーはdtの間にｒdtの確率で戦略の見直しをするのでdtの間にｒNdt人のプレーヤーが戦略の見直しを行う（Ｎは十分に大きいものとする）。

見直しを行うプレーヤーは集団中からランダムに一人、参照相手となるプレーヤーを選び出す。参照相手が自分よりも自分の効用関数でみて良い場合に、自分の戦略を参照相手の戦略に変更するものとする。ただし、以下では効用関数はどのプレーヤーについても共通であると仮定する。この

ようなマイクロプロセスを仮定するときに、集団全体ではどのようなマクロダイナミクスが生じるのであろうか。いくつかの具体的なゲームについて見てみよう。

［例］　調整ゲーム

次の調整ゲームの場合について考えよう。

自分＼相手	A	B
A	2	0
B	0	1

戦略Aのプレーヤーの割合をｘ、戦略Bのプレーヤーの割合を１－ｘとする。プレーヤーはランダムに出会ってゲームを行うとすると、最近のプレーで効用２を獲得したプレーヤーの割合はx^2、効用１を獲得したプレーヤーの割合は$(1-x)^2$、効用０を獲得したプレーヤーの割合は２ｘ(１－ｘ)となる。

微小時間dtの間にｒNdt人のプレーヤーが戦略の見直しを行うが、そのうち

x^2ｒNdt人が　　最近の効用が２
$(1-x)^2$ｒNdt人が　最近の効用が１
２ｘ(１－ｘ)ｒNdt人が 最近の効用が０

となる。ここで各プレーヤーは、参照相手の最近のゲーム結果が分かり、それと自分の最近のゲーム結果を比較して戦略を変更するかしないかを決めることとしよう。

このとき、戦略Aのプレーヤーが戦略Bに変更するのは、戦略Aで効用０のプレーヤーが、戦略Bで効用１のプレーヤーを参照相手に選んだ場合である。このときAからBへの「流出」が発生する。

見直しプレーヤーｒNdt人のうち

・戦略Aで効用０のものはｘ(１－ｘ)ｒNdt人

・これらのプレーヤーが戦略Ｂで効用１のプレーヤーを
参照相手に選ぶ確率は　$(1-x)^2$

である。したがって

$$x(1-x)rNdt\cdot(1-x)^2\text{人}$$

がdtの間にＡからＢに流出する人数となる。

一方、戦略Ｂから戦略Ａに変更するのは、戦略がＢで効用が０や１だったプレーヤーが戦略Ａで効用が２だったプレーヤーを参照する場合である。この場合、戦略Ａへの流入が発生する。

戦略Ｂのプレーヤーは全員効用が０か１なので、$rNdt$人の見直しプレーヤーのうち、

$$(1-x)rNdt\cdot x^2\text{人}$$

のプレーヤーが、dtの間に戦略をＢからＡに変更することになる。これがＡへの「流入」である。

ここで戦略Ａの人数は、Ａへの流入人数からＡからの流出人数を引いた人数だけ増減するはずである。したがって、

Ａの人数の変化＝Ａへの流入－Ａからの流出

となる。上の考察から

$$\text{Ａへの流入}=(1-x)rNdt\cdot x^2$$

$$\text{Ａへの流出}=x(1-x)rNdt\cdot(1-x)^2$$

なので

$$\text{Ａの人数の変化}=(1-x)rNdt\cdot x^2-x(1-x)rNdt\cdot(1-x)^2$$

である。

また、戦略Ａのシェアxの変化をdxとすると

$$dx=\text{Ａの人数の変化}/N$$

である。したがって

$$\text{Ａの人数の変化}=Ndx$$

となる。これを上の式に代入すると

$$Ndx=(1-x)rNdt\cdot x^2-x(1-x)rNdt\cdot(1-x)^2$$

両辺をNで割って

$$dx=(1-x)r\,dt\cdot x^2-x(1-x)r\,dt\cdot(1-x)^2$$

両辺をdtで割って

$$dx/dt=(1-x)r\cdot x^2-x(1-x)r\cdot(1-x)^2$$
$$=rx(1-x)(x-(1-x)^2)$$
$$=rx(x-1)(x^2-3x+1)$$

という微分方程式が得られる。これが街角模倣モデルのダイナミクス方程式である。

このダイナミクスを分析してみよう。２次方程式の解の公式を用いると

$$x^2-3x+1=(x-\alpha)(x-\beta)$$

ただし　$\alpha=(3-\sqrt{5})/2$　（約0.38）

$\beta=(3+\sqrt{5})/2$　（約2.62）

と因数分解できる。これよりダイナミクスのベクトル図は図１のようになる。

したがって、

$x=0$、$x=1$　が　漸近安定点

$x=(3-\sqrt{5})/2$　（約0.38）が　不安定定常点

であることがわかる。

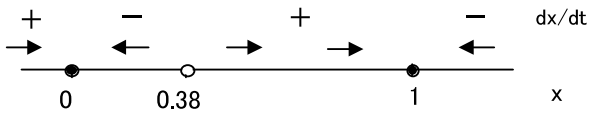

図１　街角モデルのダイナミクス

ちなみに、利得表の値を効用ではなく適応度の増減とみなしてレプリケーターダイナミクスを求めると

$$dx/dt=x(1-x)(3x-1)$$

となり

$x=0$，1　が　漸近安定点

$x=1/3$　　が　不安定定常点

となる。

この場合、レプリケーターダイナミクスと模倣ダイナミクスはおおむね同じようなダイナミクスとなる。不安定定常点（ダイナミクスの分水嶺にあたる）の位置は若干異なるが、漸近安定状態については両者は一致している。

２．４　不満・羽振りモデル

街角模倣モデルでは、見直しの頻度はすべてのプレーヤーで同じで、模倣対象者は集団からランダムに選ばれた。しかし、戦略の見直しは効用の低い人が頻繁に行い、効用の高い人はあまり行わないかもしれない。プレーヤーが自らの効用の低さに不満をもち、頻繁に他者の戦略を参照しようとすることを仮定するモデルを不満モデルと呼ぶことにする。

このとき、参照者も集団からランダムに等確率で選ばれるのではなく、「羽振り」の良い人がそうでない人よりも高い確率で選ばれるかもしれない。この場合、プレーヤーの効用が観察可能な経済的・社会的な「羽振り」と関連していることを仮定する必要があるが、なんらかのメカニズムで羽振りの良いプレーヤーが高い確率で参照されることを仮定するモデルを羽振りモデルと呼ぶことにする。

以下では、不満による模倣と羽振りのよさによる参照の両方を仮定する不満・羽振りモデルを考えることにしよう。

［仮定の追加］

以下のモデルでは、街角モデルに次の仮定を追加する。

- 最近の効用がuであったプレーヤーは、微小時間dtの間に$r(u)dt$の確率で戦略の見直しを行う。
- 最近の効用がu_1であったプレーヤーは、効用がu_2であったプレーヤーより$q(u_1)/q(u_2)$倍参照されやすい。
- $r(u)$はuについての減少関数。$q(u)$はuについての増加関数。

