

卒業研究論文

確率モデルを用いた卓球の試合分析と
選手の強さの評価

学籍番号 10D8104007K

安田 直人

中央大学理工学部情報工学科田口研究室

2014年3月

あらまし

本研究では，卓球の試合を分析して，選手の特徴を見出し，その選手がより強くなるための改善点を示すことを目的とする．卓球の試合から卓球マルコフモデルを作成し，試合のシミュレーションを行う．さらに，Bradley-Terry モデルを用いて試合結果から選手の強さの評価を行う．

キーワード：卓球，マルコフモデル，Bradley-Terry モデル

目次

第1章	はじめに.....	1
第2章	試合データ.....	2
2.1	構成.....	2
2.2	対象試合.....	3
第3章	マルコフモデルによる選手の分析.....	4
3.1	マルコフモデル.....	4
3.1.1	マルコフ連鎖と推移確率行列.....	4
3.1.2	吸収的マルコフ連鎖.....	5
3.2	卓球マルコフモデル.....	6
3.2.1	状態空間の定義.....	6
3.2.2	推移確率行列の決定.....	8
3.3	試合のシミュレーション.....	11
3.3.1	シミュレーション設定.....	11
3.3.2	シミュレーションのアルゴリズム.....	12
3.4	シミュレーション結果の分析.....	13
3.4.1	実際の試合結果との比較.....	13
3.4.2	試合の分析.....	14
第4章	Bradley-Terry モデルによる強さの推定.....	20
4.1	Bradley-Terry モデル.....	20
4.1.1	強さについて.....	20
4.1.2	BT モデルの意味.....	21
4.2	強さの推定.....	22
4.2.1	勝数の分布と尤度.....	22
4.2.2	最尤法.....	23
4.2.3	推定量に関する注意.....	25
4.3	卓球選手の強さの推定.....	26
4.3.1	対象選手.....	26
4.3.2	強さの推定値と考察.....	27
第5章	おわりに.....	30
5.1	まとめ.....	30
5.2	今後の課題.....	30
謝辞	31
参考文献	32

第1章 はじめに

近年，日本の卓球人気上昇してきている．世界選手権やオリンピックの試合はテレビのゴールデンでも放送されている．卓球王国と呼ばれる中国では国技といえる人気ぶりだが，日本でも野球やサッカーのような国民的スポーツになる日も近いといっても過言ではないのかもしれない．

2013年5月，卓球の世界選手権大会個人戦がパリで開催された．男子シングルス2回戦では日本の松平健太選手が2008年北京五輪金メダリストの馬琳(マ・リン)選手を4-1で破り，さらに4回戦では2009年W杯金メダリストのブラディミル・サムソノフ選手を4-3で破る番狂わせを演じた．準々決勝では，世界ランキング1位の中国の許昕(キョ・キン)選手に2-4で敗れ，同種目で34年ぶりのメダル獲得とはならなかったものの，ベスト8入りを果たした．松平健太選手を始め，日本人選手の今後の活躍に期待したい．

卓球は上級レベルになるとラリー戦になることも多く，目まぐるしい攻守の切り替えが行われるため，動体視力，持久力，一瞬の判断や動作などが要求される．一方で，現在のところ，卓球ではスコアブックという試合内容を記録するものが存在せず，数理的な研究もあまり進められていない．

そこで本研究では，2013年世界選手権大会個人戦の一部の試合の映像を再生し，打球の本数を打法に分けて記録したデータからマルコフモデルを用いて試合構成，攻撃パターン，勝敗の要因を分析する．また，世界選手権2009年，2011年，2013年の3年分の男子シングルス試合結果のデータからBradley-Terryモデルを用いて選手の強さの評価，選手間の強さの比較を行う．

第2章 試合データ

2.1 構成

卓球における打法の種類と意味を表 2.1 に示す。

表 2.1 打法の種類と意味

記号	打法名	意味
srv	サービス	ラリーの始まり
sho	ツツキ, ストップ	台上で下回転系のボールに対して小さく返す
flk	フリック, チキータ	台上で前進回転を与えて払うように返す
drv	ドライブ, スマッシュ	ボールに強い前進回転を与えて返す
cnt	カウンタ	相手の強打を強打して返す
blo	ブロック, ロビング	相手の強打を繋ぐように返す

本研究で使用する打法の種類について説明する。ここで、打法の種類を細かく分類しすぎると本数の合計値に差が表れにくく、映像からでは厳密に区別をつけにくいこともあり、類似している打法をいくつかまとめている。具体的には、ツツキは台上でのレシーブなどで使われる小さくカットする打法で、ストップは主に相手の短い下回転系のボールに対し、相手コートに2バウンド以上するように小さく返す打法であり、共に台上での繋ぎの守備的な打法である。ドライブは強い前進回転を与える打法で、スマッシュは強く弾くように叩き付ける打法であり、共に強打で攻撃する打法であるためまとめている。カウンタは相手が攻撃に出てくるところを防ぎ、即座に攻撃に転じ、相手の攻撃の勢いを利用して攻撃することである。カウンタをドライブ、スマッシュと区別した理由は攻撃の流れの起点になり得る重要な要素になると考えたためである。ブロックは後ろに下がらず相手の強打を返球する打法で、ロビングは後ろに下がりボールを高く打ち上げて時間を稼ぎ返球する打法であり、共に相手の強打を繋いで返球することが目的であるためまとめている。

また、フォアハンド(記号:f)とバックハンド(記号:b)の両面で攻防するため、各打法パターンは上述の打法と面を用いて"(打法の記号)_(面の記号)"のように表す。例えば、フォアのドライブを打つことを"drv_f"と表す。このように、試合中の打法と面を含めた12項目で表現する。

2.2 対象試合

本研究で扱うデータは、DVDなどの試合映像を再生し、実際に行われたプレーを記録して作成する。データの収集を行った試合は、世界卓球2013パリ男子シングルスで行われた2試合である。この2試合の詳細を表2.2に示す。松平健太 vs 馬琳の試合は最終セットのみのデータを使用している。

表 2.2 各試合の日付・選手名・結果

試合	日付	対戦相手	結果
2回戦	2013. 5. 17	松平健太 vs 馬琳	11-7
4回戦	2013. 5. 18	松平健太 vs サムソノフ	3-11, 15-13, 12-14, 11-8, 11-9, 3-11, 11-8

第3章 マルコフモデルによる選手の分析

本章では、マルコフ性をもった確率過程であるマルコフ過程[1]を用いて、試合のシミュレーションを行う。

3.1 マルコフモデル

3.1.1 マルコフ連鎖と推移確率行列

マルコフ連鎖には状態と状態間の推移という概念がある。一般に、時点 n における確率過程 X_n がとりうる値を状態といい、状態のすべての集合を状態空間という。状態 i から状態 j へ移ることを推移という。推移は、「時点 n で状態 i にいたとき、次の時点 $n+1$ に状態 j に推移する確率は、時点 $n-1$ 以前にどの状態にいたかには無関係である」という仮定に基づいている。この仮定をマルコフ性という。

離散時点の確率過程 $X_n(n=0,1,2,\dots)$ を考える。任意の $n \geq 1$ と状態空間 $S = \{i_0, i_1, \dots, i_{n-1}, j\}$ に対して

$$\begin{aligned} P\{X_n = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} \\ = P\{X_n = j | X_{n-1} = i_{n-1}\} \end{aligned} \quad (3.1)$$

が成り立つとき、確率過程 $\{X_n\}$ はマルコフ性をもつという。そしてマルコフ性をもった状態空間が離散的であるとき、確率過程 $\{X_n\}$ をマルコフ連鎖という。また、式(3.1)の推移確率が時点 n によらないとき、推移確率は定常であるといい、マルコフ連鎖は斉次的であるという。

斉次的マルコフ連鎖 X_n の推移確率を

$$p_{ij} = P\{X_n = j | X_{n-1} = i\} (1 \leq i, j \leq N) \quad (3.2)$$

とおく。 p_{ij} は確率であるから、確率の基本性質

$$0 \leq p_{ij} \leq 1, \quad \sum_{j=1}^N p_{ij} = 1 \quad (3.3)$$

を満たしている。 p_{ij} を行列の形に並べた

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix} \quad (3.4)$$

を推移確率行列と呼ぶ。式(3.3)の条件から P の行和はすべて1になる。このようにマルコフ連鎖は推移確率行列によって推移の構造を表現することができる。

3.1.2 吸収的マルコフ連鎖

図 3.1 のような推移図をもったマルコフ連鎖を考える. この推移図には推移確率を記入していないが, 矢印のあるところにはすべて正の推移確率があるものとする.

