# Markov Chain and Top Table Tennis Athletes Game Action Analysis

Yuanxin Zhang

December 15, 2021

**Abstract**

The purpose of this article is to (1) show that the average strokes in one rally in a table tennis match are around 4 to 5, (2) utilize the average 4 strokes in one rally to simulate 1000 rallies random walk to observe the most frequent game action used by the top 50 male table tennis players in the world.

Markov Chains, as one case in stochastic process, is utilized based on our purpose since it can move stepwise under defined conditions. In table tennis, we define ”condition” as ”game action”, ”stroke positions”, ”stroke directions”, or ”stroke technique”. In this article, we concentrate on game action. To quantify the entire process, the article utilizes a transition matrix(all the conditions of the transition possibilities between two correlated states construct the transition matrix.) Under the concept of the transition matrix, the specific game actions(Serve, Receive, Offense, Defense, Neutral, Control, Point, and Fault) are regarded as 8 states, and the transition between some certain two states are called transition possibilities. Therefore, the mathematical properties above constitute the key point of the modeling in this paper.

## 1 Introduction

Table tennis match consists of competitions between two persons(Women's single/Men's single) or at most 4 persons(Women's doubles/ Men's doubles/ mixed doubles). A match is played best 4 of 7 games. For each game, the first player to reach 11 points wins that game, however, a game must be won by at least a two-point margin. Each player serves two points in a row and then switches server. If in the game, the two players are tied at 10 to 10, they will serve one ball in turn. After each game, the players switch sides of the table. In the last game(i.e. 7th game), when a player scored 5 points, the players switch sides. The origin of table tennis can date back to England in the 1920s, while it was estimated that almost 300 million people play table tennis throughout the world in 1995, although there were no up-to-date statistics(Sklorz & Michaelis, 1995). Nowadays, table tennis has become a worldwide-popular sport. With the assistance of mathematical and statistical knowledge, every country is trying to utilize various strategies to conduct performance analysis and projection.

Different methodologies of strategies are utilized in different stages and different countries. The entire structure can be divided into two general analyses, which are theoretical performance analysis and practical performance analysis(Lames & McGarry, 2007). In the theoretical performance analysis, the researchers concentrate a lot on performance indices(three-phase-method), footwork analysis, momentum analysis(Double moving average), expert knowledge, and simulation approach(Markov-Chain Modeling). These methods can denote the general but optimal tactical behavior in a match; Practical performance analysis concentrates more on analyzing individual athletes to personalize training(Hohmann, Lames & Letzelter, 2002; Lames & McGarry, 2007). The structure of performance analysis is summarized by Michael Fuchs in Figure 1(Fuchs, Liu & Lanzoni, 2018)

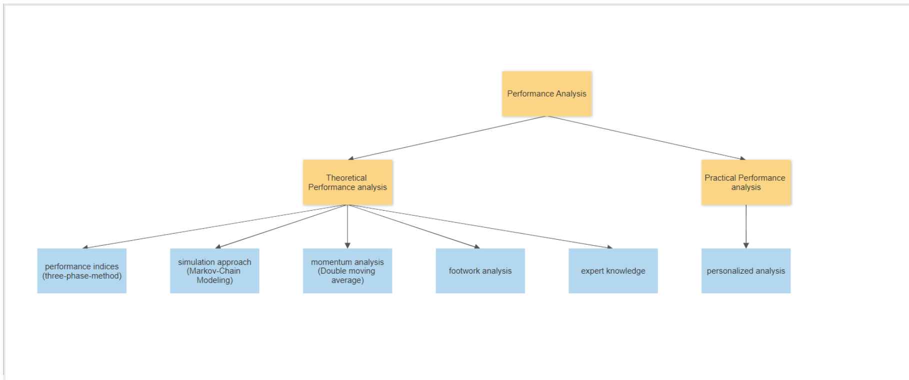



Figure 1: Overview of the general performance analysis approaches

This paper will only focus on the method of Markov Chain to employ the random walk process and estimate rough average strokes in the professional competition. Compared to other match analysis approaches, Markov Chain modeling emphasizes more on the dynamics between strokes in each rally.