まず、見直しの仮定である。最近のプレーで効用がuであったプレーヤーは、微小時間dtの間にr(u)dtの確率で戦略の見直しを行うものとする。r(u)はuについての減少関数だが、負にはならないものとしよう。このように仮定すれば効用が低いプレーヤーほど頻繁に見直しを行うことになる。

次に参照についての仮定である。最近のプレーで効用がu_1であったプレーヤーは、効用がu_2であったプレーヤーより$q(u_1)/q(u_2)$倍参照されやすいと仮定する。ここでq(u)はuについての増加関数で、常に正の値をとるものとする。このように仮定すると効用の高いプレーヤーほど参照相手として選ばれやすいことになる。

たとえば、仮にq(1)＝1、q(2)＝2、だとするならば、効用が2のプレーヤーは効用が1のプレーヤーよりも2倍の確率で他のプレーヤーに参照されることになる。もし、すべてのプレーヤーの効用が1か2で、効用1のプレーヤーの割合がx_1、効用2のプレーヤーの割合がx_2ならば、誰か一人参照相手のプレーヤーを選んだときにそのプレーヤーの効用が1である確率は

$$x_1/(x_1+2x_2)$$

となる。また参照プレーヤーの効用が2である確率は

$$2x_2/(x_1+2x_2)$$

となる。

一般には、効用u_iのプレーヤーの割合がx_iで、参照されやすさが$q(u_i)$に比例するとき、一人選んだ参照プレーヤーの効用がu_iである確率は

$$q(u_i)x_i/\Sigma_{[j]}q(u_j)x_j$$

となる。

［例］　調整ゲーム

再び次の調整ゲームについて考えよう。不満・羽振りモデルの場合、どのようなダイナミクスが生じるのであろうか。

自分＼相手	A	B
A	2	0
B	0	1

まず、戦略Aのシェアをｘ、戦略Bのシェアを１－ｘとし、ランダムマッチングでゲームが行われているものとすると、最近のプレーで

効用２のプレーヤーの割合　x^2

効用１のプレーヤーの割合　$(1-x)^2$

効用０のプレーヤーの割合　$2x(1-x)$

となる。

微小時間dtの間に効用２のプレーヤーが見直しをする確率は$r(2)dt$なので、集団全体Ｎ人のうちでは$x^2Nr(2)dt$人の効用２のプレーヤーが戦略の見直しを行うことになる（Ｎは十分に大きいものとする）。効用１や効用０のプレーヤーについても同様である。

ここで、dtの間に戦略をＡからＢに変更する人数を求めることにしよう。ＡからＢに変更するのは、戦略がＡで効用が０であったプレーヤーが、戦略がＢで効用が１であったプレーヤーを参照する場合である。

dtの間に戦略Ａで効用０のプレーヤーのうち

$x(1-x)Nr(0)dt$人

が戦略の見直しを行う。これらのプレーヤーは集団から参照相手のプレーヤーを一人選び出すが、そのプレーヤーが戦略Ｂで効用１である確率は

$q(1)(1-x)^2/(2q(0)x(1-x)+q(1)(1-x)^2+q(2)x^2)$

である。以下、分母をＱと書くことにすれば、確率は

$q(1)(1-x)^2/Q$

となる。これより、戦略Ａ効用０で戦略Ｂ効用１のプレーヤーを参照して戦略をＢに変更する人数は

$x(1-x)Nr(0)dt\cdot q(1)(1-x)^2/Q$人

となる。これが、dtの間にＡからＢに流出する人数である。

次に、ＢからＡに流入する人数を考える。戦略Ｂから戦略Ａに変更する

ケースは、戦略Ｂで効用０のプレーヤーが戦略Ａで効用２のプレーヤーを参照する場合と、戦略Ｂで効用１のプレーヤーが戦略Ａで効用２のプレーヤーを参照する場合の２通りある。

流出の場合と同様に、それぞれの人数を求めると、前者のケースでの流入は

$$x(1-x)N\,r(0)dt\cdot q(2)x^2/Q\text{人}$$

であり、後者のケースでの流入は

$$(1-x)^2N\,r(1)dt\cdot q(2)x^2/Q\text{人}$$

となる。これらの合計が微小時間dtの間に戦略Ａに流入する人数となる。

流入と流出の人数がわかれば、街角モデルと同様にＡの割合ｘについての微分方程式を立てることが出来る。

（Ａの増減）＝（Ａへの流入）－（Ａからの流出）

であるから

$$\begin{aligned}Ndx=&\,x(1-x)N\,r(0)dt\cdot q(2)x^2/Q\\&+(1-x)^2N\,r(1)dt\cdot q(2)x^2/Q\\&-x(1-x)N\,r(0)dt\cdot q(1)(1-x)^2/Q\end{aligned}$$

両辺をNdtで割ると

$$\begin{aligned}dx/dt=&(x(1-x)\,r(0)\cdot q(2)x^2+(1-x)^2\,r(1)\cdot q(2)x^2\\&-x(1-x)\,r(0)\cdot q(1)(1-x)^2)/Q\\=&\,x(1-x)(r(0)\cdot q(2)x^2+(1-x)\,r(1)\cdot q(2)x\\&-r(0)\cdot q(1)(1-x)^2)/Q\\=&\,x(1-x)((r(0)q(2)-r(0)q(1)-r(1)q(2))x^2\\&+(2\,r(0)q(1)+r(1)q(2))x-r(0)q(1))/Q\end{aligned}$$

となる。これが不満・羽振りモデルにおける模倣ダイナミクス方程式である。

これは、どのようなダイナミクスになるのであろうか。このままでは、分かりにくいので仮に

$$r(0)=0.5\quad r(1)=0.3\quad r(2)=0.1$$
$$q(0)=1\quad q(1)=3\quad q(2)=5$$

として計算してみると

$$dx/dt = x(x-1)(x^2+9x-3)/2Q$$

となる。

この場合のベクトル図は図２のようになり、

$x = 0, 1$　が　漸近安定点

$x = \alpha$　　が　不安定定常点

ただし $\alpha = (-9+\sqrt{93})/2$　（約0.32）

となる。街角モデルの場合と不満・羽振りモデルでは約0.38、不満・羽振りモデルでは約0.32）が、定性的にはほぼ同じダイナミクスとなることが分かる。

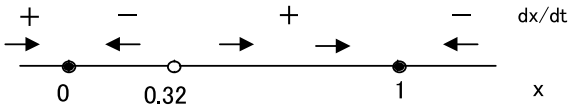

図２　不満・羽振りモデルのダイナミクス

２．５　２×２ゲームの模倣ダイナミクス

街角モデル、不満・羽振りモデルのいずれでも調整ゲームの漸近安定状態は$x = 0$と$x = 1$であった。調整ゲームは（A，A）と（B，B）がstrict Nash均衡となるゲームだが、この結果から模倣ダイナミクスではstrict Nash均衡に対応する状態が漸近安定になることが予想される。ここでは、一般の２人２戦略対称ゲーム（２×２対称ゲーム）の模倣ダイナミクスをたてることでこの予想を確認しておこう。

自分＼相手	A	B
A	a	b
B	c	d

という２×２ゲームを考える。モデルの仮定は基本的に不満・羽振りモデルと同じだが、さらに次の仮定を付け加える。

［追加の仮定］

・効用u_1のプレーヤーが効用u_2のプレーヤーを参照したときに、
戦略を効用u_2の戦略に変更する確率を$\phi(u_2-u_1)$とする。
（ただし、ϕは０から１までの値を取るu_2-u_1に関する増加関数）