マルコフ連鎖が状態 1 から出発したとする. 次の時点では状態 2, 4, 6 のいずれかの状態に推移する. ところが一度状態 2 に推移すると, その後は状態 2, 3 を行き来するだけでこの二つの状態以外に推移することはない. このようにマルコフ連鎖では状態 1, 6, 7 の間を推移していると, いつかは $C_1 = \{2,3\}, C_2 = \{4,5\}, C_3 = \{8,9\}$ の三つの集合のいずれかに推移して, 一度その集合に推移するといつまでもその中でだけ推移していて決してその集合の外に出ることはない. これが一般的なマルコフ連鎖のふるまいである. 上の C_1, C_2, C_3 のような状態の集合のことを既約な集合と呼び, 既約な集合 C は次の二つの条件で特徴付けられる.

- 1) C 中のどの二つの状態をとっても一方から他方へ何ステップかで推移することができる.
- 2) C のどの状態からも C の外の状態へ推移することはできない.

一般のマルコフ連鎖の状態空間 S は, この二つの条件を満たすいくつかの既約な集合 C_v と, それらに含まれない状態の集合 T に分割できる. 図 3.1 の例では $T = \{1,6,7\}$ であり, $S = T \cup C_1 \cup C_2 \cup C_3$ となっている.

このように, 一般のマルコフ連鎖の推移の様子は二つに分けて考えることができる. 一つは T の中で推移していずれかの C_v に推移するまでであり, もう一つは C_v への推移後である. 例えば, 図 3.1 のマルコフ連鎖でいずれかの C_v に推移するまでの様子を調べる場合, C_1, C_2, C_3 をそれぞれ一つの状態にまとめた図 3.2 のような推移図をもったマルコフ連鎖を調べればよい. 図 3.2 のように既約な集合がすべて一つの状態からなるとき, その状態を吸収状態と呼び, 集合 G で表す. G に含まれない状態を一時的状態と呼ぶ. 吸収状態, 一時的状態をもつマルコフ連鎖を吸収的マルコフ連鎖と呼ぶ.

吸収的マルコフ連鎖の推移確率行列 P の主対角線上に 1 があるとき, それに対応する状態は吸収状態である. 状態の番号を適当につけ変えて一時的状態に若い番号を与えると, 推移確率行列 P は

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \quad (3.5)$$

と表される. N をすべての状態の数, r を一時的状態の数とすると, Q は $r \times r$ 行列で T から T への推移を表し, R は $r \times (N - r)$ 行列で T から G への推移を表している. O は $(N - r) \times r$ の零行列で, I は $(N - r) \times (N - r)$ の単位行列である. 式 (3.5) を推移確率行列の標準形という.

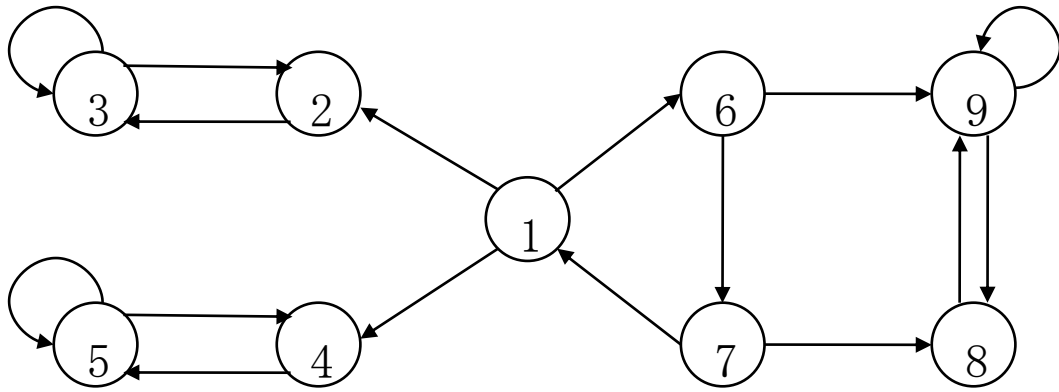


図 3.1 一般的なマルコフ連鎖の例

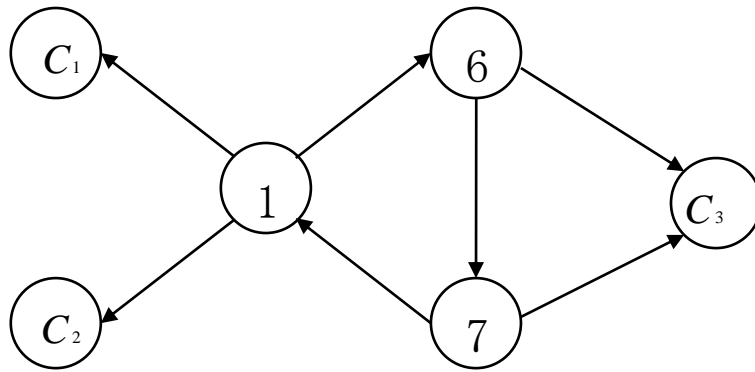


図 3.2 既約な集合をまとめた吸収的マルコフ連鎖

3.2 卓球マルコフモデル

第 2 章で得られたデータから状態空間を定義し，推移確率行列を決定し，試合のシミュレーションを行う。

3.2.1 状態空間の定義

一方の選手が相手にボールを返球する際の打法を状態として定義する．本研究で扱う状態空間を表 3.1 に示す．状態 1 から状態 12 は選手 A の打法パターンを表している．状態 13 から状態 24 は選手 B の打法パターンを表している．また，状態 25 は選手 A の得点（選手 B の失点），状態 26 は選手 B の得点（選手 A の失点）を表している．

表 3.1 状態空間

状態	選手	打法パターン
1	A	srv_f
2	A	srv_b
3	A	sho_f
4	A	sho_b
5	A	flk_f
6	A	flk_b
7	A	drv_f
8	A	drv_b
9	A	cnt_f
10	A	cnt_b
11	A	blo_f
12	A	blo_b
13	B	srv_f
14	B	srv_b
15	B	sho_f
16	B	sho_b
17	B	flk_f
18	B	flk_b
19	B	drv_f
20	B	drv_b
21	B	cnt_f
22	B	cnt_b
23	B	blo_f
24	B	blo_b
25	A	scr
26	B	scr

3.2.2 推移確率行列の決定

前項の状態空間に基づいてデータの集計を行う。状態*i*から状態*j*へ推移する打球の本数の集計結果を表 3.2 と表 3.3 に示す。行番号は推移前の状態で、列番号は推移先の状態を表しており、番号は表 3.1 の状態番号に対応している。状態空間の集合*S*は $S = \{1, 2, \dots, 26\}$ 、一時的状態の集合*T*は $T = \{1, 2, \dots, 24\}$ 、吸収状態の集合*G*は $G = \{25, 26\}$ である。

表 3.2 松平健太 vs 馬琳の本数の集計結果

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2	1	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	1	3	2	1	1
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	0	0	0	1	0	1
13	0	0	7	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	1	0	2	1	2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
23	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
24	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

表 3.3 松平健太 vs サムソノフの本数の集計結果

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	10	9	6	5	5	0	0	0	0	0	0	2
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	1	3	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	3	7	0	15	10	0	0	0	0	0	0	3
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	1	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	1	0	4	0	0	1
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	12	2	17	35	5	11	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	2	0	5	15	3	5	
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	2	5	2	4	
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	1	0	2	
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	4	1	0	1	1	1	7	
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	34	22	0	0	1	3	2	18	
13	0	0	22	0	3	8	12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	0	0	15	1	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
15	0	0	7	2	2	0	12	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
16	0	0	2	0	0	0	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
17	0	0	0	0	0	0	6	1	2	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
18	0	0	0	0	0	0	5	2	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	1	9	3	10	36	0	0	0	0	0	0	0	0	0	0	0	0	0	10	6
20	0	0	0	0	0	0	2	1	1	1	5	30	0	0	0	0	0	0	0	0	0	0	0	0	0	7	3
21	0	0	0	0	0	0	2	1	4	1	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	7	1
22	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	10	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	2
24	0	0	0	0	0	0	25	18	0	1	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	11	4
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

