オンライン・コンテンツ 3.1：
混合戦略の期待利得と最適反応の「極端な性質」を一般の p, q で理解する

図斎 大

本稿は，浅古泰史・図斎大・森谷文利『活かすゲーム理論』有斐閣の第 3 章までを読んだ読者を対象としたオンライン・コンテンツです。この文書でも引き続きモデル 3.1（PK 戦）を考えます。このモデルについて本文を読み返してください。

期待利得をどのように計算するのかというのを本文ではキーパーとキッカーそれぞれの混合戦略を $p = 0.2,\ q = 0.4$ と特定の数値例で説明しました。これを数値を特定せずに p, q という文字のままで，公式のように見せるのが，この文書の前半です。これは「混合戦略での最適反応の極端な性質」を数学的にきちんと理解することにつながります。本文では直観的に説明しましたが，この文書の後半では期待利得を p, q の式として表したのを活用して，ちょっと数学的に「極端な性質」がなぜ成り立つのかを説明します。

1 PK 戦での期待利得の計算の一般化

本文の 2.3 項ではキッカーが右サイドを攻める確率pを 0.2，キーパーが右サイドを守る確率qを 0.4 として期待利得を求め，そして表 3-4 では p, q を 0.2 刻みで利得表を示しました。ここではモデル 3.1 の PK 戦の利得表（表 3-1）を引き続き考えつつ，期待利得の計算を一般化しましょう。つまり$p,\ q$に数字を代入せず，このままで期待利得を求めていきます。「キッカーが右をとる確率」がp，そして[キーパーが右をとる確率]がqだったので，「両方がともに右をとる」という確率はpqとなります。またキッカーが左をとる確率は$1-p$になるので，「キッカーが左を，キーパーが右をとる」という確率は$(1-p)q$だとわかります。こうして戦略の組(p, q)のもとでのそれぞれの結果の実現する確率分布は下の表 A3-1 のように求められます。

表 A3-1　キッカーが戦略 p，キーパーが戦略 q をとっているときの，それぞれの結果の確率

		$1-q$	q
		キーパー	
		左	右
$1-p$	キッカー 左	$(1-p)(1-q)$	$(1-p)q$
p	右	$p(1-q)$	pq

この計算を踏まえて，本文の表 3–3 のようにそれぞれの結果からの利得と確率をまとめたのが表 A3–2 です。

表 A3–2　「キッカーの戦略 p とキーパーの戦略 q」のもとでのそれぞれの結果の起こる確率とそこからの利得

結果		確率	利得	
キッカー	キーパー		キッカー	キーパー
左	左	$(1-p)(1-q)$	0	3
左	右	$(1-p)q$	3	0
右	左	$p(1-q)$	2	1
右	右	pq	0	3

この表 A3–2 に基づいて，キッカーの期待利得を計算すると，

$$(1-p)(1-q)\cdot 0+(1-p)q\cdot 3+p(1-q)\cdot 2+pq\cdot 0 \qquad \text{(A1)}$$

と表されます。この式のp, qに任意の値，つまり任意のキッカー・キーパーの戦略を代入すれば，その戦略の組の下でのキッカーの利得が出てくるので，この式がキッカーの利得「関数」を表していると言えます。試しに式 (A1)のp, qに 0.2 と 0.4 をそれぞれ代入してみてください。すると，本文2.3項で求めた「キッカーの戦略 0.2 とキーパーの戦略 0.4」からのキッカーの期待利得1.2が出てくるはずです。同様にキーパー（第2のプレーヤー）の利得関数も，以下の式のように計算されます。

$$(1-p)(1-q)\cdot 3+(1-p)q\cdot 0+p(1-q)\cdot 1+pq\cdot 3. \qquad \text{(A2)}$$

キッカーについては式(A1)，キーパーについては式(A2)に，0.2 刻みのp, qを代入すると本文の表 3–4 を得ることができます。

2 最適反応の「極端な」性質の数学的な説明

モデル 3.1 におけるキッカーの最適反応を考えてみましょう。そこで，キーパーの戦略qを 0.8 としてみます。つまり，キーパーが左を選択する確率qが 0.8 となる混合戦略を選択しているとしましょう。一方で，キッカーが左を選択する確率を，ここではpとしたまま，変数で表現してみます。最適反応として知りたいことは，キーパーが戦略 0.8 を選択した場合に，キッカーの期待利得を最大にするようなpの値を見つけることです。キッカーの期待利得を示す式(A1)に，ここで想定しているキーパーの戦略$q=0.8$を代入すると，

$$(1-p)\ \cdot 0.2\cdot 0+(1-p)\cdot 0.8\cdot 3+p\cdot 0.2\cdot 2+p\cdot 0.8\cdot 0\ =-2p+2.4 \qquad \text{(A3)}$$

として求められます。傾きが負の 1 次式なので，このグラフは図 A3–1 (a)のようになっているとわかります。右を選択する確率pの値が高くなるほど，この期待利得は小さくなることがわかります。し

たがって，できるだけpを小さくしていくことがキッカーには好ましく，期待利得は，$p = 0$で最大になっています。すなわち，キーパーの戦略$q = 0.8$に対するキッカーの最適反応は$p = 0$である，つまり左サイドに必ず蹴る純粋戦略だということがわかります。