先ほどまでのモデルでは、参照相手の効用が高ければ相手の戦略を採用し、低ければ採用しないことを仮定していたので、戦略変更確率$\phi(u_2-u_1)$が

$u_2>u_1$　のとき　$\phi(u_2-u_1)=1$

$u_2<u_1$　のとき　$\phi(u_2-u_1)=0$

となる場合を考えていたことになる。

これをもう少し一般化して、相手との差が小さいときは戦略をまねするかどうか微妙だが、相手の効用の方がずっと良いときには高い確率でまねをすることを想定したのが、ここで追加をした仮定である。参照相手の効用を見極める上である程度の誤差がある場合や、相手のまねをすることにある程度のためらいがある場合には、この仮定のようになると考えられる。

以下、このような場合の模倣ダイナミクスを考えてみよう。

戦略Ａのシェアをｘ、戦略Ｂのシェアをｙとし、ランダムマッチングでゲームが行われているものとすると、最近のプレーで

効用ａのプレーヤーの割合　x^2

効用ｂのプレーヤーの割合　xy

効用ｃのプレーヤーの割合　xy

効用ｄのプレーヤーの割合　y^2

となる。

微小時間dtの間の見直しで戦略Ａから戦略Ｂへの変更が生じるのは、

１）効用ａのプレーヤーが効用ｃのプレーヤーを参照した場合

2）効用ａのプレーヤーが効用ｄのプレーヤーを参照した場合

3）効用ｂのプレーヤーが効用ｃのプレーヤーを参照した場合

4）効用ｂのプレーヤーが効用ｄのプレーヤーを参照した場合

の４通りである。

1）のケースで変更を行なう人数は全体の人数Nが十分に多いとして

1）　$r(a)x^2 \cdot q(c)xy/Q \cdot \phi(c-a)N$人

である。(ただし$Q=q(a)x^2+q(b)xy+q(c)xy+q(d)y^2$)

同様に、2）3）4）のケースで戦略をAからBに変更する人数はそれぞれ

2）　$r(a)x^2 \cdot q(d)y^2/Q \cdot \phi(d-a)N$人

3）　$r(b)xy \cdot q(c)xy/Q \cdot \phi(c-b)N$人

4）　$r(b)xy \cdot q(d)y^2/Q \cdot \phi(d-b)N$人

である。これらの合計が、dtの間のAからの流出人数である。

次にAへの流入人数を考える。dtの間にAへの流入が生じるのは

5）効用ｃのプレーヤーが効用ａのプレーヤーを参照した場合

6）効用ｃのプレーヤーが効用ｂのプレーヤーを参照した場合

7）効用ｄのプレーヤーが効用ａのプレーヤーを参照した場合

8）効用ｄのプレーヤーが効用ｂのプレーヤーを参照した場合

の４通りである。

流出の場合と同様に計算すると、それぞれの場合の流入人数は

5）　$r(c)xy \cdot q(a)x^2/Q \cdot \phi(a-c)N$人

6）　$r(c)xy \cdot q(b)xy/Q \cdot \phi(b-c)N$人

7）　$r(d)y^2 \cdot q(a)x^2/Q \cdot \phi(a-d)N$人

8）　$r(d)y^2 \cdot q(b)xy/Q \cdot \phi(b-d)N$人

となる。これらの合計がdtの間のAへの流入人数である。

以上を

（Aの増減)＝(Aへの流入)－(Aからの流出)

に代入して、両辺をNdtで割って整理すれば模倣ダイナミクス方程式が得られる。煩雑になるので途中の計算は省略して結果だけ示せば

$$dx/dt = x\,y\,f(x,y)/Q \quad \text{（式１）}$$

ただし、

$$\begin{aligned} f(x,y) = {} & x^2(r(c)q(a)\phi(a-c) - r(a)q(c)\phi(c-a)) \\ & + x\,y(r(c)q(b)\phi(b-c) - r(b)q(c)\phi(c-b)) \\ & + x\,y(r(d)q(a)\phi(a-d) - r(a)q(d)\phi(d-a)) \\ & + y^2(r(d)q(b)\phi(b-d) - r(b)q(d)\phi(d-b)) \end{aligned}$$

となる。これがこの場合の模倣ダイナミクス方程式である。

これは

$u_2 > u_1$　のとき　$\phi(u_2 - u_1) = 1$

$u_2 < u_1$　のとき　$\phi(u_2 - u_1) = 0$

とすると、先の不満・羽振りダイナミクスに一致するし、$r(u)$、$q(u)$を一定とすると始めの街角モデルに一致する。その意味では、このダイナミクスを一般模倣ダイナミクスと呼ぶことができる。

２．６　頂点の安定性

$f(x,y)$が煩雑であるが、式１の形自体は簡単である。この式からまず$x = 0$と$y = 0$の時は模倣ダイナミクスは定常であることが分かる。つまり、状態空間の両端（頂点）は定常である。

では、頂点の安定性はどうであろうか。これについては次の命題が成立する。

【命題１】　２×２対称ゲームの一般模倣ダイナミクスでは、戦略プロファイル（Ｉ、Ｉ）がstrict Nash均衡ならば、頂点Ｉは漸近安定である。

【証明】

戦略プロファイル（A、A）がstrict Nash均衡であるとする。このとき$a > c$である。

式１の$f(x,y)$に$x = 1$、$y = 0$を代入すると

$$f(1,0)=r(c)q(a)\phi(a-c)-r(a)q(c)\phi(c-a)$$

である。

ｒは負ではない減少関数、ｑは負ではない増加関数なので、ａ＞ｃのとき

$$r(c)>r(a)\geqq 0$$

$$q(a)>q(c)\geqq 0$$

である。

また、ａ＞ｃより

$$2a>2c$$

$$a-c>c-a$$

であり、さらにφは負ではない増加関数なので

$$\phi(a-c)>\phi(c-a)\geqq 0$$

がいえる。

これらより、

$$r(c)q(a)\phi(a-c)>r(a)q(c)\phi(c-a)$$

なので

$$f(1,0)>0$$

である。

ｆ(x,y) は連続な関数なので、ｘ＝１、ｙ＝０に十分近いｘ、ｙの範囲では

$$f(x,y)>0$$

が成立する。

これより、ｘ＝１からｘが少し小さくなってもそのずれが十分に小さければ

$$dx/dt>0$$

となり、ｘは１に収束する。したがって頂点Ａに対応するｘ＝１は漸近安定である。

戦略プロファイル（Ｂ、Ｂ）がstrict Nash均衡の場合も同様に証明できる。

【証明終わり】

この命題は、strict Nash均衡に対応する頂点が漸近安定であるという、先ほどの予測を確認するものである。これは、一般模倣ダイナミクスに関する命題なので、街角モデル、不満・羽振りモデルを含む広いクラスの模倣ダイナミクスについて成立する。