集計結果から推移確率行列 P を求める。本項で得た集計結果を行列 D ，要素を d_{ij} とし， i 行の行和 sum_i を求める。 D と sum_i ，そして式 (3.5) より P の各要素 p_{ij} は

$$p_{ij} = \begin{cases} \frac{d_{ij}}{\text{sum}_i} & (1 \leq i \leq 24, 1 \leq j \leq 26) \\ 1 & (25 \leq i \leq 26, 1 \leq j \leq 26, i = j) \\ 0 & (25 \leq i \leq 26, 1 \leq j \leq 26, i \neq j) \end{cases} \quad (3.6)$$

と表される。推移確率行列を表 3.4 と表 3.5 に示す。

表 3.4 松平健太 vs 馬琳の推移確率行列

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0.05	0.25	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.75	0	0	0	0.25	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0.17	0	0.33	0.17	0	0	0	0	0	0	0.17
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0.33	0	0	0	0	0	0.33	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.09	0	0.18	0.09	0.27	0.18	0.09	0.09	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0	0	0	0	0	0.4	0.4
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.63	0.13	0	0	0	0.13	0	0.13	0
13	0	0	0.7	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0.6	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0
16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0.08	0	0.15	0.08	0.15	0.46	0	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0
20	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
23	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.67	0
24	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

表 3.5 松平健太 vs サムソノフの推移確率行列

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.36	0.17	0.16	0.1	0.09	0.09	0	0	0	0	0	0	0.09
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.09	0	0.45	0.09	0.27	0	0	0	0	0	0	0	0.09
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.14	0.07	0.16	0	0.34	0.23	0	0	0	0	0	0	0.07
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0.33	0.33	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0.25	0	0	0	0.25	0	
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.22	0.11	0	0.11	0	0.44	0	0.11	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0.14	0.02	0.2	0.42	0.06	0.13	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.14	0.05	0.05	0	0.14	0.41	0.08	0.14	
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.28	0	0.11	0.28	0.11	0.22	
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0.17	0.17	0	0.33	
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.42	0.15	0.04	0	0.04	0.04	0.04	0.27	
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.43	0.28	0	0	0.01	0.04	0.03	0.23	
13	0	0	0.46	0	0.06	0.17	0.25	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02
14	0	0	0.68	0.05	0	0.05	0.09	0.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.09	0
15	0	0	0.24	0.07	0.07	0	0.41	0.14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0
16	0	0	0.15	0	0	0	0.54	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.08	0
17	0	0	0	0	0	0	0.33	0.06	0.11	0	0.06	0.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.22
18	0	0	0	0	0	0	0.45	0.18	0	0	0	0.36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0.01	0.12	0.04	0.13	0.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0.13	0.08
20	0	0	0	0	0	0	0.04	0.02	0.02	0.02	0.1	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0.14	0.06
21	0	0	0	0	0	0	0.09	0.05	0.18	0.05	0.09	0.18	0	0	0	0	0	0	0	0	0	0	0	0	0	0.32	0.05
22	0	0	0	0	0	0	0.33	0	0	0	0	0.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0.37	0.07	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.41	0.07
24	0	0	0	0	0	0	0.38	0.28	0	0.02	0.06	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0.06
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

3.3 試合のシミュレーション

3.2 節で作成した卓球マルコフモデルを用いて、試合のシミュレーションを行う。

3.3.1 シミュレーション設定

試合のシミュレーションは以下のルールに従う。

- 1 セットである選手が 11 点以上先取、かつ 2 選手の得点差が 2 点以上であればそのセットの勝者となる。
- 1 試合 7 セットマッチで 4 セット先取した選手がその試合の勝者となる。
- サービスは 2 本交代で、10-10 以降は 1 本交代とする。
- 促進ルールや反則行為は考えないものとする。

シミュレーションを実行する際の状態間の推移の規則について述べる。各状態間の推移は確率的に決まると仮定し、0 以上 1 以下の一様乱数 x を用いる。まず推移確率行列 P が

$$\sum_{j=1}^N p_{ij} = 1 \quad (N \text{ は列数}) \quad (3.7)$$

であることから、状態間の推移をプログラムで処理するために、以下で定義する c_{ij} を要素とする行列 C を計算する。

$$c_{ij} = \sum_{k=1}^j p_{ik} \quad (1 \leq i, j \leq 26) \quad (3.8)$$

次に、乱数 $x (0 \leq x \leq 1)$ を発生させ、 $c_{i,j-1} < x \leq c_{ij}$ のとき状態 i から状態 j へ推移させる。

吸収状態（状態 25 から状態 26）へ推移したときは、いったんマルコフモデルの推移を打ち切る。そして再び初期状態（時点 0）からモデルを実行する。

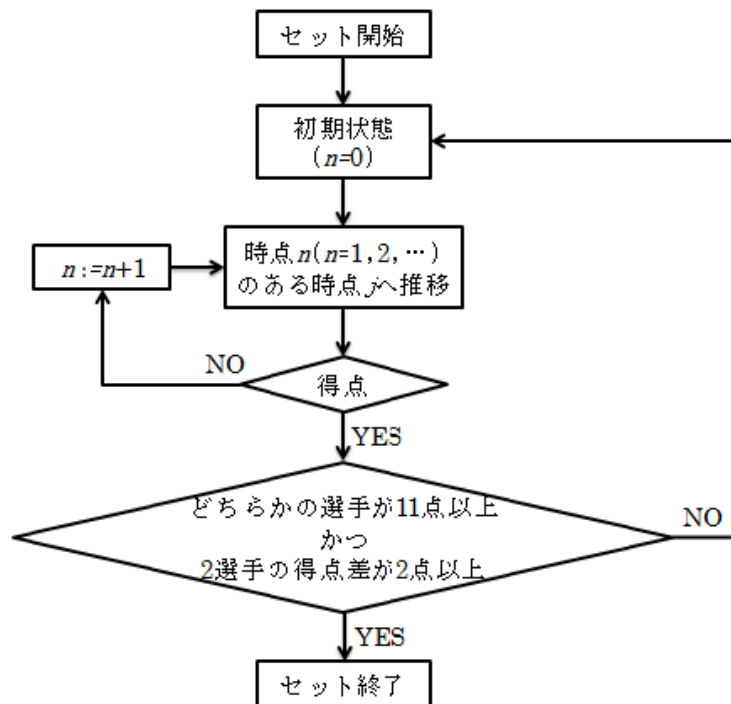


図 3.3 シミュレーションの流れ

3.3.2 シミュレーションのアルゴリズム

1 セットのシミュレーションのアルゴリズムを示す. A_{get} を選手 A の得点, B_{get} を選手 B の得点, t を合計得点とする. S を状態空間の集合 ($S = \{1, 2, \dots, 26\}$ は 3.2 節で定義), T を一時的状態の集合 ($T = \{1, 2, \dots, 24\}$), G を吸収状態の集合 ($G = \{25, 26\}$) を定義する.

- **Step1**

セット開始時の状態 $i (\in T)$ を定める. 初期状態を $i = 1$ または $i = 2$ (B からサービスの場合, $i = 13$ または $i = 14$) とする. $A_{get} = 0$, $B_{get} = 0$, $t = 0$ とする.

- **Step2**

$0 \leq x \leq 1$ の一様乱数 x を発生させて, $c_{ij-1} < x \leq c_{ij}$ を満たす状態 $j (\in S)$ へ推移させる.

- **Step3**

$j = 25 (\in G)$ ならば (a), $j = 26 (\in G)$ ならば (b), そうでなければ (c) を行う.

(a) $A_{get} = A_{get} + 1$, $t = t + 1$ と更新し, $A_{get} \geq 11$ かつ $A_{get} - B_{get} \geq 2$ ならば終了する.

(b) $Bget = Bget + 1$, $t = t + 1$ と更新し, $Bget \geq 11$ かつ $Bget - Aget \geq 2$ ならば終了する.

(c) $i = j$ として **Step2** へ戻る.

● **Step4**

$t = 4k$, $4k + 1$, $t = 2l$ ならば (e) を行う. $t = 4k + 2$, $t = 4k + 3$, $t = 2l + 1$ ならば (f) を行う. ただし, $0 \leq k \leq 4$, $l \geq 10 (\in \mathbf{Z})$ とする.

