Lecture C4. Discrete Time Markov Chain 4

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr

SEOULTECH 서울과학기술대학교 데이터사이언스학과

I. Gambler's ruin probability

Gambler's ruin

- Suppose you have \$3(=x), and bet \$1 with winning probability $p = 18/38$ until your wealth becomes 0\$(=a) or your wealth becomes \$8(=b). What is the chance of you will leave Casino with \$8? (What is the chance that you will reach b before you reach a?)

$$
\mathrm{P} = \begin{array}{c} 0(lose) \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8(won) \end{array}
\begin{pmatrix}
1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
.53 & \cdot & .47 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & .53 & \cdot & .47 & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & .53 & \cdot & .47 & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & .53 & \cdot & .47 & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & .53 & \cdot & .47 & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & .53 & \cdot & .47 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & .53 & \cdot & .47 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1
\end{pmatrix}
$$

- Result of $a = 0, b = 8, p = 18/38$

$$
\mathrm{P}^{\infty} = \begin{array}{c} 0(lose) \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8(won) \end{array}
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.92 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .08 \\
.82 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .18 \\
.72 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .28 \\
.60 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .40 \\
.48 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .52 \\
.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .67 \\
.18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .82 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

- Result of $a = 0, b = 1000, p = 18/38, x = 100.$
- What is the quantity for $\mathrm{P}^{\infty}_{100\$\to win}$?

- Result of $a = 0, b = 1000, p = 19/38, x = 100.$
- What is the quantity for $\mathrm{P}^{\infty}_{100\$\to win}$?

- Result of $a = 0, b = 10 \times 100$\$, $p = 18/38, x = 1 \times 100$ \$ (bet 100\$ for each)
- What is the quantity for $\mathrm{P}^{\infty}_{1\times 100\$\to win}$?

II. Squash

Squash

- Racket sports (court number 5 in CRC)
- Rules
  - Two players, three or five games.
  - Only the server scores points.
  - The server, on winning a rally, scores a point
  - The receiver, on winning a rally, becomes the server.
  - The player who scores nine points wins the game

- Rules (cont'd)
  - Suppose A and B are playing for the first set and $8 : \overline{7}$ now. (A's score is 8, B's score is 7, and B is serving)
  - Suppose B wins this play so that it becomes $8 : \overline{8}$.
  - Because A got to 8 first, A can decide either
    - i) This set ends at 9
    - ii) This set ends at 10
- Questions
  - Suppose the chance of A winning a play is 0.6, then should A choose i) or ii)?

- Suppose A decides "i) This set ends at 9".
- DTMC
  - Transition diagram and matrix
  - Classification of states
  - What is the chance of A winning this game?

- Suppose A decides "ii) This set ends at 10".
- DTMC

$$
P = \begin{array}{c}
lose \\
8:\overline{8} \\
\overline{8}:8 \\
8:\overline{9} \\
\overline{8}:9 \\
9:\overline{8} \\
\overline{9}:8 \\
9:\overline{9} \\
\overline{9}:9 \\
win
\end{array}
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{pmatrix}
$$

- What is the chance of A winning this game?

- What if the chance of A winning a rally is not 0.6, but for general p?

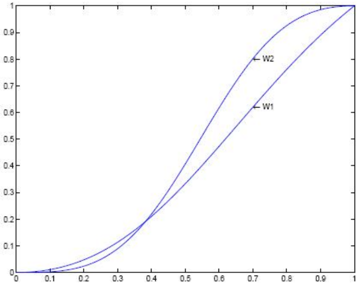


Figure 1: Probability of winning

- optimal decision
  - If $p \leq$, then choose i) ends at 9
  - Otherwise, choose ii) ends at 10
- Upon your decision, you are choosing one DTMC among the two different DTMC.

- Reference
  - Optimal Decision for the Squash Player
  - Jan Vecer, Columbia University, Department of Statistics
  - Journal of Chinese Statistical Association, 2004.
  - www.stat.columbia.edu/∼vecer/squash.ps

III. Tennis

Introduction

- DTMC for tennis game
- Used professional playing records (ATP tour 2011-2015)
- Is Markov chain a valid model for tennis game?
- What are the most important point in tennis?
- How do different court surfaces affect the model?
- How do serving ability of a player affect the model?