命題１に関連して、次の命題が成立する。

【命題２】　２×２対称ゲームの一般模倣ダイナミクスでは、戦略プロファイル（Ｉ、Ｉ）がNash均衡でなければ、頂点Ｉは不安定である。

【証明】

戦略プロファイル（Ａ、Ａ）がNash均衡でないとする。このときａ＜ｃである。

命題１と同様に

$$f(1,0)=r(c)q(a)\phi(a-c)-r(a)q(c)\phi(c-a)$$

であるが、ａ＜ｃのときは

$$r(a)>r(c)\geqq 0$$

$$q(c)>q(a)\geqq 0$$

$$\phi(c-a)>\phi(a-c)\geqq 0$$

である。したがって

$$f(1,0)<0$$

ｆ(x,y)は連続な関数なので、ｘ＝１、ｙ＝０に十分近いｘ、ｙの範囲では

$$f(x,y)<0$$

が成立する。

これよりｘ＝１は不安定である。

戦略プロファイル（Ｂ、Ｂ）がNash均衡でない場合も同様に証明できる。

【証明終わり】

この頂点の安定、不安定に関する命題はレプリケーターダイナミクスの場合と共通している。この点で模倣ダイナミクスはレプリケーターダイナミクスと良く似たダイナミクスである。

例えば、囚人のジレンマゲームでは、（C、C）はNash均衡ではなく（D、D）がstrict Nash均衡となるため、一般模倣ダイナミクスでは頂点Dが漸近安定となる。

一方、チキンゲームでは、（A、A）も（B、B）もNash均衡ではないので、頂点Aも頂点Bもどちらも不安定である。この場合は内点に漸近安定点を持つことになる。この漸近安定点の位置はｒやｑやφの値に依存するので、レプリケーターダイナミクスの内点漸近安定点に一致する保障はない。この点で、模倣ダイナミクスはレプリケーターダイナミクスと異なっているといえる。

3　試行錯誤ダイナミクス

この節では、プレーヤーが試行錯誤で戦略を変更する場合を考えてみよう。試行錯誤による学習は、他の方法による学習に比べて必要とする情報量が少ないので、情報が少ない状況で進行する学習をモデル化するのに適している。ここでは、試行錯誤学習を仮定するモデルの代表としてロスとエレブのモデル（Roth & Erev 1995、Erev & Roth 1998）を紹介する。

3．1　試行錯誤による学習

食事に行くときに「あの店はおいしかったのでまた行こう」とか「あそこはいまいちだったからやめておこう」といった形で意思決定をすることは、日常よく経験する事柄である。このように、複数の選択肢があるときに、とりあえず何か試してみて結果の良かった選択肢を高い確率で採用し、結果の悪かった選択肢を避けるタイプの学習を試行錯誤という。模倣と並んで広く用いられている学習の方略である。心理学の分野では強化学習と呼ばれることもある。

このアルゴリズムで学習をする場合、プレーヤーに必要とされる事柄は、行動の選択肢を幾つかもっていることと、行動の結果を評価する評価基準を持っていることの二つである。ある選択肢を採ったときにどうなるかを事前に知っておく必要はないし、他のプレーヤーの戦略を知る必要もない。また、他のプレーヤーの行動結果について知る必要もない。もちろん、これらの情報があればさらに効率よく学習を行なうことができるが、試行錯誤学習はこれらの情報が手に入らないときにも使用することのできる、汎用性の高い学習アルゴリズムということが出来る。

やや余談であるが、学習という現象は植物よりも動物において一般的にみられる。動物は移動して危険を避けつつえさを採るという生活様式を採用しているが、危険のありかもえさのありかも常に同じ場所とは限らず、時々刻々変化するのが普通である。この状況で生活するには、危険やえさのありかに応じて行動を変えることができたほうが有利である。学習という現象は、このような必要性を背景として進化してきたと考えられる。

試行錯誤学習は学習のなかでももっとも基本的な学習で、良い結果（えさがあった、危険を避けられた、など）をもたらす行動を強化し、悪い結果（えさがなかった、危険な目にあったなど）をもたらす行動を抑制することが原型となっている。もうすこし具体的には、神経系の中に結果の良し悪しを判定する部門があって、「良い」結果の場合にはその行動をもたらした神経細胞の結合（シナプス結合）を強化する信号を発し、「悪い」結果の場合には悪い結果をもたらしたシナプス結合を弱める信号を発して、神経結合が作り変えられていく。このようにして、「良い」結果をもたらす行動が発生しやすいように神経結合が作り変えられていくプロセスが試行錯誤学習で、人間に限らず動物が一般に行なっているタイプの学習である。

人間の場合は、「他人の経験から学ぶ」ことも多いが、「自分の経験から学ぶ」ことも依然として大きな位置を占めている。ここでは試行錯誤学習がどのようなダイナミクスをもたらすかを考えてみよう。

３．２　強化と忘却

ここでは、試行錯誤学習をモデル化する方法について考えてみよう。

まず試行錯誤の過程で変化していくものは何なのであろうか。この場合、表面的にはプレーヤーの行動が変化しているように見えるが、行動自体は変化していなくても、内心では「この店おいしくなくなってきたな」とか「あの携帯いいかも」といった評価の変化が生じている可能性がある。この内心の変化が大きくなると、実際の行動も変化すると考えられる。ロスとエレブはこの「内心の変化」を差分方程式の形で表現するモデルを考案した。

このモデルでは、プレーヤーはある戦略を採ろうとする「傾向」(propensity）を持つものと考える。この「傾向」が大きい戦略は高い確率で採用されるが、「傾向」の低い戦略は低い確率でしか採用されない。上の例でいう「内心の評価」がこの「傾向」に相当する。

プレーヤーがある戦略を実際に採ってみて、結果が良かったときには、その戦略に対する評価が上がり、その戦略を採ろうとする「傾向」が大きくなる。結果が悪かったときには、評価が下がり、その戦略を採ろうとする「傾向」が小さくなる。この効果は学習心理学の用語で強化(reinforcement）と呼ばれる。一般に高い利得は大きな強化をもたらし、低い利得や負の利得は小さな強化やマイナスの強化をもたらすことが知られている。

ただし、一度強化を受けても時間が経つと強化の効果は次第に失われていく。一度おいしいものを食べてもその印象は次第に薄れていくし、かつて嫌な体験をしてもその記憶は次第に薄れていくであろう。これが学習心理学でいう忘却の過程である。忘却についても様々な実験が行われているが、初期の忘却の速度は速く、時間が経つにつれて速度が遅くなることが一般的な傾向のようである。

この強化と忘却という過程を取り入れて、ある戦略を採ろうとする「傾向」の変化を次のようにモデル化しよう。$p_{ij}(t)$をプレーヤー i が時刻 t に戦略 j を採ろうとする傾向の大きさ、φを忘却の速さを表すパラメー

ター（ただし０＜φ＜１）、$R_{ij}(t)$を時刻ｔにｊに与えられる強化の大きさとして、時刻ｔ＋１における傾向の大きさ$p_{ij}(t+1)$が

$$p_{ij}(t+1)=(1-\phi)p_{ij}(t)+R_{ij}(t) \quad （式２）$$

となると考える。これがロス・エレブモデルの基本方程式である。

式２で忘却が無い場合を考えると、傾向$p_{ij}(t)$は強化$R_{ij}(t)$の分だけ増減する。また、強化が無い場合を考えると傾向$p_{ij}(t)$は毎期（１－φ）の割合で減少していく。このとき初期の減少は比較的すみやかで、時間が経つにつれて減少のしかたは次第に緩やかになっていく。このような性質は、学習心理学で得られた知見と少なくとも定性的に一致しているといえる。