# 2 Model Description

## 2.1 Model Introduction

If we want to build up an appropriate Markov chain to conduct the ensuing analysis, we need to figure out the match process. In this case, I will insert a structure mentioned in the article of Systematical Game Observation shown in Figure 2(Lames,1994)

In this structure, we regard a rally as one analysis unit, which a complete Markov Chain below is based on one rally. Meanwhile, we regard a stroke as one observation unit. In each stroke, they may use different tactical behaviors to gain one point. Here we describe the transition between two sequent states as "tactical behavior".

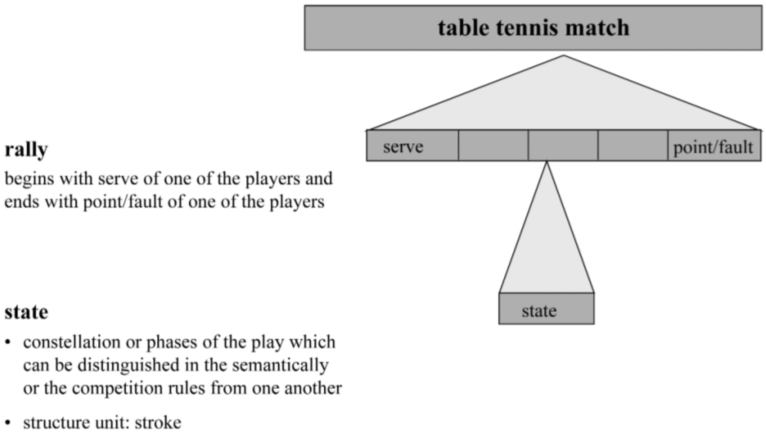



Figure 2: Table Tennis match structure by rallies and stroke

The tactical behavior can be divided into four game observation impacts, which are (1) game action, (2) stroke position, (3) stroke technique, and (4) stroke direction. All four impacts start from the server and end with "point" whomever A or B gets it. The reason to build up four game observation impacts is to consider different dynamics and strategies during the match so that the model will have a more concrete projection. However, due to the data limitation, this article will mainly focus on "game action" to investigate the most commonly used game action taken by the top male table tennis players.

Game Action is the most universal but main factor that influences a match's success, including Serve[1], Receive[2], Neutral[3], Offense[4], Defense[5], Control[6], and Point[7](Pfeiffer, Zhang & Hohmann, 2010). The property of game action is when we build up a random walk to simulate the game process, it will always begin from Serve or Receive and end at Point or Fault. The intermediate processes, such as neutral, control, defense, or offense are random.

[1]the first shot, done by the server. It begins with the ball being thrown up from the palm of a hand and stuck by the racket

[2]the return of a serve

[3]When both of the players are using attack technique(i.e. Topspin, Drive, Smash, Flip), it's regarded as 'neutral' from the side of observed player

[4]when the opponent utilizes the control technique or defense technique(i.e. Chop, Chopping short, Push, Block, Cut), but the observed player utilizes attack technique, it's regarded as 'offense' from the side of the observed player

[5]when the opponent strikes with attack technique, the observed player responses with defense technique. This is called 'Defense' on the side of observed player

[6]when both of the players use with defense or control technique, then it's called 'Control' from the side of the observed player

[7]the ended situation which one of the players earn one point in one rally

## 2.2 Markov Chain Model

In this paper, we concentrate on studying the transition probability (tactical behavior pattern) between different game actions and verify if the data we collected is rational from the statistical level. The transition probability is based on the transition between the eight discrete states of game action. As a Markov Chain, the transition has two properties. (1) We assume the current outcome is $X_n$ And consider the probability as

$$P(X_n|X_1 \cap X_2 \cap \cdots \cap X_{n-1}) = P(X_n|X_{n-1})$$

In other words, for a Markov chain, only the current trial's outcome will influence the next state, which simplifies the calculation and eliminate additional information interference. (2) the transition between different states is irrelevant with the chronological position in a match.