Dataset

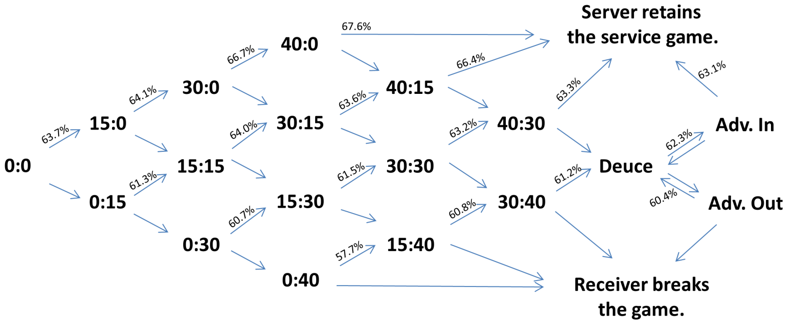
- Men's single matches in ATP tour from 2011 to 2015 are analyzed.
- The dataset includes 10,902 matches, 28,245 sets, 271,856 games, and 1,672,696 points.

Table 1. The structure of the dataset. A record for a single match is presented.

Variable	Value	Note
ID	6493708	
Date	04-Sep-14	
Tournament	Men's US Open	
Player 1	Novak Djokovic	Who serves first
Player 2	Andy Murray	
Winner	1	Player 1 or Player 2
Set 1	RSRRSSRR;SSRRSDRASRRSRR;RSSRRSSS; SRDRR;ARRSRSSS;RRSASS;SSRRSRRR;SRRSSS; RSSRRSSRAS;SRASS;SSRSRS;RSSRSA;S/DR/SR/ RR/S	A: ace S: server wins R: receiver wins D: double faults
Set 2	SRSSA;SSSS;ARRRR;SRSSS;SSAS;RRRSR; SRRRR;SRRDSSRR;ASRRSS;SSARS;SSSS; SRSRRSSRSRSS;S/RR/SR/RR/S	
Set 3	SSRSS;SSSS;RSSSS;RRSRR;SRSSRRRSRSSS; SRSSS;RSRSSS;SARRRSRR	
Set 4	SSRDRSSA;SSRSRRSRSS;SSSA;SRDSSRSS; SASDRRAS;SSSRS;SSSS;SSSS;SASRS;RRSRR	
Score	7–6(1) 6–7(1) 6–2 6–4	

DTMC diagram for a regular game

- The point winning probabilities for servers are marked.
- Q. Are they identical?
  - A. Not identical
- Q. Are they path-independent?
  - It depends··· (next page)



Test for path dependency

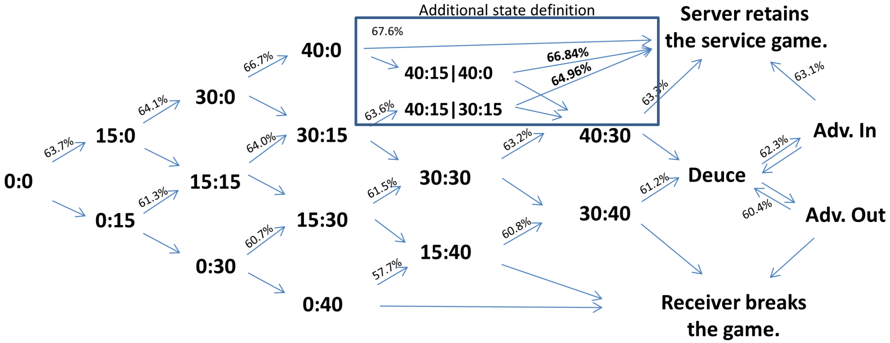
- Point winning probabilities are path-independent except for `40:15`.
- This singularity is found for all court types (Grass, Hard, and Clay). Why?

Table 3. Test of path dependency for server's point winning probabilities in regular games.

Current state	Last point won by	Server's winning prob. in the next point (%)	Number of observations	z-statistics
15:15	Server	63.97	60,429	0.12
	Receiver	63.94	62,222	
30:15	Server	63.63	78,443	0.74
	Receiver	63.4	36,960	
15:30	Server	61.86	23,157	1.46
	Receiver	61.29	44,208	
40:15	Server	66.84	73,346	5.33***
	Receiver	64.96	24,010	
30:30	Server	63.15	41,420	−0.53
	Receiver	63.32	42,057	
15:40	Server	61.38	8,672	1.31
	Receiver	60.59	25,945	
40:30	Server	63.35	52,787	0.06
	Receiver	63.33	32,736	
30:40	Server	61.38	21,043	0.55
	Receiver	61.14	30,690	
Deuce	Server	62.37	58,193	0.76
	Receiver	62.16	58,084	

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Possible remedy to build theoretically valid DTMC.



Additionally defined state prepares a legit Markov chain

Table 5. Comparison of a server's game winning probabilities between the model and the actual data.