(e) $i = 1$ または $i = 2$ (B からサービスの場合, $i = 13$ または $i = 14$) として **Step2** へ戻る.

(f) $i = 13$ または $i = 14$ (B からサービスの場合, $i = 1$ または $i = 2$) として **Step2** へ戻る.

3.4 シミュレーション結果の分析

松平健太選手が行った 2 試合について, 100 試合シミュレーションした結果を分析する.

3.4.1 実際の試合結果との比較

実際に行われた試合結果と 100 試合シミュレーションした結果を比較したものを表 3.6 に示す. 比較対象は両選手のセット取得数, 得点数, ラリー数とする. ラリー数は状態 1 から状態 26 までの出現した回数を集計したものである. シミュレーション結果の数値は平均値で表している.

表 3.6 より, 卓球マルコフモデルの整合性について考察する. 松平健太 vs 馬琳 (1 セット) の試合では, セット取得数と得点数はほぼ等しい値が得られた. ラリー数の差は 0.3 程度であり近い値が得られた. 松平健太 vs サムソノフ (7 セット) の試合では, 得点数の差は 5 程度である. セット取得数は実際の試合では松平健太選手が勝利しているが, シミュレーション結果ではサムソノフ選手が勝利している. これは, 実際の試合で松平健太選手が要所をおさえてセットを取得し接戦を制しているためであると考えられる. 過去に影響されないというマルコフモデルの本質からこのようなことが考慮されることはなく, 得点数とセット取得数に正の相関関係が見られる. このように, セット取得数は得点数の差によって左右されるため, シミュレーションの結果の整合性を考える際には考慮しないこととする. ラリー数の差は 25 程度あるが, 全ラリー数の 7%程度であり, 卓球の試合において無視できる値であると判断する. 以上より, 卓球マルコフモデルが実際の試合とほぼ等しい結果を出力することができるといえる.

表 3.6 実際の試合とシミュレーション（100 試合の平均）の比較

	松平健太 vs 馬琳		松平健太 vs サムソノフ	
	実際の試合	シミュレーション	実際の試合	シミュレーション
セット取得数	1 対 0	0.94 対 0.06	4 対 3	2.36 対 3.43
得点数	11 対 7	10.94 対 7.03	65 対 75	61.32 対 68.22
ラリー数	46 対 46	45.77 対 45.99	383 対 379	357.75 対 354.22

3.4.2 試合の分析

各試合の得失点への推移を考察する。シミュレーション結果から状態 1, 2, …, 12 から状態 25, 26 への推移と状態 13, 14, …, 24 から状態 25, 26 への推移回数を集計し、集計結果を百分率で表したものを表 3.7–表 3.10 に示す。

● 松平健太 vs 馬琳の試合分析

松平健太選手の得失点への推移を分析する。表 3.7 より、各状態から状態 25 への推移をみることで得点への推移を分析できる。また、各状態から状態 26 への推移をみることで失点への推移を分析できる。ただし、松平健太選手の得点は状態 1 から状態 12 によるものだけでなく、馬琳選手の失点によるものもある。表 3.7 には馬琳選手の失点による松平健太選手の得点は含まれていないため、表 3.8 も参照する必要がある。得点へ推移するのは状態 5, 7, 9, 10 である。これはフリック、ドライブ、カウンタであり、攻撃系の打法であることを指している。特に、得点へ推移する確率が高いのは状態 7, 9 である。このことからフォアドライブ、フォアカウンタによる得点が多いことがわかる。失点へ推移するのは状態 3, 7, 9, 12 である。攻撃系の打法であるフォアドライブ、フォアカウンタは得点もしているが、その反面、高い確率で失点にも繋がっている。一方、守備系の打法による失点では馬琳選手による攻撃を受ける失点が考えられる。

馬琳選手の得失点への推移を分析する。表 3.8 より、得点へ推移するのは状態 22 である。これはバックカウンタによるもので、カウンタによるミスがないため効果的な打法であるといえる。失点へ推移するのは状態 15, 17, 19, 23, 24 である。攻撃系の打法による失点はフォアフリックとフォアドライブである。守備系の打法であるブロックによる失点する確率が 61%と高く、松平健太選手の攻撃を受ける失点が考えられる。また、両選手ともにフォアツツキによる失点があり、サービスが有効になっていることが予想できる。

表 3.7 松平健太選手の得失点 (馬琳戦)

状態	打法	得点 (状態 25)	失点 (状態 26)
1	srv_f	0%	0%
2	srv_b	0%	0%
3	sho_f	0%	11%
4	sho_b	0%	0%
5	flk_f	11%	0%
6	flk_b	0%	0%
7	drv_f	33%	31%
8	drv_b	0%	0%
9	cnt_f	38%	33%
10	cnt_b	18%	0%
11	blo_f	0%	0%
12	blo_b	0%	25%

表 3.8 馬琳選手の得失点

状態	打法	失点 (状態 25)	得点 (状態 26)
13	srv_f	0%	0%
14	srv_b	0%	0%
15	sho_f	14%	0%
16	sho_b	0%	0%
17	flk_f	12%	0%
18	flk_b	0%	0%
19	drv_f	14%	0%
20	drv_b	0%	0%
21	cnt_f	0%	0%
22	cnt_b	0%	100%
23	blo_f	43%	0%
24	blo_b	18%	0%

松平健太選手と馬琳選手の打法の割合を図 3.4 に示す。松平健太選手と馬琳選手を比較し、それぞれの選手の特徴を分析する。

松平健太選手の打法の割合は全体的にバランスがよく、苦手な打法がないように思える。特に、馬琳選手よりフリックの割合が高く、台上技術に優れているため先に攻撃することができる。また、ブロックの割合は馬琳選手よりも高いため、攻撃されがちであるが、カウンタの割合も高いため、相手のボールに合わせて自分の攻撃に転換することができる。

一方、馬琳選手はドライブの割合が多くを占めていて、ブロックの割合は松平健太選手ほど高くない。このことから、松平健太選手より攻撃的な戦術であるといえる。

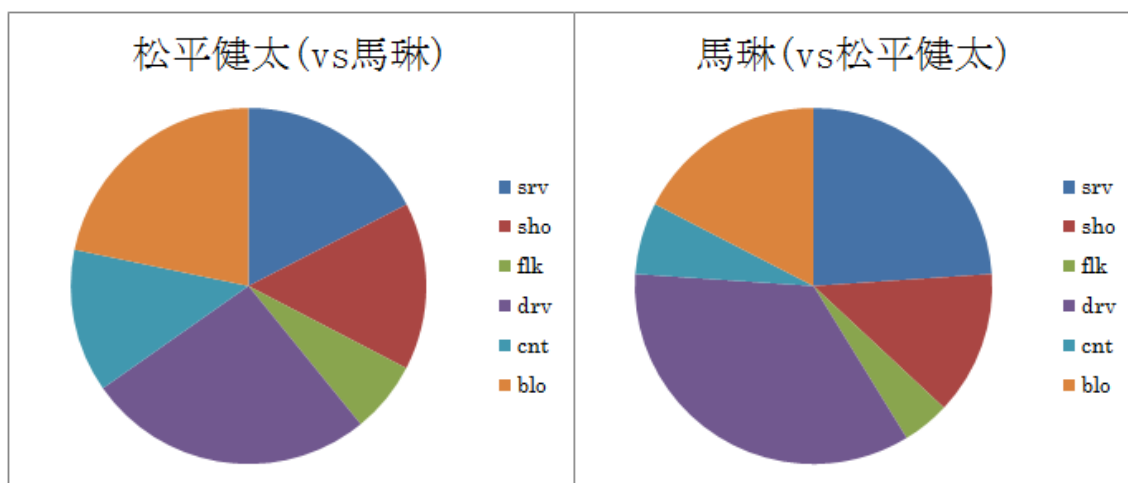


図 3.4 打法の割合 (松平健太 vs 馬琳)