３．３　傾向性と行動

次に、心理的傾向と実際の行動との関係を考えておこう。Ａを採りたくてＢを採りたくない人は大体Ａを採るであろうし、逆の人は大体Ｂを採るであろう。しかし、場合によってはＡも採りたいしＢも採りたいというケースや、ＡもＢも採りたくないというケースも考えられる。これらの場合、プレーヤーは迷ったあげく、ＡまたはＢをとる、あるいはどちらも採らないという行動をとるであろう。

どちらもとらないという選択肢がないものとすると、プレーヤーはＡまたはＢを確率的にとると考えるのが、妥当な推論となる。ＡまたはＢをとりたいと思う傾向が同程度ならば、ＡやＢをとる確率は等しいと考えられるし、どちらかを取りたい傾向がやや大きい場合は、そちらをとる確率がやや大きくなるであろう。この状況を近似的にモデル化するために、「プレーヤーは戦略ｉをとりたいと思う傾向に比例した確率で戦略ｉをとる」という仮定がおかれることが多い。

この仮定のもとで、傾向性と行動の関係がどうなるかを考えておこう。まず２戦略の場合を考える。時刻ｔでプレーヤーｉが戦略Ａをとりたいと思う傾向を$p_{ia}(t)$、プレーヤーｉが戦略Ｂをとりたいと思う傾向を$p_{ib}(t)$とするとき、プレーヤーｉがＡを取る確率$x_{ia}(t)$とＢを取る確率$x_{ib}(t)$はそれぞれ

$$x_{ia}(t) = p_{ia}(t)/(p_{ia}(t)+p_{ib}(t))$$
$$x_{ib}(t) = p_{ib}(t)/(p_{ia}(t)+p_{ib}(t)) \qquad (式3)$$

となる。$x_{ij}(t)$は$p_{ij}(t)$に比例しているし、$x_{ia}(t)+x_{ib}(t)=1$となり確率の要件を満たしている。

同様に、戦略が1からnまでn個ある場合は

$$x_{ij}(t) = p_{ij}(t)/\Sigma p_{ik}(t)$$

となる。やはり、$x_{ij}(t)$は$p_{ij}(t)$に比例しているし、$\Sigma x_{ij}(t)=1$となり確率の要件を満たしている。

ただしこの場合、$p_{ij}(t)<0$となる場合があると確率が負になってしまって都合が悪い。そこで式2で、強化$R_{ij}(t)$は正または0だと考えるか、負の強化がある場合は$p_{ij}(t)=0$となると、それ以上$p_{ij}(t)$は減らないことを仮定する必要がある。

3．4　傾向性の挙動

ここまでが、ロス・エレブモデルの基本的な枠組みである。この枠組みでダイナミクスを考えると、どのようなことが起こるのであろうか。傾向性パラメーターp_{ij}の挙動、戦略採用確率x_{ij}の挙動の順に調べていこう。

まず、傾向性パラメーターの挙動を大まかに考えてみよう。式2において、強化パラメーター$R_{ij}(t)$がもし一定の値だとすると、p_{ij}はどのようになるのであろうか。

このとき

$$p_{ij}(t+1) = (1-\phi)p_{ij}(t)+R_{ij}$$
$$= p_{ij}(t)+R_{ij}-\phi p_{ij}(t)$$

より

$$p_{ij}(t+1)-p_{ij}(t) = R_{ij}-\phi p_{ij}(t)$$

である。

これより、

$R_{ij} > \phi p_{ij}(t)$　のときp_{ij}は増加
$R_{ij} < \phi p_{ij}(t)$　のときp_{ij}は減少

することがわかる。

p_{ij}が定常となるのは、$R_{ij}=\phi p_{ij}(t)$、つまり、

$$p_{ij}(t)=R_{ij}/\phi \qquad (式4)$$

のときである。p_{ij}がこれより大きいとp_{ij}は減少、小さいと増加するので、式4のp_{ij}は漸近安定である。

これより、強化R_{ij}が一定の場合は傾向パラメーター$p_{ij}(t)$は強化の値を忘却パラメーターϕで割った値に収束することが分かる。

これが傾向性パラメーターの大まかな挙動である。強化値が大きいほど収束値は大きくなるし、忘却の割合が大きいほど収束値は小さくなる。強化値は実際には一定ではなく、ゲームの繰り返しごとに変化するはずであるが、強化値の平均が大きいと収束値が大きく、小さいと小さくなることが予想される。

3．5　採用確率の挙動

次に、プレーヤーiによる戦略jの採用確率x_{ij}の挙動を考えよう。ここでは、簡単のために戦略はAとBの二つの場合を考える。このとき

$$x_{ia}(t)=p_{ia}(t)/(p_{ia}(t)+p_{ib}(t))$$

である。2戦略では、$x_{ia}(t)$の挙動を考えれば十分なので以下ではこれを単に$x(t)$と表記する。また、傾向性についても添え字iを省略して書くことにすると

$$x(t)=p_{a}(t)/(p_{a}(t)+p_{b}(t))$$

となる。

ここで計算上の工夫として

$$k(t)=p_{a}(t)+p_{b}(t)$$

と置くことにする。このとき

$$x(t)=p_{a}(t)/k(t)$$

$$p_{a}(t)=x(t)k(t)$$

$$p_{b}(t)=(1-x(t))k(t)$$

となる。

さて、以下では時刻ｔと時刻ｔ＋１の間のｘの変化Δｘを考えることにしよう。

$$\Delta x = x(t+1) - x(t)$$
$$= p_a(t+1)/k(t+1) - x(t)$$
$$= ((1-x(t))p_a(t+1) - x(t)p_b(t+1))/k(t+1) \qquad (式5)$$

である。

ここで、

$$p_a(t+1) = (1-\phi)p_a(t) + R_a$$
$$= (1-\phi)x(t)k(t) + R_a$$
$$p_b(t+1) = (1-\phi)p_b(t) + R_b$$
$$= (1-\phi)(1-x(t))k(t) + R_b$$

なので、これを式５の分子に代入すると

$$式5の分子 = (1-x(t))((1-\phi)x(t)k(t) + R_a)$$
$$-x(t)((1-\phi)(1-x(t))k(t) + R_b)$$
$$= (1-x(t))(1-\phi)x(t)k(t) + (1-x(t))R_a$$
$$-x(t)(1-\phi)(1-x(t))k(t) - x(t)R_b$$
$$= (1-x(t))R_a - x(t)R_b$$

となる。

以上より

$$\Delta x = ((1-x(t))R_a - x(t)R_b)/k(t+1) \qquad (式6)$$

となることが分かる。式６を見れば戦略Ａの採用確率ｘ(t)の変化を知ることができる。

３．６　調整ゲームのダイナミクス

式６は具体的にはどのようなダイナミクスなのであろうか。次の調整ゲームを例にとって考えてみよう。

自分＼相手	A	B
A	3	1
B	1	2

ここで、利得表の数値はそれぞれの場合に戦略ＡやＢを採ろうとする傾向性に与えられる強化値を表しているものとする。例えば、自分が戦略Ａ、相手が戦略Ａを採ったときに自分が戦略Ａを採ろうとする傾向性に与えられる強化値R_aは３である。このとき、戦略Ｂは採用されていないので、戦略Ｂを採ろうとする傾向性に与えられる強化値R_bは０である。