	Model	Actual data			Comparison	
State	Server's winning prob. (%)	Server's winning prob. (%)	Number of obs.	Standard deviation (%)	Diff. of prob. (%)	z-statistics
0:0	78.91	78.91	271,856	0.08	0	0.01
0:15	62.12	62.21	98,603	0.15	−0.09	−0.57
15:0	88.46	88.41	173,253	0.08	0.05	0.59
0:30	39.72	40.12	38,174	0.25	−0.4	−1.61
15:15	76.28	76.26	122,651	0.12	0.02	0.18
30:0	95.29	95.17	111,031	0.06	0.12	1.81*
0:40	15.57	16.08	15,017	0.3	−0.51	−1.69*
15:30	55.37	55.23	67,365	0.19	0.14	0.75
30:15	88.06	88.16	115,403	0.1	−0.1	−1.02
40:0	98.9	98.78	74,071	0.04	0.12	2.99***
15:40	26.97	27.11	34,617	0.24	−0.14	−0.59
30:30	73.16	73.13	83,477	0.15	0.03	0.22
40:15|30:15	96.65	96.7	73,346	0.07	−0.05	−0.75
40:15|40:0	96.46	96.23	24,010	0.12	0.23	1.93*
30:40	44.37	44.48	51,733	0.22	−0.11	−0.51
40:30	89.9	89.83	85,523	0.1	0.07	0.67
Deuce	72.45	72.1	116,277	0.13	0.35	2.69***
Adv. Out	43.78	43.37	43,878	0.24	0.41	1.71*
Adv. In	89.83	89.5	72,403	0.11	0.33	2.86***

*p < .1; **p < .05; ***p < 0.01.

Reference

- Sim, M. & Choi, D.* (2019) The winning probability of a game and the importance of points in tennis matches. Research Quarterly for Exercise and Sport. (SSCI & SCIE)

IV. High-frequency financial data

What is OB pressure?

- OB pressure can be seen as a random walk in two-dimensional space.
- The current location of the particle indicates OB state.
- The particle moves to NEWS direction by an OB event.
- Price changes when it touches an axis.

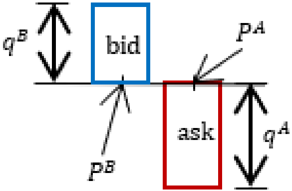


그림 1: Schematic diagram of order book

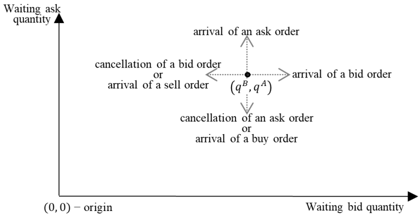


그림 2: Analogy of a 2D random walk

Hidden liquidity

- The presence of HL effectively shifts the both axis.
- That is, HL works as `buffer` of the axis.
- We incorporate the estimated HL and generate OB pressure matrix.

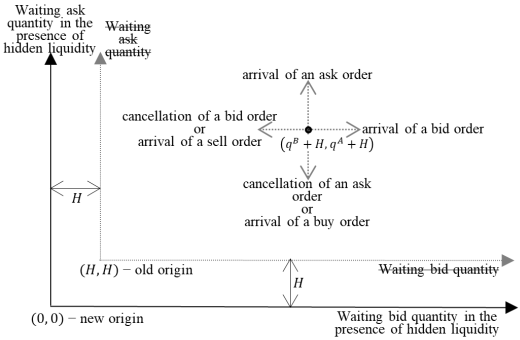


그림 3: HL buffers the axis

Reference

- Sim, M. K., & Deng, S. (2020). Estimation of level-I hidden liquidity using the dynamics of limit order-book. *Physica A: Statistical Mechanics and its Applications*, 540, 122703. (SCIE)

V. Stock price - binomial tree

Stock price - binomial tree

- Let X_n be the closing price of the stock at n-th day.
- Let $p = \mathbb{P}(X_{n+1} = x + 1 | X_n = x)$, and $1 - p = \mathbb{P}(X_{n+1} = x - 1 | X_n = x)$
- Consider future evolutions, starting with $X_0 = 100$.

- Consider an European call option which matures at day 5 with exercise price 101.
- (If you possess one unit of the call option, then at the day 5, you have a right to buy the stock at 101 dollars.)
- If $X_5 = 103$, then you can buy the stock at 101 and sell at 103. In this case, you earn 2 dollar.
- If $X_5 = 99$, then you still can buy the stock at 101. But you would not do it because you can buy a stock at 99 dollars. (Possessing call option is the "right" not the "obligation")
- i.e., the payoff of a call option is $(X_5 - 101)^+$

- What is the expected payoff for the option, when $p = 0.6$?

Summary

- The behavior of stock price's movement is often believed as random walk.
- The random walk leads to binomial evolution through the discrete time.
- At the far end from the beginning, the binomial expansions converges to normal distribution - In other words, applicable is the famous theorem of "normal approximation of binomial distribution."

```
"If I only had an hour to chop down a tree, I would spend the first 45 minutes sharpening my axe. -
A. Lincoln"
```