● 松平健太 vs サムソノフの試合分析

松平健太選手の得失点への推移を分析する。表 3.9 より、得点へ推移するのは状態 5, 7, 8, 9, 11, 12 である。攻撃系の打法はフォアフリック、両ドライブ、フォアカウンタを指している。その中でもドライブは両面合わせて得点の 55% と高く得点源になっていることがわかる。また、攻撃だけでなく守備系の打法であるブロックによる得点も 23% と比較的高く、攻守ともに得点することがわかる。一方、失点へ推移するのは状態 4, 5 以外のすべてである。特に、状態 12 のバックブロックによる失点が 34% と高く、サムソノフ選手の攻撃を防げず失点することが予想される。状態 7 のフォアドライブによる失点が 21% である。失点は得点ほど高くないため、より攻めることができれば得点につながると考えられる。

次に、サムソノフ選手の得失点への推移を分析する。表 3.10 より、得点へ推移する状態は状態 13, 17, 19, 20, 21, 23, 24 である。攻撃系な打法はフォアフリック、両ドライブ、フォアカウンタを指している。特に、フォアフリックから得点に推移する確率は、すべての打法による得点の 20% と高く、さらに失点は 0% と非常に効果的な打法であることがわかる。できるだけ先に攻撃をすることで得点できる確率を増やすことができる。両ドライブから得点に推移する確率の合計は 42% と両ドライブが得点源になっている。一方、失点へ推移するのは、状態 14, 15, 16, 19, 20, 21, 23, 24 である。失点に関しては失点する確率が高い状態 19, 20, 21, 23, 24 について分析する。攻撃系の打法は両ドライブ、フォアカウンタを指している。フォアカウンタから得点する確率より失点する確率の方が高いためリスクを負ってまで攻撃する必要がないと考えられる。守備系の打法は両ブロックを指している。バックブロックから失点に推移する確率 21% と得点に推移する 18% に差が小さく、サムソノフ選手の守備系の打法は攻撃的な打法に近いことがわかる。

サムソノフ選手と松平健太選手を比較し、分析を行う。松平健太選手は攻撃系の打法と守備系の打法による得点率と失点率の差が大きいことがわかる。一方、サムソノフ選手は攻撃系の打法と守備系の打法による得点率と失点率の差が小さいことがわかる。このことから、松平健太選手は守備系の打法をなるべく打たず、攻撃系の打法を増やすような戦略を要する。サムソノフ選手は攻守ともに得点することができるため、対戦相手の特徴に合わせて戦法を変えることができる。

両選手ともに共通していえることは、サービスやツツキによるもったいないミスを防ぐことで失点を減らし、できるだけ先に攻撃することができれば得点に繋がる機会が増えると予想できる。

表 3.9 松平選手の得失点（サムソノフ戦）

状態	打法	得点（状態 25）	失点（状態 26）
1	srv_f	0%	3%
2	srv_b	0%	2%
3	sho_f	0%	6%
4	sho_b	0%	0%
5	flk_f	9%	0%
6	flk_b	0%	2%
7	drv_f	34%	21%
8	drv_b	21%	9%
9	cnt_f	13%	7%
10	cnt_b	0%	4%
11	blo_f	7%	12%
12	blo_b	16%	34%

表 3.10 サムソノフ選手の得失点

状態	打法	失点（状態 25）	得点（状態 26）
13	srv_f	0%	6%
14	srv_b	4%	0%
15	sho_f	4%	0%
16	sho_b	2%	0%
17	flk_f	0%	20%
18	flk_b	0%	0%
19	drv_f	20%	27%
20	drv_b	13%	15%
21	cnt_f	15%	6%
22	cnt_b	0%	0%
23	blo_f	21%	9%
24	blo_b	21%	18%

松平健太選手とサムソノフ選手の打法の割合を図 3.5 に示す。松平健太選手とサムソノフ選手を比較し、それぞれの選手の特徴を分析する。

両選手のそれぞれの打法の割合は類似していることが見て取れる。つまり、2人のプレースタイルが似ているということである。その中でも、サービスの割合がほぼ等しいことは自明だが、ツッツキ、ドライブ、カウンタの割合もほぼ等しい割合であることがわかる。フリックを比較すると、若干ではあるがサムソノフ選手の方が高い割合を占めている。これは、相手のサービスやツッツキから先に攻撃に繋げることができていることを表している。さらに、ブロックの割合を比較すると松平健太選手の方が高いことがわかる。以上のことから、サムソノフ選手の方が攻撃する機会が多いことがわかる。松平健太選手は、ツッツキをする場面でフリックをすることができれば自分から攻撃に繋げることができる。さらに、自分の武器であるカウンタの割合を増やすことができれば、より試合展開が変わっていたのかもしれない。

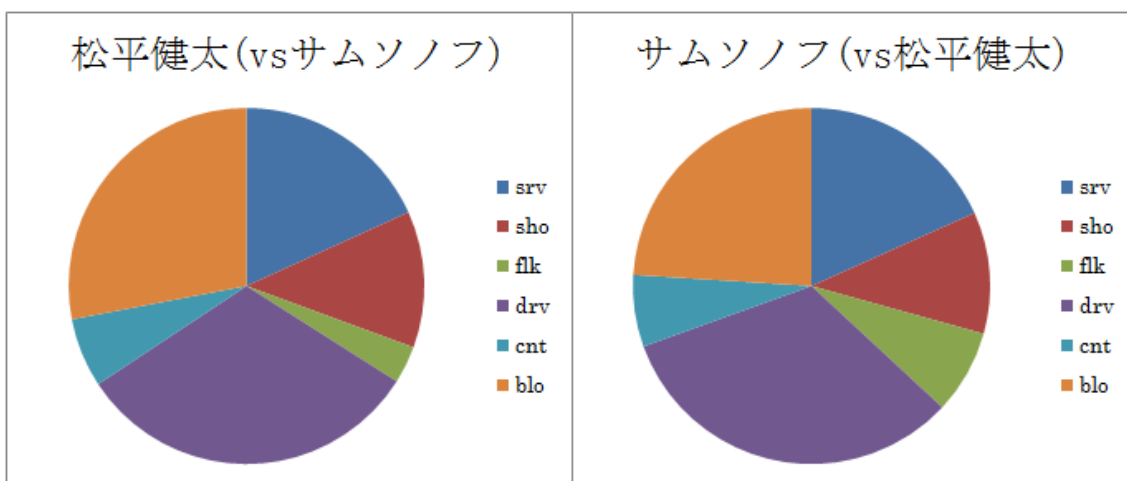


図 3.5 打法の割合 (松平健太 vs サムソノフ)

第4章 Bradley-Terry モデルによる強さの推定

球技をはじめとする2人の選手あるいは2つのチームが対戦する多くのスポーツでは、リーグ戦などの形で互いに何回か対戦し、その結果に基づいて各選手あるいはチームの強さを決めることがよくある。

本章では、世界選手権の試合結果から Bradley-Terry モデル[2]を用いて合理的に各選手の強さを推定する。

4.1 Bradley-Terry モデル

4.1.1 強さについて

一般に、スポーツにおいては強い方が必ず勝つとは限らない。実際に、強いはずの選手が負ける番狂わせがなく、一方が常に勝つものと決まっていれば、ゲームの面白さはなくなってしまう。したがって、ある選手が強いということは、その選手が必ず勝つということではなく、その選手の勝つ確率が大きいことを意味すると考える。

全部で m 人の選手が互いに何回か対戦するものとし、そのうちの選手 i が選手 j に勝つ確率を

$$p_{ij} \quad (i \neq j; i, j = 1, 2, \dots, m)$$

と表す。ここでは引き分けはないものと仮定すると

$$p_{ij} + p_{ji} = 1 \quad (i \neq j) \tag{4.1}$$

が成り立つ。

ここで、これらの確率から強さというものが一義的に定められるための条件について考える。たとえば $m = 3$ として

$$\begin{aligned} p_{12} &< \frac{1}{2} < p_{21} \\ p_{23} &< \frac{1}{2} < p_{32} \\ p_{31} &< \frac{1}{2} < p_{13} \end{aligned} \tag{4.2}$$

という関係が成り立つ場合、選手2が1より強く、3が2より強く、1が3より強い、いわゆる「三すくみ」の状態になって、3選手間の強さの順序さえ簡単には決められないことになる。

各選手の強さが一義的に定められるためには、それらの確率の間に、単に不等式 (4.2) が成立してはならないというだけでなく、より強い関係が成立しなければならない。そのために自然な条件として考えられるのは、各選手に対応して m 個の量 $\pi_1, \pi_2, \dots, \pi_m$ が存在して、すべての i, j の組み合わせに対して確率 p_{ij} が

$$p_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \quad (i \neq j) \quad (4.3)$$