このゲームを自分と相手の二人で繰り返してプレーする場合を考える。自分がＡを採る確率をｘ、相手がＡを採る確率をｙとすると、自分と相手の戦略の組み合わせ（戦略プロファイル）は

確率ｘｙで　（Ａ，Ａ）
確率ｘ(１－ｙ)で　（Ａ，Ｂ）
確率(１－ｘ)ｙで　（Ｂ，Ａ）
確率(１－ｘ)(１－ｙ)で　（Ｂ，Ｂ）

となる。これよりR_a、R_bの値は

確率ｘｙで　$R_a=3$、$R_b=0$
確率ｘ(１－ｙ)で　$R_a=1$、$R_b=0$
確率(１－ｘ)ｙで　$R_a=0$、$R_b=1$
確率(１－ｘ)(１－ｙ)で　$R_a=0$、$R_b=2$

である。

したがって、式６よりΔｘは

確率ｘｙで　$3(1-x)/((1-\phi)k(t)+3)$
確率ｘ(１－ｙ)で　$(1-x)/((1-\phi)k(t)+1)$
確率(１－ｘ)ｙで　$-x/((1-\phi)k(t)+1)$
確率(１－ｘ)(１－ｙ)で　$-2x/((1-\phi)k(t)+2)$

となる。

分母が少しずつ異なるが、ダイナミクスの概略を知るために大体同じ値

と考えてΔｘの期待値を求めると、次のようになる。ちなみに、φが十分小さな値のときには、式４より$k(t)=p_a(t)+p_b(t)$が大きな値となるので、分母が同じ値と考えても近似的には差し支えない。

このとき、分母を一定としたときのΔｘの期待値は

$$\Delta x\text{の期待値}=(xy\cdot 3(1-x)+x(1-y)\cdot(1-x)-(1-x)y\cdot x-(1-x)(1-y)\cdot 2x)/\text{分母}$$
$$=x(1-x)(3y+(1-y)-y-2(1-y))/\text{分母}$$
$$=x(1-x)(3y-1)/\text{分母}\qquad(式7)$$

という簡単な式になる。

同様な計算をΔｙについても行なうと

$$\Delta y\text{の期待値}=y(1-y)(3x-1)/\text{分母}\qquad(式8)$$

となる。この二つの式が調整ゲームを自分と相手の二人で繰り返し行なった場合の試行錯誤ダイナミクス方程式となる。

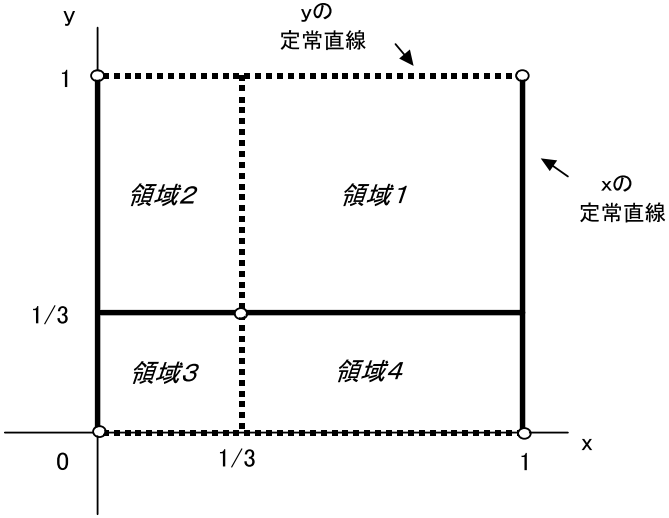

図３　試行錯誤ダイナミクスの定常直線と定常点

式７、式８から、ベクトル図を描いてみよう。このダイナミクスは自分がＡを採る確率ｘと相手がＡを採る確率ｙの二つの変数を持つので、状態空間は直線ではなく平面となる。またｘとｙの範囲は

$0 \leqq x \leqq 1$

$0 \leqq y \leqq 1$

であるので、状態空間は（０、０）、（０、１）、（１、０）、（１、１）を頂点とする正方形になる。

まず定常点を考えると、式７からΔｘの期待値＝０となるのは

$x = 0$、$x = 1$、$y = 1/3$

である。これを図３の太線で示す。また、式８からΔｙの期待値＝０となるのは

$y = 0$、$y = 1$、$x = 1/3$

である。これを図３の点線で示す。この太線と点線の交点がダイナミクスの定常点であるが、それは状態空間の４頂点と（1/3、1/3）の５点となる。

次にダイナミクスの方向を矢印で記入する。定常直線で状態空間は四つの領域に分割されるが、それぞれの領域で

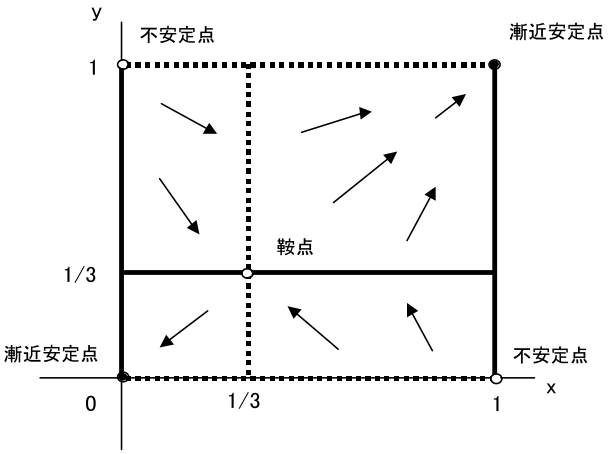

図４　試行錯誤ダイナミクスのベクトル図

	Δｘの期待値	Δｙの期待値
領域１	＋	＋
領域２	＋	－
領域３	－	－
領域４	－	＋

となるので、それぞれ右上、右下、左下、左上に矢印を記入すると図４のベクトル図が出来上がる。これが、二人調整ゲームの試行錯誤ダイナミクスを示すベクトル図である。

これより、

（０、０）、（１、１）が漸近安定点

（０、１）、（１、０）が不安定点

（1/3、1/3）が鞍点

となることが分かる。（１、１）は両者がＡ、（０、０）は両者がＢを採る状態で、調整ゲームのstrict Nash均衡である。したがって、試行錯誤ダイナミクスでも、strict Nash均衡が漸近安定となることが予想される。

３．７　２×２ゲームのダイナミクス

上の結果は、Δｘ やΔｙの期待値を求めることで得られた結果である。実際には、ｘやｙは確率的に変化するので期待値の方向にそのまま変化していくわけではなく、変化の方向にはある程度の散らばりがある。しかし、期待値の方向がダイナミクスの収束を示しているときには、変化の方向にある程度の散らばりがあっても、それが大きすぎない限り実際のダイナミクスもやはり収束する。したがって、Δｘ やΔｙの期待値を求めて分析する前節の方法は、ダイナミクスの収束が起きる場合には有用である。