## 2.3 Data Collection

In this study, I utilize the data from the article "Performance diagnosis through mathematical simulation in table tennis in left and right handed shakehand and penholder players" due to the time limitation(Hohmann, Zhang & Koth, 2004). 152 matches of the world's top 50 male players were evaluated over a 3-year (1997-2000) period. Figure 3 shows the Markov Chain form of the 8 states and their transition probabilities.

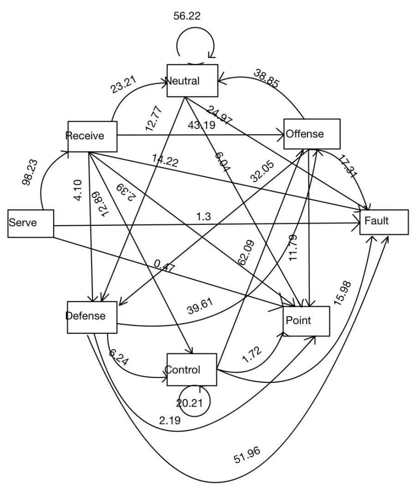



Figure 3: Markov Chain for Game Actions in 152 observed table tennis matches of the top 50 male world class players

However, to simplify this process, we construct Table 1 to show the data more directly. In Table 1, the matrix denotes that the point rate of each game action is smaller than the fault rate of each game action(Columns 7&8). Specifically, "Defense" has the highest fault rate (51.96%)and antepenultimate point rate (2.19%); ”Offense” has the highest point rate (11.79%) and a relatively low fault rate (17.31%).

| Game Action | Receive | Neutral | Offense | Defense | Control | Point | Fault |
|---|---|---|---|---|---|---|---|
| **Serve** | 98.23 | | | | | 0.47 | 1.30 |
| **Receive** | | 23.21 | 43.19 | 4.10 | 12.89 | 2.39 | 14.22 |
| **Neutral** | | 56.22 | | 12.77 | | 6.04 | 24.97 |
| **Offense** | | 38.85 | | 32.05 | | 11.79 | 17.31 |
| **Defense** | | | 39.61 | | 6.24 | 2.19 | 51.96 |
| **Control** | | | 62.09 | | 20.21 | 1.72 | 15.98 |

Table 1: Mean transition probability in 152 matches of the top 50 male athletes

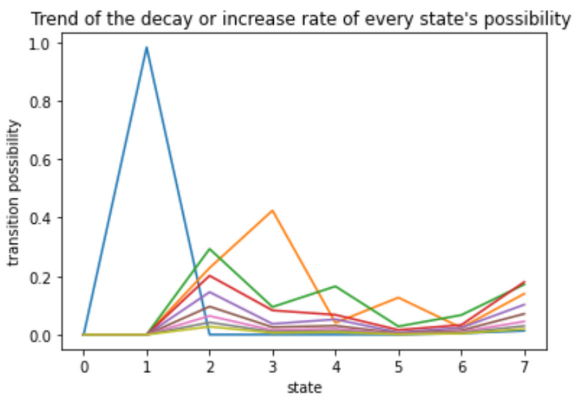
### 2.4 Simulation

Now we utilize the transition probability to simulate the random walk process. The random walk process can help us confirm the probability of each state after n strokes. Pursuant to the table tennis rule, the start state must be Serve or Receive, while the end state must be Point or Fault. Therefore, the initial probability matrix of each state should be $\pi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, or $\pi_0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. Here, to simplify the process, we solely analyze on the first case as the serve. In the Olympics website, the table tennis average strokes in one rally is around 4 to 5. By Markov Chain, when n = 1,2,3$\cdots$, the formula of $\pi_0 A^n = \pi_n$ is utilized to calculate the probability of every state after n strokes. We assume that the end probability matrix should be $\pi_n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & a & b \end{bmatrix}$, and $a+b=1$ after n strokes. If the data is rational, we should get n $\approx$ 4 or 5. The following codes denote this verification process.