あるいは

$$\frac{p_{ij}}{p_{ji}} = \frac{\pi_i}{\pi_j} \quad (i \neq j) \quad (4.4)$$

が成立するものと想定することである。このとき、 π_i は選手 i の「強さ」を表わすものと解釈できる。この想定は、Bradley and Terry によるもので、Bradley-Terry モデル[2]と呼ばれる。以下ではそれを BT モデルと略称する。

4.1.2 BT モデルの意味

選手 i と j が勝負を決めるのに、直接対戦をせず、別の選手 k との対戦結果によって決めることにする。つまり、選手 i が j に勝つことを $i \rightarrow j$ と表わすことにして、 $i \rightarrow k$, $k \rightarrow j$ ならば i が j に勝ったものとし、 $i \leftarrow k$, $k \leftarrow j$ ならば j が i に勝ったものとする。また、 $i \rightarrow k$, $j \rightarrow k$ または $i \leftarrow k$, $j \leftarrow k$ のときは引き分けとして、同じことを勝負がつくまで続ける。

このとき、選手 i が j に勝つ確率と負ける確率の比は $p_{ik}p_{kj} : p_{ki}p_{jk}$ であるから、選手 i が j に勝つ確率は

$$\frac{p_{ik}p_{kj}}{p_{ik}p_{kj} + p_{ki}p_{jk}}$$

となる。

選手 i, j, k の間に「カモ・苦手」関係がないということは、 i と j が直接対戦したとき i が j に勝つ確率が、 k を介して i, j の勝負を決めるとき i が j に勝つ確率に等しいことを意味する。このとき

$$p_{ij} = \frac{p_{ik}p_{kj}}{p_{ik}p_{kj} + p_{ki}p_{jk}} \quad (4.5)$$

すなわち

$$\frac{p_{ij}}{p_{ji}} = \frac{p_{ik}p_{kj}}{p_{ki}p_{jk}} \quad (4.6)$$

という関係が成り立つ。

すべての i, j, k について式 (4.6) が成立するものとすれば

$$\pi_i = \frac{p_{im}}{p_{mi}} \quad (i = 1, 2, \dots, m-1), \quad \pi_m = 1 \quad (4.7)$$

とおくと、すべての $i, j (i \neq j)$ について

$$\frac{p_{ij}}{p_{ji}} = \frac{p_{im}p_{mj}}{p_{mi}p_{jm}} = \frac{\pi_i}{\pi_j} \quad (4.8)$$

が成立する。このように式 (4.4) が得られる。

また、式 (4.6) は

$$p_{ij}p_{jk}p_{ki} = p_{ji}p_{ik}p_{kj} \quad (4.9)$$

と書き直すことができる。式 (4.9) は、3 選手がそれぞれ 1 回ずつ勝負したとき、 $i \rightarrow j$, $j \rightarrow k$, $k \rightarrow i$ という形の「三すくみ」が起こる確率とが等しいことを意味している。もし本当に三すくみ的な状況があればこれらの確率の一方が他方より大きくなる。したがって、式 (4.9) はそのような状況がないことを意味することになる。

4.2 強さの推定

4.2.1 勝数の分布と尤度

m 選手の間で、何回か試合が行われたとき、その結果から各選手の強さ π_i を推定する問題を考える。選手 i, j の間の試合数を $n_{ij} (= n_{ji})$ とし、 i の j に対する勝ち数を確率変数

$$X_{ij} \quad (i \neq j; i, j = 1, 2, \dots, m)$$

で表す。引き分けはないものと仮定したので

$$X_{ij} + X_{ji} = n_{ij} \quad (i \neq j) \quad (4.10)$$

となる。

ここで、 X_{ij} が2項分布 $B(n_{ij}, p_{ij})$ に従うものとすれば

$$\Pr\{X_{ij} = x_{ij}\} = \frac{n_{ij}!}{x_{ij}!x_{ji}!} p_{ij}^{x_{ij}} p_{ji}^{x_{ji}} \quad (X_{ij} = 0, 1, \dots, n_{ij}) \quad (4.11)$$

となる。したがって、それらの同時確率は

$$\Pr\{X_{ij} = x_{ij}; i \neq j; i, j = 1, 2, \dots, m\} = \prod_{i < j} \prod \left(\frac{n_{ij}!}{x_{ij}!x_{ji}!} p_{ij}^{x_{ij}} p_{ji}^{x_{ji}} \right) \quad (4.12)$$

となる。

さらに、BT モデル (4.3) が成り立つときには、式 (4.12) は

$$\begin{aligned} \Pr\{X_{ij} = x_{ij}; i \neq j; i, j = 1, 2, \dots, m\} &= \prod_{i < j} \prod \left(\frac{n_{ij}!}{x_{ij}!x_{ji}!} \cdot \frac{\pi_i^{x_{ij}} \pi_j^{x_{ji}}}{(\pi_i + \pi_j)^{n_{ij}}} \right) \\ &= \prod_{i < j} \prod \left(\frac{n_{ij}!}{x_{ij}!x_{ji}!} \cdot \frac{1}{(\pi_i + \pi_j)^{n_{ij}}} \right) \cdot \prod_{i=1}^m \pi_i^{t_i} \end{aligned} \quad (4.13)$$

となる.

ここで

$$t_i = \sum_{j \neq i} x_{ij} \quad (4.14)$$

は選手*i*の総勝数である.

式 (4.13) を未知母数 $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_m)$ の関数とみなすとき, それを尤度 (関数) と呼ぶ. 未知母数 $\boldsymbol{\pi}$ に関係のない部分を const. とおき, 改めて確率変数を含む形に書くと, この場合の尤度は

$$L = \text{const.} \times \prod_{i < j} (\pi_i + \pi_j)^{-n_{ij}} \cdot \prod_{i=1}^m \pi_i^{T_i} \quad (4.15)$$

となる. つまり, この場合の尤度は個々の X_{ij} を明示的に含まず, 各選手の総勝数

$$T_i = \sum_{j \neq i} X_{ij} \quad (4.16)$$

だけの関数として表される. 言い換えれば, BT モデルのもとでは (T_1, T_2, \dots, T_m) が十分統計量となる. 逆に (T_1, T_2, \dots, T_m) が十分統計量となるのは BT モデルが成り立つ場合に限るということも証明できる. すなわち, この場合に各選手の強さを求めるためには, 試合数 n_{ij} のほかに各選手の総勝数 T_i を知ればよいということになる.

4.2.2 最尤法

強さ $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_m)$ の推定に最尤法を用いる. つまり, 式 (4.15) の尤度 L を最大にする $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_m)$ を求めて, $\boldsymbol{\pi}$ に対する推定量 (最尤推定量) とする.

式 (4.3) の p_{ij} の値はすべての π_i を定数倍しても変化しないから, π_i の値を一義的に定めるためには何らかの基準化が必要となる. そのために, k を適当な定数として

$$\sum_{i=1}^m \pi_i = k \quad (4.17)$$

とおく.

この場合の最尤方程式はラグランジュの未定乗数法によって求められる. すなわち, 式 (4.15) の対数 (対数尤度) を

$$l = \log L = \sum_{i=1}^m T_i \log \pi_i - \sum_{i < j} n_{ij} \log(\pi_i + \pi_j) + \text{const.} \quad (4.18)$$

とし, λ をラグランジュ乗数とすれば

$$\begin{aligned} \frac{\partial}{\partial \pi_i} \left\{ l - \lambda \left(\sum_{i=1}^m \pi_i - k \right) \right\} &= 0 \\ \frac{\partial}{\partial \lambda} \left\{ l - \lambda \left(\sum_{i=1}^m \pi_i - k \right) \right\} &= 0 \end{aligned}$$

より

$$\frac{T_i}{\hat{\pi}_i} - \sum_{j \neq i} \frac{n_{ij}}{\hat{\pi}_i + \hat{\pi}_j} - \lambda = 0 \quad (i = 1, 2, \dots, m) \quad (4.19)$$

$$\sum_{i=1}^m \hat{\pi}_i = k \quad (4.20)$$

が得られる.