ここでは、前節の方法を用いて一般の２×２対称ゲームのダイナミクスを考えてみよう。次のゲームを考える。ａ、ｂ、ｃ、ｄはそれぞれの場合の強化を表す正の数である。

自分＼相手	A	B
A	a	b
B	c	d

自分がAを採る確率をｘ、相手がAを採る確率をｙとする。このとき、自分がAやBを採ろうとする傾向パラメーターに与えられる強化値R_a、R_bはそれぞれ

確率ｘｙで　$R_a=a$、$R_b=0$
確率ｘ(１－ｙ)で　$R_a=b$、$R_b=0$
確率(１－ｘ)ｙで　$R_a=0$、$R_b=c$
確率(１－ｘ)(１－ｙ)で　$R_a=0$、$R_b=d$

となる。

このとき、Δｘは式６より

確率ｘｙで　$a(1-x)/((1-\phi)k(t)+a)$
確率ｘ(１－ｙ)で　$b(1-x)/((1-\phi)k(t)+b)$
確率(１－ｘ)ｙで　$-cx/((1-\phi)k(t)+c)$
確率(１－ｘ)(１－ｙ)で　$-dx/((1-\phi)k(t)+d)$

となる。

ここでも計算の簡略化のために、これらの分母を一定値とみなすと

$$\Delta x\text{の期待値}=(xy\cdot a(1-x)+x(1-y)\cdot b(1-x)-(1-x)y\cdot cx-(1-x)(1-y)\cdot dx)/\text{分母}$$
$$=x(1-x)(ay+b(1-y)-cy-d(1-y))/\text{分母}$$

となる。

ここでａｙ＋ｂ(１－ｙ)は自分が戦略Aを採ったときの期待効用u_{1a}に等しい。また、ｃｙ＋ｄ(１－ｙ)は自分が戦略Bを採ったときの期待効用u_{1b}である。これらを用いると

$$\Delta x\text{の期待値}=x(1-x)(u_{1a}-u_{1b})/\text{分母}\quad\text{（式９）}$$

同様にΔｙの期待値について計算すると

$$\Delta y\text{の期待値}=y(1-y)(u_{2a}-u_{2b})/\text{分母}\quad\text{（式10）}$$

となる。ただし、u_{2a}は相手がAを採ったときの相手の期待効用、u_{2b}は相手がBを採ったときの相手の期待効用である。

式９、式10が一般の２×２対称ゲームのダイナミクス方程式である。これより、試行錯誤ダイナミクスでは

１）　（０、０）、（０、１）、（１、０）、（０、０）の各頂点が定常点

２）　期待効用の大きい戦略の採用確率が増加することがわかる。

２）は分母が一定とみなした場合の近似的な結果であるが、分かりやすい性質である。

３．８　頂点の安定性

式９、式10から状態空間の頂点が定常であることが分かるが、頂点の安定性についてはどうであろうか。これについては、次の命題が成立する。

【命題３】　二人二戦略ゲームで戦略プロファイル（Ｉ、Ｊ）がstrict Nash均衡であれば、（Ｉ、Ｊ）は試行錯誤ダイナミクスの漸近安定状態である。

【証明】

戦略プロファイル（A、A）がstrict Nash均衡の場合を考えると、このときａ＞ｃである。

$$f_1(x、y)=u_{1a}-u_{1b}$$
$$f_2(x、y)=u_{2a}-u_{2b}$$

を考えると

$$f_1(1、1)=a-c>0$$
$$f_2(1、1)=a-c>0$$

である。f_1、f_2は連続関数なので、これよりｘ＝１、ｙ＝１の近傍に

$$f_1(x、y)>0$$
$$f_2(x、y)>0$$

となる領域が存在する。式９、式10より、この領域では

Δｘの期待値＞０

Δｙの期待値＞０

なのでｘ＝１、ｙ＝１、つまり戦略プロファイル（Ａ、Ａ）を採る状態は漸近安定である。（Ｂ、Ｂ）がstrict Nash均衡の場合も同様に証明できる。

次に戦略プロファイル（Ａ、Ｂ）がstrict Nash均衡の場合を考えると、このときｃ＞ａ、ｂ＞ｄである。これより

$f_1(1、0)=b-d>0$

$f_2(1、0)=a-c<0$

となる。f_1、f_2は連続関数なのでｘ＝１、ｙ＝０の近傍に

$f_1(x、y)>0$

$f_2(x、y)<0$

となる領域が存在する。この領域では

Δｘの期待値＞０

Δｙの期待値＜０

なのでｘ＝１、ｙ＝０、つまり戦略プロファイル（Ａ、Ｂ）を採る状態は漸近安定である。（Ｂ、Ａ）がstrict Nash均衡の場合も同様に証明できる。

【証明終わり】

この命題は、頂点に対応する戦略プロファイルがstrict Nash均衡であれば、その頂点が漸近安定であることを示している。逆に、頂点に対応する戦略プロファイルがNash均衡でなければ、次の命題が成立する。

【命題４】 二人二戦略ゲームで戦略プロファイル（Ｉ、Ｊ）がNash均衡でなければ、（Ｉ、Ｊ）は試行錯誤ダイナミクスで不安定である。

【証明略】

これらの命題は、静学解を吟味することで試行錯誤ダイナミクスにおける安定性を判定できる場合があることを示している。

３．９　二人モデルと集団モデル

３．６から３．９までで考察してきたモデルは二人のプレーヤーが繰り返して何回かゲームを行なうことを想定したモデルである。これに対し、レプリケーターダイナミクスや模倣ダイナミクスで想定してきたのは、プレーヤーの集団があってその中からプレーヤーがランダムに選ばれてゲームを行なう状況であった。

後者では集団中の戦略のシェア（頻度）を考え、シェアの変化を考えるというアプローチが可能であったのに対し、前者ではプレーヤーが二人しかいないため「戦略のシェア」を考えるアプローチは余り意味がない。そこで、前者では「戦略のシェア」の代わりに各々のプレーヤーがある戦略を採用する確率を考え、この「採用確率」のダイナミクスを考える、というアプローチを取った。

このように、両者はダイナミクスを考えるという点では共通しているが、想定する状況や、変化するものの中味が大きく異なっているので、異なるタイプのモデルと考えておいた方がよい。ここでは、前者のタイプのモデルを二人モデル（一般には非集団モデル）、後者のタイプのモデルを集団モデルと呼ぶことにしよう。

社会の中である考え方や行動のパターンが広まっていったり消滅したりする現象をモデル化するには、集団モデルが適している。一方、夫婦の間で役割分担が出来上がっていくプロセスや、寡占企業同士の駆け引き、隣国同士の相互作用などを考えるときには、ランダムマッチングモデルを用いるよりもここで紹介した二人モデルの方が適している。

このように、いずれのモデルにも分析に適した社会現象がある。試行錯誤ダイナミクスについてはここまで二人モデルのみを紹介してきたが、では集団モデル型の試行錯誤モデルは考えられないのであろうか。以下の節ではこの点について考えてみよう。