**Here shows the code**

```
import numpy as np
import matplotlib.pyplot as plt
from numpy.linalg import matrix_power
# normalize the state
state = {
    0: "Serve",
    1: "Receive",
    2: "Neutral",
    3: "Offense",
    4: "Defense",
    5: "Control",
    6: "Point",
    7: "Fault"}
```

```
# here we build up the transition matrix shown in the table above
A = np.array([
[0.0000, 0.9823, 0.0000, 0.0000, 0.0000, 0.0000, 0.0047, 0.0130],
[0.0000, 0.0000, 0.2321, 0.4319, 0.0410, 0.1289, 0.0239, 0.1422],
[0.0000, 0.0000, 0.5622, 0.0000, 0.1277, 0.0000, 0.0604, 0.2497],
[0.0000, 0.0000, 0.3885, 0.0000, 0.3205, 0.0000, 0.1179, 0.1731],
[0.0000, 0.0000, 0.0000, 0.3961, 0.0000, 0.0624, 0.0219, 0.5196],
[0.0000, 0.0000, 0.0000, 0.6209, 0.0000, 0.2021, 0.0172, 0.1598],
[0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000],
[0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000]
])
# we gain the possibilities of each state after 1 or 2
or 3 strokes
pi_0 = [1,0,0,0,0,0,0,0]
for n in range(1,10):
    power_A =matrix_power(A, n)
    pi_n = pi_0@power_A
    plt.plot(col, pi_n)

>>>Output:
```




In figure 4, the vertical trend based on each state denotes that 0: serve and 1:receive decays to 0 which accords with our expectation; 6:point is slowly rising while 7: fault is rapidly increasing. which also adopt our assumption before. And the difference between 2-3 strokes and 4-5 strokes in some certain state indicates an obvious gap(which means the state possibilities decay or increase abruptly, and we neglect the ensuing change of strokes possibilities). Therefore, I conclude when n = 4 to 5, the possibility of every state have almost accorded with our assumption which is $\pi_n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & a & b \end{bmatrix}$ So the 4 to 5 average strokes in one rally denoted by Olympics website is rational.

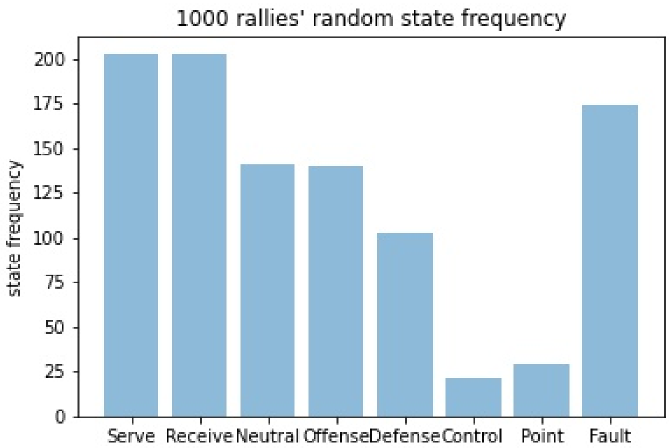
In the process above, we find out the average strokes in one rally is approximately 4 or 5. Under the 5 strokes case, we randomly go through 1000 rallies to figure out the most frequent game action taken by the top athletes.

**Here shows the code.**

(The codes above are still available here)

```
i = 0
times=np.zeros((1,8))
while i<=1000:
    n = 5 # According to the research, the average
    strokes in one rally is 4 to 5. Here we take 5.
    start_state = 0

    prev_state = start_state
    temp=np.zeros((1,8))
    temp[0,start_state] = temp[0,start_state]+ 1

    while n-1:
        if prev_state != 6 and prev_state!= 7:

            curr_state = np.random.choice
            ([0,1,2,3,4,5,6,7], p = A[prev_state])

            temp[0,curr_state] =  temp[0,curr_state] +1
            prev_state = curr_state
            n-=1
            if n ==1:
                if prev_state==6 or prev_state==7:
                    times+=temp

        else:
            break
    i+=1
print(times)
temp=times.tolist()
performance=temp[0]

## Here we graph 1000 rallies' game action distribution
objects = ('Serve', 'Receive', 'Neutral',
'Offense', 'Defense', 'Control', 'Point', 'Fault')
y_pos = np.arange(len(objects))
```

```
y_pos=y_pos.tolist()
plt.bar(y_pos, performance, align='center', alpha = 0.5)
plt.xticks(y_pos, objects)
plt.ylabel('state frequency')
plt.title('1000 rallies\' random state frequency')

plt.show()

>>>Output:
[[203. 203. 141. 140. 103.  22.  29. 174.]]
```