ここで

$$\hat{p}_{ij} = \frac{\hat{\pi}_i}{\hat{\pi}_i + \hat{\pi}_j} \quad (i \neq j; i, j = 1, 2, \dots, m) \quad (4.21)$$

とおくと, 式 (4.19) から

$$T_i = \sum_{j \neq i} n_{ij} \hat{p}_{ij} + \lambda \hat{\pi}_i \quad (i = 1, 2, \dots, m) \quad (4.22)$$

となる. これをすべての i について加えれば

$$\begin{aligned} \sum_{i=1}^m T_i &= \sum_{i < j} \sum_{j < i} n_{ij} \\ &= \sum_{i < j} \sum_{j < i} n_{ij} (\hat{p}_{ij} + \hat{p}_{ji}) + \lambda \sum_i \hat{\pi}_i \\ &= \sum_{i < j} \sum_{j < i} n_{ij} + \lambda k \end{aligned} \quad (4.23)$$

となり, $\lambda = 0$ となることがわかる. したがって, 最尤方程式は

$$\sum_{j \neq i} n_{ij} \frac{\hat{\pi}_i}{\hat{\pi}_i + \hat{\pi}_j} = T_i \quad (i = 1, 2, \dots, m) \quad (4.24)$$

$$\sum_{i=1}^m \hat{\pi}_i = k \quad (4.25)$$

となる.

この方程式を解くには, 次のようにするのが簡単である. 式 (4.24) は

$$\sum_{j \neq i} \frac{n_{ij}}{\hat{\pi}_i + \hat{\pi}_j} = \frac{T_i}{\hat{\pi}_i} \quad (i = 1, 2, \dots, m) \quad (4.26)$$

と書くことができるので, $\boldsymbol{\pi}^{(0)} = (\pi_1^{(0)}, \pi_2^{(0)}, \dots, \pi_m^{(0)})$ を適当な 1 組の初期近似値として, この左辺の和に対する近似値

$$r_i^{(0)} = \sum_{j \neq i} \frac{n_{ij}}{\hat{\pi}_i^{(0)} + \hat{\pi}_j^{(0)}} \quad (i = 1, 2, \dots, m) \quad (4.27)$$

を計算する. そして

$$\hat{\pi}_i^{(1)} = \frac{T_i}{r_i^{(0)}} \quad (4.28)$$

とにおいて, 式 (4.25) を満たす新しい近似値 $\boldsymbol{\pi}^{(1)}$ を次式 (4.29) により求める.

$$\hat{\pi}_i^{(1)} = \frac{k\hat{\pi}_i^{(1)}}{\sum_{j=1}^m \hat{\pi}_j^{(1)}} \quad (i = 1, 2, \dots, m) \quad (4.29)$$

このことを $\hat{\pi}_i^{(k)}$ ($i = 1, 2, \dots, m$)の値が収束するまで繰り返す.

4.2.3 推定量に関する注意

$n_{ij} = n$, つまり, すべての選手が同じ回数だけ対戦する場合について考える. このとき, 式 (4.24) から

$$T_i - T_k = \frac{n(\hat{\pi}_i - \hat{\pi}_k)}{\hat{\pi}_i + \hat{\pi}_k} + \sum_{j \neq i, j \neq k} \frac{n(\hat{\pi}_i - \hat{\pi}_k)\hat{\pi}_j}{(\hat{\pi}_i + \hat{\pi}_j)(\hat{\pi}_k + \hat{\pi}_j)} \quad (4.30)$$

が得られる. この右辺は $(\hat{\pi}_i - \hat{\pi}_k) \times (\text{正の量})$ という形であるから, π_i の大きさの順序は勝数 T_i , あるいは勝率 $T_i/n(m-1)$ の順序と同じになる. したがってこの場合は, 勝数で順位を決めることは合理的であると考えられる.

これに対して n_{ij} が一定でないときには, 強さ $\hat{\pi}_i$ の順序は必ずしも勝率による順位と一致しない. このことは当然であり, 強い選手とたくさん対戦した選手は, たとえ勝率が高くなくても, 必ずしも弱いということにはならないはずである. しかしこの場合でも, どの相手に勝ってどの相手に負けたかということは問題にする必要がなく, それぞれの相手との試合数と総勝数だけで強さが定まることは 4.2.1 で述べたとおりである. 総勝数とそれぞれの相手との対戦数が同じならば強い相手に勝ったということは他方では弱い相手に負けたことを意味するので, 当の選手が強いことにはならないからである.

全勝や全敗の選手がいる場合を考える. 選手 1 が全勝だった場合, $x_{1j} = n_{1j}$ ($j = 2, 3, \dots, m$)であるから,

$$\sum_{j=2}^m n_{1j} = \sum_{j=2}^m x_{1j} = T_1 \quad (4.31)$$

となる. したがって, 式 (4.24) が $i = 1$ で解をもつためには, $n_{1j} > 0$ の選手, つまり選手 1 と対戦したすべての選手 j について

$$\frac{\hat{\pi}_1}{\hat{\pi}_1 + \hat{\pi}_j} = 1 \quad (4.32)$$

すなわち $\hat{\pi}_j = 0$ が成り立たなければならない. このとき, 式 (4.24) の第 j 式が解をもつためには, $T_j = 0$ となる必要がある. つまり, 全勝の選手 1 と対戦した選手がすべて全敗でなければ式 (4.24) は解をもたない. 特に, 選手 1 と対戦した選手同士が対戦していれば解は存在しない. したがって, 一般的には全勝の選手がいるときは各選手の強さは決定できないことになる. 一方, 選手 1 が全敗だったとすれば, $T_1 = 0$, したがって, 式 (4.24) から $\hat{\pi}_1 = 0$ となる. さらにこのとき, 式 (4.24) は

$$\sum_{j \neq i, j=1} n_{ij} \frac{\hat{\pi}_i}{\hat{\pi}_i + \hat{\pi}_j} = T_i - n_{1i} \quad (i = 2, \dots, m) \quad (4.32)$$

と書き直される。つまり、全敗の選手がいるときは、その選手の強さは0であり、それ以外の選手の強さの計算においては、全敗の選手に対する勝利は無視される。

4.3 卓球選手の強さの推定

4.3.1 対象選手

本研究では、世界選手権 2009 年, 2011 年, 2013 年の 3 年分の男子シングルの試合結果のデータを扱う。世界選手権の試合はいずれもリーグ戦ではなくトーナメント方式で行われるため同じ選手と対戦する機会が少ない。そのため、今回対象とする選手を、世界選手権 2009 年, 2011 年, 2013 年の 3 大会のうち 2 回以上ベスト 16 まで勝ち残った 13 選手として対戦表を作成する。表 4.1 に 2014 年 1 月現在の卓球選手の世界ランキングの昇順で対象選手と対戦成績を示す。0 の欄は今大会では直接対戦していないことを指す。

表 4.1 対象選手の対戦成績（世界ランキングの昇順）

	名前	1	2	3	4	5	6	7	8	9	10	11	12	13
1	馬 龍		0	0	5-12	4-1	4-2	0	4-1	0	0	0	0	0
2	許 昕	0		0-4	0	0	0	0	0	1-4	0	4-2	0	3-4
3	張 繼科	0	4-0		8-4	0	4-1	0	0	4-2	0	0	4-0	0
4	王 皓	12-5	0	4-8		0	0	0	0	4-0	4-0	0	0	0
5	オフチャロフ	1-4	0	0	0		2-4	0	0	0	0	0	0	0
6	ボル	2-4	0	1-4	0	4-2		0	0	0	0	0	0	4-1
7	サムソノフ	0	0	0	0	0	0		0	0	0	3-4	0	3-4
8	馬 琳	1-4	0	0	0	0	0	0		3-4	0	5-7	4-2	0
9	王 励勤	0	4-1	2-4	0-4	0	0	0	4-3		0	0	0	4-3
10	水谷 隼	0	0	0	0-4	0	0	0	0	0		0	0	0
11	松平 健太	0	2-4	0	0	0	0	4-3	7-5	0	0		0	0
12	朱 世赫	0	0	0-4	0	0	0	0	2-4	0	0	0		0
13	陳 杞	0	4-3	0	0	0	1-4	4-3	0	3-4	0	0	0	

4.3.2 強さの推定値と考察

$n_{ij}(=n_{ji})$ を選手 i, j の間の総セット数, T_i を選手 i の選手 j に対するセット取得数（勝数と呼ぶ）として、最尤方程式を解く。対戦成績と各選手の強さの推定値 $\hat{\pi}_i$ の降順に並びかえたものを表 4.2 に示す。 $\hat{\pi}_i$ の計算にあたっては、 $k = 1000$ として行った。