３．10　集団型試行錯誤ダイナミクス

２×２ゲーム

自分＼相手	A	B
A	a	b
B	c	d

をランダムマッチングで行なっているプレーヤーの集団を考える。各プレーヤーは試行錯誤で戦略を修正しているものとする。このとき、集団内の戦略採用頻度はどのように変化するのであろうか。

プレーヤーｉが戦略ｊを採ろうとする傾向性$p_{ij}(t)$が

$$p_{ij}(t+1)=(1-\phi)p_{ij}(t)+R_{ij}(t)$$

にしたがって変化しているとする。このとき、プレーヤーｉが戦略Ａを採る確率を$x_i(t)$とすると、これの変化Δx_iは

$$\Delta x_i=((1-x_i(t))R_{ia}-x_i(t)R_{ib})/(p_{ia}(t+1)+p_{ib}(t+1))$$

となる。ここまでは３．５節の考察と同じである。

ここで、集団中のx_iの平均としてｘを考えると、集団中からランダムに一人選んだ相手が戦略Ａを採る確率はｘとなる。このとき、プレーヤーｉがＡやＢを採ろうとする傾向パラメーターに与えられる強化値R_{ia}、R_{ib}はそれぞれ

確率$x_i x$で	$R_{ia}=a$、$R_{ib}=0$
確率$x_i(1-x)$で	$R_{ia}=b$、$R_{ib}=0$
確率$(1-x_i)x$で	$R_{ia}=0$、$R_{ib}=c$
確率$(1-x_i)(1-x)$で	$R_{ia}=0$、$R_{ib}=d$

となる。

これより、Δx_iは

確率$x_i x$で	$(1-x_i)a/((1-\phi)k_{ia}(t)+a)$
確率$x_i(1-x)$で	$(1-x_i)b/((1-\phi)k_{ia}(t)+b)$
確率$(1-x_i)x$で	$-x_i c/((1-\phi)k_{ia}(t)+c)$

確率$(1-x_i)(1-x)$で　$-x_i d/((1-\phi)k_{ia}(t)+d)$

となる。

分母が少しずつ異なるが、ここでもこれらをおおむね同じと考えてΔx_iの期待値を求めると

$$\Delta x_i\text{の期待値の分子}=x_i x(1-x_i)a+x_i(1-x)(1-x_i)b$$
$$-(1-x_i)x x_i c-(1-x_i)(1-x)x_i d$$
$$=x_i(1-x_i)(xa+(1-x)b-xc-(1-x)d)$$

ここで$xa+(1-x)b$はプレーヤーiが戦略Aを採ったときの期待効用u_aに等しく、$xc+(1-x)d$は戦略Bを採ったときの期待効用u_bに等しいので、

$$\Delta x_i\text{の期待値}=x_i(1-x_i)(u_a-u_b)/\text{分母}\qquad(式11)$$

となる。

これより$u_a>u_b$のときx_iの期待値は増加し、$u_a<u_b$のときx_iの期待値は減少することがわかる。ところで、この結果は任意のプレーヤーiについて成立する。ここで

$$\Delta x=\Sigma\Delta x_i/N$$

であるので、任意のプレーヤーについて$\Delta x_i>0$ならば$\Delta x>0$、$\Delta x_i<0$ならば$\Delta x<0$である。

以上より、

$u_a>u_b$　のとき　$\Delta x>0$

$u_a<u_b$　のとき　$\Delta x<0$

であることがわかる。

３．11　集団型試行錯誤ダイナミクスの性質

式11から、集団型試行錯誤ダイナミクスの性質について幾つかの命題が導ける。まず、状態空間の頂点の定常性についてである。ちなみに、２戦略集団モデルでは、集団の中で戦略Aが採用される確率の平均xが分かれば、戦略Bが採用される確率の平均$1-x$も分かるので、状態空間は$0\leqq x\leqq 1$の線分（単位単体）となる。状態空間の頂点は、単位単体の両端$x=0$と

$x=1$である。

状態空間の頂点について次の命題が成立する。

【命題5】 2×2対称ゲームの集団型試行錯誤ダイナミクスにおいて、状態空間の頂点は定常点である。

【証明】

$x=0$の場合を考えると、$x=x_i/N$と$x_i \geqq 0$より、$x=0$のときは任意のプレーヤーiについて$x_i=0$である。このとき式11より任意のiについて$\Delta x_i=0$である。したがって$\Delta x=0$。ゆえに$x=0$は定常点である。

$x=1$の場合は、$0 \leqq x_i \leqq 1$より任意のiについて$x_i=1$である。したがって、任意のiについて$\Delta x_i=0$なので$\Delta x=0$。ゆえに$x=1$は定常点である。

【証明終わり】

二人モデルの場合と同様に、集団モデルの場合も状態空間の頂点は定常点である。頂点の安定性については次の命題が成立する。

【命題6】 2×2対称ゲームにおいて、戦略プロファイル（Ⅰ，Ⅰ）がstrict Nash均衡であるならば、戦略Ⅰに対応する頂点は、集団型試行錯誤ダイナミクスにおいて漸近安定である。

【証明】

戦略プロファイル（A，A）がstrict Nash均衡の場合を考える。このとき$a>c$である。式11において$f(x)=u_a-u_b$とおくと

$$f(1)=a-c>0$$

である。

$f(x)$は連続な関数なので、$f(1)>0$より$x=1$の近傍に

$$f(x)>0$$

となる領域があることがわかる。したがって、ｘが１から少しずれても、ずれが十分小さければ

Δx_iの期待値＞０

となる。ゆえに、戦略Ａに対応する頂点ｘ＝１は漸近安定である。

戦略プロファイル（Ｂ，Ｂ）がstrict Nash均衡の場合も同様に証明できる。

【証明終わり】

【命題７】　２×２対称ゲームにおいて、戦略プロファイル（Ｉ，Ｉ）がNash均衡でなければ、戦略Ｉに対応する頂点は、集団型試行錯誤ダイナミクスにおいて不安定である。

【証明略】

例えば、戦略プロファイル（Ａ，Ａ）がNash均衡でなければ、ａ＜ｃとなることから証明できる。

このように、各プレーヤーが試行錯誤で戦略を修正する場合の平均戦略採用頻度に関しても、レプリケーターダイナミクスや模倣ダイナミクスと同様の命題が成立する。

４　まとめ

本論文では学習ダイナミクスのうち、模倣ダイナミクスと試行錯誤ダイナミクスについて検討してきた。これらはいずれも、プレーヤーが利得構造について知らない場合にも実現可能なダイナミクスである。２×２ゲームについての分析の結果、いずれのダイナミクスにおいてもstrict Nash均衡に対応する状態が漸近安定となり、Nash均衡ではない状態が不安定となることが示された。

模倣ダイナミクスと試行錯誤ダイナミクスは現実の社会的相互作用においても機能している可能性が高いダイナミクスなので、これらのダイナミ

クスモデルの整備によって、より広い範囲の社会現象がゲーム理論を用いて分析できるようになることが期待される。

文献

Erev,I. & Roth, A. E. 1998. "Predictiong How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria" *The American Economic Review* 88 (4) 848-881

石原英樹,金井雅之 2002.『進化的意思決定』朝倉書店

Roth,A.E. & Erev, I. 1995. "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term" *Games and Economic Behavior* 8 164-212