Random Walks with Memory Applied to Grand Slam Tennis Matches Modeling

Ing. Tomáš Kouřim

Institute of Information Theory and Automation, AS CR Prague

2.7.2019

Outline

1. Prepare mathematical model
2. Apply it on tennis modeling
3. Use it in real life betting

Outline

1. Prepare mathematical model
2. Apply it on tennis modeling
3. Use it in real life betting

Outline

1. Prepare mathematical model
2. Apply it on tennis modeling
3. Use it in real life betting

Random walk

Definition

A man starts from a point O and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $r + \delta r$ from his starting point, O.

[Karl Pearson: The problem of the random walk. (1905)]

"Drunken sailor?"

Random walk

Definition

A man starts from a point O and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $r + \delta r$ from his starting point, O.

[Karl Pearson: The problem of the random walk. (1905)]

"Drunken sailor?"

Random walk properties

- $n-$dimensional, on a matrix, graph, finite or infinite set
- Discrete random process
- Self avoiding, reinforced
- Brownian motion, polymer creation, games simulation, sports simulation

Random walks with varying transition probabilities

- Random walk with memory
- Memory coefficient $\lambda \in (0, 1)$ affecting the transition probabilities
- First step of the walk X_1 depends on an initial transition probability p_1
- Further steps depending on a transition probability p_t evolving as

$$X_{t-1} = 1 \rightarrow p_t = \lambda p_{t-1}$$

$$X_{t-1} = 0 \rightarrow p_t = 1 - \lambda(1 - p_{t-1})$$

- "Success punished"

Random walks with varying transition probabilities

- Random walk with memory
- Memory coefficient $\lambda \in (0, 1)$ affecting the transition probabilities
- First step of the walk X_1 depends on an initial transition probability p_1
- Further steps depending on a transition probability p_t evolving as

$$X_{t-1} = 1 \rightarrow p_t = \lambda p_{t-1}$$

$$X_{t-1} = 0 \rightarrow p_t = 1 - \lambda(1 - p_{t-1})$$

- "Success punished"

Alternative definitions

- “Success rewarded”
- Different coefficients for different events
- Generally n possible steps and m different coefficients λ affecting the transition probabilities
- Possible applications in
  - sports modeling
  - reliability and survival analysis
  - medical research
- Discrete alternative to random processes with memory

Alternative definitions

- "Success rewarded"

$$X_{t-1} = 1 \to p_t = 1 - \lambda(1 - p_{t-1})$$

$$X_{t-1} = 0 \to p_t = \lambda p_{t-1}$$

- Different coefficients for different events
- Generally n possible steps and m different coefficients λ affecting the transition probabilities
- Possible applications in
  - sports modeling
  - reliability and survival analysis
  - medical research
- Discrete alternative to random processes with memory

Alternative definitions

- "Success rewarded"
- Different coefficients for different events

$$X_{t-1} = 1 \rightarrow p_t = \lambda_1 p_{t-1}$$

$$X_{t-1} = 0 \rightarrow p_t = 1 - \lambda_2(1 - p_{t-1})$$

- Generally n possible steps and m different coefficients λ affecting the transition probabilities
- Possible applications in
  - sports modeling
  - reliability and survival analysis
  - medical research
- Discrete alternative to random processes with memory

Alternative definitions

- "Success rewarded"
- Different coefficients for different events
- Generally n possible steps and m different coefficients λ affecting the transition probabilities

$$p_t = f(p_{t-1}, X_{t-1}, \lambda_1, \dots, \lambda_m)$$

- Possible applications in
  - sports modeling
  - reliability and survival analysis
  - medical research
- Discrete alternative to random processes with memory

Alternative definitions

- "Success rewarded"
- Different coefficients for different events
- Generally n possible steps and m different coefficients λ affecting the transition probabilities
- Possible applications in
  - sports modeling
  - reliability and survival analysis
  - medical research
- Discrete alternative to random processes with memory

Alternative definitions

- "Success rewarded"
- Different coefficients for different events
- Generally n possible steps and m different coefficients λ affecting the transition probabilities
- Possible applications in
  - sports modeling
  - reliability and survival analysis
  - medical research
- Discrete alternative to random processes with memory

Application in tennis

- Matches, sets, games, points or strokes can be considered steps of a discrete random walk
- Data suggest that tennis evolves according to the random walk with varying transition probabilities, namely to the option with "Success rewarded"
- Sets considered as steps of the random walk
- Men Grand Slam tournaments - played as *best-of-five*

Application in tennis

- Matches, sets, games, points or strokes can be considered steps of a discrete random walk
- Data suggest that tennis evolves according to the random walk with varying transition probabilities, namely to the option with “Success rewarded”
- Sets considered as steps of the random walk
- Men Grand Slam tournaments - played as *best-of-five*