From the figure of "1000 rallies' random state frequency", top athletes are easy to receive the ball as the initial walk "serve" is defined. Among the game actions, Neutral and Offense are more frequent to utilize during one rally; while Control is the least used by top athletes. Meanwhile, the reason why "serve" only runs 203 out of 1000 rallies is I reduce the case of which the last stroke is not "point" or "fault". This raises the credibility of different game action distributions.

## 3 Results

The study analyzes one aspect of tactical behavior called game action. I first use the transition probability from the 152 matches of the top 50 male athletes to construct a Markov Chain. In the first simulation, the formula $\pi_0 A^n = \pi_n$ is utilized to test the possibility of every current state after n = 1,2,3,$\cdots$ And the result shows that when n = 4 or 5, there is a sudden decay or increase in the possibility of every state, so we neglect the ensuing calculation and conclude that the average strokes in one table tennis rally is 4 or 5.

In the second simulation, I utilize the first conclusion which employs a 5-step random walk and construct 1000 rallies to figure out the most common game actions used in the table tennis matches. The result denotes the top athletes as a serve almost never lose the received ball, and they are willing to act in a ’Neutral’ or ’Offense’ action instead of "Defense’ action; They seldom take ’Control’ action. Pursuant to the explanation of "Neutral" and "Offense", in other words, the top athletes are more willing to take aggressive technique to win the match.

# 4 Conclusions

This article first introduces the mainstream performance analysis approaches in table tennis and analyze the Markov Chain approach. Then We construct a table tennis match procedure to describe the how a rally begins and ends. In the process of describing four aspects of tactical behaviors, we identify a concentration on ”game action” and collect the relevant game action data. Finally, we utilize two simulations. First simulation verifies the rationality of the average 4-5 strokes in one rally; Second simulation utilizes the conclusion from the first simulation, as well as using the random walk to distinguish the most frequently used game action among the top 50 male athletes in table tennis. Eventually, we conclude that top male athletes are more willing to take ”Nuetral” or ”Offense” action, which can be inferred that they use aggressive techniques commonly to ensure the winning probability.

## References


[1] Sklorz, M., and R. Michaelis. *"Richtig Tischtennis [Correct Table Tennis]."* München: BLV Sportpraxis Top (1995).

[2] Lames, Martin, and Tim McGarry. *"On the search for reliable performance indicators in game sports."* International Journal of Performance Analysis in Sport 7.1 (2007): 62-79. DOI: 10.1080/24748668.2007.11868388

[3] Hohmann, A., Lames, M., and Letzelter, M. *Einfuehrung in die Trainingswissenschaft* [Introduction to the science of training]. Wiebelsheim: Limpert; 2002.

[4] Michael Fuchs, Ruizhi Liu, Ivan Malagoli Lanzoni, Goran Munivrana, Gunter Straub, Sho Tamaki, Kazuto Yoshida, Hui Zhang, and Martin Lames *Table tennis match analysis: a review*, Journal of Sports Sciences, 36:23, 2653-2662; 2018,

DOI: 10.1080/02640414.2018.1450073

[5] Lames, M., *Systematische Spielbeobachtung* [Systematical Game Observation], Philippka, Münster, 1994.

[6] Hohmann, Andreas et al. *"Performance diagnosis through mathematical simulation in table tennis in left and right handed shakehand and penholder players."*, 2004.