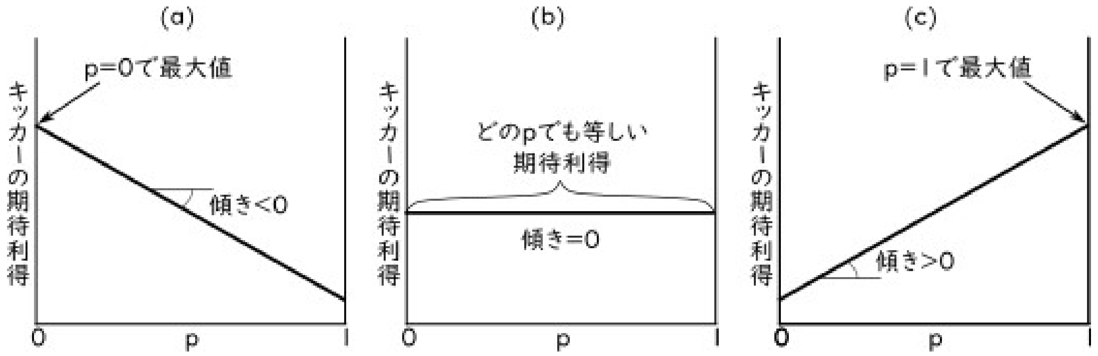

図 A3.1. キッカーの期待利得の全パターン

同様に，$q = 0.6$（右を確率 0.6 で選択）と$q = 0.2$（右を確率 0.2 で選択）それぞれへのキッカーの最適反応を求めてみてください。それぞれを式(A1)に代入して，グラフを描いて，そこから期待利得を最大にするpを探せばよいわけです。$q = 0.8$と同じく$q = 0.6$に対してもグラフは図 A3-1 (a)のようになり$p = 0$が最適反応になったと思います。一方で，$q = 0.2$に対しては，期待利得を示すグラフは図 A3-1 (c)のようになり，右を選択する確率pの値が高くなるほど，期待利得が大きくなっていくことがわかります。よって，$p = 1$が最適反応になります。

ただどちらにしてもキーパーの戦略を特定した後の，キッカーの期待利得をグラフにすると直線になります。それを確認するために式(A1)を

$$\{-0(1-q) - 3q + 2(1-q) + 0q\}p + \{\,0(1-q) + 3q\,\} = (2-5q)\ p + 3q$$

と変形しましょう。ここでqはキーパーが決めているのでキッカーは変えられません。ですから，p のみがキッカーが動かせる変数であり，上式では■や■で囲ったところはキッカーにとっては動かせない数，すなわり「定数」になります。したがってこの期待利得の式は，キッカーにとっては■p + ■という 1 次関数の式だと言えます。（例えば$q = 0.8$を代入すると，■ $= -2$と■ $= 2.4$という体数になり，この式■p + ■は$-2p\ + 2.4$となり(A3)と一致します。）なので，グラフにすると直線，特に■が傾きで■が切片になります。そして，■の中が負なら傾きが負でグラフは図 A3-1 (a)のようになり$p = 0$で最大値，■の中が正なら傾きが正でグラフは図 A3-1 (c)のようになり$p = 1$で最大値になるわけです。まとめると，傾きがゼロでない限りは，$p = 0$か$p = 1$という端，つまり純粋戦略で期待利得は最大になります。実際，この2つの端で期待利得が異なれば傾きはゼロではありません。このことから性質 A)が導かれます。

qを 0 から 1 に連続的に増やしていくと■の中，つまり傾きが正から負に変わっていきますね。もちろんその途中で傾きがゼロになります。このときのキッカーの利得のグラフは図 A3-1 (b)のようになり，どのpでも同じ期待利得になることが見てとれます。このときのqは■の中をゼロとした式，つまり$2-5q=0$を解くことで$q=0.4$だとわかります。上の式変形で$\{-0(1-q)-3q+2(1-q)+0q\}$が$(2-5q)$になっていた，つまりこの2つは同じです。なので，$2-5q=0$というのは前者の方をゼロにしたもの，つまり

$$0(1-q)q+3q=2(1-q)+0q$$

とも書けますが，これは左辺が$p=0$からの，右辺が$p=1$からの期待利得を表しています。つまり，$q=0.4$は，この2つの純粋戦略からの期待利得が同じになることから導かれることから注意してください。これが性質 B)の状況です。

以上の議論をまとめると，$q=0.4$に対しては，キッカーにとってどんな戦略pでも最適反応になり，そのqを境目として，それよりもqが小さいと$p=1$が，qが大きいと$p=0$が最適反応になり，本文で示した結果になります。

ここまで qに具体的な数字を与えて性質 A)，B)を導きましたが，この議論の肝は，qを定数として与えると利得関数がpの 1 次関数になること，そしてその傾きの中がまたqの 1 次関数になっていることです。そしてそれぞれの変数に関して 1 次関数になっているのは，偶々ではなく，利得表（本文の表 3-1）のような事後の結果の利得と表 A3-1 のような 2 者の混合戦略での確率の積を掛け合わせて期待利得を求めること自体から来ています。したがって，性質 A)，B)は違うゲームでも成立する一般的な性質なのです。