Application in tennis

- Matches, sets, games, points or strokes can be considered steps of a discrete random walk
- Data suggest that tennis evolves according to the random walk with varying transition probabilities, namely to the option with "Success rewarded"
- Sets considered as steps of the random walk
- Men Grand Slam tournaments - played as *best-of-five*

Data

- *www.oddsportal.com*
- Sport matches, results and odds
- Some statistics
- Ten seasons of tennis data since 2009
- 4255 matches, 432 players
- Novak Djokovic played 188 times

Model application

- Two parameters, p_1 and λ
- p_1 used to predict result of the first set
- Once set is finished, apply λ and previous p_i to obtain p_{i+1}

Odds to probability

- Pinnacle Sports bookmaker - market leading oddsmaker
- Closing first set odds as a base
- Bookmaker margin removal

$$g_j(a_j) = \frac{a_j \cdot (M-1)}{(M-1) + a_j \cdot (1 - \frac{1}{f_j(a_j)}) \cdot S(a)}$$

$$f_j(a_j) = a_j \cdot (1 - S(a))$$

$$S(a) = 1 - \sum_{i=1}^{M} \frac{1}{a_i}$$

Optimal λ

- Given set of matches
- Optimal λ found using maximal likelihood estimates

$$L = \prod_{i=1}^{N} (x_i p_i + (1 - x_i)(1 - p_i))$$

$$L_l = \sum_{i=1}^{N} log(x_i p_i + (1 - x_i)(1 - p_i))$$

Model evaluation

- One year training, next year testing $\rightarrow$ 9 splits
- λ optimized for each training set and applied to corresponding testing set
- Ljapunov CLT yields:

$$y = \frac{\sqrt{N}(\bar{x} - \hat{p})}{\hat{\sigma}} \sim \mathcal{N}(0, 1)$$

$$H_0 : \bar{p} = \hat{p} \quad vs. \quad H_1 : \bar{p} \neq \hat{p}$$

Model evaluation

- Bookmaker's favorite predicted
- CLT assumes independence → only 1 set per match can be in testing set → 36 testing sets
- Optimal testing
  - every subset
  - 95% of subsets within 95% confidence interval
- Real life testing
  - per tournament
  - grouped by initial probabilities p_0
  - 93.9% of subsets within 95% confidence interval

Model evaluation

- Bookmaker's favorite predicted
- CLT assumes independence → only 1 set per match can be in testing set → 36 testing sets
- Optimal testing
  - every subset
  - 95% of subsets within 95% confidence interval
- Real life testing
  - per tournament
  - grouped by initial probabilities p_0
  - 93.9% of subsets within 95% confidence interval

Model evaluation

- Bookmaker's favorite predicted
- CLT assumes independence → only 1 set per match can be in testing set → 36 testing sets
- Optimal testing
  - every subset
  - 95% of subsets within 95% confidence interval
- Real life testing
  - per tournament
  - grouped by initial probabilities p_0
  - 93.9% of subsets within 95% confidence interval

Results

- In-play betting tool
- Betting against Tipsport bookmaker
- Men Grand Slam tournaments - Wimbledon 2019
- If $p_i > \frac{1}{odd_i}$ then bet $p_i \cdot BU$
- Results available at *www.tomaskourim.com*

Results

- In-play betting tool
- Betting against Tipsport bookmaker
- Men Grand Slam tournaments - Wimbledon 2019
- If $p_i > \frac{1}{odd_i}$ then bet $p_i \cdot BU$
- Results available at *www.tomaskourim.com*

Next steps

- Model implementation
  - λ optimization
  - p_1 optimization
- Model improvement
  - Other versions of random walk with memory
  - Combination with other approaches
- Model testing
  - Model evaluation granularity
  - Performance on a larger dataset
  - Betting module for more bookmakers
  - Application of the model to *best-of-three* matches
- Application in other domains

Next steps

- Model implementation
  - λ optimization
  - p_1 optimization
- Model improvement
  - Other versions of random walk with memory
  - Combination with other approaches
- Model testing
  - Model evaluation granularity
  - Performance on a larger dataset
  - Betting module for more bookmakers
  - Application of the model to *best-of-three* matches
- Application in other domains

Next steps

- Model implementation
  - λ optimization
  - p_1 optimization
- Model improvement
  - Other versions of random walk with memory
  - Combination with other approaches
- Model testing
  - Model evaluation granularity
  - Performance on a larger dataset
  - Betting module for more bookmakers
  - Application of the model to *best-of-three* matches
- Application in other domains

Next steps

- Model implementation
  - λ optimization
  - p_1 optimization
- Model improvement
  - Other versions of random walk with memory
  - Combination with other approaches
- Model testing
  - Model evaluation granularity
  - Performance on a larger dataset
  - Betting module for more bookmakers
  - Application of the model to *best-of-three* matches
- Application in other domains

Thank you.

tomaskourim.com

Thank you.

tom@skourim.com