張繼科選手（2011年・2013年の金メダリスト）は勝数が高く敗数が低いので勝率も高く強さ $\hat{\pi}_i$ は 346.38 と圧倒的に強いことがわかる。王皓選手（2009年の金メダリスト）の勝数は張繼科選手と同じ 24 であるが、敗数が多いため強さ $\hat{\pi}_i$ は 247.74 となっている。

松平健太選手の勝率は 0.520 であり、勝率の順位でいうと第 3 位になるが強さ $\hat{\pi}_i$ では 25.73 とあまり高くないことがわかる。これは、馬龍選手と松平健太選手を比較するとわかるように、勝数と敗数の差が互いに 1 であるが、勝数が多い馬龍選手の方が強さ $\hat{\pi}_i$ が高くなり、勝数の少ない松平健太選手の方が強さ $\hat{\pi}_i$ の値は小さくなるからである。

敗数が 8 であるオフチャロフ選手、サムソノフ選手、朱世赫選手の 3 選手について比較する。オフチャロフ選手と朱世赫選手の勝数は 3 勝と 2 勝で、決して高くない。それに対してサムソノフ選手は 6 勝している。しかし、オフチャロフ選手の強さ $\hat{\pi}_i$ は 33.30,

サムソノフ選手の強さ $\hat{\pi}_i$ は23.54であり、勝数の少ないオフチャロフ選手の方が高いことがわかる。表4.1からオフチャロフ選手が対戦した相手を見ると、馬龍選手とボル選手であり、ともに強さ $\hat{\pi}_i$ が高い相手からセットを奪っている。サムソノフ選手が対戦した相手を見ると、松平健太選手と陳杞選手であり、 $\hat{\pi}_i$ はそれほど高くない相手と接戦になっているからであると考えられる。また、朱世赫選手が対戦した相手を見ると、張継科選手と馬琳選手であり、取得した2セットは $\hat{\pi}_i$ があまり高くない馬琳選手である。このことから、単純に多く勝てば強いということではなく、より強い選手に勝つ方が $\hat{\pi}_i$ が大きくなり強いということがいえ、互いに対戦回数異なる場合は対戦相手の強さも考慮されていることがわかる。水谷隼選手は勝数が0であり、4.2.3項より強さ $\hat{\pi}_i = 0$ となり、水谷隼選手と対戦した王皓選手の4セット分の勝利は無視される。

表 4.2 対戦成績と各選手の強さの推定値 $\hat{\pi}_i$

	名前	国名	勝数	敗数	勝率	世界ランク	強さ $\hat{\pi}_i$
3	張 継科	中国	24	7	0.774	3	346.38
4	王 皓	中国	24	13	0.649	4	247.78
1	馬 龍	中国	17	16	0.515	1	113.71
6	ボル	ドイツ	11	11	0.500	8	73.75
9	王 励勤	中国	14	15	0.483	12	49.59
5	オフチャロフ	ドイツ	3	8	0.273	6	33.30
13	陳 杞	中国	12	15	0.444	22	27.92
11	松平 健太	日本	13	12	0.520	16	25.73
2	許 昕	中国	8	14	0.364	2	23.90
8	馬 琳	中国	13	17	0.433	11	23.73
7	サムソノフ	ベラルーシ	6	8	0.428	10	23.54
12	朱 世赫	韓国	2	8	0.200	19	10.83
10	水谷 隼	日本	0	4	0.000	13	0.00

表 4.1 を強さ $\hat{\mu}_i$ の降順に並びかえたものを表 4.3 に示す. 表 4.3 の下三角は自分より強い相手と対戦することになり, 上三角は自分より弱い相手と対戦することになる. 松平健太 vs 許昕の試合をみると 2-4 で松平健太選手が格下である許昕選手に敗戦している. 強さ $\hat{\mu}_i$ はそれぞれ 25.73 と 23.90 と近いが, 「カモ・苦手」関係があるといえる. それ以外の対戦では, 「カモ・苦手」関係が見られず, 強さ $\hat{\mu}_i$ の降順であることがいえる.

表 4.3 対象選手と対戦成績 (強さ $\hat{\mu}_i$ の降順)

	名前	3	4	1	6	9	5	13	11	2	8	7	12	10
3	張 繼科		8-4	0	4-1	4-2	0	0	0	4-0	0	0	4-0	0
4	王 皓	4-8		12-5	0	4-0	0	0	0	0	0	0	0	4-0
1	馬 龍	0	5-12		4-2	0	4-1	0	0	0	4-1	0	0	0
6	ボル	1-4	0	2-4		0	4-2	4-1	0	0	0	0	0	0
9	王 励勤	2-4	0-4	0	0		0	4-3	0	4-1	4-3	0	0	0
5	オフチャロフ	0	0	1-4	2-4	0		0	0	0	0	0	0	0
13	陳 杞	0	0	0	1-4	3-4	0		0	4-3	0	4-3	0	0
11	松平 健太	0	0	0	0	0	0	0		2-4	7-5	4-3	0	0
2	許 昕	0-4	0	0	0	1-4	0	3-4	4-2		0	0	0	0
8	馬 琳	0	0	1-4	0	3-4	0	0	5-7	0		0	4-2	0
7	サムソノフ	0	0	0	0	0	0	3-4	3-4	0	0		0	0
12	朱 世赫	0-4	0	0	0	0	0	0	0	0	2-4	0		0
10	水谷 隼	0	0-4	0	0	0	0	0	0	0	0	0	0	

第5章 おわりに

5.1 まとめ

本研究では、マルコフモデルを用いて卓球の試合をモデル化した。両選手の打法、得失点を状態として定義し、推移確率行列を作成し、その推移確率に従って試合のシミュレーションを行った。シミュレーションで得られた結果と実際の試合のセット取得数、得点数、ラリー数を比較したところ、シミュレーションは実際の試合にほぼ等しい値を得ることができた。したがって、本研究で作成した卓球マルコフモデルは実際の試合をほぼ再現できたといえる。

また、シミュレーションの結果から各選手の試合構成と特徴を分析した。松平健太選手の特徴および改善点を簡単に述べる。打法の割合は全体的にバランスがよく、攻撃、守備の両方で苦手な打法が少ないことがわかった。その中でも、フリックやカウンタの割合が高く、守備面だけでなく積極的に攻撃することができる。サービスやツッツキによるもったいないミスを防ぐことで失点を減らし、フリックなどの台上技術を磨き、できるだけ先に攻撃することができれば得点に繋がる機会が増えると予想できる。そして、どのような相手であっても、自分の武器であるカウンタを打つことができれば、より自分にとって楽な試合展開になるのかもしれない。

次に、Bradley-Terry モデルを用いて世界選手権 3 年分の試合結果から選手の強さを合理的に評価することができた。今回は、互いに対戦回数が異なる場合の強さの推定値 $\hat{\mu}_i$ を比較することで、選手の強さの評価を行った。評価結果から、単純に多く勝てば強いということではなく、より強い選手に勝つ方が強さ $\hat{\mu}_i$ の値が大きくなることがいえ、対戦相手の強さも考慮されていることがわかった。

5.2 今後の課題

今後の課題を以下に示す。

- 今回、卓球マルコフモデルでは 2013 年 5 月の 2 試合のデータを扱ったが、データ数が少ないので、できるだけ多くの試合のデータをとる必要がある。
- さらに遡ったデータを用いて、選手の成長の過程をみる。
- 選手の強さの評価において、世界選手権だけでなく様々な大会の試合結果を集め、より精度の高い比較を行う。

謝辞

本研究を進めるにあたり，多くのご指導ご助言をいただいた中央大学工学部情報工学科の田口東教授に深く感謝いたします。また，多くのご助言とご協力をいただいた中央大学工学部情報工学科の南さつき助教，田口研究室の大学院生の方々，学部生の皆様に深く感謝いたします。

参考文献

- [1] 森村英典, 高橋幸雄, マルコフ解析, OR ライブラリー18, 日科技連出版社, 東京, 1979.
- [2] 竹内啓, 藤野和健, スポーツの数理科学, 共立出版社, 1998.
- [3] 長野光, “高校バレーボールにおける選手の評価とチーム力の分析”, 中央大学工学部情報工学科卒業研究論文, 2011.